Probability of Price Crashes, Rational Speculative Bubbles, and the Cross-Section of Stock Returns

Jeewon Jang*
Jangkoo Kang†

Abstract

A recent paper by Conrad, Kapadia, and Xing (2014) shows that stocks with a high probability of extreme positive payoffs (jackpots) subsequently earn low returns. We find that stocks with a high probability of extreme *negative* returns (crashes) earn abnormally low average returns and that the cross-sectional return predictability of crash probability completely subsumes the jackpot effect. We also find that stocks with high crash probability clearly underperform, regardless of level of institutional ownership or variations in investor sentiment. Moreover, institutional demand for stocks with high crash probability increases until their prices arrive at the peak of overvaluation. Our evidence contradicts the presumption that sophisticated investors are always willing to trade against mispricing, and it suggests that the crash effect we find could arise partially from rational speculative bubbles, not entirely from sentiment-driven overpricing.

JEL classification: G11; G12

Keywords: Price crashes; Cross-section of stock returns; Anomalies; Institutional investors; Rational speculative bubbles

---

* Corresponding author: College of Business, Chosun University, 309 Pilmundaero, Dong-Gu, Gwangju 61452, Republic of Korea; Phone: +82-62-230-6822; E-mail: jeewonjang@chosun.ac.kr

† College of Business, Korea Advanced Institute of Science and Technology (KAIST), 85 Hoegiro, Dongdaemun-Gu, Seoul 02455, Republic of Korea; Phone: +82-2-958-3521; E-mail: jkkang@business.kaist.ac.kr

Probability of Price Crashes, Rational Speculative Bubbles, and the Cross-Section of Stock Returns

Abstract

A recent paper by Conrad, Kapadia, and Xing (2014) shows that stocks with a high probability of extreme positive payoffs (jackpots) subsequently earn low returns. We find that stocks with a high probability of extreme *negative* returns (crashes) earn abnormally low average returns and that the cross-sectional return predictability of crash probability completely subsumes the jackpot effect. We also find that stocks with high crash probability clearly underperform, regardless of level of institutional ownership or variations in investor sentiment. Moreover, institutional demand for stocks with high crash probability increases until their prices arrive at the peak of overvaluation. Our evidence contradicts the presumption that sophisticated investors are always willing to trade against mispricing, and it suggests that the crash effect we find could arise partially from rational speculative bubbles, not entirely from sentiment-driven overpricing.

JEL classification: G11; G12

Keywords: Price crashes; Cross-section of stock returns; Anomalies; Institutional investors; Rational speculative bubbles

1. Introduction

It is well established in the finance literature that various firm characteristics are able to predict stock returns in the cross-section. Conrad, Kapadia, and Xing (2014; hereafter CKX) recently document that stocks with a high predicted probability of extreme positive returns (jackpots) earn abnormally low returns on average. CKX provide evidence that individual investors with a preference for skewed, lottery-like payoffs bid up the prices of stocks with a high jackpot probability, which subsequently leads to their low returns. Moreover, the authors show that the jackpot effect is stronger in stocks with higher arbitrage costs, indicating that the high arbitrage costs prevent rational investors from arbitraging away the jackpot effect. Their findings are in line with recent empirical studies that document that a variety of cross-sectional anomalies yield abnormal profits from taking short positions in overpriced stocks and become stronger when limits to arbitrage enable the overpricing to persist for a while (Nagel, 2005; Campbell, Hilscher, and Szilagyi, 2008; Stambaugh, Yu, and Yuan, 2012, 2015).[1]

Inspired by CKX, this paper explores the cross-sectional relation between the predicted probability of extreme outcomes and future stock returns from a different standpoint. We examine whether the probability of extremely *low* returns is also related to the cross-section of stock returns, while CKX focus only on the effect of extremely high returns. We are interested in the event of extremely low returns (i.e., price crashes) for two reasons. First, an extensive literature documents that aggregate stock market returns exhibit negative skewness, or asymmetric volatility, that is, a higher volatility associated with negative returns (French, Schwert, and Stambaugh, 1987; Nelson, 1991; Campbell and Hentschel, 1992; Engle and Ng, 1993; Glosten, Jagannathan, and Runkle, 1993; Bekaert and Wu, 2000). More simply, examples of

[1] Nagel (2005) documents that the underperformance of stocks with a high market-to-book, analyst forecast dispersion, turnover, or volatility is most pronounced among stocks with low institutional ownership. Campbell, Hilscher, and Szilagyi (2008) find that the puzzling low returns on financially distressed firms are more pronounced for stocks with informational or arbitrage-related frictions. Stambaugh, Yu, and Yuan (2012; hereafter SYY) provide evidence that a broad set of cross-sectional anomalies can be partially explained by sentiment-driven overpricing. Stambaugh, Yu, Yuan (2015) document that the idiosyncratic volatility puzzle is more pronounced among overpriced stocks, especially for stocks that are less easily shorted.

the largest price changes in the stock market are usually declines rather than rises, including the crash of October 1987 and the crash of 2008 to 2009. This asymmetric distribution of stock returns leads us to investigate whether the possibility of price crashes, rare but more frequent than jackpots, significantly affects stock returns. In addition, since investors could be more concerned about negative shocks in stock prices than positive shocks, the effect of crashes, not just the effect of jackpots alone, is worthy of attention.

Second, recent empirical asset pricing studies document that various cross-sectional anomalies arise mainly from the underperformance of stocks with particular characteristics, indicating that overpricing is more prevalent than underpricing in the stock market (Nagel, 2005; SYY; Avramov et al., 2013).[2] These findings imply that firm characteristics that reflect the extent of overpricing could predict future returns in the cross-section. Our conjecture is that the predicted probability of price crashes measures the degree of overpricing. In other words, stocks with a higher probability of price crashes tend to be relatively more overpriced today because the probability would increase as the stock price approaches the peak of price bubbles, and thus they would earn lower returns in the next period.

This study therefore investigates the cross-sectional relation between the predicted probability of price crashes and future stock returns and explores the sources of the cross-sectional relation that differentiate it from other cross-sectional anomalies, particularly the jackpot effect of CKX. To predict the ex ante probability of price crashes for individual stocks, we employ a generalized logit model that extends the binary logit framework adopted by CKX. Specifically, we regard a firm's future extreme positive (jackpots) or extreme negative returns (crashes) as a ternary event, not defining each of the extreme events as a binary one, to prevent the predicted probability of crashes from being mixed up with that of jackpots. By building a model jointly predicting two mutually exclusive events, we can distinguish the cross-sectional return predictability of crash probability from that of jackpot probability. The estimation

---

[2] These findings are consistent with the arbitrage asymmetry noted by Stambaugh et al. (2015), which means that investors willing to buy an underpriced stock are reluctant or unable to sell short an overpriced stock due to short-sale constraints.

results of the generalized logit model show that a firm's high past returns, high volatility, high skewness, low age, few tangible assets, and high sales growth predict high probabilities of the firm's future extreme returns in both directions, but the sensitivity of each predictor is larger in magnitude for future crashes than for future jackpots.

By constructing decile portfolios sorted on the predicted probability of price crashes, we find that stocks with higher crash probability subsequently earn lower average returns. The long–short strategy of going long the lowest and short the highest crash probability decile yields a Fama–French (1993) three-factor alpha of 0.79% per month, with a *t*-statistic of 3.60, which is largely due to the substantial underperformance of the highest decile. The highest crash probability portfolio consists of small and growth stocks with high prior returns in recent months, low liquidity, and high turnover. Although stocks with high crash probability share very similar characteristics with stocks with high jackpot probability, the negative cross-sectional relation between crash probability and future returns remains strong, even if jackpot probability is controlled for, indicating that the crash probability effect is clearly distinct from the jackpot effect. Moreover, in both portfolio- and firm-level analyses, we find that the cross-sectional return predictability of crash probability still holds after controlling for a large set of variables, including the idiosyncratic volatility of Ang et al. (2006), the maximum daily return of Bali, Cakici, and Whitelaw (2011), the analyst forecast dispersion of Diether, Malloy, and Scherbina (2002), and firm characteristics such as size, the book-to-market ratio, past returns, liquidity, and turnover, all of which are closely related to both crash probability and stock returns in the cross-section.

Since the abnormally low returns on stocks with high crash probability cannot be understood from the efficient market view, which predicts that the low returns will be quickly exploited by rational arbitrageurs, we investigate whether the underperformance of stocks with high crash probability becomes stronger when limits to arbitrage become higher. Consistent with previous empirical studies on anomalies, as well as theories on limited arbitrage, the crash probability effect is particularly pronounced among small, low-priced, illiquid, and low analyst coverage stocks, which indicates that the low average returns of stocks with high crash probability result from persistent overpricing and high arbitrage costs. On the other hand, we find evidence that stocks with high crash probability are overpriced regardless of the firms' institutional ownership, which clearly differentiates the crash probability effect from previously documented anomalies, including the jackpot effect of CKX.

Our finding that the overpricing of stocks with high crash probability is not arbitraged away even among stocks owned largely by institutions contradicts the standard limited arbitrage view, which predicts that overpricing is inversely related to institutional ownership because sophisticated investors always trade against mispricing unless arbitrage is limited by frictions such as noise trader risk or short-sale constraints.[3] Alternatively, reasoning along the line of theories on rational speculation can help to understand our empirical finding. For example, DeLong et al. (1990b) document that rational traders could bid up the prices of securities above fundamental values when they anticipate positive feedback traders buying the securities at even higher prices in the next period. Abreu and Brunnermeier (2002, 2003) document that rational investors optimally choose to ride a price bubble even if they are aware of the overvaluation, when coordinated selling pressure is necessary to burst the bubble but other arbitrageurs are not yet likely to attack the bubble. These theories consistently imply that trading against mispricing is not always the optimal strategy for rational investors and these investors even contribute to price deviations from fundamentals.[4] The theories provide an explanation of why overpricing is not arbitraged away even among stocks mainly held by sophisticated institutions.

We find further evidence that supports the rational speculation view but refutes the standard limited arbitrage view. First, we find that the overpricing of stocks with high crash probability is prevalent irrespective of market-wide sentiment, particularly for stocks owned largely by institutions. This indicates that the crash probability effect is not entirely driven by investor sentiment, which is clearly different

---

[3] Standard theories on limited arbitrage include those of Miller (1977), Harrison and Kreps (1978), Delong et al. (1990a), Dow and Gorton (1994), and Shleifer and Vishny (1997).

[4] Empirical documentation also supports this line of theories. Brunnermeier and Nagel (2004) document that hedge funds invested heavily in technology stocks during the technology bubble, not exerting any correcting force on their prices, although this does not seem to be the result of unawareness of the bubble. Griffin et al. (2011) provide evidence that institutions bought technology stocks more aggressively than individuals did during the technology bubble, which shows that rational investors sometimes fail to trade against mispricing.

from other cross-sectional anomalies examined by SYY. Second, by examining changes in institutional holdings prior to forming decile portfolios, we find that stocks with higher crash probability have been bought more heavily by institutions during a recent six-quarter period. Moreover, institutional holdings for a firm increase continuously up until the firm enters into the highest decile sorted by crash probability while they slow down afterward. This indicates that institutional investors have a tendency to buy an overvalued security until its price reaches the peak of the bubble. Our findings are consistent with the rational speculation view that predicts that sophisticated institutions may not trade against overpricing but could drive asset prices farther away from fundamentals. In addition, our results suggest that the cross-sectional return predictability of crash probability may be at least partially due to rational speculative bubbles driven by institutions and not entirely to overpricing driven by the sentiment of individual investors.

Finally, we compare the cross-sectional return predictability of crash probability to that of related anomaly variables. When constructing the double-sorted portfolios based on crash and jackpot probabilities, the negative cross-sectional relation between jackpot probability and portfolio returns is completely eliminated, while the relation between crash probability and portfolio returns is unaffected. Moreover, we find that stocks with high jackpot probability are overpriced only during periods of high investor sentiment, which indicates that the jackpot effect of CKX is due to sentiment-driven overpricing, unlike the crash probability effect. These different sources of overpricing for the two related probabilities offer a clue as to why the jackpot effect is completely subsumed by the crash probability effect.

In comparison with the failure probability of Campbell et al. (2008; hereafter CHS), the negative cross-sectional relation between crash probability and future returns becomes insignificant when controlling for failure probability at the portfolio level. Nevertheless, we provide evidence that crash probability is still significantly related to the cross-section of stock returns when orthogonalized to failure probability, implying that it contains additional information about future returns not reflected in failure probability. The sharpest contrast between the two probabilities is that stocks in the highest decile sorted by crash probability experience large price increases before the entry into the highest decile and then their prices plummet after portfolio formation, whereas the prices of stocks in the highest decile sorted by failure probability decline steadily before and after the entry into the highest decile. The sudden reversal of the sharply increasing returns of stocks with high crash probability is consistent with growing price bubbles followed by bursting of them. This evidence suggests that the cross-sectional return predictability of crash probability could be attributed to the ability to time the peaks of price bubbles, as well as the ability to pick relatively overpriced stocks.

Our study is related to an extensive literature on the relation between firm characteristics and the cross-section of stock returns. Although standard asset pricing theories state that riskier assets should command higher expected returns, there is considerable evidence that the cross-sectional pattern of stock returns is related to various firm characteristics, such as firm size (Banz, 1981), the book-to-market ratio (Stattman, 1980), past returns (Jegadeesh and Titman, 1993), idiosyncratic volatility (Ang et al., 2006), failure probability (CHS), maximum daily returns (Bali et al., 2011), and jackpot probability (CKX). Several empirical studies document that the cross-sectional relations, unexplained by the efficient market view, are strong only when individuals own a large fraction of shares (Nagel, 2005; CHS; CKX; Stambaugh et al., 2015). In addition, SYY provide comprehensive evidence that 11 cross-sectional anomalies that survive adjustments for risk exposure are caused by sentiment-driven overpricing. Our empirical findings on the effect of crash probability clearly differ from theirs in that overpricing is also present among stocks owned largely by institutions and not associated with variations in investor sentiment. Our evidence casts doubt on the common presumption underlying both the efficient market view and the standard limited arbitrage view that sophisticated traders always stabilize stock prices as arbitrageurs by trading against mispricing whenever possible.

On the other hand, Edelen, Ince, and Kadlec (2016) examine seven well-known stock return anomalies and document that institutional investors have a strong propensity to buy stocks classified as overvalued, which means that they trade contrary to anomaly prescriptions. Our empirical evidence on the crash probability effect is consistent with their results. However, our study differs from theirs in that we suggest an alternative explanation based on rational speculation, whereas they suggest that agencyinduced preferences drive portfolio managers to seek stock characteristics associated with poor long-run performance. Although our evidence does not rule out their explanation, we argue that our empirical findings can be better understood in the context of the rational speculation view that presumes that sophisticated traders can destabilize stock prices by riding price bubbles even if arbitrage is not limited (DeLong et al., 1990b; Abreu and Brunnermeier, 2002, 2003).

The remainder of this paper proceeds as follows. Section 2 describes the data, defines the event of price crashes, and presents the empirical framework for predicting the ex ante probability of price crashes. Section 3 shows empirical results on the relation between the predicted probability of price crashes and the cross-section of stock returns. Section 4 explores the sources behind the cross-sectional relation and discusses its distinctive features. Section 5 compares the cross-sectional return predictability of crash probability to other related anomalies. Section 6 concludes the paper.

2. Generalized Logit Model of Price Crashes

2.1 Data

We obtain the monthly and daily stock files from the Center for Research in Security Prices (CRSP) database over the period January 1926 through December 2015. Our sample includes NYSE, AMEX, and NASDAQ ordinary common stocks with CRSP share codes 10 and 11. Stocks with an end-of-month price below $5 per share are excluded from the sample. The monthly and daily stock returns are adjusted for any delisting returns provided in the CRSP database. Accounting variables to calculate various firm characteristics are obtained from annual Compustat data. We match the annual accounting data for the end of year $t-1$ with the monthly CRSP data for June of year t to May of year $t+1$. Since the variables calculated using the Compustat data start in 1951, our analysis is restricted to the period June 1952 to December 2015.

We obtain data on institutional ownership from the Thomson Reuters Institutional (13f) Holdings database. Since data on the quarterly institutional holdings start in the first quarter of 1980, the analysis based on institutional holdings data is confined to the period after 1980. We also use the unadjusted file from the Institutional Brokers Estimates Systems (I/B/E/S) database, which provides data on analyst earnings forecasts starting in January 1976.

2.2 Definition of price crashes

We call the event that a stock price bubble bursts a price crash. In other words, we refer to the low-probability event of large negative returns as a price crash throughout the paper. To define the event of price crashes specifically, we follow CKX in defining jackpot returns as extreme positive returns. A price crash is defined as the event of log returns less than −70% over the next 12 months. The cutoff of −70% corresponds roughly to a capital loss of 50% if the dividend yield is negligible. Symmetrically, we define a jackpot as achieving log returns greater than 70% over the next 12 months, which corresponds roughly to a capital gain of 100%.[5]

The difference between our study and that of CKX is that we mainly consider the left tail, not the right tail, of a return distribution, even though both studies investigate the effect of expected extreme returns on subsequent stock returns. The following reasons motivate us to examine the event of extremely low returns. First, a stylized fact is that the largest movements in the stock market are often declines rather than rises. For example, during the period June 1952 to December 2015, seven out of the 10 largest daily returns in the CRSP value-weighted index were negative.[6] This asymmetry suggests that the rare event of extremely negative returns, which is relatively more frequent than jackpots, could affect stock prices significantly but differently, because investors might care more about negative shocks than positive

[5] Although CKX define jackpot returns as annual log returns greater than 100%, they show that the jackpot effect they find is robust to different cutoffs.

[6] Of the ten largest daily movements, three occurred in October 1987 and the remaining seven occurred in October 2008, both of which correspond to notorious periods of stock market declines. Moreover, the three positive returns were due to the recovery following large negative shocks, rather than independent price increases.

shocks. Second, it has been documented that the abnormal profits based on various stock market anomalies come from taking short positions in overpriced stocks (Nagel, 2005; SYY; Avramov et al., 2013; Stambaugh et al., 2015). Generally, due to short-sale constraints, it is more difficult to exploit arbitrage profits by selling overpriced stocks than by buying underpriced stocks, which leads to the prevalence of overpricing in the stock market. This implies that firm characteristics related to overpricing should have cross-sectional return predictability. We argue that stocks with high probability of price crashes are the most overpriced currently and should subsequently have low returns, since the probability of price crashes would become higher as the stock price approaches to the peak of the bubble. We therefore investigate the cross-sectional return predictability of the probability of price crashes in this paper. Moreover, we investigate whether the return predictability of the crash probability is distinguished from the jackpot effect documented by CKX.

2.3 Prediction of price crashes with a generalized logit model

We employ a generalized logit model to estimate the ex ante probability of extreme negative returns, or crash probability, extending the binary logit model of CKX. We basically follow the method of CKX in predicting price crashes but regard a firm's realization of extreme positive or negative returns as a ternary event, rather than considering each extreme event as a binary one. Since highly volatile stocks are likely to have high probabilities of both crashes and jackpots, the two probabilities may be positively correlated, which makes it difficult to identify the effects of crashes and jackpots separately. This difficulty would be compounded if we predicted two mutually exclusive events as if they are independent binary events. Since one of our goals is to distinguish the effect of crash probability on stock returns from the jackpot effect, we define price crashes and jackpots as exclusive but mutually dependent and jointly predict the probabilities of the two events.

Specifically, we model the probabilities of crashes and jackpots over the next 12 months with the following distribution:

$$\Pr_t(Y_{i,t,t+12} = -1) = \frac{\exp(\alpha_{-1} + \beta_{-1} X_{i,t})}{1 + \exp(\alpha_{-1} + \beta_{-1} X_{i,t}) + \exp(\alpha_1 + \beta_1 X_{i,t})} \ ,$$
$$\Pr_t(Y_{i,t,t+12} = 1) = \frac{\exp(\alpha_1 + \beta_1 X_{i,t})}{1 + \exp(\alpha_{-1} + \beta_{-1} X_{i,t}) + \exp(\alpha_1 + \beta_1 X_{i,t})} \ , \tag{1}$$

where $Y_{i,t,t+12}$ is a ternary variable that equals −1 if firm i's log return during months $t + 1$ to $t + 12$ is less than −70%, one if the same return is greater than 70%, and zero otherwise. $X_{i,t}$ denotes a vector of explanatory variables known at the end of month t. Equation (1) indicates that an increase in the value of $\alpha_{-1} + \beta_{-1}X_{i,t}$ or $\alpha_1 + \beta_1 X_{i,t}$ predicts a higher probability of a crash or jackpot, respectively, over the next 12 months.

We follow CKX to choose explanatory variables to predict extreme positive and extreme negative returns. They include past return (*RET12*), total volatility (*TVOL*), total skewness (*TSKEW*), firm size (*SIZE*), detrended turnover (*DTURN*), firm age (*AGE*), tangible assets (*TANG*), and sales growth (*SALESG*). The first five variables are employed following previous empirical studies on skewness. Chen, Hong, and Stein (2001) forecast skewness based on cross-sectional regression specifications and find that an increase in trading volume and positive past returns predict negative skewness. Boyer, Mitton, and Vorkink (2010) document that firm characteristics such as firm size and idiosyncratic volatility are also important predictors of future idiosyncratic skewness. In addition, CKX add three new variables to predict future jackpot returns and find that young and rapidly growing firms with fewer tangible assets are likely to experience extremely high returns. Based on these findings, we use the same variables to predict crashes, since we assume that the variables related to skewness and extreme outcomes could forecast extreme returns in both directions with either the same or opposite signs. Annual accounting data such as tangible assets and sales growth at year-end are matched with monthly observations from June of the next year to ensure all information is observable at the beginning of the 12-month period over which a crash or jackpot is measured. Further details on these variables are provided in the Appendix.

Table 1 shows the in-sample estimation results of our generalized logit model. The coefficients on all the predictors are statistically significant at the 5% significance level, with the exception of *DTURN* as a predictor of crashes. The signs of the estimated parameters for crashes and jackpots are identical for all the variables except *SIZE*, which implies that stocks with high crash probability also tend to have high jackpot probability. The result indicates that stocks with higher past returns, higher volatility, higher skewness, lower detrended turnover, lower age, fewer tangible assets, and higher sales growth tend to have higher probabilities of future extreme returns in both directions. For firm size (*SIZE*), the estimated coefficients for crashes and jackpots have opposite signs to each other. Price crashes are more likely to occur as firm size increases, whereas jackpots become more probable as firm size decreases. One notable thing is that, for most of the explanatory variables, the coefficient for crash probability is larger in magnitude than that for jackpot probability, which suggests that those variables are predictors of crashes rather than of jackpots. We also report the changes in odds ratios for a one standard deviation change in each variable, where the odds ratio of each extreme event (crashes or jackpots) is defined as the probability of the event divided by the probability of a non-extreme event. The result shows that total volatility (*TVOL*) has the largest impact on the odds ratios of both crashes and jackpots: A one standard deviation increase in *TVOL* increases the odds ratio of crashes by 85.5% and that of jackpots by 61.8%. For crashes, the next most influential variables are *AGE*, *TANG*, and *SALESG*. For jackpots, *SIZE* and *AGE* are relatively important predictors, which is consistent with the results of CKX. The pseudo-R^2 value of our generalized logit model, proposed by Nagelkerke (1991), is 12.8%.

3. Probability of Price Crashes and the Cross-Section of Stock Returns

3.1 Portfolios sorted on the predicted probability of crashes

In this section, we examine the cross-sectional relation between the probability of price crashes and subsequent returns. In particular, we construct decile portfolios sorted on the predicted probability of price crashes and observe the returns on each portfolio realized in the subsequent month. To avoid look-ahead bias, we re-estimate our generalized logit model using only historically available data for every month, recursively, over expanding estimation windows, starting from June 1952. We then calculate the out-of-sample predicted probability of price crashes in month t with a set of estimated parameters from the window ending in month $t - 12$ to ensure that the out-of-sample predictors are not observed before the period over which future extreme returns are measured. We call the out-of-sample predicted probability the crash probability (*CRASHP*). Likewise, we calculate the out-of-sample predicted jackpot probability (*JACKPOTP*). Based on *CRASHP*, we form decile portfolios at the end of each month t and calculate the monthly returns on each decile realized in month $t + 2$. We allow one month of waiting between the portfolio ranking and holding periods to eliminate the effect of short-term reversals. To make sure the crash probabilities are reliably predicted, we use the predicted series from November 1971 and thus the portfolio returns begin in January 1972.

Table 2 shows the mean monthly returns and risk-adjusted returns on decile portfolios sorted by crash probability. Panel A reports the value-weighted portfolio returns and Panel B reports the equal-weighted portfolio returns. In Panel A, the mean value-weighted returns from decile 1 to decile 8 do not show a monotonic pattern. However, portfolio returns decrease sharply from decile 8 to decile 10, with the lowest return of 0.36% per month in decile 10. This cross-sectional pattern is commonly observed in previous studies (Diether et al., 2002; Ang et al., 2006; Bali et al., 2011; SYY; CKX), indicating that stocks in the highest decile are likely to be considerably overpriced. The return difference between decile 10 and decile 1 is −0.64% per month, with a t-statistic of −1.68. To control for conventional risk factors, we also report CAPM alphas, Fama–French (1993) three-factor alphas, and Carhart (1997) four-factor alphas.[7] The alphas on the zero-cost portfolio buying the highest and selling the lowest *CRASHP* decile are −1.16% (t-statistic = −3.57) for the CAPM, −0.79% (t-statistic = −3.60) for the three-factor model, and −0.50% (t-statistic = −2.55) for the four-factor model, all statistically significant. These results show that the cross-sectional difference in returns between the highest and lowest *CRASHP* portfolios cannot be explained by risk factors, largely due to abnormally low returns on the highest *CRASHP* decile portfolio. The equal-

---

[7] We obtain data on the monthly factor portfolio returns for the Fama–French three-factor model and the Carhart four-factor model from Kenneth French's website, at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

weighted portfolio returns in Panel B exhibit very similar but stronger patterns, confirming the finding that stocks with high crash probability subsequently earn lower returns.

To investigate the economic significance of the cross-sectional return predictability of *CRASHP*, we examine the cumulative profits of buy-and-hold trading strategies for each decile sorted on *CRASHP* during the period January 1972 to December 2015. Specifically, we calculate the cumulative values of the value-weighted decile portfolios when a dollar is invested in each portfolio at the beginning of the period and held through the period with monthly rebalancing based on one-month-lagged *CRASHP*. The result is shown in Figure 1. At the end of December 2015, the final value of the investment in the lowest *CRASHP* decile becomes $127.37, which is approximately 1.8 times the value of the CRSP value-weighted index, $69.07. In contrast, the same strategy investing in the highest *CRASHP* decile yields $0.53, which is 1/130th of the passive benchmark. This result shows that the highest *CRASHP* decile significantly underperforms, implying that stocks with high crash probability are overpriced.

3.2 Crash probability and firm characteristics

In the previous section, we find that stocks with a higher predicted probability of crashes earn significantly lower returns, on average, which is not accounted for by risk factors. Since the literature documents that various firm characteristics are related to the cross-section of stock returns, our finding of the cross-sectional return predictability of crash probability could be due to other variables closely linked to crash probability. In this section, we examine various characteristics of portfolios sorted on crash probability and investigate whether controlling for these characteristics affects the cross-sectional relation between crash probability and stock returns.

Table 3 presents firm characteristics and other related variables for decile portfolios sorted by *CRASHP*. These include the jackpot probability (*JACKPOTP*), the failure probability (*FAILP*) of CHS, idiosyncratic volatility (*IVOL*), idiosyncratic skewness (*ISKEW*), the maximum (*MAX*) and minimum (*MIN*) daily returns in month *t* of Bali et al. (2011), the price per share (*PRC*), institutional ownership (*IO*), analyst coverage (*COVER*), analyst forecast dispersion (*DISP*), the market beta (*BETA*), the log of market capitalization (*SIZE*), the book-to-market ratio (*BM*), the return in month t (*REV*), the return over month $t-6$ to month $t-1$ (*MOM*), illiquidity (*ILLIQ*), and share turnover (*TURN*). The definitions of these variables are given in the Appendix. We report the cross-sectional mean of each variable winsorized at 25% and 75% for each decile, averaged across months over the period January 1972 to December 2015.

The first set of characteristics includes variables related to volatility, skewness, or extreme payoffs, all of which are expected to be highly associated with *CRASHP*. Both *JAKCPOTP* and *FAILP* are very analogous to *CRASHP* in that they are the ex ante probability of a rare event predicted by a logit model. As expected, both *JACKPOTP* and *FAILP* increase monotonically from the lowest to the highest *CRASHP* decile. The cross-sectional correlations of these variables with *CRASHP* are 0.79 and 0.30, respectively. The high correlations, combined with the negative relations between each of the two variables and the cross-section of returns documented by CKX and CHS, raise the possibility that the cross-sectional return predictability of *CRASHP* we document just mimics the effect of the jackpot or failure probability. We address this concern in Section 5. In addition, idiosyncratic volatility (*IVOL*) and skewness (*ISKEW*) increase with *CRASHP*, which indicates that stocks with a high *CRASHP* experience highly volatile and skewed returns prior to portfolio formation. The patterns in *MAX* and *MIN* across deciles are also clear and consistent with *IVOL* and *ISKEW*.

The next set of variables is considered as measures of limits to arbitrage. It is known in the literature that the price per share (*PRC*) is related inversely to trading costs, institutional ownership (*IO*) indicates investor sophistication or short-sale constraints, and analyst coverage (*COVER*) and analyst forecast dispersion (*DISP*) measure information uncertainty (Bhardwaj and Brooks, 1992; Hong, Lim, and Stein, 2000; Nagel, 2005; Kumar and Lee, 2006; Zhang, 2006). Table 3 shows that stocks with a higher *CRASHP* tend to have lower prices, lower institutional ownership, lower analyst coverage, and higher analyst forecast dispersion. The strong associations of *CRASHP* with these variables consistently indicate that stocks with a high *CRASHP* have high arbitrage costs and mispricing is thus not likely to be eliminated. The relation between the *CRASHP* effect and limits to arbitrage is discussed in Section 4.

The final set of variables comprises firm characteristics known as determinants of stock returns in the cross-section. The market beta (*BETA*) increases from the lowest to the highest *CRASHP* decile. On the other hand, firm size (*SIZE*) and the book-to-market ratio (*BM*) clearly decrease, indicating that the highest *CRASHP* decile consists of small and growth stocks, on average. This result is consistent with that of Fama and French (1996), who find abnormally low returns on the smallest and lowest book-to-market portfolio. The return over month t (*REV*) tends to increase and the past return over months $t-6$ to $t-1$ (*MOM*) also tends to increase from the lowest to the highest *CRASHP* decile, although the patterns are not monotonic. These results imply that stocks with a high *CRASHP* are likely to experience price rises recently prior to portfolio formation. A possible interpretation is that these patterns could be evidence of the overpricing of high-*CRASHP* stocks, if we assume that part of the appreciation could originate from price bubbles and not from changes in fundamental values. Meanwhile, Amihud's (2002) illiquidity (*ILLIQ*) shows that stocks become less liquid as the predicted probability of crashes increases. Share turnover (*TURN*) also displays a strongly increasing pattern across deciles. This result is consistent with the idea that a large trading volume causes a price increase since it affects stock's visibility and subsequent demand (Gervais, Kaniel, and Mingelgrin, 2001).

In sum, the results reported in Table 3 show clear and strong relations of *CRASHP* with various firm characteristics, which indicates that the negative relation of *CRASHP* with subsequent stock returns could depend largely on such characteristics that determine the cross-section of stock returns. To investigate the role of the characteristics on the effect of crash probability, we report in Table 4 the risk-adjusted returns on *CRASHP*-sorted portfolios after controlling for each of the characteristics. The control variables are the same as those reported in Table 3, with two exceptions: residual institutional ownership (*RIO*) and residual analyst coverage (*RCOVER*), replacing institutional ownership (*IO*) and analyst coverage (*COVER*), respectively, and defined as the residuals of the logit of *IO* or *COVER* from the cross-sectional regression on *SIZE* and squared *SIZE*, following Nagel (2005).

Specifically, we first sort stocks into quintile based on a control characteristic at the end of each month t and then, within each quintile, we sort stocks based on *CRASHP* into quintile portfolios. We calculate the monthly value-weighted returns on each intersection of the two sorts realized in month $t + 2$ and report the returns on each *CRASHP* quintile averaged across the control quintiles, after adjusting for the Fama–French three risk factors. Table 4 shows that controlling for various characteristics excluding *FAILP* does not alter the previous finding of the crash probability effect. In particular, the risk-adjusted returns decrease from the lowest to the highest *CRASHP* quintile, with a sharp drop to −0.34% per month in quintile 5, after controlling for *JACKPOTP*. The zero-cost portfolio buying the highest and selling the lowest *CRASHP* quintile yields −0.50% (*t*-statistic = −4.24) per month. In spite of the high correlation between the two variables and the jackpot effect documented in CKX, this evidence shows the cross-sectional return predictability of *CRASHP* is distinguished from the jackpot effect. When controlling for *FAILP*, however, we find the decreasing pattern across quintiles weakens, which leads to an insignificant return difference of −0.22% (*t*-statistic = −1.27) between the highest and lowest *CRASHP* portfolios. This result raises the possibility that the effect of *CRASHP* can be completely eliminated by the failure probability puzzle. We address this issue and provide evidence against this argument in Section 5.

For the other 15 characteristics, the risk-adjusted returns on the zero-cost portfolios controlling for a characteristic range from −0.30% to −0.70% per month on average, which is somewhat reduced in magnitude in comparison with the results in Table 2 but still statistically significant. For example, the zero-cost portfolio buying the highest and selling the lowest *CRASHP* quintile earns −0.30% (*t*-statistic = −2.51) when controlling for *IVOL*, −0.34% (*t*-statistic = −2.81) when controlling for *MAX*, −0.36% (*t*-statistic = −2.16) when controlling for *DISP*, −0.70% (*t*-statistic = −5.33) when controlling for *SIZE*, −0.44% (*t*-statistic = −2.80) when controlling for *BM*, and −0.52% (*t*-statistic = −3.12) when controlling for *REV*. In particular, controlling for the variables whose negative cross-sectional relations to stock returns could result in seeming return predictability of *CRASHP* does not attenuate the cross-sectional pattern of the *CRASHP*-sorted portfolios.

3.3 Firm-level cross-sectional regressions

In the previous section, we show that the relation between crash probability and stock returns is robust to controlling for each of the various firm characteristics closely associated with the cross-section of both crash probability and stock returns. In this section, we use firm-level regression analyses to examine whether the effect of the crash probability remains robust even after simultaneously controlling for the characteristics. Table 5 reports the Fama and MacBeth (1973) estimates for the cross-sectional regression coefficients of stock returns in month t on subsets of the firm characteristics in month $t-1$. Each row of Table 5 represents a different specification of the cross-sectional regression.

The first and second rows of Table 5 present the results of the cross-sectional regressions on *CRASHP* with and without control characteristics, respectively. They confirm the negative and significant relation between *CRASHP* and future stock returns, which is not diminished by the inclusion of a set of control variables such as *BETA*, *SIZE*, *BM*, *REV*, *MOM*, *ILLIQ*, and *TURN*. The average slope coefficient on *CRASHP* in the second row is −6.82 with a t-statistic of −6.32. Multiplying the spread in the mean *CRASHP* between decile 1 and decile 10, reported in Table 3, by the average slope yields a monthly premium of 1.20%, which is substantially larger in magnitude than the return differences reported in Table 2.

To compare the cross-sectional return predictability of *CRASHP* to that of other similar characteristics, given the common control characteristics, *JACKPOTP*, *FAILP*, *IVOL*, *ISKEW*, *MAX*, *MIN*, and *DISP* are each included in turn in the cross-sectional regressions. For *JACKPOTP*, the third and fourth rows of Table 5 confirm the finding of CKX, where stocks with a higher *JACKPOTP* earn subsequently lower returns. When included together with *CRASHP*, however, the coefficient on *JACKPOTP* turns out to be positive, while the negative coefficient on *CRASHP* is unaffected. After we control for the common variables, *JACKPOTP* even loses its predictability for stock returns, but the coefficient on *CRASHP* is close to that in the second row. This finding indicates that the jackpot effect documented by CKX seems to be subsumed by *CRASHP*, although the two probabilities share very similar features. We complement this evidence in Section 5.

The next four rows of Table 5 compare the cross-sectional predictability of *CRASHP* and *FAILP*. In Table 4, we find that the *CRASHP*-sorted portfolios do not yield a significant return difference after controlling for *FAILP*. In contrast, the predictability of *CRASHP* does not disappear when *FAILP* is included in the firm-level regression as well, though the coefficient is slightly reduced. This result suggests that *CRASHP* contains information on future stock returns differentiated from the information contained in *FAILP*, despite the high correlation between the two variables. Meanwhile, the results on *IVOL* are subsequently shown in the next rows. Both *CRASHP* and *IVOL* retain their predictability for future returns after controlling for each other, but the coefficients on both variables become smaller in magnitude.

We also consider idiosyncratic skewness (*ISKEW*) as a variable related to extreme payoffs of stocks. Although stocks with a high *CRASHP* tend to have a high *ISKEW*, the negative relation between *CRASHP* and future returns is not affected by *ISKEW* in our analysis. In addition, we compare the effect of *CRASHP* with the effects of *MAX* and *MIN* following Bali et al. (2011). When the common firm characteristics are not controlled for, *MAX* negatively predicts stock returns for both specifications with and without *CRASHP*, but the coefficients on *MIN* are not significant. In contrast, when the common characteristics are controlled for, *MIN* becomes highly significant for both cases with and without *CRASHP*, whereas the *MAX* effect becomes insignificant by including *CRASHP* in the model. Overall, the cross-sectional effects of *MAX* and *MIN* are not robust to model specifications. On the other hand, the negative coefficient on *CRASHP* is sustained and is not influenced by *MAX* or *MIN*. Finally, the last four rows in Table 5 present the effect of analyst forecast dispersion (*DISP*). The negative *CRASHP*–return relation is unchanged, though the size of the coefficient shrinks, by controlling for *DISP*.

To sum up, the cross-sectional regressions at the firm level provide strong evidence of a negative cross-sectional relation between crash probability and future stock returns. The return predictability of *CRASHP* is maintained even after controlling simultaneously for various related characteristics that also predict the cross-section of stock returns. From both the portfolio-level and firm-level evidence, we confirm that the cross-sectional effect of crash probability is accounted for by neither various firm characteristics nor risk factors in the literature.

4. Probability of Price Crashes and Sources of Overpricing

4.1 Crash probability effects and limits to arbitrage

Our findings in Section 3 imply that low average returns on stocks with high crash probability cannot be explained by the efficient market view, which states that any mispricing cannot persist because rational arbitrageurs quickly trade against mispricing to eliminate deviations from fundamental values, although some irrational agents cause mispricing in the market. Instead, the cross-sectional return predictability of crash probability is consistent with theories that allow for limits to arbitrage. Therefore, in this section we investigate whether the overpricing of stocks with high crash probability is stronger as limits to arbitrage are higher. We use five variables as measures of limits to arbitrage: firm size (*SIZE*), the price per share (*PRC*), illiquidity (*ILLIQ*), analyst coverage (*COVER*), and institutional ownership (*IO*). The first three variables are known to be related to trading costs, lower analyst coverage indicates higher information uncertainty and lower institutional ownership indicates lower investor sophistication and greater short-sale constraints (Bhardwaj and Brooks, 1992; Hong, Lim, and Stein, 2000; Nagel, 2005; Kumar and Lee, 2006; Zhang, 2006; Fama and French, 2008). Following CHS and CKX, we classify sample stocks as high and low limits-to-arbitrage subgroups for each of the five variables each month and then, within each subgroup, we construct quintile portfolios sorted by *CRASHP*. The monthly value-weighted portfolio returns on each quintile for each subgroup are measured two months later after portfolio formation.

When we classify stocks based on the limits-to-arbitrage variables other than *SIZE*, we control for firm size to disentangle the separate effect of those variables from that of firm size. Specifically, for the *SIZE* subgroups, we sort stocks into terciles based on *SIZE* and classify the smallest and largest terciles as the small and big subgroups, respectively. For each of the other four variables, we first sort stocks into terciles based on *SIZE* and then, within each size tercile, we sort stocks into terciles based on the limits-to-arbitrage variable of interest. Then, stocks included in the lowest *PRC* (*ILLIQ*, *COVER*, *IO*) terciles within each size tercile are classified as the low price (liquid, low analyst coverage, low institutional ownership) subgroup, and stocks included in the highest *PRC* (*ILLIQ*, *COVER*, *IO*) terciles within each size tercile are classified as the high price (illiquid, high analyst coverage, high institutional ownership) subgroup.

Table 6 presents returns on the *CRASHP*-sorted portfolios for different subgroups classified by the limits-to-arbitrage variables. Panel A shows the zero-cost portfolio returns buying the highest and selling the lowest *CRASHP* quintile for each subgroup. For firm size, although the average return on the zero-cost portfolio is negative for both the small and big groups, it is statistically significant only for small stocks. For example, the mean monthly returns and three-factor alphas are −0.75% (*t*-statistic = −2.73) and −1.01% (*t*-statistic = −5.27) per month for small firms and −0.16% (*t*-statistic = −0.55) and −0.21% (*t*-statistic = −1.26) per month for large firms, respectively. This result indicates that the effect of crash probability is clearly observed in small firms but not in large firms. When we classify stocks by price per share or liquidity, we find very similar results that the crash probability effect is strong in low-priced and illiquid stocks but not in high-priced and liquid stocks. These results imply that the negative cross-sectional relation between *CRASHP* and stock returns is due to mispricing and delayed arbitrage.

For the next two limits-to-arbitrage variables, the difference in the zero-cost portfolio returns between the high and low limits-to-arbitrage groups is less clear. For analyst coverage, the zero-cost portfolio return is negative and statistically significant for high-coverage firms as well, when controlling for the Fama–French three factors, and it becomes insignificant after additionally controlling for the momentum factor. Nevertheless, the zero-cost portfolio earns more negative returns on average, in low-coverage firms than in high-coverage firms, which is consistent with the prediction of the standard theories on limits to arbitrage. The most interesting patterns are observed for subgroups classified by institutional ownership. For high-institutional-ownership stocks, the mean return on the zero-cost portfolio is −0.39% per month, which is lower than the mean return for low-institutional-ownership stocks. Although the average return becomes insignificant and less negative than in low-institutional-ownership stocks when adjusted by the four-factor model, this seems inconsistent with previous studies that document that various cross-sectional anomalies are concentrated among stocks with low institutional ownership (Nagel, 2005; CHS; CKX; Stambaugh et al., 2015).

To closely examine the relation between the effect of crash probability and limits to arbitrage, we also show the returns on the highest and lowest *CRASHP* quintiles for different subgroups classified by the limits-to-arbitrage variables in Panels B and C of Table 6, respectively. Focusing on the highest *CRASHP* quintile shown in Panel B, we show the average raw and risk-adjusted returns on the highest *CRASHP* quintile are clearly lower in stocks with high limits to arbitrage than in stocks with low limits to arbitrage, when limits to arbitrage are measured by firm size or price per share. For subgroups divided by liquidity and analyst coverage, the three- and four-factor alphas are more negative and significant in illiquid and low-coverage stocks than in liquid and high-coverage stocks. These findings obviously support the view that low average returns on stocks with high crash probability result from delayed arbitrage against the overpricing of those stocks. In contrast, we find no evidence that the overpricing of high-*CRASHP* stocks is prevalent among stocks with lower institutional ownership. Surprisingly, the three-factor alpha on the highest *CRASHP* quintile is lower for high-institutional-ownership firms than for low-institutional ownership-firms, and the four-factor alpha is nearly the same for both groups but significantly different from zero for only high-institutional-ownership firms. This result is evidently not consistent with the literature that argues that overpricing is persistent among stocks with low institutional ownership but not present among stocks with high institutional ownership.

The literature predicts that overpricing is inversely related to institutional ownership for two reasons. First, it is more difficult to arbitrage away mispricing as a lower proportion of shares is owned by institutions, since individuals are likely to be unsophisticated, noise traders and cause prices to deviate from fundamental values, and this *noise trader risk* deters rational arbitrageurs from exploiting mispricing (DeLong et al., 1990a). Second, since the main suppliers of stock loans to short sellers are institutional investors, stocks with lower institutional ownership are more costly to sell short (D'Avolio, 2002; Nagel, 2005). This prevents the overpricing of stocks owned mainly by individuals from being readily eliminated. The common premise of these two reasons is that sophisticated, professional investors unexceptionally act as rational arbitrageurs, which means that they always trade against mispricing unless arbitrage is restricted by frictions in financial markets. This premise leads to the conclusion that overpricing should be less persistent among stocks held mainly by institutions because institutions must take a short position in overpriced stocks whenever they become aware of the overvaluation. However, our findings in Table 6 cannot be explained by this line of literature.

An alternative explanation for our empirical results can be found in theories that view sophisticated investors as rational speculators. According to DeLong et al. (1990b), rational traders can buy and bid up prices above fundamental values because they anticipate that positive feedback traders will buy the securities at even higher prices in the next period. In Abreu and Brunnermeier (2002, 2003), since a price bubble bursts only if arbitrageurs are able to coordinate their selling strategies due to capital constraints, rational investors, who know that the asset is overvalued but that other arbitrageurs are not likely to trade against the bubble yet, maximize their profits by riding the bubble. These theories consistently imply that trading against mispricing is not always optimal for rational investors and, moreover, they even contribute to price movements away from fundamentals. We argue that these theories predict that overpricing, or a price bubble, is not necessarily related to institutional ownership, which is differentiated from the prediction of the standard literature on limits to arbitrage. This is because sophisticated traders may not sell short overpriced stocks even though they can, but they may even ride the bubble; hence the overpricing of stocks owned mainly by institutions may not be arbitraged away for a while. Therefore, our findings in this section are more consistent with the rational speculation view than with the standard limited arbitrage view.

4.2 Crash probability effects across different sub-periods

In the previous section, we find evidence that the overpricing of stocks with high crash probability is less likely to be eliminated when limits to arbitrage are higher generally, but the overpricing still persists even if the stocks are owned largely by institutions. This evidence is definitely inconsistent with the standard literature on limits to arbitrage that assumes that sophisticated traders always trade against mispricing unless the arbitrage trades are impeded. Alternatively, this evidence appears to be consistent with the literature on rational speculation. The rational speculation view further suggests that the overpricing of high-crash-probability stocks could be partly driven by rational speculators, not entirely by behavioral, sentiment-driven investors, whereas the standard limited arbitrage view predicts that price deviations from fundamentals are driven solely by noise traders.

In this section, we follow SYY to provide further investigation. SYY document that various cross-sectional anomalies are stronger following periods of high investor sentiment, since the stocks in the short-leg of each anomaly-based strategy, relatively overpriced compared to the stocks in the long-leg, are more overpriced when sentiment is high.[8] The findings of SYY indicate that the broad set of anomalies examined reflects sentiment-driven overpricing, at least partially. To investigate whether the overpricing of high-crash-probability stocks is associated with variations in market-wide sentiment, we use the monthly market-wide sentiment index constructed by Baker and Wurgler (2006), obtained from Jeffrey Wurgler's website.[9] Specifically, we classify each month from January 1972 to October 2015 as a high-sentiment period if the sentiment index in the previous month is above the median value over the sample period and as a low-sentiment period otherwise.[10] We then separately calculate the average value-weighted returns on decile portfolios sorted by *CRASHP* for the high- and low-sentiment periods. We report the average returns adjusted by the Fama–French three-factor model following each period in Panel A of Table 7.

During the months following high investor sentiment, the long–short strategy that goes long the lowest *CRASHP* decile and short the highest *CRASHP* decile yields 1.01% (*t*-statistic = 3.37) per month, on average, when adjusted for the three risk factors. Moreover, the profit comes from considerably low

---

[8] In their work, the highest-performing decile of each anomaly is called the long-leg, and the lowest-performing decile is called the short-leg.

[9] See http://people.stern.nyu.edu/jwurgler/. We use the data updated as of March 31, 2016.

[10] Since the monthly series of the sentiment index ends September 2015, the sample period is restricted.

returns on the short-leg, the highest *CRASHP* decile, of −0.82% (*t*-statistic = −2.98). On the other hand, a similar pattern is also observed during the months following low investor sentiment, although the slope of decreasing returns across deciles is substantially reduced. During low-sentiment periods, the average return on the long–short strategy is 0.57% (*t*-statistic = 2.07) per month, much less than the average return during high-sentiment periods but still statistically significant. As a result, the difference in the long–short strategy returns between high- and low-sentiment periods does not differ statistically from zero. This result is not consistent with that of SYY. In addition, the insignificant difference in the returns on the highest *CRASHP* decile between the two periods shows why our result differs. The stocks in the highest *CRASHP* decile are more overpriced during high-sentiment periods, but they are also overpriced during low-sentiment periods. This finding suggests that, even though variations in market-wide sentiment could partially drive the cross-sectional return predictability of *CRASHP*, the overpricing of high-*CRASHP* stocks could also be substantially driven by other factors as well.

To examine whether the overpricing of high-*CRASHP* stocks is influenced by variations in economic states but not by market-wide sentiment, we further classify each month into two sub-periods based on each of alternative state indicators. We choose three state variables: the National Bureau of Economic Research (NBER) recession indicator; the market excess return, defined as the CRSP value-weighted return in excess of the one-month T-bill rate; and the liquidity innovation of Pastor and Stambaugh (2003). The average risk-adjusted returns on the *CRASHP*-sorted decile portfolios following expansion and recession periods, up and down market periods, and high- and low-liquidity periods are reported in Panels B to D of Table 7, respectively. In short, the long–short strategy based on *CRASHP* earns highly significant abnormal returns for any sub-period, with only one exception of the low-liquidity period in which the long–short return is only marginally significant. Consequently, there is no significant difference in the long–short portfolio returns across states, no matter what state variables are used. When focusing on the highest *CRASHP* decile, we find no evidence that the overpricing of the highest *CRASHP* decile is pronounced in one particular state. On the one hand, these findings show that the negative cross-sectional relation between *CRASHP* and stock returns is driven neither by hedging demand against changes in economic states nor by investor sentiment. On the other hand, the results in Table 7 show that the overpricing of high-*CRASHP* stocks we find is a very strong and robust phenomenon across different sub-periods.

Our findings are inconsistent with both the predictions based on the standard limited arbitrage theories and the empirical findings of SYY. Meanwhile, our analyses do not provide a direct test of the rational speculation view in the bubble literature. Nevertheless, the results in both the previous section and this section seem to be more consistent with and better explained by the rational speculation view. To supplement our findings, we perform the same analysis for the size-controlled sub-groups classified by institutional ownership in the previous section. Specifically, for each of the institutional ownership subgroups constructed as in Table 6, we sort stocks on *CRASHP* into quintile portfolios and calculate the monthly value-weighted returns on each portfolio realized two months after portfolio formation. Table 8 presents the average risk-adjusted returns on the lowest and highest *CRASHP* quintiles and their differences for each subgroup, following high- and low-investor sentiment periods.

For the low-institutional-ownership group, the stocks in the highest *CRASHP* quintile earn a much lower alpha of −0.76% (*t*-statistic = −2.52) than the lowest *CRASHP* quintile following high-sentiment periods, while the difference between the highest and lowest *CRASHP* quintiles is not significant following low-sentiment periods. This result indicates that the overpricing of stocks with a high *CRASHP* is apparent when investor sentiment is high but not when investor sentiment is low, if the stocks are owned mainly by non-institutional investors. On the contrary, for the high-institutional-ownership group, the highest *CRASHP* quintile yields significantly low average returns, regardless of investor sentiment. As a result, the difference between high- and low-sentiment periods is not significant for either the long–short strategy that goes long the lowest and short the highest *CRASHP* quintile or the short-leg itself. This finding indicates that, if stocks are owned mainly by institutions, the overpricing of high-*CRASHP* stocks is not associated with variations in market-wide sentiment.

Our findings in Table 8, integrated with those in Tables 6 and 7, suggest an explanation that the cross-sectional return predictability of crash probability may be due to both sentiment-driven overpricing and rational speculative bubbles. For stocks largely held by individuals, the overpricing of high-*CRASHP* stocks is substantially driven by market-wide sentiment. This finding is very similar to and consistent with those of previous empirical studies on cross-sectional anomalies. More interesting is the finding that stocks with a high *CRASHP* are overpriced even when they are largely owned by sophisticated investors, regardless of market-wide sentiment. This finding differentiates the crash probability effect from the other various anomalies examined previously. Although our evidence does not rule out other alternative explanations, the crash probability effect appears to be explained better by the rational speculation view than by the standard limited arbitrage view.

4.3 Crash probability effects and changes in institutional holdings

In the previous section, we find that the low average returns on stocks with high crash probability may not be entirely driven by the sentiment of noise traders, which seems to contradict the standard limited arbitrage view. We suggest an alternative explanation, where the overpricing of stocks with high crash probability may be, at least partially, driven by the rational speculation of institutional investors. In this section, to find evidence supporting the rational speculation view, we examine changes in institutional holdings prior to the formation of portfolios sorted by crash probability. Following Edelen et al. (2016), we focus on the number of institutions holding a stock, as well as the percent of shares held by institutions, since the change in the number of institutions is more likely to track new and closed positions, whereas the change in institutional ownership includes adjustments to ongoing positions. We also choose the six-quarter period prior to portfolio formation to measure changes in institutional holdings.

Specifically, at the end of each quarter, the change in the number of institutions ($\Delta INUM$) during the prior six quarters is scaled by the average number in the same market capitalization decile as of the beginning of the six-quarter period. Stocks are sorted into decile portfolios based on crash probability at the end of each quarter. For each decile, we calculate the time-series and cross-sectional averages of both the change in the number of institutions ($\Delta INUM$) and the change in institutional ownership (ΔIO), plotted in Panel A of Figure 2. The two measures for changes in institutional holdings mirror each other very closely. The changes in institutional holdings during the prior six quarters strongly increase as the crash probability becomes higher from decile 1 to decile 9, and they are nearly constant for decile 9 and decile 10. Although the increasing pattern is not monotonic for the two highest *CRASHP* deciles, the change in institutional holdings is undeniably greater for the highest *CRASHP* decile than for the eight lowest *CRASHP* deciles. Interestingly, we report in Table 3 that stocks with higher crash probability have lower institutional ownership. These findings imply that, in spite of the lowest level of institutional holdings, stocks in the highest *CRASHP* decile have been bought most heavily by institutions during the most recent six quarters.

In Panel B of Figure 2, we investigate how the level of institutional holdings for a firm changes around the quarter when the firm enters into the highest *CRASHP* decile. The number of institutions (*INUM*) is scaled by the average number in the same market capitalization decile each quarter. For firms entering into the highest *CRASHP* decile each quarter, we present the average number of institutions (*INUM*) and the average institutional ownership (*IO*) of the firms in excess of the mean of each measure for all firms that have non-missing institutional holdings data that quarter, for the six quarters before and after the entry into the highest decile. The number of institutions (*INUM*) increases steadily during the six quarters prior to the entry into the highest decile, consistent with the results in Panel A of Figure 2. Particularly, it increases more rapidly during the latest two quarters. In contrast, the number of institutions remains almost flat after the firms enter into the highest *CRASHP* decile. Institutional ownership (*IO*) also displays an increasing pattern during the pre-entry period, but peaks one quarter earlier than the entry into the highest decile and slows down afterward. Both measures for institutional holdings indicate that institutional investors show a strong propensity to buy stocks with high crash probability, which are relatively overpriced, until the stock prices reach the peaks of the bubbles.

The results in this section evidently support the rational speculation hypothesis as an explanation for the crash probability effect. Our findings indicate that stocks with higher crash probability exhibit greater institutional demand in the cross-section and that institutions are net buyers for stocks close to but not yet at the peak of overvaluation. These findings imply that sophisticated institutions may drive the prices of currently overpriced stocks even farther away from fundamentals through rational speculation, as predicted by DeLong et al. (1990b) and Abreu and Brunnermeier (2002, 2003). The results throughout Section 4 provide strong evidence that the underperformance of high-crash-probability stocks may arise, at least partially, from rational speculative bubbles driven by sophisticated traders.

5. Probability of Price Crashes and Other Cross-Sectional Anomalies

5.1 Crash probability effects and jackpot effects

In Section 3 we find strong evidence that stocks with a higher probability of future extreme negative returns earn significantly lower returns, on average, while CKX document that stocks with a higher probability of extreme positive returns also do so. Moreover, we find in untabulated results that stocks with higher jackpot probability have smaller sizes, higher past returns in the short- and intermediate-term periods, lower liquidity, higher turnover, and lower institutional ownership, all of which are very similar to the properties of stocks with higher crash probability.[11] These similarities raise the question of whether the cross-sectional return predictability of crash probability subsumes or is subsumed by the jackpot effect of CKX. We find, in Section 3, that the crash probability effect is not subsumed by the jackpot effect in that, in both the portfolio- and firm-level analyses, the negative cross-sectional relation between *CRASHP* and future returns is not eliminated after controlling for jackpot probability. In contrast, the return predictability of *JACKPOTP* disappears when *CRASHP* is added to the cross-sectional regression model. In this section, we provide further evidence on the relation between crash and jackpot probabilities by comparing their cross-sectional return predictability and discuss their similarities and differences.

Table 9 shows the returns on portfolios sorted by jackpot probability and how they are influenced by crash probability. Panel B reports the risk-adjusted returns on the *JACKPOTP*-sorted quintile portfolios

[11] One difference is that stocks with high jackpot probability have relatively higher book-to-market ratios, while stocks with high crash probability are growth stocks.

after controlling for *CRASHP* and, for comparison, Panel A shows the same returns on quintile portfolios sorted by *JACKPOTP* alone. As documented by CKX, we find in Panel A that the returns on the *JACKPOTP*-sorted portfolios clearly decrease from quintile 1 to quintile 5, with a particularly low return of −0.42% (*t*-statistic = −2.58) per month on quintile 5. However, in Panel B, the decreasing pattern in portfolio returns disappears after controlling for *CRASHP*; thus the zero-cost portfolio buying the highest and selling the lowest *JACKPOTP* quintile yields an insignificant alpha of 0.07% (*t*-statistic = 0.77). This result is consistent with the results of the firm-level regressions in Table 5. Panel C of Table 9 also reports the risk-adjusted returns on portfolios sorted first by *CRASHP* and then by *JACKPOTP*. In short, in each row classified by *CRASHP*, sorting on *JACKPOTP* does not produce a significant cross-sectional difference in returns. In contrast, for each column ranked by *JACKPOTP*, the highest *CRASHP* portfolio earns substantially low returns, which leads to a negative and significant return difference between the highest and lowest *CRASHP* portfolios. These findings confirm that the return predictability of jackpot probability is completely subsumed by that of crash probability, even though the two probabilities share very similar characteristics.

A distinctive feature of the crash probability effect is that the overpricing of high-*CRASHP* stocks is robust regardless of firms' institutional ownership, as documented in Section 4. In contrast, CKX document that the jackpot effect is much stronger in firms with low institutional ownership. In Section 4, we also find evidence that the overpricing of high-*CRASHP* stocks is not related to variations in investor sentiment, particularly for stocks owned mainly by institutions. To investigate the association of the jackpot effect with variations in market-wide sentiment, we perform the same analysis, following SYY, on the *JACKPOTP*-sorted quintile portfolios and report the average risk-adjusted returns following high- and low-sentiment periods in Panel D of Table 9. The results clearly differ from those for the *CRASHP*-sorted portfolios reported in Table 7. During the month following high-sentiment periods, the portfolio returns show a sharply decreasing pattern across quintiles, yielding a significant alpha of 0.96% (*t*-statistic = 4.37) on the long–short strategy that goes long the lowest and short the highest *JACKPOTP* quintile. During the month following low-sentiment periods, however, the portfolio returns are relatively flat across quintiles and the same long–short strategy earns nearly zero returns. The difference in returns between high- and low-sentiment periods clearly indicates that the jackpot effect can be attributed to the sentiment-driven overpricing of high-*JACKPOTP* stocks, like the other 11 anomalies examined by SYY. On the other hand, our findings in Section 4 suggest that the crash probability effect is not completely driven by market-wide sentiment and may be partially attributed to rational speculative bubbles driven by sophisticated traders. This different source of overpricing seems to enable the crash probability effect to completely subsume the jackpot effect.

Finally, to compare the *CRASHP*- and *JACKPOTP*-sorted portfolios, we plot the factor loadings of the excess returns on decile portfolios sorted by either *CRASHP* or *JACKPOTP* on the Carhart four factors in Figure 3. The two sets of decile portfolios have similar patterns in MKT, SMB, and HML loadings, but somewhat different patterns in WML loadings. The loadings on SMB increase as both crash and jackpot probabilities increase, while the HML loadings decrease as both probabilities increase. In particular, the highest *CRASHP* decile loads very negatively on HML, compared to the highest *JACKPOTP* decile, indicating that it contains more growth stocks. On the other hand, the loadings on WML appear to be decreasing across both sets of decile portfolios, but the shape of the decreasing pattern is different. Whereas the pattern is convex for the *JACKPOTP*-sorted portfolios, it is rather concave for the *CRASHP*-sorted portfolios. As a result, the WML loadings sharply drop from decile 8 to decile 10 when sorted by *CRASHP*, but they increase slightly when sorted by *JACKPOTP*. The large negative WML loading of the highest *CRASHP* decile is surprising because the highest *CRASHP* decile includes past winner stocks on average, as reported in Table 3. Our interpretation is that stocks that have recently experienced price rises are likely to be near the peak of price bubbles, if the price rises do not entirely arise from changes in fundamentals; hence, they have a high probability of bursting of the bubbles. If the bubbles burst, as expected, after forming the portfolios, then these stocks would experience large negative returns in the holding periods of the portfolios. Consequently, the portfolio returns would move along with past losers, which leads to a negative WML loading. This interpretation suggests that the superior predictability of crash probability for future returns could stem from the ability of timing the peak of price bubbles, as well as the ability of picking relatively overpriced stocks.

5.2 Crash probability effects and the distress puzzle

In this section, we discuss the difference between the cross-sectional return predictability of crash probability and failure probability. Our measure of crash probability and the failure probability constructed by CHS are similar in that both variables are the ex ante probabilities of negative and rare events and the two events of price crashes and firms' failure may be accompanied by one another. In Section 3, we find that the return difference between the highest and lowest *CRASHP* portfolios becomes insignificant after controlling for *FAILP*. In addition, in untabulated results, we find that the return difference between the highest and lowest *FAILP* portfolios is not affected by controlling for *CRASHP*. These findings raise the concern that the negative relation between crash probability and future returns just mimics the distress puzzle documented by CHS, which is more prevalent and fundamental. However, the firm-level cross-sectional regressions differ by showing that the return predictability of *CRASHP* is retained, even after controlling for *FAILP*. Given these conflicting findings, we further investigate whether crash probability, compared to failure probability, contains additional information for future returns that leads to the difference in the return predictability of the two probabilities.

To this end, we construct new variables by orthogonalizing the crash and failure probabilities to each other to entirely eliminate the shared information between the two probabilities. Specifically, we regress *CRASHP* on *FAILP* each month in the cross-section and refer to the residuals as the residual crash probability (*RCRASHP*). Likewise, we also regress *FAILP* on *CRASHP* and refer to the residuals as the residual failure probability (*RFAILP*). The variable *RCRASHP* is highly correlated with *CRASHP* but not correlated with *FAILP*, and *RFAILP* is highly correlated with *FAILP* but not correlated with *CRASHP*. Then, we construct decile portfolios sorted on *RCRASHP* and *RFAILP*, respectively, at the end of month t, and calculate the monthly value-weighted returns on each portfolio realized in month $t + 2$. Table 10 reports the mean returns, CAPM alphas, Fama–French three-factor alphas, and Carhart four-factor alphas on the *RCRASHP*- and *RFAILP*-sorted portfolios in Panels A and B, respectively.

Panel A of Table 10 shows that the mean returns do not show a clear pattern from decile 1 to decile 6 but decrease from decile 6 to decile 10, which is comparable to the pattern in the *CRASHP*-sorted portfolios reported in Table 2. The alphas on the zero-cost portfolio buying the highest and selling the lowest *RCRASHP* deciles are −1.02% (*t*-statistic = −3.17) for the CAPM, −0.58% (*t*-statistic = −2.38) for the three-factor model, and −0.44% (*t*-statistic = −2.14) per month for the four-factor model, respectively, all of which are statistically significant. Moreover, these significant return differences are mainly due to substantially low returns on the highest *RCRASHP* decile. On the other hand, Panel B of Table 10 shows that the cross-sectional relation between *RFAILP* and future returns is rather weak. Interestingly, the return difference between the highest and lowest *RFAILP* deciles is not enough to reject a zero return. In sum, the results in Table 10 confirm that crash probability has distinct information, not contained in failure probability, to predict future returns, while the return predictability of failure probability becomes weak when orthogonalized to crash probability.

In Figure 4, we compare the loadings of the excess returns on each of the *CRASHP*- and *FAILP*-sorted portfolios on the Carhart four factors. Whereas the loadings on MKT and SMB of the two sets of decile portfolios do not differ much from each other, the loadings on HML and WML display a significant gap between the *CRASHP*- and *FAILP*-sorted portfolios. The loadings on HML are close to an inverse U-shape across the *FAILP* deciles and they are positive for all the portfolios except the two extreme deciles, whose loadings are slightly negative but insignificant. This finding is in sharp contrast to the clearly decreasing pattern of the *CRASHP*-sorted portfolios, which is consistent with the highest *FAILP* decile containing more value stocks, compared to the highest *CRASHP* decile containing growth stocks.[12] For the loadings on WML, although they tend to decrease across deciles for both the *CRASHP*- and *FAILP*-sorted portfolios, the WML loadings for the *FAILP*-sorted portfolios display an even steeper decline than

[12] The 25% winsorized mean of the book-to-market ratio of the highest *FAILP* decile is 0.830 on average, while that of the highest *CRASHP* decile is 0.538, as reported in Table 3.

those for the *CRASHP*-sorted ones. Consistent with this negative relation between WML loadings and failure probability, we find in untabulated results that the average past returns decline sharply from the lowest to the highest *FAILP* decile. These findings indicate that stocks in the highest *FAILP* decile closely resemble momentum losers, while the highest *CRASHP* decile includes momentum winners, despite its negative loading on WML, as discussed in Section 5.1.

When we compare various characteristics between the two sets of decile portfolios, the most remarkable difference is that the average returns over different horizons from one to 12 months prior to portfolio formation are positively associated with crash probability but negatively so with failure probability, while the returns subsequent to portfolio formation are negatively associated with both probabilities. To determine the fundamental difference between the return predictability of *CRASHP* and that of *FAILP* by focusing on the top deciles, we conduct an event study in which the event is defined as the entry of a stock into the highest *CRASHP* or *FAILP* decile, and we examine the daily average abnormal returns around the event. Specifically, decile portfolios are constructed by sorting stocks based on each of *CRASHP* and *FAILP* at the end of each month, and the end of the month when a stock enters into the highest decile is defined as the event day. Then, abnormal returns (ARs) and cumulative abnormal returns (CARs) are calculated over the event window, comprised of 126 pre-event trading days, the event day, and 126 post-event trading days, based on the market model estimated from the window of 126 trading days prior to the event window. Figure 5 presents the average ARs and CARs over the event window for the highest *CRASHP* decile in Panel A and those for the highest *FAILP* decile in Panel B.

We can tell at a glance that the two panels in Figure 5 are very different from each other. For the highest decile of *CRASHP*, the CAR increases slowly but obviously during a horizon from $d - 126$ to $d - 21$, shoots up from $d - 21$ to the event day d, and then steadily declines during the post-event days. In contrast, for the highest decile of *FAILP*, the CAR continues to decrease over the event window, more steeply in particular during the month prior to the event. These patterns are consistent with the previous findings that the highest *CRASHP* decile contains past winners and the highest *FAILP* decile contains past losers. Moreover, these patterns imply that the sources of return predictability may differ completely between the two probability measures. The continuing decline in the returns of the distressed firms suggests that their overpricing results from underreaction to bad news. On the other hand, the sudden reversal of sharply increasing returns in the highest *CRASHP* decile, which seems far removed from the underreaction story, could be regarded as growing price bubbles followed by bursting of them. Of course, this study does not explain what makes stocks overvalued fundamentally and why those stocks are so. Nevertheless, our findings in Figure 5 supports the interpretation that the strong cross-sectional relation between crash probability and future returns could be due to both of the ability to time the peaks of the bubbles and the ability to pick the most overpriced stocks.

5.3 Crash probability effects and various cross-sectional anomalies

In this section, we ask whether the crash probability effect is linked to the cross-sectional anomalies examined by SYY and Stambaugh et al. (2015). SYY and Stambaugh et al. (2015) examine 11 cross-sectional anomalies known to survive adjustments for risk exposure and provide evidence that these 11 anomalies are at least partially caused by sentiment-driven overpricing. Since our finding of the return predictability of crash probability is also primarily due to the overpricing of stocks with high crash probability, there could be concerns that our finding is only another form of one of numerous anomalies confirmed by the literature. Therefore, we investigate the role of the 11 anomalies on the cross-sectional return predictability of crash probability and provide evidence that our finding is an independent phenomenon of these anomalies. Since two of the 11 anomaly variables—the failure probability of CHS and the momentum of Jegadeesh and Titman (1993)—are examined in Sections 3 and 5.2, respectively, we investigate only the remaining nine anomaly variables here. These include Ohlson's (1980) O-score (*OSCR*); the net stock issues (*NSI*) of Ritter (1991); the composite equity issues (*CEI*) of Daniel and Titman (2006); the total accruals (*TA*) of Sloan (1996); the net operating assets (*NOA*) of Hirshleifer et al. (2004); the gross profitability (*GP*) of Novy-Marx (2013); the asset growth (*AG*) of Cooper, Gulen, and Schill (2008); the return on assets (*ROA*) of Fama and French (2006); and the abnormal capital investment (*ACI*) of Titman, Wei, and Xie (2004). The definitions of these variables are given in the Appendix.

Table 11 presents the risk-adjusted returns on the *CRASHP*-sorted portfolios after controlling for each of the nine anomaly variables. The procedure of portfolio construction is identical to that for Table 4. Overall, the zero-cost portfolio buying the highest and selling the lowest *CRASHP* quintile yields negative three-factor alphas for all the controlling variables, six of which are statistically significant. The insignificant alphas are −0.17% (*t*-statistic = −1.58) when controlling for *CEI*, −0.30% (*t*-statistic = −1.64) when controlling for *TA*, and −0.25% (*t*-statistic = −1.36) when controlling for *AG*. For the six variables yielding significant results, the three-factor alphas on the zero-cost portfolios range from −0.29% to −0.58% per month, which is comparable to the results in Table 4.

Table 12 presents the results of firm-level cross-sectional regressions including together crash probability and each of the nine anomaly variables. As in Table 5, the set of control characteristics includes *BETA*, *SIZE*, *BM*, *REV*, *MOM*, *ILLIQ*, and *TURN*. Excluding *OSCR* and *AG*, the anomaly variables have statistically significant predictability, with signs consistent with those documented by previous studies, that is, positive for *GP* and *ROA* and negative otherwise. The cross-sectional predictability of these anomaly variables is not affected by crash probability and other firm characteristics. More importantly, our measure of crash probability has a highly significant and negative coefficient in all the specifications considered. Compared to the results in Table 5, the magnitudes of the coefficients on *CRASHP* are not affected much by the anomaly variables. The results in Table 12 confirm that the cross-sectional return predictability of crash probability is not altered by controlling for the broad set of anomaly variables examined by SYY, strengthening the robustness of our findings.

6. Conclusion

Motivated by the recent study of CKX, this study investigates the cross-sectional relation between the predicted probability of price crashes and future stock returns and finds strong evidence that differentiates our results from other cross-sectional anomalies examined extensively in the literature. First, we find that stocks with a high probability of price crashes earn substantially low returns on average, which is not accounted for by various firm characteristics known as the determinants of the cross-section of stock returns. Second, we find evidence that the overpricing of stocks with high crash probability is not arbitraged away even among stocks owned largely by institutions. Moreover, the cross-sectional return predictability of crash probability is not associated with variations in market-wide sentiment, particularly for stocks with high institutional ownership. These findings cast doubt on the role of institutional investors as rational arbitrageurs as assumed in the standard literature on limited arbitrage, which predicts that overpricing is inversely related to institutional ownership and that mispricing is primarily driven by noise traders. Third, we find that stocks with higher crash probability exhibit greater institutional demand and institutional investors have a tendency to buy stocks close to but not yet at the peak of overvaluation. This evidence suggests that the overpricing of high-crash-probability stocks could also be driven by sophisticated institutions, not solely by behavioral, noise traders. Finally, the cross-sectional return predictability of jackpot probability documented by CKX is completely eliminated when controlling for crash probability, while the crash probability effect is not altered by jackpot probability. In addition, we provide evidence that the crash probability effect clearly differs from a broad set of cross-sectional anomalies, including the financial distress puzzle of CHS.

Although our evidence does not rule out other possible explanations, our distinctive findings can be well understood along the line of theories on rational speculation that presumes that sophisticated traders do not always stabilize asset prices but destabilize them by riding a price bubble to maximize profits. We suggest that the underperformance of stocks with high crash probability could be result partially from rational speculative bubble driven by sophisticated traders, not entirely from sentiment-driven overpricing by noise traders. The different sources of overpricing seem to enable the crash probability effect to survive after controlling for other related anomaly variables. Our evidence also suggests that the superior return predictability of crash probability could stem from both the ability of timing the peak of price bubbles and the ability of picking overpriced stocks.

Appendix: Variable Definitions

The definitions of all the variables in this study are provided in alphabetical order.

ACI: Abnormal capital investment at the end of June in year t to May in year $t + 1$ is defined as capital expenditure for year $t - 1$ divided by the previous three-year average of capital expenditure, minus one. The capital expenditure (Compustat annual item CAPX) is scaled by sales (item SALE) for the same year.

AG: Asset growth at the end of June in year t to May in year $t + 1$ is the annual change in total assets (Compustat annual item AT) for year $t - 1$ divided by total assets for year $t - 2$.

AGE: Firm age is the number of years since the firm's first appearance on the CRSP monthly stock file.

BETA: Market beta is estimated from the time-series regression of daily excess returns on market excess returns, defined as the CRSP value-weighted returns in excess of the one-month T-bill rate, in each month. To avoid issues related to nonsynchronous trading, we use the lag and lead of the market excess return as well as the current market return. Then, the market beta is defined as the sum of the three slope coefficients on the lag, lead, and current market excess returns. We exclude stocks with fewer than 15 daily excess returns a month.

BM: Book-to-market ratio at the end of June in year t to May in year $t + 1$ is the ratio of book equity for the end of year $t - 1$ to market equity at the end of December in year $t - 1$. The book equity is defined as stockholders' equity (Compustat annual item SEQ), plus deferred taxes and investment tax credit (item TXDITC) if available, minus the book value of preferred stocks. The book value of preferred stocks is the preferred stock redemption value (item PSTKRV) if available, or the liquidating value (item PSTKL) if available, or the par value (item PSTK).

CEI: Composite equity issues at the end of June in year t to May in year $t + 1$ is measured as the log of the ratio of market equity at the end of June in year t to market equity at the end of June in year $t - 5$, minus the cumulative log return from the end of June in year $t - 5$ to the end of June in year t.

COVER: Analyst coverage is the log of one plus the number of estimates for earnings forecasts (I/B/E/S unadjusted file item NUMEST) for the current fiscal year (item FPI = 1).

CRASHP: Crash probability at the end of month t is the probability of log returns less than −70% from month $t + 1$ to month $t + 12$, predicted out-of-sample from the generalized logit model specified in Equation (1).

DISP: Analyst forecast dispersion of Diether, Malloy, and Scherbina (2002) is the standard deviation of earnings forecasts (I/B/E/S unadjusted file item STDEV) divided by the absolute value of the mean of earnings forecasts (item MEANEST) for the current fiscal year (item FPI = 1).

DTURN: Detrended turnover is the six-month average of monthly share turnover subtracted by the prior 18-month average.

FAILP: Following Campbell, Hilscher, and Szilagyi (2008), failure probability is defined by

$$FAILP_{i,t} = \frac{1}{1+\exp\left(-DIST_{i,t}\right)} ,$$

where

$$\begin{aligned} DIST_{i,t} = {} & -9.164 - 20.264 \times NIMTAAVG_{i,t} + 1.416 \times TLMTA_{i,t} - 7.129 \times EXRETAVG_{i,t} \\ & + 1.411 \times SIGMA_{i,t} - 0.045 \times RSIZE_{i,t} - 2.132 \times CASHMTA_{i,t} + 0.075 \times MB_{i,t} - 0.058 \times PRICE_{i,t} , \end{aligned}$$

$$NIMTAAVG_{i,t} = \frac{1-\phi}{1-\phi^{12}}\left(NIMTA_{i,t} + \phi NIMTA_{i,t-1} + \cdots + \phi^{11} NIMTA_{i,t-11}\right) ,$$

$$EXRETAVG_{i,t} = \frac{1-\phi}{1-\phi^{12}}\left(EXRET_{i,t} + \phi EXRET_{i,t-1} + \cdots + \phi^{11} EXRET_{i,t-11}\right) ,$$

and $\phi = 2^{-1/3}$. *NIMTA* is net income (Compustat quarterly item NIQ) divided by the sum of market equity (price per share times the number of outstanding shares) and total liabilities (item LTQ), *EXRET* is the firm's log return in excess of the log return on S&P 500 index, *TLMTA* is total liabilities divided by the sum of market equity and total liabilities, *SIGMA* is the annualized three-month rolling sample standard deviation of daily returns, *RSIZE* is the log ratio of market equity to the total market value of the S&P 500 index, *CASHMTA* is cash and short-term investments (item CHEQ) divided by the sum of market equity and total liabilities, *MB* is the market-to-book equity ratio, and *PRICE* is the log of the price per share. We winsorize each of the explanatory variables at their fifth and 95th percentiles of the pooled distribution of all firm–month observations.

GP: Gross profitability at the end of June in year t to May in year $t + 1$ is measured as total revenue (Compustat annual item REVT) minus the cost of goods sold (item COGS) for year $t - 1$, scaled by total assets (item AT).

ILLIQ: Illiquidity at the end of month t denotes the price impact measure proposed by Amihud (2002), defined as the average of the ratio of the absolute daily return to the dollar trading volume in month t.

IO: Institutional ownership is the fraction of shares owned by institutions, obtained from the Thomson Reuters Institutional (13f) Holdings database.

ISKEW: Idiosyncratic skewness at the end of month t is defined as the skewness of the residuals from the time-series regression of daily excess returns from month $t - 5$ to month t on the Fama–French three factors. We exclude stocks with fewer than 50 daily excess returns during the six-month period.

IVOL: Idiosyncratic volatility at the end of month t is defined as the standard deviation of the residuals from the time-series regression of daily excess returns in month t on the Fama–French three factors. We exclude stocks with fewer than 15 daily excess returns each month.

JACKPOTP: Jackpot probability at the end of month t is the probability of log returns greater than 70% from month $t + 1$ to month $t + 12$, predicted out-of-sample from the generalized logit model specified in Equation (1).

MAX: Maximum daily return within month t. We exclude stocks with fewer than 15 daily returns each month.

MIN: Negative of the minimum daily return within month t. We exclude stocks with fewer than 15 daily returns each month.

MOM: Momentum, measured at the end of month t, is defined as the cumulative return from month $t - 6$ to month $t - 1$.

NOA: Net operating assets at the end of June in year t to May in year $t + 1$ is defined as operating assets minus operating liabilities for year $t - 1$, scaled by total assets (Compustat annual item AT) for year t

− 2. Operating assets is total assets minus cash and short-term investments (item CHE). Operating liabilities is total assets minus debt in current liabilities (item DLC) minus long-term debt (item DLTT) minus minority interest (item MIB) minus preferred stock (item PSTK) minus common equity (item CEQ).

NSI: Net stock issues at the end of June in year t to May in year $t + 1$ is measured as the log of the ratio of split-adjusted shares outstanding at the end of year $t - 1$ to the split-adjusted shares outstanding at the end of year $t - 2$, where split-adjusted shares outstanding is common shares outstanding (Compustat annual item CSHO) times the adjustment factor (item AJEX).

OSCR: Following Ohlson (1980), the O-score is defined by

$$\begin{aligned} OSCR = &-1.32 - 0.407 \times ASIZE + 6.03 \times TLTA - 1.43 \times WCTA + 0.0757 \times CLCA \\ &-1.72 \times OENEG - 2.37 \times NITA - 1.83 \times FUTL + 0.285 \times INTWO - 0.521 \times CHIN \end{aligned},$$

where *ASIZE* is the log of total assets (Compustat annual item AT) divided by the GDP deflator, *TLTA* is the ratio of total liabilities (item LT) to total assets, *WCTA* is the ratio of working capital (item WCAP) to total assets, *CLCA* is the ratio of current liabilities (item LCT) to current assets (item ACT), *OENEG* is one if total liabilities exceed total assets and zero otherwise, *NITA* is the ratio of net income (item NI) to total assets, *FUTL* is funds from operations (item FOPT) divided by total liabilities, *INTWO* is one if net income has been negative for the last two years and zero otherwise, and *CHIN* is the difference in net income between the current and previous years divided by the sum of the absolute values of the two values.

RCOVER: Residual analyst coverage is defined as the residuals from the cross-sectional regression of analyst coverage (*COVER*) on firm size (*SIZE*) and squared firm size ($SIZE^2$) for each month.

RCRASHP: Residual crash probability is defined as the residuals from the cross-sectional regression of the predicted probability of price crashes (*CRASHP*) on failure probability (*FAILP*) for each month.

RET12: Measured for each stock at the end of month t as the log return on the stock from month $t - 11$ to month t.

REV: Short-term reversal represents the return on a stock over month t.

RFAILP: Residual failure probability is defined as the residuals from the cross-sectional regression of failure probability (*FAILP*) on the predicted probability of price crashes (*CRASHP*) for each month.

RIO: Residual institutional ownership is defined as the residuals from the cross-sectional regression of the logit of institutional ownership (*IO*) on firm size (*SIZE*) and squared firm size ($SIZE^2$) for each month, following Nagel (2005). If the values of *IO* are below 0.0001 or above 0.9999, they are replaced with 0.0001 or 0.9999, respectively.

ROA: Return on assets at the end of month t is defined as income before extraordinary items (Compustat item IBQ) for the most recently announced quarter, divided by one-quarter-lagged total assets (item ATQ).

SALESG: Sales growth at the end of June in year t to May in year $t + 1$ is the log difference between sales (Compustat annual item SALE) at the end of year $t - 1$ and sales at the end of year $t - 2$.

SIZE: Firm size is defined as the log of the price per share times the number of shares outstanding.

TA: Total accruals at the end of June in year t to May in year $t + 1$ is defined as the annual change for year $t - 1$ in non-cash working capital minus depreciation (Compustat annual item DP), scaled by the last two-year average of total assets (item AT). Non-cash working capital is non-cash current assets, measured as current assets (item ACT) minus cash and short-term investments (item CHE), minus current liabilities (item LCT) less debt in current liabilities (item DLC) and income taxes payable (item TXP).

TANG: Tangible assets at the end of June in year t to May in year $t + 1$ are property, plant, and equipment (Compustat annual item PPEGT) divided by total assets (item AT) for the end of year $t - 1$.

TSKEW: Total skewness at the end of month t is calculated using daily log returns from month $t - 5$ to month t. We exclude stocks with fewer than 50 daily log returns during the six-month period.

TURN: Turnover is the ratio of the monthly trading volume to the number of shares outstanding.

TVOL: Total volatility at the end of month t is the standard deviation of daily log returns from month $t - 5$ to month t. We exclude stocks with fewer than 50 daily log returns during the six-month period.

References

Abreu, D., Brunnermeier, M.K., 2002. Synchronization risk and delayed arbitrage. *Journal of Financial Economics* 66, 341–360.

Abreu, D., Brunnermeier, M.K., 2003. Bubbles and crashes. *Econometrica* 71, 173–204.

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

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61, 259–299.

Avramov, D., Chordia, T., Jostova, G., Philipov, A., 2013. Anomalies and financial distress. *Journal of Financial Economics* 108, 139–159.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61, 1645–1680.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99, 427–446.

Banz, R.W., 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9, 3–18.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* 13, 1–42.

Bhardwaj, R.K., Brooks, L.D., 1992. The January anomaly: effects of low share price, transaction costs, and bid–ask bias. *Journal of Finance* 47, 553–575.

Boyer, B., Mitton, T., Vorkink, K., 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23, 169–202.

Brunnermeier, M.K., Nagel, S., 2004. Hedge funds and the technology bubble. *Journal of Finance* 59, 2013–2040.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31, 281–318.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *Journal of Finance* 63, 2899–2939.

Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.

Chen, J., Hong, H., Stein, J.C., 2001. Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61, 345–381.

Conrad, J., Kapadia, N., Xing, Y., 2014. Death and jackpot: why do individual investors hold overpriced stocks? *Journal of Financial Economics* 113, 455–475.

Cooper, M.J., Gulen, H., Schill, M.J., 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63, 1609–1652.

Daniel, K.D., Titman, S., 2006. Market reactions to tangible and intangible information. *Journal of Finance* 61, 1605–1643.

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

DeLong, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990a. Noise trader risk in financial markets. *Journal of Political Economy* 98, 703–738.

DeLong, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990b, Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance* 45, 379–395.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross-section of stock returns. *Journal of Finance* 57, 2113–2141.

Dow, J., Gorton, G., 1994. Arbitrage chains. *Journal of Finance* 49, 819–849.

Edelen, R.M., Ince, O.S., Kadlec, G.B., 2016. Institutional investors and stock return anomalies. *Journal of Financial Economics* 119, 472–488.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749–1778.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.

Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance*

51, 55–84.

Fama, E.F., French, K.R., 2006. Profitability, investment, and average returns. *Journal of Financial Economics* 82, 491–518.

Fama, E.F., French, K.R., 2008. Dissecting anomalies. *Journal of Finance* 63, 1653–1678.

Fama, E.F., MacBeth, J.D., 1973. Risk, return and equilibrium: empirical tests. *Journal of Political Economy* 81, 607–636.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19, 3–29.

Gervais, S., Kaniel, R., Mingelgrin, D.H., 2001. The high-volume return premium. *Journal of Finance* 56, 877–919.

Glosten, L., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Griffin, J.M., Harris, J.H., Shu, T., Topaloglu, S., 2011. Who drove and burst the tech bubble? *Journal of Finance* 66, 1251–1290.

Harrison, M.J., Kreps, D.M., 1978. Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics* 92, 323–336.

Hirshleifer, D., Hou, K., Teoh, S.H., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38, 297–331.

Hong, H., Lim, T., Stein, J.C., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance* 55, 265–295.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* 48, 65–91.

Kumar, A., Lee, C.M.C., 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451–2486.

Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. *Journal of Finance* 32, 1151–1168.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. *Journal of*

Financial Economics 78, 277–309.

Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692.

Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.

Newey, W.K., West, K.D., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance-matrix. *Econometrica* 55, 703–708.

Novy-Marx, R., 2013. The other side of value: the gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18, 109–131.

Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 642–685.

Ritter, J.R., 1991. The long-run performance of initial public offerings. *Journal of Finance* 46, 3–27.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. *Journal of Finance* 52, 35–55.

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

Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: investor sentiment and anomalies. *Journal of Financial Economics* 104, 288–302.

Stambaugh, R.F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70, 1903–1948.

Stattman, D., 1980. Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers* 4, 25–45.

Titman, S., Wei, K., Xie, F., 2004. Capital investments and stock returns. *Journal of Financial and Quantitative Analysis* 39, 677–700.

Zhang, X.F., 2006. Information uncertainty and stock returns. *Journal of Finance* 61, 105–136.

Table 1. Parameter estimates of the prediction model for extreme stock returns

This table reports the estimated parameters of the generalized logit model specified in Equation (1) for predicting extreme stock returns. A crash is defined as log returns less than −70% over the next 12 months and a jackpot is defined as log returns greater than 70% over the next 12 months. The variable *RET12* is the log return over the past 12 months; *TVOL* and *TSKEW* are the standard deviation and skewness, respectively, of daily log returns over the past six months; *SIZE* is the log of market capitalization; *DTURN* is the detrended turnover, defined by the six-month average of monthly share turnover minus the prior 18-month average; *AGE* is the number of years since the firm's first appearance on the CRSP monthly stock file; *TANG* is tangible assets divided by total assets; and *SALESG* is the sales growth over the prior year. All explanatory variables are constructed to be observable at the beginning of the 12-month period over which a crash or jackpot is measured. The sample includes all available firm–month observations during the period June 1952 to December 2015.

	Crash			Jackpot			
Variable	Coefficient	z-Statistic	% change in odds ratio for a σ change	Coefficient	z-Statistic	% change in odds ratio for a σ change	R^2
Intercept	-3.974	-139.196		-2.116	-65.002		0.128
RET12	0.040	5.880	1.70 %	0.042	5.047	1.80 %	
TVOL	44.107	198.846	85.50 %	34.350	130.569	61.80 %	
TSKEW	0.039	12.934	4.30 %	0.031	8.620	3.40 %	
SIZE	0.044	21.218	9.00 %	-0.131	-52.621	-22.40 %	
DTURN	-0.003	-1.301	-0.30 %	-0.047	-14.580	-5.40 %	
AGE	-0.028	-77.500	-36.00 %	-0.014	-38.955	-20.70 %	
TANG	-0.484	-44.295	-17.10 %	-0.199	-16.996	-7.40 %	
SALESG	0.357	46.670	13.10 %	0.178	17.598	6.30 %	

Table 2. Returns on portfolios sorted by the out-of-sample predicted probability of price crashes

This table reports the returns on decile portfolios sorted on the predicted probability of price crashes. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*). The monthly portfolio returns on each decile realized in month *t* + 2 are calculated and the mean returns, CAPM alphas, Fama–French three-factor alphas, and Carhart four-factor alphas are reported. Column 1 shows the results for the lowest *CRASHP* decile, column 10 is for the highest *CRASHP* decile, and column 10-1 is for the zero-cost portfolio buying the highest and selling the lowest *CRASHP* decile. Panel A reports the value-weighted returns and Panel B reports the equal-weighted returns. The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015.

	1	2	3	4	5	6	7	8	9	10	10-1
Panel A: Value-weighted portfolios sorted on *CRASHP*											
Mean	0.99	0.84	0.94	1.01	1.05	1.03	0.91	0.97	0.76	0.36	-0.64
	(5.85)	(4.45)	(4.56)	(4.27)	(4.56)	(4.38)	(3.35)	(2.90)	(2.29)	(0.78)	(-1.68)
CAPM α	0.20	-0.02	0.05	0.06	0.08	0.01	-0.15	-0.19	-0.42	-0.96	-1.16
	(3.15)	(-0.34)	(0.60)	(0.88)	(1.11)	(0.13)	(-1.76)	(-1.18)	(-2.34)	(-3.44)	(-3.57)
Three-factor α	0.15	-0.08	0.02	0.05	0.13	0.10	-0.02	0.01	-0.23	-0.64	-0.79
	(2.96)	(-1.38)	(0.32)	(0.74)	(1.90)	(1.05)	(-0.25)	(0.12)	(-1.95)	(-3.29)	(-3.60)
Four-factor α	0.08	-0.06	-0.02	-0.01	0.10	0.15	0.01	0.05	-0.09	-0.41	-0.50
	(1.65)	(-1.08)	(-0.35)	(-0.14)	(1.40)	(1.59)	(0.16)	(0.42)	(-0.75)	(-2.40)	(-2.55)
Panel B: Equal-weighted portfolios sorted on *CRASHP*											
Mean	1.13	1.16	1.23	1.30	1.34	1.28	1.28	1.17	0.94	0.40	-0.73
	(6.69)	(6.17)	(6.11)	(5.92)	(5.68)	(5.35)	(4.83)	(4.00)	(2.83)	(1.04)	(-2.36)
CAPM α	0.33	0.32	0.36	0.39	0.39	0.30	0.25	0.09	-0.21	-0.86	-1.19
	(3.40)	(2.62)	(2.77)	(2.86)	(2.78)	(2.08)	(1.58)	(0.53)	(-1.16)	(-3.70)	(-4.41)
Three-factor α	0.13	0.09	0.14	0.17	0.19	0.12	0.12	0.00	-0.23	-0.78	-0.91
	(1.91)	(1.30)	(1.97)	(2.71)	(2.79)	(1.82)	(1.68)	(0.06)	(-2.90)	(-6.41)	(-6.05)
Four-factor α	0.11	0.13	0.17	0.21	0.22	0.16	0.19	0.08	-0.09	-0.53	-0.64
	(1.82)	(2.15)	(2.83)	(4.11)	(3.62)	(2.58)	(2.48)	(1.20)	(-1.22)	(-4.65)	(-4.46)

Table 3. Characteristics of portfolios sorted by crash probability

This table reports various firm characteristics for each decile portfolio sorted on the predicted probability of price crashes. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*). Then, we calculate the 25% winsorized means of various characteristics for each decile in each month and report their time-series averages. The firm characteristic variables are the crash probability (*CRASHP*), jackpot probability (*JACKPOTP*), failure probability (*FAILP*), idiosyncratic volatility (*IVOL*), idiosyncratic skewness (*ISKEW*), the maximum daily return in month *t* (*MAX*), the minimum daily return in month *t* (*MIN*), the price per share (*PRC*), institutional ownership (*IO*), analyst coverage (*COVER*), analyst forecast dispersion (*DISP*), the market beta (*BETA*), the log of market capitalization (*SIZE*), the book-to-market ratio (*BM*), the return in month *t* (*REV*), the return over months $t-6$ to $t-1$ (*MOM*), illiquidity (*ILLIQ*), and share turnover (*TURN*). The sample period is January 1972 to December 2015 for all the variables but *FAILP,* which begins August 1972; *IO,* which begins April 1980; and *COVER* and *DISP*, which begin January 1976.

	1	2	3	4	5	6	7	8	9	10
CRASHP	0.006	0.014	0.021	0.029	0.038	0.048	0.060	0.078	0.106	0.182
JACKPOTP	0.010	0.018	0.023	0.028	0.032	0.038	0.043	0.051	0.061	0.080
FAILP (%)	0.031	0.031	0.032	0.033	0.035	0.037	0.040	0.045	0.053	0.086
IVOL	1.131	1.330	1.495	1.637	1.785	1.953	2.158	2.414	2.769	3.492
ISKEW	0.187	0.254	0.301	0.321	0.351	0.379	0.402	0.428	0.471	0.598
MAX	2.972	3.472	3.862	4.218	4.603	5.034	5.580	6.259	7.210	9.187
MIN	2.694	3.090	3.410	3.684	3.975	4.315	4.728	5.244	5.941	7.259
PRC	33.137	27.256	24.504	22.696	21.557	19.729	17.873	16.291	14.349	11.880
IO	0.517	0.505	0.507	0.510	0.511	0.497	0.478	0.457	0.423	0.357
COVER	2.474	2.126	1.987	1.914	1.877	1.835	1.787	1.734	1.701	1.633
DISP	0.043	0.048	0.051	0.050	0.050	0.054	0.058	0.065	0.079	0.114
BETA	0.669	0.691	0.713	0.757	0.816	0.878	0.966	1.072	1.200	1.358
SIZE	7.204	6.258	5.880	5.648	5.536	5.384	5.254	5.114	4.949	4.776
BM	0.897	0.850	0.817	0.802	0.751	0.723	0.691	0.652	0.602	0.538
REV	0.925	0.878	0.920	0.983	1.111	1.145	1.243	1.299	1.348	1.776
MOM	6.380	6.009	6.712	7.065	7.843	8.300	8.691	8.928	8.315	7.533
ILLIQ	0.032	0.077	0.140	0.216	0.258	0.293	0.319	0.372	0.419	0.514
TURN	0.062	0.058	0.061	0.067	0.075	0.083	0.093	0.107	0.124	0.166

Table 4. Returns on portfolios sorted by crash probability after controlling for other characteristics

This table reports the risk-adjusted returns on quintile portfolios sorted on the predicted probability of price crashes after controlling for each of various firm characteristics. At the end of each month t, stocks are sorted into quintile, based on a control variable, and then, within each quintile, the stocks are sorted into quintile portfolios based on crash probability (*CRASHP*). The monthly value-weighted portfolio returns on each intersection of the two sorts realized in month $t + 2$ are calculated and the returns on each of the controlled *CRASHP* quintiles are defined as the average returns across the control quintiles. The table shows the portfolio returns adjusted for the risk factors of the Fama–French three-factor model. The control variables include jackpot probability (*JACKPOTP*), failure probability (*FAILP*), idiosyncratic volatility (*IVOL*), idiosyncratic skewness (*ISKEW*), the maximum daily return in month t (*MAX*), the minimum daily return in month t (*MIN*), the price per share (*PRC*), residual institutional ownership (*RIO*), residual analyst coverage (*RCOVER*), analyst forecast dispersion (*DISP*), the market beta (*BETA*), the log of market capitalization (*SIZE*), the book-to-market ratio (*BM*), the return in month t (*REV*), the return over months $t-6$ to $t-1$ (*MOM*), illiquidity (*ILLIQ*), and share turnover (*TURN*). The numbers in parentheses are t-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015 for all the control variables but *FAILP,* which begins October 1972; *RIO*, which begins June 1980; and *RCOVER* and *DISP*, which begin March 1976.

	Ranking on *CRASHP*					
	1	2	3	4	5	5-1
Controlling for						
JACKPOTP	0.16	0.09	0.02	-0.09	-0.34	-0.50
	(2.04)	(1.86)	(0.43)	(-1.52)	(-3.55)	(-4.24)
FAILP	-0.07	-0.14	-0.23	-0.19	-0.29	-0.22
	(-1.13)	(-1.93)	(-2.64)	(-2.26)	(-1.99)	(-1.27)
IVOL	0.01	-0.13	-0.09	-0.07	-0.29	-0.30
	(0.16)	(-2.06)	(-1.48)	(-0.84)	(-2.84)	(-2.51)
ISKEW	0.06	0.10	0.11	0.01	-0.47	-0.53
	(1.33)	(2.02)	(2.00)	(0.07)	(-3.23)	(-3.06)
MAX	0.06	-0.08	0.06	-0.09	-0.29	-0.34
	(1.11)	(-1.83)	(0.94)	(-1.25)	(-2.65)	(-2.81)
MIN	0.01	0.00	-0.06	-0.11	-0.30	-0.31
	(0.20)	(-0.04)	(-1.14)	(-1.41)	(-2.89)	(-2.59)
PRC	-0.07	-0.06	-0.10	-0.27	-0.57	-0.50
	(-1.07)	(-0.72)	(-1.47)	(-3.48)	(-4.97)	(-3.62)
RIO	0.05	0.01	0.07	-0.09	-0.41	-0.47
	(0.81)	(0.25)	(1.24)	(-1.09)	(-2.56)	(-2.31)
RCOVER	0.09	-0.02	0.07	0.00	-0.37	-0.46
	(1.77)	(-0.34)	(1.22)	(0.01)	(-2.40)	(-2.49)
DISP	-0.01	-0.05	-0.05	-0.01	-0.37	-0.36
	(-0.20)	(-0.81)	(-0.88)	(-0.16)	(-2.53)	(-2.16)
BETA	0.06	-0.02	-0.07	-0.01	-0.46	-0.52
	(1.42)	(-0.56)	(-1.44)	(-0.18)	(-4.07)	(-3.87)
SIZE	0.17	0.12	0.10	-0.01	-0.53	-0.70
	(2.79)	(2.10)	(1.85)	(-0.27)	(-5.67)	(-5.33)
BM	0.08	0.01	0.05	-0.10	-0.36	-0.44
	(1.45)	(0.21)	(1.05)	(-1.50)	(-2.74)	(-2.80)
REV	0.10	0.04	0.04	-0.05	-0.42	-0.52
	(2.04)	(0.69)	(0.81)	(-0.78)	(-3.05)	(-3.12)
MOM	0.03	-0.06	-0.01	-0.04	-0.53	-0.56
	(0.65)	(-1.17)	(-0.32)	(-0.64)	(-4.31)	(-3.99)
ILLIQ	0.11	0.13	0.05	-0.03	-0.48	-0.59
	(1.82)	(2.25)	(0.97)	(-0.63)	(-5.05)	(-4.38)
TURN	0.11	0.08	0.04	-0.03	-0.46	-0.57
	(2.30)	(1.59)	(0.83)	(-0.52)	(-4.65)	(-5.13)

Table 5. Firm-level cross-sectional regressions

This table reports the time-series averages of the coefficients from the cross-sectional regressions of stock returns on the lagged values of various characteristics. The variable *CRASHP* and *JACKPOTP* are the probabilities of extreme negative and extreme positive returns over the next 12 months, respectively, predicted out-of-sample from the generalized logit model in Equation (1). The other firm characteristics include failure probability (*FAILP*), idiosyncratic volatility (*IVOL*), idiosyncratic skewness (*ISKEW*), the maximum daily return in month t (*MAX*), the minimum daily return in month t (*MIN*), analyst forecast dispersion (*DISP*), the market beta (*BETA*), the log of market capitalization (*SIZE*), the book-to-market ratio (*BM*), the return in month t (*REV*), the return over months $t-6$ to $t-1$ (*MOM*), illiquidity (*ILLIQ*), and share turnover (*TURN*). The numbers in parentheses are t-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015 for all control variables but *FAILP*, which begins September 1972, and *DISP,* which begins February 1976.

CRASHP	*JACKPOTP*	*FAILP*	*IVOL*	*ISKEW*	*MAX*	*MIN*	*DISP*	*BETA*	*SIZE*	*BM*	*REV*	*MOM*	*ILLIQ*	*TURN*
-6.23														
(-4.43)														
-6.82								0.02	-0.12	0.07	-0.03	0.01	-0.03	0.86
(-6.32)								(0.56)	(-3.10)	(1.62)	(-5.92)	(4.58)	(-3.15)	(1.72)
	-4.71													
	(-1.63)													
	-9.87							0.01	-0.15	0.12	-0.03	0.01	-0.02	-0.11
	(-3.31)							(0.34)	(-3.97)	(2.28)	(-5.84)	(4.95)	(-2.05)	(-0.27)
-8.63	6.42													
(-6.64)	(2.32)													
-6.52	0.48							0.02	-0.10	0.07	-0.03	0.01	-0.04	0.74
(-5.48)	(0.13)							(0.65)	(-2.49)	(1.59)	(-6.05)	(4.25)	(-3.15)	(1.56)
		-0.29												
		(-0.45)												
		-2.68						-0.01	-0.08	0.17	-0.03	0.01	-0.04	0.12
		(-2.74)						(-0.21)	(-1.90)	(2.61)	(-5.88)	(2.91)	(-3.72)	(0.11)
-5.59		0.18												
(-4.10)		(0.31)												
-5.74		-2.12						0.02	-0.13	0.10	-0.03	0.01	-0.03	0.74
(-5.46)		(-2.78)						(0.53)	(-3.24)	(2.01)	(-5.92)	(3.12)	(-3.19)	(0.93)
			-0.24											
			(-4.67)											
			-0.24					0.01	-0.11	0.12	-0.03	0.01	-0.02	0.94
			(-6.51)					(0.31)	(-3.17)	(2.19)	(-5.23)	(4.54)	(-2.10)	(1.92)
-4.61			-0.12											
(-3.75)			(-3.51)											
-5.64			-0.13					0.02	-0.14	0.07	-0.03	0.01	-0.02	1.52
(-5.67)			(-4.83)					(0.80)	(-3.83)	(1.55)	(-5.54)	(4.55)	(-2.27)	(2.75)

(*continued*)

(Table 5 *continued*)

CRASHP	*JACKPOTP*	*FAILP*	*IVOL*	*ISKEW*	*MAX*	*MIN*	*DISP*	*BETA*	*SIZE*	*BM*	*REV*	*MOM*	*ILLIQ*	*TURN*
				0.04										
				(1.66)										
				0.00				-0.01	-0.06	0.16	-0.03	0.01	-0.04	-0.67
				(-0.17)				(-0.33)	(-1.41)	(2.59)	(-5.84)	(4.36)	(-3.80)	(-1.37)
-6.46				0.05										
(-4.61)				(2.30)										
-6.86				0.01				0.02	-0.12	0.07	-0.03	0.01	-0.03	0.91
(-6.38)				(0.32)				(0.54)	(-3.06)	(1.62)	(-5.95)	(4.37)	(-3.14)	(1.79)
					-0.07	-0.01								
					(-5.44)	(-0.86)								
					-0.02	-0.10		0.03	-0.11	0.12	-0.04	0.01	-0.02	1.11
					(-2.07)	(-8.94)		(1.02)	(-2.94)	(2.19)	(-7.22)	(4.53)	(-2.22)	(2.22)
-4.92					-0.05	0.02								
(-4.03)					(-5.45)	(1.12)								
-5.63					0.00	-0.08		0.04	-0.14	0.07	-0.04	0.01	-0.02	1.73
(-5.55)					(0.62)	(-8.24)		(1.29)	(-3.79)	(1.52)	(-7.46)	(4.52)	(-2.27)	(3.05)
							-0.18							
							(-1.91)							
							-0.22	-0.02	-0.08	0.07	-0.02	0.01	-0.12	-0.76
							(-2.08)	(-0.39)	(-1.97)	(1.07)	(-4.89)	(4.29)	(-2.53)	(-1.52)
-4.22							-0.16							
(-2.62)							(-1.76)							
-5.35							-0.22	0.00	-0.13	-0.01	-0.03	0.01	-0.09	0.18
(-4.29)							(-2.00)	(-0.01)	(-3.20)	(-0.26)	(-5.30)	(3.96)	(-1.97)	(0.44)

Table 6. Crash probability effects and limits to arbitrage

This table reports the returns on portfolios sorted on the predicted probability of price crashes for subgroups classified by limits to arbitrage. For firm size (*SIZE*), at the end of each month *t*, stocks are sorted into terciles, based on *SIZE*, and the smallest and largest terciles are classified as the small and big subgroups, respectively. For other limits-to-arbitrage variables, stocks are sorted into terciles based on *SIZE* first and then, within each size tercile, the stocks are sorted into terciles based on their price per share (*PRC*), illiquidity (*ILLIQ*), analyst coverage (*COVER*), or institutional ownership (*IO*). Stocks contained in the lowest *PRC* (*ILLIQ*, *COVER*, or *IO*) terciles within each size tercile are classified as the low price (liquid, low analyst coverage, or low institutional ownership) subgroup. Stocks contained in the highest *PRC* (*ILLIQ*, *COVER*, or *IO*) terciles within each size tercile are classified as the high price (illiquid, high analyst coverage, or high institutional ownership) subgroup. Then, within each of the 10 subgroups, quintile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*) at the end of month *t*. The monthly value-weighted returns on each portfolio realized in month $t + 2$ are calculated and the mean returns, CAPM alphas, Fama–French three-factor alphas, and Carhart four-factor alphas are reported. Panel A reports the returns on the zero-cost portfolio buying the highest and selling the lowest *CRASHP* quintile for each subgroup. Panels B and C report returns on the highest and lowest *CRASHP* quintiles for each subgroup, respectively. The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015 for all the limits-to-arbitrage variables but *COVER*, which begins March 1976; and *IO*, which begins June 1980.

	Firm size		Price		Liquidity		Analyst coverage		Institutional ownership	
	Small	Big	Low	High	Illiquid	Liquid	Low	High	Low	High
Panel A: Highest − lowest *CRASHP* quintile										
Mean	-0.75	-0.16	-0.89	0.12	-0.54	-0.24	-0.42	-0.22	-0.36	-0.39
	(-2.73)	(-0.55)	(-2.39)	(0.41)	(-2.04)	(-0.73)	(-1.28)	(-0.64)	(-0.78)	(-1.27)
CAPM α	-1.12	-0.56	-1.35	-0.24	-0.88	-0.72	-0.92	-0.77	-1.07	-0.81
	(-4.55)	(-2.25)	(-4.01)	(-0.95)	(-3.79)	(-2.55)	(-3.27)	(-2.35)	(-2.75)	(-2.79)
Three-factor α	-1.01	-0.21	-0.98	0.12	-0.61	-0.32	-0.64	-0.46	-0.64	-0.49
	(-5.27)	(-1.26)	(-4.87)	(0.59)	(-4.05)	(-1.60)	(-3.16)	(-2.18)	(-2.17)	(-2.23)
Four-factor α	-0.87	-0.05	-0.77	0.14	-0.44	-0.08	-0.61	-0.08	-0.48	-0.25
	(-4.68)	(-0.31)	(-3.93)	(0.79)	(-3.17)	(-0.45)	(-3.23)	(-0.39)	(-1.61)	(-1.32)

(*continued*)

(Table 6 *continued*)

	Firm size		Price		Liquidity		Analyst coverage		Institutional ownership	
	Small	Big	Low	High	Illiquid	Liquid	Low	High	Low	High
Panel B: Highest *CRASHP* quintile										
Mean	0.55	0.82	0.19	1.05	0.62	0.69	0.74	0.75	0.63	0.65
	(1.49)	(2.21)	(0.43)	(2.81)	(1.80)	(1.67)	(1.92)	(1.78)	(1.21)	(1.62)
CAPM α	-0.60	-0.38	-1.10	-0.11	-0.54	-0.60	-0.63	-0.65	-0.83	-0.74
	(-2.39)	(-1.95)	(-4.01)	(-0.51)	(-2.90)	(-2.52)	(-2.89)	(-2.30)	(-2.44)	(-3.24)
Three-factor α	-0.75	-0.08	-0.99	0.25	-0.52	-0.23	-0.53	-0.37	-0.49	-0.53
	(-5.39)	(-0.62)	(-5.45)	(1.46)	(-4.36)	(-1.35)	(-3.44)	(-1.96)	(-1.75)	(-2.93)
Four-factor α	-0.59	0.02	-0.59	0.17	-0.32	-0.06	-0.54	-0.03	-0.33	-0.33
	(-4.19)	(0.17)	(-3.43)	(1.14)	(-3.04)	(-0.34)	(-3.72)	(-0.14)	(-1.18)	(-2.04)
Panel C: Lowest *CRASHP* quintile										
Mean	1.30	0.97	1.08	0.93	1.16	0.92	1.16	0.97	0.98	1.04
	(6.15)	(5.74)	(5.45)	(5.41)	(6.32)	(5.33)	(6.47)	(5.61)	(5.92)	(4.61)
CAPM α	0.52	0.18	0.25	0.13	0.34	0.12	0.29	0.12	0.24	0.07
	(3.27)	(2.75)	(1.82)	(1.93)	(2.69)	(1.95)	(2.27)	(1.72)	(2.61)	(0.64)
Three-factor α	0.25	0.13	-0.01	0.13	0.09	0.08	0.12	0.09	0.16	-0.04
	(2.59)	(2.57)	(-0.16)	(2.04)	(1.18)	(1.87)	(1.38)	(1.90)	(2.26)	(-0.41)
Four-factor α	0.28	0.07	0.18	0.03	0.11	0.03	0.07	0.05	0.16	-0.08
	(3.17)	(1.39)	(2.01)	(0.43)	(1.60)	(0.64)	(0.93)	(1.14)	(1.88)	(-0.89)

Table 7. Crash probability effects across different sub-periods

This table reports the average risk-adjusted returns on decile portfolios sorted on the predicted probability of price crashes following sub-periods classified by several binary state indicators. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*) and the monthly value-weighted portfolio returns on each decile realized in month *t* + 2 are calculated. The average risk-adjusted returns following sub-periods are the estimates of α_1 and α_2 in the regression

$$r_t - r_f = \alpha_1 D_{1,t-1} + \alpha_2 D_{2,t-1} + \beta MKT_t + \gamma SMB_t + \delta HML_t + \varepsilon_t \ ,$$

where $D_{1,t-1}$ and $D_{2,t-1}$ are dummy variables indicating each of the two states in month $t-1$ and r_t is the portfolio return in month *t*. Panel A shows the results for high- and low-investor sentiment periods classified by the median value of the sentiment index of Baker and Wurgler (2006). Panel B shows the results for expansion and recession periods classified by the NBER recession indicator. Panel C shows the results for up- and down-market periods classified based on the sign of the CRSP value-weighted return in excess of the one-month T-bill rate. Panel D shows the results for high- and low-liquidity periods classified by the sign of the liquidity innovation of Pastor and Stambaugh (2003). The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to October 2015 in Panel A, January 1972 to September 2014 in Panel B, and January 1972 to December 2015 in Panels C and D.

	1	2	3	4	5	6	7	8	9	10	10-1
Panel A: High- vs. low-sentiment periods											
High sentiment	0.19	0.01	0.13	0.00	0.14	0.14	0.05	-0.17	-0.34	-0.82	-1.01
	(2.96)	(0.15)	(1.44)	(-0.04)	(1.36)	(1.08)	(0.47)	(-0.97)	(-1.96)	(-2.98)	(-3.37)
Low sentiment	0.10	-0.17	-0.07	0.12	0.12	0.06	-0.08	0.20	-0.14	-0.46	-0.57
	(1.55)	(-2.24)	(-0.90)	(1.35)	(1.36)	(0.51)	(-0.70)	(1.21)	(-0.88)	(-1.95)	(-2.07)
Difference	0.08	0.18	0.20	-0.12	0.02	0.08	0.13	-0.37	-0.21	-0.36	-0.44
	(0.98)	(1.83)	(1.83)	(-0.85)	(0.13)	(0.52)	(0.82)	(-1.58)	(-0.90)	(-1.07)	(-1.18)
Panel B: Expansion vs. recession periods											
Expansion	0.14	-0.09	0.02	0.08	0.15	0.09	0.01	0.06	-0.29	-0.56	-0.70
	(2.60)	(-1.46)	(0.29)	(1.02)	(2.17)	(0.91)	(0.07)	(0.49)	(-2.47)	(-2.68)	(-3.00)
Recession	0.20	-0.04	0.09	-0.03	0.04	0.19	-0.08	-0.23	-0.13	-0.97	-1.17
	(1.51)	(-0.21)	(0.54)	(-0.18)	(0.20)	(0.96)	(-0.40)	(-0.59)	(-0.41)	(-1.92)	(-2.03)
Difference	-0.06	-0.05	-0.07	0.10	0.11	-0.10	0.09	0.29	-0.16	0.41	0.46
	(-0.39)	(-0.21)	(-0.40)	(0.64)	(0.52)	(-0.46)	(0.42)	(0.76)	(-0.51)	(0.77)	(0.77)
Panel C: Up- vs. down-market periods											
Up market	0.12	-0.06	-0.03	0.19	0.15	0.02	0.02	-0.02	-0.39	-0.65	-0.78
	(1.88)	(-0.81)	(-0.30)	(1.88)	(1.87)	(0.19)	(0.17)	(-0.14)	(-2.70)	(-2.84)	(-2.99)
Down market	0.18	-0.11	0.09	-0.14	0.10	0.19	-0.07	0.06	-0.01	-0.62	-0.80
	(2.54)	(-0.99)	(0.88)	(-1.22)	(0.88)	(1.89)	(-0.51)	(0.35)	(-0.07)	(-2.37)	(-2.65)
Difference	-0.05	0.06	-0.11	0.33	0.05	-0.17	0.09	-0.08	-0.38	-0.03	0.02
	(-0.58)	(0.40)	(-0.89)	(1.93)	(0.36)	(-1.10)	(0.48)	(-0.44)	(-1.61)	(-0.11)	(0.06)
Panel D: High- vs. low-liquidity periods											
High liquidity	0.15	-0.06	-0.02	0.10	0.10	0.01	0.03	-0.01	-0.23	-0.75	-0.89
	(2.46)	(-0.94)	(-0.28)	(1.00)	(1.20)	(0.12)	(0.23)	(-0.11)	(-1.53)	(-3.71)	(-4.39)
Low liquidity	0.15	-0.10	0.08	-0.01	0.16	0.21	-0.08	0.05	-0.23	-0.50	-0.65
	(1.61)	(-1.02)	(0.64)	(-0.13)	(1.46)	(1.61)	(-0.67)	(0.30)	(-1.09)	(-1.74)	(-1.82)
Difference	0.00	0.04	-0.10	0.11	-0.05	-0.19	0.10	-0.07	0.00	-0.25	-0.25
	(-0.02)	(0.36)	(-0.65)	(0.74)	(-0.39)	(-1.36)	(0.63)	(-0.40)	(0.00)	(-0.86)	(-0.70)

Table 8. Crash probability effects, institutional ownership, and investor sentiment

This table reports the average risk-adjusted returns on portfolios sorted by the predicted probability of price crashes for subgroups classified by institutional ownership, following high- and low-investor sentiment periods. At the end of each month *t*, stocks are sorted into terciles, based on firm size (*SIZE*) first, and then, within each size tercile, the stocks are sorted into terciles based on institutional ownership (*IO*). The stocks contained in the lowest and highest *IO* terciles within each size tercile are classified as the low-institutional-ownership and high-institutional-ownership subgroups, respectively. Then, within each subgroup, quintile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*) and the monthly value-weighted returns on each portfolio realized in month $t + 2$ are calculated. The average risk-adjusted returns following high- and low-sentiment periods are the estimates of α_1 and α_2 in the regression

$$r_t - r_f = \alpha_1 D_{1,t-1} + \alpha_2 D_{2,t-1} + \beta MKT_t + \gamma SMB_t + \delta HML_t + \varepsilon_t \ ,$$

where $D_{1,t-1}$ and $D_{2,t-1}$ are dummy variables indicating whether the level of the sentiment index of Baker and Wurgler (2006) in month $t - 1$ is above and below the median, respectively, and r_t is the portfolio return in month *t*. The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is June 1980 to October 2015.

	Low institutional ownership			High institutional ownership		
	Lowest *CRASHP* quintile	Highest *CRASHP* quintile	Highest − lowest	Lowest *CRASHP* quintile	Highest *CRASHP* quintile	Highest − lowest
High sentiment	0.19	-0.76	-0.95	0.03	-0.48	-0.51
	(1.76)	(-2.52)	(-2.96)	(0.26)	(-2.05)	(-1.82)
Low sentiment	0.09	-0.07	-0.17	-0.13	-0.58	-0.46
	(0.97)	(-0.20)	(-0.41)	(-0.81)	(-2.04)	(-1.23)
Difference	0.09	-0.69	-0.78	0.15	0.10	-0.06
	(0.61)	(-1.80)	(-1.89)	(0.91)	(0.27)	(-0.12)

Table 9. Jackpot effects and crash probability

This table reports the returns on portfolios sorted by the predicted probability of jackpot returns. Panel A reports the risk-adjusted returns on quintile portfolios sorted by jackpot probability (*JACKPOTP*). Panel B reports the risk-adjusted returns on quintile portfolios sorted by *JACKPOTP* after controlling for crash probability (*CRASHP*). Panel C reports the risk-adjusted returns on double-sorted portfolios based on *CRASHP* and *JACKPOTP*. Panel D reports the average risk-adjusted returns on quintile portfolios sorted by *JACKPOTP* following periods of high and low investor sentiment. In Panel A, at the end of each month *t*, quintile portfolios are constructed by sorting stocks based on *JACKPOTP* and the monthly value-weighted portfolio returns on each quintile realized in month *t* + 2 are calculated. In Panels B and C, at the end of each month *t*, stocks are sorted into quintiles based on *CRASHP* and then, within each quintile, the stocks are sorted into quintile portfolios based on *JACKPOTP*. The monthly value-weighted portfolio returns on each intersection of the two sorts realized in month *t* + 2 are calculated. In Panel B, the returns on each of the controlled *JACKPOTP* quintiles are defined as the average returns across the *CRASHP* quintiles. In Panel C, the returns on each intersection of the two quintile sorts are reported. Panels A to C show the portfolio returns adjusted for the risk factors of the Fama–French three-factor model. In Panel D, the average risk-adjusted returns following periods of high and low sentiment are the estimates of α_1 and α_2 in the regression

$$r_t - r_f = \alpha_1 D_{1,t-1} + \alpha_2 D_{2,t-1} + \beta MKT_t + \gamma SMB_t + \delta HML_t + \varepsilon_t \ ,$$

where $D_{1,t-1}$ and $D_{2,t-1}$ are dummy variables indicating high- and low-sentiment periods in month $t-1$, respectively, classified by the median value of the sentiment index of Baker and Wurgler (2006), and r_t is the value-weighted return on each of the *JACKPOTP*-sorted quintiles in month *t*. The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015 for all the panels but Panel D, which ends October 2015.

	Ranking on *JACKPOTP*					
	1	2	3	4	5	5-1
Panel A: Value-weighted portfolios sorted by *JACKPOTP*						
Three-factor α	0.05	0.09	-0.09	-0.17	-0.42	-0.48
	(1.97)	(1.54)	(-1.23)	(-1.38)	(-2.58)	(-2.62)
Panel B: Value-weighted portfolios sorted by *JACKPOTP* controlling for *CRASHP*						
Three-factor α	-0.03	-0.03	-0.08	-0.06	0.04	0.07
	(-0.65)	(-0.50)	(-1.45)	(-0.91)	(0.46)	(0.77)
Panel C: Three-factor alphas on value-weighted portfolios sorted by *JACKPOTP* within *CRASHP* subgroups						
Low *CRASHP*	0.08	0.09	-0.06	0.04	0.01	-0.08
	(1.57)	(1.52)	(-0.80)	(0.51)	(0.08)	(-0.80)
2	0.04	0.02	0.11	0.14	0.19	0.15
	(0.67)	(0.26)	(1.33)	(1.63)	(1.77)	(1.35)
3	0.09	0.09	0.06	0.07	0.32	0.22
	(1.38)	(1.36)	(0.81)	(0.69)	(2.45)	(1.43)
4	-0.03	0.09	0.05	0.06	0.19	0.22
	(-0.32)	(0.85)	(0.48)	(0.53)	(1.37)	(1.55)
High *CRASHP*	-0.33	-0.44	-0.57	-0.62	-0.50	-0.17
	(-2.24)	(-2.66)	(-2.88)	(-3.25)	(-2.40)	(-0.90)
5-1	-0.41	-0.53	-0.51	-0.66	-0.50	
	(-2.34)	(-3.02)	(-2.03)	(-3.03)	(-2.16)	
Panel D: Three-factor alphas on portfolios sorted by *JACKPOTP* following high- vs. low-sentiment periods						
High sentiment	0.10	0.04	-0.20	-0.47	-0.85	-0.96
	(2.76)	(0.59)	(-1.73)	(-2.67)	(-4.33)	(-4.37)
Low sentiment	0.00	0.14	0.02	0.12	-0.01	-0.02
	(0.11)	(1.76)	(0.15)	(0.78)	(-0.07)	(-0.08)
Difference	0.10	-0.09	-0.22	-0.58	-0.84	-0.94
	(2.09)	(-0.95)	(-1.35)	(-2.75)	(-3.22)	(-3.21)

Table 10. Failure probability and crash probability

This table reports the returns on decile portfolios sorted on the residual predicted probability of price crashes and the residual failure probability. The residual crash probability (*RCRASHP*) is defined as the residuals from the cross-sectional regression of the predicted probability of price crashes (*CRASHP*) on the failure probability (*FAILP*) each month. The residual failure probability (*RFAILP*) is defined as the residuals from the cross-sectional regression of failure probability (*FAILP*) on crash probability (*CRASHP*) each month. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on each of *RCRASHP* and *RFAILP*, respectively. The monthly portfolio returns on each decile realized in month $t + 2$ are calculated and the mean returns, CAPM alphas, Fama–French three-factor alphas, and Carhart four-factor alphas are reported. Panel A reports the value-weighted returns on the *RCRASHP*-sorted portfolios and Panel B reports the value-weighted returns on the *RFAILP*-sorted portfolios. The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is October 1972 to December 2015.

	1	2	3	4	5	6	7	8	9	10	10-1
Panel A: Value-weighted portfolios sorted on *RCRASHP*											
Mean	1.03	0.99	0.90	0.95	1.04	1.14	0.98	0.91	0.77	0.48	-0.55
	(5.16)	(4.84)	(4.29)	(4.01)	(4.56)	(4.79)	(3.75)	(2.60)	(2.31)	(1.05)	(-1.48)
CAPM α	0.19	0.12	0.00	0.00	0.07	0.14	-0.07	-0.24	-0.41	-0.83	-1.02
	(1.98)	(1.17)	(-0.05)	(0.06)	(0.94)	(1.21)	(-0.71)	(-1.35)	(-2.30)	(-3.03)	(-3.17)
Three-factor α	0.10	0.04	-0.02	-0.01	0.12	0.20	0.03	-0.02	-0.21	-0.48	-0.58
	(1.05)	(0.40)	(-0.32)	(-0.15)	(1.41)	(1.87)	(0.28)	(-0.12)	(-1.58)	(-2.45)	(-2.38)
Four-factor α	0.16	0.09	-0.07	-0.07	0.10	0.24	0.06	0.02	-0.09	-0.28	-0.44
	(1.65)	(0.85)	(-0.95)	(-0.86)	(1.25)	(2.18)	(0.58)	(0.15)	(-0.66)	(-1.61)	(-2.14)
Panel B: Value-weighted portfolios sorted on *RFAILP*											
Mean	0.83	0.89	1.00	1.06	1.03	1.02	0.91	1.03	1.01	0.73	-0.10
	(2.04)	(2.90)	(3.51)	(4.12)	(4.44)	(4.44)	(4.23)	(4.48)	(4.05)	(2.52)	(-0.33)
CAPM α	-0.39	-0.21	-0.07	0.06	0.07	0.07	-0.01	0.15	0.11	-0.25	0.14
	(-1.50)	(-1.21)	(-0.49)	(0.54)	(0.85)	(1.00)	(-0.10)	(1.31)	(0.73)	(-1.68)	(0.43)
Three-factor α	-0.03	0.06	0.11	0.20	0.10	0.10	-0.05	0.07	-0.01	-0.45	-0.42
	(-0.17)	(0.41)	(0.79)	(1.93)	(1.40)	(1.35)	(-0.57)	(0.62)	(-0.05)	(-3.37)	(-1.70)
Four-factor α	0.06	0.07	0.07	0.18	0.10	0.09	-0.10	0.09	0.13	-0.19	-0.25
	(0.34)	(0.58)	(0.55)	(1.74)	(1.07)	(1.24)	(-1.00)	(0.86)	(0.99)	(-1.37)	(-1.21)

Table 11. Returns on portfolios sorted by crash probability after controlling for anomaly variables

This table reports the risk-adjusted returns on quintile portfolios sorted on the predicted probability of price crashes after controlling for each of various anomaly variables. At the end of each month *t*, the stocks are sorted into quintile based on an anomaly variable and then, within each quintile, the stocks are sorted into quintile portfolios based on crash probability (*CRASHP*). The monthly value-weighted portfolio returns on each intersection of the two sorts realized in month *t* + 2 are calculated and the returns on each of the controlled *CRASHP* quintiles are defined as the average returns across the control quintiles. The table shows the portfolio returns adjusted for the risk factors of the Fama–French three-factor model. The anomaly variables include Ohlson's (1980) O-score (*OSCR*), the net stock issues (*NSI*) of Ritter (1991), the composite equity issues (*CEI*) of Daniel and Titman (2006), the total accruals (*TA*) of Sloan (1996), the net operating assets (*NOA*) of Hirshleifer et al. (2004), the gross profitability (*GP*) of Novy-Marx (2013), the asset growth (*AG*) of Cooper et al. (2008), the return on assets (*ROA*) of Fama and French (2006), and the abnormal capital investment (*ACI*) of Titman et al. (2004). The numbers in parentheses are *t*-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015.

	Ranking on *CRASHP*					
	1	2	3	4	5	5-1
Controlling for						
OSCR	0.07	0.07	0.06	0.02	-0.40	-0.47
	(1.27)	(1.23)	(1.19)	(0.27)	(-2.75)	(-2.73)
NSI	0.04	0.04	0.08	0.02	-0.26	-0.30
	(0.86)	(1.04)	(1.67)	(0.26)	(-2.46)	(-2.32)
CEI	0.00	0.01	0.11	0.12	-0.17	-0.17
	(-0.04)	(0.22)	(1.85)	(1.71)	(-2.06)	(-1.58)
TA	0.04	0.04	0.11	-0.03	-0.27	-0.30
	(0.65)	(0.96)	(2.14)	(-0.36)	(-1.78)	(-1.64)
NOA	0.10	0.05	0.17	-0.07	-0.36	-0.46
	(2.71)	(1.29)	(2.75)	(-0.93)	(-2.32)	(-2.74)
GP	0.04	-0.06	0.08	-0.05	-0.54	-0.58
	(1.06)	(-1.62)	(1.49)	(-0.49)	(-3.45)	(-3.36)
AG	0.08	0.04	0.09	-0.04	-0.16	-0.25
	(1.58)	(0.86)	(2.02)	(-0.54)	(-1.09)	(-1.36)
ROA	-0.05	-0.14	-0.09	-0.22	-0.59	-0.54
	(-0.93)	(-2.15)	(-1.39)	(-2.47)	(-4.83)	(-3.97)
ACI	0.06	-0.04	0.18	0.02	-0.23	-0.29
	(1.23)	(-0.82)	(3.41)	(0.19)	(-1.94)	(-2.01)

Table 12. Firm-level cross-sectional regressions with anomaly variables

This table reports the time-series averages of the coefficients from the cross-sectional regressions of stock returns on the lagged values of various anomaly variables and firm characteristics. The variable *CRASHP* is the probability of extreme negative returns over the next 12 months, predicted out-of-sample from the generalized logit model in Equation (1). The anomaly variables include Ohlson's (1980) O-score (*OSCR*), the net stock issues (*NSI*) of Ritter (1991), the composite equity issues (*CEI*) of Daniel and Titman (2006), the total accruals (*TA*) of Sloan (1996), the net operating assets (*NOA*) of Hirshleifer et al. (2004), the gross profitability (*GP*) of Novy-Marx (2013), the asset growth (*AG*) of Cooper et al. (2008), the return on assets (*ROA*) of Fama and French (2006), and the abnormal capital investment (*ACI*) of Titman et al. (2004). The other firm characteristics include the market beta (*BETA*), the log of market capitalization (*SIZE*), the book-to-market ratio (*BM*), the return in month t (*REV*), the return over months $t-6$ to $t-1$ (*MOM*), illiquidity (*ILLIQ*), and share turnover (*TURN*). The numbers in parentheses are t-statistics adjusted using Newey–West (1987) standard errors with 12 lags. The sample period is January 1972 to December 2015.

CRASHP	*OSCR*	*NSI*	*CEI*	*TA*	*NOA*	*GP*	*AG*	*ROA*	*ACI*	*BETA*	*SIZE*	*BM*	*REV*	*MOM*	*ILLIQ*	*TURN*
-6.40	0.02															
(-3.90)	(0.78)															
-6.84	-0.02									0.02	-0.10	0.07	-0.03	0.01	-0.03	0.88
(-5.18)	(-1.17)									(0.57)	(-2.67)	(1.23)	(-6.62)	(3.78)	(-2.62)	(1.15)
-5.70		-1.36														
(-3.51)		(-3.59)														
-6.40		-1.16								0.02	-0.10	0.08	-0.03	0.01	-0.03	0.88
(-4.91)		(-4.25)								(0.52)	(-2.51)	(1.35)	(-6.59)	(3.80)	(-2.92)	(1.13)
-5.21			-0.37													
(-3.42)			(-4.01)													
-6.05			-0.34							0.02	-0.10	0.07	-0.03	0.01	-0.03	1.11
(-4.79)			(-3.96)							(0.51)	(-2.46)	(1.19)	(-6.67)	(3.71)	(-3.04)	(1.41)
-5.97				-0.70												
(-3.57)				(-2.03)												
-6.60				-0.67						0.01	-0.10	0.08	-0.03	0.01	-0.03	0.87
(-4.97)				(-2.31)						(0.42)	(-2.49)	(1.31)	(-6.59)	(3.76)	(-2.70)	(1.11)
-6.19					-0.44											
(-3.81)					(-2.64)											
-6.88					-0.42					0.01	-0.10	0.08	-0.03	0.01	-0.03	0.99
(-5.22)					(-2.94)					(0.44)	(-2.53)	(1.49)	(-6.65)	(3.55)	(-2.61)	(1.24)
-6.46						0.50										
(-3.87)						(2.75)										
-6.83						0.46				0.01	-0.09	0.12	-0.03	0.01	-0.03	0.84
(-5.11)						(2.69)				(0.38)	(-2.19)	(1.95)	(-6.72)	(3.79)	(-2.70)	(1.03)

(*continued*)

(Table 12 *continued*)

CRASHP	*OSCR*	*NSI*	*CEI*	*TA*	*NOA*	*GP*	*AG*	*ROA*	*ACI*	*BETA*	*SIZE*	*BM*	*REV*	*MOM*	*ILLIQ*	*TURN*
-5.74							-0.21									
(-3.48)							(-1.47)									
-6.55							-0.16			0.02	-0.09	0.08	-0.03	0.01	-0.03	1.05
(-4.92)							(-1.59)			(0.52)	(-2.37)	(1.34)	(-6.63)	(3.81)	(-2.69)	(1.33)
-5.63								7.11								
(-3.51)								(3.89)								
-6.24								9.52		0.01	-0.11	0.14	-0.03	0.01	-0.03	0.94
(-4.87)								(4.80)		(0.47)	(-2.65)	(2.33)	(-6.83)	(3.11)	(-2.77)	(1.14)
-6.11									-0.09							
(-3.68)									(-3.05)							
-6.91									-0.08	0.02	-0.11	0.07	-0.03	0.01	-0.03	1.02
(-5.19)									(-3.69)	(0.52)	(-2.72)	(1.21)	(-6.58)	(3.81)	(-2.66)	(1.27)

Figure 1. Cumulative profit of trading strategies based on crash probability

This figure shows the cumulative profits of the trading strategies based on the predicted probability of price crashes. At the beginning of each month *t* from January 1972 to December 2015, decile portfolios are formed by sorting stocks based on crash probability (*CRASHP*) in month $t - 2$. Then, the cumulative value of the value-weighted buy-and-hold strategy on each decile is calculated, with an initial investment of $1 at the beginning of January 1972. P1 and P10 denote the lowest and highest *CRASHP* deciles, respectively, P1−P10 is the difference between the two portfolios, and RM denotes the CRSP value-weighted portfolio.

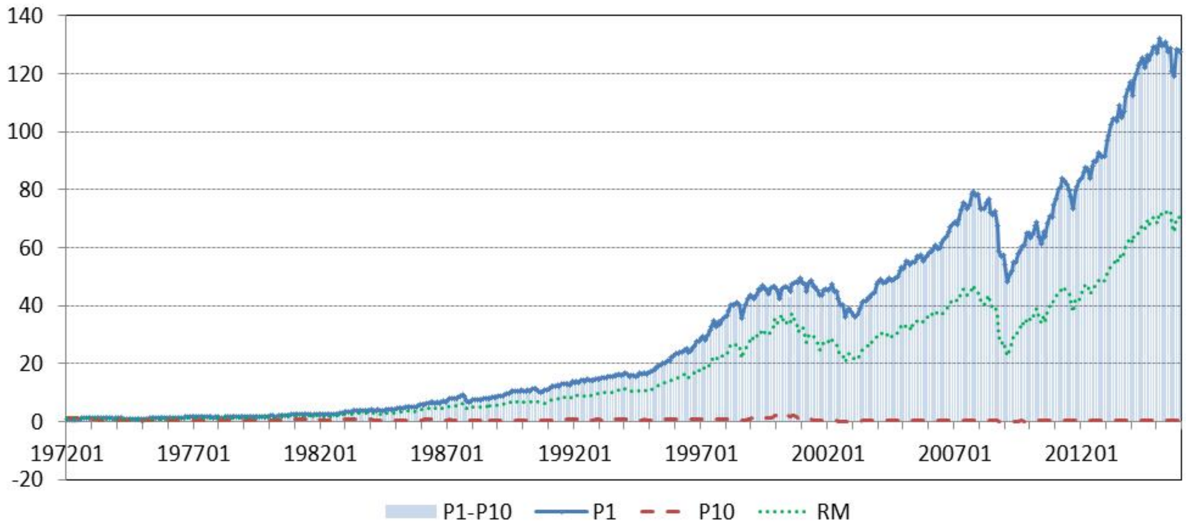
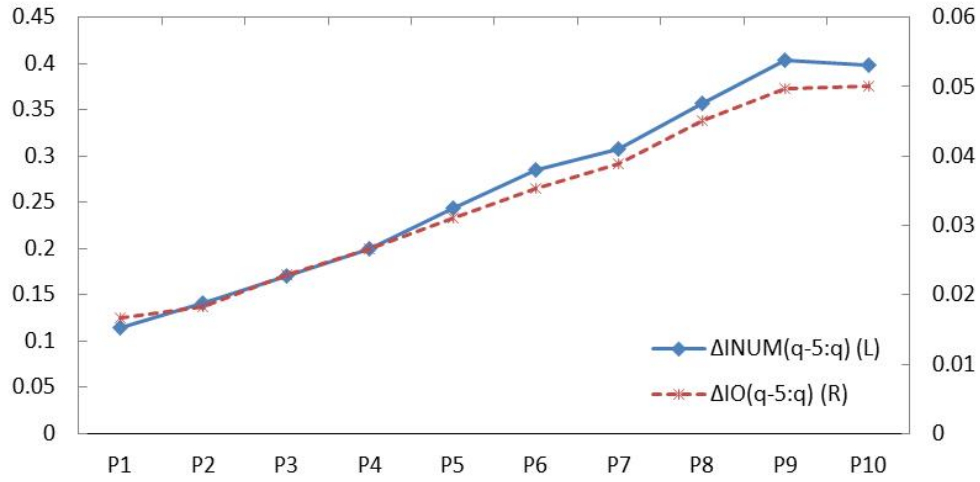


Figure 2. Crash probability effects and changes in institutional holdings

This figure shows the relation between changes in institutional holdings and the predicted probability of price crashes. At the end of each quarter, decile portfolios are constructed by sorting stocks based on crash probability (*CRASHP*). The number of institutions (*INUM*) is defined as the number of institutions holding a stock, scaled by the average number of institutions in the same market capitalization decile each quarter (or as of the beginning of the period over which changes in institutional holdings are measured), and institutional ownership (*IO*) is defined as the fraction of shares owned by institutions each quarter. In Panel A, we plot the time-series and cross-sectional average of changes in institutional holdings during the six quarters prior to portfolio formation for each decile portfolio, where changes in institutional holdings are measured by both changes in the number of institutions (Δ*INUM*, left axis) and changes in institutional ownership (Δ*IO*, right axis). In Panel B, for firms that enter into the highest *CRASHP* decile each quarter, we present the average number of institutions (*INUM*, left axis) and the average institutional ownership (*IO*, right axis) of the firms in excess of the mean of each measure for all firms that have institutional holdings data that quarter, for six quarters before and after the entry into the highest *CRASHP* decile. We only include those firms that have non-missing data for the 13-quarter window around the event. The sample period is September 1981 to December 2015.

Panel A: Changes in institutional holdings for *CRASHP*-sorted portfolios



Panel B: Institutional holdings around the entry into the highest *CRASHP* decile

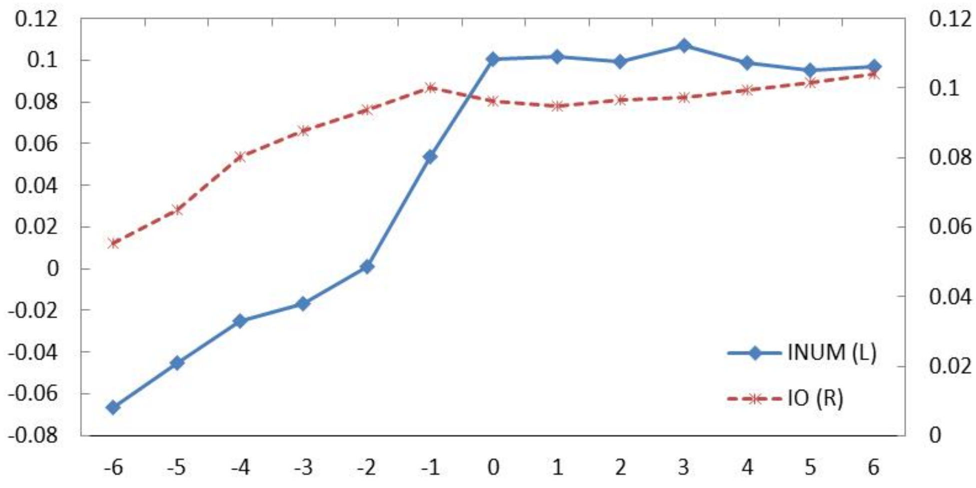


Figure 3. Loadings on Carhart's four factors for portfolios sorted by crash and jackpot probability

This figure plots the loadings on the factor portfolio returns in the Carhart four-factor model of decile portfolio returns sorted by the predicted probabilities of price crashes and jackpot returns. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on each of the crash probability (*CRASHP*) and the jackpot probability (*JACKPOTP*). The monthly value-weighted portfolio returns on each decile realized in month *t* + 2 are calculated and the monthly returns in excess of the one-month T-bill rate are regressed on the Carhart four risk factors of MKT, SMB, HML, and WML. The sample period is January 1972 to December 2015.

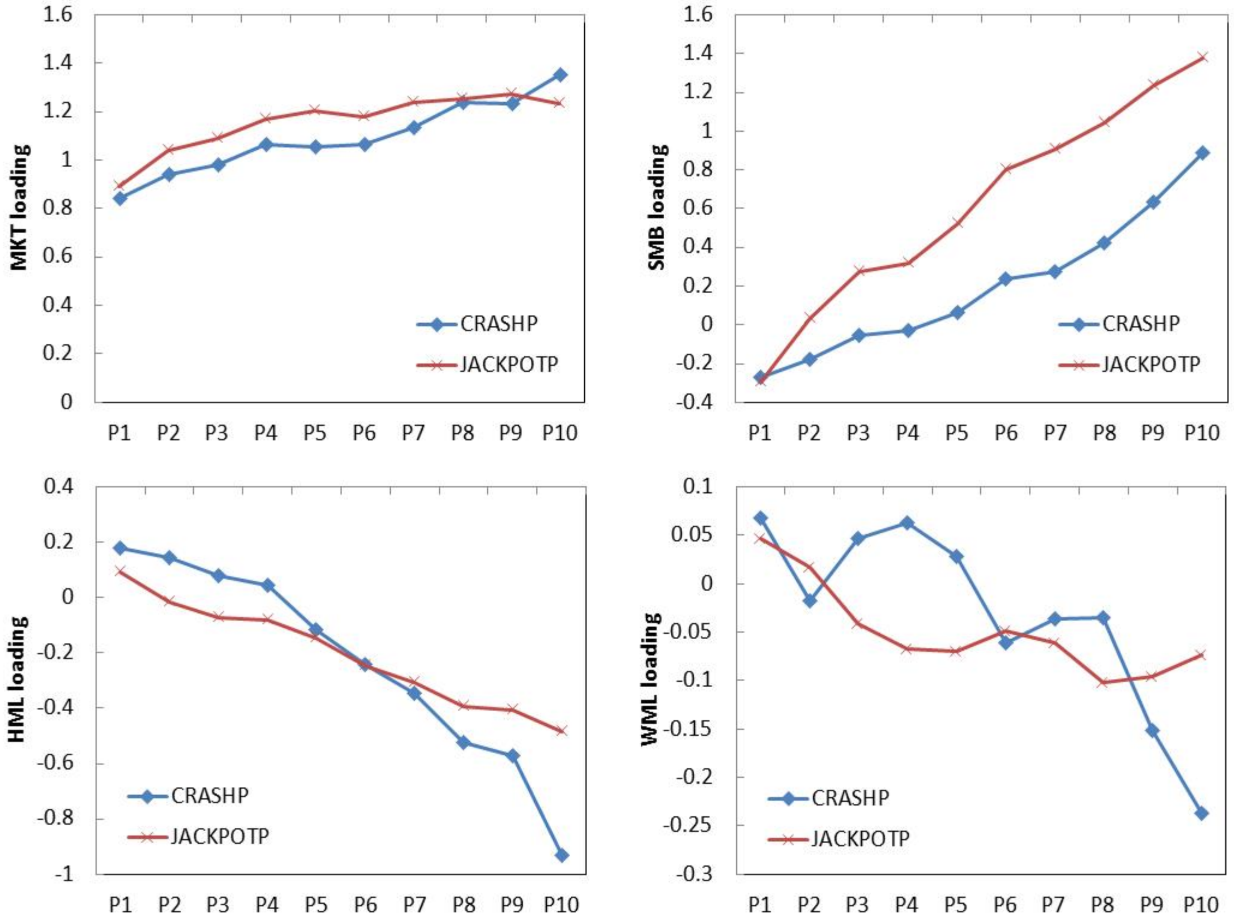


Figure 4. Loadings on Carhart's four factors for portfolios sorted by crash and failure probability

This figure plots the loadings on the factor portfolio returns in the Carhart four-factor model of decile portfolio returns sorted by the predicted probabilities of price crashes and failure. At the end of each month *t*, decile portfolios are constructed by sorting stocks based on each of the crash probability (*CRASHP*) and the failure probability (*FAILP*). The monthly value-weighted portfolio returns on each decile realized in month *t* + 2 are calculated and the monthly returns in excess of the one-month T-bill rate are regressed on the Carhart four risk factors of MKT, SMB, HML, and WML. The sample period is January 1972 to December 2015 for the *CRASHP*-sorted portfolios and October 1972 to December 2015 for the *FAILP*-sorted portfolios.

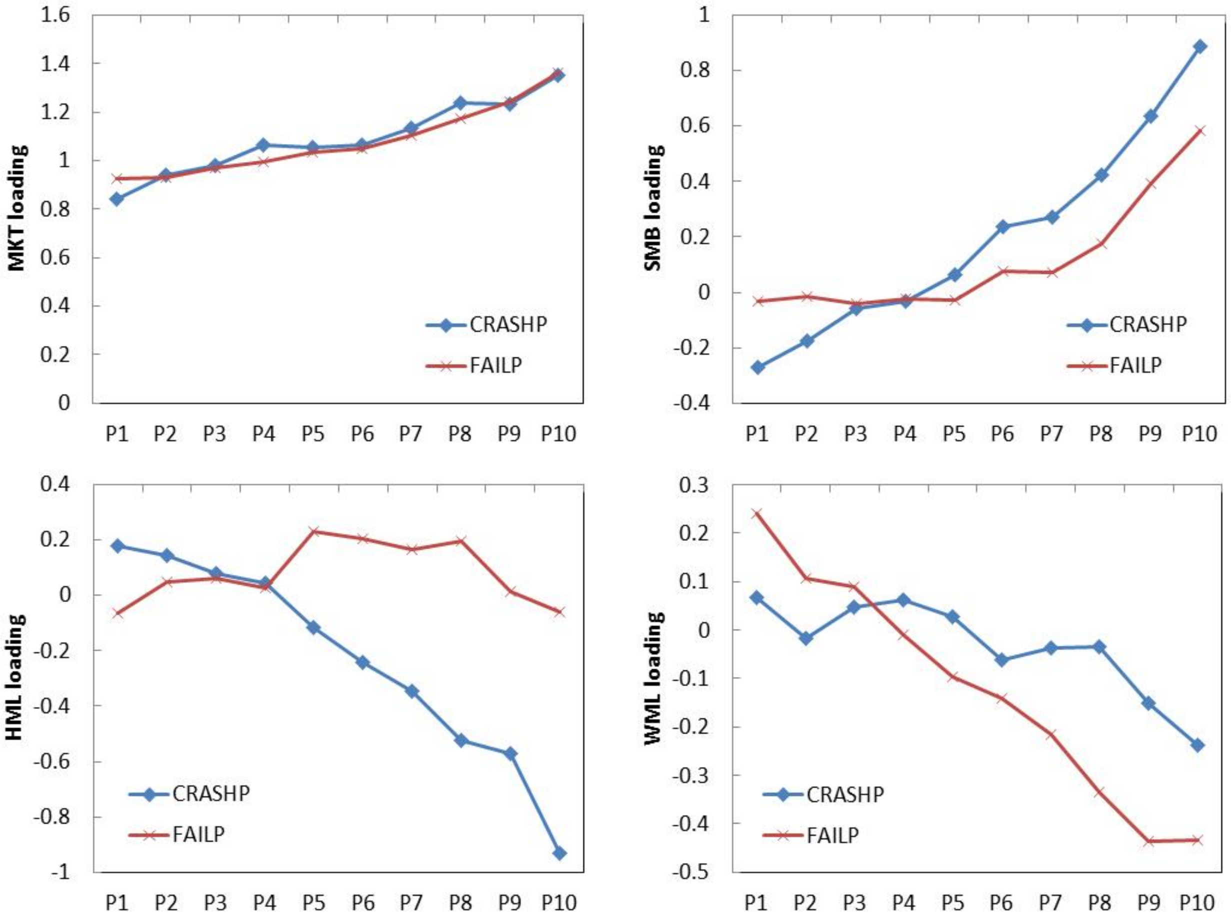
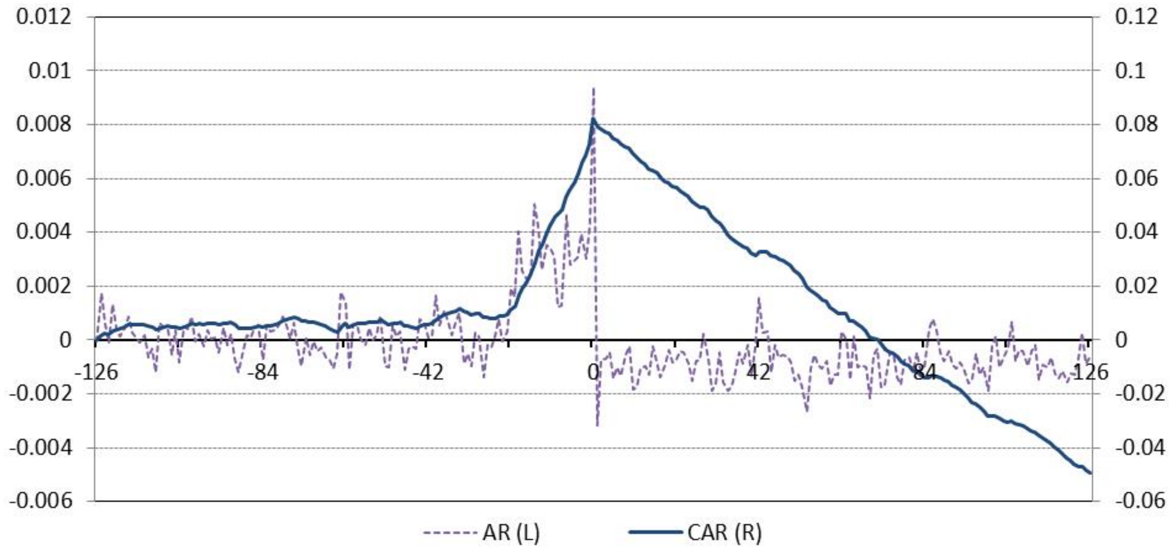


Figure 5. Cumulative abnormal returns around entry into the top decile portfolio sorted by crash and failure probability

This figure shows the daily abnormal returns and cumulative abnormal returns around the entry into the top decile portfolio sorted based on each of the predicted probabilities of price crashes and firms' failure. At the end of each month, decile portfolios are constructed by sorting stocks based on either the crash probability (*CRASHP*) or the failure probability (*FAILP*). The event of interest is defined as the entry of a firm into the highest *CRASHP* decile (Panel A) or *FAILP* decile (Panel B). We present the abnormal returns (AR, left axis) and cumulative abnormal returns (CAR, right axis) during an event window of 253 trading days, comprised of 126 pre-event days, the event day, and 126 post-event days. We calculate ARs and CARs based on a market model estimated from the window of 126 trading days prior to the event window. The sample period is January 1972 to December 2015 for the *CRASHP*-sorted portfolios and August 1972 to December 2015 for the *FAILP*-sorted portfolios.

Panel A: Highest *CRASHP* decile



Panel B: Highest *FAILP* decile