Corporate Fraud Risk and Stock Market Performance*

Lena Jaroszek, Alexandra Niessen-Ruenzi and Stefan Ruenzi

January 2016

Abstract

In this paper we investigate the impact of fraud risk—measured by the probability for earnings overstatements—on a firms future stock market performance. Based on an out-of-sample estimation of individual firms' fraud risk, we find that stocks with higher fraud risk earn significantly lower stock market returns. A trading strategy going long in stocks of firms with low fraud risk and short in stocks of firms with high fraud risk delivers a statistically significant alpha of more than 10% per year. This result is robust to controlling for differences in firms' liquidity, downside risk, or investor preferences. Our results suggest that the market does not efficiently price corporate fraud risk. Limits of arbitrage do not explain our results. Furthermore, abnormal returns are higher after periods of high sentiment, suggesting that the return patterns documented here constitute an anomaly.

JEL-Classification Codes: G320, G340, G14, G12

Keywords: Corporate Fraud; Earnings Overstatements; Stock Returns

*All authors are from the University of Mannheim. Jaroszek is also at the Centre for European Economic Research (ZEW) Mannheim, Germany. We have benefited from comments by Ilia Dichev, Katharina Lewellen, Karin Thorburn and seminar participants at the University of New South Wales, University of Mannheim, University of Hannover and University of Tübingen. All errors are our own. Please address all correspondence to jaroszek@zew.de, tel. ++49-(0)621-1235-380, Centre for European Economic Research, L7, 1, 68131 Mannheim, Germany.

Corporate Fraud Risk and Stock Market Performance

January 2016

Abstract

In this paper we investigate the impact of fraud risk—measured by the probability for earnings overstatements—on a firms future stock market performance. Based on an out-of-sample estimation of individual firms' fraud risk, we find that stocks with higher fraud risk earn significantly lower stock market returns. A trading strategy going long in stocks of firms with low fraud risk and short in stocks of firms with high fraud risk delivers a statistically significant alpha of more than 10% per year. This result is robust to controlling for differences in firms' liquidity, downside risk, or investor preferences. Our results suggest that the market does not efficiently price corporate fraud risk. Limits of arbitrage do not explain our results. Furthermore, abnormal returns are higher after periods of high sentiment, suggesting that the return patterns documented here constitute an anomaly.

JEL-Classification Codes: G320, G340, G14, G12

Keywords: Corporate Fraud; Earnings Overstatements; Stock Returns

1 Introduction

Fraud detection is one of the most severe corporate events and strongly affects investors. Dyck, Morse, and Zingales (2014) estimate average corporate fraud costs to be as high as 22% of firm value. In prominent fraud cases such as Enron, shareholders even lost their entire investments (after fraud detection, the Enron stock plummeted from $83 to almost zero (Healy and Palepu, 2003)). In addition to direct fraud costs due to legal penalties, Karpoff, Lee, and Martin (2008) estimate the reputational costs for firms involved in fraudulent activities to be about 7.5 times higher. Furthermore, corporate fraud leads to a loss of investor confidence in the stock market. Giannetti and Wang (2016) show that after the revelation of corporate fraud, households lose trust in the stock market and decrease their holdings in fraudulent as well as non-fraudulent firms.

Nevertheless, Dichev, Graham, Harvey, and Rajgopal (2013) report that, according to the CFOs they surveyed, about 20% of firms intentionally misrepresent earnings and thus mislead their investors about the true value of the firm. Therefore, investors should be keen on predicting a firm's propensity to commit accounting fraud. Survey results by Brazel, Jones, Thayer, and Warne (2015) show that this is indeed the case for at least a subset of investors.

In this paper, we use publicly available information to predict a firm's propensity to commit accounting fraud in the form of earnings overstatements. Then, we investigate the impact of firms' fraud propensity on future stock market returns. It is not clear ex-ante how stock market returns of firms with a high propensity to commit fraud differ from stock market returns of other firms. If the market is not able to predict fraud risk correctly, the negative expected price effects would not be reflected in today's prices. Thus, the stocks of firms with a high fraud propensity would eventually be overvalued and subsequently deliver lower returns on average. In contrast, if fraud risk is correctly taken into account by the market, firms with a high fraud propensity would deliver the same returns or even higher returns than other firms. Capital market theory would predict the same level of returns if fraud risk is completely diversifiable from the investors' point of view. However, if fraud risk is systematic, a higher propensity to commit fraud would indicate a higher systematic risk of that firm which eventually mandates a return premium in equilibrium. It is not implausible that fraud risk has a systematic component. Fraud risk would, for example, be systematic if managers are more likely to engage in fraudulent behavior during recessions or if fraud detection clusters in periods of market downturns.

In our empirical analysis, we first predict each individual firm's propensity to commit accounting fraud. We build on a model developed by Dechow, Ge, Larson, and Sloan (2011) which is based on firms' financial statements. We start with a large set of accounting-based variables that are potential predictors of fraud and which also serve as the starting point of the analysis in Dechow, Ge, Larson, and Sloan (2011). We then closely follow their procedure and exclude variables which are insignificant in predicting fraud in a stepwise regression calibration process. To calibrate the model, we use SEC "Accounting and Auditing Enforcement Releases" (AAER). The calibrated model can then be used to calculate a scaled probability of committing fraud by overstating earnings, labeled as "F-score".[1] The F-score aggregates information on accruals quality, financial performance, and market related incentives for identifying misstatements. Since its calibration focuses on variables that can be obtained from firms' published financial statements alone, this approach can easily be replicated and applied by investors.

We re-calibrate the F-score model over time on a 15-year rolling-window basis and use it to estimate each individual firm's F-score over time. This procedure allows us to also analyze whether and how the determinants of fraud change over time. Then, we sort firms into decile portfolios according to their propensity to commit fraud as proxied by their lagged F-score from this model. Portfolios are then re-balanced monthly, based on F-scores which are updated using the most recent publicly available data for the respective firm.

We find that firms in the decile with the lowest propensity to commit fraud outperform firms in the decile with the highest propensity to commit fraud by a significantly positive 1.11% per month. We obtain similar results if we compute equal- or value-weighted differences in risk-adjusted returns of these portfolios measured as the alpha of the CAPM, the Fama and French (1993) three factor model, or the Carhart (1997) four factor model, respectively. In all cases, a trading strategy going

[1] We use the terms (financial) misstatements and fraud interchangeably. However, Dechow, Ge, Larson, and Sloan (2011) point out that *"although fraud is often implied by an "Accounting and Auditing Enforcement Release" (AAER), accused firms typically do not admit or deny guilt with respect to the SEC allegations."*

long in stocks of firms with low fraud risk and short in stocks of firms with high fraud risk delivers a statistically significant alpha of at least 10% p.a. As an alternative model specification, we run multivariate regressions controlling for firm characteristics as well as proxies for distress risk and for the accruals anomaly. We still find that firms with a high propensity to commit fraud deliver significantly lower subsequent returns. This finding holds irrespective of whether the firms that are actually investigated by the SEC are included in portfolio.

Our results are robust if we control for additional systematic risk factors, such as those in the q-factor model by Hou, Xue, and Zhang (2015), the recently suggested Fama and French (2015) five-factor model, or the betting against beta model by Frazzini and Pedersen (2014). The results are also robust to including the gross profitability factor from Novy-Marx (2013), the Asness, Frazzini, and Pedersen (2013) quality-minus-junk factor, and the max return factor by Bali, Cakici, and Whitelaw (2011). Furthermore, we show that our results are not driven by the exposure to systematic tail risk (Kelly and Jiang, 2014; Chabi-Yo, Ruenzi, and Weigert, 2015), or to systematic liquidity risk (Pastor and Stambaugh, 2003; Sadka, 2006). They also withstand in a battery of further robustness tests.

Taken together, the impact of fraud risk on returns does not seem to be driven by differences in the exposure of high and low fraud risk firms to known systematic risk factors. Together with the observation that low fraud risk firms generate higher risk-adjusted returns than high fraud risk firms, this suggests that the market does not correctly take into account information on fraud risk. Consequently, the most likely explanation for our findings is that they are the reflection of a market inefficiency.

Market inefficiencies are more likely to sustain if there are limits of arbitrage (Shleifer and Vishny, 1997). Thus, one potential explanation for our main result is that the stocks that our trading strategy would mandate to trade are subject to severe limits of arbitrage, such that investors cannot easily trade these stocks and exploit the fraud-induced mispricing. We run several checks and find at best very weak evidence that our main effect is stronger among firms with more severe limits of arbitrage. We also document that our strategy still delivers abnormal risk-adjusted returns of about 8% per year if portfolios are only rebalanced on an annual (rather than a monthly) frequency.

This finding alleviates concerns that investors might simply not be able to exploit the anomaly documented here due to excessive trading costs, suggesting that the patters we document constitute an anomaly.

Stambaugh, Yu, and Yuan (2012) argue that if a certain return anomaly like the one we document is really due to mispricing, abnormal returns should be stronger after periods of high investor sentiment which we indeed find to be the case in our setting.

Our paper contributes to several strands of the literature. First, we contribute to the literature on corporate fraud prediction (Dyck, Morse, and Zingales, 2014; Johnson, Ryan, and Tian, 2009; Yu and Yu, 2011), in particular to the strand that combines multiple criteria into an aggregate measure to identify companies with higher-than-average fraud probability (Dechow, Ge, Larson, and Sloan, 2011; Dimmock and Gerken, 2012).[2] We contribute to this literature by not only calculating an aggregate fraud propensity measure for each firm in our sample that investors could easily replicate, but by doing so in a rolling-window analysis allowing us to analyze how the importance of various variables in predicting fraud varies over time.

Second, we contribute to the literature focusing on the ex-post stock price reaction to news about firms involved in dubious actions. An early study by Feroz, Park, and Pastena (1991) examines 224 AAERs issued between 1982 and 1989 and focuses on market returns around disclosures of reporting violations. In an event study they find disclosures to be associated with average two-day abnormal returns of –13%. Dechow, Sloan, and Sweeney (1996) focus on firms that were subject to SEC investigations because of earnings management. Palmrose, Richardson, and Scholz (2004) analyze the stock price reaction to earnings restatements, and Gande and Lewis (2009) investigate the wealth effects of shareholder-initiated class action lawsuits, all documenting significantly negative announcement returns.[3] We differ from these papers by not focusing on announcement returns following actual fraud incidents, but rather showing that a higher-than-average fraud probability

[2] Complementing models which predict earnings overstatements include Kim and Skinner (2012), who analyze how to predict securities litigation risk by supplementing the commonly used industry membership with measures of firm characteristics such as size, growth, and stock volatility. Furthermore, Nguyen, Iqbal, and Shiwakoti (2015) construct an aggregate "ESCORE" to predict earnings management (but not earnings manipulation).

[3] Corporate fraud adversely affects other firms, too: According to Fich and Shivdasani (2007), even firms with board interlocks to a firm sued for fraudulent behavior exhibit significant valuation declines at the time of the lawsuit filing.

calculated based on publicly observable information is associated with lower stock market returns in the future. To our knowledge, there is only one contemporaneous paper investigating the predictive ability of the probability of earnings overstatements on future stock returns: consistent with our results, Beneish, Lee, and Nichols (2013) find that a high probability of earnings overstatements predicts lower future returns. However, our paper goes one step further by examining the role of sentiment and limits of arbitrage to assess whether an implementable trading strategy based on our findings would be profitable and to better understand the sources of these return patterns. Furthermore, we differ from their paper methodologically (i) by updating the calibration model over time, which is an important feature of our analysis, (ii) by controlling for the exposure to recently suggested systematic risk factors, and (iii) by using the more popular Dechow, Ge, Larson, and Sloan (2011) F-score model to predict fraud (instead of the model and fixed coefficient estimates of Beneish (1999) as used in their paper).

2 Data and methodology

Our analysis is based on three data sources. We obtain financial statement information from Standard & Poor's Compustat North America annual and quarterly filings.[4] Returns and market related information are from CRSP's monthly stock files. We restrict the sample to shares listed on NYSE, AMEX or NASDAQ. Information on earnings overstatements is included in the SEC Accounting and Auditing Enforcement Releases (AAERs). These releases are issued during or at the conclusion of an investigation for alleged accounting and/or auditing misconduct. The advantage of using AAERs to identify fraudulent firms is that AAERs are reliable since the SEC will in most cases correctly identify manipulating firms. Karpoff, Koester, Lee, and Martin (2014) compare different databases with information on financial misstatements to a set of more than 1,000 hand-collected financial misrepresentation incidences. In this comparison, AAERs are shown to cover the widest scope of financial misconduct and they are also most likely to prompt SEC enforcement for financial misrepresentation. Therefore, the type I error rate will be low. However, given its budget

[4]Quarterly filings are only used to augment the annual information by the filings' publication dates (Compustat item rdq). We do not use any other information from quarterly filings due to lower data coverage for some items.

constraints, the SEC is likely to prosecute more obvious cases (for example, restatements, large write-offs, large stock price declines).[5] This fact might lead to a selection bias in the type of firms that are convicted of corporate fraud. As a result, many manipulating firms will remain undetected (which might lead to a less selective calibration and increase in type II errors). However, this is not an issue for our later asset pricing results as it would only introduce some noise in the firm-sorting process. The disadvantage of using AAERs is that they are typically published months after the initial public announcement of financial misconduct. However, the late publication is also not of importance in our analyses since we are not interested in the ex-post stock price reaction to fraud revelation but in the ex-ante impact of fraud risk on expected stock returns.

Note that in this paper, fraud risk refers to the probability that a firm misleads its investors about the true value of the firm by overstating its earnings. Thus, while other types of fraudulent behavior such as collusion or cartel agreements might be beneficial for investors as long as they remain undetected, earnings overstatements are clearly undesirable from the investors point of view.

We use the 2013 version of the AAER database which is obtained from the Center for Financial Reporting and Management (CFRM) at Berkeley and covers all AAERs which were issued between 1980 and 2009.[6] From these data, we collect information on 1,468 earnings overstatements affecting annual financial statements.[7] We are able to merge 1,351 earnings overstatements to firms in Compustat.

We construct our data set in two steps: First, we recalibrate the F-score according to the procedure described in detail in Dechow, Ge, Larson, and Sloan (2011) based on AAER information and Compustat data. Dechow, Ge, Larson, and Sloan (2011) use the full sample period to calibrate the F-score. We do not follow this procedure, as it could introduce a look-ahead bias in our asset pricing tests. Thus, we deviate from the original paper in that we use a 15-year rolling (overlapping)

[5] Kedia and Rajgopal (2011) find that the SEC is more likely to investigate firms located closer to its offices, which is consistent with the view that the SEC has constrained resources.

[6] See http://www.sec.gov/divisions/enforce/friactions.shtml

[7] Of the 1,825 entries in the data file we had to exclude 314 items because of missing company identifiers and further 43 incidents which covered earnings understatements.

window approach where the first (last) window uses information from 1980 to 1994 (1995 to 2009) to calibrate the model. The window is then always rolled forward by one year and re-calibrated. In the next step we use the most recent publicly available accounting data of a firm (that still does not belong to the respective model calibration window) and calculate its F-score based on the estimated model. Hypothetically, this procedure allows us to calculate F-scores per firm from 1995 onwards (the year in which our earliest calibration period ends). However, to make sure that our out-of-sample return analysis is based on publicly available data from a meaningful number of firms, we start our analysis in January 1996. Since AAER misstatements are identified only up to 2009 in our database, in our asset pricing tests we carry forward the 2009 F-score calibration estimates to the years 2010 to 2012. Penny stocks, i.e., stocks with a previous month's price below USD 1, are excluded from the sample.[8]

The final dataset that we use in our asset pricing tests includes more than 600,000 firm-month observations and covers a time period from 1996 to 2012.

2.1 F-score calibration

To measure a firm's probability to commit fraud, we closely follow Dechow, Ge, Larson, and Sloan (2011) and use AAERs on earnings overstatements as our measure of corporate fraud. To predict fraudulent activities, we estimate logit regressions with three groups of independent variables: variables reflecting accruals quality, measures of financial performance, and market related incentives. These variables can be easily obtained from financial statements. We follow Dechow, Ge, Larson, and Sloan (2011) and winsorize all explanatory variables (except for issuance of debt or equity) at the 1% and 99% level.

According to Healy (1985), earnings misstatements are mainly predicted by accrual components of firms' annual statements. Therefore, five measures of accruals are included as independent variables. Among them are two comprehensive accruals measures: working capital accruals as described in Allen, Larson, and Sloan (2013), and accruals according to Richardson, Sloan, Soliman, and Tuna

[8]Our results are similar if we filter out stocks with a price below USD 5 or exclude firms in the first NYSE size decile from our sample.

(2005) (RSST accruals) which extends the definition of working capital accruals by including long-term net operating assets. Furthermore, two specific accruals accounts are included - changes in receivables and changes in inventories. These accounts influence figures closely followed by investors. While changes in receivables can improve sales growth, both accrual accounts influence gross profit through their effect on revenue recognition and cost of goods sold. The fifth measure is the share of "soft" assets which are defined as assets that are subject to assumptions and forecasts. Soft assets are neither cash nor property, plant and equipment (PP&E) and a higher share of soft assets likely provides more opportunities to manipulate short-term earnings.

To measure the impact of a firm's financial performance on the propensity to commit fraud, we include two measures designed to address the suspicion that managers try to hide deteriorating performance (as suggested by, e.g., Dechow, Sloan, and Sweeney, 1996; Beneish, 1999): changes in cash sales and changes in return on assets. Market related incentives to manipulate earnings figures are present for firms which issue debt or equity as they might try to push-up market prices around the issuance of new securities so that they can achieve higher issuance prices. Therefore, we also include the total volume of equity and debt issuance as an additional control variable.

The F-score calibration is based on logit regressions which yield, for each firm, the propensity to overstate earnings. Specifically, the dependent variable in our fraud prediction regression is equal to one if a firm is charged with earnings overstatement for a certain year and thus listed in an AAER, and zero otherwise. Independent variables are measured in first differences. Dechow, Ge, Larson, and Sloan (2011) provide two potential and non-exclusive reasons for including first differences rather than levels of accruals: First, managers may use earnings management techniques that are still allowed under GAAP before resorting to outright manipulation eventually identified by the SEC. Secondly, growing accruals in years preceding a misstatement may reflect overinvestment which leads to misstatements when growth expectations do not materialize.

To calibrate the F-score we run backward elimination logistic regressions with a conservative significance level for elimination of 15% as in Dechow, Ge, Larson, and Sloan (2011). However, we calibrate F-score parameters based on 15-year rolling (overlapping) windows instead of using the full sample available, which would otherwise introduce a look-ahead bias in our later asset pricing tests. The calibrated coefficients based on the latest time horizon available in our data (i.e., 1995 to 2009) yield the following estimation equation for a firm's propensity to commit fraud:

$$F = -7.397 + 0.491 * (rsst\_acc) + 2.090 * (ch\_rec) + 1.550 * (soft\_assets) + 0.066 * (ch\_cs) + (-0.262) * (ch\_roa) + 1.724 * (issue)$$

These estimates can be compared to the values calibrated by Dechow, Ge, Larson, and Sloan (2011) which are as follows:

$$F = -7.893 + 0.790 * (rsst\_acc) + 2.518 * (ch\_rec) + 1.191 * (ch\_inv) + 1.979 * (soft\_assets) + 0.171 * (ch\_cs) + (-0.932) * (ch\_roa) + 1.029 * (issue)$$

Differences between coefficients in our rolling-window calibration and coefficients in Dechow, Ge, Larson, and Sloan (2011) might be due to the fact that re-calibrating the fraud prediction model accounts for structural changes in accounting principles (e.g., as reaction to newly introduced regulation as the Sarbanes-Oxley Act). This reasoning is supported by Figure 1 which displays coefficients of our fraud prediction model over time.

Coefficients in Figure 1 indicate that the components predicting corporate fraud are time varying and some explanatory variables even drop out or enter the prediction model in certain years only. For most variables, the coefficient estimates decrease in absolute values over time, except for the percentage of soft assets, and total debt and equity issuance. Still, t-values increase for all coefficients except for the change in inventories and change in cash sales. Change in inventories consequently drops out of the F-score model for more recent periods. Change in cash sales is insignificant for one calibration horizon for which working capital accruals become a significant predictor.

Given the time varying nature of variables predicting corporate fraud and the need to avoid look-ahead bias in our asset pricing tests, we use the rolling-window F-score calibration as described above in our main analysis. However, in our later robustness analyses, we also consider a setting in which the F-score model is calibrated based on the initial horizon from 1980 to 1994 and held constant afterwards.

2.2 Descriptive statistics

To validate that our F-score measure indeed captures corporate fraud risk, we sort firms into deciles according to their F-score and count the number of firms that are prosecuted by the SEC for alleged corporate fraud in each of the deciles in the following year. Results are displayed in Figure 2.

As expected, we observe that the number of firms that committed fraud increases over the F-score deciles. While there are only 11 (2% of) firms with fraudulent activities in the lowest F-score decile (which reflects low fraud risk), there are 96 (21% of) firms with fraudulent activities in the highest F-score decile (which reflects high fraud risk). This indicates that, while not perfectly predicting corporate fraud, our F-score measure is reasonably correlated with future fraudulent activities by firms.[9] The figure also shows the F-score cutoff values for the respective decile portfolios. For example, firms in the lowest (highest) F-score decile have a F-score smaller than 0.44 (larger than 1.74).

In the next step, we investigate whether firms with high and low fraud risk differ significantly in terms of their annual statements, accounting performance, or equity/debt issuance activities. Descriptive statistics on all explanatory variables as well as univariate comparisons between high and low fraud risk firms are provided in Table 1.

While columns (1) to (6) in Table 1 report summary statistics on our whole sample of firms (excluding firm-year observations that have missing observations for some of the variables which are needed to compute the F-score) and columns (7) to (9) compare characteristics of above and below median F-score firms. The average F-score value is close to one which is comforting given that an F-score value of 1.00 indicates that the misstatement probability is equal to the unconditional fraud expectation. The recalibrated F-score values compare reasonably to the original results from Dechow, Ge, Larson, and Sloan (2011). In the original paper the authors choose a cut-off value of 1.4 to identify firms with high fraud probability. Specifically, a value of 1.4 implies that the probability for a misstatement is 1.4 times higher than the unconditional probability. In their sample and

[9] The total number of misstatement firms for which we have all financial statement information necessary to calculate the F-score is 461.

calibration model, about 20% of firms show F-score values of above 1.4. In our larger sample and with recalibrated F-score values, the fraction of firms with F-score values above 1.4 amounts to nearly 30%. The average F-score in the sample of above median F-score firms is 2.4 times higher than in the sample of below average F-score firms. Accruals quality is lower for high (i.e., above median) F-score firms. For example, RSST accruals amount to 7.6% of assets for firms with an above median fraud risk and are therefore 9.8 percentage points higher compared to below median fraud risk firms than for low fraud risk firms. Similarly, changes in receivables are 4.3 percentage points higher for high fraud risk firms compared to low fraud risk firms. The percentage of soft assets as a share of total assets is 33.3 percentage points higher for high fraud risk firms. Furthermore, firms with above median F-score also have higher Sloan (1996) accruals. While Sloan (1996) accruals are not included in the F-score (since they highly overlap with RSST accruals), we use them as a control in our later multivariate analysis.

With respect to our financial performance variables, change in cash sales is more than twice as large for high as compared to low fraud risk firms. This finding is consistent with results in Dechow, Ge, Larson, and Sloan (2011) who argue that misstating firms grow their capital bases and increase the scale of their business operations which simultaneously increases cash and credit sales. Further, firms may misstate sales through transaction management. We find an economically small difference for changes in ROA which is in line with Dechow, Ge, Larson, and Sloan (2011).

Concerning market related firm characteristics, the difference in the book-to-market ratio is insignificant, i.e., high fraud risk firms are not significantly more likely to be growth or value firms. Nearly all firms with above median fraud risk have issued debt or equity in the previous year compared to 84.4% for below median fraud risk firms. High fraud risk firms tend to be larger which might stem from the F-score being calibrated on SEC AAERs and the SEC may focus their examinations on large firms. For the asset pricing tests, this may be beneficial because strategies based on large firm investments can better be implemented. High fraud risk firms also have on average higher stock prices and their dividend yields are lower. Although we do not find a significant difference for above and below median F-score firms with respect to their book-to-market ratios, firms with above median fraud risk have higher past sales growth and therefore tend to be glamour stocks according to the definition by Lakonishok, Shleifer, and Vishny (1994). While below median fraud risk firms experience an average sales growth of 101% over a five-year horizon this rate is around 116% for higher fraud risk firms. Also, we observe higher asset growth for high fraud risk firms which could be associated with potential underperformance of stock returns according to Cooper, Gulen, and Schill (2008). Somewhat surprisingly, above median F-score firms show on average slightly lower financial distress risk which we measure according to Campbell, Hilscher, and Szilagyi (2008). Similar to their results, the average distress probabilities are relatively low. For high fraud risk firms financial distress risk is on average 0.1%. All univariate differences (except for the difference in book to market) are statistically significant at the 1% level.

3 Impact of corporate fraud risk on stock market performance

In our main analysis, we use firm-specific fraud risk derived from the F-score calibration to investigate whether stock market returns of firms with high fraud risk differ from those with low fraud risk. We first examine raw returns and factor-model based risk-adjusted returns of portfolios consisting of firms with different levels of fraud risk (Section 3.1) and then check whether our results are robust with respect to various aspects of our empirical design (Section 3.2).

3.1 Portfolio evidence

We start our empirical investigation by sorting firms into deciles according to their previous month's F-score. The F-score model used to determine an individual firm's F-score is calibrated based on a moving 15 year window as described in Section 2.1. The calibration and the F-score calculation for the individual firms only include annual data until the most recent annual statement. This procedure ensures that our measure of fraud risk is solely based on publicly available information to which all investors would have access at the time we sort firms into portfolios.[10] Thus, our trading strategy is fully implementable. The average number of firms in each decile is 294, ranging from a

[10] We rely on Compustat's data item rdq to derive financial statements' publication lags and impute missing values. Our results are robust to using a fixed publication horizon of six months.

minimum of 147 in January 1996 (due to calibration and publication delay) up to a maximum of 381 in June 1999. In the next step, we compute value-weighted average returns less the risk-free rate for each decile and examine a long-short strategy that goes long in firms with low fraud risk (decile 1) and short in firms with high fraud risk (decile 10). Results are presented in Panel A of Table 2.

We find that firms with low fraud risk (decile 1) earn positive returns amounting to 0.775% over the subsequent month, while there is a negative return of -0.334% for firms with high fraud risk (decile 10). The return difference is 1.11% per month and statistically significant at the 1%-level, indicating that a trading strategy of going long in firms with low fraud risk and short in firms with high fraud risk delivers an annual return of about 13%. In untabulated results we found that this finding is obtained even if firms that are actually targeted by the SEC for alleged fraud are excluded from the sample, showing that firms with a high probability of engaging in earnings overstatements are subject to negative price changes subsequently, irrespective of whether they are actually investigated by the SEC or not.

To make sure that our result based on raw returns is not driven by any systematic difference between high and low fraud risk firms in their exposure to the well-known systematic risk factors, we also compute portfolio alphas based on the CAPM (controlling for the exposure to the market factor) in Panel B, the Fama and French (1993) model (additionally controlling for exposures to the size and value factor) in Panel C, and the Carhart (1997) model (additionally controlling for exposure to the momentum factor) in Panel D. Based on the CAPM, a zero-investment strategy buying those stocks with the lowest fraud propensity and selling stocks with the highest fraud propensity generates an average monthly alpha of 1.16% which is significant at the 1% level. Including the size (SMB) and the book-to-market (HML) factor slightly reduces the monthly alpha to 1.03%, adding the momentum factor yields a monthly alpha of 0.89%. In all cases, the abnormal returns earned by the long-short trading strategy remain highly significant and economically large even after adding additional risk factors. While the asset pricing models used in Table 2 price intermediate fraud risk portfolios fairly well (as indicated by the insignificant alpha estimates for most of those portfolio) high fraud firms persistently show significantly negative alphas. However, although results in the alpha regressions are mainly driven by the highest fraud risk portfolio, it should be noted that this portfolio is still large (containing on average more than 300 firms). Furthermore, the low fraud risk portfolio earns a positive albeit insignificant alpha such that both contribute to the significant abnormal return of the long-short trading strategy.

Our results thus far are based on value-weighted portfolios. To check whether they are only driven by a small-number of very large firms, we now repeat the portfolio analysis based on equal-weighted portfolio returns. Results are reported in Table 3.

Compared to results based on value-weighted returns, we observe very similar patterns if returns are equal-weighted for each fraud risk decile. The positive alphas for the low fraud risk portfolio (decile 1) now turn significant, while we still observe negative alphas for the high fraud risk portfolio (decile 10). In economic terms, alphas for the low fraud risk portfolio double in magnitude, if the portfolio return is equal- as opposed to value-weighted. For high fraud risk firms, equal weighted portfolio returns yield smaller alphas, but they are still statistically significant for the Fama and French (1993) and Carhart (1997) model. Most importantly, as with the value-weighted portfolios, a long-short strategy investing in those firms with the lowest fraud risk and shorting those stocks with the highest fraud risk yields highly significant alphas across all model specifications. They range from 0.90% (Carhart model) to 1.03% (CAPM) and are all significant at the 1% level.

Taken together, results from the portfolio analysis show that a trading strategy investing in stocks with low fraud risk firms and selling stocks with high fraud risk firms yields persistently positive alphas of about 0.9% to 1% per month which translates into an economically significant outperformance of 11% to 12% per year. These findings suggest that the market does not correctly price corporate fraud risk expectations that can be formed based upon publicly available data. The sign of the abnormal performance of our trading strategy clearly shows that investors do not earn a risk premium as a compensation for holding stocks of firms with high fraud risk.

3.2 Robustness of main result

To ensure the robustness of our portfolio results, we apply several modifications to our estimation strategy. For the sake of brevity, in Table 4 we only report results from these modifications for the lowest (decile 1) and highest (decile 10) fraud risk portfolio as well as the long-short portfolio return based on the Carhart model.

In Panels A and B, we show results for different rolling and expanding windows used to calibrate the F-score. Columns (1) to (3) of Panel A repeat the baseline analysis from Table 2 for ease of comparison. In columns (4) to (6), we estimate the F-score based on a 10 year (instead of a 15 year) rolling window. This allows us to start the asset pricing tests in 1991 (instead of 1996) already, since we only need 10 years to calibrate the F-score. Shortening the rolling window arguably makes F-score estimations more noisy. Nevertheless, we still observe a positive alpha of 0.66% per month for the long-short portfolio which is significant at the 5% level. Thus, our results are robust to a variation in the time period used to calibrate the F-score.

In the second set of results, we use expanding instead of rolling windows. This approach allows for more information to enter the F-score model in the later sample years and thus should make it more precise. At the same time, by including relatively old data in the calibration of the model, the change in the importance of certain variables (see Figure 1) might not be captured in a timely manner.

Results in columns (1) to (3) in Panel B show an even stronger effect on the Carhart alpha of the long-short portfolio for the baseline model using an initial 15 year but expanding window, which now amounts to 0.96% per month. It is still significant at the 1% level. If we start the estimation with a 10 year window which subsequently expands (columns (4) to (6)), we observe an alpha of 0.74% which is also slightly larger than the alpha based on a fixed length 10 year rolling window in Panel A and again statistically significant at the 1% level.

In Panel C, the F-score model is calibrated as in our baseline models, i.e., we use a 15 year rolling window, but we modify the data updating frequency and the portfolio rebalancing procedure. In columns (1) to (3), we use fixed coefficients to compute the F-score and do not allow them to vary over time. Instead, we only use the first 15-year window once to determine the F-score model. With this procedure, we follow Dechow, Ge, Larson, and Sloan (2011) more closely, who also use one set of coefficients to estimate their F-score. However, in contrast to Dechow, Ge, Larson, and Sloan (2011), we still do not use the entire time period available to calibrate the F-score. By using coefficients that are obtained from the first 15 year estimation, there is still no look-ahead bias in our asset pricing tests (that only start subsequent to the first 15 year period). Results in Table 4 show that there still is a significantly positive alpha for the long-short portfolio that amounts to 0.86% per month. Not surprisingly, this result is weaker in economic and statistical terms, which would be expected due to the less precise estimation of the F-score which, particularly for recent years, should be measured with a lot of noise due to the time gap between the computation of coefficients underlying the F-score and the asset pricing tests. This finding shows that it is important to take into account the time-varying nature of the impact of various variables for fraud prediction (see also Figure 1).

Finally, in columns (4) to (6), we rebalance portfolios yearly instead of monthly. Our main result is not affected: a portfolio that is long in low fraud risk firms and short in high fraud risk firms delivers a positive alpha of 0.63% per month which is still economically large and remains significant at the 5% level.

Overall, our portfolio results portray a consistent picture suggesting that high fraud risk firms deliver significantly lower returns as compared to low fraud risk firms. The portfolio strategy can be implemented based solely on public information and in all specifications delivers statistically significant and economically meaningful abnormal returns.

4 Alternative explanations

In the following, we investigate alternative explanations for the observation that stocks of high fraud firms generate lower alphas than stocks of low fraud firms. We first analyze the impact of individual firm characteristics (Section 4.1) on stock returns. Specifically, we investigate whether our results are driven by anomalies already known in the literature. Among others, we include Sloan

(1996) accruals and financial distress risk according to Campbell, Hilscher, and Szilagyi (2008) into a multivariate regression analysis. We then turn to an analysis of alternative factor models to check whether our results are driven by the exposure of our trading strategy returns to systematic risk factors other than the Carhart (1997) factors (Section 4.2). We then investigate whether our results are due to limits of arbitrage (Section 4.3) and examine the impact of sentiment on our results (Section 4.4). Finally, we discuss the implementability of a trading strategy trying to exploit our findings and assess the impact of transaction costs in more detail (Section 4.5).

4.1 Impact of firm characteristics on stock returns

Brennan, Chordia, and Subrahmanyam (1998) show that several firm characteristics have an impact on stock returns, and Daniel and Titman (1997) warn against only using factor-sensitivities to explain the cross-section of stock returns. Therefore, we now turn to a multivariate regression analysis to explore whether firm characteristics drive returns of firms with high and low fraud risk and thus explain our results. Specifically, we regress monthly excess stock returns over the risk-free rate on a variable indicating to which F-score decile a firm belongs, and the same set of firm characteristics as examined in Brennan, Chordia, and Subrahmanyam (1998).[11] Furthermore, we follow Lilienfeld-Toal and Ruenzi (2014) and include lagged asset growth, a dummy variable indicating whether a firm belongs to a high-tech industry, five-year sales growth, and a dummy variable indicating S&P 500 index membership. All independent variables are lagged to mitigate endogeneity problems.

Besides firm characteristics that are typically considered in the cross sectional asset pricing literature, we enrich the model by also including Sloan (1996) accruals and a measure of financial distress according to Campbell, Hilscher, and Szilagyi (2008). Sloan (1996) finds a long-short strategy investing in stocks of firms with low accruals and selling stocks of firms with high accruals to earn positive size-adjusted abnormal returns in the year following portfolio formation. For each firm, we calculate the accruals measure according to the methods described in Sloan (1996) and

[11] Using deciles instead of F-scores directly minimizes noise by reducing the impact of extreme F-scores in individual stock level regressions. Nevertheless, our later results also hold if we use the F-score value itself instead of sorting into deciles. However, analyzing F-score deciles facilitates interpretation of regression coefficient estimates.

Chen, Novy-Marx, and Zhang (2011). In order to control for financial distress risk, we follow Campbell, Hilscher, and Szilagyi (2008) and predict a firm's one year ahead default probability based on a broad set of accounting and market variables. Since the financial distress puzzle is driven by the firms with the highest financial distress risk, we include an indicator variable marking those firms which are among the highest 10% of the distress risk distribution in a given month.[12] Results are presented in Table 5.

Column (1) of Table 5 reports results from a regression that includes month fixed effects. Standard errors are clustered at the firm level. The main independent variable, *fraud decile*, ranges from 1 to 10 and indicates the fraud decile, to which a firm belongs in a given month.[13] Similar to our portfolio analysis, we find that higher fraud risk leads to lower subsequent stock returns. The impact of the fraud decile variable is negative and statistically significant at the 1%-level, confirming our earlier findings from the portfolio analysis that high fraud risk stocks have significantly lower returns. The coefficient estimate indicates that firms that belong to the lowest decile with respect to their F-scores achieve returns that are on average 0.41% per month higher than those of firms that belong to the highest F-score decile.

With respect to control variables, we find that value firms and small firms generate higher subsequent returns confirming known return patterns. We also confirm the findings from Cooper, Gulen, and Schill (2008) by showing that firms with high asset growth have significantly lower subsequent stock returns. Furthermore, high accruals firms earn significantly lower returns which is in line with the evidence provided in Sloan (1996). Lastly, we can also confirm the financial distress puzzle (Campbell, Hilscher, and Szilagyi, 2008) by showing that firms with the highest distress risk deliver lower returns.

To control for potentially time varying industry effects, we include combined month-industry fixed effects in column (2).[14] We still observe a significantly negative impact of fraud risk on subsequent

[12]If we include the level of financial distress risk instead of the high distress risk indicator the coefficient estimate on this variable is insignificant whereas our results on fraud risk are still unaffected.

[13]Fraud deciles are formed each month as in our portfolio analysis in Section 3.

[14]We use Fama-French 49 industries and obtained the assignment of SIC codes to industry groups from Kenneth French's data library.

monthly stock returns which is of similar magnitude as before and statistically significant at the 1%-level. This result shows that our main finding is not just driven by industry effects.

As an alternative to controlling for industry-time fixed effects, we follow Daniel, Grinblatt, Titman, and Wermers (1997) and control for a stock's characteristic based benchmark by including DGTW-month combined fixed effects in column (3). In order to determine to which characteristic benchmark portfolio a firm belongs we use the breakpoints of Daniel, Grinblatt, Titman, and Wermers (1997) which dependently sort stocks into 125 portfolios according to size quintiles, book-to-market quintiles, and momentum quintiles.[15] The results for combined month-DGTW effects are slightly stronger as compared to the model including combined month-industry fixed effects. They now indicate an outperformance of firms in the lowest fraud risk decile of about 5.5% per year as compared to firms in the highest fraud risk decile.

In column (4), we go even one step further and include firm-fixed effects in addition to time-fixed effects. Including firm fixed effects controls for the impact of all non-time varying firm characteristics on returns and is a very restrictive specification as any impact of the F-score can now only be driven by within-firm time-series variation in firm-level fraud risk. However, even in this case results remain remarkably strong. The coefficient estimate is only slightly reduced as compared to our baseline-specification from column (1) and is still significant at the 5%-level. The finding that much of the return-effect is driven by time-series variation in the F-score also underlines the importance of our approach to update the F-score calibration model and computation over time.

In the next step, in addition to including month and firm fixed effects, we cluster standard errors two-dimensionally along the firm and time dimensions in column (5), as suggested by Petersen (2009). This does not affect the coefficient estimate but reduces statistical significance. However, even in this conservative specification, the impact of fraud risk on stock returns is still statistically significant at the 10%-level.

Finally, in column (6), we run traditional Fama and MacBeth (1973) (FMB) regressions, where we

[15] We use the traditional DGTW benchmark breakpoints. The DGTW benchmarks are available via http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm and are also described in Wermers (2003) for the period until December 2010 which shortens our sample in the respective analysis.

compute coefficients on each independent variable from cross-sectional regressions separately for each month, and then determine the final parameter values based on the mean and statistical significance of the time-series statistics of these monthly estimates. This does not affect our main result, which remains comparable to the pooled panel regressions in terms of both, economic magnitude and statistical significance.

Taken together, results from the multivariate firm level analysis confirm our previous portfolio level evidence showing that our findings cannot be explained by differences in other firm characteristics.

4.2 Systematic risk factors

An alternative explanation for our results could be that fraud risk is highly correlated with other systematic risk factors that might have a negative return premium. If the exposure of high and low fraud risk firms to systematic risk factors (that are not yet included in the Carhart (1997) four factor model) differs, the significant alphas we document might be driven by these differences in exposures rather than by fraud risk per se. To analyze this possibility, in Table 6 we provide an in-depth analysis of the impact of a large number of new systematic risk factors that are currently debated in the empirical asset pricing literature.

We generally follow the same approach as in Table 2 and regress value-weighted returns from the long-short fraud portfolio on the Fama and French (1993) factors, but then include additional explanatory factors (except in the case of the q-factor model of Hou, Xue, and Zhang (2015) that does not include the traditional factors). In column (1), we implement the recently introduced Fama and French (2015) five factor model, in which one of the two newly introduced factors captures robust-minus-weak (rmw) operating profitability (minus interest expenses) and the other one conservative-minus-aggressive (cma) investments measured by the growth of total assets. We find a positive monthly alpha of 0.98% which is significant at the 1% level and even larger than the alpha obtained from the Carhart (1997) four factor model. In column (2), we alternatively estimate the related q-factor model recently suggested by Hou, Xue, and Zhang (2015). Our main result again holds.

Adding the gross profitability factor from Novy-Marx (2013) (column (3)) leads to even stronger results with an alpha of 1.11% per month. In column (4), we add the Asness, Frazzini, and Pedersen (2013) quality-minus-junk factor, which aggregates several profitability, growth, safety and payout measures into one score, and in column (5) we add the betting against beta factor by Frazzini and Pedersen (2014).[16] Again, we obtain strong results with a monthly alpha of 1.10% and 0.91%, respectively.

In the next step, we turn to several asset pricing models which focus on investors' preferences or biases: (i) we include the max return factor by Bali, Cakici, and Whitelaw (2011) (column (6)) which addresses investors' preferences for lottery like stocks, (ii) we control for the exposure to tail risk according to Kelly and Jiang (2014) (column (7)), and crash risk assessed by the lower tail dependence (LTD) factor according to Chabi-Yo, Ruenzi, and Weigert (2015) (column (8)) which addresses preferences against stocks with high crash risk.[17] In all specifications, we still obtain a significantly positive alpha on the long-short fraud portfolio indicating that high fraud risk firms deliver lower returns than low fraud risk firms. The economic magnitudes are similar to results from previous sections and range between 0.85% per month and 0.94% per month. Fraud risk seems to be unrelated to the low beta anomaly and cannot be attributed to lottery stock or crash risk preferences.

Finally, we also consider systematic liquidity risk factors as suggested by Pastor and Stambaugh (2003) and Sadka (2006)[18] and find them to also be unrelated to the negative impact of corporate fraud risk on stock returns (columns (9) and (10)).

Overall, the evidence presented in Table 6 shows that corporate fraud risk does not seem to be subsumed by other known systematic risk factors, suggesting that market inefficiencies might explain our main result. The following section explores this conjecture in more detail.

[16]We obtain the rmw and cma factor from Kenneth Fench's data library, the q-factors were provided by the authors, the gross profitability factor is downloaded from Robert Novy-Marx's data library and the quality-minus-junk and betting-against-beta factors are from Lasse Pedersen's website.

[17]The lottery factor was downloaded from Nusret Cakici's data library, tail risk is calculated according to the original paper, and the LTD factor was provided by the authors of Chabi-Yo, Ruenzi, and Weigert (2015).

[18]Data on the liquidity factors is obtained from Robert Stambaugh's Wharton website and Ronnie Sadka's website.

4.3 Limits of arbitrage

Rather than the market not pricing fraud risk correctly, it is possible that market participants are aware of the abnormal returns we document, but cannot profit from this "anomaly" due to limits of arbitrage (Shleifer and Vishny, 1997). If limits of arbitrage are severe, abnormal returns can persist even if market participants are aware of their existence. Therefore, we now investigate whether alphas of high fraud risk firms are lower among firms with more severe limits of arbitrage. In the literature, several proxies for limits of arbitrage are suggested. We focus on three different types of limits of arbitrage: differences in liquidity, volatility, and information asymmetry. If limits of arbitrage drive our findings, we should see that the effect is mainly driven by firms with low liquidity levels, high volatility, and high information asymmetries as it is more difficult and costly for arbitrageurs to take advantage of profit opportunities among such firms. Results based on double sorts according to F-scores and the various proxies for limits of arbitrage are presented in Table 7.

In columns (1) and (2), we investigate whether our results are stronger for firms with low liquidity levels. As liquidity proxies we use the Corwin and Schultz (2012) bid-ask spread proxy,[19] as well as the Amihud (2002) illiquidity ratio and divide our sample firms into high (low) limits of arbitrage portfolios depending on whether their liquidity level is below (above) median liquidity in the sample in a specific month. For the Corwin and Schultz (2012) spread as well as for the Amihud (2002) illiquidity ratio, we observe that alphas of the long-short portfolio are still significantly positive for both, the high and the low limits of arbitrage sample. Contrary to the expectation, the effect is even slightly weaker for the high limits of arbitrage (i.e., low liquidity) sample, but the difference between the samples is statistically insignificant.

In columns (3) and (4), limits of arbitrage are defined to be high (low) if a firm's total return volatility or its idiosyncratic return volatility is higher (lower) than the median. In both cases, we find that our main effect is now indeed stronger for the subsample with high limits of arbitrage compared to the subsample with low limits of arbitrage. However, the differences between the samples are again insignificant for both volatility measures.

[19] We are grateful to Shane Corwin for providing the data.

Finally, in columns (5) and (6), we use analyst coverage as a proxy for limits of arbitrage. In column (5) we compare firms with above and below median number of analysts following the firm and in column (6) we compare firms without any analyst coverage to firm with at least one analyst following them. High analyst coverage should reduce information asymmetries and thus lead to lower limits of arbitrage. We find that our results do not differ significantly between subsamples of high and low limits of arbitrage based on analyst coverage. While alphas are significantly positive for the subsample with low limits of arbitrage, they are only significantly positive for the subsample with high limits of arbitrage if the number of analysts covering a firm is used to define the strength of information asymmetries (column (5)), but not if the existence of analyst coverage is used (column (6)).

These results are different from the Sloan (1996) accruals anomaly since Mashruwala, Rajgopal, and Shevlin (2006) document that the accruals anomaly is concentrated in firms with high idiosyncratic stock return volatility. Therefore, risk-averse arbitrageurs may shy away from investing in stocks with extreme accruals because their higher idiosyncratic volatility makes the investment risky. Comparable evidence is provided by Campbell, Hilscher, and Szilagyi (2008) with respect to the financial distress puzzle which is also more pronounced for stocks with informational or liquidity-related frictions.

4.4 Investor sentiment

Our results so far suggest that the high abnormal returns of low as compared to high fraud risk firms is not due to systematic risk. Also, our analysis considering limits of arbitrage reveals that market participants could potentially exploit this "anomaly". These findings suggest that the return patterns we document are due to temporary misvaluations caused by the trades of uninformed investors that are not able to correctly price high and low fraud risk firms relative to each other.

Stambaugh, Yu, and Yuan (2012) study long-short strategies based on 11 anomalies and argue that if these anomalies are due to mispricing their returns should be larger after periods of high investor sentiment. Their argument, which assumes overpricing to be more frequent than underpricing due to short-sale impediments, also applies to the case at hand: If the negative return on high fraud-risk firms is due to mispricing, we should find this effect to be more pronounced in periods following high investor sentiment. Results from an analysis similar to Stambaugh, Yu, and Yuan (2012), in which we compare strategy returns after periods of high and low sentiment, are presented in Table 8.

Our results reveal a striking pattern and are in line with a strong impact of investor sentiment: First, the return on a long-short strategy is highly significant and also large in economic terms only in periods following high levels of investor sentiment. Second, the return on the high fraud risk portfolio (which is the short leg in our trading strategy) is large and negative in periods following high sentiment. These results are in line with investors' views being overly optimistic during periods of high sentiment leading to an overvaluation particularly of high fraud risk firms (which uninformed investors are not able to correctly price) and the subsequent correction of this temporal overpricing.

4.5 Implementability and after-cost performance

Our main results are based on monthly rebalancing of firms into portfolios according to their propensity to commit fraud. If the propensity of firms to commit fraud varies strongly over time, monthly rebalancing leads to high trading costs for investors, because portfolio compositions would change frequently. Thus, a long-short trading strategy which buys stocks of firms with low fraud risk and sells stocks of firms with high fraud risk might not be profitable in this case due to high trading costs. Therefore, we now investigate whether firms' propensity to commit fraud is stable over time or whether portfolio compositions change frequently over our sample period.[20] In Figure 3 we sort stocks into deciles in year 0 and follow the same stocks from this starting portfolio over the subsequent four years. Then, in each of the subsequent years, we calculate the equal-weighted average of the F-scores per decile.

We observe that the propensity to commit fraud is persistent over time. Stocks that are sorted

[20]Note that, in our setting, F-scores can change due to two reasons: first because of the yearly update of the F-score calibration model (even if firm-individual input parameters are constant) and second because of changes in the firm-level input parameters (even if the calibration model remains unchanged).

into high fraud deciles in a given year have high average F-scores in the subsequent four years as well. Thus, even a monthly portfolio rebalancing frequency does not fundamentally change portfolio compositions in the short term and thus should not lead to very high transaction costs.

This is also supported by results shown earlier in Panel C of Table 4. There, we still observe significant alphas if portfolios are rebalanced annually instead of monthly. If the frequency of portfolio rebalancing is reduced from monthly to yearly, the profitability of the zero-investment strategy is reduced by about one third, compared to monthly rebalancing, but the long-short portfolio still earns significant and economically large alphas of about 8% per year. This should be more than enough to cover any potential trading costs of reasonable magnitude (that are required due to yearly rebalancing) and to still achieve meaningful positive trading strategy returns.

Another friction that could prevent investors from taking advantage from the observed return patterns could be short-sale constraints some stocks are subject to, which is often the case for smaller stocks. However, our results hold if we exclude firms below the NYSE lowest size decile from our analysis (see 2) and the high fraud risk firms that an investor would need to short are significantly larger than the low fraud risk firms (see Table 1).

Finally, one could argue that the calibration of the F-score model is cumbersome and unsophisticated investors might find it hard to update the model frequently, e.g., due to lack of access to the needed data on AAERs. However, our results in Panel C of Table 4 also show that we still obtain significantly positive alphas of our trading strategy if we sort firms into fraud deciles based on an F-score that uses constant coefficients (that is, the F-score calibrated a single time based on data from the time period 1980-1994). There, we calibrate the F-score only once based on financial statement information from the time period 1980-1994 and hold the coefficients of the model constant afterwards. This approach could be implemented by an investor without information about recent AAERs but who knows the coefficients of the calibrated model at just one point in time (comparable to how investors often calculate Altman's Z or Ohlson's O scores based on the calibration equation of the original articles).

5 Conclusion

This paper investigates whether investors can earn a positive risk premium for holding stocks of firms with a high probability to commit accounting fraud. In contrast, we find robust evidence that firms with a high probability to commit fraud exhibit significantly lower stock market performance as compared to firms that are less likely to commit fraud according to an accounting-based fraud prediction model. We can preclude that limits of arbitrage explain the different pricing patterns of firms likely to commit fraud. Also, the result is not driven by a correlation of the propensity to commit fraud with other traditional and recently proposed systematic risk factors. We interpret our results as evidence that investors do not price corporate fraud risk efficiently. In line with this view, we find that the mispricing of fraud risk is highest after periods of high investor sentiment.

Predicting accounting fraud is crucial for investors. If a firm can trick the market by reporting overstated earnings, stock prices will be temporarily inflated. However, they will eventually drop, either due to the long-term revelation of the true firm value or because the firm comes under SEC investigation, both of which lead to negative abnormal returns for investors. Thus, investors should pay close attention to the determinants of fraud. Our paper provides an easy to implement way for them to estimate individual firms' fraud risk and to eventually consider it in their investment decision.

References

Allen, E. J., C. R. Larson, and R. G. Sloan (2013): "Accrual reversals, earnings and stock returns," *Journal of Accounting and Economics*, 56(1), 113–129.

Amihud, Y. (2002): "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets*, 5(1), 31–56.

Asness, C., A. Frazzini, and L. H. Pedersen (2013): "Quality minus junk," Working Paper.

Baker, M., and J. Wurgler (2006): "Investor sentiment and the cross-section of stock returns," *The Journal of Finance*, 61(4), 1645–1680.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011): "Maxing out: Stocks as lotteries and the cross-section of expected returns," *Journal of Financial Economics*, 99(2), 427–446.

Beneish, M. D. (1999): "The detection of earnings manipulation," *Financial Analysts Journal*, 55(5), 24–36.

Beneish, M. D., C. M. C. Lee, and D. C. Nichols (2013): "Earnings manipulation and expected returns," *Financial Analysts Journal*, 69(2), 57–82.

Brazel, J. F., K. L. Jones, J. Thayer, and R. C. Warne (2015): "Understanding investor perceptions of financial statement fraud and their use of red flags: evidence from the field," *Review of Accounting Studies*, 20(4), 1373–1406.

Brennan, M. J., T. Chordia, and A. Subrahmanyam (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49(3), 345–373.

Campbell, J. Y., J. Hilscher, and J. Szilagyi (2008): "In Search of Distress Risk," *The Journal of Finance*, 63(6), 2899–2939.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *The Journal of Finance*, 52(1), 57–82.

Chabi-Yo, F., S. Ruenzi, and F. Weigert (2015): "Crash Sensitivity and the Cross-Section of Expected Stock Returns," Working Paper.

Chen, L., R. Novy-Marx, and L. Zhang (2011): "An Alternative Three-Factor Model," Working Paper.

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *The Journal of Finance*, 63(4), 1609–1651.

Corwin, S. A., and P. Schultz (2012): "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices," *The Journal of Finance*, 67(2), 719–760.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997): "Measuring Mutual Fund Performance With Characteristic-Based Benchmarks," *The Journal of Finance*, 52(3), 1035–1058.

Daniel, K., and S. Titman (1997): "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *The Journal of Finance*, 52(1), 1–33.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan (2011): "Predicting Material Accounting Misstatements," *Contemporary Accounting Research*, 28(1), 17–82.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney (1996): "Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC," *Contemporary Accounting Research*, 13(1), 1–36.

Dichev, I. D., J. R. Graham, C. R. Harvey, and S. Rajgopal (2013): "Earnings Quality: Evidence From the Field," *Journal of Accounting and Economics*, 56(2), 1–33.

Dimmock, S. G., and W. C. Gerken (2012): "Predicting fraud by investment managers," *Journal of Financial Economics*, 105(1), 153–173.

Dyck, A., A. Morse, and L. Zingales (2014): "How pervasive is corporate fraud?," Working Paper.

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

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *The Journal of Political Economy*, pp. 607–636.

Feroz, E. H., K. Park, and V. S. Pastena (1991): "The Financial and Market Effects of the SEC's Accounting and Auditing Enforcement Releases," *Journal of Accounting Research*, (1), 107–142.

Fich, E. M., and A. Shivdasani (2007): "Financial Fraud, Director Reputation, and Shareholder Wealth," *Journal of Financial Economics*, 86(2), 306–336.

Frazzini, A., and L. H. Pedersen (2014): "Betting against beta," *Journal of Financial Economics*, 111(1), 1–25.

Gande, A., and C. M. Lewis (2009): "Shareholder Initiated Class Action Lawsuits: Shareholder Wealth Effects and Industry Spillovers," *Journal of Financial and Quantitative Analysis*, 44(4), 823–850.

Giannetti, M., and T. Y. Wang (2016): "Corporate scandals and household stock market participation," *Journal of Finance, forthcoming.*

Gompers, P. A., J. L. Ishii, and A. Metrick (2003): "Corporate Governance and Equity Prices," *Quarterly Journal of Economics*, 118(1), 107–155.

Healy, P. M. (1985): "The effect of bonus schemes on accounting decisions," *Journal of Accounting and Economics*, 7(1), 85–107.

Healy, P. M., and K. G. Palepu (2003): "The Fall of Enron," *Journal of Economic Perspectives*, 17(2), 3–26.

Hou, K., C. Xue, and L. Zhang (2015): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28(3), 650–705.

Johnson, S. A., H. E. Ryan, and Y. S. Tian (2009): "Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter," *Review of Finance*, 13(1), 115–145.

Karpoff, J. M., A. Koester, D. S. Lee, and G. S. Martin (2014): "Database Challenges in Financial Misconduct Research," *Georgetown McDonough School of Business Research Paper*, (2012-15).

Karpoff, J. M., D. S. Lee, and G. S. Martin (2008): "The Cost to Firms of Cooking the Books," *Journal of Financial and Quantitative Analysis*, 43(3), 581–611.

KEDIA, S., AND S. RAJGOPAL (2011): "Do the SEC's enforcement preferences affect corporate misconduct?," *Journal of Accounting and Economics*, 51(3), 259–278.

KELLY, B., AND H. JIANG (2014): "Tail Risk and Asset Prices," *Review of Financial Studies*, 27(10), 2841–2871.

KIM, I., AND D. J. SKINNER (2012): "Measuring Securities Litigation Risk," *Journal of Accounting and Economics*, 53(1), 290–310.

LAKONISHOK, J., A. SHLEIFER, AND R. W. VISHNY (1994): "Contrarian Investment, Extrapolation, and Risk," *The Journal of Finance*, 49(5), 1541–1578.

LILIENFELD-TOAL, U. V., AND S. RUENZI (2014): "CEO Ownership, Stock Market Performance, and Managerial Discretion," *The Journal of Finance*, 69(3), 1013–1050.

MASHRUWALA, C., S. RAJGOPAL, AND T. SHEVLIN (2006): "Why Is The Accrual Anomaly Not Arbitraged Away? The Role of Idiosyncratic Risk and Transaction Costs," *Journal of Accounting and Economics*, 42(1-2), 3–33.

NGUYEN, N. T. M., A. IQBAL, AND R. K. SHIWAKOTI (2015): "There's No Smoke Without Fire: Does the Context of Earnings Management Contain Information About Future Stock Returns?," Working Paper.

NOVY-MARX, R. (2013): "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108(1), 1–28.

PALMROSE, Z.-V., V. J. RICHARDSON, AND S. SCHOLZ (2004): "Determinants of Market Reactions to Restatement Announcements," *Journal of Accounting and Economics*, 37(1), 59–89.

PASTOR, L., AND R. F. STAMBAUGH (2003): "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111(3), 642–685.

PETERSEN, M. A. (2009): "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches," *Review of Financial Studies*, 22(1), 435–480.

RICHARDSON, S. A., R. G. SLOAN, M. T. SOLIMAN, AND I. TUNA (2005): "Accrual reliability, earnings persistence and stock prices," *Journal of Accounting and economics*, 39(3), 437–485.

SADKA, R. (2006): "Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk," *Journal of Financial Economics*, 80(2), 309–349.

SHLEIFER, A., AND R. W. VISHNY (1997): "The Limits of Arbitrage," *The Journal of Finance*, 52(1), 35–55.

SLOAN, R. (1996): "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *The Accounting Review*, 71(3), 289–315.

STAMBAUGH, R. F., J. YU, AND Y. YUAN (2012): "The Short of It: Investor Sentiment and Anomalies," *Journal of Financial Economics*, 104(2), 288–302.

WERMERS, R. (2003): "Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence," Working Paper.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

YU, F., AND X. YU (2011): "Corporate Lobbying and Fraud Detection," *Journal of Financial and Quantitative Analysis*, 46(6), 1865–1891.

Figure 1: F-score coefficient estimates

The figure displays the coefficient estimates for the variables considered in Dechow, Ge, Larson, and Sloan (2011) F-score model 1. Following the approach in the original paper we calibrate the coefficients performing backward elimination and excluding variables with a significance level below 15%. Contrary to Dechow, Ge, Larson, and Sloan (2011), who perform only one regression for their full sample period from 1980 to 2005, we estimate 15 year rolling window regressions based on the years 1980 to 2009. The graphs show the evolution of the coefficient estimates for each variable in the model for the time windows moving forward. For detailed variable definitions please refer to the Appendix.

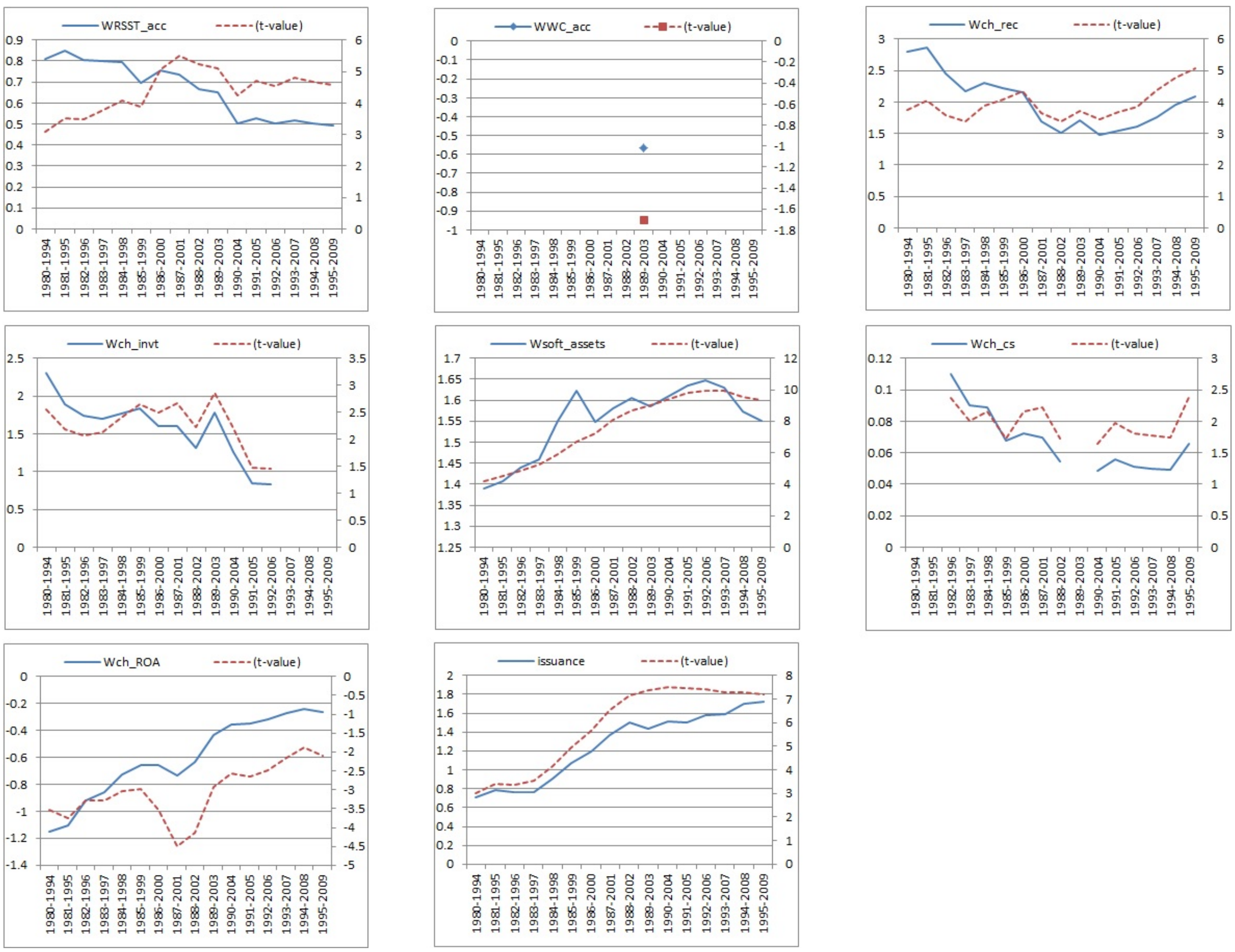

Figure 2: Distribution of misstatement firms across F-score deciles

This table examines detection rates of 416 misstatement firm years in the final data sample (with available return and F-score information). F-score parameters are calibrated on a 15 year rolling window. The statistics are based on the time period 1996 to 2009 (the horizon for which information about AAER is available and the F-score can be calibrated based on past data). Firms are sorted into decile portfolios PF1 to PF10 according to F-score values based on previous year (most recent publicly available) balance sheet information. The bars show the respective number and percentage of firms prosecuted by the SEC for each portfolio. The rhombs indicate the lower cutoff value of the F-score regarding the respective decile.

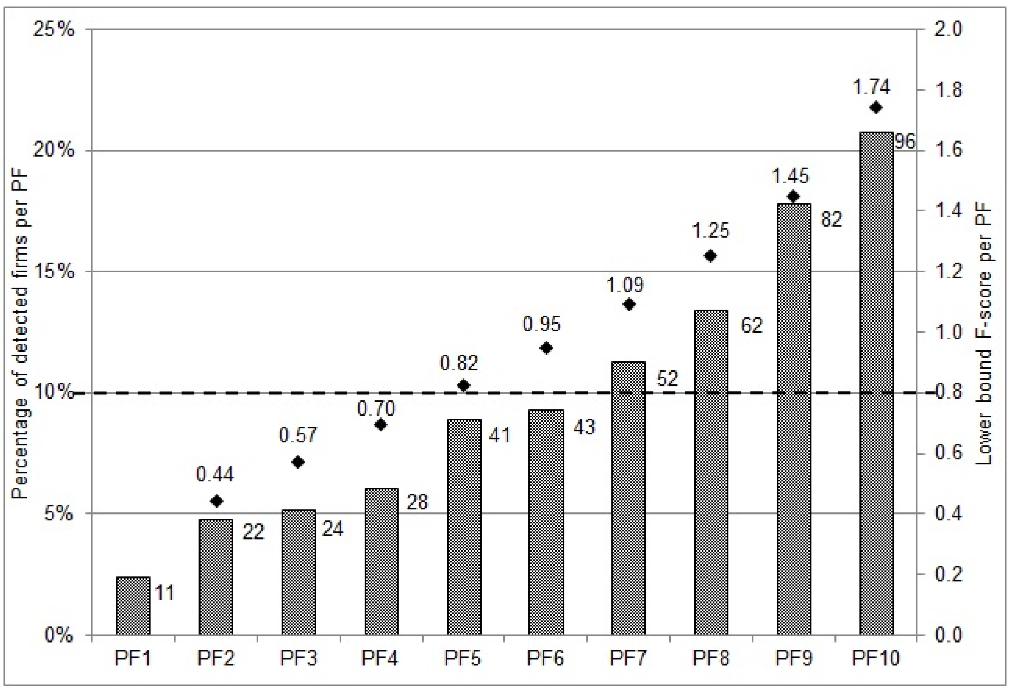

Figure 3: Persistence of F-score

The figure displays the development of the average equal-weighted F-score in ten decile portfolios. Firms are sorted into deciles based on their F-score in year t = 1. We follow the stocks in each portfolio during the subsequent four years and calculate for each year the equal-weighted average of F-scores.

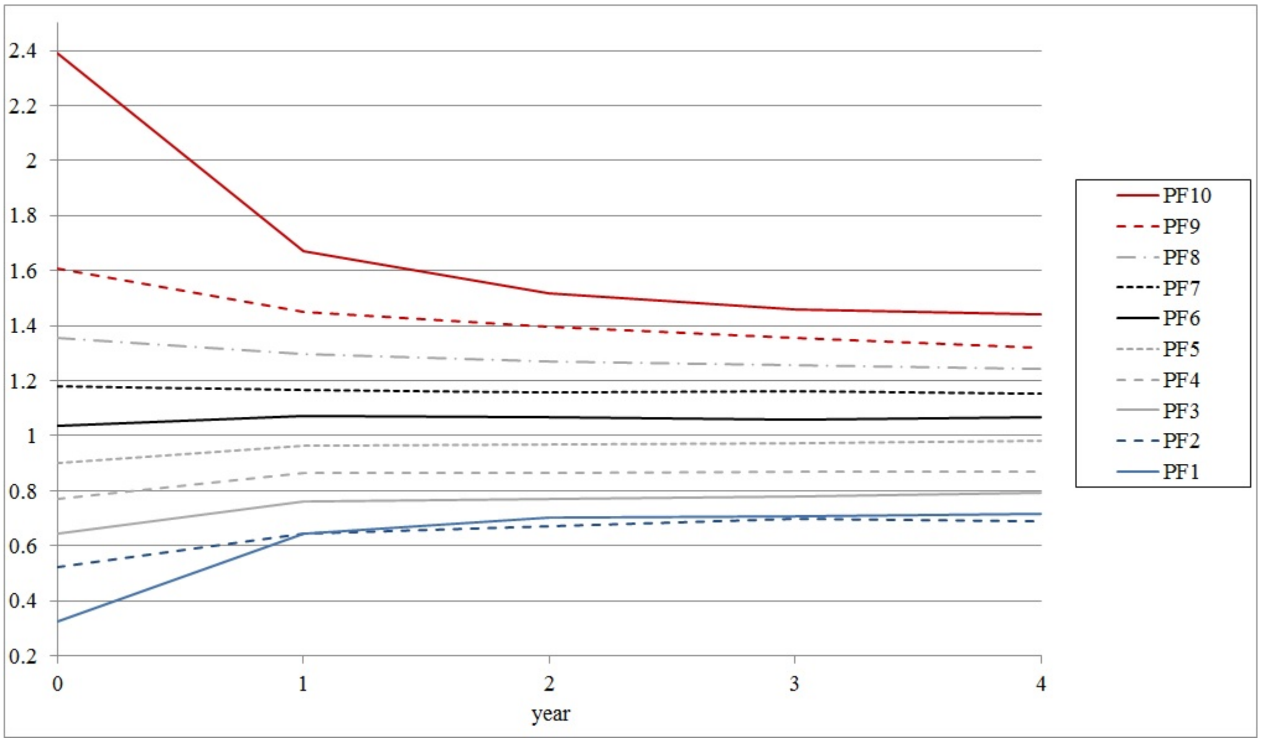

Table 1: Descriptive statistics

This table displays descriptive statistics for our sample including firm years with return information and all variables entering the F-score model. F-score parameters are calibrated on a 15 year moving window, and statistics in column (9) are based on the time period 1996 to 2012. In accordance with the asset pricing tests, penny stocks are excluded (i.e. if the price is below 1 USD, the stock is excluded in the subsequent month). In columns (7) to (9), we split the sample depending on whether a firm's F-score is below or above the median F-score of all firm-years included in columns (1) to (6). All variables which enter the F-score (accruals quality variables and performance variables) are winsorized at the 1% and 99% level. Sloan accruals are computed as in Sloan (1996) and distress risk is computed as in Campbell, Hilscher, and Szilagyi (2008). All variables are described in detail in the Appendix. Distress risk is shown as a percentage number while all other variables are given in decimal values. The number of observations on sales growth is lower, because this variable requires 4 years of lagged data.

	Sample statistics						Sample split by median F-score		
				Percentiles			low F	high F	Difference
	N	mean	SD	p1	p50	p99	mean	mean	in means
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
F-Score	51,754	1.056	0.633	0.140	0.951	3.281	0.616	1.496	0.880***
Accruals quality variables									
Work. cap. accruals	51,754	0.008	0.086	-0.274	0.005	0.286	-0.016	0.032	0.048***
RSST accruals	51,754	0.027	0.190	-0.605	0.027	0.691	-0.022	0.076	0.098***
Sloan accruals	51,677	-0.041	0.094	-0.350	-0.039	0.249	-0.070	-0.011	0.059***
Δ receivables	51,754	0.013	0.065	-0.189	0.007	0.242	-0.008	0.035	0.043***
Δ inventory	51,754	0.008	0.049	-0.150	0.000	0.197	-0.005	0.022	0.027***
% Soft-assets	51,754	0.54	0.242	0.036	0.566	0.952	0.373	0.706	0.333***
Performance variables									
Δ cash sales	51,754	0.166	0.712	-0.974	0.082	3.336	0.109	0.223	0.114***
Δ ROA	51,754	-0.004	0.146	-0.566	0.000	0.544	0.001	-0.008	-0.009***
Market related variables									
Issuance	51,754	0.921	0.270	0.000	1.000	1.000	0.844	0.997	0.153***
BM	48,628	0.696	4.327	-0.444	0.517	4.087	0.691	0.702	0.011
Size	51,482	2,624	13,924	4.222	238.7	46,189	2,444	2,803	359.0***
Price	51,482	19.42	29.67	1.000	12.54	89.60	17.76	21.06	3.300***
Yield	51,752	0.007	0.017	0.000	0.000	0.077	0.008	0.006	-0.002***
Sales growth	43,899	1.087	2.743	-0.885	0.389	19.911	1.007	1.164	0.157***
Asset growth	51,754	0.164	0.524	-0.601	0.065	2.991	0.124	0.203	0.079***
Distress risk	50,945	0.115	0.242	0.011	0.044	1.164	0.128	0.102	-0.026***

Table 2: Return analysis of value-weighted fraud portfolios

This table shows asset pricing tests for stocks of firms sorted into deciles according to their F-scores. The sample runs from 1996 to 2012. Penny stocks are excluded (i.e. if the price in the preceding month is below 1 USD an observation is excluded). Portfolios are formed when financial statement information becomes public (based on Compustat variable rdq date). Portfolios are sorted such that PF10 contains firms with the highest fraud probability. Portfolio returns are calculated as value-weighted averages. t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

Panel A: Raw returns	low F PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	high F PF10	long-short PF1-10
	0.775*	0.712*	0.559*	0.674**	0.621*	0.584*	0.587*	0.473	0.53	-0.334	1.109***
	(1.82)	(1.87)	(1.66)	(1.98)	(1.81)	(1.79)	(1.72)	(1.35)	(1.29)	(-0.71)	(3.62)
Panel B: CAPM											
α	0.268	0.259	0.156	0.244*	0.187	0.158	0.151	0.024	-0.003	-0.895***	1.163***
	(1.19)	(1.29)	(0.91)	(1.68)	(1.22)	(1.33)	(1.12)	(0.18)	(-0.02)	(-3.45)	(3.80)
MRP	1.089***	0.974***	0.867***	0.925***	0.930***	0.916***	0.937***	0.964***	1.146***	1.207***	-0.118*
	(20.72)	(20.05)	(21.45)	(22.92)	(24.23)	(34.27)	(34.92)	(27.91)	(32.47)	(18.09)	(-1.73)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.723	0.727	0.740	0.827	0.815	0.871	0.841	0.846	0.871	0.727	0.017
Panel C: Fama French 3F											
α	0.269	0.163	0.156	0.303**	0.195	0.148	0.236*	0.033	0.009	-0.758***	1.026***
	(1.34)	(0.83)	(0.93)	(2.09)	(1.30)	(1.24)	(1.85)	(0.23)	(0.06)	(-3.61)	(3.35)
MRP	0.984***	0.965***	0.841***	0.928***	0.928***	0.923***	0.939***	0.943***	1.124***	1.088***	-0.105
	(19.40)	(19.91)	(20.40)	(23.01)	(25.90)	(33.94)	(38.11)	(29.57)	(29.74)	(20.67)	(-1.44)
SMB	0.374***	0.203***	0.093*	-0.114**	-0.007	-0.006	-0.158***	0.061	0.058	0.177**	0.197
	(4.32)	(3.55)	(1.85)	(-2.45)	(-0.19)	(-0.15)	(-3.23)	(1.09)	(0.98)	(2.20)	(1.51)
HML	-0.176*	0.178**	-0.045	-0.113**	-0.017	0.030	-0.166***	-0.054	-0.062	-0.472***	0.296***
	(-1.95)	(2.53)	(-0.75)	(-2.27)	(-0.28)	(0.73)	(-3.39)	(-0.87)	(-1.19)	(-4.76)	(2.71)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.793	0.747	0.748	0.836	0.816	0.872	0.860	0.850	0.875	0.806	0.071
Panel D: Carhart 4F											
α	0.214	0.114	0.125	0.297**	0.207	0.119	0.243*	0.029	0.059	-0.673***	0.887***
	(1.07)	(0.57)	(0.74)	(2.03)	(1.36)	(1.00)	(1.82)	(0.21)	(0.40)	(-3.26)	(2.87)
MRP	1.018***	0.996***	0.861***	0.931***	0.920***	0.941***	0.935***	0.945***	1.091***	1.034***	-0.016
	(18.86)	(18.03)	(20.52)	(23.88)	(23.99)	(34.89)	(32.96)	(24.45)	(29.90)	(19.75)	(-0.19)
SMB	0.361***	0.192***	0.086*	-0.115**	-0.004	-0.013	-0.156***	0.060	0.070	0.197**	0.164
	(4.38)	(3.21)	(1.75)	(-2.44)	(-0.12)	(-0.32)	(-3.22)	(1.11)	(1.25)	(2.58)	(1.35)
HML	-0.150	0.200***	-0.030	-0.110**	-0.023	0.044	-0.169***	-0.052	-0.086*	-0.513***	0.362***
	(-1.54)	(2.76)	(-0.54)	(-2.26)	(-0.36)	(1.10)	(-3.26)	(-0.91)	(-1.75)	(-5.21)	(3.35)
UMD	0.080	0.071	0.046	0.008	-0.018	0.043*	-0.011	0.006	-0.075*	-0.125**	0.205***
	(1.49)	(1.51)	(1.22)	(0.29)	(-0.56)	(1.84)	(-0.29)	(0.16)	(-1.94)	(-2.57)	(2.98)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.798	0.752	0.750	0.836	0.816	0.874	0.860	0.850	0.879	0.815	0.133

Table 3: Return analysis of equal-weighted fraud portfolios

This table shows asset pricing tests for stocks of firms sorted into deciles according to their F-scores. The sample runs from 1996 to 2012. Penny stocks are excluded (i.e. if the price in the preceding month is below 1 USD an observation is excluded). Portfolios are formed when financial statement information becomes public (based on Compustat variable rdq date). Portfolios are sorted such that PF10 contains firms with the highest fraud probability. Portfolio returns are calculated as equal-weighted averages. t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

Panel A: Raw returns	low F PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	high F PF10	long-short PF1-10
	1.227**	0.905*	0.958**	0.937**	1.117**	0.937**	0.986**	0.861*	0.649	0.256	0.971***
	(2.51)	(1.96)	(2.12)	(2.07)	(2.45)	(2.11)	(2.20)	(1.95)	(1.45)	(0.52)	(4.60)
Panel B: CAPM											
α	0.707**	0.373	0.426*	0.391*	0.572**	0.398*	0.452**	0.330	0.111	-0.321	1.027***
	(2.26)	(1.45)	(1.77)	(1.75)	(2.49)	(1.85)	(1.99)	(1.51)	(0.50)	(-1.19)	(5.05)
MRP	1.119***	1.144***	1.147***	1.172***	1.173***	1.159***	1.150***	1.140***	1.156***	1.240***	-0.121***
	(17.74)	(20.54)	(22.40)	(24.99)	(22.16)	(24.61)	(19.75)	(20.71)	(21.23)	(21.32)	(-3.05)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.583	0.687	0.716	0.750	0.739	0.759	0.732	0.745	0.747	0.697	0.036
Panel B: Fama French 3F											
α	0.502**	0.142	0.240	0.216	0.352**	0.190	0.199	0.100	-0.101	-0.485**	0.987***
	(2.39)	(0.80)	(1.58)	(1.47)	(2.41)	(1.37)	(1.44)	(0.67)	(-0.70)	(-2.60)	(4.78)
MRP	0.941***	1.032***	1.011***	1.052***	1.067***	1.059***	1.057***	1.058***	1.053***	1.086***	-0.145***
	(15.91)	(23.31)	(24.23)	(27.22)	(25.21)	(29.02)	(25.75)	(26.09)	(26.78)	(24.27)	(-2.84)
SMB	1.000***	0.813***	0.816***	0.742***	0.771***	0.727***	0.783***	0.706***	0.747***	0.842***	0.158
	(8.89)	(12.03)	(12.99)	(10.72)	(10.41)	(10.96)	(12.25)	(8.61)	(10.60)	(8.44)	(1.07)
HML	0.112	0.271***	0.142**	0.147**	0.262***	0.246***	0.347***	0.323***	0.250***	0.071	0.041
	(1.42)	(4.14)	(2.29)	(2.47)	(4.01)	(4.24)	(5.30)	(4.50)	(3.97)	(0.95)	(0.51)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.825	0.859	0.901	0.901	0.898	0.907	0.903	0.889	0.901	0.867	0.067
Panel D: Carhart 4F											
α	0.565***	0.231	0.343**	0.336**	0.488***	0.306**	0.338***	0.233*	0.034	-0.329*	0.894***
	(2.63)	(1.30)	(2.39)	(2.45)	(3.72)	(2.40)	(2.71)	(1.71)	(0.25)	(-1.86)	(4.29)
MRP	0.900***	0.975***	0.945***	0.975***	0.980***	0.985***	0.968***	0.972***	0.966***	0.986***	-0.085*
	(15.45)	(20.77)	(22.20)	(25.23)	(26.26)	(26.65)	(28.34)	(27.27)	(24.22)	(20.09)	(-1.69)
SMB	1.014***	0.834***	0.840***	0.770***	0.802***	0.754***	0.815***	0.737***	0.778***	0.879***	0.136
	(8.87)	(13.16)	(14.30)	(12.87)	(12.54)	(12.99)	(14.96)	(11.67)	(15.17)	(10.81)	(1.03)
HML	0.082	0.228***	0.093	0.090	0.198***	0.191***	0.282***	0.260***	0.186***	-0.003	0.085
	(1.07)	(3.69)	(1.57)	(1.59)	(3.65)	(3.66)	(5.51)	(4.27)	(3.69)	(-0.05)	(1.13)
UMD	-0.094*	-0.131***	-0.151***	-0.177***	-0.201***	-0.170***	-0.205***	-0.196***	-0.198***	-0.231***	0.137**
	(-1.92)	(-3.55)	(-4.86)	(-5.02)	(-6.73)	(-4.99)	(-5.93)	(-4.72)	(-4.49)	(-4.00)	(2.19)
months	204	204	204	204	204	204	204	204	204	204	204
R^2	0.830	0.870	0.916	0.922	0.924	0.927	0.932	0.916	0.928	0.897	0.125

Table 4: Robustness checks

Panels A and B of this table show results based on Carhart (1997) alphas for modified F-score calibrations. In Panel A, columns (1) to (3) show results for calibrations based on 15 year rolling windows, while estimations in columns (4) to (6) are based on 10 year rolling windows. In Panel B, the time horizons are expanded every year with a starting window of 15 (10) years in columns (1) to (3) ((4) to (6)). Panel C presents results for two portfolio rebalancing approaches: In columns (1) to (3), the F-score coefficients are estimated once based on the earliest 15 year horizon and then kept constant afterwards. Still, portfolios are formed monthly as in the baseline specification. In columns (4) to (6), the F-score is calibrated based on a 15-year rolling window (as in the baseline specification), but portfolio formation is conducted on a yearly frequency only. The analysis runs from 1996 to 2012 for the 15 year calibration horizons, and from 1991 to 2012 for the 10 year calibration horizons. All portfolio returns are value-weighted averages. Penny stocks are excluded (i.e. if the stock price in the preceding month is below 1 USD). Portfolios are formed when financial statement information becomes public (based on Compustat variable rdq date). t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

Panel A: Modified rolling windows

	15-year rolling window			10-year rolling window		
	low F (1)	high F (2)	low-high (3)	low F (4)	high F (5)	low-high (6)
α	0.214	-0.673***	0.887***	0.146	-0.513***	0.659**
	(1.07)	(-3.26)	(2.87)	(0.87)	(-2.90)	(2.49)
MRP	1.018***	1.034***	-0.016	0.973***	1.087***	-0.114
	(18.86)	(19.75)	(-0.19)	(19.95)	(23.09)	(-1.49)
SMB	0.361***	0.197**	0.164	0.465***	0.173**	0.292**
	(4.38)	(2.58)	(1.35)	(5.60)	(2.22)	(2.26)
HML	-0.150	-0.513***	0.362***	-0.017	-0.540***	0.522***
	(-1.54)	(-5.21)	(3.35)	(-0.26)	(-6.21)	(4.28)
UMD	0.080	-0.125**	0.205***	0.125***	-0.094**	0.220***
	(1.49)	(-2.57)	(2.98)	(2.85)	(-2.04)	(2.95)
months	204	204	204	264	264	264
R^2	0.798	0.815	0.133	0.786	0.815	0.211

Panel B: Expanding estimation windows

	15-year expanding window			10-year expanding window		
	low F (1)	high F (2)	low-high (3)	low F (4)	high F (5)	low-high (6)
α	0.275	-0.687***	0.962***	0.207	-0.535***	0.742***
	(1.38)	(-3.32)	(3.17)	(1.23)	(-2.98)	(2.83)
MRP	1.000***	1.037***	-0.037	0.969***	1.063***	-0.094
	(18.62)	(21.08)	(-0.49)	(20.04)	(22.52)	(-1.29)
SMB	0.381***	0.125	0.256**	0.425***	0.148*	0.277**
	(4.68)	(1.58)	(2.02)	(5.71)	(1.97)	(2.28)
HML	-0.122	-0.495***	0.372***	-0.079	-0.500***	0.421***
	(-1.25)	(-5.09)	(3.49)	(-0.89)	(-5.69)	(4.28)
UMD	0.079	-0.153***	0.232***	0.060	-0.128***	0.189***
	(1.46)	(-3.03)	(3.30)	(1.15)	(-2.64)	(2.72)
months	204	204	204	264	264	264
R^2	0.795	0.810	0.168	0.781	0.798	0.165

Panel C: Modified trading strategy

	constant coefficients			yearly rebalancing		
	low F (1)	high F (2)	low-high (3)	low F (4)	high F (5)	low-high (6)
α	0.125	-0.734***	0.858***	0.365	-0.268	0.633**
	(0.56)	(-3.52)	(2.73)	(1.53)	(-1.34)	(2.07)
MRP	1.121***	1.023***	0.097	1.004***	1.030***	-0.026
	(19.72)	(18.96)	(1.25)	(17.50)	(20.49)	(-0.35)
SMB	0.524***	0.192**	0.332**	0.380***	0.243***	0.137
	(6.50)	(2.34)	(2.60)	(3.62)	(3.40)	(0.99)
HML	-0.196*	-0.521***	0.325***	-0.007	-0.459***	0.452***
	(-1.88)	(-5.24)	(2.92)	(-0.09)	(-5.97)	(3.88)
UMD	0.120*	-0.175***	0.295***	0.103*	-0.209***	0.312***
	(1.92)	(-3.68)	(4.17)	(1.82)	(-4.45)	(3.94)
months	204	204	204	204	204	204
R^2	0.806	0.814	0.188	0.732	0.839	0.218

Table 5: Multivariate firm-level evidence

This table presents multivariate firm-level regressions of F-score deciles and control variables on stock returns. In columns (1) to (5), we run panel regression models. Column (6) presents Fama and MacBeth (1973) estimates and shows average coefficients and time-series standard errors for 204 monthly cross-sectional regressions from January 1996 to December 2012. The analysis with DGTW fixed effects in column (3) ends in December 2010 due to the availability of DGTW data. In all regressions, the dependent variable is a firm's monthly stock return minus the risk free rate. We use control variables as in Gompers, Ishii, and Metrick (2003) and add Sloan (1996) accruals as well as an indicator for high financial distress risk according to Campbell, Hilscher, and Szilagyi (2008). All variables are described in detail in the Appendix. F-score parameters are calibrated on a 15 year rolling window. The full time-horizon available spans 1980-2009. For 2010 - 2012 the parameters from 2009 are used. Penny stocks are excluded (i.e. if the stock price in the preceding month is below 1 USD). Portfolios are formed when financial statement information becomes public (based on Compustat data item rdq date). Portfolios are sorted such that PF10 contains firms with the highest probability of committing fraud. t-statistics are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

	Month FE (1)	Month × industry FE (2)	Month × DGTW FE (3)	Month + firm FE (4)	Month +firm FE (5)	Fama-MacBeth (1973) (6)
Fraud decile$_{t-1}$	-0.041***	-0.040***	-0.046***	-0.036**	-0.036*	-0.040**
	(-4.29)	(-3.85)	(-4.39)	(-2.02)	(-1.70)	(-2.13)
NASDUM$_{t-1}$	1.176***	0.877***	1.196***	2.906***	2.906***	0.807**
	(5.32)	(3.99)	(4.91)	(5.14)	(3.34)	(2.21)
SP500$_{t-1}$	-0.085	-0.059	0.064	-0.360*	-0.360	0.019
	(-1.15)	(-0.81)	(0.62)	(-1.70)	(-1.22)	(0.11)
Tech Dum$_{t-1}$	0.126**	-0.272*	0.182***	-0.171	-0.171	0.094
	(2.41)	(-1.89)	(3.23)	(-0.69)	(-0.73)	(0.66)
logBM$_{y-1}$	0.330***	0.348***	0.283***	0.361***	0.361***	0.209***
	(9.82)	(10.25)	(5.85)	(5.02)	(3.17)	(3.27)
logSize$_{t-2}$	-0.158***	-0.147***	-0.205***	-2.905***	-2.905***	-0.072
	(-4.10)	(-3.77)	(-3.84)	(-25.03)	(-11.42)	(-0.59)
Price$_{t-2}$	0.002**	0.001**	0.000	0.012**	0.012*	-0.001
	(2.21)	(2.20)	(0.28)	(2.10)	(1.96)	(-0.48)
logNYDVOL$_{t-2}$	0.144***	0.132***	0.188***	0.210***	0.210	0.054
	(5.43)	(4.82)	(6.00)	(4.23)	(1.52)	(0.61)
logNADVOL$_{t-2}$	0.057**	0.067***	0.096***	-0.022	-0.022	0.001
	(2.20)	(2.58)	(3.13)	(-0.49)	(-0.15)	(0.01)
Yield$_{y-1}$	0.737	2.252	0.587	-0.505	-0.505	-3.731
	(0.56)	(1.61)	(0.39)	(-0.23)	(-0.14)	(-1.55)
logret$_{t-2,t-3}$	-0.208	-0.482***	-0.543**	-0.173	-0.173	0.350
	(-1.21)	(-2.78)	(-2.55)	(-0.96)	(-0.17)	(0.83)
logret$_{t-4,t-6}$	-0.254**	-0.307***	-0.437***	-0.175*	-0.175	0.005
	(-2.52)	(-2.95)	(-3.57)	(-1.67)	(-0.38)	(0.02)
logret$_{t-7,t-12}$	0.150*	0.143	-0.322**	0.378***	0.378	0.243
	(1.67)	(1.53)	(-2.36)	(3.89)	(1.04)	(1.04)
Sales growth$_{y-6,y-1}$	-0.032***	-0.040***	-0.027**	0.026	0.026	-0.028*
	(-2.68)	(-3.56)	(-2.23)	(1.39)	(1.25)	(-1.95)
Asset growth$_{y-1}$	-0.462***	-0.391***	-0.492***	-0.156**	-0.156	-0.376***
	(-7.41)	(-6.34)	(-7.42)	(-2.17)	(-1.46)	(-4.90)
Sloan accruals$_{y-1}$	-1.514***	-1.552***	-1.527***	-1.362***	-1.362**	-1.136**
	(-3.92)	(-4.08)	(-3.74)	(-2.81)	(-2.43)	(-2.55)
CHS distress$_{t-1}$, p90	-0.441***	-0.438***	-0.282*	-0.368**	-0.368*	-0.416*
	(-3.11)	(-3.15)	(-1.79)	(-2.22)	(-1.70)	(-1.83)
Month Fixed Effects	Yes	No	No	Yes	Yes	Yes
Firm Fixed Effects	No	No	No	Yes	Yes	No
Combined FE	No	Yes	Yes	No	No	No
Clustered SE, firm	Yes	Yes	Yes	Yes	Yes	No
Clustered SE, month	No	No	No	No	Yes	No
N (months)	204	204	180	204	204	204
N	457,612	457,612	412,622	457,612	457,612	457,612
R^2	0.125	0.184	0.175	0.150	0.150	0.062

Table 6: Analysis of different factor models

This table shows asset pricing tests for the F-score long-short portfolio on different factor models. The long-short portfolio is based on alphas of the value-weighted return portfolios of low minus high fraud risk firms. The time horizon included depends on the availability of the asset pricing factors. F-score parameters are calibrated based on a 15 year moving window. The full time-horizon available spans 1980 to 2009. For 2010 and 2011 the parameters from 2009 are used. Penny stocks are excluded (i.e. if the price in the preceding month is below 1 USD). Portfolios are formed when financial statement information becomes public (based on Compustat data item rdq date). All variables are described in detail in the Appendix. t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

	Fama/French 5 (1)	q-Factors (2)	Gross Profitability (3)	Quality-Junk (4)	Betting-against-Beta (5)
α	0.981***	0.872***	1.105***	1.101***	0.908***
	(3.07)	(2.61)	(3.66)	(3.59)	(2.73)
MRP	-0.076	-0.059	-0.058	-0.120	-0.011
	(-1.00)	(-0.64)	(-0.75)	(-1.34)	(-0.14)
SMB	0.055		0.129	0.034	0.158
	(0.42)		(1.15)	(0.24)	(1.26)
HML	0.099		0.311***	0.357***	0.350***
	(0.60)		(2.92)	(3.28)	(3.02)
UMD			0.204***	0.235***	0.202***
			(3.13)	(3.08)	(2.74)
rmw	-0.235				
	(-1.40)				
cma	0.592**				
	(2.54)				
q_size		0.168			
		(1.36)			
q_invst		0.665***			
		(3.91)			
q_profit		-0.052			
		(-0.35)			
pmu			-0.408***		
			(-2.77)		
qmj				-0.302*	
				(-1.74)	
bab					0.015
					(0.17)
	Lottery Factor (6)	**Kelly Tail Risk (7)**	**Lower Tail Dep. (8)**	**PS Liquidity (9)**	**Sadka Liquidity (10)**
α	0.853**	0.936***	0.871**	0.853**	0.866**
	(2.38)	(3.16)	(2.40)	(2.43)	(2.52)
MRP	-0.025	0.007	0.039	0.026	0.035
	(-0.25)	(0.09)	(0.29)	(0.32)	(0.40)
SMB	0.137	0.182	0.146	0.173	0.175
	(0.80)	(1.56)	(1.11)	(1.39)	(1.40)
HML	0.405**	0.366***	0.364***	0.388***	0.396***
	(2.23)	(3.35)	(3.05)	(3.47)	(3.37)
UMD	0.209**	0.168**	0.194***	0.215***	0.221***
	(2.19)	(2.46)	(2.66)	(3.05)	(3.04)
max_vw	0.036				
	(0.29)				
Kelly_vw_5_1		-0.121			
		(-1.10)			
LTD_vw_5_1			-0.082		
			(-0.66)		
ps_liqu_td				0.015	
				(0.22)	
sadka_liqu					-0.210
					(-0.37)
N (months)	168	204	168	180	180
R^2	0.129	0.141	0.131	0.139	0.140

Table 7: Limits of arbitrage analysis

This table shows estimation results of the Carhart (1997) four-factor model for value-weighted long-short fraud portfolios. Stocks are selected into portfolios using unconditional double-sorts based on firms' F-score and proxies for limits of arbitrage. We obtain long-short fraud portfolio returns as the difference between highest and lowest F-score decile for above and below median values of the respective proxy variable (except in column (7)). The proxy is named in the header of each column. The proxies on information asymmetry are based on I/B/E/S number of analysts forecasting 1-year earnings per share. We obtain long-short portfolio returns as the difference between firms with below or above median analyst coverage (column (6)) and for no versus any analyst coverage (column (7)). F-score parameters are calibrated on a 15 year rolling window. The full time-horizon available spans 1980 to 2009. For 2010 and 2011 the parameters from 2009 are used. Penny stocks are excluded (i.e. if the price in the preceding month is below 1 USD). Portfolios are formed when financial statement information becomes public (based on Compustat data item rdq date). Portfolios are sorted on F-score such that PF10 contains firms with the highest probability of committing fraud. t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

	Liquidity		Risk		Information Asymmetry	
	Corwin/Schultz Spread (1)	Amihud Illiq. Ratio (2)	Volatility (3)	Idiosyncr. Volatility (4)	Number of Analysts (5)	Any Analyst Coverage (6)
high limits of arbitrage	0.998*	0.602**	1.356***	1.084**	0.903*	0.542
$(PF1 - PF10)$	(1.95)	(2.57)	(3.10)	(2.35)	(1.92)	(1.06)
low limits of arbitrage	1.143***	1.107***	0.835**	1.069***	1.008***	1.008***
$(PF1 - PF10)$	(3.68)	(3.41)	(2.34)	(3.06)	(3.14)	(3.19)
Difference	-0.145	-0.505	0.521	0.015	-0.105	-0.466
	(-0.26)	(-1.33)	(1.02)	(0.03)	(-0.21)	(-0.87)
N (months)	204	204	204	204	204	204

Table 8: The impact of investor sentiment

Panel A of this table reports value-weighted raw returns (over the risk free rate) and Panel B reports Carhart (1997) model excess returns on ten F-score sorted portfolios. We distinguish periods following months of high or low investor sentiment. High and low sentiment periods are identified by the median level of the sentiment index of Baker and Wurgler (2006), obtained from the author's homepage. F-score parameters are calibrated on a 15 year rolling window. The full time-horizon available spans 1980 to 2009. For 2010 and 2011, the parameters from 2009 are used. The asset pricing analysis covers 1996 to 2012. Penny stocks are excluded (i.e. if the price in the preceding month is below 1 USD). Portfolios are formed when financial statement information becomes public (based on Compustat data item rdq date). Portfolios are sorted such that PF10 contains firms with the highest probability of earnings overstatement. t-statistics for White (1980) standard errors are given in parentheses. Significance levels are indicated as *** 1%-level, ** 5%-level, * 10%-level.

	low F PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	high F PF10	long-short PF1-10
Panel A: Raw Returns											
$HighSent$	0.094	0.007	0.083	0.058	-0.053	0.026	-0.062	-0.311	-0.24	-1.979**	2.073***
	(0.13)	(0.01)	(0.16)	(0.10)	(-0.10)	(0.05)	(-0.11)	(-0.54)	(-0.35)	(-2.32)	(3.92)
$LowSent$	1.411***	1.549***	1.052**	1.376***	1.295**	1.209**	1.259***	1.135***	1.232**	1.300***	0.111
	(2.99)	(3.09)	(2.10)	(2.97)	(2.60)	(2.27)	(3.45)	(2.86)	(2.33)	(3.10)	(0.63)
Panel B: Carhart 4F											
$\alpha_{HighSent}$	0.317	0.040	0.266	0.376	0.192	0.181	0.312	-0.072	0.168	-1.289***	1.606***
	(0.95)	(0.13)	(0.96)	(1.60)	(0.74)	(1.02)	(1.22)	(-0.31)	(0.68)	(-3.61)	(2.96)
$\alpha_{LowSent}$	-0.031	0.359	-0.009	0.324	0.193	0.117	0.219	-0.047	-0.142	-0.079	0.048
	(-0.12)	(1.40)	(-0.03)	(1.51)	(1.04)	(0.61)	(1.49)	(-0.25)	(-0.72)	(-0.29)	(0.12)

A Appendix

A.1 Variable definitions

Asset Growth The year-on-year percentage growth rate of current assets of a firm over the preceding fiscal year (from fiscal year ending *y-2* to *y-1*) from Compustat (item act).

logBM The natural logarithm of the ratio of book value of common equity (previous fiscal year from Compustat) to market value of common equity (market cap calculated form CRSP data for the same month when previous fiscal year book equity information is available). Book value of common equity is the sum of book common equity (Compustat item ceq) and deferred taxes (Compustat item txdb).

logNADVOL The natural logarithm of the dollar volume of trading in month *t-2* for stocks traded on the NASDAQ. Approximated as stock price at the end of month *t-2* multiplied by share volume in month *t-2*.

logNYDVOL The natural logarithm of the dollar volume of trading in month *t-2* for stocks traded on the NYSE or Amex. Approximated as stock price at the end of month *t-2* multiplied by share volume in month *t-2*.

logret2_3 The natural logarithm of the compounded gross returns for months *t-2* and *t-3*.

logret4_6 The natural logarithm of the compounded gross returns for months *t-6* through *t-4*.

logret7_12 The natural logarithm of the compounded gross returns for months *t-12* through *t-7*.

logSize The natural logarithm of the firm's market capitalization in thousand USD from CRSP stocks at the end of month *t-2*.

NASDUM A dummy variable equal to one if the firm traded on NASDAQ at the end of month *t-1* and zero otherwise.

Price The price in USD at which the firm's stock is traded from the CRSP stock database at the end of month *t-2*.

SP500 A dummy variable equal to one for constituents of the S&P 500 as of the end of month *t-1*. Data are from CRSP S&P 500 constituent file.

Sales Growth The sales growth (Compustat item sale) of a firm over the past five fiscal years *y-6* through *y-1*.

Tech Dummy A dummy variable taking the value of one, if a firm belongs to a tech industry (defined as in Anderson and Reeb (2003) as firms belonging to the two-digit SIC codes 34, 36, 38 and 73), and zero otherwise at the end of month *t-1*.

Yield The dividend yield of a firm computed as ratio dividends per share by ex-date to price at fiscal year close. According to Compustat user guide. Measured at the end of the previous fiscal year.

A.2 Variables entering the F-score calibration

This table shows definitions of the variables included in the F-score estimation (see Dechow, Ge, Larson, and Sloan (2011)). Variables in lower case letters refer to Compustat abbreviations while upper case letters refer to auxiliary variables.

Variable	Calculation
WC accruals	$[(\Delta act_y - \Delta che_y) - (\Delta lct_y - \Delta dlc_y - \Delta txp_y)]/[0.5 * (at_y + at_{y-1})]$
RSST accruals	$(\Delta WC_y + \Delta NCO_y + \Delta FIN_y)/[0.5 * (at_y + at_{y-1})]$ $WC = act - che - lct + dlc$ $NCO = at - act - ivao - lt + lct + dltt$ $FIN = ivst + ivao - dltt - dlc - pstk$
Δ receivables	$\Delta rect_y/[0.5 * (at_y + at_{y-1})]$
Δ inventory	$\Delta invt_y/[0.5 * (at_y + at_{y-1})]$
% soft-assets	$(at_y - ppent_y - che_y)/at_y$
Δ cash sales	$(sale_y - \Delta rect_y)/(sale_{y-1} - \Delta rect_{y-1}) - 1$
Δ ROA	$ib_y/[0.5 * (at_y + at_{y-1})] - ib_{y-1}/[0.5 * (at_{y-1} + at_{y-2})]$
issuance	Indicator variable which is equal to 1 if a firm issued equity ($sstk > 0$) or debt ($dltis > 0$).