# Extremal Quantiles and Stock Price Crashes

Panayiotis C. Andreou *†
Sofia Anyfantaki‡
Esfandiar Maasoumi§
Carlo Sala¶

**July 2023**

**Abstract**

We employ extreme value theory to identify stock price crashes, featuring low-probability events that produce large, firm-specific negative outliers in the conditional distribution. Traditional methods employ approximations under Gaussian assumptions and central moments. This is inherently imprecise and susceptible to misspecifications, especially for tail events. We instead propose new definitions and measures for crash risk based on conditional extremal quantiles (CEQ) of idiosyncratic stock returns. CEQ provide information on quantile specific impact of covariates, and shed light on prior empirical puzzles and shortcomings in identifying crashes. Additionally, to capture the magnitude of crashes, we provide an expected shortfall analysis of the losses due to crash. Our findings have important implications for a burgeoning literature in financial economics that relies on traditional approximations.‡‡

**Keywords:** Stock price crashes; Extremal quantiles; Extreme value theory; Quantile regression
**JEL Codes:** C14; D81; G11; G12; G32.

*Department of Finance, Accounting and Management Science, Cyprus University of Technology, Lemesos, Cyprus; e-mail: panayiotis.andreou@cut.ac.cy.

†Durham University Business School, Durham University, Mill Hill Lane, Durham DH1 3LB, UK.

‡Bank of Greece, Economic Analysis & Research Department, 21, El. Venizelos Ave, 10250, Athens, Greece; e-mail: SAnyfantaki@bankofgreece.gr. The views expressed in this article are those of the authors and not necessarily reflect those of the Bank of Greece or the Eurosystem.

§Department of Economics, Emory University, 1602 Fishburne Drive, Atlanta, GA, 30322; e-mail: esfandiar.maasoumi@emory.edu

¶ESADE Business School, Ramon LLull University, Avenida de Torreblanca 59, 08172, Sant Cugat, Barcelona, Spain; e-mail: carlo.sala@esade.edu. Financial support from the AGAUR - SGR 2017-640 grant and from the Spanish Ministry of Science and Innovation - PID2019-1064656GB-I00/AEI/10.13039/501100011033 are gratefully acknowledged.

‡‡ This contribution honors the memory of Michael McAleer, a leading econometrician, a long time Associate Editor of ER, and a great friend. Mike thought deeply about volatility, Value at Risk (VaR), Basel rules, and financial econometrics, generally. A very few examples are , Chang, Jiménez-Martín, Maasoumi, McAleer, and Pérez-Amaral (2019), Hoti, Maasoumi, McAleer, and Slottje (2009) and Maasoumi and McAleer (2006). We thank the editor, R. Taylor and two referees for helpful comments and guidance.

# 1 Introduction

A burgeoning body of literature in financial economics has examined firm-specific stock price crashes, defined as low-probability events that produce a large and negative outlier in the conditional distribution of idiosyncratic returns.[1] These studies predominantly use the *traditional method* under (implicit or explicit) Gaussian assumptions, and rely on central moments of conditional distributions to define stock price crashes. In this paper, we advocate an alternative approach to define crash measures based on conditional extremal quantiles (CEQ) of idiosyncratic returns. Estimating the conditional quantiles of a response variable given a set of covariates, quantile regression analysis is a useful tool for tail events. The proposed methodology allows firm or industry specific covariate effects, which may be different in the central and tail regions of the conditional distribution of idiosyncratic returns. As such, our approach departs from the existing literature by challenging the common practice of defining firm-specific stock price crashes as extreme tail events, identified with dichotomous distance measures defined as $\lambda$-standard-deviations below the mean of idiosyncratic returns.

We employ the inference tools of Chernozhukov and Fernández-Val (2011) for extremal conditional quantile models. It is widely recognized that left-tail events are rare and challenging to predict accurately. In comparison to regular quantile regression, extremal quantile regression has been demonstrated to be a superior method for capturing extreme tail events such as crashes, as highlighted by Chernozhukov and Fernández-Val (2011).

It is worth noting that the CEQ model does not impose strong assumptions on the distribution function of the underlying error term. Unlike the traditional method, the CEQ model estimates allow for data-dependent and quantile-specific effects. This flexibility enables the model to effectively capture nonlinearities in the data, as demonstrated by Chernozhukov

[1] Some notable examples, *inter alia*, are: Chen, Hong, and Stein (2001); Jin and Myers (2006); Hutton, Marcus, and Tehranian (2009); Kim, Li, and Zhang (2011); Callen and Fang (2015); Andreou, Antoniou, Horton, and Louca (2016); Kim, Wang, and Zhang (2016); Andreou, Louca, and Petrou (2017); Chang, Chen, and Zolotoy (2017); Ertugrul, Lei, Qiu, and Wan (2017); Cheng, Li, and Zhang (2020); Li and Zeng (2019).

and Fernández-Val (2011). This provides a heterogeneous model for the conditional probability of stock price crashes. In other words, the CEQ model takes into account covariates that exhibit different quantile-specific effects on crashes, for different industries and firm types. This feature allows for a more nuanced and comprehensive understanding of the factors influencing the occurrence and severity of stock price crashes.

In our empirical application, we investigate idiosyncratic weekly returns as data dependent deviations from market and/or industry returns, for the period 2000 to 2019, covering common stocks traded on NYSE, AMEX and NASDAQ. We perform a data driven idiosyncratic "excess return" analysis. The projected dynamic and other time series market and industry effects is accounted for and removed to define idiosyncratic returns (see Section 2 for further elaboration).

We provide a comparative analysis between the proposed CEQ method and its Gaussian counterpart. Assuming a Gaussian distribution of weekly returns, we would expect to see 0.1% of sample firms crashing in any given week based on a 3.09-standard-deviation definition of crash, as in Hutton et al. (2009). This corresponds to a *theoretical* crash probability of $1-(1-0.001)^{52} = 5.07\%$ over the course of a year. Andreou, Lambertides, and Magidou (2022) find that between 1950 and 2019 the *empirical* frequency of idiosyncratic stock price crashes for US-listed firms has steadily increased from 5.5% to an astonishing 27%. The huge disparity between the empirical *vs.* theoretical thresholds is what Andreou et al. (2022) coin as the *stock price crash risk puzzle*, and call for more research to rationalize it.

We find that the occurrence of stock price crashes is more prevalent than would be expected under a Gaussian (or implicitly Gaussian) assumption for idiosyncratic returns. In our sample of 62,657 firm-year observations, we identify 12,456 firm-year observations or 20.28% of the total sample as crashes based on the widely used 3.09-standard-deviation definition as introduced by Hutton et al. (2009). As previously documented by Andreou et al. (2022), we observe that the percentage of stock price crashes has risen over the last decade based on this definition, which is consistent with the notion that factors such as exploding firm specific leverage cause returns to deviate from a normal distribution.

Kelly and Jiang (2014) argue that a single process governs tail risk fluctuations for all assets. Building on this notion, we adopt a similar approach by pooling firm-year data at the industry level. We annually estimate the CEQ of pooled weekly idiosyncratic returns, conditioning on a comprehensive set of covariates commonly employed in specifications for negative skewness (Chen et al., 2001). Focusing on the $0.1th$ quantile, we classify a firm as experiencing a stock price crash in a given year if, during that year, at least one weekly return falls below the estimated CEQ. In other words, a stock price crash is defined as a return that is unexpected - an extreme event - conditional of stock/firm specific characteristics. Surprisingly, our findings reveal that the number of firm-year crashes identified by $CEQ$ estimate is significantly lower. Specifically, we identify only 6,447 firm-year observations or 10.48% of the total sample as crashes.

In addition to examining the frequency of crashes, we also estimate their magnitude through an expected shortfall analysis. Acting as a proxy for the losses, the expected shortfall analysis allows us to evaluate the monetary impact of the miss-estimation. Interestingly, we find that the $CEQ$ losses are much larger and significant compared to the ones estimated using under the Gaussian approach.

One of the key assumptions behind CEQ regression is that the covariates can impact the extremal quantiles and the central quantiles very differently (Chernozhukov, 2005). Chen et al. (2001)– under Gaussian assumption–find that firms that experience larger increases in turnover relative to trend, are predicted to have larger negative skewness, with the effect of turnover being statistically and economically significant. Interestingly, using CEQ to estimate extreme risk reveals a different story. We find that the detrended level of turnover has a significant negative coefficient on the lower tail confirming the findings of Jiang, Wu, Zhou, and Zhu (2020), that higher turnover implies a lower crash probability. This indicates that the traditional method may not accurately capture information of the extreme left-tail events, as it is based on a sample that, likely, does not include observations of these events.

In connection with the above, we also support previous evidence that methods relying on the extreme value theory provide more accurate tails approximations than the Gaussian one. Our empirical findings show that the coefficients and their respective 90% confidence intervals (CIs) are more unstable at the tails. We find that, into the tails, Gaussian approximate CIs are often narrower than extremal CIs, especially for values of the quantile index less than 0.08. These findings reveal that there is additional significant estimation uncertainty into the lower tails of returns.

Lastly, in a separate application, we confirm the results of Andreou et al. (2022) who highlight that the exceedingly high and rising (over time) frequency of crashes cannot be adequately explained by either financial reporting opacity or overinvestment, identified in earlier literature as the two main channels of crash risk (Hutton et al., 2009; Kim et al., 2011; Callen and Fang, 2015; Andreou et al., 2016; Kim et al., 2016; Andreou et al., 2017). Yet, using the CEQ crash estimates we reach two interesting results. First, the frequency of crashes remains relatively constant across years, and does not present the difficult-to-rationalize upward trending behavior that fuels the stock price crash risk puzzle. Second, within a pooled cross-sectional regression setting, we find that both opacity and overinvestment are positive and statistically related to the one-year-ahead stock price crashes. This evidence is well-aligned with the agency models of Jin and Myers (2006) and Benmelech, Kandel, and Veronesi (2010) that theorize opacity and overinvestment as the channels that managers strategically exploit to camouflage bad news. Our results might offer some explanations to help explaining the empirical evidence in Andreou et al. (2022) who, with a similar regression analysis, find that both opacity and overinvestment are statistically nonsignificant. In this vein, we suggest that Andreou et al. (2022) reach statistically nonsignificant results because, the traditional approach they use for estimating crashes, is possibly generating several "false-positive" occurrences.

Our findings align with the understanding that a stock price crash entails the likelihood of observing a substantial negative outlier specific to the distribution of equity returns. To the best of our knowledge, this study is the first one to investigate the benefits of a non-Gaussian quantile approach to examine stock price crashes in a cross-sectional context.

Our study also contributes to the existing literature on the relationship between crash risk and its determinants, an area of research that, as aforementioned, has garnered considerable attention and continues to thrive in recent years (see, for instance, footnote 1). As a main contribution, we offer a novel method that can be explored by future studies to estimate stock price crashes, thereby extending the toolkit available to researchers in this field. Additionally, our empirical findings provide valuable evidence that could contribute, to some extent, to the rationalization of the stock price crash risk puzzle documented by Andreou et al. (2022).

Finally, while we do not directly test the hypothesis that tail risk explains differences in expected returns across stocks, our results offer insights that can contribute to this line of inquiry. The recent empirical asset pricing literature has placed significant emphasis on exploring the connections between left tail risk and the cross-section of expected stock returns, highlighting the relevance of our research in this domain (see subsection 2.3 for related literature).

The remainder of the paper is organized as follows. Section 2 presents the methodology for estimating the newly introduced CEQ crash measure, along with a discussion of the conventional crash risk measures commonly employed in the literature. Additionally, this section provides an overview of the relevant literature related to our study. In Section 3, we present the empirical application of our methodology. Section 4 further extends the empirical analysis by examining the opacity and overinvestment channels. Finally, Section 5 concludes the paper, summarizing the key findings and highlighting the contributions of our research.

# 2 Measuring stock price crashes

## *2.1 The conditional extremal quantiles method*

Chernozhukov (2005) extends the extreme value theory to develop extreme quantile regression models for the tails of conditional distributions, and broaden the properties of the Koenker and Bassett Jr (1978) quantile regression estimator. Chernozhukov and Fernández-Val (2011) made use of self-normalized quantile regression statistics to develop feasible and practical inferential methods for the quantile regression of Chernozhukov (2005), when the quantile index is either low (close to zero), or high (close to one). Many important applications of quantile regression involve the study of extremal phenomena (for a review see Chernozhukov, Fernández-Val, and Kaji, 2016). In financial economics, a prominent example of the use of CEQ is the conditional value-at-risk analysis (Chernozhukov and Umantsev, 2001; Chernozhukov and Du, 2006; Chernozhukov and Fernández-Val, 2011).

We now provide the conditional extremal quantile regression crash measure ($CEQ-CRASH$). The Appendix reviews some basics for CEQ and briefly presents the estimation and inference results of Chernozhukov and Fernández-Val (2011), which we employ and rely upon in this paper.

Let $Y$ be a real random response variable of interest with a continuous distribution function $F_Y(y) = Pr[Y \leq y]$. A $\tau$-quantile of $Y$ is $Q_Y(\tau) = \inf\{y : F_Y(y) > \tau\}$ for some $\tau \in (0,1)$. Let $X$ be a $d-$dimensional vector of covariates related to $Y$ (typically transformations of original regressors, including a constant), and $F_Y(y|x) = Pr[Y \leq y|X = x]$ denote the conditional distribution function of $Y$ given $X = x$. The conditional $\tau$-quantile of $Y$ given $X = x$ is then $Q_Y(\tau|x) = \inf\{y : F_Y(y|x) > \tau\}$ for some $\tau \in (0,1)$.

The conditional quantile function $Q_Y(\tau|x)$ is called the $\tau$-quantile regression function, and can be used to measure the effect of covariates on outcomes, both at the center and at the upper and lower tails of the outcome distribution. Whenever the probability index $\tau$ is either close to zero or close to one, then it is called the conditional extremal $\tau$-quantile.

The most common model for $Q_Y(\tau|x)$ is the linear model:

$$Q_Y(\tau \mid x) = x'\beta(\tau) \text{ for all } \tau \in (0,\eta] \text{ and some } \eta \in (0,1], \tag{1}$$

and for every $x$ in the support of $X$.

Let $(Y, X)$ be a sample of size $T$. The estimator of the conditional quantile function is:

$$\hat{Q}_Y(\tau|x) = x'\hat{\beta}(\tau), \tag{2}$$

where the $\tau$-quantile estimator of $\beta(\tau)$ solves:

$$\hat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}} \sum_{t=1}^{T} \rho_\tau \left(Y_t - X_t'\beta\right), \tag{3}$$

where $\rho_\tau(u) = (\tau - 1(u < 0))\, u$ is the asymmetric absolute deviation function of Fox and Rubin (1964). The median case was introduced by De Laplace (1818) and the general quantile formulation by Koenker and Bassett Jr (1978).

The sampling conditions (C3) and (C4) of Chernozhukov and Fernández-Val (2011) hold (see the Appendix). Thus, the sequence $\{(Y_t, X_t)\}_{t=1}^{T}$ is assumed to be either independent and identically distributed (i.i.d.), or stationary and weakly-dependent with extreme events satisfying a non-clustering condition.[2]

Moreover, assuming that the regularity condition (C1) of Chernozhukov and Fernández-Val (2011) holds, covariates may impact the extremal quantiles and the central quantiles very differently as well as covariates may have a differential impact across various extremal quantiles. The stronger condition (C2) of Chernozhukov and Fernández-Val (2011) imposes the existence and Pareto-type behavior of the conditional quantile density function enabling inference.

We examine the extremal quantiles of the cross section of "excess firm specific" returns distribution within each year. Our notion of excess returns is embodied in *Eq.* (6) below. It may be interpreted as a "partialing out", first step, that is a projection operator. It purges the expected (contemporaneous and recent) market and industry return effects which

[2] The non-clustering condition is of the Meyer (1973) type and states that the probability of two extreme events co-occurring at nearby dates is much lower than the probability of just one extreme event. This assumption is convenient because it leads to limit distributions of extremal quantile regression estimators as if independent sampling had taken place. The plausibility of the non-clustering assumption is an empirical matter.

embody common market and dynamic effects. The next step is a quantile regression of the latter idiosyncratic residuals on further firm-specific components and controls. In other words, the conditional extremal $\tau$-quantile regression estimates from *Eqs.* (2) and (3), with $\tau$ close to zero, are of interest. A stock price crash is then defined as a return that is an unexpected, extreme event, conditional on stock/firm specific characteristics.

In particular, we follow Kelly and Jiang (2014) who argue that a single process governs tail risk fluctuations for all assets and pool the data per Fama and French (1997) 12 industries basis.[3] We then use these pooled idiosyncratic weekly returns to estimate industry-year coefficients using *Eq.* (3) and focusing on the $0.1th$ quantile. This is a linear index model with quantile-specific coefficients, allowing a great deal of heterogeneity. Naturally, firms belonging to different industries will have different coefficients, as does the same firm in different years.

Once we have the estimated coefficients $\hat{\beta}(\tau)$ for each industry and year by solving *Eq.* (3), we may estimate the industry specific threshold $\hat{Q}_Y(\tau|x)$ using *Eq.* (2). As for the vector of covariates this includes the following control variables: stock return volatility, stock returns, stock detrended turnover, firm size, and market-to-book ratio; these are the control variables described in Chen et al. (2001). More explanation on the data and the variables is given in the next section.

The last step includes the identification of crashes. We define $CEQ - CRASH_{j,t}$ as the likelihood of an idiosyncratic, *extreme left-tail event* measured with a binary variable set equal to one if, within fiscal year $t$, the firm $j$ experiences at least one "crash week", *i.e.*, a large and negative idiosyncratic return that falls below the 0.1% industry specific threshold

[3] Due to data limitation issues we cannot perform our analysis on a per-firm basis. However, we performed also the analysis by (a) pooling per-year-industry and (b) pooling data per-year and then take the average over all years. We find that our findings are not sensitive to the way we split the data. All robustness checks are available upon request.

estimate $\hat{Q}_Y(\tau|x)$ from *Eq.* (2), and zero otherwise. Specifically,

$$CEQ - CRASH_{j,t} = \begin{cases} 1 & \text{if } \exists w \in \{1, 2, \ldots, n\} : R_{j,w} < \hat{Q}_Y(\tau|x) \\ 0, & otherwise \end{cases}, \tag{4}$$

where $R_{j,w}$ are the idiosyncratic weekly returns as estimated in *Eq.* (5) below, and $w = 1, 2, \ldots, n$ are the weeks within a fiscal year $t$.

## 2.2 The traditional method

Jin and Myers (2006) define a stock price crash as a large and negative outlier in a firm's residual stock return that occurs, at most, once in 100 periods. An extensive literature in financial economics posits a price crash as the incidence of an *extreme left-tail event* in the distribution of idiosyncratic returns (see, for example, footnote 1). In replicating the traditional way, we follow Hutton et al. (2009) and Andreou et al. (2022) and compute moment-based crashes as $\lambda$-standard-deviations below the mean of idiosyncratic returns.

Let again $w = 1, 2, \ldots, n$ be the weeks within a fiscal year $t$. The *idiosyncratic return*, $R_{j,w}$, for firm $j$ in week $w$ is defined as:

$$R_{j,w} = \ln(1 + \epsilon_{j,w}), \tag{5}$$

where $\epsilon_{j,w}$ is a residual return coming from an index model regression as defined by *Eq.* (6). These returns are log-transformed so as to treat for potential positive skewness in raw returns and enable us to symmetrically identify extreme left- *vs.* right-tail events.

Specifically, $\epsilon_{j,w}$ is estimated as the excess return from an expanded market and industry regression, as follows:

$$r_{j,w} = \alpha_j + \sum_{i=-2}^{i=+2} \beta_{i,j} r_{MKT,w+i} + \sum_{i=-2}^{i=+2} \gamma_{i,j} r_{IND,w+i} + \epsilon_{j,w}, \tag{6}$$

where $r_{j,w}$ is firm $j's$ stock return, $r_{MKT,w}$ is the CRSP value-weighted market index return, and $r_{IND,w}$ is the Fama and French (1997) value-weighted 48-industry index return in week $w$. We include up to two lead and lag weekly return terms for the market and industry indices, to control for booms and busts that might happen around the week of interest allowing us to measure the firm's idiosyncratic return with higher precision. To preclude look-ahead bias that accounts for the effect of earnings release when the subsequent crash risk measures are matched with financial statements data, *Eq.* (6) is estimated over the 52-week window ending 13 weeks after the fiscal year-end.[4]

We measure $CRASH_{j,t}$ with a binary variable set equal to one if within fiscal year $t$ the firm $j$ experiences at least one "crash week", *i.e.*, a large negative idiosyncratic return that falls more than $\lambda$ standard deviations below its mean return, and zero otherwise. Specifically,

$$CRASH_{j,t} = \begin{cases} 1 & \text{if } \exists w \in \{1, 2, \ldots, n\} : R_{j,w} < \mu_{j,t} - \lambda * \sigma_{j,t} \\ 0, & otherwise \end{cases}, \tag{7}$$

where $\mu_{j,t}$ and $\sigma_{j,t}$ are, respectively, the mean and standard deviation of the idiosyncratic returns over the weeks that fall within fiscal year $t$.

Following Hutton et al. (2009), $\lambda$ is set equal to 3.09 to generate a frequency of 0.1% extreme left-tail events as per the normal distribution. Although, the $CRASH$ measure appears to align well with the theoretical concept of a crash that is an idiosyncratic, large negative outlier in the distribution of returns, as we show in the empirical application section, the average percentage of crashes has steadily increased and it is much more prevalent than would be expected under a Gaussian (or implicitly Gaussian) assumption for firm-specific

[4] Excess return is typically computed as deviation from a given risk free return. Here, idiosyncratic weekly return is computed as deviation from a statistically determined, stable, weekly market and industry return. An interpretation is that we are removing a linear projection expected value of market and/or industry returns. This is a *partialling out* of returns that accounts for the expected value of market and common industry factors, before a quantile regression is conducted on other conditioning covariates. An alternative approach would be a single step estimation of quantiles, controlling for quantile effects of market and industry weekly returns. Another approach may first estimate the conditional distribution of weekly returns, controlling for all desired covariates simultaneously, by a method such as *distribution regressions.*

returns.

On the other hand, the continuous measure of crash risk introduced by Chen et al. (2001) aims to capture the negative asymmetry of a firm's stock return distribution, indicating stocks that are merely more "crash prone", *i.e.*, subject to a more left-skewed distribution. Specifically, the negative coefficient of skewness ($NCSKEW$) for each firm-year is calculated as the negative value of the ratio of the third moment of idiosyncratic weekly returns to the standard deviation of idiosyncratic weekly returns raised to the third power. Since Chen et al. (2001) seminal study there has been an increasing interest among researchers to understand the factors that drive negative skewness in stock returns induced by stock price crashes.[5]

However, negative asymmetry in returns may arise due to several less extreme negative returns (*i.e.*, negative returns of moderate size), something that does not necessarily comply with the notion that stock price crash risk represents the likelihood of extreme negative idiosyncratic return outliers (Andreou, Andreou, and Lambertides, 2021). In fact, in our data we notice several cases where skewness is negative (indicating that this stock is prone to a crash) without any extreme firm-specific left-tail outcomes. In a similar vein, $NCSKEW$ often turns positive despite occurrence of a crash, noted upon a closer examination. What happens, in practice, is that a firm (within a year) experiences both a "crash" and a "jump" week, which in turn implies that skewness may be positively driven by the "jump" component, despite the occurrence of a large extreme negative return.

All in all, traditional crash risk measures as routinetely employed in prior studies seem to imply a misspecification. In this respect, our main contribution is to develop a tail risk measure as described in subsection 2.1, which we can subsequently employ to (properly) identify stock price crashes.

[5] Examples of such papers include Hutton et al. (2009); Kim et al. (2011); Callen and Fang (2015); Andreou et al. (2016); Kim et al. (2016); Andreou et al. (2017); Chang et al. (2017); Ertugrul et al. (2017); Kim, Wang, and Zhang (2019); Li and Zeng (2019); Andreou et al. (2022).

## *2.3 Related literature*

Studies have shown that stock prices are more susceptible to extreme negative movements than positive ones (French, Schwert, and Stambaugh, 1987; Campbell and Hamao, 1992; Bekaert and Wu, 2000). While extreme positive returns tend to benefit investors, extreme negative returns can be harmful for stock portfolios, especially for under-diversified retail individual investors (Barber and Odean, 2008; Andreou et al., 2021). There is still an ongoing question of how to measure return asymmetry empirically.

Using the third moment of daily idiosyncratic returns may not fully capture asymmetry. As noted by Meijer (2000) it does not consider situations where a return distribution may have zero skewness but further asymmetry, with a negative (or positive) fifth moment. Ghysels, Plazzi, and Valkanov (2016) propose a quantile-based measure of conditional skewness or asymmetry of asset returns. In fact, quantile-based measures of skewness have a long history, beginning with their introduction in the influential text by Bowley (1920). Jiang et al. (2020) propose two asymmetry measures based on the distribution function of the data that as they show offer additional insights beyond what can be deduced from the third central moment. In this respect, although not directly related, we are connected to the literature that seeks to develop asymmetry measures that go beyond skewness.

At the same time, the available evidence regarding the impact of skewness on stock returns is mixed and inconclusive (see Bali, Engle, and Murray, 2016, for an excellent survey). Research, such as Boyer, Mitton, and Vorkink (2010) and Bali and Murray (2013), have found a negative correlation between idiosyncratic skewness and firm characteristics such as size, book to market, and turnover. In contrast, Jiang et al. (2020) asymmetry measures are significantly linked to all firm characteristics, and the correlations are in the same direction as those between firm characteristics and skewness, with the exception of turnover. Kumar (2009) has found a similar correlation between the turnover ratio and the lottery-type stocks, which have much higher turnover ratios than other stocks.

Chernozhukov (2005) examines extreme quantile regression models for the tails of conditional distributions, and analyzes the properties of the Koenker and Bassett Jr (1978) quantile regression estimator. He argues that conventional asymptotic inference, based on the Gaussian limiting distributions, is not valid for extremal quantile regression and provides practical inferential methods when the quantile $\tau \in (0, 1)$ is either very low (close to zero) or high (close to one). Chernozhukov and Fernández-Val (2011) further show that the extreme value distribution provides a better approximation to the distribution of extremal sample quantiles than the normal distribution. Prominent applications of extremal quantile regression are for production frontiers, determinants of low infant birth weights, auction models and conditional value-at-risk (for a review see Chernozhukov et al., 2016).

A number of statistical tests have been developed to detect jumps in asset prices. Boswijk, Laeven, and Yang (2018) have developed and implemented testing procedures, to detect the presence of self-excitation in jumps. Their model shares a common feature with Bollerslev and Todorov (2014) in that the occurrence of price jumps, whether directly or through the volatility channel, exerts a positive feedback on the price jump intensity process. This, in turn, induces a higher conditional probability of future price jumps. Such (local and clustered) increases of tail risk constitute exactly the phenomenon of self-excitation in jumps. On the other hand, Corradi, Silvapulle, and Swanson (2018) provide a model-free jump test for the null hypothesis of zero jump intensity. Their test in contrast to Boswijk et al. (2018) detects jump self-excitation in the data generating process.

While these papers do not consider actual crash incidence, our study focuses on the empirical incidence of crash risk. However, we acknowledge that this literature is suggestive of the presence of temporal variation in the jump tail index. Moreover, Bollerslev and Todorov (2014) have demonstrated that the tail index parameter may be uniquely identified, and in turn estimated, from a cross-section of deep out-of-the-money short-maturity options at a given point in time without making any assumptions about the temporal variation in the overall jump intensity process. They achieve this by "pooling" the option data to recover an average of the tail shape parameters over the relevant time-intervals. Therefore, they rely on the information in the cross-section of options for identifying the temporal variation in the jump tails.

Finally, our study has implications for the asset pricing literature as we are motivated by the need for a reliable measure of idiosyncratic tail risk over time, which is based on certain aspects of asset pricing theories. Specifically, Kelly and Jiang (2014) assume that extreme return events obey a power law, and that there is a different level of firm-specific tail risk across assets, but only a single process governs tail risk fluctuations for all assets. Using empirical analysis, they estimate the common time-varying component of return tails and demonstrate its significant predictive power for the aggregate market returns. Van Oordt and Zhou (2016) use extreme value theory to estimate tail betas to measure the sensitivity of stock return to market downturns and find no evidence of a premium associated with tail betas.

# 3 Empirical application

The data are drawn from the Center for Research in Security Prices (CRSP) for the period 2000 to 2019. We exclude financial service firms (SIC 6000-6999) and utilities (SIC 4900-4999) because the financial characteristics in these industries are not the same as in other industries. Large data sets are crucial to the accuracy of extreme value estimates because only a small fraction of data is informative about the tail distribution.

Unless otherwise specified, residual weekly returns are estimated using the expanded index model in *Eq.* (6). Then, idiosyncratic weekly returns are computed as the natural logarithm of one plus the residual return, as in *Eq.* (5).[6] The final sample with complete information for computing stock price crashes consists of 62,657 firm-year observations.

One of Chen et al. (2001) key forecasting variables is the recent deviation of turnover from its trend. It turns out that firms that experience larger increases in turnover relative to

[6] The quantile regressions are based on the 52 idiosyncratic weekly returns pooled over all stocks within a given industry, in a given year. We have repeated all computations using only the industry returns (no market returns) in the index model; the results are quite robust.

trend experience more negative skewness; *Dturnover* denotes the detrended turnover which is the average monthly share turnover in a given stock, defined as shares traded divided by shares outstanding over the period. Chen et al. (2001) find that when past returns have been high, skewness is forecasted to become more negative, while stocks with low ratios of book value to market value are also forecasted to have more negative skewness. Following their arguments, in the context of a bubble model, high past returns or a low book-to-market value imply that the bubble has been building up for a long time, so that there is a larger drop when it pops and prices fall back to fundamentals. In this respect, we also include *Returns*, which are the cumulative returns on a given stock also measured over the fiscal year $t$, and *Market to Book* calculated as the market to book value of equity at the end of fiscal year $t$. Chen et al. (2001) control for volatility to address the concern that skewness might be correlated with volatility (Campbell and Hamao, 1992). More specifically, our control variables include *Sigma* which is the standard deviation of a given stock's weekly returns, measured over the fiscal year $t$. We also control for *Size* calculated as the logarithm of the firm's stock market capitalization at the end of period $t$.

For the traditional crash risk measure we compute $CRASH$ using *Eq.* (7) as well as $NCSKEW$ for completeness purposes. For CEQ regression analysis we pool the data per Fama and French (1997) 12 industries basis to estimate industry-year coefficients using *Eq.* (3) and using the $0,1th$ quantile estimates of *Eq.* (2), we estimate the $CEQ - CRASH$ indicator as of *Eq.* (4).[7]

## *3.1 Identification of crashes*

Table 1 presents yearly information about the incidence and magnitude of crashes when the traditional binary $CRASH$ risk measure is used using *Eq.* (7).

Interestingly, consistent with Hutton et al. (2009) and Andreou et al. (2022), it seems

[7] We estimate the conditional extreme quantile following in step the code as per the Koenker (2016) quantreg package, as well as the code from Chernozhukov and Du (2008) and Chernozhukov and Fernández-Val (2011).

that crashes are more prevalent than would have been expected under normality of firm-specific returns. In particular, the sample consists of 62,657 firm-year observations, of which 12,456 firm-years or 20.28% are classified as crashes which is far above the 5.07% under the Gaussian assumption. Also, the average weekly return of crashes throughout the period of investigation is substantial and equals −23%. Both the prevalence and the magnitude of the crashes indicate that stock price crashes are events with adverse consequences for the shareholders of a firm.

**[Insert Table 1, here]**

The large disparity between the empirical *vs.* theoretical incidence of crashes, and more importantly the persistent upward trend in the crash occurrences is what Andreou et al. (2022) have coined as the "stock price crash risk puzzle". Our main argument is that the empirical observation of the upsurge in crash frequencies is a consequence of inadequate definitions and measurement of crash risk. Below we show that our estimates could provide partly an answer to this puzzle, although other reasons should also be taken into account for such high crash propensity.

Table 2 presents the same yearly information about the incidence and magnitude of crashes but when the crash risk measure is the $CEQ - CRASH$. The response variable $Y$ are the idiosyncratic weekly returns pooled over all stocks within a given industry, in a given year. The control variables are those described above (Chen et al., 2001) that have been proposed by literature as having a predictive power in explaining the occurrence of a stock price crash. In any case, we have performed various robustness checks on the inclusion of controls and the results remain the same.[8]

Based on the $0.1th$ $CEQ-CRASH$ estimate, 6,447 firm-years or 10.48% are classified as crashes. These findings are also in line with prior evidence showing that simple returns are not normally distributed and exhibit negative skewness (Harvey and Siddique, 2000; Chen et al., 2001). Although the number of firm-years classified as crashes are far less when using

[8] All robustness checks are available upon request.

the $CEQ - CRASH$ measure, the average returns during crashes throughout the period of investigation is again substantial and equals $-24\%$. These two observations suggest that our crash risk definition captures only the substantially negative events for firms' stock prices.

**[Insert Table 2, here]**

Table 3 presents the same yearly information about the incidence and magnitude of crashes as in Table 1 but the threshold used is adjusted to be comparable with the one in Table 2. Hence, in *Eq.* (7) $\mu_R$ and $\sigma_R$ are the mean and standard deviation of the pooled per industry weekly idiosyncratic returns. Again crashes are more prevalent than would have been expected under normality of firm-specific returns. In particular, 18.87% are classified as crashes and the average weekly return of crashes throughout the period of investigation is more substantial and equals $-27\%$.

We first note that under our approach, there is no regular pattern between the percentages of firm-year crashes and market crashes. This suggests that the quantile method takes better account of pertinent information in the tails and lead to more reliable results. Furthermore, the traditional notion of crash creates a known upward bias, *i.e.*, the absolute magnitude of a return needed to qualify as a crash is smaller for larger firms (Hutton et al., 2009).[9] This is an artifact of its functional definition, or approximation that associates crashes with volatility as measured by variance. Our quantile method corrects for this bias. The number of recorded crashes and their upward trend still support the notion of increasing incidence, beyond the theoretical expectations. But the inflated view of this is moderated by our improved concept and definitions.There is a generally objectionable aspect to associating more crashes to more volatile assets prices. While such price movements may visit tail areas, it can be argued that characterization of these events as "crashes" is dubious. A relatively stable asset that improbably crashes to very rare low levels is a more credible view of a "crash".

**[Insert Table 3, here]**

[9] Note that we have checked whether excluding firm-years with a stock price less than $1 at the end of the fiscal year artificially creates the upward-trending frequency of crashes observed and the number of crashes is fairly steady.

### 3.2 Expected shortfall analysis

While both $CEQ-CRASH$ and $CRASH$ tells us about the *frequency* of crashes a serious risk analysis also investigates the *magnitude* of crashes. Classical risk management literature usually proxies the frequency and magnitude of crashes with two main indicators, the Value at Risk (VaR) and the Expected Shortfall (ES) (Jorion, 2007). The difference between these two risk measures lies in how they handle risk assessment. VaR primarily focuses on the frequency of losses, while CVaR complements VaR by incorporating the average magnitude of potential losses. Despite their user-friendly nature, both VaR and CVaR are not exempt from drawbacks. The VaR, due to its lack of coherence, penalizes diversification efforts (Artzner, Delbaen, Eber, and Heath, 1999). On the other hand, CVaR estimates, being non-elicitable, pose challenges for backtesting purposes (Gneiting, 2011). Due to these factors, the Basel Committee on Banking Supervision has published the consultative paper "Fundamental Review of the Trading Book"[10], formally recommending financial institutions to employ CVaR as the measure for market risk, while utilizing VaR for back-testing purposes.

The necessity of conducting a expected shortfall analysis arises from the importance of understanding post-threshold outcomes, as similar frequencies of events may exhibit vastly different magnitudes, particularly in cases of heavy-tailed distributions. To quantify our ES measure, we focus solely on weeks characterized by market crashes, utilizing either the $CRASH$ or the $CEQ-CRASH$ metric. Our approach involves computing a dollar-based ES indicator, which captures the disparity between a firm's market capitalization during the crash event week and its market capitalization in the preceding week. This resulting value represents our approximate estimation of ES, indicating the dollar value lost during the event.

Figure 1 showcases the ES values for both $CEQ-CRASH$ (depicted in blue) and $CRASH$ (depicted in red). The results demonstrate a notable discrepancy in both the frequency and magnitude of crashes between the two indicators. While the $CEQ-CRASH$

[10] https://www.bis.org/publ/bcbs265.pdf

indicator exhibits a lower frequency of crashes, its ES values are considerably larger, particularly in recent years. These findings validate the existence of highly leptokurtic returns, which are more accurately captured by the $CEQ - CRASH$ indicator in terms of both frequency and magnitude.

**[Insert Figure 1, here]**

### *3.3 Conditional extremal quantile estimates*

Figure 2 plots the CEQ estimates along with 90% pointwise CIs. The solid lines represent the extremal CIs and the dashed lines the normal CIs. The extremal CIs are computed by the extremal subsampling method described in Chernozhukov et al. (2016). The normal CIs are based on the normal approximation with the standard errors computed with the method proposed by Powell (1991).

The coefficients of the figures are derived as described in the previous section. However, to enable a simple visualization across time and industries, we take the average of the coefficients as follows: first, we average the coefficients of a specific industry along the time axis; then, we take the average of the Fama and French 12 industries.

Figure 2 shows that *Size* measured using market capitalization turns out positive and strongly significant in the left extreme, while it is negative and significant in the right extreme. *Market to Book* is negative and significant in contrast to Chen et al. (2001) except for the right tail. Interestingly, we find that the detrended level of turnover (*Dturnover*) has a negative significant coefficient at the left tail, then becomes nonsignificant in the middle range, and then significantly positive at the right tail. It turns out *Dturnover* appears to have an impact (negative/positive) on extremal quantiles, whereas it has no effect on the typical quantiles. This is in contrast to the results of Chen et al. (2001), and highlights the importance of our methodology, whilst at the same time lends credence to the contribution of our paper in the crash risk literature. *Sigma* is negative significant and then positive significant at the two extremes, as expected. Finally, *Stock Return* (and its lagged values) do not show any statistical significance, another piece of evidence in disagreement to the results reported in Chen et al. (2001) who find that return terms are always positive and strongly significant.

**[Insert Figure 2, here]**

Figure 3 plots the median bias-corrected QR estimates along with 90% pointwise CIs for the lower tail. The bias correction is also implemented using extremal subsampling with the same specifications (Chernozhukov and Fernández-Val, 2011; Chernozhukov et al., 2016). Due to the median bias-correction, the coefficient estimates are slightly different from Figure 2. As before, lagged returns have no impact on extreme risk. *Market to Book* has a significant negative impact on left extremal quantiles. More surprisingly, the detrended level of turnover has negative and highly significant coefficients.

Comparing the CIs produced by the extremal inference and the normal inferences in Figure 2 show that they closely match in the central region, while it reveals that the normal CIs are often narrower than the extremal CIs in the tails. This discrepancy reveals that when the situation is extremal the normal CIs on the tails substantially underestimates the sampling variation and hence it might lead to a substantial undercoverage in the CIs. Figure 2 reveals that differences between central and extremal inference occur only sufficiently far in the tails. In fact, as shown in Figure 3 normal CIs are indeed much more narrow than extremal CIs at $\tau < 0.10$ for stock returns, past stock returns and market to book. The discrepancies between extremal CIs and central CIs for the coefficient on *Dturnover* arise mostly when $\tau < 0.05$. This validates the use of extremal quantiles.

**[Insert Figure 3, here]**

# 4 The effect of opacity and overinvestment on future stock price crashes

In the absence of appropriate monitoring, CEOs might undertake actions that maximize their own wealth to the detriment of shareholder welfare (*e.g.*, Jin and Myers, 2006; Hutton et al., 2009; Kothari, Shu, and Wysocki, 2009; Bhagat and Bolton, 2013; Callen and Fang, 2015; Andreou et al., 2016). In this vein, the breadth of the crash risk literature holds the view that the underlying reason that triggers stock price crashes is the hoarding of bad news mechanism, which is fueled by self-interested executives who strategically camouflage bad news via the financial reporting opacity and overinvestment channels. More specifically, the renowned agency model of Jin and Myers (2006) argues that information asymmetry between managers and shareholders, combined with investors' incompletely secured property rights that result in lack of transparency, enables managers to accumulate bad news. Accordingly, the lack of transparency (*i.e.*, opacity) increases, along with the amount of concealed negative information in accordance. However, the managers' ability to conceal bad news is not unlimited. Therefore, when the hoarded bad news crosses a tipping point, negative information comes out all at once. As a result, the accumulation of bad news leads to stock price crashes that suddenly spur in the market.

In the same agency spirit, the theory of Benmelech et al. (2010) draws motivation from the argument that CEOs, aiming to protect and/or increase the component of firm performance which directly affects their financial rewards, exploit information asymmetries to manifest management's self-interested behavior and persistently hide bad news by engaging in overinvestment. Particularly, when the growth rate of investment opportunities starts to decline, concerns about their personal wealth can incentivize CEOs to conceal adverse outcomes from shareholders. As a result, CEOs do not reveal the bad news to the investors in a timely fashion to retain their expectations and accordingly the level of stock price. According to this paradigm, CEOs are engaged in value-destroying investment decisions, at least temporarily, until the revelation of the real growth rate of the firm's investment opportunities, which triggers a stock price crash.

The theoretical explanations derived from the agency theory arguments, accentuate financial reporting opacity and overinvestment as the channels underpinning the relation between the hoarding of bad news and stock price crashes. These two channels enable managers to persistently camouflage bad news that has an adverse effect on their firm's economic fundamentals, in the hope that subsequent events will turn in their favor to avoid experiencing the underlying negative consequences upon its public revelation to investors.

Following the above theoretical motivations, a large strand of studies venture to explain stock price crashes by identifying determinants consistent with the agency viewpoint. Assessing the literature, Andreou et al. (2022) conducted a qualitative meta-analysis of 94 papers published since 2009 in prestigious finance, business and accounting journals and show the over-reliance of the extant literature on explanations rooted in the opacity and overinvestment channels as described by the agency theory. Admittedly the empirical research has been very prolific in using opacity and overinvestment as the channels through of which managers strategically exploit to fuel the bad news mechanism. Notwithstanding, Andreou et al. (2022) argue about the inefficacy of opacity and overinvestment to rationalize the stock price crash risk phenomenon. Their investigation supports that both channels have attenuated in the past two decades for the average US-listed firm, whereas, in stark contrast, the frequency of stock price crashes has notably surged (as also shown in Table 3).

There are valid reasons to justify the attenuated role of the agency paradigm in explaining stock price crashes in the US markets. For instance, investors have witnessed at the dawn of the new millennium an extensive list of corporations that collapsed due to managerial misconduct, which resulted in policymakers increasing regulation on financial reporting and other business practices at publicly traded companies. Ergo, in the past two decades there has been an upsurge of corporate governance regulation, laws and exchange listing standards such as the Regulation Fair Disclosure in 2000, the Sarbanes-Oxley Act of 2002, the Dodd-Frank Act of 2010, and the Corporate Governance Reform and Transparency Act of 2017 (Bhagat and Bolton, 2013; DeFond and Zhang, 2014). Accordingly, recent studies suggest that not only accrual-based earnings management experienced a significant decline after the passage of Sarbanes-Oxley Act (Cohen, Dey, and Lys, 2008), but that the new regulatory regime resulted in an improved corporate-governance system. It is also a fact that in the recent years public company boards have been facing an increasing demand for corporate governance effectiveness and quality. Investors—especially institutional investors and activists—continually exert pressure for constant development on governance trends.

Collectively, it is then natural to assume that the adoption of important regulation and laws in the last two decades have significantly contributed in the strengthening of corporate governance functions aiming to combat managerial opportunism and helped to protect shareholders' welfare. The overall improvement in firms' corporate governance should have limited the leeway for managers to persistently conceal negative information regarding their firms' economic fundamentals to benefit themselves at the expense of shareholders through opacity and overinvestment. These developments could possibly explain to a great extent the evidence in Andreou et al. (2022) who report nonsignificant results for the crash-opacity and crash-overinvestment relations, especially in the post-Sarbanes-Oxley Act period. However, *absence of evidence is not evidence of absence*, thus we cannot rule out that more accurate measures of stock price crashes might reveal different results.

We perform logit regression analyses to examine the relation between opacity, overinvestment and crash risk. The regression specifications are as follows:

$$Pr(Y_{j,t+1} = 1|x) = F(\alpha + \beta_1 Opacity_{j,t} + \beta_2 Overinvestment_{j,t} + \gamma \boldsymbol{X}_j + FE_{year} + FE_{ind}) \quad (8)$$

where $Y_{j,t+1}$ is either $CRASH$ or $CEQ - CRASH$ measured in fiscal year $t+1$, $F$ is the logistic transformation, and all explanatory variables are measured in year $t$.[11] *Opacity*

[11] For the needs of this analysis, following Andreou et al. (2022) we impose additional filtering rules, particularly, keeping common stocks (*i.e.,* share codes 10 and 11) traded in NYSE, AMEX and NASDAQ, excluding firm-years with a stock price less than \$1 at the end of the fiscal year and having fewer than 26

is calculated as the three-year moving sum of the absolute value of annual discretionary accruals, whereby discretionary accruals are estimated based on the modified Jones model (Dechow, Sloan, and Sweeney, 1995). Higher values of *Opacity* associate with firms that are more likely to be managing reported earnings to camouflage bad news. *Overinvestment* is calculated as the three-year moving sum of abnormal component of investment, whereby the expected levels of investment is estimated following the methodology in Richardson (2006). The vector $\boldsymbol{X}$ includes the following control variables: $log(Total\ assets)$ measuring the natural logarithm of total assets; $Firm\ age$ calculated as the natural logarithm of the number of years that the firm is covered in the Compustat universe; $Market\ to\ Book$ calculated as the market to book value of equity; $Zscore$ calculated as the fitted value using the updated coefficients of the model proposed by Altman; $Return\ on\ equity$ calculated as the ratio of income before extraordinary items to book value of equity; and $Ncskew$ calculated as the negative of the third moment of idiosyncratic weekly returns within the fiscal year, divided by the associated standard deviation of firm-specific weekly returns raised to the third power−idiosyncratic weekly returns are estimated using *Eqs* (5) and (6). Finally, $FE_{year}$ feature year-fixed effects and $FE_{ind}$ industry-fixed effects where we use the Fama and French 12 industry classifications.

Table 4 shows summary statistics of the variables included in the logistic regression analysis. The mean value of $CRASH$ is 0.211, suggesting that about 21% of these firm-year observations experience at least one crash event. Similarly, the mean value of $CEQ-CRASH$ is 0.094 suggesting that 9.4% of these firm-year observations experience at least one crash event under our modified crash measure. With respect to the two channels, the mean value (standard deviation) of *Opacity* is 0.231 (0.229) and the mean value (standard deviation) of *Overinvestment* is 0.020 (0.233). All variables have distribution characteristics very similar to those reported in Andreou et al. (2022) and previous literature.

**[Insert Table 4, here]**

weeks of stock returns in a fiscal year, and dropping firm-year observations without available information in Compustat for computing the financial variables.

Table 5 reports the logistic regression results as per *Eq.* (8). The results in models (1)-(3) where the $CRASH$ measure is employed are in accordance to the evidence reported in Andreou et al. (2022), whereby the coefficients of *Opacity* and *Overinvestment* are non-significant. Interestingly, though, in models (4)-(6) where the $CEQ - CRASH$ measure is employed both *Opacity* and *Overinvestment* turn out positive and statistically significant.The evidence in models (4)-(6) is not conclusive that $CEQ - CRASH$ is the best approach, but it is an indication that the traditional method of operationalizing $CRASH$ is possibly generating many "false-positive" crashes which are disconnected from finance theories.

**[Insert Table 5, here]**

# 5 Conclusions

A variable coefficient conditional quantile model of tail risk demonstrates the value of more precise notions and counting of crashes, in terms of both frequency and magnitude. We find that this approach makes a difference, showing a smaller frequency of firm-specific crashes, and consequential magnitudes. Equipped with extremal inference, we venture far into the tails and study extremely low return quantiles. Some of our findings differ sharply from previous results for typical non-extremal quantiles. The traditional moment-based results and puzzles are shown to be artifacts of the approximations that presume (often implicitly) Gaussian distributions for stock returns. This is empirically unsupported. Our findings are robust to a set of alternative data analyses and modeling decisions, some dictated by data limitations. The findings also shed light on the actual distribution of returns based on quantile estimation and rigorous inferential techniques. Our paper is the first, in the cross section setting, to examine the benefits of a non-Gaussian quantile approach, reaching findings that also apply to all other moment based approximations, with or without Gaussianity.

The jump literature and similar, try to model existence of jumps, while other studied may try to explain why these jumps occur. Much of this literature is focused on time series aspects of returns. In contrast, we are characterizing the cross section elements of idiosyncratic returns, after removing of market and industry effects. This is more in the nature of a counting crashes, *i.e.*,the frequency of crashes that are firm specific. Since we are not trying to explain the time path of returns, idiosyncratic or otherwise, time series modeling issues are not central. Dynamic short-term effects are known to be separated−albeit related to−from long run, perhaps firm-specific managerial effects, which cross section studies identify. Ours is a cross sectional analysis, checked over many points in time. Of course, we acknowledge the importance of time series properties of variables.

# References

Andreou, C. K., Andreou, P. C., Lambertides, N., 2021. Financial distress risk and stock price crashes. *Journal of Corporate Finance* 67, 101870.

Andreou, P. C., Antoniou, C., Horton, J., Louca, C., 2016. Corporate governance and firm-specific stock price crashes. *European Financial Management* 22, 916–956.

Andreou, P. C., Lambertides, N., Magidou, M., 2022. A critique of the agency theory viewpoint of stock price crash risk: The opacity and overinvestment channels. *British Journal of Management* .

Andreou, P. C., Louca, C., Petrou, A. P., 2017. CEO age and stock price crash risk. *Review of Finance* 21, 1287–1325.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Mathematical finance* 9, 203–228.

Bali, T. G., Engle, R. F., Murray, S., 2016. Empirical asset pricing: The cross section of stock returns. John Wiley & Sons.

Bali, T. G., Murray, S., 2013. Does risk-neutral skewness predict the cross-section of equity option portfolio returns? *Journal of Financial and Quantitative Analysis* 48, 1145–1171.

Barber, B. M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785–818.

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

Benmelech, E., Kandel, E., Veronesi, P., 2010. Stock-based compensation and CEO (dis) incentives. *The Quarterly Journal of Economics* 125, 1769–1820.

Bhagat, S., Bolton, B., 2013. Director ownership, governance, and performance. *Journal of Financial and Quantitative Analysis* 48, 105–135.

Bollerslev, T., Todorov, V., 2014. Time-varying jump tails. *Journal of Econometrics* 183, 168–180.

Boswijk, H. P., Laeven, R. J., Yang, X., 2018. Testing for self-excitation in jumps. *Journal of Econometrics* 203, 256–266.

Bowley, A., 1920. Elements of Statistics, 4th Edn (New York, Charles Scribner). *Bowley4Elements of Statistics1920* .

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

Callen, J. L., Fang, X., 2015. Religion and stock price crash risk. *Journal of Financial and Quantitative Analysis* 50, 169–195.

Campbell, J. Y., Hamao, Y., 1992. Predictable stock returns in the United States and Japan: A study of long-term capital market integration. *The Journal of Finance* 47, 43–69.

Chang, C.-L., Jiménez-Martín, J.-Á., Maasoumi, E., McAleer, M., Pérez-Amaral, T., 2019. Choosing expected shortfall over VaR in Basel III using stochastic dominance. *International Review of Economics & Finance* 60, 95–113.

Chang, X., Chen, Y., Zolotoy, L., 2017. Stock liquidity and stock price crash risk. *Journal of Financial and Quantitative Analysis* 52, 1605–1637.

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

Cheng, C. A., Li, S., Zhang, E. X., 2020. Operating cash flow opacity and stock price crash risk. *Journal of Accounting and Public Policy* 39, 106717.

Chernozhukov, V., 2005. Extremal quantile regression. *The Annals of Statistics* 33, 806–839.

Chernozhukov, V., Du, S., 2006. Extremal quantiles and value-at-risk. MIT Department of Economics Working Paper.

Chernozhukov, V., Du, S., 2008. pExtremal Quantiles and Value'at'Risk, q The New Palgrave Dictionary of Economics. *271r292* .

Chernozhukov, V., Fernández-Val, I., 2011. Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies* 78, 559–589.

Chernozhukov, V., Fernández-Val, I., Kaji, T., 2016. Extremal quantile regression: An overview. *arXiv Preprint arXiv:1612.06850* .

Chernozhukov, V., Umantsev, L., 2001. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* 26, 271–292.

Cohen, D. A., Dey, A., Lys, T. Z., 2008. Real and accrual-based earnings management in the pre-and post-Sarbanes-Oxley periods. *The accounting review* 83, 757–787.

Corradi, V., Silvapulle, M. J., Swanson, N. R., 2018. Testing for jumps and jump intensity path dependence. *Journal of Econometrics* 204, 248–267.

De Laplace, P. S., 1818. Théorie analytique des probabilités: Supplément a la théorie analytique des probabilités: Février 1818, vol. 2. Courcier.

Dechow, P. M., Sloan, R. G., Sweeney, A. P., 1995. Detecting earnings management. *Accounting Review* 70, 193–225.

DeFond, M., Zhang, J., 2014. A review of archival auditing research. *Journal of Accounting and Economics* 58, 275–326.

deHaan, L. F. M., 1970. On regular variation and its application to the weak convergence of sample extremes.

Ertugrul, M., Lei, J., Qiu, J., Wan, C., 2017. Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis* 52, 811–836.

Fama, E. F., French, K. R., 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.

Fox, M., Rubin, H., 1964. Admissibility of quantile estimates of a single location parameter. *The Annals of Mathematical Statistics* pp. 1019–1030.

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

Ghysels, E., Plazzi, A., Valkanov, R., 2016. Why invest in emerging markets? The role of conditional return asymmetry. *The Journal of Finance* 71, 2145–2192.

Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics* pp. 423–453.

Gneiting, T., 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106, 746–762.

Harvey, C. R., Siddique, A., 2000. Conditional skewness in asset pricing tests. *The Journal of Finance* 55, 1263–1295.

Hoti, S., Maasoumi, E., McAleer, M., Slottje, D., 2009. Measuring the volatility in US treasury benchmarks and debt instruments. *Econometric Reviews* 28, 522–554.

Hutton, A. P., Marcus, A. J., Tehranian, H., 2009. Opaque financial reports, R2, and crash risk. *Journal of Financial Economics* 94, 67–86.

Jiang, L., Wu, K., Zhou, G., Zhu, Y., 2020. Stock return asymmetry: Beyond skewness. *Journal of Financial and Quantitative Analysis* 55, 357–386.

Jin, L., Myers, S. C., 2006. R2 around the world: New theory and new tests. *Journal of Financial Economics* 79, 257–292.

Jorion, P., 2007. Value at risk: the new benchmark for managing financial risk. The McGraw-Hill Companies, Inc.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. *The Review of Financial Studies* 27, 2841–2871.

Kim, C., Wang, K., Zhang, L., 2019. Readability of 10-K reports and stock price crash risk. *Contemporary Accounting Research* 36, 1184–1216.

Kim, J.-B., Li, Y., Zhang, L., 2011. Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics* 100, 639–662.

Kim, J.-B., Wang, Z., Zhang, L., 2016. CEO overconfidence and stock price crash risk. *Contemporary Accounting Research* 33, 1720–1749.

Koenker, R., 2016. quantreg: Quantile Regression. R package version 5.21.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society* pp. 33–50.

Kothari, S. P., Shu, S., Wysocki, P. D., 2009. Do managers withhold bad news? *Journal of Accounting Research* 47(1), 241–276.

Kumar, A., 2009. Who gambles in the stock market? *The Journal of Finance* 64, 1889–1933.

Li, Y., Zeng, Y., 2019. The impact of top executive gender on asset prices: Evidence from stock price crash risk. *Journal of Corporate Finance* 58, 528–550.

Maasoumi, E., McAleer, M., 2006. Multivariate stochastic volatility: An overview. *Econometric Reviews* 25, 139–144.

Meijer, E., 2000. An asymmetric distribution with zero skewness. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.2531847* .

Meyer, R., 1973. A Poisson-type limit theorem for mixing sequences of dependent'rare'events. *The Annals of Probability* 1, 480–483.

Powell, J., 1991. oEstimation of Monotonic Regression Models under Quantile Restrictions, p in Nonparametric and Semiparametric Methods in Econometrics, ed. by W. Barnett, J. Powell, and G. Tauchen.

Richardson, S., 2006. Over-investment of free cash flow. *Review of Accounting Studies* 11(2), 159–189.

Van Oordt, M. R., Zhou, C., 2016. Systematic tail risk. *Journal of Financial and Quantitative Analysis* 51, 685–705.

# Appendix

## A. Basics and notation

We review some basics on extremal conditional quantile regression using the notation of Chernozhukov and Fernández-Val (2011).

Let a real random response variable of interest $Y$ with a continuous distribution function $F_Y(y) = Pr[Y \leq y]$. A $\tau$-quantile of $Y$ is $Q_Y(\tau) = \inf\{y : F_Y(y) > \tau\}$ for some $\tau \in (0,1)$. Let $X$ be a $d-$dimensional vector of covariates related to $Y$ (typically transformations of original regressors, including a constant), and $F_Y(y|x) = Pr[Y \leq y|X = x]$ denote the conditional distribution function of $Y$ given $X = x$. The conditional $\tau$-quantile of $Y$ given $X = x$ is then $Q_Y(\tau|x) = \inf\{y : F_Y(y|x) > \tau\}$ for some $\tau \in (0,1)$.

The conditional quantile function $Q_Y(\tau|x)$ is called the $\tau$-quantile regression function and can be used to measure the effect of covariates on outcomes, both at the center and at the upper and lower tails of the outcome distribution. Whenever the probability index $\tau$ is either close to zero or close to one then it is called the conditional extremal $\tau$-quantile.

The theory for extremal quantile regression assumes that the tails of the conditional distribution of the outcome variable have Pareto-type behavior, i.e. (lower) tails decay approximately as a power function (Gnedenko, 1943; deHaan, 1970).

Define a random variable $U$ as $U \equiv Y$, if the lower end-point of the support of $Y$ is $-\infty$, and $U \equiv Y - Q_Y(0)$, if the lower end-point of the support of $Y$ is $Q_Y(0) > -\infty$. The quantile function of $U$, denoted by $Q_U$ then has lower end-point $Q_U(0) = -\infty$ or $Q_U(0) = 0$.

Then, the assumption that the distribution function $F_U$ and its quantile function $Q_U$ exhibit Pareto-type behavior in the tails can be formally stated as the following two equivalent conditions:

$$F_U(u) \sim \bar{L}(u) \cdot u^{-1/\xi} \, as \, u \searrow Q_U(0), \tag{9}$$

$$Q_U(\tau) \sim L(\tau) \cdot \tau^{-\xi} \, as \, \tau \searrow 0, \tag{10}$$

for some real number $\xi \neq 0$, where $\bar{L}(u)$ is a nonparametric, slowly-varying function at $Q_U(0)$, and $L(\tau)$ is a nonparametric slowly-varying function at 0. The prime examples of slowly-functions are the constant function $L(y) = L$ and the logarithmic function. The number $\xi$ is the so-called extreme value index. The extreme value index is very important since this generally controls the tail behavior of the distribution function. If $\xi < 0$, the right-tail is short, i.e. the right endpoint is finite. This class is called the Weibull class and contains among others the uniform and reverse Burr CDFs. If $\xi > 0$, the right-tail is heavy. Examples in this class (Frechet class) are the Pareto, Burr, Student's and other. If $\xi = 0$, the right-tail is of an exponential type, and the right endpoint can then be either finite or infinite. This class (Gumbel class) encompasses the exponential, normal, log-normal, gamma and classical Weibull CDFs.

## B. Inference

Let the conditional quantile function of $Y$ given $X = x$ given by the linear model:

$$Q_Y(\tau \mid x) = x'\beta(\tau), \; for \, all \, \tau \in (0, \eta], and \, some \, \eta \in (0, 1], \tag{11}$$

and for every $x$ in the support of $X$. This linear model is computationally convenient and it has good approximation properties with varying quantile effects.

Let $(Y, X)$ a sample of size $T$. Then, the $\tau$-quantile estimator of $Q_Y(\tau|x)$ and $\beta(\tau)$ solves:

$$\hat{Q}_Y(\tau|x) = x'\hat{\beta}(\tau),$$

$$\hat{\beta}(\tau) \in \arg\min_{\beta \in \mathbb{R}} \sum_{t=1}^{T} \rho_\tau (Y_t - X_t'\beta),$$

where $\rho_\tau(u) = (\tau - 1(u < 0))\, u$ is the asymmetric absolute deviation function of Fox and Rubin (1964). The median case was introduced by De Laplace (1818) and the general quantile formulation by Koenker and Bassett Jr (1978).

The analysis of the properties of the estimators of extremal quantiles relies on extreme value theory which uses sequences of quantile indexes $\{\tau_T\}_{T=1}^{\infty}$ that change with the sample size $T$. A sequence of quantile index and sample size pairs $\{\tau_T, T\}_{T=1}^{\infty}$ is said to be an extreme order sequence if $\tau_T \searrow 0$ and $\tau_T T \to k \in (0, \infty)$ as $T \to \infty$; an intermediate order sequence if $\tau_T \searrow 0$ and $\tau_T T \to \infty$ as $T \to \infty$; and a central order sequence if $\tau_T$ is fixed as $T \to \infty$. The extreme order sequence leads to an extreme value law in large samples, whereas the intermediate and central sequences lead to normal laws. The extreme value law provides a better approximation to the extremal quantile regression estimators.

Given the basics of the previous section, the main assumption for inference is that the response variable $Y$ is conditioned on a varying parameter linear index $X\beta_e$, with Pareto-type tails as formally stated in conditions (C1) and (C2) of Chernozhukov and Fernández-Val (2011). This allows covariates to impact the extremal quantiles and the central quantiles very differently as well as it allows for a differential impact of covariates across various extremal quantiles. This implies conditional heteroskedasticity that is common in economic applications.

Moreover, the sampling conditions of Chernozhukov and Fernández-Val (2011) hold. Specifically, data are either i.i.d. or stationary and weakly dependent with extreme events satisfying a non-clustering condition. In particular, the sequence $\{(Y_t, X_t)\}_{t=1}^{T}$ is assumed to form a stationary, strongly mixing process with geometric mixing rate that satisfies the condition that curbs clustering of extreme events.[12]. This assumption leads to limit distributions as if independent sampling had taken place. The assumption (C3) of Chernozhukov and Fernández-Val (2011) about compactness of $X$, non-generacy and non-lattice also holds.

[12] The non-clustering condition is of the Meyer (1973) type and states that the probability of two extreme events co-occurring at nearby dates is much lower than the probability of just one extreme event. For example, it assumes that a large market crash is not likely to be immediately followed by another large crash.

Given the above assumptions the canonically -normalized QR statistic is given by:

$$\hat{Z}_T(k_T) := A_{\mathrm{T}}\left(\widehat{\beta}(\tau_T) - \beta(\tau_T)\right) for\ A_{\mathrm{T}} := 1/Q_U(1/T).$$

This is generally infeasible for inference because it depends on the unknown canonical normalization constant $A_T$. This constant can only be estimated consistently under strong parametric assumptions and an additional estimation procedure.

The self-normalized QR statistic is given by:

$$Z_T(k_T) := \mathcal{A}_{\mathrm{T}}\left(\widehat{\beta}(\tau_T) - \beta(\tau_T)\right) for\ \mathcal{A}_{\mathrm{T}} := \frac{\sqrt{k_T}}{\left(\bar{X}_T'\widehat{\beta}(m\tau_T) - \widehat{\beta}(\tau_T)\right)},$$

where $\bar{X}_T = T^{-1}\sum_{t=1}^{T} X_t$ and $m$ is a real number such that $k(m-1) > d_x$ for any integer $k \geq 1$ and $k_T = \tau_T T \to k$, as $T \to \infty$. This is always feasible because it uses a normalization that only depends on the data.

Chernozhukov and Fernández-Val (2011) show that for $k_T \to k > 0$, as $T \to \infty$.

$$\hat{Z}_T(k_T) \to_d \hat{Z}_\infty(k)$$

where for $\chi = 1$ if $\xi < 0$ and $\chi = -1$ if $\xi > 0$,

$$\hat{Z}_\infty(k) := \chi \cdot \arg\min_{z\in\mathbb{R}}\left[-kE[X]'z + \sum_{t=1}^{\infty}\left\{\mathcal{X}_t'z - \chi\left(\Gamma_t^{-\xi} - k^{-\xi}\right)\mathcal{X}_t'\gamma\right\}_+\right],$$

where $\{\mathcal{X}_t, \mathcal{X}_t, ...\}$ is an i.i.d. sequence with distribution $F_X$;$\{\mathit{\Gamma}_1, \mathit{\Gamma}_2, ...\} := \{\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, ...\}$; $(\mathcal{E}_1, \mathcal{E}_2, ...)$ is an i.i.d sequence of standard exponential variables that is independent of $\{\mathcal{X}_t, \mathcal{X}_t, ...\}$ ;and $\{y\}_+ := max\{0, y\}$. Furthermore,

$$Z_T(k_T) \to_d Z_\infty(k) := \frac{\sqrt{k}\hat{Z}_\infty(k)}{E[X]'\left(\hat{Z}_\infty(mk) - \hat{Z}_\infty(k)\right) + \chi \cdot (m^{-\xi} - 1)k^{-\xi}}$$

Consider general linear functions of the coefficient vector $\hat{\beta}(\tau)$, $\psi'\hat{\beta}(\tau)$; for some nonzero vector $\psi \in \mathbb{R}^{d_x}$, based on the extreme value approximations of $\psi'\hat{Z}_{\infty}(k)$ and $\psi' Z_{\infty}(k)$. Chernozhukov and Fernández-Val (2011) construct asymptotically median-unbiased estimators and a $(1-\alpha)\%$ confidence interval for $\psi'\beta(\tau)$. Computation of the critical values of the limit extreme value distributions can be achieved through an analytical computation which requires a consistent estimation of $\xi$ based on a regression analogue of the Pickands or Hill estimator and of the scale parameter. In practice, it is convenient to compute the quantiles of the extreme value distributions using simulation or resampling methods, instead of an analytical method. Subsampling has the advantage that it does not require the estimation of $\xi$, and is consistent under general assumptions of Chernozhukov and Fernández-Val (2011). The extremal bootstrap although again needs to consistently estimate $\xi$ and $\gamma$, it is computationally less demanding than the analytical inference procedure (Chernozhukov et al., 2016).

# Figures

**Figure 1.** Expected shortfall

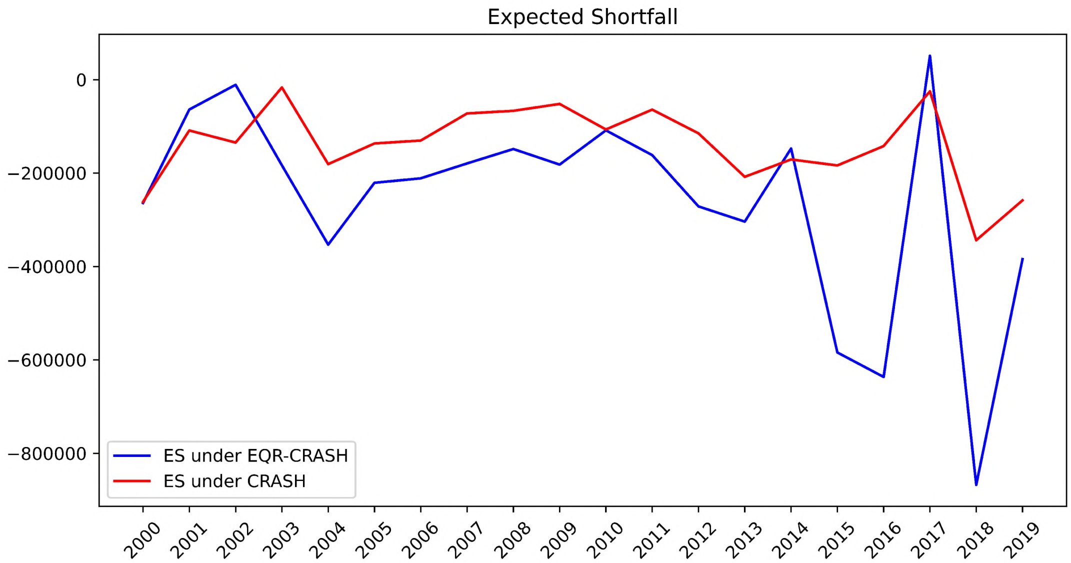


**Figure 2.** Conditional extremal quantile estimates.

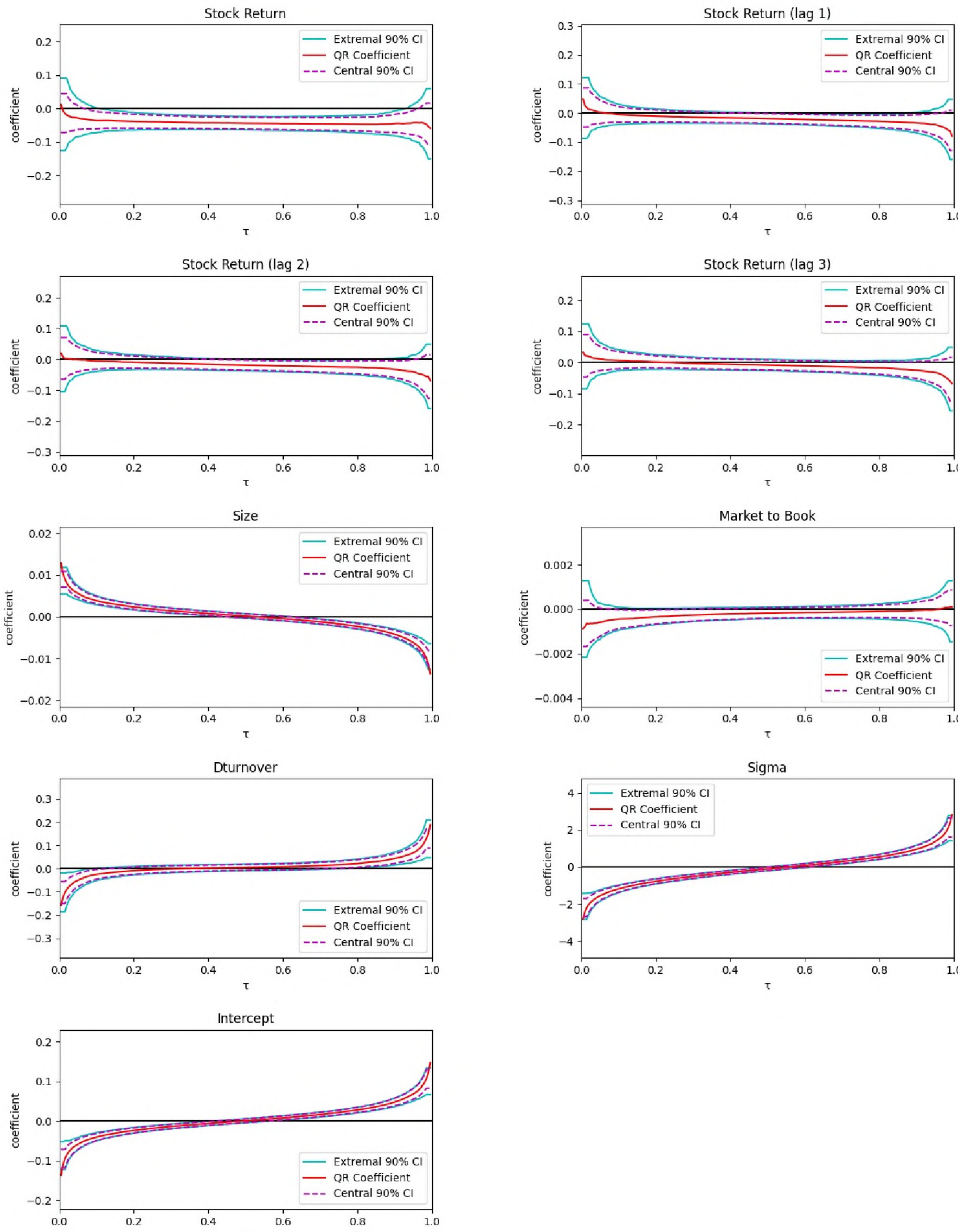
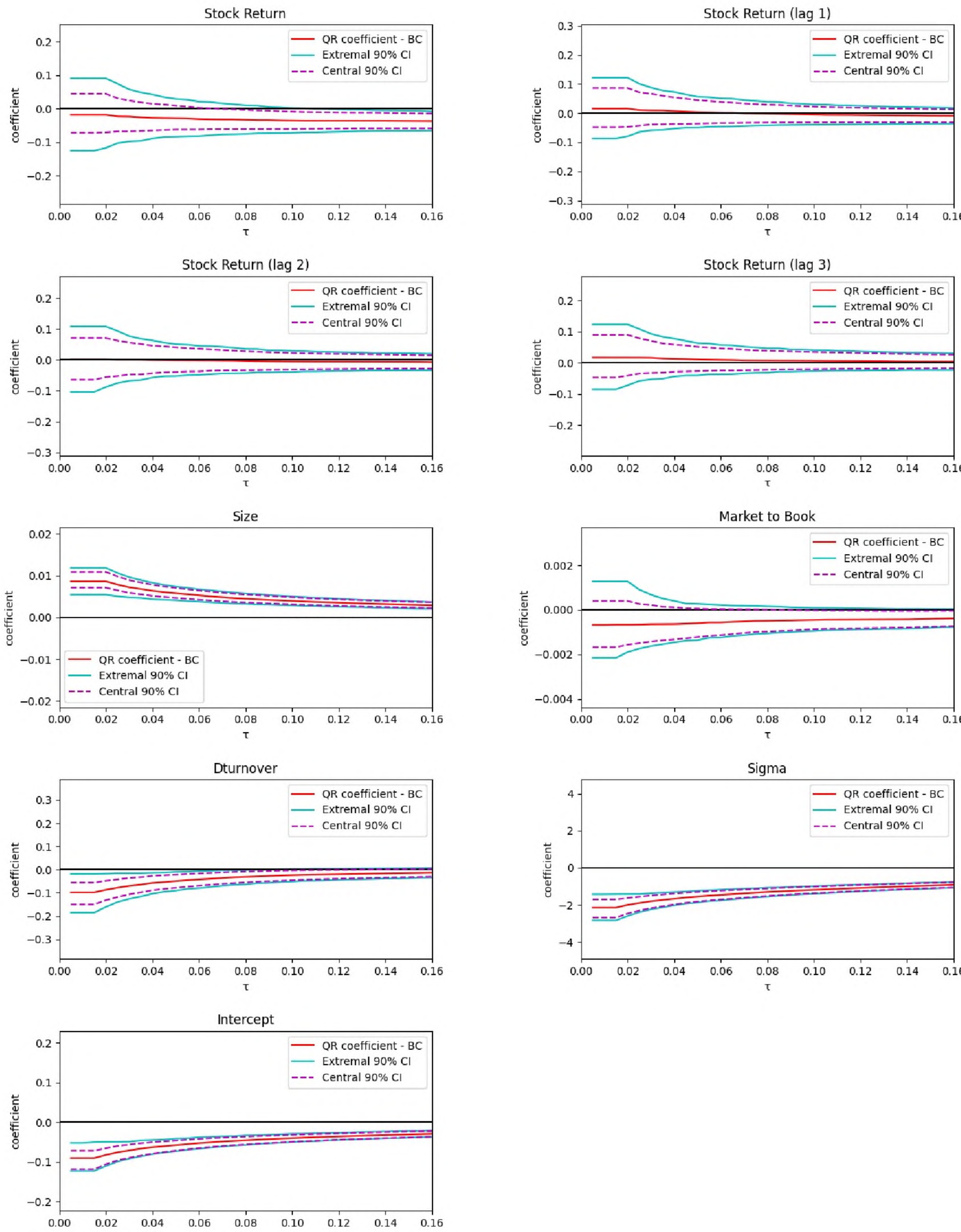


**Figure 3.** Conditional extremal quantile estimates for the lower tails



# Tables

**Table 1. Yearly incidence and magnitude of stock price crashes under CRASH**
Entries report yearly information about the incidence and magnitude of crashes. The crash risk measure is a binary variable that equals 1 when firm experiences at least one crash week during the fiscal year, and zero otherwise. A crash week is identified when the firm-specific weekly return is 3.09 standard deviations below the average firm-specific weekly returns for the entire fiscal year where 3.09 is chosen to generate a frequency of 0.1% in the normal distribution.

| Year | Number of observations | Number of crashes | Percentage of crashes | Average returns during crashes | Standard deviation of returns during crashes | Mean Ncskew during crashes |
|---|---|---|---|---|---|---|
| 2000 | 4,706 | 710 | 15.09 | -0.34 | 0.16 | 0.11 |
| 2001 | 4,161 | 647 | 15.56 | -0.28 | 0.15 | 0.16 |
| 2002 | 3,785 | 729 | 19.26 | -0.29 | 0.15 | 0.13 |
| 2003 | 3,542 | 564 | 15.92 | -0.20 | 0.11 | 0.23 |
| 2004 | 3,533 | 730 | 20.67 | -0.22 | 0.13 | 0.42 |
| 2005 | 3,449 | 700 | 20.29 | -0.21 | 0.11 | 0.51 |
| 2006 | 3,359 | 654 | 19.47 | -0.21 | 0.13 | 0.42 |
| 2007 | 3,310 | 667 | 20.15 | -0.25 | 0.13 | 0.41 |
| 2008 | 3,115 | 638 | 20.18 | -0.28 | 0.16 | -0.05 |
| 2009 | 2,924 | 451 | 15.42 | -0.21 | 0.14 | 0.13 |
| 2010 | 2,820 | 489 | 17.34 | -0.19 | 0.12 | 0.47 |
| 2011 | 2,738 | 508 | 18.55 | -0.21 | 0.13 | 0.24 |
| 2012 | 2,654 | 579 | 21.82 | -0.20 | 0.14 | 0.52 |
| 2013 | 2,664 | 584 | 21.92 | -0.19 | 0.12 | 0.51 |
| 2014 | 2,755 | 619 | 22.47 | -0.21 | 0.13 | 0.41 |
| 2015 | 2,718 | 655 | 24.09 | -0.24 | 0.15 | 0.41 |
| 2016 | 2,611 | 608 | 23.29 | -0.23 | 0.17 | 0.36 |
| 2017 | 2,582 | 661 | 25.60 | -0.21 | 0.13 | 0.57 |
| 2018 | 2,621 | 652 | 24.88 | -0.24 | 0.16 | 0.55 |
| 2019 | 2,610 | 611 | 23.41 | -0.27 | 0.16 | 0.67 |
| Totals | 62,657 | 12,456 | 20.28 | -0.23 | 0.14 | 0.36 |

**Table 2. Yearly incidence and magnitude of stock price crashes**

Entries report yearly information about the incidence and magnitude of crashes. The crash risk measure is the CEQ-CRASH. Focusing on the $0.1th$ quantile estimate our adjusted crash measure is that if a firm $i$ for a given fiscal year $t$ experiences at least one firm-specific weekly return below a CEQ estimated threshold then it is classified as a crashed firm.

| Year | Number of observations | Number of crashes | Percentage of crashes | Average returns during crashes | Standard deviation of returns during crashes | Mean Ncskew during crashes |
|---|---|---|---|---|---|---|
| 2000 | 4,706 | 342 | 7.27 | -0.35 | 0.18 | -0.51 |
| 2001 | 4,161 | 392 | 9.42 | -0.30 | 0.17 | -0.36 |
| 2002 | 3,785 | 360 | 9.51 | -0.29 | 0.19 | -0.69 |
| 2003 | 3,542 | 352 | 9.94 | -0.22 | 0.13 | -0.47 |
| 2004 | 3,533 | 341 | 9.65 | -0.23 | 0.15 | -0.02 |
| 2005 | 3,449 | 330 | 9.57 | -0.22 | 0.14 | 0.04 |
| 2006 | 3,359 | 369 | 10.99 | -0.20 | 0.15 | -0.01 |
| 2007 | 3,310 | 332 | 10.03 | -0.26 | 0.17 | -0.09 |
| 2008 | 3,115 | 318 | 10.21 | -0.31 | 0.20 | -0.93 |
| 2009 | 2,924 | 276 | 9.44 | -0.23 | 0.16 | -1.08 |
| 2010 | 2,820 | 277 | 9.82 | -0.21 | 0.14 | -0.26 |
| 2011 | 2,738 | 325 | 11.87 | -0.22 | 0.14 | -0.29 |
| 2012 | 2,654 | 314 | 11.83 | -0.22 | 0.14 | 0.09 |
| 2013 | 2,664 | 280 | 10.51 | -0.19 | 0.13 | -0.23 |
| 2014 | 2,755 | 309 | 11.21 | -0.22 | 0.17 | -0.16 |
| 2015 | 2,718 | 280 | 10.30 | -0.26 | 0.17 | -0.33 |
| 2016 | 2,611 | 332 | 12.72 | -0.18 | 0.16 | -0.10 |
| 2017 | 2,582 | 277 | 10.73 | -0.22 | 0.14 | 0.04 |
| 2018 | 2,621 | 331 | 12.63 | -0.21 | 0.17 | -0.04 |
| 2019 | 2,610 | 310 | 11.88 | -0.25 | 0.17 | 0.07 |
| Totals | 62,657 | 6,447 | 10.48 | -0.24 | 0.16 | -0.27 |

**Table 3. Yearly incidence and magnitude of stock price crashes under adjusted CRASH** Entries report yearly information about the incidence and magnitude of crashes. The crash risk measure is a binary variable that equals 1 when firm experiences at least one crash week during the fiscal year, and zero otherwise. A crash week is identified when the firm-specific weekly return is 3.09 standard deviations below the average industry weekly returns for the entire fiscal year where 3.09 is chosen to generate a frequency of 0.1% in the normal distribution.

| Year | Number of observations | Number of crashes | Percentage of crashes | Average returns during crashes | Standard deviation of returns during crashes | Mean Ncskew during crashes |
|---|---|---|---|---|---|---|
| 2000 | 4,706 | 942 | 20.02 | -0.37 | 0.13 | -1.35 |
| 2001 | 4,161 | 833 | 20.02 | -0.32 | 0.13 | -1.46 |
| 2002 | 3,785 | 797 | 21.06 | -0.33 | 0.13 | -1.45 |
| 2003 | 3,542 | 519 | 14.65 | -0.24 | 0.11 | -1.79 |
| 2004 | 3,533 | 568 | 16.08 | -0.26 | 0.12 | -1.61 |
| 2005 | 3,449 | 626 | 18.15 | -0.24 | 0.10 | -1.61 |
| 2006 | 3,359 | 601 | 19.89 | -0.24 | 0.12 | -1.64 |
| 2007 | 3,310 | 695 | 20.99 | -0.28 | 0.12 | -1.76 |
| 2008 | 3,115 | 731 | 23.47 | -0.33 | 0.15 | -1.89 |
| 2009 | 2,924 | 436 | 14.91 | -0.25 | 0.15 | -2.19 |
| 2010 | 2,820 | 385 | 13.65 | -0.23 | 0.12 | -1.86 |
| 2011 | 2,738 | 542 | 19.79 | -0.25 | 0.11 | -1.59 |
| 2012 | 2,654 | 526 | 19.82 | -0.24 | 0.11 | -1.67 |
| 2013 | 2,664 | 432 | 16.22 | -0.23 | 0.12 | -1.93 |
| 2014 | 2,755 | 568 | 20.62 | -0.25 | 0.11 | -1.78 |
| 2015 | 2,718 | 606 | 22.30 | -0.28 | 0.13 | -1.70 |
| 2016 | 2,611 | 492 | 18.84 | -0.27 | 0.16 | -1.63 |
| 2017 | 2,582 | 486 | 18.82 | -0.26 | 0.13 | -1.85 |
| 2018 | 2,621 | 476 | 18.16 | -0.29 | 0.15 | -2.04 |
| 2019 | 2,610 | 574 | 21.99 | -0.32 | 0.15 | -1.77 |
| Totals | 62,657 | 11,835 | 18.87 | -0.27 | 0.13 | -1.73 |

**Table 4. Summary statistics**

This table presents summary statistics of the stock price crash measures ($CRASH$ and $CEQ-CRASH$), *Opacity*, *Overinvestment*, and control variables. The CRSP-Compustat data set covers the period 2000-2019. The number of observations for each variable corresponds to the number of non-missing observations for the variables included in the regression models. For variable definitions and details of their computation, see Andreou et al. (2022).

| Variable | Number of observations | Mean | Std Dev | Lower quartile | Median | Upper quartile |
|---|---|---|---|---|---|---|
| CRASH | 33,698 | 0.211 | 0.408 | 0.000 | 0.000 | 0.000 |
| CEQ-CRASH | 33,698 | 0.094 | 0.292 | 0.000 | 0.000 | 0.000 |
| Opacity | 33,698 | 0.231 | 0.229 | 0.022 | 0.160 | 0.279 |
| Overinvestment | 33,698 | 0.020 | 0.233 | -0.423 | -0.016 | 0.087 |
| log(Total assets) | 33,698 | 6.533 | 1.957 | 1.483 | 6.508 | 7.880 |
| Firm age | 33,698 | 3.034 | 0.636 | 1.609 | 2.996 | 3.555 |
| Zscore | 33,698 | 5.557 | 2.011 | 4.556 | 4.863 | 5.624 |
| Market to Book | 33,698 | 3.269 | 3.809 | -7.790 | 2.218 | 3.735 |
| Return on equity | 33,698 | 0.034 | 0.423 | -2.539 | 0.093 | 0.167 |
| Detrended turnover | 33,698 | 0.001 | 0.020 | -0.058 | 0.000 | 0.007 |
| Ncskew | 33,698 | -0.016 | 0.783 | -4.822 | -0.053 | 0.361 |

**Table 5. The effect of opacity and overinvestment on future stock price crashes**

This table reports logistic regression estimates for the relation between opacity, overinvestment, and stock price crashes. Estimates are derived using the CRSP-Compustat universe for the period 2000−2019. The dependent variable in models (1)-(3) is $CRASH$ estimated as per *Eq.* (7) and measured in fiscal year $t+1$. The dependent variable in models (4)-(6) is $CEQ-CRASH$ estimated as per *Eqs.* (2) and (3) and measured in fiscal year $t+1$. The explanatory variables are measured in fiscal year $t$ or earlier. The estimates include a constant and different conditional fixed effects (as indicated at the bottom of the table) whose coefficients are suppressed. Industry fixed effects are defined based on the Fama–French 12-industry classification. All continuous variables are winsorized at the $1^{st}$ and $99^{th}$ percentiles. Robust standard errors clustered at the firm level are shown in parentheses. The symbols ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% level, respectively.

| | CRASH | | | EQR-CRASH | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Opacity | 0.074 | | 0.065 | 0.295*** | | 0.271** |
| | (0.066) | | (0.067) | (0.107) | | (0.109) |
| Overinvestment | | 0.067 | 0.058 | | 0.193** | 0.164* |
| | | (0.062) | (0.063) | | (0.092) | (0.093) |
| log(Total assets) | 0.037*** | 0.033*** | 0.035*** | 0.084*** | 0.070*** | 0.079*** |
| | (0.009) | (0.009) | (0.009) | (0.021) | (0.021) | (0.021) |
| Firm age | -0.098*** | -0.099*** | -0.098*** | -0.091** | -0.097** | -0.090** |
| | (0.027) | (0.027) | (0.027) | (0.044) | (0.044) | (0.044) |
| Market to Book | 0.009** | 0.009** | 0.009** | -0.007 | -0.006 | -0.007 |
| | (0.004) | (0.004) | (0.004) | (0.006) | (0.006) | (0.006) |
| Zscore | 0.011 | 0.012 | 0.012 | 0.014 | 0.017 | 0.016 |
| | (0.008) | (0.008) | (0.008) | (0.014) | (0.014) | (0.014) |
| Return on equity | 0.083** | 0.082** | 0.085** | -0.204*** | -0.216*** | -0.201*** |
| | (0.036) | (0.036) | (0.036) | (0.047) | (0.048) | (0.048) |
| Detrended turnover | 0.066 | 0.024 | 0.032 | 2.074* | 1.962* | 1.976* |
| | (0.707) | (0.711) | (0.710) | (1.072) | (1.081) | (1.070) |
| Ncskew | 0.036*** | 0.036*** | 0.036*** | 0.018 | 0.018 | 0.018 |
| | (0.017) | (0.017) | (0.017) | (0.027) | (0.027) | (0.027) |
| Ncskew ($t-1$) | 0.063*** | 0.063*** | 0.063*** | -0.001 | -0.001 | -0.001 |
| | (0.018) | (0.018) | (0.018) | (0.026) | (0.026) | (0.026) |
| Fixed effects | Year, Industry | Year, Industry | Year, Industry | Year, Industry | Year, Industry | Year, Industry |
| No. of observations | 33,285 | 33,285 | 33,285 | 33,285 | 33,285 | 33,285 |
| Pseudo R-squared | 0.012 | 0.012 | 0.012 | 0.059 | 0.059 | 0.059 |