Option-Implied Bounds for the Crash Probability of a Stock

Ran Shi

October, 2021

Forward-looking crash probability of an individual stock

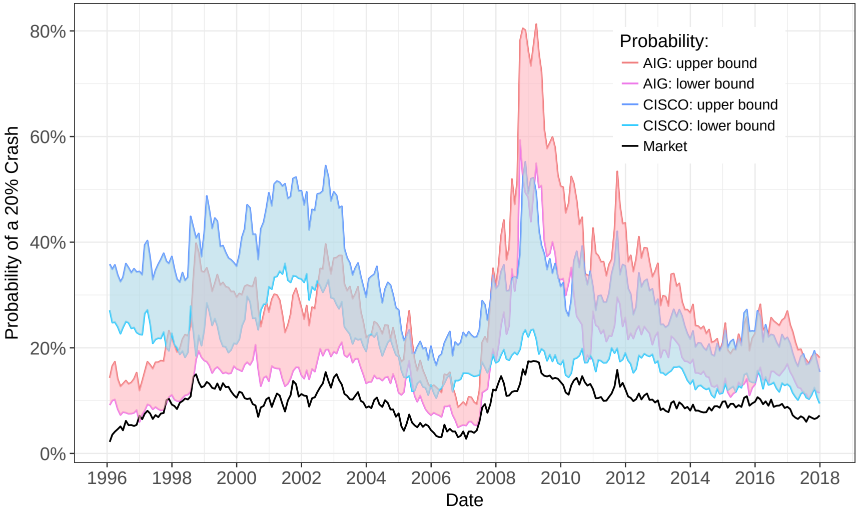
- An enormous body of literature on predicting expected returns: $\mathbb{E}[R_i]$
- Relatively thin on predicting a crash: $\mathbb{P}[R_i \leq q]$
  - Chen-Hong-Stein (2001), Daniel-Klos-Rottke (2017), Greenwood-Shleifer-You (2017), etc:

    Mainly characteristics-driven and some are only about "skewness"
  - Bates (1991): a case study of the Black Monday using option prices
- Some measures of downside risk:
  - Ang-Chen-Xing (2006), Kelly-Jiang (2014), Lu-Murray (2018), etc.
  - Do not directly answer the question above
- Martin (2017) answers the question for the market: $\mathbb{P}[R_m \leq q]$

This paper

- Forecasting the crash probability of a stock
  - Not a point forecast, forecasting bounds
  - Probabilistic bounds, not statistical confidence intervals: 100% coverage
  - Directly computed from real-time option prices
  - No distributional assumption
  - No parameter estimation
  - Theoretically sharp and empirically tight
  - Drives out characteristics in-sample
  - Out-performs a data snooper out-of-sample
- Extensions: an upper bound on the expected return of a stock

Probabilities of a 20% crash in one year



Theory and method (1)

- A one-period CRRA investor chooses to hold the market: her first-order condition implies that there is a stochastic discount factor
$$M \propto R_m^{-\gamma}$$
- Let $M = R_m^{-\gamma}/\lambda$, λ is a constant
- The physical expectation of any random payoff X can be rewritten under the risk-neutral probabilities:
$$\mathbb{E}[X] = \mathbb{E}[\underbrace{M\lambda R_m^{\gamma}}_{\equiv 1} X] = \lambda\mathbb{E}[M(R_m^{\gamma}X)] = \frac{\lambda}{R_f}\mathbb{E}^*[R_m^{\gamma}X],$$
where $\mathbb{E}^*[\cdot]$ represents the risk-neutral expectation

Theory and method (2)

- We have

$$\mathbb{E}[X] = \frac{\lambda}{R_f}\mathbb{E}^*[R_m^\gamma X]$$

Let $X = 1$,

$$1 = \frac{\lambda}{R_f}\mathbb{E}^*[R_m^\gamma]$$

Dividing each side:

$$\mathbb{E}[X] = \frac{\mathbb{E}^*[R_m^\gamma X]}{\mathbb{E}^*[R_m^\gamma]}$$

- A general framework for a payoff contingent on R_i: $X = h(R_i)$
  - $X = \mathcal{I}\{R_i \leq q\}$, crash probability of a stock: $\mathbb{E}[\mathcal{I}\{R_i \leq q\}] = \mathbb{P}[R_i \leq q]$

Theory and method (3)

- For a payoff contingent on R_i:
$$\mathbb{E}[h(R_i)] = \frac{\mathbb{E}^*[R_m^\gamma h(R_i)]}{\mathbb{E}^*[R_m^\gamma]}$$
$$= \frac{\int x^\gamma h(y)\,\mathrm{d}Q_{mi}(x, y)}{\int x^\gamma\,\mathrm{d}Q_m(x)}$$
  - Q_{mi}: risk-neutral joint distribution of R_m and R_i
  - Q_m: risk-neutral marginal distribution of R_m
  - Q_i: risk-neutral marginal distribution of R_i (not appearing here)
  - Probability CDFs
- Breeden and Litzenberger (1978)
  - Recovering risk-neutral (marginal) distributions from option prices

Theory and method (4)

- Breeden-Litzenberger for the marginal distribution Q of returns
$$\text{put}(K) = \frac{1}{R_f} \int_0^\infty \max(K - xS_0, 0)\, \mathrm{d}Q(x)$$
Integrating by parts
$$\text{put}(K) = \frac{S_0}{R_f} \int_0^{\frac{K}{S_0}} Q(x)\, \mathrm{d}x$$
Taking derivatives with regard to K
$$Q\left(\frac{K}{S_0}\right) = R_f \text{put}'(K),$$
Applying the put-call parity
$$Q\left(\frac{K}{S_0}\right) = R_f \text{call}'(K) + 1$$
- Q_i: stock option; Q_m: S&P 500 index option B-L detail

Theory and method (5)

Recall that

$$\mathbb{E}[h(R_i)] = \frac{\mathbb{E}^*[R_m^\gamma h(R_i)]}{\mathbb{E}^*[R_m^\gamma]} = \frac{\int x^\gamma h(y)\,\mathrm{d}Q_{mi}(x, y)}{\int x^\gamma\,\mathrm{d}Q_m(x)}$$

- The denominator is known invoking Breeden-Litzenberger
- What to do about the numerator?
  - Need to quantify the joint distribution

Theory and method (6)

- How to quantify $Q_{mi}(x, y)$?
  - Multivariate Breeden-Litzenberger?
  - Methodologically feasible (Martin, 2018)
  - Need a very large number of traded options written on $(S_m + \alpha S_i)$ (Ross, 1976; Martin, 2018)
    - Many different values of α!
    - With different strikes for each of them
    - Deep market
  - In practice, it is almost impossible to fully characterize the joint distribution!
- No deterministic answer to $\mathbb{E}[h(R_i)]$

Theory and method (7)

- But we can bound it
- What do we know about Q_{mi}?
  - Both the marginals Q_m and Q_i, using the Breeden-Litzenberger approach
- What remains unknown is the dependence structure between the stock return and the market return
- Use copula theory to obtain sharp bounds (Sklar, 1959)

Theory and method (8)

Dissecting ingredients in the joint distribution

$$\begin{tikzcd}
& Q_{mi}(x,\ y) & \\
Q_m(x) \arrow[ur] & ? \arrow[u] & Q_i(y) \arrow[ul]
\end{tikzcd}$$

Theory and method (8)

Dissecting ingredients in the joint distribution

$$
\begin{tikzcd}
 & Q_{mi}(x,\ y) & \\
Q_m(x) \arrow[ur] & C(\cdot,\cdot) \arrow[u] & Q_i(y) \arrow[ul]
\end{tikzcd}
$$

Theory and method (8)

Dissecting ingredients in the joint distribution

$$
\begin{tikzcd}
 & Q_{mi}(x,\ y) \equiv C(Q_m(x), Q_i(y)) & \\
Q_m(x) \arrow[ur] & C(\cdot,\cdot) \arrow[u] & Q_i(y) \arrow[ul]
\end{tikzcd}
$$

Theory and method (9)

- A copula function glues two marginals together
- It is a mapping from the square $[0,1]^2$ to $[0,1]$
- It is the joint distribution of two arbitrary $U[0,1]$ random variables
- Now that $Q_{mi}(x,y) = C(Q_m(x), Q_i(y))$, changing variables in the integral, the numerator $\mathbb{E}^*[R_m^\gamma h(R_i)]$ is

$$\int x^\gamma h(y)\,\mathrm{d}Q_{mi}(x,y) = \int_{[0,1]^2} \underbrace{\left[Q_m^{-1}(u)\right]^\gamma h\left(Q_i^{-1}(v)\right)}_{\text{a known function}} \mathrm{d}C(u,v)$$

- Our goal: bounding this integral

Theory and method (10)

A useful theorem:

Theorem (Fréchet-Hoeffding bounds)

If $C(u, v)$ is a copula function, then

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v), \quad \text{for all } (u, v) \in [0, 1]^2.$$

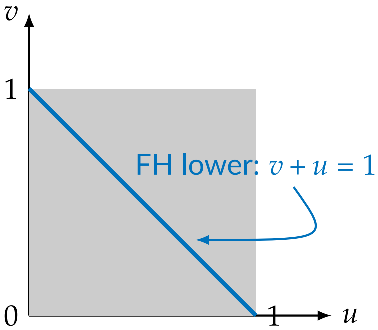
and an easy-to-handle special case: h is non-decreasing

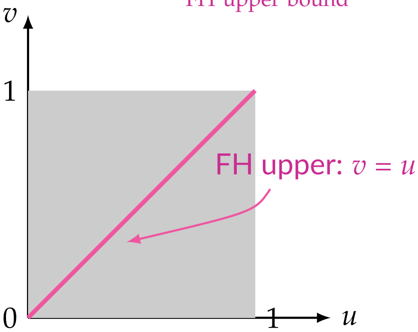
Theory and method (11)

If $h(\cdot)$ is non-decreasing,

$$\min \int_{[0,1]^2} \text{something}\, \mathrm{d}C(u,v) = \int_{[0,1]^2} \text{something}\, \mathrm{d}\Bigg(\underbrace{\min C(u,v)}_{\text{FH lower bound}} \Bigg)$$

$$\max \int_{[0,1]^2} \text{something}\, \mathrm{d}C(u,\ v) = \int_{[0,1]^2} \text{something}\, \mathrm{d}\Bigg(\underbrace{\max C(u,\ v)}_{\text{FH upper bound}} \Bigg)$$





Theory and method (12)

Sharp bounds on the expectation of a payoff contingent on R_i

If function $h : \mathbb{R}_+ \mapsto \mathbb{R}$ is non-decreasing, then

$$\frac{\int_0^1 \left[Q_m^{-1}(u)\right]^{\gamma} h\left(Q_i^{-1}(1-u)\right) \mathrm{d}u}{\int_0^1 \left[Q_m^{-1}(u)\right]^{\gamma} \mathrm{d}u} \leq \mathbb{E}[h(R_i)] \leq \frac{\int_0^1 \left[Q_m^{-1}(u)\right]^{\gamma} h\left(Q_i^{-1}(u)\right) \mathrm{d}u}{\int_0^1 \left[Q_m^{-1}(u)\right]^{\gamma} \mathrm{d}u}$$

- Widening effects of γ: $\nearrow$ the upper bound and $\searrow$ the lower bound
- $\gamma = 0$: both equal $\mathbb{E}^*[h(R_i)]$
- $\gamma = \infty$: $\mathbb{E}[h(R_i)] \in [h(0), h(\infty)]$
  - Example: if $h(R_i) = \mathcal{I}\{R_i > q\}$, $h(0) = 0$, $h(\infty) = 1$
- Any $\gamma \in [1, 2]$ gives tight bounds and similar results

Theory and method (13)

- Interpretations: countermonotonicity and comonotonicity under the risk-neutral distribution
- When achieving the lower bound
$$Q_m(R_m) + Q_i(R_i) = 1 \quad \Rightarrow R_i = Q_i^{-1}(1 - Q_m(R_m)) \triangleq \mathcal{L}(R_m)$$
the stock return is a monotonic decreasing function of the market return; the risk-neutral beta: $\beta^* \approx \mathcal{L}'(R_m^*) < 0$
- When achieving the upper bound,
$$Q_m(R_m) = Q_i(R_i) \quad \Rightarrow R_i = Q_i^{-1}(Q_m(R_m)) \triangleq \mathcal{U}(R_m)$$
the stock return is a monotonic increasing function of the market return; the risk-neutral beta: $\beta^* \approx \mathcal{U}'(R_m^*) > 0$

Theory and method (14)

- F-H bounds are sharp in two dimension: a mathematical fact
- An example
  - Consider $\gamma = 1$ and $h(R_i) = R_i$
  - Now $\mathbb{E}[R_i] = \mathbb{E}^*[R_m R_i]/\mathbb{E}^*[R_m] = \mathbb{E}^*[R_m R_i]/R_f$
  - An upper bound on the expected return based on *Fréchet-Hoeffding*

$$\mathbb{E}[R_i] \leq \underbrace{\frac{\mathbb{E}^*\left[R_m Q_i^{-1}(Q_m(R_m))\right]}{R_f}}_{FH}$$

  - Another upper bound based on *Cauchy-Schwartz*:

$$\mathbb{E}[R_i] = \frac{\mathrm{cov}^*[R_m,\ R_i] + \mathbb{E}^*[R_m]\mathbb{E}^*[R_i]}{R_f} \leq \underbrace{\frac{\sqrt{\mathrm{var}^*[R_m]\mathrm{var}^*[R_i]} + R_f^2}{R_f}}_{CS}$$

Theory and method (15)

Under the risk-neutral probability

- Normal benchmark: $\mathrm{var}^*[R_m] = \sigma_m^2$ and $\mathrm{var}^*[R_i] = \sigma_i^2$

$$FH \equiv CS = \frac{R_f^2 + \sigma_m \sigma_i}{R_f},$$

  FH bounds are trivial

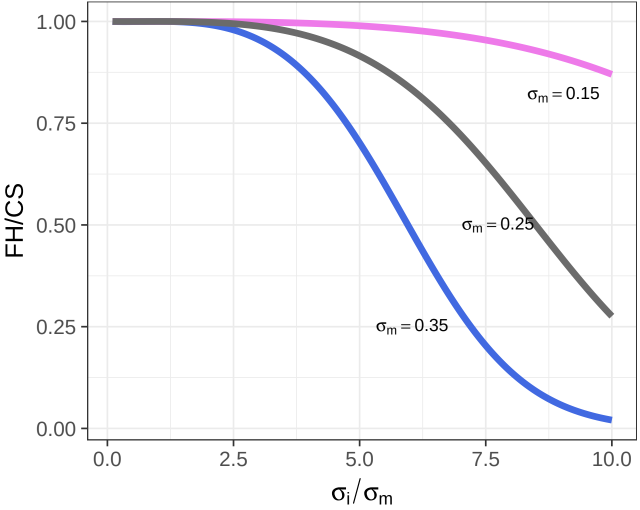
- Lognormal case: $\mathrm{var}^*[\log R_m] = \sigma_m^2$ and $\mathrm{var}^*[\log R_i] = \sigma_i^2$

$$FH = R_f \exp(\sigma_m \sigma_i),$$

$$CS = R_f + R_f \sqrt{\left[\exp(\sigma_m^2) - 1\right]\left[\exp(\sigma_i^2) - 1\right]}$$

Theory and method (16)

Let $\sigma_m = 15\%, 25\%, 35\%$ and $\sigma_i = k\sigma_m, k \in [1/10,\ 10]$



Summary of theory and method

- Till now, I have introduced a methodological framework for bounding quantities like $\mathbb{E}[h(R_i)]$ using real-time option prices
- The bounds are sharp in theory, cannot be improved without
  - New data: not available
  - Additional assumptions: might be fragile
- Are these bounds tight in forecasting crash probabilities?
  - Being close to the true expectation
  - The interval between the bounds covering the true expectation
  - An empirical quest

Data

- S&P 500 index and stock constituents from **Compustat**
- Firm characteristics from **Compustat**
- Option implied volatilities from **OptionMetrics**
  - Underlying stocks belonging to the S&P 500 index
  - Monthly from $1996/01$ to $2017/12$
  - Maturing in $1, 3, 6$ and 12 months
  - The total number of firms included is over $1,000$
  - On average around 480 firms each month
  - Over $120,000$ firm-month observations per maturity
- Risk-free rates from **OptionMetrics**
- Price, return, and volume data from **CRSP**

Computation

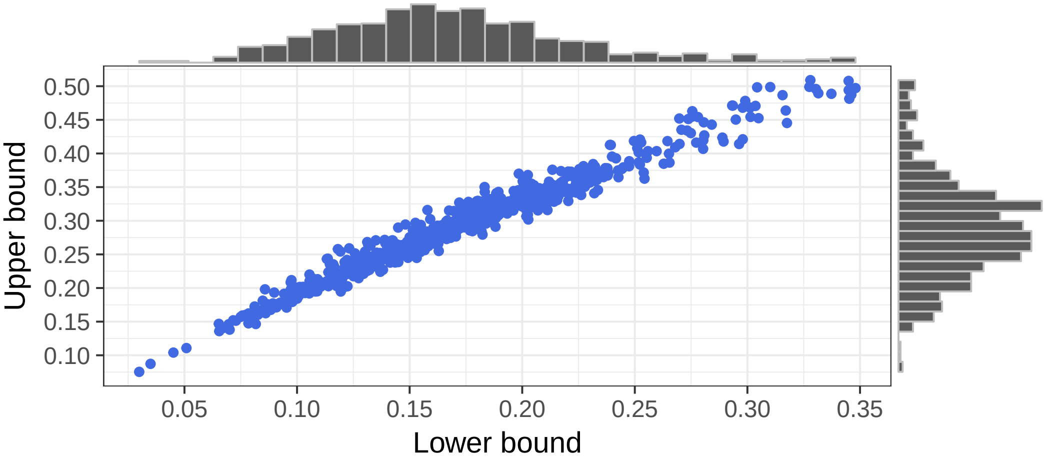
- Set $\gamma = 1.5$
- The bounds are directly applicable to $h(R_i) = \mathcal{I}\{R_i > q\}$
- Crash probability bounds: $\mathbb{P}[R_i \leq q] = 1 - \mathbb{E}[\mathcal{I}\{R_i > q\}]$
- Focus on a crash of $5\%, 10\%$ and 20%: $q = 95\%, 90\%$ and 80%
- For stock i at month t, compute the lower and upper bounds for $\tau = 1, 3, 6$ and 12 months, denoted by

  $$\mathrm{ProbLower}_{i,t}(\tau, q), \quad \mathrm{ProbUpper}_{i,t}(\tau, q)$$

  which forecast $\mathbb{P}[R_{i,t\to t+\tau} \leq q]$, the (time-$t$ conditional) probability of a $(1-q)\%$ crash at time $t+\tau$, for stock i

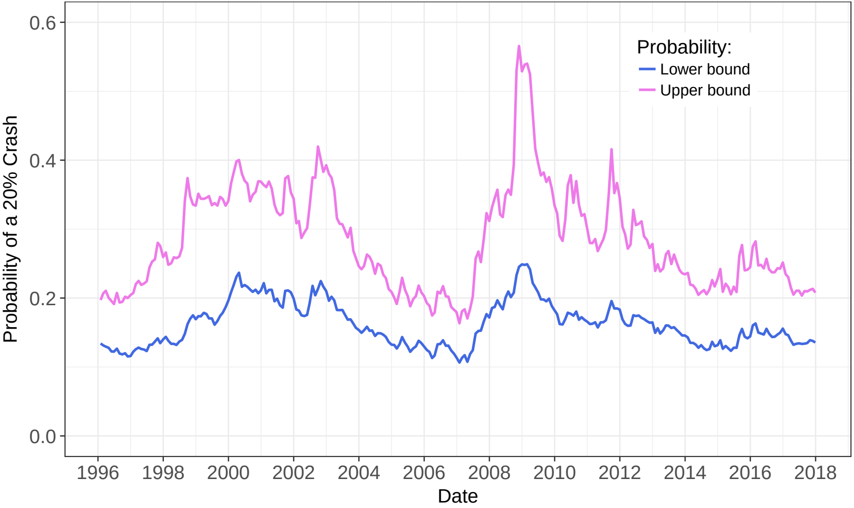
Summarizing the data: variations among firms

Probability of a 20% crash (one-year horizon), time-series median (obs. ≥ 48)



Summarizing the data: variations across time

Probability of a 20% crash (one-year horizon), cross-sectional median



Option implied bounds are tight: Tests

- If both the bounds are close to the true probability of crashes, panel regressions like

$$\mathcal{I}[R_{i,t\to t+\tau} \leq q] = \alpha^L + \beta^L \,\mathrm{ProbLower}_{i,t}(\tau, q) + \varepsilon_{i,t+\tau},$$

  or

$$\mathcal{I}[R_{i,t\to t+\tau} \leq q] = \alpha^U + \beta^U \,\mathrm{ProbUpper}_{i,t}(\tau, q) + \varepsilon_{i,t+\tau},$$

  would give us

$$\alpha^L = \alpha^U = 0, \quad \beta^L \approx 1, \quad \beta^U \approx 1$$

  for any horizon τ and return threshold q

- Ideally, $\beta^L \gtrapprox 1$ and $\beta^U \lessapprox 1$ so that the bounds also offer good coverage

Option implied bounds are tight: Results (1)

Down by 5%, i.e., $q = 0.95$

	Lower bound				Upper bound			
Mon.	1	3	6	12	1	3	6	12
α	-0.037	-0.049	-0.080	-0.021	-0.045	-0.042	-0.034	-0.009
	(0.016)	(0.021)	(0.026)	(0.028)	(0.022)	(0.037)	(0.046)	(0.051)
	[0.017]	[0.030]	[0.041]	[0.043]	[0.030]	[0.067]	[0.061]	[0.074]
β	1.129	1.144	1.239	1.051	0.923	0.813	0.717	0.578
	(0.083)	(0.078)	(0.087)	(0.092)	(0.090)	(0.104)	(0.112)	(0.108)
	[0.083]	[0.137]	[0.124]	[0.230]	[0.128]	[0.205]	[0.181]	[0.164]
R^2-adj.	3.99%	2.54%	2.42%	1.77%	4.10%	2.35%	1.71%	1.26%

Option implied bounds are tight: Results (2)

Down by 10%, i.e., $q = 0.9$

	Lower bound				Upper bound			
Mon.	1	3	6	12	1	3	6	12
α	-0.023	-0.039	-0.063	-0.068	-0.031	-0.041	-0.044	-0.068
	(0.007)	(0.011)	(0.015)	(0.018)	(0.009)	(0.017)	(0.026)	(0.033)
	[0.009]	[0.012]	[0.029]	[0.036]	[0.008]	[0.018]	[0.041]	[0.061]
β	1.121	1.165	1.257	1.195	0.861	0.791	0.717	0.634
	(0.093)	(0.079)	(0.081)	(0.081)	(0.084)	(0.079)	(0.086)	(0.087)
	[0.122]	[0.091]	[0.159]	[0.142]	[0.074]	[0.071]	[0.140]	[0.138]
R^2-adj.	7.05%	5.02%	4.27%	3.45%	7.04%	4.64%	3.22%	2.60%

Option implied bounds are tight: Results (3)

Down by 20%, i.e., $q = 0.8$

	Lower bound				Upper bound			
Mon.	1	3	6	12	1	3	6	12
α	-0.005	-0.014	-0.021	-0.045	-0.011	-0.017	-0.017	-0.051
	(0.002)	(0.005)	(0.008)	(0.009)	(0.003)	(0.007)	(0.012)	(0.016)
	[0.002]	[0.008]	[0.015]	[0.025]	[0.003]	[0.008]	[0.023]	[0.040]
β	1.034	1.152	1.193	1.105	0.703	0.680	0.602	0.519
	(0.136)	(0.105)	(0.097)	(0.085)	(0.104)	(0.076)	(0.074)	(0.065)
	[0.154]	[0.184]	[0.201]	[0.151]	[0.116]	[0.083]	[0.110]	[0.162]
R^2-adj.	6.88%	6.84%	5.90%	5.31%	6.67%	6.08%	4.46%	3.99%

Option implied bounds are tight: Robustness (1)

- Logistic regressions
- Transform:

$$\text{LogOdds} = \log\left(\frac{\text{Prob}}{1 - \text{Prob}}\right)$$

- Model:

$$\mathcal{I}\{R \leq q\} \sim \text{Bernoulli}(p)$$
$$p = \frac{\exp(\alpha + \beta\,\text{LogOdds})}{1 + \exp(\alpha + \beta\,\text{LogOdds})},$$

- Panel regressions
- Looking for $\alpha = 0$, $\beta^L \gtrapprox 1$ and $\beta^U \lessapprox 1$ across all horizons (τ) and thresholds (q)

Option implied bounds are tight: Robustness (2)

Down by 20%, i.e., $q = 0.8$, logistic regression

	Lower bound				Upper bound			
Mon.	1	3	6	12	1	3	6	12
α	0.232	0.408	0.416	0.311	−0.023	−0.476	−0.782	−1.330
	[0.179]	[0.282]	[0.281]	[0.223]	[0.324]	[0.266]	[0.282]	[0.208]
β	1.148	1.206	1.218	1.362	1.342	1.153	1.031	1.065
	[0.031]	[0.090]	[0.086]	[0.166]	[0.104]	[0.137]	[0.147]	[0.232]
R^2-Adj.*	17.88%	11.35%	7.99%	7.01%	18.60%	10.58%	6.62%	5.59%

*: Adjusted McFadden's pseudo-R^2

Option implied bounds are tight: Robustness (3)

Controlling for eighteen characteristics:

- lagret^1, $\text{lagret}^{2,12}$ (mom)
- vol^1, vol^3, vol^6
- β, logsize, bm
- cape, prof (gross prof./asset), inv (inv./asset), lev (debt/asset), accrl (accrual/asset), debida (debt/EBITDA), cash (cce/curr. liability)
- tnover, dtnover (turnover ratio and detrended turnover ratio)
- shortint (short interest)

Option implied bounds are tight: Robustness (4.a)

Down by 20%, i.e., $q = 0.8$, without controls

	Lower bound				Upper bound			
Mon.	1	3	6	12	1	3	6	12
α	-0.005	-0.014	-0.021	-0.045	-0.011	-0.017	-0.017	-0.051
	(0.002)	(0.005)	(0.008)	(0.009)	(0.003)	(0.007)	(0.012)	(0.016)
	[0.002]	[0.008]	[0.015]	[0.025]	[0.003]	[0.008]	[0.023]	[0.040]
β	1.034	1.152	1.193	1.105	0.703	0.680	0.602	0.519
	(0.136)	(0.105)	(0.097)	(0.085)	(0.104)	(0.076)	(0.074)	(0.065)
	[0.154]	[0.184]	[0.201]	[0.151]	[0.116]	[0.083]	[0.110]	[0.162]
R^2-adj.	6.88%	6.84%	5.90%	5.31%	6.67%	6.08%	4.46%	3.99%

Option implied bounds are tight: Robustness (4.b)

Down by 20%, i.e., $q = 0.8$, with controls

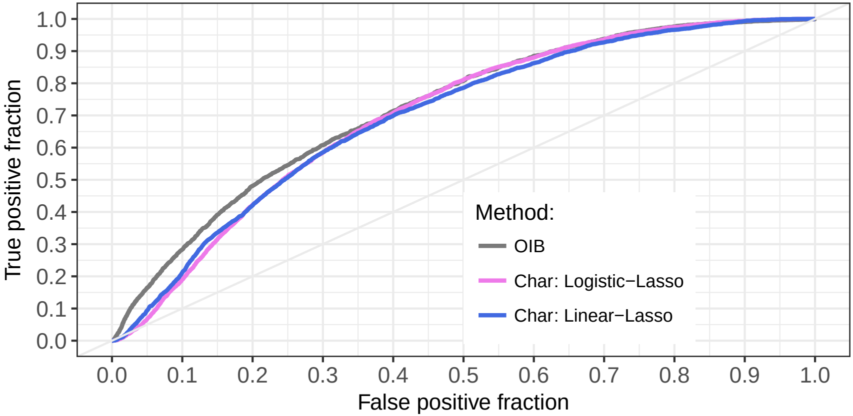
	Lower bound				Upper bound			
Mont.	1	3	6	12	1	3	6	12
α	0.014	0.100	0.107	0.110	0.004	0.087	0.107	0.106
	(0.021)	(0.051)	(0.070)	(0.070)	(0.022)	(0.053)	(0.072)	(0.072)
	[0.030]	[0.068]	[0.146]	[0.102]	[0.022]	[0.052]	[0.085]	[0.113]
β	0.802	1.062	1.418	1.046	0.529	0.460	0.434	0.287
	(0.159)	(0.174)	(0.187)	(0.153)	(0.152)	(0.109)	(0.103)	(0.075)
	[0.127]	[0.190]	[0.246]	[0.274]	[0.081]	[0.151]	[0.103]	[0.151]
R^2-adj.	5.35%	5.57%	5.31%	4.92%	5.24%	5.03%	4.09%	3.95%

Option-implied bounds forecast crashes

- Competitor: “Fishing” desperately
  - Include all eighteen characteristics
  - Train predicative models using data from 1996 to 2010
    - Both linear and logistic models
    - Variable selection using Lasso
    - Tuning parameters for Lasso: 10-fold CV using the training sample
  - Out-of-sample forecasting: 2011-2017
  - Have to remove any observations with missing data
- Option-implied bounds are directly used to forecast
- Gauging performances: ROC curves

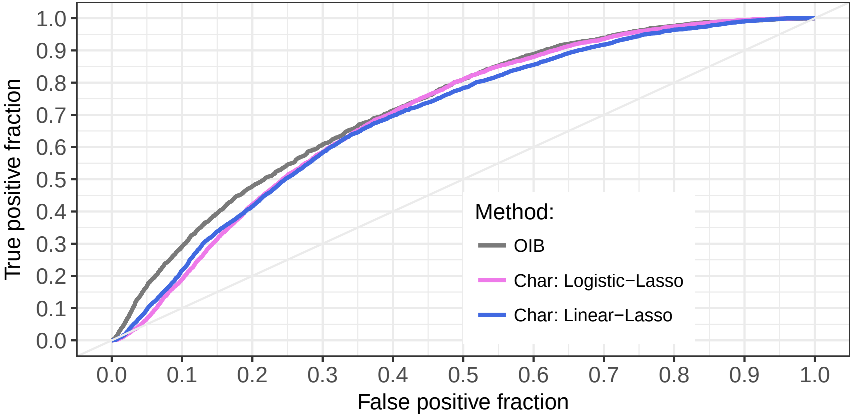
Option-implied bounds forecast crashes

Down by 20% in one year



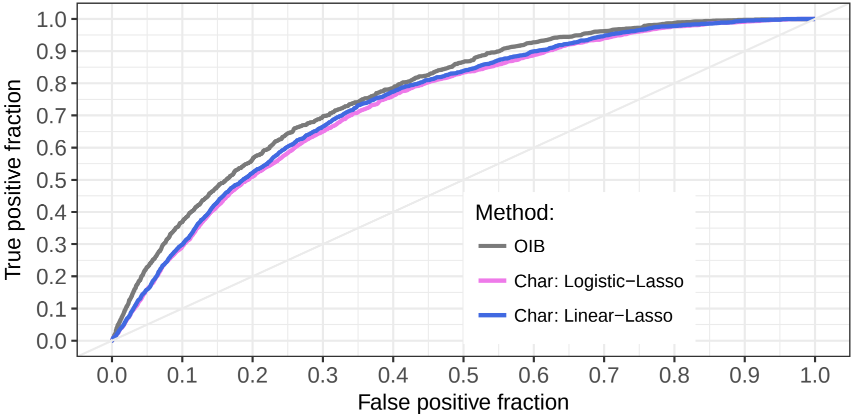
Option-implied bounds forecast crashes

Down by 20% in six months



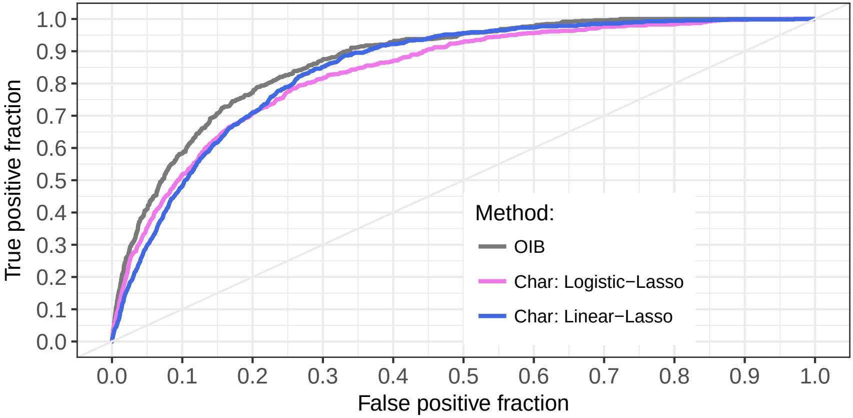
Option-implied bounds forecast crashes

Down by 20% in three months



Option-implied bounds forecast crashes

Down by 20% in one month



Summary

- A methodological framework offering sharp bounds for $\mathbb{E}[h(R_i)]$ using option prices
  - Analytically tractable solution when h is monotonic
  - Computational solution does exist for a general h (available in the paper)
  - Maybe of its own interest
- One variable to rule them all
  - Bounding the crash probability of a stock
  - Forecasting crashes
- New disciplined approach with lots of applications

Extension: An upper bound on the expected return (1)

Let $h(R_i) = R_i$ and consider the case when $\gamma = 1$, recall that

$$\mathbb{E}[R_i] \leq \frac{\mathbb{E}^*[R_m Q_i^{-1}(Q_m(R_m))]}{R_f}$$
$$= \frac{\mathbb{E}^*[g_i(S_m)]}{R_f}$$

assuming no dividend ($R_m = S_m/S_0$), where

$$g_i(S_m) = \frac{S_m}{S_0} Q_i^{-1}\left(Q_m\left(\frac{S_m}{S_0}\right)\right)$$

Extension: An upper bound on the expected return (2)

For any twice differentiable function $g(x)$ defined on $\mathbb{R}_+$,

$$g(S) = g(F) + g'(F)(S - F) + \int_0^F g''(K)(K - S)^+ \, \mathrm{d}K + \int_F^\infty g''(K)(S - K)^+ \, \mathrm{d}K$$

Taking expectations with regard to S

$$\mathbb{E}^*[g(S)] = g(F) + g'(F)\underbrace{(\mathbb{E}^*[S] - F)}_{=0}$$
$$+ \int_0^F g''(K) R_f \mathrm{put}(K) \, \mathrm{d}K + \int_F^\infty g''(K) R_f \mathrm{call}(K) \, \mathrm{d}K$$

- Recall that $\mathbb{E}[R_i] \leq \mathbb{E}^*[g_i(S_m)]/R_f$
- Apply the result above to $g_i(S_m)$
- $g_i(F_m) = R_f Q_i^{-1}\left(Q_m(R_f)\right)$

Extension: An upper bound on the expected return (3)

With time horizon being T, define the comonotonic volatility index (CVIX) as

$$\mathrm{CVIX}_i(T) = \frac{1}{R_f T} \left\{ \int_0^{F_m} g_i''(K)\mathrm{put}_m(K)\,\mathrm{d}K + \int_F^{\infty} g_i''(K)\mathrm{call}_m(K)\,\mathrm{d}K \right\}$$

An upper bound for the expected return of a stock, $\mathbb{E}[R_i]$

$$\frac{\mathbb{E}[R_i]}{TR_f} \leq \frac{Q_i^{-1}(Q_m(R_f))}{TR_f} + \mathrm{CVIX}_i(T)$$

Extension: An upper bound on the expected return (4)

Testing the theory by running the following regression

$$\frac{R_{i,t\to t+\tau}}{TR_f} = \alpha_t + \beta\,\mathrm{CVIX}_{i,t}(T) + \epsilon_{i,t+\tau},$$

- Time-fixed effect for the $\frac{Q_i^{-1}(Q_m(R_f))}{TR_f}$ term
- $T = 12/\tau$ for monthly data
- Ideally $\beta = 1$

Extension: An upper bound on the expected return (5)

Mont.	1	3	6	12
β	0.924	0.902	1.044	0.543
	(0.728)	(0.510)	(0.433)	(0.215)
	[0.740]	[0.647]	[0.775]	[0.386]
R^2-Adj.	21.47%	23.74%	24.10%	24.24%
Projected R^2-Adj.	0.00%	0.23%	0.80%	0.99%

Appendix: Risk-neutral marginals (1)

- American options and dividend payments: decomposing the premium as

$$\begin{aligned}&\text{option pre.(EU wo. dividend)}\\ =&\text{option pre.(AM w. dividend)} + \underbrace{\text{dividend adj.} - \text{early exercise}}_{\text{Binomial tree with dividend projection}}\end{aligned}$$

  - Proprietary dividend yield projection by OptionMetrics: q
  - Fit a Binomial tree for each combination of maturity lengths and strike prices: $u = e^{\sigma \Delta t}$, $d = e^{-\sigma \Delta t}$, $R = e^{(r_f - q)\Delta t}$
  - Use the implied volatilities to compute the Black-Scholes prices
  - $\approx$ computing the dividend adj. and early exercise premium under B-S

Appendix: Risk-neutral marginals (2)

- Microstructural concerns
  - Drop all options with non-standard settlement
  - The midpoints of the bid/ask price must be above the intrinsic value (remove all missing observations)
  - The dispersion parameter of the volatility surface must be smaller than 0.1 (a proprietary measure computed by OptionMetrics to monitor the fineness of the volatility surface)
  - Only use out-of-the-money options, that is,

$$Q\left(\frac{K}{S_0}\right) = \begin{cases} R_f \mathrm{put}'(K), & K \leq R_f S_0 \\ R_f \mathrm{call}'(K) + 1, & K > R_f S_0 \end{cases}$$

Appendix: Risk-neutral marginals (3)

- Taking derivatives: $\text{put}'(K)$ and $\text{call}'(K)$
  - Observables
    - Let $x_i = K_i$
    - Let $y_i = \text{put}(K_i)$ if $K_i \leq R_f S_0$ and $y_i = \text{call}(K_i) - S_0 + K_i / R_f$ if $K_i > R_f S_0$
  - Nonparametric shape-constrained fitting

$$\min_{f \in \mathcal{F}} \left\{ \sum_i [y_i - f(x_i)]^2 + \frac{1}{2} \lambda \|f\|_2^2 \right\}$$

where $\mathcal{F} \triangleq \{f \in \mathcal{C}(\mathbb{R}) : f > 0, f' > 0, f'' > 0\}$ to rule out arbitrage opportunities.

The tuning parameter λ is chosen by cross-validation

Appendix: Risk-neutral marginals (4)

- Tails of the risk-neutral distributions are hard (or even impossible) to pin down exactly in a model-free manner
  - Due to the lack of very deep out-of-the-money options
  - We observe options with Delta in $[0.2, 0.8]$ most of the time
  - Assume that the implied volatilities hold constant for deep out-of-the-money options
    - They all equal the implied volatility at the nearest strike prices
  - Asymptotic justifications: let $x = \log\left(\frac{K}{S_0 R_f}\right)$ be the log-moneyness,
    - Let $d_1 = \frac{x}{\sigma(x)\sqrt{\tau}} - \frac{\sigma(x)\sqrt{\tau}}{2}$ and $d_2 = -\frac{x}{\sigma(x)\sqrt{\tau}} - \frac{\sigma(x)\sqrt{\tau}}{2}$, then

$$-\frac{1}{K\sqrt{\tau}}\frac{\Phi(d_1)}{\phi(d_1)} \leq \frac{\partial\sigma(K)}{\partial K} \leq \frac{1}{K\sqrt{\tau}}\frac{\Phi(d_2)}{\phi(d_2)}$$

Appendix: Risk-neutral marginals (5)

$$-\frac{1}{K\sqrt{\tau}}\frac{\Phi(d_1)}{\phi(d_1)} \leq \frac{\partial\sigma(K)}{\partial K} \leq \frac{1}{K\sqrt{\tau}}\frac{\Phi(d_2)}{\phi(d_2)}$$