Duration of Stock Market Crashes

Jose Antunes-Neto[*] and Ravi Jagannathan[†]

Kellogg School of Management

PRELIMINARY DRAFT

Abstract

In this paper, we investigate the episodes of stock market crashes in the United States since 1926. We define a crash as a period of time during which the stock market (measured by the CRSP Index) falls by 10% over its recent peak value. 35 episodes of crashes were identified during this period. When looking at the time to recovery for each one of these crashes, we see large heterogeneity. In this paper, we investigate the determinants of this heterogeneity. More specifically, we look to explain why some episodes take too long to recover, as others take shorter than expected. We provide economic and statistical explanations for the causes of these episodes

Keywords: Stock Market Crashes, Bubbles, Duration, Risk
JEL Codes:

[*]jose.neto@kellogg.northwestern.edu
[†]rjaganna@kellogg.northwestern.edu

1 Introduction

In this paper, we investigate the episodes of stock market crashes in the United States since 1926. We define a crash as a period of time during which the stock market (measured by the CRSP Index) falls by 10% over its recent peak value. 35 episodes of crashes were identified during this period. When looking at the time to recovery for each one of these crashes, we see large heterogeneity. In this paper, we investigate the determinants of this heterogeneity. More specifically, we look to explain why some episodes take too long to recover, as others take shorter than expected. We provide economic and statistical explanations for the causes of these episodes.

We test this by comparing with simulated data using the empirical distribution of returns. Comparing real crashes with similar sized simulated events, we see that some crashes take significantly longer to recover than what is expected. Most specifically, the crashes that follow the Great Depression of 1930's, the Dot-Com bubble and the 2008 financial crisis have a smaller recovery rate. On the other hand, a series of other crashes identified by our methodology are faster to return to their pre-crash levels. We then formulate a simple model for the crash duration which allow us to formulate testable hypothesis for the factors that are linked with longer and shorter crashes. This model is estimated both in structural (via Maximum Likelihood) and reduced form (via Ordinary Least Squares). We find that the presence of economic downturns, such as bubble bursts and downward business cycles induce longer lived crashes.

WRITE LITERATURE REVIEW Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This paper is structured as follows: Section 2 exhibits the methodology used to identify the crashes and runs a quick overview on the characteristic of some special events. It also presents the simulated study that is used to compare the observed with expected duration of such events. Section 3 describes the theoretical model for the duration of crashes and the estimation methods used, for which the data utilized in this paper is described in Section 4.Section 5 shows the main findings of this article and Section 6 concludes.

2 Methodology

Let X_t denote the process of the cumulative market log-returns (thereafter index) at time t. We define a crash as an event in which the market index fall below a certain threshold $c > 0$ relative to its previous peak:

$$\mathcal{C} := \left\{ X_t - \overline{X}_{t-1} < -c \right\} \qquad \overline{X}_t := \max_{t-\kappa \leq s \leq t} X_s$$

where the peak $\overline{X}_t$ is defined over a rolling window of $\kappa > 0$ days.

Now consider that a crash was identified at time $t = t_0$. The recovery from that crash is defined as the event in which the market returns to its peak level pre-crash, i.e.

$$\mathcal{R}_{t_0} := \left\{ X_t - \overline{X}_{t_0} \geq 0 \right\}$$

2.1 Data

For the market index, we use the daily CRSP value-weighted index from 1926 to 2024. Data is collected from Kenneth French's website. To identify the crashes, we set the threshold value to $c = 0.10$ and the lookback window to $\kappa = 252$ business days (1 year).

We identify 35 crashes in the sample period. Figure 1 displays the market index and the identified crashes in shaded areas. Table 1 shows the date the crash was first identified, where the relative peak occurred and the date the market finally recovered from that drawdown as well as the duration of the crash (in days). We see that, out of the 35 crash periods, the longest events, namely the Great Depression, the dot-com bubble, the financial crisis of 2008 are mostly associated with bubbles in the market.

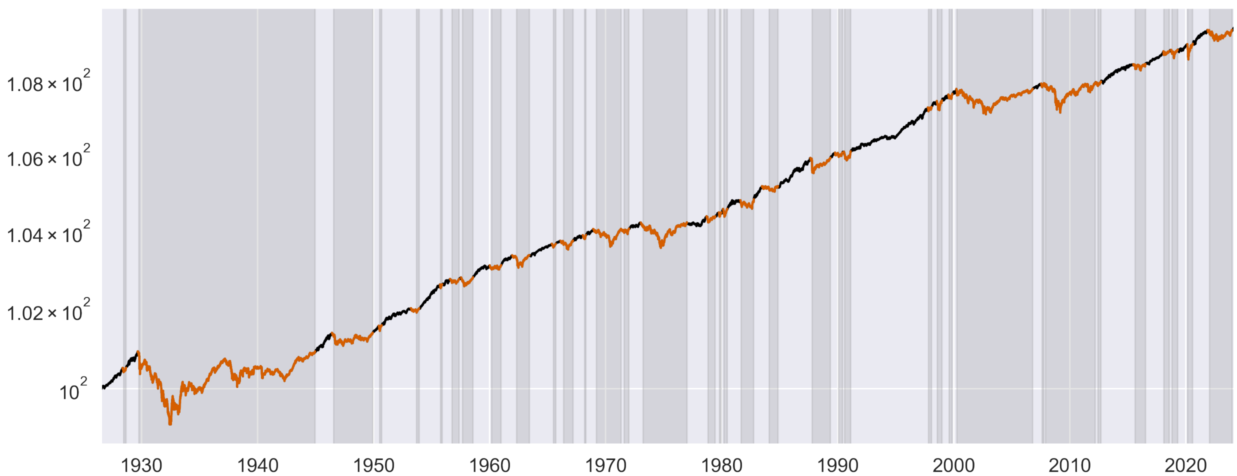

Figure 1: CRSP Index and Identified Crashes

If we compare between these crashes, we can find heterogeneities in the duration. Comparing how long did the crash last with the total loss suffered by the market before recovery, we find that bubble events, such as the dot-com bubble and the financial crisis lasted longer than similar events in terms of drawdown, such as the 1973 stock market crash. This is depicted in Figure 2, which shows the graph between total loss (defined by the return to through) and the duration of the crash (in months).

2.2 Distribution

We verify whether these crash events can be replicated by simulated returns. This tests whether the market returns display a different distribution in crash states compared to non-crash states. For this, we simulated 100.000 different paths of the same length as the real data using the empirical distribution of returns in the full sample. The large number of simulation is necessary to allow us to replicate the most severe crashes, which have close to zero probability of occurrence. This resulted into 5.213.516 crashes (around 53 crashes per simulation on average). For each one of these paths, We identify the crashes and recoveries as before . The results are shown in Figure 3.

To compare real against simulated events, I estimated the conditional expected duration of a crash given its total loss using the simulated data and fit it to the real crashes. This gives us how long the simulated paths expected the real crashes to last. Using bootstrap, I also calculate the 95% confidence interval for this estimate. The results are shown in Figure 4 and Table 2.

Table 1: Crash and Recovery Dates NAME THEM?

Peak	Crash	Recovery	Duration
1928-05-14	1928-06-12	1928-08-30	.0f
1929-09-03	1929-10-04	1944-12-16	.0f
1946-05-29	1946-07-23	1949-12-05	.0f
1950-06-12	1950-07-11	1950-09-13	.0f
1953-03-19	1953-09-14	1953-12-04	.0f
1955-09-23	1955-10-11	1955-12-01	.0f
1956-08-02	1956-10-01	1957-05-14	.0f
1957-07-15	1957-08-26	1958-07-28	.0f
1960-01-05	1960-03-07	1960-12-30	.0f
1961-12-12	1962-04-30	1963-06-03	.0f
1965-05-13	1965-06-28	1965-09-14	.0f
1966-02-09	1966-05-17	1967-03-10	.0f
1968-01-12	1968-03-05	1968-04-18	.0f
1968-11-29	1969-03-14	1971-04-28	.0f
1971-04-28	1971-08-04	1972-01-06	.0f
1973-01-11	1973-03-21	1977-01-03	.0f
1978-09-12	1978-10-23	1979-06-13	.0f
1979-10-05	1979-10-22	1979-12-18	.0f
1980-02-13	1980-03-07	1980-06-26	.0f
1981-08-11	1981-09-04	1982-10-08	.0f
1983-06-22	1984-02-08	1984-11-07	.0f
1987-08-25	1987-10-15	1989-05-16	.0f
1989-10-09	1990-01-25	1990-05-31	.0f
1990-07-16	1990-08-06	1991-02-12	.0f
1997-10-07	1997-10-27	1998-02-03	.0f
1998-07-17	1998-08-04	1998-12-24	.0f
1999-07-16	1999-08-10	1999-11-17	.0f
2000-03-24	2000-04-13	2006-10-24	.0f
2007-07-13	2007-08-15	2007-10-08	.0f
2007-10-09	2007-11-26	2012-03-14	.0f
2012-04-02	2012-06-01	2012-09-07	.0f
2015-06-23	2015-08-24	2016-07-11	.0f
2018-01-26	2018-02-08	2018-07-26	.0f
2018-09-20	2018-10-24	2019-04-24	.0f
2020-02-19	2020-02-27	2020-08-06	.0f
2021-11-16	2022-01-21	2023-12-27	.0f

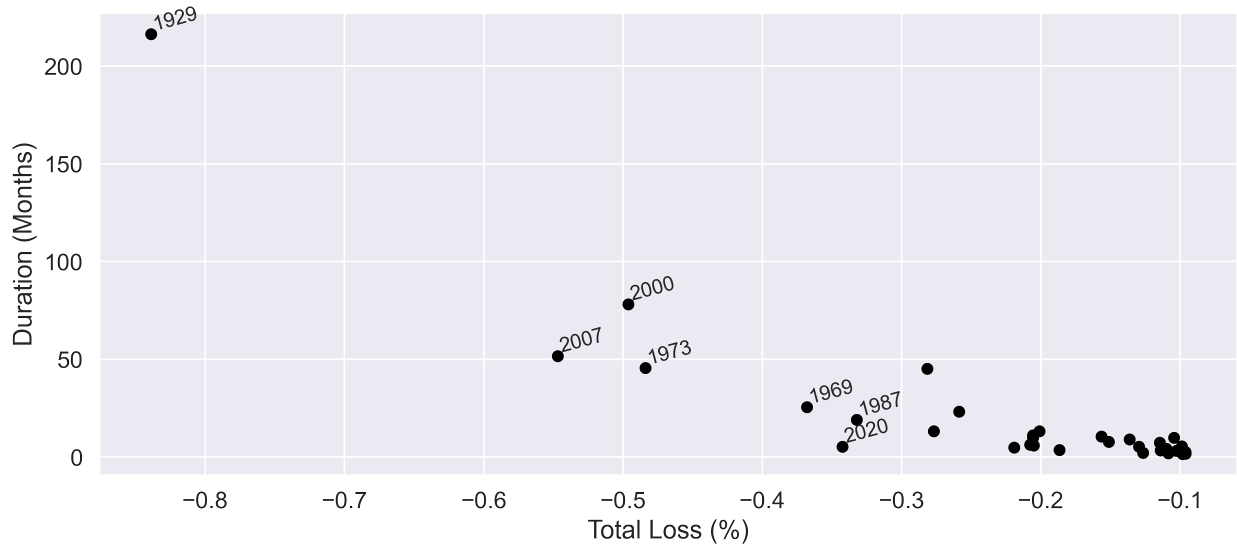

Figure 2: Duration vs. Total Loss

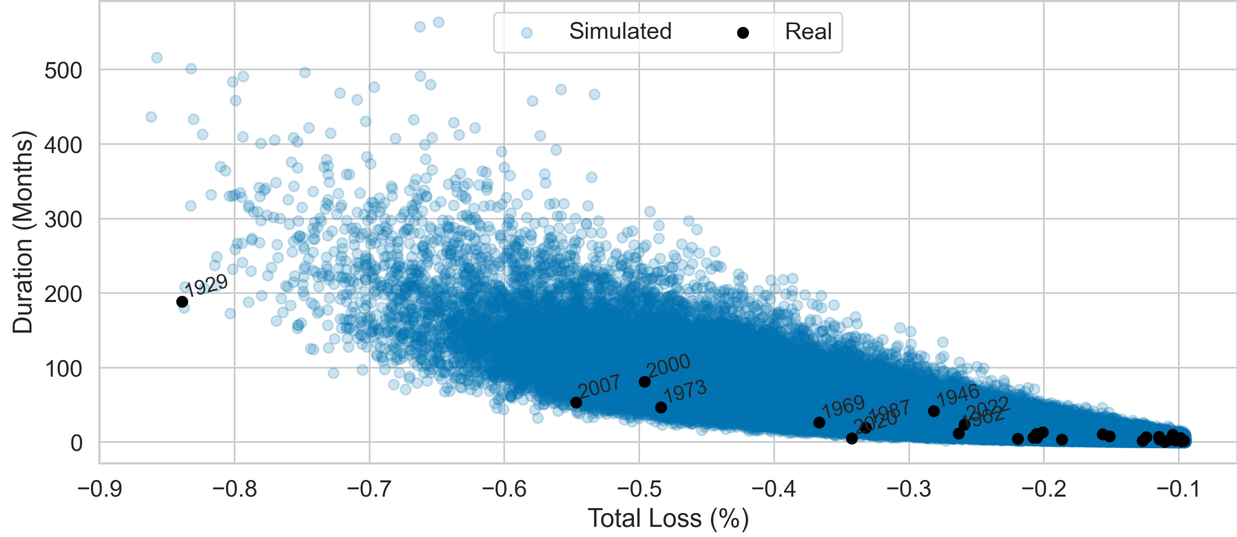

Figure 3: Duration vs. Total Loss in Simulated Data

Table 2: Simulation Fit

Drawdown	Loss to Bottom	Duration	Mean	Lower	Upper
1928-05-14	-0.11	77	192.33	192.20	192.47
1929-09-03	-1.83	3988	6486.58	6345.06	6631.19
1946-05-29	-0.33	918	312.87	312.61	313.14
1950-06-12	-0.14	66	202.01	201.87	202.15
1953-03-19	-0.11	185	194.56	194.42	194.69
1955-09-23	-0.10	48	189.99	189.86	190.12
1956-08-02	-0.12	202	197.01	196.87	197.14
1957-07-15	-0.23	269	244.81	244.62	245.00
1960-01-05	-0.11	257	193.00	192.87	193.14
1961-12-12	-0.32	383	307.25	306.99	307.52
1965-05-13	-0.10	87	189.97	189.84	190.10
1966-02-09	-0.23	281	244.87	244.69	245.06
1968-01-12	-0.10	68	190.72	190.59	190.85
1968-11-29	-0.46	627	455.52	455.06	455.99
1971-04-28	-0.14	180	203.27	203.13	203.42
1973-01-11	-0.66	1036	909.33	908.01	910.63
1978-09-12	-0.16	195	213.33	213.17	213.48
1979-10-05	-0.12	51	194.62	194.48	194.75
1980-02-13	-0.21	95	232.70	232.53	232.87
1981-08-11	-0.22	302	241.73	241.55	241.92
1983-06-22	-0.15	359	206.29	206.15	206.44
1987-08-25	-0.40	449	384.07	383.71	384.43
1989-10-09	-0.12	167	195.12	194.98	195.25
1990-07-16	-0.23	150	246.35	246.16	246.54
1997-10-07	-0.10	84	191.11	190.98	191.25
1998-07-17	-0.25	113	254.59	254.39	254.78
1999-07-16	-0.12	87	196.77	196.64	196.91
2000-03-24	-0.69	1716	987.15	985.67	988.64
2007-07-13	-0.10	60	189.97	189.84	190.10
2007-10-09	-0.79	1155	1392.45	1389.95	1395.01
2012-04-02	-0.11	113	192.02	191.89	192.16
2015-06-23	-0.17	273	215.87	215.71	216.02
2018-01-26	-0.10	128	190.92	190.79	191.06
2018-09-20	-0.23	153	244.49	244.31	244.68
2020-02-19	-0.42	120	402.57	402.18	402.96
2021-11-16	-0.30	550	288.51	288.27	288.74

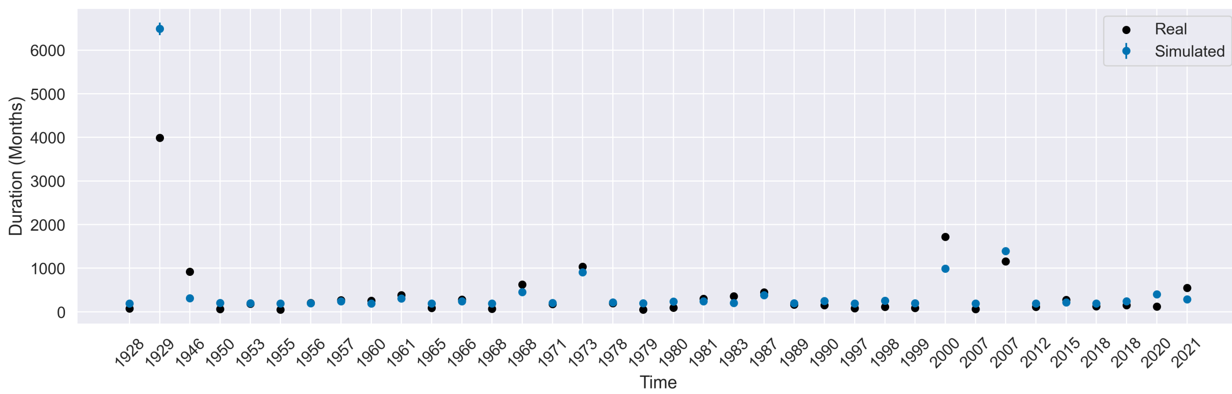

Figure 4: Real vs. Fitted Duration

3 Model

In this section, we formulate a model for the duration of crashes. We start by defining an unconditional model, which is then extended to allow for explained variables in the duration process. This model implies testable hypothesis for the factors that correlate with longer and shorter crashes and enable us to estimate the model and test for the significance of a set of proposed factors.

3.1 Unconditional Model

Let X_t denote the market security price at time t. Assume that it follows:

$$X_t = X_0 + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW_s \tag{1}$$

where W_s denotes a Standard Brownian Motion and μ and σ denote the constant drift and volatility of the price process, respectively.

Assume that a crash is detected at time $t = 0$ with $X_0 = 0$ without loss of generality. The crash is defined to last as long as the price process is below a certain threshold α relative to its initial value. Duration T then is defined as:

$$T = \min\{t : X_t \geq \alpha\} \tag{2}$$

Following Lancaster (1990), we know that the distribution of T is given by an Inverse Gaussian distribution with parameters (α, μ, σ) and probability density function given by:

$$f(t) = \frac{\alpha}{\sigma\sqrt{2\pi t^3}} \exp\left\{-\frac{(\alpha - \mu t)^2}{2\sigma^2 t}\right\} \quad \text{for } t \geq 0 \tag{3}$$

The moments of the duration distribution are given by:

$$\mathbb{E}(T) = \frac{\alpha}{\mu} \qquad \mathbb{V}(T) = \frac{\alpha\sigma^2}{\mu^3} \tag{4}$$

3.2 Conditional Model

Now assume that there exists a set of K observed factors $F = (F_j)_{j=1}^{K}$ known at $t = 0$ that determine the drift and volatility of process $\mu(F), \sigma(F)$ with a linear factor structure:

$$\mu(F) = F'\gamma \qquad \sigma(F) = F'\lambda \tag{5}$$

Assuming that ν and η are jointly normally distributed and independent of the process T, we can easily rewrite that, conditional of F, the duration process follows an Inverse Gaussian distribution with parameters $(\alpha, \gamma, \lambda)$:

$$f(t) = \frac{\alpha}{\sigma(F)\sqrt{2\pi t^3}} \exp\left\{-\frac{(\alpha - \mu(F)t)^2}{2\sigma(F)^2 t}\right\} \quad \text{for } t \geq 0 \tag{6}$$

Moreover, the moments of the duration distribution can be written as a function of the factors:

$$\mathbb{E}(T|F) = \frac{\alpha}{\mu(F)} = \frac{\alpha}{F'\gamma} \qquad \mathbb{V}(T|F) = \frac{\alpha\sigma(F)^2}{\mu(F)^3} = \frac{\alpha\,(F'\lambda)^2}{(F'\gamma)^3} \tag{7}$$

3.3 Estimation

We estimate the model in two ways. From (6), we can derive the Maximum Likelihood Estimator for the parameters $(\alpha, \gamma, \lambda)$ by maximizing the log-likelihood function:

$$\mathcal{L}(\alpha, \gamma, \lambda) = \sum_{i=1}^{N} \log f(t_i) \propto \sum_{i=1}^{N} -\log\left(F_i'\gamma\right) - \frac{3}{2}\log T_i - \frac{(\alpha - F_i'\beta T_i)^2}{2\,(F_i'\gamma)^2\, T_i} \tag{8}$$

Alternatively, the results of (7) can be used to estimate the parameters by OLS. For this, we start by noting that the CEF can be rewritten as

$$\mathbb{E}(T|F) = \frac{\alpha}{F'\gamma} = \alpha\left(\sum_{j=1}^{K} F_j\gamma_j + \sum_{j=1}^{K}\sum_{l=1}^{K} F_jF_l\gamma_j\gamma_l + \sum_{j=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} F_jF_lF_m\gamma_j\gamma_l\gamma_m + \ldots\right) \tag{9}$$

which can be approximated by a polinomial of F of sufficiently large degree p and we estimate the parameters by running OLS on the following:

$$T_i = \alpha + \sum_{j=1}^{K} F_{ij}\tilde{\gamma}_j + \sum_{j=1}^{K}\sum_{l=1}^{K} F_{ij}F_{il}\tilde{\gamma}_j\tilde{\gamma}_l + \sum_{j=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} F_{ij}F_{il}F_{im}\tilde{\gamma}_j\tilde{\gamma}_l\tilde{\gamma}_m + \cdots + \varepsilon_i \tag{10}$$

In which the error variable ε_i is heteroskedastic by definition following (7).

Finally, by performing a Taylor Expansion of the variable T, we have that

$$\begin{aligned} \mathbb{E}\log T &\approx \log\,\mathbb{E}(T) - \frac{1}{2}\,\mathbb{V}(T) \\ &\approx \log\alpha - \log F'\gamma - \frac{1}{2}\frac{\alpha\,(F'\lambda)^2}{(F'\gamma)^3} \end{aligned} \tag{11}$$

so a similar model as in (10) can be estimated for the log-duration process.

4 Data

5 Results

6 Conclusion

A Appendix

References

Robert F Engle and Jeffrey R Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.

Pablo A Guerron-Quintana, Tomohiro Hirano, and Ryo Jinnai. Bubbles, crashes, and economic growth: Theory and evidence. *American Economic Journal: Macroeconomics*, 15(2):333–371, 2023.

Tony Lancaster. *The econometric analysis of transition data*. Number 17. Cambridge university press, 1990.

Tyler Muir. Financial crises and risk premia. *The Quarterly Journal of Economics*, 132(2):765–809, 2017.

Peter CB Phillips, Shuping Shi, and Jun Yu. Testing for multiple bubbles. 2012.

Peter CB Phillips, Shuping Shi, and Jun Yu. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International economic review*, 56(4):1043–1078, 2015a.

Peter CB Phillips, Shuping Shi, and Jun Yu. Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4):1079–1134, 2015b.