What can inform us about the arrival of daily extreme stock market events?

Vasiliki Chatzikonstanti[1]
Michail Karoglou[2],
Christian Wolff[3]
Sotiria Xanalatou[4]

Abstract

We examine what accelerates and decelerates the arrival of extreme stock markets events at monthly frequency for the US stock market. By modeling the duration of an extreme stock market event as a time-to-event counting process, we make use of all S&P500 shares over the last fifty years to examine how key financial indicators and the principal macroeconomic variables can capture whether, and if so to what extent, the likelihood of occurrence of such events rising or falling. Our findings suggest that there are notable differences between positive and negative stock market events and that macroeconomic conditions are clearly much more unambiguous in indicating whether the likelihood of such events rises or falls.

Keywords: black swans, S&P500, survival analysis, extreme market events

JEL Classification:

[1] Athens University of Economics and Business, Athens, 11362, Greece; email vchatzikonstanti@aueb.gr
[2] Aston Business School, Birmingham, B4 7ET, UK; email m.karoglou@aston.ac.uk
[3] University of Luxembourg, Luxembourg, L-272, Luxembourg; email christian.wolff@uni.lu
[4] University of Luxembourg, Luxembourg, L-272, Luxembourg; email xanalatousotiria@gmail.com

Introduction

If there was some standardized history of a stock market, it would indisputably contain a very lengthy chapter dedicated to its crashes and other extreme events. It is no surprise therefore that both the institutional and individual investors and, in turn, the academic community, the mainstream press and regulatory bodies have an unsaturated desire to learn more about them. And they do so irrespective of whether their underlying goal is to determine what causes or simply when they will take place either to profit from their occurrence or to insure against them or to somehow prevent them from happening in the first place.

As one would expect, the underlying literature is immense. However, a crude albeit convenient, for our purposes, way to steer through it is to consider the research focus. In this spirit, the overwhelming majority of the research in this literature is dedicated to examining individual dramatic events, involving a specific large firm, a major institution, an entire market, a whole economy or even a broader economic system. The other surprisingly more modest part of the research in this literature is dedicated to examining these extreme stock market events as an integral part of the ever-eluding stochastic process of the stock market returns. In it, the notorious black swan events, also the inspiration behind the proliferation of the so-called black swan funds, are at the centerstage.

In this paper we examine if, and to what extent, some of the widely available financial indicators of firms and principal macroeconomic variables can be used to determine the likelihood of daily extreme stock market events in the short and longer investment horizon. Using all the available data for the shares included in the S&P500 from the introduction of the index, we adopt survival analysis techniques with time-varying regressors to explore what contributes to, and to what extent, to the arrival of an extreme stock market event. Our findings suggest that (i) there are three distinct periods that characterize the probability of the arrival of an extreme stock market event; (ii) that these periods are persistent across the different industries; (iii) that firm specific factors seem to be much less relevant into letting us know as to whether the probability of a next extreme stock market event taking place is higher or lower in sharp contrast to the macro-variables, and especially economic growth and inflation. Finally, we also find substantial differences between positive and negative extreme stock market events.

The remainder of this paper is as follows. Section 2 positions our work in the existing literature and provides the key theoretical underpinnings of our investigation. Section 3 discusses our data and methodology. Section 4 presents empirical results and Section 5 their discussion. Finally, Section 6 concludes our study.

Theoretical Underpinnings

The literature on extreme stock market events is unsurprisingly voluminous. Although it is overwhelmingly focused on negative extreme events such as stock market crashes and financial crises, a plethora of statistical and econometric models developed with the objective to foresee the latter. The so-called Early Warning Systems (hereafter EWS) literature contains a remarkably broad set of methods developed for exactly this purpose. Nevertheless, the performance of such methods so far suggest that we are still far from establishing a generally satisfactory approach (see for example Berg et al 2005 and references therein). The 2007/8 financial crisis and the ensuing sovereign debt crisis have demonstrated beyond any doubt that a broader set of methods and techniques are imperative to be considered.

Following Dastkhan (2019), we overview the existing EWS literature by splitting it into two categories. The first category is agnostic about the underlying stochastic processes that generate extreme events. The second category tries instead to pin-point explicit features of these stochastic processes with the intention to using them for predictions.

Specifically, the first category draws on the dramatic advances in computation power and is founded upon the use of high-dimensional non-linear quantitative methods, such as artificial neural network. For example, Yoon and Park (2014) applied pattern recognition techniques to introduce a stock market instability index while Döpke et al (2017) used boosted regression trees to re-examine the usefulness of selected leading indicators in predicting recessions and Dastkhan (2019) used a network representation of assets, based on the values of the forward-looking conditional value-at-risk (CoVaR) model. More relevant to our investigation is the work of Chatzis et al. (2018) who used extensive variations of deep learning algorithms to assess their performance in forecasting tail events in the global stock markets. Likewise, Samitas et al (2020) combined network analysis and machine-learning algorithms to predict the possibility of contagion risk during shock events and crisis periods in stock exchange markets. The key characteristic of this

strand of literature is twofold: first, apart from the importance, it acknowledges the urgency of predicting extreme events in financial markets or whole economies, and, second, it allows for the treatment of the underlying processes that bring forth an extreme event as a 'black box.'

The second category goes to the opposite direction and contains the more voluminous and established literature typically consisting of studies that make use of a specific econometric technique. It contains studies that mainly belong in either the so-called signals modeling approach or the limited dependent variable modelling approach (typically some probit/logit variant). The key characteristic of these strands of literature is that it attempts to avoid treating extreme events as black boxes; but, in order to establish a stable relation between indicators and crises, it is forced to make quite rigid modelling assumptions to address the inherent endogeneity problem that stems from the fact that the behavior of the indicator variables is affected both by the crisis itself and the policies undertaken so as to mitigate it.

The signals approach, popularized by Kaminsky et al (1998), Kaminsky and Reinhart (1999) and Goldstein, Kaminsky et al. (2000), is at its core a non-parametric threshold approach that aims to identify thresholds for certain vulnerability indicators. The latter indicators are transformed into binary signals of financial or economic distress and the level of thresholds is determined based on the minimization of the noise-to-signal ratio. Various such indicators can be then combined to form a composite indicator (Kaminsky, 1999). This strand of literature contains a plethora of applications and extensions such as those of Brüggemann and Linne (2002), Edison (2003) and El-Shagi et al. (2013). More relevant to our work is the work of Alessi and Detken (2011) who used a quasi-real time signaling approach to predict asset price booms.

In the same category belong the binary choice models that attempt to identify a stable relationship between a set of indicators and a binary crisis variable. This relationship is then used for forecasting the crisis variable. Notable examples of this approach that uses logit models include Frankel and Rose (1996), Berg and Pattillo (1999), Eichengreen et al. (1995), Kumar et al (2003) as well as Duca and Peltonen (2013) who developed a framework for predicting periods of financial instability. Bussière and Fratzscher (2006) adopt a multinomial logit model to extend the binary choice from two to more states, namely crisis, post-crisis, and tranquil periods in order to enhance the efficiency of the

EWS and, more relevant to this work, so do Li et al (2015) who built up an EWS for equity market crises.

Other EWS studies that belong in the same category make use of some other econometric model. For example, Abiad (2003) predicts a crisis based on Markov-switching models with time-varying transition probabilities to bypass the a priori dating of each crisis episode that is inherent in the other approaches. Closer to our work, Faranda et al. (2015) identify early warning indicators of financial crises as deviation in the ARMA space from a reference model. Likewise, Gresnigt et al (2015) determine the probability of the occurrence of an extreme stock market event within a certain time period through a model designed initially for earthquake sequences, the so-called Epidemic-Type Aftershock Sequence model (ETAS) – treating, in effect, the period around a financial market crash as analogous to the seismic activity around earthquakes.

Beyond the scope of the Early Warning Systems literature, that we hereby presented, we also view our study as one that relates to a series of seminal studies in disaster risk. In his seminal work on *Rare disasters and asset markets*, Barro (2006) models economic disasters as downward jumps in per capita gross domestic product (GDP). In particular, Barro (2006) characterizes economic disasters by steep declines of at least 15% in real per capita GDP for twenty O.E.C.D. countries[5]. We are inspired by the work of Barro (2006) in terms of his approach of the probability distribution of disaster (rare) events. The measure by which he implies observed probabilities of economic disasters basically renders instants of time of those disasters consecutive, rather than fixed. As he addresses GDP sharp declines based on the greater war periods in the 20th century, Barro (2006) shifts disaster risk (modeling) attention to a probability of entering a particular state of the world, relative to the individual dramatic event (GDP sharp declines). Our duration modeling approach exploits sharp declines on market returns (market crashes), so as to assess the predictive power of factors relative to the probability of entering a rare event state of the world. The latter state of the world is defined by individual (daily) dramatic events at both tails of the distribution of stock returns. Entrance to such a state state of the world stems from survivor and hazard rate functions (presented in our Methodology section). By analogy, the time spells of duration analysis are the main object, by which we observe the predicted time to surviving a current state of the world, and consequently, entering another. The interpretation of time spells/durations is, actually, the time-to-default.

[5] O.E.C.D.: Organization for Economic Cooperation and Development

Also, Barro (2006) calibration of the disaster parameter, that relates to the probability of a default, is contingent upon the occurrence of a disaster. Similarly, our model is one that assumes that, the probability of observing a market crash is contingent upon the rate at which a time spell is completed. In other words, that the probability of observing a market crash is contingent to the occurrence of past market crashes. Moreover, Barro and Ursúa (2009) examine conditional probabilities of dramatic events of large consumption declines given the historical information set of past stock market crashes. Their predictions are based on cumulative return drops of 25% and average declines in consumption, through the lens of 100 macroeconomic contractions. What they highlight, is the change in the conditional probability of a transition to a state, that satisfies particular conditions of a macroeconomic contraction, given past stock market rare events. Our duration data approach directly targets at future market crashes prediction, by looking at conditional probabilities of each (time unit) observation. Thus, we assume the relative importance of the macro economy critical for our model.

Wachter (2013) extends Barro's constant disaster probability and power utility model, by introducing rather stochastic probability, and recursive preferences. The model allows for the intensity of jumps to follow a square root process (Cox et al., 1985), and, while accommodating reasonable risk aversion levels, it explains the quantitative puzzle on equities returns (Mehra and Prescott, 1985). Implied disaster probabilities of the model, as well as the matching of stock returns volatility to excess stock returns volatility, is based on a variation of the price-dividend ratio (dividend yield). Our work can also be seen as part of an empirical counter-exercise of the work of Wachter (2013) in the following way: we address the question of how well do prominent in existing literature financial and macroeconomic indicators perform in predicting the intensity (probability) of negative and positive stock market rare events. The latter are widely known as *Black Swans* [6]. While we rely on conditional probabilities of entering an analogue of a market crash state of the world based on past observations, we shift attention to identifying the factors that impact those conditional probabilities, such that they fit historical data. Our dependent variable is the intensity function, that reflects the imminence of an extreme market event, and, thus

[6]In his book "The Black Swans: the impact of the highly improbable", Taleb (2008) discusses that *Black Swans* are mainly characterized by three elements: their probability of occurring is unknown, their impact is largely significant and, by consequence devastating, and they can be foreseen in hindsight.

the probability of a market extreme event[7]. Our independent variables, financial and macroeconomic, consist of the explanatory variables assigned to the model. In other words, our study is an empirical exercise, whereby, financial indicators and proxies of the macroeconomic environment, can adequately predict a measure of rare events. The latter measure is by default constructed to address empirical data over consecutive time intervals.

Further recent work on the link between rare disaster risk and market performance, also accounts for the disaster probability, through the scope of transition probabilities from one state of the world, to another. Berkman et al. (2011) show that a conditional measure of rare disaster probability correlates with the stock market, and drives stock returns. The financial indicators that are shown by Berkman et al. (2011) governing disaster risk, relate to stock market performance, and are summarized to the earnings-price ratio and the dividend yield. In addition, in their "horse race" estimation exercise among prominent financial forecasting variables and their own dynamic tail risk measure, Kelly and Jiang (2014) also find that the (aggregate) dividend-price ratio succeeds on predicting future stock market returns[8]. Relying, thus, on evidence of prominent predictors in existing literature studies, we have selected and incorporated in our model both financial and macroeconomic predictor variables. Our selection is based on the results and correlations of the above-mentioned variables in extreme events prediction. In order to simplify the terminology used in our results and analysis, and following Taleb (2008) terminology, we hereafter refer to positive and negative extreme returns values as *positive Black Swans* and *negative Black Swans*, respectively.

Overall, this study conducts an empirical investigation, to determine what factors can be used to explain the duration of the period, measured in number of days, between two consecutive Black Swans. The latter is done for all individual shares of the S&P500 index at daily frequency. We collect the periods, that in the context of survival analysis are denoted as ‘runs’, from every single individual share included in the S&P500. Discussion at sectoral level is based upon the grouping of these constituent shares, according to the industry they belong to.

[7]Details on the dependent variable of our duration model follow in the Methodology Section.

[8]Market returns in the study of Kelly and Jiang (2014) refer to CRSP value-weighted market index returns over one-month.

Data

Our dataset comprises of the monthly stock market (log-) returns of all constituent shares included in the Standard and Poor's 500 (S&P500) stock market index which tracking the stock performance of 500 large companies listed on stock exchanges in the United States and spans the period January 1973 until February 2019 (563 months for each company) tentatively earlier than the dramatic events around the Covid-19 pandemic. Naturally, over all these decades, the companies included in the index have changed and therefore for consistency we have decided to include in our analysis all the data of the companies that were once included in the index over our sample period (533 firms in total). For each of these shares we have also collected the respective price-to-earnings ratio, the price-to-book ratio and the dividend yield all key financial indicators that are used extensively in static and dynamic asset valuation models. Finally, we also make use of the principal macroeconomic variables that form the basis of all modern neo-Keynesian dynamic stochastic general equilibrium (DSGE) models that make use of monthly data namely the industrial production index growth, the predominant proxy of the monthly GDP growth, inflation, and the federal funds interest rate.[9]

Table 1 provides an overview of the aggregate properties of the variables in our dataset.

Insert Table 1 around here

Methodology

We argue that the most direct way to determine what contributes, and to what extent, towards the arrival of an extreme stock market event is to consider the dual problem of determining what contributes, and to what extent, towards lengthening or shortening the period between two successive such events. To this end, we model the duration of the period until a next extreme event (i.e. the period since the last extreme event) in days as a time-to-event counting process, which in turn enables us to draw our inferences by applying survival analysis methods. Given that, to our knowledge, this is not only the first

[9] All data were collected from Refinitiv's Datastream database.

time that this aspect of extreme stock market events is examined but also that this is the first time that this modelling approach is introduced in this strand of literature, in what follows we provide a comprehensive set of details to explain it.

In particular, by treating the period between two successive extreme events as a random variable, which we denote as *T*, the probability of such a period continuing without an extreme event until time *t* is modelled through the survival function $S(t)= S(T=t) = P(T>t) = 1 - F(T\leq 0)$ where *F(.)* is the cumulative distribution function. Consequently, a survival analysis involves examining the properties of *S(t)* and the factors that determine it.

Typically, empirical investigations that make use of survival analysis begin with a visual depiction of the non-parametric product-limit estimator of *S(t)*, more commonly known as the Kaplan-Meier or KM method (Kaplan and Meier, 1958). This is often combined with reporting the results of one or more proportional hazards regressions, generally known as Cox regressions, Cox proportional hazard model or the *Cox model* which was first conceived by Cox (1979). This is a powerful semi-parametric approach that, unlike the univariate KM method, enables us to incorporate covariates into the survival function and estimate and test for their possible impact. In fact, this method, and its later incarnations, proved so successful that in many disciplines made it almost synonymous to survival analysis.

A significant limitation of the mainstream methods is the fact that they are based on time-invariant covariates. However, in some settings, such as in our case, it is essential for some or all covariates to be allowed to vary over time. To address this, alternative estimators need to be considered.

The most popular non-parametric, and asymptotically equivalent with the KM, estimator is the multiple decrements estimator that was first conceived by Allen (1978), and which is currently known as *the Nelson-Aalen* or NA estimator, of *S(t)* which is given by:

$$\widetilde{S}(t)=e^{-\widetilde{H}(t)} \tag{1}$$

where the integrated hazard or cumulative hazard rate estimator is given by:

$$\widetilde{H}(t)=\sum_{t_i\leq t}\frac{d_i}{n_i} \tag{2}$$

where d_i is the number of fund managers that change posts at time t_i, and n_i is the total number of extreme stock return events at time t_i.

In the case of the semi-parametric approach, the Cox model is expanded to accommodate time-varying covariates and Andersen and Gill (1982) were the first to offer us estimates of the survival function conditional on x_i covariates:

$$S(t \vee x_1, x_2, \ldots) = S_0(t)^{\exp(\beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots)} \quad (3)$$

Eq. (3) is conveniently expressed through the so-called hazard ratio as:

$$\frac{h(t \vee x_1, x_2, \ldots)}{h_0(t)} = \exp(\beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots) \quad (4)$$

and, estimated specifically by using the natural logs of eq. (4):

$$\ln \frac{h(t \vee x_1, x_2, \ldots)}{h_0(t)} = \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots \quad (5)$$

where *h(t)* is the hazard function which captures the probability that the next extreme stock market event will take place given that it has not happened up to time *t*. The hazard function is linked to the survival function through the identity *h(t) = f(t) / S(t)* where *f(t)* is the probability density function of the *t* random variable.

The hazard ratio can be extended to the case of time-varying covariates, since the partial likelihood is built by considering only what happens at each 'failure time' (i.e., the time of an extreme stock market event taking place) and therefore, at the limit, it can account for every single time period (in our case, day). This is why, even from its conception, this extension, which bridges the gap between survival analysis and counting processes, has been used to test the so-called assumption of proportional hazards.

Therefore, the only remaining issue that is left for us to determine is; and the explicit definition of what we denote as an extreme stock market event.

In terms of the covariates (*x*'s in the above equations) that might accelerate or delay the arrival of an extreme stock market event, in general, we consider firm specific factors and factors that capture the contemporary state of the macroeconomy. The former includes the dividend yield as well as the price-to-book and price-to-earnings ratios as well as the stock return of the respective share to control for the possible effect of the magnitude and size of the extreme stock market event itself. The latter includes the values of industrial production growth, inflation rate and federal funds rate. Consequently, the main semi-parametric model we estimate is given by:

$$\ln \frac{h(t | x_1, x_2, \ldots)}{h_0(t)} = f(\textit{firm specific factors}, \textit{state of the macroeconomy})$$

where *f(.)* is a linear function of the variables discussed above. The exponent of the coefficients of these variables shows whether, and to what extent, does each variable accelerate (hazard ratio greater than 1) or delay (hazard ratio less than 1) the arrival of an extreme stock market event.

Finally, with respect to how we determine extreme stock market events we make use of the customary practice of classifying them as the *3-sigma* events, also known as black swans.[10] In other words, we consider a stock return observation as an extreme stock market event if its standardized unexpected return value is beyond three standard deviations above and below the (unconditional) mean of the respective series.[11]

Empirical results

In this section, we first present the results from the non-parametric approach and then the ones from the semi-parametric approach.

The non-parametric analysis

Figure 1 presents the *Nelson-Allen* estimates for the duration until the next extreme stock market event takes place over time.

Insert Figure 1 around here

As expected, and we can visually observe, the probability of not having an extreme stock market event decreases over time. However, this probability is far from being linear and, most interestingly, can be vividly distinguished into three periods. The first, that contains the observations of approximately the first month after the last extreme stock market event, is the shortest and steepest period with the probability of having no extreme stock market event dropping at its conclusion to less than 50%. The second period lasts for a bit more than two extra quarters and is a rather moderate with the probability of having no extreme stock market event reaching a bit more than 5% in the end. In other words, our

[10] Chatzikonstanti and Karoglou (2022) explore in detail the concept and its distinction from and relation with stock market crashes, structural changes, extreme values, and outliers. For our purposes it is worth noting that they can be considered as a stricter subset of extreme values or tail events.
[11] For robustness, we have also explored another known candidate, the *6-sigma* events, but our inference has remained unchanged.

data suggests that the probability of having two successive extreme stock market events within a period of about eight months is about 95%. The final period is almost flat.

Figure 2 presents the same *Nelson-Allen* estimates but also makes a distinction across the different industries.

Insert Figure 2 around here

What is particularly noteworthy is that despite some small discrepancies across industries, the overall structure is surprisingly similar. Nine out of the forty sectors have lower survival curves (Gas, Water and Multi-utilities, the Health Care Providers, Investment Banking & Brokerage services, Life Insurance, Non-life Insurance, Software and Computer Services, Real Estate Investment Trusts, Telecommunications Service Providers, and Travel and Leisure), which suggests that they exhibit extreme stock market events more often than the other sectors; and fourteen out of the forty sectors have higher survival curves (Beverages, Consumer Services, Electricity, Finance and Credit Services, Food Producers, General Industrials, Industrial Engineering, Industrial Transportation, Media, Oil, Gas and Coal, Personal Care, Drug and Grocery Stores, Personal Goods, and Retailers), which suggests that they exhibit extreme stock market events less often than the other sectors. However, the three periods structure we have discussed before is clearly present in all cases, and, most importantly, the duration of each of these periods is uncannily similar across almost all industries.

Figure 3 takes the non-parametric analysis a step further and adds the distinction between positive and negative extreme stock market events.

Insert Figure 3 around here

The graph depicts a clear pattern: the probability of having a positive extreme market events is substantially smaller than the probability of having a negative one. In other words, negative tail events are much more likely to occur than positive ones. This could effectively be one of the reasons of the well-documented small albeit statistically significant skewness in the US stock market returns.

The semi-parametric analysis

Table 2 presents the results of our semi-parametric approach for our main model. What we can observe is that a particularly striking feature of the results is that, apart from

the stock return, the firm specific factors are found marginally significant, in sharp contrast to macrovariables that capture the state of the economy.

Insert Table 2 around here

In terms of the firm specific factors, the size and magnitude of the return are clearly linked to the probability of an extreme stock market event taking place but, at this level of analysis, this appears to be simply capturing the fact that the more extreme observations are much less likely to occur than the less extreme ones. Also, the utilization of the net assets of a firm (Price-to-Book ratio) seems to only marginally be linked to a higher chance of an extreme stock market event for the corresponding firm. The effect of the dividend yield is similar but more pronounced, and statistically significant. In contrast, a higher (lower) market value of a firm with respect to its earnings (Price-to-Earnings ratio) is associated to a lower (higher) chance of an extreme stock market event but only at 10% significance level.

In terms of the factors that determine the state of the macroeconomy, the growth series demonstrates clearly the most dramatic effect. Indeed, positive (negative) growth means substantially smaller (larger) probability of an extreme stock market event. The interest rate has also a similar effect although much more moderate. In contrast, higher (lower) inflation is linked to larger (smaller) probability of an extreme stock market event.

Table 3 presents the results of our semi-parametric approach if in our main model we make a distinction between positive and negative stock market events by introducing a positive stock market event dummy and interacting it with the other variables. What we can observe is that the previous results are very similar to the new ones; but there are substantial differences in terms of magnitude of the effect between positive and negative extreme stock markets events.

Insert Table 3 around here

In terms of the firm specific factors, the size and magnitude of the return are as before clearly linked to the probability of an extreme stock market event taking place and from the corresponding interaction term we can see that more extreme positive stock market events take place faster than negative ones. This seems to be linked to the fact that positive

events are on average less extreme than the negative ones, which is also indicated when we look at the positive hazard ratio of the dummy on its own, and therefore serves as a partial explanation of the well-established negative skewness in the US stock returns. Also, the utilization of the net assets of a firm (Price-to-Book ratio) seems to be linked only to a higher chance of a positive extreme stock market event for the corresponding firm taking place but although statistically significant its effect is very small. The marginally significant effect of the dividend yield has completely disappeared this time while a higher (lower) market value of a firm with respect to its earnings (Price-to-Earnings ratio) is associated to a lower (higher) chance of an extreme stock market event but only for the positive events.

In terms of the factors that determine the macroeconomic conditions they are almost identical to those that we found in the main model. What we also find is that the effect of inflation is reversed when it is about positive extreme stock market events, although the statistical evidence is significant only at 10% level. Also, the effect of the interest rate is more moderate for positive events.

Overall, the semi-parametric analysis provides evidence to support that the state of the macroeconomy is a key determinant of the probability of a stock market event taking place. Despite the substantial variation across the different macrovariables, since economic growth, primarily, and, much more modestly, interest rates seem to have the opposite effect than inflation, the results are very explicit. In contrast, firm specific variables seem to matter very little and primarily for positive extreme stock market events. Therefore, we can conclude that based on our main model the probability of an extreme stock market event taking place is clearly affected by the state of the macroeconomy.

Discussion

Conclusions

References

Abiad, A., (2003). Early Warning Systems: A Survey and a Regime-Switching Approach. IMF Working Paper 32.

Alessi, L., & Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. European Journal of Political Economy, 27, 520-533.

Barro, R. (2006). Rare disasters and asset markets in the twentieth century. The Quarterly Journal of Economics 121.(3), 823–866.

Barro, R. J. and J. F. Ursúa (2009). Rare disasters, asset prices and welfare costs. American Economic Review 99.(1), 243–264.

Berg, A., Borensztein, E., Pattillo, C. (2005). Assessing early warning systems: how have they worked in practice? IMF Staff Paper 52,462–502.

Berg, A., Pattillo, C., (1999). Predicting currency crises: the indicators approach and an alternative. Journal of International Money and Finance 18, 561–586.

Berkman, H., B. Jacobsen, and J. B. Lee (2011). Time-varying rare disaster risk and stock returns. Journal of Financial Economics 101, 313–332.

Brüggemann, A., Linne, T., (2002). Are the Central and Eastern European Transition Countries Still Vulnerable to a Financial Crisis? Results from the Signals Approach. IWH Discussion Paper 157.

Bussière, M., Fratzscher, M., (2006). Towards a new early warning system of financial crises. Journal of International Money and Finance 25, 953–973.

Chatzis, S., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques, Expert systems with applications 112, 353-371.

Cox, J. C., J. C. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. Econometrica 53, 385–408.

Dastkhan, H. (2019). Network based early warning system to predict financial crisis. International Journal of Finance & Economics, 1-23.

Döpke, J. , Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. International Journal of Forecasting, 33 745–759.

Duca, M. L., & Peltonen, T. A. (2013). Assessing systemic risks and predicting systemic events. Journal of Banking & Finance 37, 2183-2195.

Edison, H., (2003). Do indicators of financial crises work? An evaluation of an early warning system. International Journal of Finance and Economics 8, 11–53.

Eichengreen, Barry, Andrew Rose and Charles Wyplosz, 1995, “Exchange Market Mayhem: The Antecedents and Aftermaths of Speculative Attacks”, Economic Policy, 21, pp. 249-312.

Faranda, D., Flavio, M. , Giachino, E. , Vaienti, S. , & Dubrulle, B. (2015). Early warnings indicators of financial crises via auto regressive moving average models. Communications in Nonlinear Science and Numerical Simulation 29, 233–239.

Frankel, J.A., Rose, A.K., (1996). Currency crashes in emerging markets: an empirical treatment. Journal of International Economics 41, 351–366.

Goldstein, Morris, Graziela L. Kaminsky and Carmen M. Reinhart, Assessing Financial Vulnerability, An Early Warning System for Emerging Markets, June 2000, Institute for International Economics.

Gresnigt, F., Kole, E., & Franses, P. H. (2015). Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. Journal of Banking & Finance 56, 123-139.

Kaminsky, G.L., (1999). Currency and Banking Crises: The Early Warnings of Distress. IMF Working Paper 99/178.

Kaminsky, G.L., Lizondo, S., Reinhart, C.M., (1998). Leading indicators of currency crises. IMF Staff Papers 45, 1–48.

Kaminsky, G.L., Reinhart, C.M., (1999). The Twin crises: the causes of banking and Balance-of-Payments problems. American Economic Review 89 (3), 473–500.

Kelly, B. and H. Jiang (2014). Tail risk and asset prices. The Review of Financial Studies 27.(10), 2841–2871.

Kumar, M., Moorthy, U., Perraudin, W., (2003). Predicting emerging market currency crashes. Journal of Empirical Finance 10, 427–454.

Li, W. X., Chen, C. C. S., & French, J. J. (2015). Toward an early warning system of financial crises: What can index futures and options tell us?. The Quarterly Review of Economics and Finance, 55 87-99.

Mehra, R. and E. Prescott (1985). The equity premium puzzle. The Journal of Monetary Economics 15, 145–161.

Samitas, A., Kampouris, E., & Kenourgios, D. (2020). Machine learning as an early warning system to predict financial crisis. International Review of Financial Analysis 71, 101-507.

Schoenfeld, D. (1982). “Partial residuals for the proportional hazards regression model”. In:Biometrika 69.1, pp. 239–241.

Taleb, N. N. (2008). The Black Swan: the impact of the highly improbable. Penguin.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market
volatility ? Journal of Finance 68.(3), 987–1035.

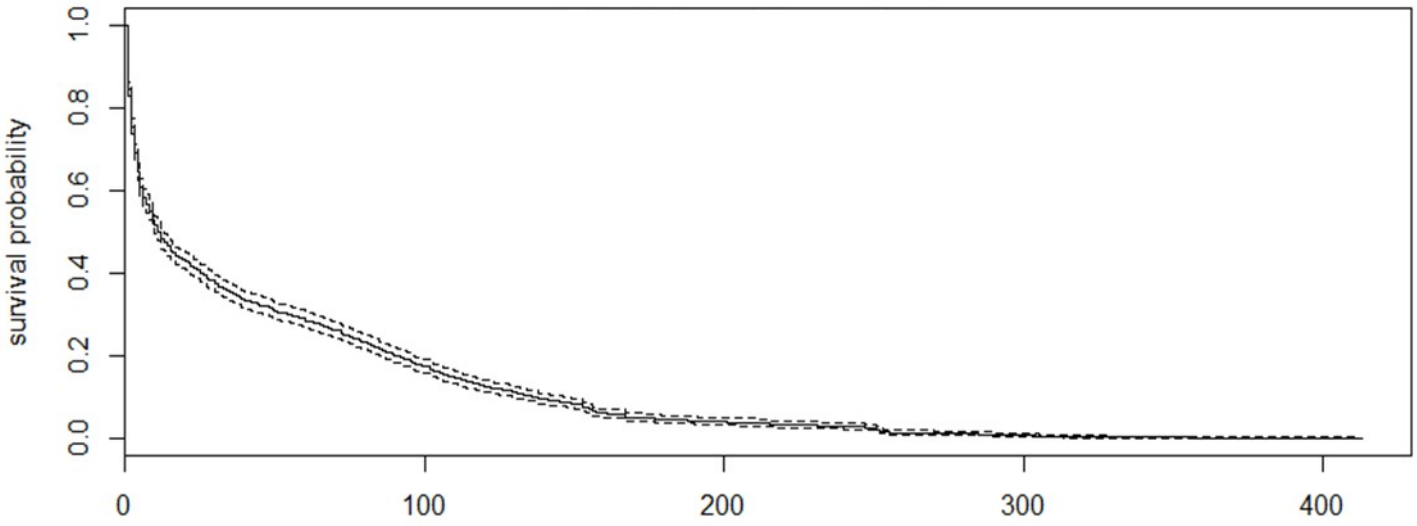
Yoon, W., & Park, K. (2014). A study on the market instability index and risk warning levels in early warning system for economic crisis. Digital Signal Processing, 29, 35–44.

Tables and Figures

Table 1. Summary statistics

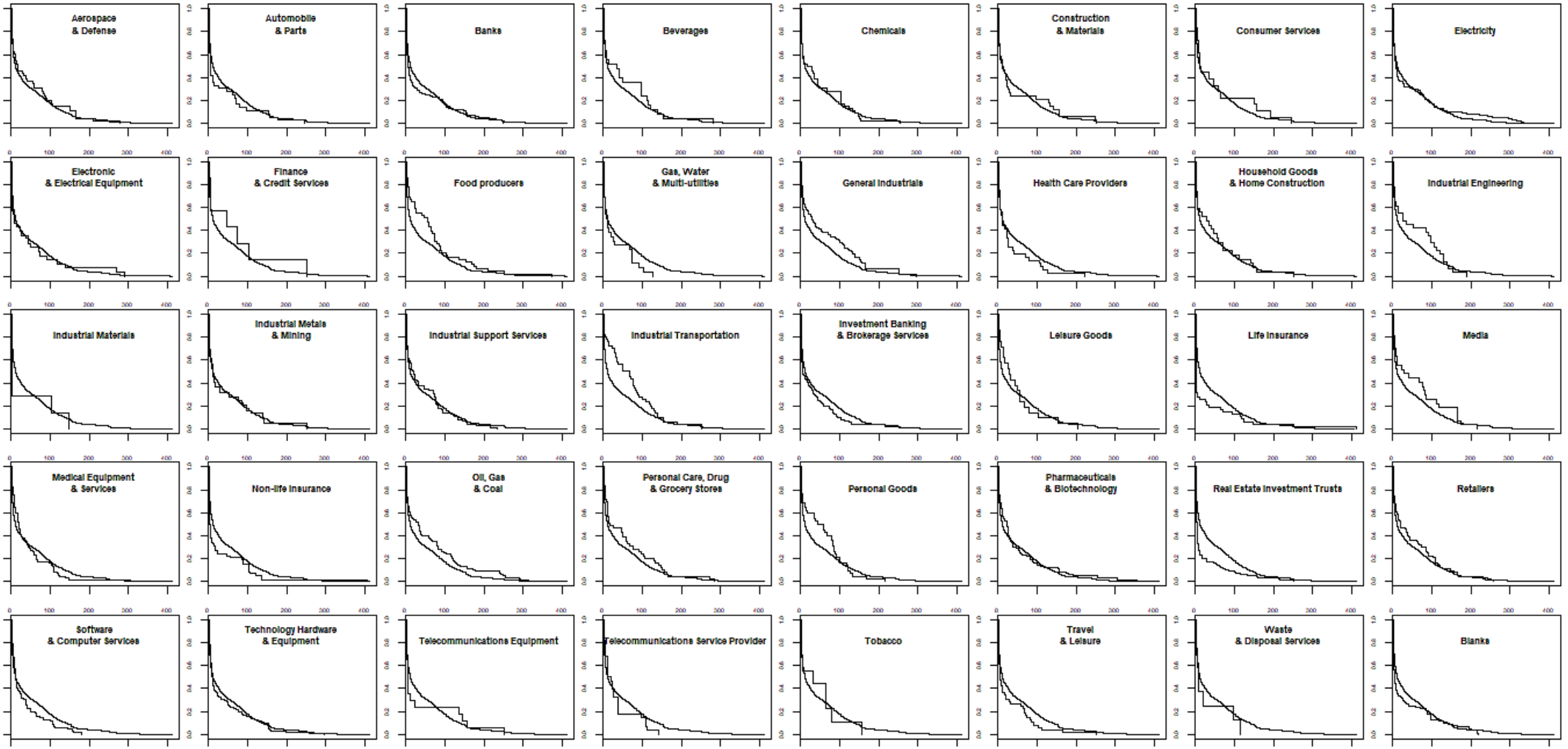
	Mean	Std.Dev.	Skewness	Kurtosis
Stock returns	0.01	0.10	-0.09	37.13
Price-to-book value	2.79	36.91	-33.22	2090.4
Dividend Yield	0.31	0.39	-0.9	4.82
Price-to-earnings	1.28	0.32	1.49	14.79
Growth	0.001	0.003	-1.35	9.53
Inflation	-0.001	0.16	0.01	5.64
Interest rate	4.8	3.91	0.87	3.84

Figure 1: N-A estimate for the duration until the next extreme stock market event



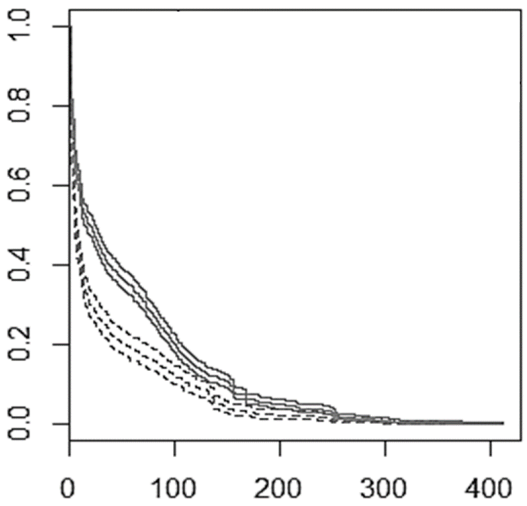
Notes: This figure depicts the probabilities of the number of days that will take until the next extreme stock market event will take place as determined by the non-parametric Nelson-Allen estimates. The dotted lines indicate the 95% confidence interval.

Figure 2: N-A estimate for the duration until the next extreme stock market event for each industry and for the whole market



Notes: This figure depicts the probabilities of the number of days that will take until the next extreme stock market event of each industry will take place as determined by the non-parametric Nelson-Allen estimates. For comparison, it also includes the respective curve for the overall market as depicted in Figure 1.

Figure 3: Distinguishing extreme market events



Notes: This figure depicts the probabilities and 95% confidence intervals of the number of days that will take until the next negative (grey lines) and positive (dotted lines) extreme stock market event will take place as determined by the non-parametric Nelson-Allen estimates.

Table 2. Estimation results of main model

	Coeff.	Std.Error	Hazard ratios
Stock returns	-3.10***	0.24	0.05
Price-to-book value	0.002*	0.001	1.002
Dividend Yield	0.17**	0.08	1.19
Price-to-earnings	-0.19*	0.11	0.83
Growth	-83.32***	6.43	<0.01%
Inflation	0.85***	0.16	2.35
Interest rate	-0.16***	0.01	0.86

Notes: This table contains presents the coefficient estimates, standard errors and hazard ratios from the semi-parametric Cox model for our main model. An '*', '**', and '***' denote statistical significance at the 10%, 5% and 1% level, respectively. The estimation was based on 1137 extreme stock market events (from 63475 observations), with Concordance 0.724 (Std.Error 0.009), Likelihood Ratio Test 787.9 (p-value<0.001), Wald test 804.3 (p-value<0.001), and Score (logrank) test 853.5 (p-value<0.001).

Table 3. Estimation results when distinguishing between positive and negative event

	Coeff.	Std.Error	Hazard ratios
(1) Stock returns	-3.36***	0.28	0.03
(2) Price-to-book value	0.002	0.001	1.002
(3) Dividend Yield	0.11	0.09	1.11
(4) Price-to-earnings	0.04	0.13	1.05
(5) Growth	-81.84***	7.33	<0.01%
(6) Inflation	1.05***	0.19	2.85
(7) Interest rate	-0.19***	0.01	0.83
(8) D(positive event)	1.02***	0.36	2.77
(9) Stock Returns·D(positive event)	0.910**	0.372	2.483
(10) Price-to-book value·D(positive event)	0.001***	0	1.01
(11) Dividend Yield·D(positive event)	0.28	0.18	1.32
(12) Price-to-earnings·D(positive event)	-0.79***	0.26	0.45
(13) Growth·D(positive event)	-0.43	13.91	0.65
(14) Inflation·D(positive event)	-0.64*	0.35	0.53
(15) Interest rate·D(positive event)	0.07***	0.02	1.07

Notes: This table contains presents the coefficient estimates, standard errors and hazard ratios from the semi-parametric Cox model for our main model. An '*', '**', and '***' denote statistical significance at the 10%, 5% and 1% level, respectively – grey entries are not statistically significant even at 10% level. The estimation was based on 1137 extreme stock market events (from 63475 observations), with Concordance 0.742 (Std.Error 0.009), Likelihood ratio test 858.4 (p-value<0.001), Wald test 875 (p-value<0.001), and Score (logrank) test 943.7 (p-value<0.001).