Received: 27 April 2018 | Revised: 30 December 2019 | Accepted: 18 June 2020
DOI: 10.1002/ijfe.2317

RESEARCH ARTICLE

WILEY

Can black swans be tamed with a flexible mean-variance specification?

Vasiliki Chatzikonstanti[1] | **Michail Karoglou**[2]

[1]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

[2]Aston Business School, Aston University, Birmingham, UK

Correspondence
Vasiliki Chatzikonstanti, Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands.
Email: v.chatzikonstanti@rug.nl

Abstract

We examine the homogeneity of the highly improbable returns, what practitioners and the mainstream economic press also call black swan events. By setting up a simple framework and using the benchmark stock market indices of all OECD countries, we find that the frequency of black swans varies greatly over the last two decades often with dramatic changes that can be related to major economic events. Moreover, during the global financial crisis, black swans were substantially more frequent for most countries even after controlling for the level of volatility. This implies that, despite the plethora of appropriate financial instruments to counter this effect, during an obvious economic turmoil, stock markets are still more likely to experience highly improbable events.

KEYWORDS

black swans, latent non-linearities, stock returns, structural breaks

1 | INTRODUCTION

Remarkable in many ways is that, despite the spectacular econometric advances of the last few decades, the ever-debated distribution of stock market returns remains an excitedly controversial issue. At its core, the presence and abnormal frequency of extreme returns, currently beyond the reach of existing theoretical structures, reveal intermittent model failures with real consequences in areas of the utmost interest to institutional and individual financial market practitioners and researchers, such as asset pricing and risk management. This is what the Goldman Sachs' Chief Financial Officer was referring to when in August 2007 he famously lamented "*We were seeing things that were 25 standard deviation moves, several days in a row*." Market practitioners label these improbable stock market events black swans.[1] And the quick spread of the so-called "black swan" funds is one of the latest manifestations of their unyielding interest in them.

Using a parsimonious mean–variance specification, we test a predominant assumption in the respective literature namely the homogeneity of black swans against the possibility that they are clustered over time. One such cluster of black swans would be made apparent only by comparing it to its neighbouring cluster, effectively by discerning what we call here *black swan swarms*[2]: periods during which (at least) the frequency of black swans is different from the periods just before and after. In other words, our primary objective is to robustly test for the presence of black swan swarms which implies that the distribution of extreme events varies over time and is driven by the arrival of news – although in our methodology, this clustered time variation is based on no assumptions as to whether the news process causes extreme events or extreme events cause (changes in) the news process or both.

One possibility why black swans may be clustered in such a way may be that news follows a process that resembles the succession of swarms each of which may differ in size from its neighbouring ones but also in the number of black swans that they contain. In other words, the explanation for processes that yield black swan swarms may lie either in the arrival process of extraordinary news or in market dynamics in response to the news.[3]

If information comes in clusters, each with its own characteristics, then the frequency of black swans in asset returns or prices may also exhibit clusters. Alternatively, the mechanisms through which expectational differences of traders with heterogeneous priors and private information are resolved may go through distinct phases if, for example, the mechanisms by which processing information and responding to it changes due to changes in the legislative framework or technological means including the available financial instruments.[4] In either case, market dynamics lead to black swan swarms. In either case, market expectation based on public information could be unbiased at every time point, hence consistent with market efficiency and with a view that major sources of disturbances (extraordinary news) are changes in country-specific fundamentals.

Using a comprehensive set of daily stock market returns from the benchmark indices of all OECD countries (34 in total[5]) over a long period, we find that when we account for the potential heterogeneity of black swans: (a) highly improbable events are dramatically less frequent, (b) the country-specific element of the black swans distribution becomes substantially less prominent, and the distributions of positive and negative black swans seem dramatically more similar; (c) the frequency of black swans changes over time, and in some cases, these changes are quite dramatic even for contiguous segments; (d) the frequency of black swans is (at least) uncorrelated to the levels of volatility, across different measures of the latter. These four findings suggest that our notion of different black swan swarms is well justified and can complement existing approaches of modelling tail events. Consequently, the framework we propose may well be directly relevant not only to the valuation practices of "black swan" funds but also to the pricing of a large class of derivatives, for example, out-of-the-money vanilla options and to the estimation of jump processes, given for example the sensitivity of the Poisson rate estimate to the sample size. When we apply this framework to examine the aggregate effect of market participants' behaviour during the 2007/8 crisis, we find that black swans become notably more frequent irrespective of the levels of volatility, an effect that could be attributed to self-fulfilling expectations. Moreover, if we view black swans as model failures, then our findings suggest that these failures do come in clusters but for practical purposes, a flexible mean–variance specification that proxies the retrospective model (which in turn can be viewed as a proxy of the actual albeit unknown stochastic process) seems capable to keep them low and stable over the medium term even when such dramatic events take place.

The remainder of the paper is organized as follows. Section 2 provides the rationale of our approach and distinguishes it from the conventional routes that could have been used potentially in the literature for the same purpose while Section 3 overviews the data. Section 4 describes the methodology and Section 5 presents our results. Finally, Section 6 contains our concluding remarks.

2 | THEORETICAL UNDERPINNINGS

Unsurprisingly, there are several major strands of the academic literature that are tangent to the study of black swans, such as the literature on jump and tail risk. In fact, if one associates asset price movements with the (unknown) news process, then the entire literature of modelling the news process and its impact on economic and financial variables becomes relevant. This is for example what Ehrmann and Fratzscher (2003) do when they look at monetary policy announcements made by the Federal Reserve, Bundesbank and ECB and show their importance in capturing the links between monetary policy and money markets. Or what Fratzscher and Straub (2013) do, within a structural VAR setup, when they link fluctuations in asset prices to news shocks, which they interpret as anticipated changes to technology, and show their substantial effect on trade balances. Or what Dergiades, Milas, and Panagiotidis (2015) do when they illustrate the predictive ability of the Google search queries and social media posts for the short run movements of the financial crises. Or even, within the early warning system (EWS) literature, what Bussiere and Fratzscher (2006) do when they caution against the use of the typical binomial-logit EWS, and in favour of its multinomial-logit extension, due to the former's biased estimation and, in turn, worse ability to anticipate financial crises by failing to distinguish between tranquil and crisis/post-crisis periods.

And yet, so far, the attention has been predominantly the size of specific extreme movements, typically catastrophic events such as stock market crashes. Interestingly, and despite the fact that it has long been shown that the tail behaviour of returns is fundamentally different from the remainder of the return distribution (with evidence as early as Akgiray & Booth, 1988 and Jansen & De Vries, 1991), it is hardly challenging to notice a crucial implicit albeit strong assumption that is made in the vast majority of the relevant papers. Black swans (overall, positive or negative) are assumed as a relatively homogeneous group of observations that, at most, may differ in size – which often prompts each specific case study. We examine how the frequency of black swans evolves in the presence of structural changes and/or latent non-linearities with the purpose of providing some historical grounding of their likelihood of occurrence that can be used for inference. By doing so, as we demonstrate in Section 5.1, we can readily gain further intuition about the operations of financial markets without resorting to some possible alternative methods that, apart from being much more sensitive and demanding in terms of data, their asymptotic properties cannot overcome the natural scepticism that they inspire because of their mostly unknown albeit intuitively very demanding in terms of sample sizes properties.[6]

In the remainder of this section, we explain how the notion of black swans is distinguished from similar notions; then we discuss what the homogeneity of black swans' assumption entails and why it should be tested.

2.1 | What is (and is not) a black swan?

To set ideas let us define r_t, the daily (log)returns process, to be given by the very general:

$$r_t = \mu_t + \sigma_t \cdot u_t, \tag{1}$$

where the (conditional) mean μ_t and/or variance σ_t^2 of the return process may be constant or change over time. The standardized unexpected return is denoted by the u_t terms, which by construction have mean equal to zero and variance equal to one. The customary practice is to classify an observation as a black swan if its standardized unexpected return value is beyond a threshold value, namely three standard deviations above and below the mean. In other words, r_t is a black swan if $|u_t| < 3$.[7]

For market practitioners, the notion of black swan is used primarily as an established heuristic to readily identify dramatic stock return movements during trading days with no reference to any specific model – not least for the purposes of communicating to one another their modelling approach and performance.[8] Likewise, by being both straightforward to calculate and model-free, it proved particularly appealing to the popular press: extreme stock market returns effortlessly grab the attention of the general public; hence they are a favourite source of prominent news headlines and elaborate financial/economic analyses about their possible causes and consequences all while remaining tacit about the underlying modelling intricacies. And maybe for exactly that reason, market participants do care about them, and react to them, and in turn yield their striking real effects. However, amongst academics, it is a notion that is often confused with four fuzzily defined (albeit distinctly popular in the finance and econometric literature) notions: the notion of stock market crashes; the notion of outliers; the notion of extreme value and the notion of breaks.

The notion of black swans overlaps but does not coincide with the notion of stock market crashes. Indeed, a stock market crash is unlikely to take place with returns that are not extreme. Therefore, a stock market crash is typically a black swan, albeit one that is principally characterized by its extreme size, often reflecting or entailing a systemic collapse, rather than its infrequent occurrence, which is primarily the emphasis of the black swan notion. In such cases, for example, portfolio choice models such as the one proposed by Liu and Loewenstein (2013) are effectively about addressing the possibility of negative black swans for optimal portfolio selection. However, black swans may well be positive, and even when negative, they may also refer to magnitudes that do not involve catastrophic drawdowns, such as those that often follow speculative stock market bubbles, but which nevertheless have attracted the attention of the financial press and the general public.

The notion is also often confused with the notion of outliers, which are widely viewed as extreme observations as well, even though typically attributed to measurement errors in contrast to black swans, the frequency of which is often used as a model failure diagnostic. In both cases, the magnitude/size of the extreme observation is a decisive determinant. However, there is a key difference between the two notions. Outliers can be the source of substantial econometric modelling concerns, especially in the context of time series modelling,[9] but despite the lack of a firm mathematical definition, they are in essence an econometric tool which aims to robustify the evidence from an analysis in favour or against some underlying theory. This is why there is a plethora of outlier-identification methods each of which is built upon some set of assumptions to make it suitable for the targeted family of models which may also be multivariate so as to account for any possible cross-correlations or at least interdependence between the variates. In contrast, black swans in financial markets are both defined and identified as the returns that are at least three standard deviations away from the (unconditional) mean of a specific portfolio or series,[10] and they constitute an integral part of the "normal" operation of the markets although typically beyond the counterfactual normality assumption.

The notion of extreme value is the third notion that is often confused with the notion of black swan. Following Pickands III (1975) and Balkema and de Haan (1974), who proved that for a large class of distribution functions, the tails are approximated by the Generalized Pareto distribution, this strand of the literature continues to grow. When it comes to the notion itself, we can observe that from a certain perspective, there can be a lot of overlap between the two notions because the sample estimates of the (conventionally parametric) extreme value distribution parameters are obtained by fixing the frequency of extreme observations.[11] In practice, this means that the threshold that classifies observations as extreme values is found through a grid search over the outermost quantiles of the empirical density, or selected in such a way so as to provide a "reasonable" number of observations to estimate the tail distribution parameters. In contrast, for distributions with finite mean and variance, we can use Chebychev's inequality to obtain an upper boundary for the probability of a black swan occurrence; but the lower boundary is 0, that is, $0 \leq \text{Prob}(|X - \mu| \leq 3\sigma) \leq 1/3^2 \approx 11.1\%$. In other words, unlike extreme values, which by definition must exist for the tail distribution to be proxied in a sample, black swans may well not exist. Consequently, even in sufficiently large samples, the defined extreme values may not be extreme enough to be considered extraordinary and thus be classified as black swans. This is, for example, the case for the normal distribution (about 0.3% in total for both sides of such samples would be classified as black swans) and for the Student-t distribution with more than 2 degrees of freedom (about 1.2% in total for both sides of such samples would be classified as black swans). In this respect, for financial markets data, the notion of black swans can be viewed as a very conservative version of extreme values. However, the difference is more than a simple statistical particularity because it explains why black swan events are perceived to be closer to being the quantitative equivalent of extraordinary events. The reverse, having black swans that do not classify as extreme values, is also true at least theoretically – because in samples, we can always calculate mean and variance estimates. This is, for example, the case for some distributions with undefined mean or variance (e.g., Cauchy, Pareto or Student-t distribution with 2 degrees of freedom).

The notion of breaks is the fourth notion that is often confused with black swans. In fact, had we adopted a less explicit definition for the notion of black swan, a break could also be viewed as a black swan, the first or, more fittingly in our context, the *leader* of each black swan swarm. Indeed, a break as a manifestation of a structural change is quite sporadic and highly improbable. And although they are barely a handful of observations over two decades of data to have any impact upon our inference, if we view them as black swans, it is interesting to mention three features that justify the use of different terms.

First, a break signifies a change in the stochastic process without, however, any indication as to what this change involves and for how long. For example, it could be a change in the unconditional mean of the stochastic process or in the unconditional variance or in the memory of the process or some other characteristic or even a combination of all these. In contrast, a one-observation spike is only about the magnitude of the change.

Second, and building on the above, it is far from being straightforward for a market participant with superior information only about the timing of a break to construct a successful strategy to "beat" the market. S/he would need to also know, for example, what feature of the underlying process will this break change, how long will it take until a suitable strategy can yield economic profits, if and when a next break will take place and so on. In sharp contrast, a market participant who can predict, even roughly, when a one-observation spike will take place can easily obtain abnormal economic gains with, say, a combination of out-of-the money options.

Third, a break may be due to some important economic event that attracts the attention of the public and/or the mainstream economic press.[12] But, it may well be a much more latent event that can be identified only in retrospection. Or it may be what a much more complex and/or unknown data generating process yields whenever we attempt to approximate its behaviour at a particular point in time. In fact, these are some of the main reasons why statistical tools that reveal such events ex-post are vital for any ex-ante modelling.[13] In contrast, in today's world, it is rather hard to imagine an unlikely one-observation change that will not attract the attention of the public and/or the mainstream press. Apparently, distinguishing breaks from black swans is not only an econometric convenience of separation of concerns.[14]

2.2 | The assumption of homogeneity of black swans

In terms of the mainstream academic literature, its long-challenged foundational normality assumption dictates that black swan events should be expected to be exceptionally rare, occurring at daily frequency at most once every 370 trading days. In contrast, the distribution-free Chebychev's theorem implies that they could occur on average even every 9 trading days. Interestingly, a generally accepted "good" fit of the empirical distribution, such as the student-t distribution with 5 degrees of freedom, sets the frequency to a bit more than twice per year.

Evidently, this remarkable difference is the reason why there are so many radically different instruments for and approaches to asset pricing and risk management – which in turn offers a simple explanation as to the undiminished and often avid interest of market practitioners and the research community.[15] It also explains why despite the fact that the notion of black swans, a stylized fact with inherently no theoretical foundations, is at the core of the "battle" of models, an often used criterion to evaluate the real-world applicability of theories that aim to explain the nature of the returns distribution.

The most straightforward and, for that reason, mainstream approach of a large strand of the literature is to assume that the answer to the non-normality puzzle lies in describing the distribution of stock returns as a mixture of normal distributions (see e.g., Kim & Kon, 1996 and references therein). In that way, the gap between the empirical evidence and the theoretical views of informationally efficient markets is elegantly bridged. The new problem that arises is to determine how the mixture of normals is composed.

A non-parametric approach to define the mixture of normals that appears to gain in popularity lately, primarily due to various albeit fundamental deficiencies of parametric specifications, uses data-driven methods to identify multiple breaks in the mean and/or volatility dynamics.[16] The appeal of this approach is especially enhanced by a well-established empirical finding: the aforementioned conditional volatility models are typically estimated with implausibly high levels of volatility persistence, which is consistent with the presence of multiple structural changes that are not taken into account (see e.g., Hillebrand, 2005, and references therein, who shows the direct link between omitted breaks and high levels of persistence). Incorporating some break detection procedure into the existing financial modelling paradigms has already been recognized as essential (see e.g., Kim & Kon, 1999).

Nonetheless, the possibility of breaks, irrespective of whether they are actually attributed to structural changes and/or some ephemeral effect of some unknown nonlinear stochastic process, inevitably questions the validity of the assumption that the frequency of black swans remains unchanged. In other words, the frequency of black swans in one segment of a given sample may or may not be the same to the one of its contiguous segment. In the former case, we can infer that the black swans in the two segments belong in the same black swan swarm. In the latter case, however, we have to accept that they belong in different black swan swarms – which also indicates that the particular breakdate can be thought of as a reasonable proxy of the start of a period that is characterized by a higher or lower black swans' frequency.

What makes this observation most critical is that its inevitable consequence is to challenge the robustness and validity of analyses that lack provision for differences between black swan swarms – effectively making the rather strong assumption that there is a single black swan swarm. And although in case of trivial changes in the frequencies of black swans, a possible bias-reducing method would be to adopt a rolling window of a short time span,[17] under the rather strong assumption that the properties of any estimation undertaken remain unaffected from the width of the sub-sample, in all other cases such an approach is misleading.

For example, the homogeneity of black swans assumption may well lead to substantial mispricing of a large class of out-of-the-money options; averaging out the frequencies of black swans inevitably underestimates (if the most recent swarm features more black swans per time unit than the one just before) or overestimates (if the most recent swarm features less black swans per time unit than the one just before) the actual probability for a black swan to occur. The main aspiration of this paper is to empirically investigate the validity of this assumption and, en route, to offer a simple structure upon which the many affected areas can address the possibility of different black swan swarms.[18]

3 | DATA

The dataset consists of all the reported daily closing values of stock market indices of 34 OECD countries obtained from Thompson Reuters, Datastream. The sample period runs from as early as January 1, 1965 to May 20, 2016 but not for all countries because most benchmark indices were introduced at different times. Table 1 provides some descriptive statistics for our stock index (log) returns also demonstrating the well-documented negative skewness and especially leptokurtosis that yield the characteristic non-normality of stock market returns.

4 | METHODOLOGY

Capturing the heterogeneity of black swans demands an explicit definition of a black swan and a structure that captures the potential heterogeneity. However, these two

TABLE 1 Data overview

	Australia	Austria	Belgium	Canada	Chile	Czech Rep.	Denmark
Obs.	6,255	7,923	9,493	8,950	6,883	5,772	6,904
Mean	0.02%	0.02%	0.03%	0.02%	0.06%	−0.002%	0.03%
St.dev.	0.95%	1.33%	0.98%	1.03%	1.12%	1.33%	1.17%
Skewness	−0.4	−0.3	−0.2	−0.7	0.2	−0.4	−0.3
Kurtosis	5.7	7.6	10.4	12.7	6.8	12.2	5.8
	Finland	**France**	**Germany**	**Greece**	**Hungary**	**Estonia**	**Iceland**
Obs.	7,665	7,531	13,406	7,210	6,622	5,209	6,101
Mean	0.03%	0.01%	0.02%	0.01%	0.05%	0.04%	0.02%
St.dev.	1.61%	1.38%	1.22%	1.86%	1.61%	1.51%	1.71%
Skewness	−0.4	−0.1	−0.2	−0.1	−0.5	−1.01	−45.3
Kurtosis	9.01	5.3	7.5	5.9	11.8	24.7	2,825.9
	Ireland	**Israel**	**Italy**	**Japan**	**Korea**	**Luxembourg**	**Mexico**
Obs.	8,707	7,586	4,797	13,406	10,798	4,534	7,404
Mean	0.03%	0.04%	−0.01%	0.02%	0.03%	−0.01%	0.08%
St.dev.	1.22%	1.45%	1.56%	1.24%	1.50%	1.68%	1.52%
Skewness	−0.6	−0.4	−0.1	−0.4	−0.3	0.2	0.1
Kurtosis	11.1	6.3	3.8	10.3	8.3	59.4	7.5
	Netherlands	**New Zealand**	**Norway**	**Poland**	**Portugal**	**Slovakia**	**Slovenia**
Obs.	8,709	4,015	7,665	5,764	6,101	5,918	2,383
Mean	0.03%	0.01%	0.03%	0.01%	0.01%	0.02%	−0.01%
St.dev.	1.32%	0.69%	1.50%	1.80%	1.16%	1.44%	1.66%
Skewness	−0.3	−0.5	−0.97	−0.2	−0.4	1.5	−0.1
Kurtosis	8.3	5.6	15.6	5.4	7.2	41.9	292.6
	Spain	**Sweden**	**Switzerland**	**Turkey**	**UK**	**USA**	
Obs.	7,402	7,664	7,014	7,142	9,731	13,406	
Mean	0.02%	0.04%	0.03%	0.13%	0.03%	0.02%	
St.dev.	1.38%	1.43%	1.14%	2.60%	1.09%	1.01%	
Skewness	−0.1	0.01	−0.4	−0.04	−0.5	−1.04	
Kurtosis	5.8	4.6	8.2	4.5	21.8	28.4	

Note: Obs. refers to the number of observations in the sample; St.dev. refers to the sample standard deviation; and *Kurtosis* refers to the excess kurtosis observed in the respective sample.

issues are not only inherently interrelated but, most importantly, they are also depended upon the stochastic process that governs the realizations of stock returns for which, as we have briefly illustrated in Section 2, the research community has not succeeded yet in providing a model that is accepted, if not universally at least by some majority. These constitute the core challenges that we need to surmount.

Our approach is to err at the side of caution and be as agnostic as possible about the dynamics that govern the stock returns. For that reason, we assume (a) that the aforementioned structure can be proxied non-parametrically by splitting our samples in continuous segments that are homogenous in terms of mean and variance dynamics (irrespective of whether this segmentation is attributed to actual structural changes taking place in the underlying stochastic process or to latent non-linearities the impact of which in determining the statistical properties of the underlying stock returns can be captured to some extent by this segmentation)[19] and (b) that the definition of black swans is segment-specific. Subsequently, for (a) we identify breaks in the mean and/or volatility dynamics using a battery of break tests and for (b) we identify black swans within each segment. Then, a black swan swarm is identified whenever the frequency of the black swans within a segment differs from that of its neighbouring segments.

With respect to the number and timing of the potential breaks in the mean and/or volatility dynamics, we identify them using the Nominating-Awarding procedure of Karoglou (2010) with the proposed battery of break tests which require at most β-mixing conditions (see Appendix S1 for more details). There are several alternative procedures that exist in the literature which could have been used instead. However, most of them, and often the most popular ones, demand α-mixing or even uniform-mixing conditions making them intrinsically inappropriate for use with high-frequency stock market returns. To our knowledge, so far, the rest are typically isomorphic functions of one or more of the tests in the battery of tests we adopt here so the interested reader will most likely end up, if not with the same, at least with very similar to ours results.

With respect to the identification of black swans within each segment, we follow the widespread 3-sigma customary practice (namely an observation with value beyond the three standard deviations above or below the mean threshold value – following the notation of our main model, $|u_t| < 3$). To robustify our results, apart from the 4-sigma and 6-sigma threshold (i.e., $|u_t| < 4$ and $|u_t| < 6$, respectively), we have also considered an alternative definition of the threshold namely the six and seven times the interquartile range (see, e.g., Stock & Watson, 2005 and Breitung & Eickmeier, 2011). This last type of threshold is even closer in spirit to using the extreme value theory approach although it is still based on sidestepping the estimation of the extreme value distributions.[20] Predictably, the different thresholds yield very similar results, and for that reason, we report only those of the first and more widely accepted definition of black swans.

The remainder of this section describes our main model and how we use it to draw inference.

4.1 | The main model

Schematically, for r_t the daily (log)returns, our main model, which is kept agnostic about the presence of mean and/or volatility persistence by leaving unspecified all respective terms, could be given by the very general:

$$r_t = \begin{cases} \mu_0 + \sigma_0 \cdot u_{0,t} & \text{for } 0 \le t < \tau_1, \\ \mu_1 + \sigma_1 \cdot u_{1,t} & \text{for } \tau_1 \le t < \tau_2, \\ \ldots & \ldots \\ \mu_n + \sigma_n \cdot u_{n,t} & \text{for } \tau_n \le t < T \end{cases}$$

where T is the sample size, n denotes the number of breaks, which occur at dates τ_1, τ_2, ..., τ_n when the (unconditional) mean μ and/or variance σ^2 of the return process change. The standardized unexpected return is denoted by the u_t terms, which by construction have mean equal to zero and variance equal to one.

At this point, and in anticipation of the discussion on breaks that follows, it is worth noting that we make minimal assumptions about the underlying dynamics for both specifications, namely that only for any two adjacent segments (denoted as j and $j + 1$) it holds that only $\mu_j \neq \mu_{j+1}$ or only $\sigma_j \neq \sigma_{j+1}$ or both $(\mu_j, \sigma_j) \neq (\mu_{j+1}, \sigma_{j+1})$. By doing so, we bypass the pitfalls of fully specifying the underlying model, hence making our results relevant to a wide range of modelling paradigms, while preserving statistically endorsed changes in the first two moments (as explained later) – even if only for the given samples.

In this paper, by arguing that blacks swans may not be a homogenous group of observations, we are effectively claiming that there may be one or more $u_{i,t}$ terms that differ substantially from its neighbouring ones, at least in terms of the frequency of its extreme observations. This suggests that all (or a subset) of the identified breaks can serve as an indicator of changes either in the arrival process of extraordinary news or in market dynamics in response to the news. Consequently, we hypothesize that there is an explicit relationship between the two, namely that breaks in the mean and/or volatility dynamics can be used to proxy the beginning and end of black swan swarms, effectively identifying them (but without necessarily causing them).[21]

This setup enables us to have a very straightforward way of testing the assumption of homogeneity of black swans: if in a series, we can identify even one break in the mean and/or volatility dynamics that also entails a (statistically significant with, say, the standard t-test for proportions) change in the frequency of black swans, then for that series, the assumption does not hold. Most importantly, the inference that we can draw from this non-parametric approach[22] is based on minimal assumptions that ground it even within the prevalent mean–variance paradigm. And yet, it is consistent with the strands of the finance literature that emphasizes the importance of moments beyond the second one since it remains valid even if the underlying unknown stochastic process is actually, say, some regime-switching conditional kurtosis or piecewise mixture of jump-diffusion process because it would be capturing locally the average effect of these processes.

Introducing mean and/or volatility persistence within each segment in the main model would not necessarily invalidate our inference about the presence or absence of different black swan swarms.[23] However, for robustness, we are also looking at the same segments in the standardized residuals of the full-sample ARMA-AP(G)ARCH. In

TABLE 2 Major economic events around the identified breakdates

Break date	Major events	Break date	Major events	Break date	Major events	Break date	Major events
Australia		**Austria**		**Belgium**		**Canada**	
October 25, 2001	+44 days of the September 11 attacks	October 6, 1992	—	March 7, 1991	—	February 23, 1988	—
July 25, 2007	Financial crisis of 2007–08	July 26, 2007	Financial crisis of 2007–08	July 31, 1997	+29 days of the Asian Financial Crisis of 1997 (July 2, 1997)	October 27, 1997	October 27, 1997 mini-crash
July 17, 2009	European sovereign debt crisis	November 5, 2009	European sovereign debt crisis	July 24, 2007	Financial crisis of 2007–08	August 19, 2009	European sovereign debt crisis
January 6, 2012		August 7, 2012		January 15, 2008		January 4, 2012	
November 28, 2014		June 5, 2015	2015–16 Chinese stock market crash	May 22, 2009	European sovereign debt crisis	September 19, 2014	
June 26, 2015	2015–16 Chinese stock market crash	**Czech**		October 2, 2014		August 19, 2015	2015–16 Chinese stock market crash
October 6, 2015		June 22, 1998	−26 days of the Russian Financial crisis of 1998 (July 17, 1998)	**Germany**		February 18, 2016	
Chile		June 11, 2010	European sovereign debt crisis	June 7, 1985	—	**Estonia**	
June 9, 1998	−38 days of the Russian Financial crisis of 1998 (July 17, 1998)	March 7, 2012		July 21, 1997	+19 days of the Asian Financial Crisis of 1997 (July 2, 1997)	October 20, 1998	+93 days of the Russian Financial crisis of 1998 (July 17, 1998)
May 16, 2000	Dot com bubble	August 20, 2015	2015–16 Chinese stock market crash	June 17, 2003	—	December 1, 2011	European sovereign debt crisis
December 1, 2011	European sovereign debt crisis	January 4, 2016		January 15, 2008	Financial crisis of 2007–08	June 4, 2013	
May 31, 2013		**France**		July 16, 2009	European sovereign debt crisis	**Greece**	
February 17, 2014		April 27, 1988	—	August 6, 2012		September 25, 2001	+10 days of the September 11 attacks
July 1, 2015	2015–16 Chinese stock market crash	August 3, 1998	+17 days of the Russian Financial crisis of 1998 (July 17, 1998)	October 10, 2014		June 23, 2008	Financial crisis of 2007–08
Finland		April 11, 2003	—	**Ireland**		October 14, 2014	European sovereign debt crisis

TABLE 2 (Continued)

Break date	Major events	Break date	Major events	Break date	Major events	Break date	Major events
August 20, 1992	—	January 15, 2008	Financial crisis of 2007–08	February 9, 1988	—	August 26, 2015	2015–16 Chinese stock market crash
October 20, 1997	−7 days of the October 27, 1997 mini-crash	May 14, 2009	European sovereign debt crisis	July 24, 2007	Financial crisis of 2007–08	**Denmark**	
July 18, 2003	—	August 6, 2012		July 9, 2010	European sovereign debt crisis	July 10, 1997	+8 days of the Asian Financial Crisis of 1997 (July 2, 1997)
July 25, 2007	Financial crisis of 2007–08	September 23, 2014		June 7, 2012		August 8, 2007	Financial crisis of 2007–08
July 17, 2009	European sovereign debt crisis	June 22, 2015	2015–16 Chinese stock market crash	October 2, 2014		June 30, 2009	European sovereign debt crisis
August 8, 2012		October 6, 2015		**Korea**		March 26, 2015	—
April 17, 2015	—	**Iceland**		March 12, 1986	—	**Hungary**	
Italy		August 24, 2004	—	April 30, 2003	2003 invasion of Iraq (March 20–May 1, 2003)	July 7, 1993	—
April 8, 2003	—	December 10, 2008	Financial crisis of 2007–08	July 21, 2009	European sovereign debt crisis	April 1, 1999	Dot com bubble
September 4, 2008	Financial crisis of 2007–08	March 4, 2011	European sovereign debt crisis	September 17, 2012		January 23, 2012	European sovereign debt crisis
May 22, 2009	European sovereign debt crisis	October 10, 2014		June 29, 2015	2015–16 Chinese stock market crash	**Luxembourg**	
August 8, 2012		January 26, 2016	2015–16 Chinese stock market crash	February 19, 2016		March 26, 2003	2003 invasion of Iraq (March 20–May 1, 2003)
September 23, 2014		**Japan**		**New Zealand**		May 9, 2007	Financial crisis of 2007–08
June 29, 2015	2015–16 Chinese stock market crash	January 31, 1975	—	January 7, 2008	Financial crisis of 2007–08	August 3, 2010	European sovereign debt crisis
Mexico		February 21, 1990	—	August 25, 2009	European sovereign debt crisis	August 8, 2012	
October 23, 1997	−4 days of the October 27, 1997 mini-crash	January 4, 2008	Financial crisis of 2007–08	July 3, 2013		June 29, 2015	2015–16 Chinese stock market crash
January 5, 2001	Dot com bubble	May 20, 2009	European sovereign debt crisis	June 10, 2015	2015–16 Chinese stock market cras	**Norway**	

(Continues)

TABLE 2 (Continued)

Break date	Major events	Break date	Major events	Break date	Major events	Break date	Major events
October 16, 2001	+35 days of the September 11 attacks	August 21, 2015	2015–16 Chinese stock market crash	**Slovakia**		December 15, 1992	—
July 24, 2009	European sovereign debt crisis	**Netherlands**		July 4, 1994	—	May 12, 2006	—
December 2, 2009		April 8, 1988	—	February 10, 2003	—	August 4, 2009	European sovereign debt crisis
January 19, 2012		July 16, 1997	+14 days of the Asian Financial Crisis of 1997 (July 2, 1997)	September 9, 2008	Financial crisis of 2007–08	July 24, 2012	
March 13, 2013		July 8, 2003	—	June 29, 2010	European sovereign debt crisis	October 1, 2014	
October 16, 2013		January 15, 2008	Financial crisis of 2007–08	October 28, 2013		August 20, 2015	2015–16 Chinese stock market crash
November 28, 2014		July 16, 2009	European sovereign debt crisis	**Switzerland**		February 29, 2016	
February 13, 2015	—	December 21, 2011		August 24, 1992	—	**Slovenia**	
July 22, 2015	2015–16 Chinese stock market crash	October 1, 2014		July 6, 1998	-11 days of the Russian Financial crisis of 1998 (July 17, 1998)	November 16, 2012	European sovereign debt crisis
Poland		June 22, 2015	2015–16 Chinese stock market crash	June 24, 2004	—	June 18, 2014	
June 6, 1995	—	October 6, 2015		July 25, 2008	Financial crisis of 2007–08	July 21, 2015	2015–16 Chinese stock market crash
February 6, 2002	—	March 2, 2016		April 7, 2010	European sovereign debt crisis	**Turkey**	
May 27, 2010	European sovereign debt crisis	**Portugal**		December 3, 2012		October 28, 1998	+3 months of the Russian Financial crisis of 1998 (July 17, 1998)
August 10, 2012		February 26, 2003	—	October 5, 2015	2015–16 Chinese stock market crash	April 15, 2004	—
June 17, 2015	2015–16 Chinese stock market crash	January 9, 2008	Financial crisis of 2007–08	**USA**		May 24, 2010	European sovereign debt crisis
Israel		December 9, 2008		March 20, 1974	1973–1974 Oil Crisis	**Sweden**	

TABLE 2 (Continued)

Break date	Major events	Break date	Major events	Break date	Major events	Break date	Major events
January 23, 1991	—	**UK**		March 27, 1998	−4 months of the Russian Financial crisis of 1998 (July 17, 1998)	December 23, 1993	—
April 3, 1995	—	February 9, 1989	—	October 19, 1998	+3 months of the Russian Financial crisis of 1998 (July 17, 1998)	October 26, 1998	+3 months of the Russian Financial crisis of 1998 (July 17, 1998)
April 19, 2001	—	July 15, 1999	Dot com bubble	July 19, 2010	European sovereign debt crisis	April 9, 2004	—
September 17, 2009	European sovereign debt crisis	July 19, 2010	European sovereign debt crisis	December 21, 2012		July 24, 2008	Financial crisis of 2007–08
December 13, 2011		December 17, 2012		**Spain**		July 6, 2010	European sovereign debt crisis
January 18, 2012		July 11, 2014		April 8, 2004	—	July 3, 2013	
August 20, 2015	2015–16 Chinese stock market crash	September 23, 2015	2015–16 Chinese stock market crash	January 15, 2009	European sovereign debt crisis	October 2, 2015	2015–16 Chinese stock market crash
September 30, 2015				September 10, 2013			

Note: The table lists major economic events that may be associated with the breakdates that have been identified by the Nominating-Awarding procedure.

TABLE 3 Summary of the number of breaks and frequency of black swans

	Australia	Austria	Belgium	Canada	Chile	Czech	Denmark	Estonia	Finland
1. Number of segments	8	6	7	8	7	6	5	4	8
2. Black swans in full samples	1.31% (82)	1.75% (139)	1.58% (150)	1.56% (140)	1.42% (98)	1.4% (81)	1.36% (94)	1.96% (102)	1.75% (134)
3. Black swans in segmented samples	.78% (49)	1.58% (125)	1.32% (125)	1.22% (109)	1.16% (80)	1.37% (79)	1.23% (85)	1.71% (89)	1.11% (85)
4. Difference	*51.5%*	*10.6%*	*18.2%*	*25.0%*	*20.3%*	*2.5%*	*10.1%*	*13.6%*	*45.5%*
5. Negative black swans in full samples	.78% (49)	1.1% (87)	.91% (86)	.91% (81)	.71% (49)	.92% (53)	.77% (53)	.94% (49)	.87% (67)
6. Negative black swans in segmented samples	.54% (34)	.95% (75)	.68% (65)	.69% (62)	.58% (40)	.85% (49)	.74% (51)	.83% (43)	.57% (44)
7. Difference	*36.5%*	*14.8%*	*28.%*	*26.7%*	*20.3%*	*7.8%*	*3.8%*	*13.1%*	*42.1%*
8. Positive black swans in full samples	.53% (33)	.66% (52)	.67% (64)	.66% (59)	.71% (49)	.49% (28)	.59% (41)	1.02% (53)	.87% (67)
9. Positive black swans in segmented samples	.24% (15)	.63% (50)	.63% (60)	.53% (47)	.58% (40)	.52% (30)	.49% (34)	.88% (46)	.53% (41)
10. Difference	*78.8%*	*3.9%*	*6.5%*	*22.7%*	*20.3%*	*−6.9%*	*18.7%*	*14.2%*	*49.1%*

	France	Germany	Greece	Hungary	Iceland	Ireland	Israel	Italy	Japan
1. Number of segments	10	8	5	4	6	6	8	7	6
2. Black swans in full samples	1.35% (102)	1.37% (184)	1.72% (124)	1.51% (100)	.13% (8)	1.94% (169)	1.37% (104)	1.56% (75)	1.47% (197)
3. Black swans in segmented samples	1.17% (88)	1.04% (140)	1.51% (109)	1.46% (97)	1.46% (89)	1.61% (140)	1.33% (101)	1.% (48)	1.27% (170)
4. Difference	*14.8%*	*27.3%*	*12.9%*	*3.0%*	*−240.9%*	*18.8%*	*2.9%*	*44.6%*	*14.7%*
5. Negative black swans in full samples	.8% (60)	.78% (104)	.94% (68)	.82% (54)	.11% (7)	1.11% (97)	.76% (58)	.98% (47)	.86% (115)
6. Negative black swans in segmented samples	.65% (49)	.54% (72)	.76% (55)	.71% (47)	.74% (45)	.93% (81)	.79% (60)	.6% (29)	.71% (95)
7. Difference	*20.3%*	*36.8%*	*21.2%*	*13.9%*	*−186.1%*	*18.%*	*−3.4%*	*48.3%*	*19.1%*
8. Positive black swans in full samples	.56% (42)	.6% (80)	.78% (56)	.69% (46)	.02% (1)	.83% (72)	.61% (46)	.58% (28)	.61% (82)
9. Positive black swans in segmented samples	.52% (39)	.51% (68)	.75% (54)	.76% (50)	.72% (44)	.68% (59)	.54% (41)	.4% (19)	.56% (75)
10. Difference	*7.4%*	*16.3%*	*3.6%*	*−8.3%*	*−378.4%*	*19.9%*	*11.5%*	*38.8%*	*8.9%*

	Korea	Luxembourg	Mexico	Netherlands	New Zealand	Norway	Poland	Portugal	Slovakia
1. Number of segments	7	6	12	11	5	8	6	4	6
2. Black swans in full samples	1.76% (190)	1.54% (70)	1.66% (123)	1.72% (150)	1.22% (49)	1.59% (122)	1.56% (90)	1.44% (88)	1.64% (97)
3. Black swans in segmented samples	1.43% (154)	.99% (45)	1.46% (108)	1.24% (108)	.87% (35)	1.49% (114)	1.15% (66)	1.25% (76)	2.01% (119)
4. Difference	*21.0%*	*44.2%*	*13.0%*	*32.9%*	*33.6%*	*6.8%*	*31.%*	*14.7%*	*−20.4%*
5. Negative black swans in full samples	.91% (98)	.93% (42)	.81% (60)	1.% (87)	.72% (29)	1.02% (78)	.83% (48)	1.02% (62)	.86% (51)
6. Negative black swans in segmented samples	.69% (75)	.64% (29)	.72% (53)	.79% (69)	.5% (20)	.97% (74)	.71% (41)	.79% (48)	1.12% (66)

TABLE 3 (Continued)

	Korea	Luxembourg	Mexico	Netherlands	New Zealand	Norway	Poland	Portugal	Slovakia
7. Difference	*26.7%*	*37.0%*	*12.4%*	*23.2%*	*37.2%*	*5.3%*	*15.8%*	*25.6%*	*−25.8%*
8. Positive black swans in full samples	.85% (92)	.62% (28)	.85% (63)	.72% (63)	.5% (20)	.57% (44)	.73% (42)	.43% (26)	.78% (46)
9. Positive black swans in segmented samples	.73% (79)	.35% (16)	.74% (55)	.45% (39)	.37% (15)	.52% (40)	.43% (25)	.46% (28)	.9% (53)
10. Difference	*15.2%*	*56.%*	*13.6%*	*48.0%*	*28.8%*	*9.5%*	*51.9%*	*−7.4%*	*−14.2%*

	Slovenia	Spain	Sweden	Switzerland	Turkey	UK	USA
1. Number of segments	4	4	8	8	4	7	6
2. Black swans in full samples	.76% (18)	1.46% (108)	1.58% (121)	1.64% (115)	1.69% (121)	1.55% (151)	1.37% (184)
3. Black swans in segmented samples	.76% (18)	1.35% (100)	1.19% (91)	1.45% (102)	1.22% (87)	1.42% (138)	1.19% (159)
4. Difference	*0.0%*	*7.7%*	*28.5%*	*12.0%*	*33.0%*	*9.%*	*14.6%*
5. Negative black swans in full samples	.46% (11)	.82% (61)	.86% (66)	.98% (69)	.83% (59)	.74% (72)	.67% (90)
6. Negative black swans in segmented samples	.46% (11)	.82% (61)	.61% (47)	.91% (64)	.67% (48)	.67% (65)	.59% (79)
7. Difference	*0.0%*	*0.0%*	*34.0%*	*7.5%*	*20.6%*	*10.2%*	*13.0%*
8. Positive black swans in full samples	.29% (7)	.63% (47)	.72% (55)	.66% (46)	.87% (62)	.81% (79)	.7% (94)
9. Positive black swans in segmented samples	.29% (7)	.53% (39)	.57% (44)	.54% (38)	.55% (39)	.75% (73)	.6% (80)
10. Difference	*0.0%*	*18.7%*	*22.3%*	*19.1%*	*46.4%*	*7.9%*	*16.1%*

Note: (1) reports the number of segments as determined by the identified breaks in each series; (2) and (3) report the number of black swans relative to the sample size (and in absolute terms in brackets) in each series before and after breaks are taken into account respectively; (4) reports the relative percentage difference between (2) and (3) using the log-difference approximation. (5), (6) and (7) are as (2), (3) and (4) respectively but only for the negative black swans. Similarly, (8), (9) and (10) are as (2), (3) and (4) respectively but only for the positive black swans. The greyed cells indicate a rise in the overall number of black swans when segmenting the sample.

TABLE 4 Summary of the frequency of black swans in the residuals of the full-sample best-fit ARMA-AP(G)ARCH

	Australia	Austria	Belgium	Canada	Chile	Czech	Denmark	Estonia	Finland
1. Number of segments	8	6	7	8	7	6	5	4	8
2. Black swans in full samples	.69% (43)	.88% (70)	1.1% (104)	.86% (77)	.78% (54)	1.37% (79)	.74% (51)	1.4% (73)	.85% (65)
3. Black swans in segmented samples	.48% (30)	1.06% (84)	1.03% (98)	.84% (75)	.74% (51)	.78% (45)	.61% (42)	1.46% (76)	.81% (62)
4. Difference	*36.%*	*−18.2%*	*5.9%*	*2.6%*	*5.7%*	*56.3%*	*19.4%*	*−4.%*	*4.7%*
5. Negative black swans in full samples	.48% (30)	.57% (45)	.72% (68)	.55% (49)	.41% (28)	.81% (47)	.32% (22)	.71% (37)	.57% (44)
6. Negative black swans in segmented samples	.4% (25)	.69% (55)	.7% (66)	.55% (49)	.42% (29)	.49% (28)	.38% (26)	.84% (44)	.52% (40)
7. Difference	*18.2%*	*−20.1%*	*3.%*	*0.0%*	*−3.5%*	*51.8%*	*−16.7%*	*−17.3%*	*9.5%*
8. Positive black swans in full samples	.21% (13)	.32% (25)	.38% (36)	.31% (28)	.38% (26)	.55% (32)	.42% (29)	.69% (36)	.27% (21)
9. Positive black swans in segmented samples	.08% (5)	.37% (29)	.34% (32)	.29% (26)	.32% (22)	.29% (17)	.23% (16)	.61% (32)	.29% (22)
10. Difference	*95.6%*	*−14.8%*	*11.8%*	*7.4%*	*16.7%*	*63.3%*	*59.5%*	*11.8%*	*−4.7%*
	France	**Germany**	**Greece**	**Hungary**	**Iceland**	**Ireland**	**Israel**	**Italy**	**Japan**
1. Number of segments	10	8	5	4	6	6	8	7	6
2. Black swans in full samples	.64% (48)	.72% (96)	1.04% (75)	1.1% (73)	.2% (12)	1.3% (113)	.95% (72)	.71% (34)	.9% (120)
3. Black swans in segmented samples	.6% (45)	.63% (84)	.89% (64)	1.12% (74)	1.41% (86)	1.37% (119)	.95% (72)	.71% (34)	.91% (122)
4. Difference	*6.5%*	*13.4%*	*15.9%*	*−1.4%*	*−196.9%*	*−5.2%*	*0.0%*	*0.0%*	*−1.7%*
5. Negative black swans in full samples	.45% (34)	.42% (56)	.51% (37)	.63% (42)	.11% (7)	.64% (56)	.58% (44)	.5% (24)	.54% (72)
6. Negative black swans in segmented samples	.4% (30)	.33% (44)	.44% (32)	.68% (45)	.72% (44)	.7% (61)	.61% (46)	.52% (25)	.54% (72)
7. Difference	*12.5%*	*24.1%*	*14.5%*	*−6.9%*	*−183.8%*	*−8.6%*	*−4.4%*	*−4.1%*	*0.0%*
8. Positive black swans in full samples	.19% (14)	.3% (40)	.53% (38)	.47% (31)	.08% (5)	.65% (57)	.37% (28)	.21% (10)	.36% (48)
9. Positive black swans in segmented samples	.2% (15)	.3% (40)	.44% (32)	.44% (29)	.69% (42)	.67% (58)	.34% (26)	.19% (9)	.37% (50)
10. Difference	*−6.9%*	*0.0%*	*17.2%*	*6.7%*	*−212.8%*	*−1.7%*	*7.4%*	*10.5%*	*−4.1%*
	Korea	**Luxembourg**	**Mexico**	**Netherlands**	**New Zealand**	**Norway**	**Poland**	**Portugal**	**Slovakia**
1. Number of segments	7	6	12	11	5	8	6	4	6
2. Black swans in full samples	.83% (90)	.86% (39)	.96% (71)	.83% (72)	.57% (23)	.85% (65)	.82% (47)	.9% (55)	1.62% (96)
3. Black swans in segmented samples	.82% (89)	.9% (41)	.92% (68)	.83% (72)	.5% (20)	.85% (65)	.85% (49)	.85% (52)	1.88% (111)
4. Difference	*1.1%*	*−5.%*	*4.3%*	*0.0%*	*14.%*	*0.0%*	*−4.2%*	*5.6%*	*−14.5%*
5. Negative black swans in full samples	.48% (52)	.49% (22)	.65% (48)	.56% (49)	.37% (15)	.5% (38)	.5% (29)	.46% (28)	1.05% (62)
6. Negative black swans in segmented samples	.46% (50)	.55% (25)	.57% (42)	.55% (48)	.32% (13)	.53% (41)	.54% (31)	.39% (24)	1.12% (66)
7. Difference	*3.9%*	*−12.8%*	*13.4%*	*2.1%*	*14.3%*	*−7.6%*	*−6.7%*	*15.4%*	*−6.3%*

TABLE 4 (Continued)

	Korea	Luxembourg	Mexico	Netherlands	New Zealand	Norway	Poland	Portugal	Slovakia
8. Positive black swans in full samples	.35% (38)	.37% (17)	.31% (23)	.26% (23)	.2% (8)	.35% (27)	.31% (18)	.44% (27)	.57% (34)
9. Positive black swans in segmented samples	.36% (39)	.35% (16)	.35% (26)	.28% (24)	.17% (7)	.31% (24)	.31% (18)	.46% (28)	.76% (45)
10. Difference	*−2.6%*	*6.1%*	*−12.3%*	*−4.3%*	*13.4%*	*11.8%*	*0.0%*	*−3.6%*	*−28.%*

	Slovenia	Spain	Sweden	Switzerland	Turkey	UK	USA
1. Number of segments	4	4	8	8	4	7	6
2. Black swans in full samples	.84% (20)	.84% (62)	.73% (56)	.74% (52)	1.09% (78)	.91% (89)	.77% (103)
3. Black swans in segmented samples	.84% (20)	.73% (54)	.64% (49)	.7% (49)	.92% (66)	1.05% (102)	.97% (130)
4. Difference	*0.0%*	*13.8%*	*13.4%*	*5.9%*	*16.7%*	*−13.6%*	*−23.3%*
5. Negative black swans in full samples	.46% (11)	.53% (39)	.44% (34)	.51% (36)	.66% (47)	.45% (44)	.48% (64)
6. Negative black swans in segmented samples	.5% (12)	.47% (35)	.42% (32)	.5% (35)	.62% (44)	.51% (50)	.6% (80)
7. Difference	*−8.7%*	*10.8%*	*6.1%*	*2.8%*	*6.6%*	*−12.8%*	*−22.3%*
8. Positive black swans in full samples	.38% (9)	.31% (23)	.29% (22)	.23% (16)	.43% (31)	.46% (45)	.29% (39)
9. Positive black swans in segmented samples	.34% (8)	.26% (19)	.22% (17)	.2% (14)	.31% (22)	.53% (52)	.37% (50)
10. Difference	*11.8%*	*19.1%*	*25.8%*	*13.4%*	*34.3%*	*−14.5%*	*−24.8%*

Note: The row numbers are identical to those of Table 4, to facilitate comparison. Similarly, the greyed cells indicate a rise in the overall number of black swans when segmenting the sample.

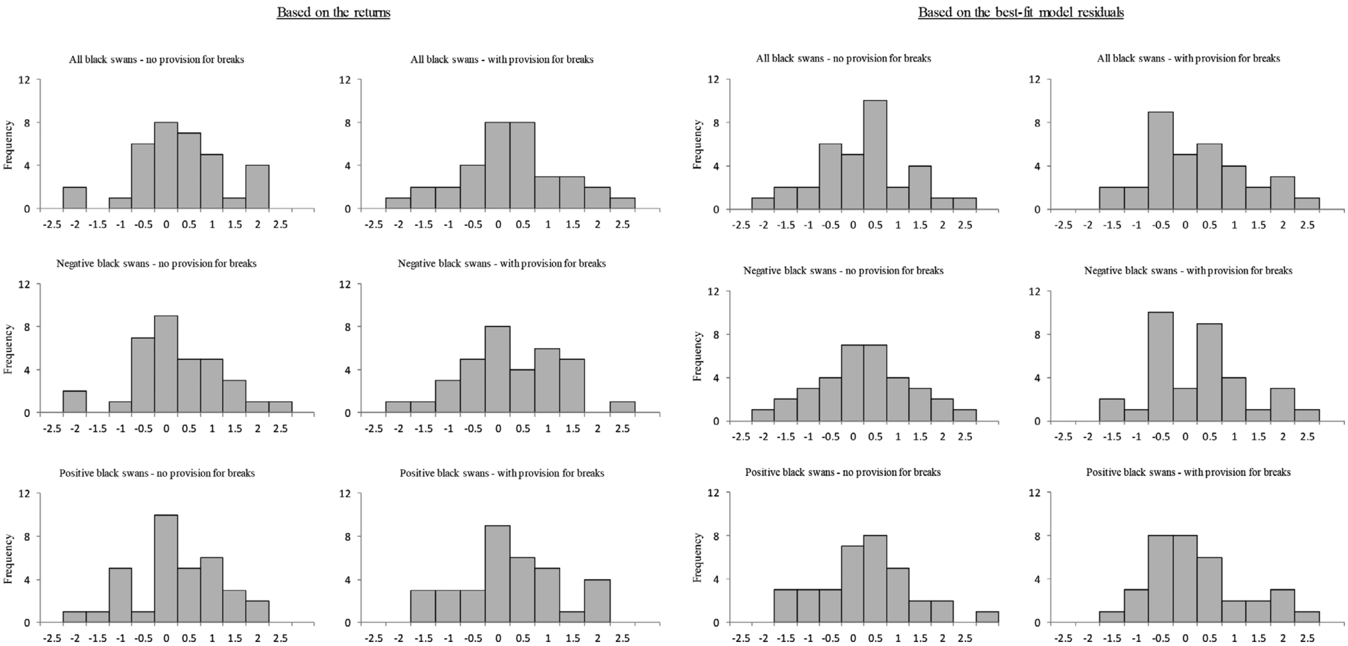

FIGURE 1 Histograms of all, negative only, and positive only black swans. Note: All histograms are based on the standardized values of all 34 markets to facilitate comparison across graphs. Each of these histograms has been obtained by first pooling and then standardizing the respective black swans frequencies of all 34 countries. See also footnote 30 for further details

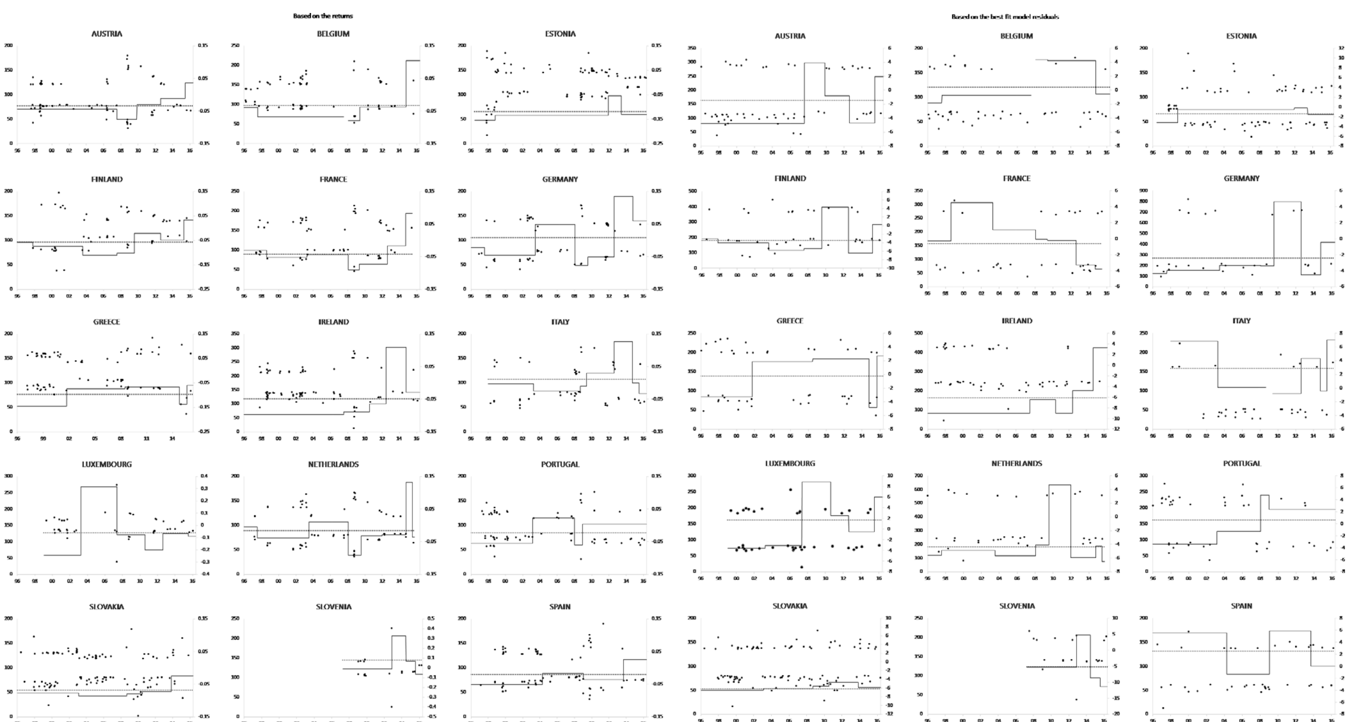

FIGURE 2 (a): Average number of days until a black swan appears over time – eurozone. (b): Average number of days until a black swan appears over time – rest of the world. Note: The left-hand side plots are based on the returns series, while the right-hand side plots are based on the residuals that the best fit ARMA-APGARCH model yields in each case. The dotted lines represent the average days of one black swan to appear for each country and the dots the identified black swans. Discontinuities are due to the absence of black swans. It is worth noting that if the black swans was a homogenous group of observations then no changes would take place across the different segments

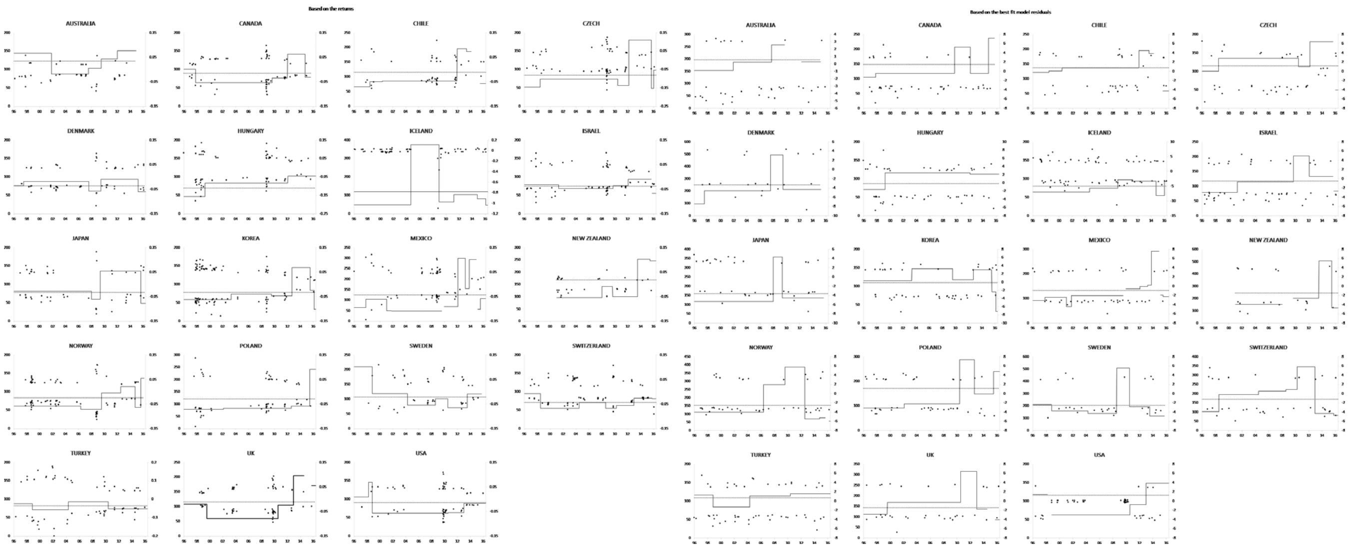

FIGURE 2 (Continued)

this way, even if we assume that the identified breaks are the result of size distortions of the tests due to the conditional mean–variance persistence,[24] we can still test the presence or absence of different black swan swarms. In such case, our model can be given by[25]:

$$r_{i,t} = m_i + \sum_{j=1}^{4} \varphi_j \cdot r_{i,t-k} + \sum_{k=1}^{4} \psi_k \cdot \varepsilon_{i,t-k} + \varepsilon_{i,t},$$
$$\varepsilon_{i,t} = \sigma_{i,t} \cdot u_{i,t},$$
$$\sigma_{i,t}^d = \omega + \sum_{l=1}^{6} a_l \cdot \varepsilon_{i,t-l}^d + \sum_{s=1}^{6} \beta_s \cdot \sigma_{i,t-s}^d + \gamma \cdot \varepsilon_{i,t-1}^d \cdot h_t$$

where the index $i = \{1, 2, ..., n\}$ denotes the segment and $h_t = 1$ if $\varepsilon_{i,t-1} < 0$ otherwise 0. The standardized return term $u_{i,t}$ is as before.[26] Unlike our main model, this flexible specification explicitly captures the possibility of mean persistence of stock returns (when φ's and ψ's are non-zero), with agents that may exhibit herding behaviour which may or may not cause symmetric volatility clustering (when α's and β's and/or γ's are non-zero).

Nevertheless, the reader should bear in mind that by fitting retrospectively such a flexible model onto our data, we are effectively averaging out much of the variability of the series hence disguising many black swans as typical observations. Still, there is some value to this exercise because the bias that it incorporates is different in nature from the bias of assuming a single black swan swarm and therefore can provide insights when the latter is juxtaposed against the multiple black swan swarm proposition. Moreover, it can be considered as a proxy of the best possible ex-ante model that an economic agent can employ for each series and therefore illustrates if there is any added value to considering multiple black swan swarms.

5 | EMPIRICAL RESULTS

In general, we find that the stochastic behaviour of all indices yields about three to eleven breaks during the sample period, roughly one every one and a half to four years on average.[27] The predominant feature of the underlying segments is that mainly changes in variance are found statistically significant. Finally, there are several breakdates that are identical to all series and others that are very close to one another, which apparently signify economic events with a global impact.

Table 2 provides a detailed account of some possible associations that can be drawn between major economic events and the identified breakdates, when changes in the frequency of black swans take place.[28] It appears that dates for the extraordinary events of the Asian financial crisis of 1997, the global financial crisis of 2007/08, the European sovereign-debt crisis that followed and the 2015–16 Chinese stock market crash are very clearly identified in most stock return series and with very little or no variability. Other less spectacular events such as the Russian financial crisis of 1998 or the dot com bubble can also be associated with breakdates that have been identified in some series.

Overall, there are three findings that are particularly insightful. The first finding is about the effect of distinguishing between black swan swarms when determining the total number of black swans. Table 3 summarizes the relevant results.

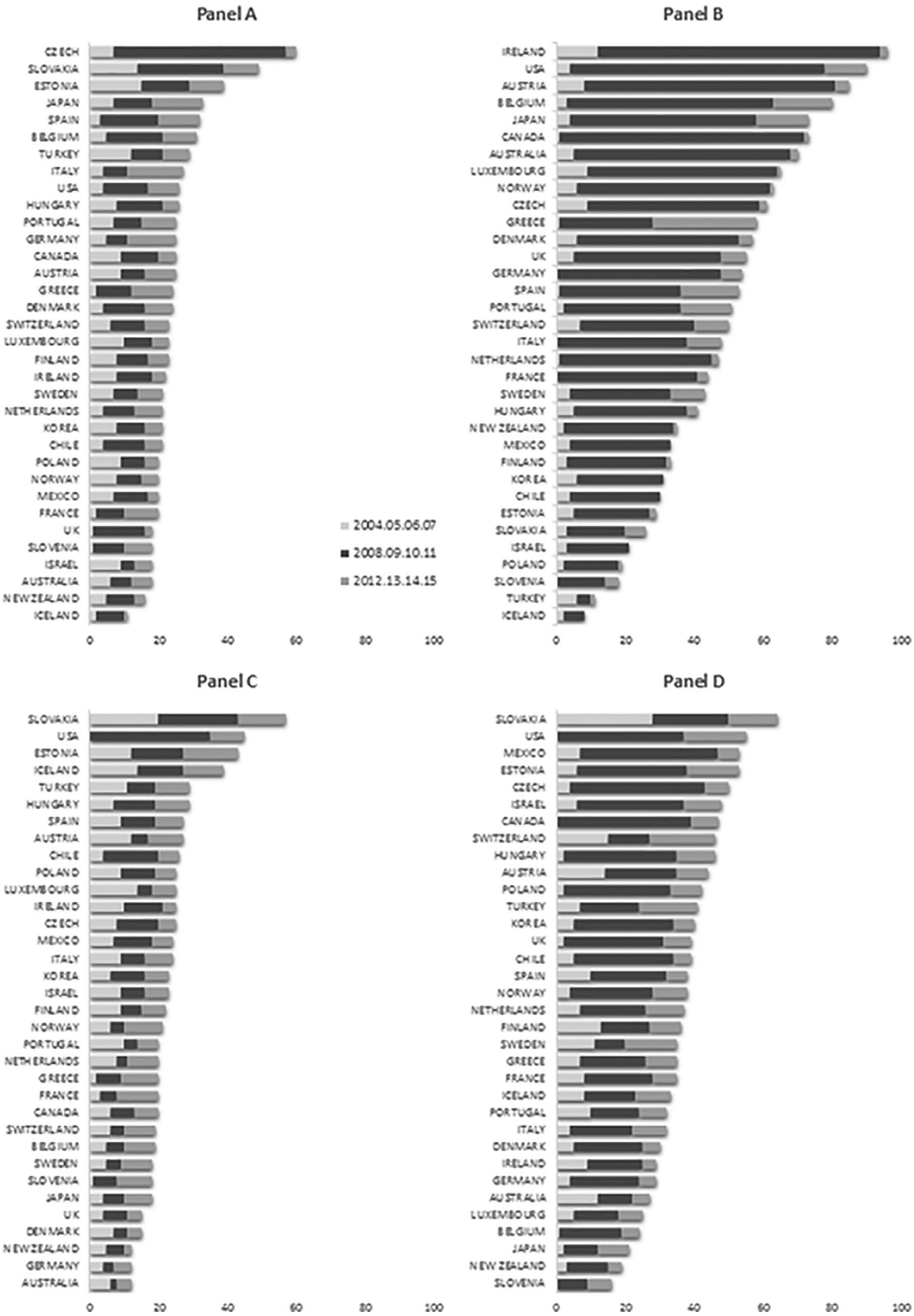

FIGURE 3 Number of black swans per country around the 2007/8 crisis. Note: The left panels (panels a and c) are based on the residuals from the best fit ARMA-APGARCH model in each series, and the right panels (panels b and d) are based on the returns series. The top panels (panels a and b) impose a single black swan swarm; the bottom panels (panels c and d) allow for multiple black swan swarms

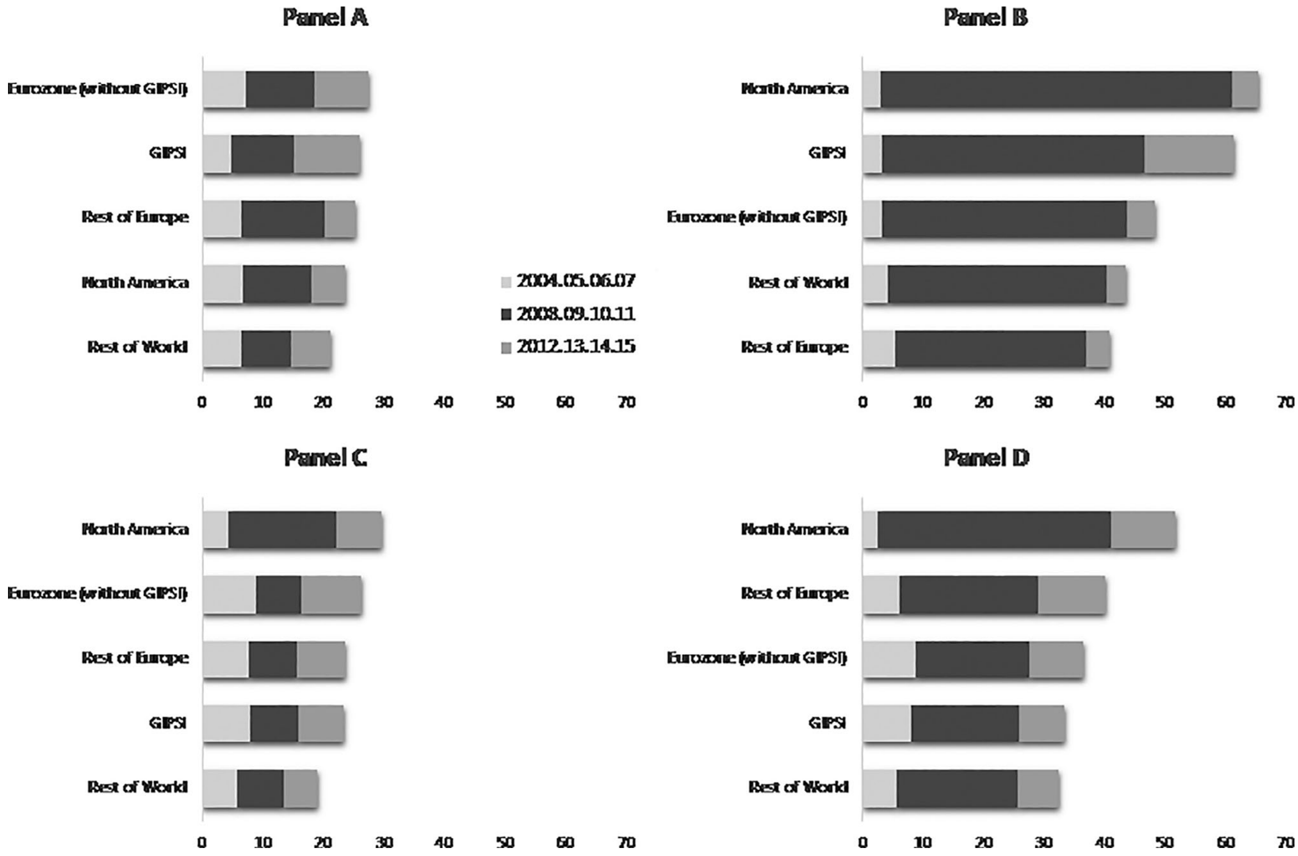

FIGURE 4 Average number of black swans per group around the 2007/8 crisis. Note: The left panels (panels a and c) are based on the residuals from the best-fit ARMA-APGARCH model in each series, and the right panels (panels b and d) are based on the returns series. The top panels (panels a and b) impose a single black swan swarm; the bottom panels (panels c and d) allow for multiple black swan swarms

We observe that with the exceptions of Slovenia and Iceland, the impact of distinguishing the different black swan swarms yields on average about 22% less black swans in total. In other words, highly improbable events are dramatically less frequent when there is provision for possible breaks in the mean and/or volatility dynamics. The identified breaks are quite sporadic; we identify 3–11 breaks in each stock market index, despite the fact that the data span about three decades on average. On aggregate, black swans constitute on average 1.56% of all trading days when there is no provision for breaks, a figure which drops to 1.26% when breaks are taken into account. The case of Iceland is especially informative as to the extent to which latent non-linearities and/or structural changes, such as the banking collapse of 2008 which was also identified by the Nominating-Awarding procedure, can severely bias inference about the frequency of black swans. Table 4 confirms the same finding even though not always and much more moderately.

The effect also persists even when making the distinction between negative and positive black swans[29]; in fact, it appears that it is notably more pronounced with the positive black swans (21% reduction of the frequency of positive black swans as compared to the 19% decrease of the frequency of the negative black swans). Interestingly, the partitioning of black swans into negative and positive reveals that there is a higher proportion of negative black swans, irrespective of whether there is or there is no provision for breaks (accounting for 55 and 56% of the overall black swans respectively) which also offers a partial explanation as to the negative asymmetry of the stock market returns.[30] As before, the same finding is also observed in Table 4 although not as prominently especially in regards to negative black swans.

The second finding involves the broad relationship between the negative and positive black swans across markets. Figure 1 presents the respective histograms for all identified black swans, for only the negative black swans that have been identified and for only the positive black swans that have been identified (standardized to facilitate comparison). Each of these histograms has been obtained by first pooling and then standardizing the respective black swan frequencies of all 34 countries.[31]

The histograms suggest that the impact of distinguishing the different black swan swarms yields a noticeably less different set of stock markets. This

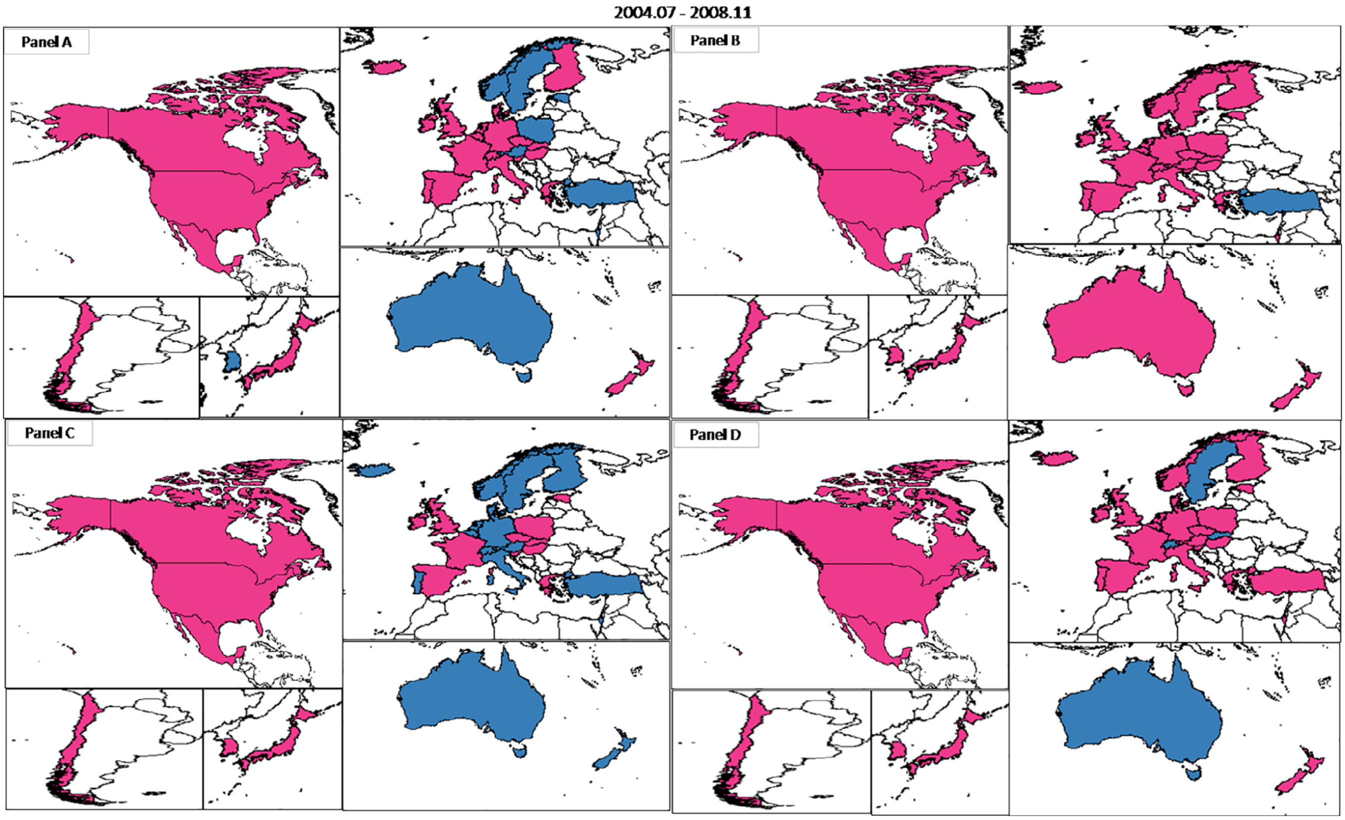

FIGURE 5 More (in pink), equal (in grey) or less (in blue) black swans during 2008–2011 than during 2004–2007? Note: The left panels (panels a and c) are based on the residuals from the best fit ARMA-APGARCH model in each series, and the right panels (panels b and d) are based on the full samples returns series. The top panels (panels a and b) impose a single black swan swarm; the bottom panels (panels c and d) allow for multiple black swan swarms. Grey indicates statistically insignificant changes at 5% level [Colour figure can be viewed at wileyonlinelibrary.com]

effectively implies that the country-specific element when examining black swans becomes less prominent when breaks in the mean and/or volatility dynamics are taken into account. Furthermore, the distributions of positive and negative black swans seem more similar although not for the residuals from the best-fit ARMA-APGARCH model.

The third finding involves the homogeneity of black swan swarms, that is, to what extent the frequency of black swans changes over time as each economy enters into a different regime. Figure 2 presents the respective graphs (for illustration purposes, Figure 2a for the Eurozone economies and Figure 2b for the rest of the world) by expressing the underlying relative frequency of a black swan into the corresponding average days until one appears, as obtained from the reciprocal of the underlying frequency.[32]

For almost all stock markets, the frequency of black swans changes over time, and in some cases, these changes are quite dramatic even for contiguous segments. In fact, there are very few exceptions to the rule that the identified black swan swarms are quite different to one another.[33] It appears that identifying breaks in the mean and/or volatility dynamics can indeed be a reliable proxy for capturing changes in the frequencies of black swans since only in the cases of Denmark, Israel, Slovakia and Turkey, the breaks do not always seem to have identified very different black swan swarms across the selected sample. In all other markets, they capture either only dramatic changes or also gradual ones that end up very different from where they started. From a different albeit relevant perspective, these empirical findings provide further support to the importance of incorporating some break detection procedure into the existing financial modelling paradigms for ex-post but especially for ex-ante analysis and decision-making, what Kim and Kon (1999), amongst others, emphatically urge the research community and market practitioners to do. We revisit the same figures in the next section.

In sum, the notion of black swan swarms seems robustly justified. The simple, and suitable for a large class of modelling paradigms, approach we adopt, the use

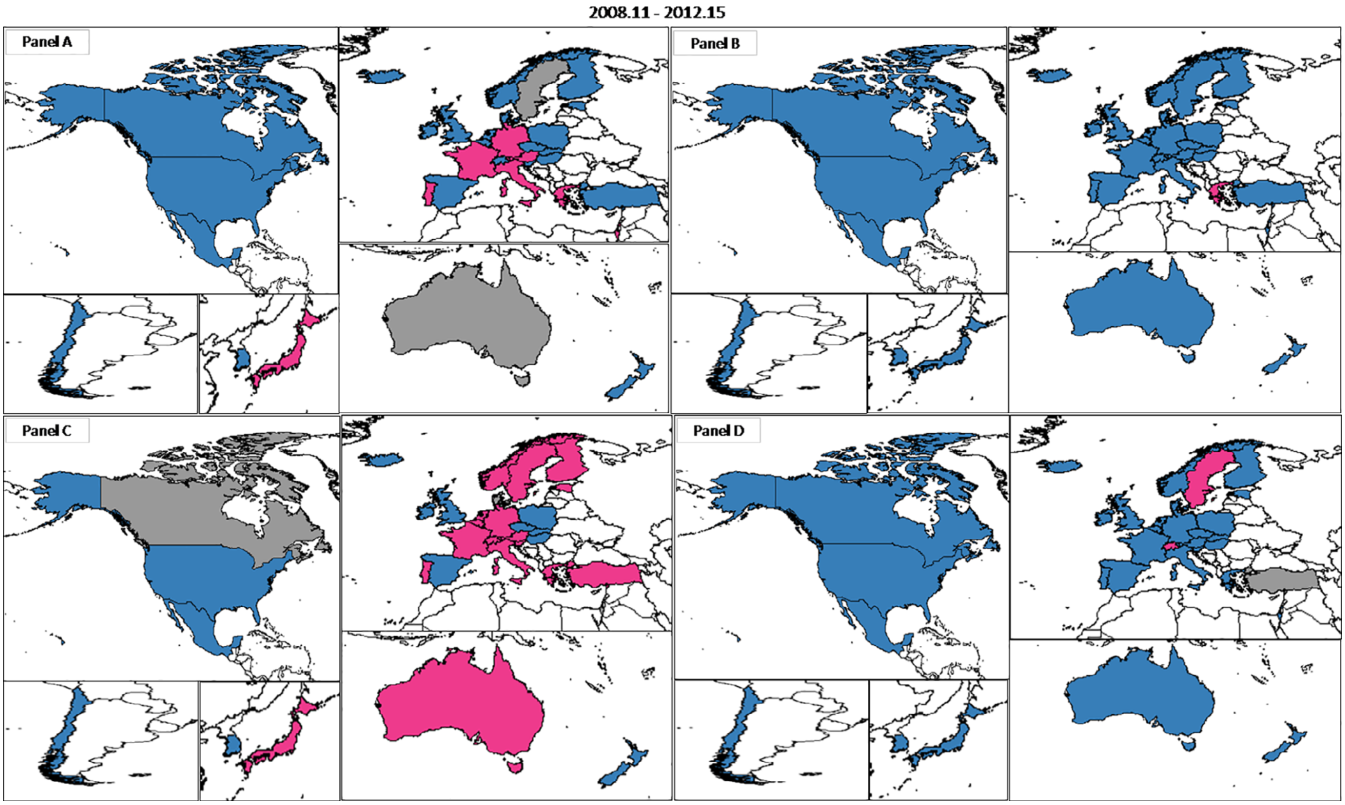

FIGURE 6 More (in pink), equal (in grey) or less (in blue) black swans during 2012–2015 than during 2008–2011? [Colour figure can be viewed at wileyonlinelibrary.com]

of a data-driven break-detection procedure to proxy the different swarms, when employed on a comprehensive set of stock market returns yields several plausible econometric results all while remaining quite agnostic about the underlying stochastic processes and the likely interdependencies amongst the series. Consequently, the use of the notion can be used to improve our intuition about the operation of financial markets. In the section that follows, we make use of this notion to examine what can we observe about the aggregate behaviour of investors and traders before and after the global financial crisis.

5.1 | The frequency of black swans before and after the 2007/8 financial crisis

The 2007/8 financial crisis, with an "official" starting date the collapse of the Lehman Brothers, is widely considered by many economists as one of the worst financial crises in history, often compared to the Great Depression of the 1930s. One of its most characteristic features is its worldwide effect part of which is the sovereign debt crisis that succeeded it, granting it rather justifiably the title global financial crisis. Consequently, it is an economic phenomenon that can exemplify how the notion of black swan swarms can deepen our understanding of the response of financial markets to such shocks.

What we do here is to look at the differences in the black swans around the 2007/8 crisis. Figure 3 compares the frequencies of black swans before the 2007/8 crisis with the frequencies of black swans after the 2007/8 crisis.

Panels a and c, which are based on the standardized residuals obtained from the best-fit ARMA-AP(G) ARCH model, show mixed results which suggest that the recent period is not much different from the longer period before. Allowing either for one black swan swarm or for multiple, the frequency of black swans before and after the 2007/08 crisis is statistical significant in 10 and 8 series, respectively, out of 34. This should actually be expected given that, as we note in the respective section, by fitting such a model, we are averaging out much of the variability in the series, effectively imposing that the whole sample exhibits the same stochastic behaviour. When we look at panels b and d, we see more notably that the recent period is not the same as the longer period, at least in terms of the (average) frequencies of black swans. The

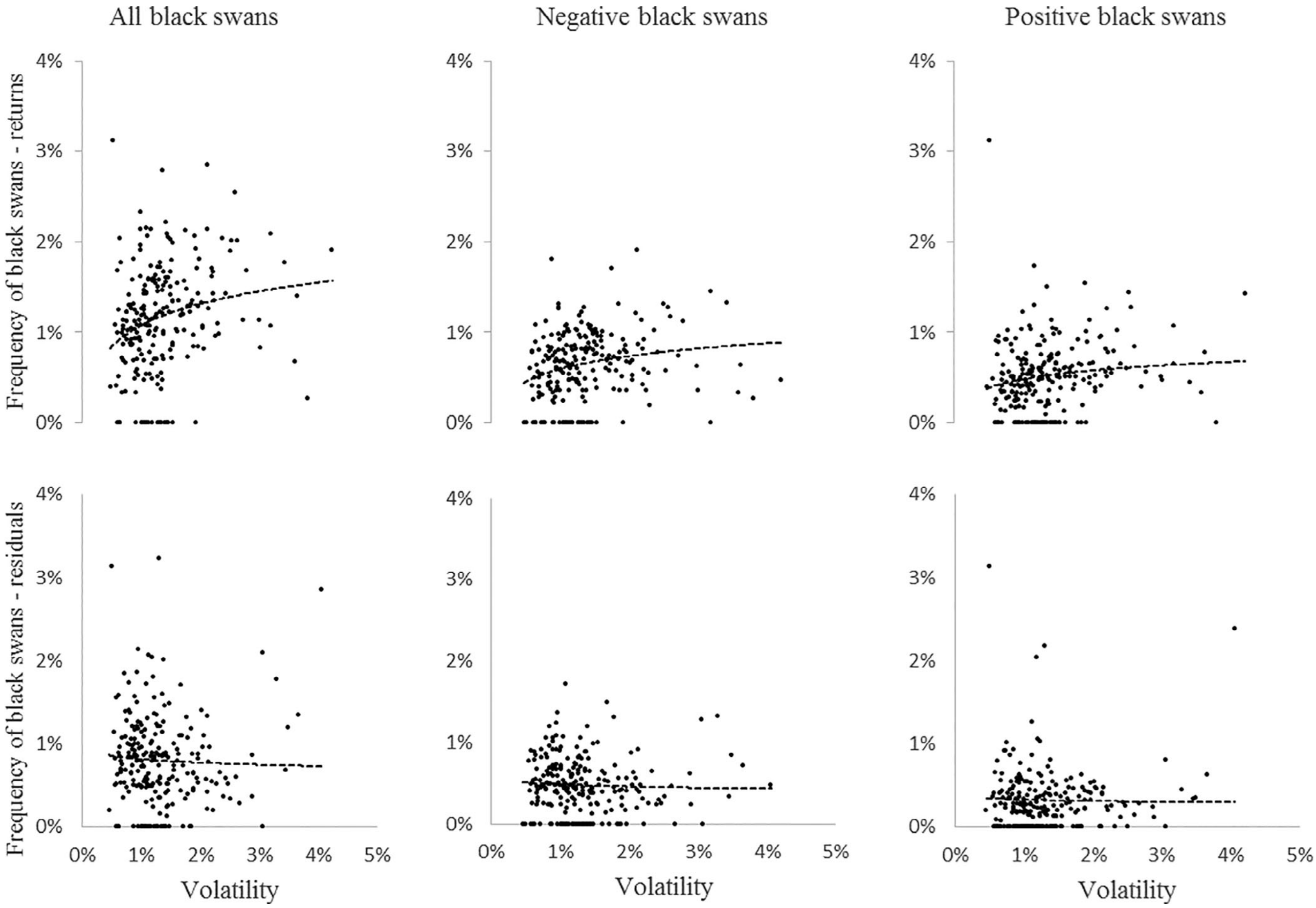

FIGURE 7 Scatterplots of volatility against the frequency of black swans. Note: The six scatterplots depict the combinations of a volatility estimate and the frequency of black swans for every identified segment. Four different volatility estimators have been examined namely: (a) using the sample standard deviation, which is also the one depicted; (b) using the Bartlett kernel; (c) using the Quadratic Spectral kernel (both [b] and [c] implemented with the Newey-West automatic bandwidth selection procedure); and (d) using the VARHAC kernel of den Haan and Levin (1998). All volatility estimators yield the same results and for that reason only (a) is depicted. The six regression lines are based on the log volatility to conform to the positivity constraint and all have statistically insignificant elasticities (slopes) at 5% level (the Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors are based on the Bartlett kernel with the Newey-West automatic bandwidth selection procedure). The same outcome (i.e., non-significance of elasticities) is observed even when we make a distinction between pre- and post-crisis periods

frequency of black swans before and after the 2007/08 crisis is statistical different in 32 out of 34 countries under the assumption of a single black swan swarm and in 24 out 34 countries when allowing for multiple black swan swarms. In particular, we see that in most (but not all) countries, black swans are more frequent that what they used to be. In other words, it seems that the black swan swarms after the 2007/8 crisis contain some pieces of information about the still on-going global economic turmoil.

When we revisit Figure 2a,b and focus on the post 2007/8 period, we can observe that in many markets, the average number of days that a black swan appears changes and often dramatically, but the so-called lead–lag effects make it hard to discern whether these changes are associated with the period of the crisis. Therefore, to draw a clear conclusion as to the effect of the 2007/8 crisis, we focus on three 4-year periods namely the 2004/7, 2008/11 and 2012/15.

Figures 3 and 4 illustrate this information for each of the four specifications we consider, while Figures 5 and 6 graph the statistically significant differences. The results are quite straightforward: in the 4 years, after the start of the 2007/8 crisis, the number of black swans was higher. This could effectively be interpreted as a natural consequence of the increased systemic risk that stock market participants expect and experience; apparently stock markets are more likely to generate extreme returns when their participants expect them.[34] Also, the specifications that condition on black swan swarms can yield substantially different results that those that do not both in the number of black swans and in the period that occurred. The US and Czech Republic markets are particularly telling: both the number of black swans and their timing can be overwhelmingly different, effectively reinforcing the message of Kim and Kon (1999) that incorporating a break-detection process in financial modelling should be considered essential.

What makes this finding particularly interesting is that if we plot the frequency of black swans against a volatility measure, as we do in Figure 7, we cannot identify any significant positive or negative relationship between the two. The absence of a relationship between the two remains irrespective of whether we look at all the black swans or only the positive or negative ones, not even when we condition the analysis on only the pre 2007/8 period and in the post 2007/8 period. In other words, a higher volatility cannot be associated with either a higher or a lower frequency of black swans. Therefore, the fact that in the post 2007/8 crisis period, the frequency of black swans rises cannot be attributed to the underlying higher or lower volatility.

6 | CONCLUSIONS

This study examines empirically the homogeneity of highly improbable events, black swan events, and in particular whether their frequency changes over time. Our analysis endeavours to remain as agnostic as possible about the underlying distribution of stock returns, and for that reason, it is built upon the notion of breaks in the mean and/or volatility dynamics to capture structural changes in the stock markets and to moderate the inevitable bias of latent nonlinearities that might be present in the underlying stock returns. Subsequently, we introduce the notion of black swan swarm to describe the frequency of black swans within a homogenous, in terms of mean and volatility dynamics, periods of time which enables us to investigate how the frequency of black swans evolves.

All 34 stock market return series display a large number of black swans, which is considerably reduced when we relax the assumption of black swan homogeneity, a feature that is prominent even when black swans are distinguished into negative and positive, whereas the proportion of negative is higher. Moreover, the country-specific statistical features become less pronounced, and the distribution of negative and positive black swans less dissimilar. Finally, the evolution of the frequency of black swans reveals a smaller likelihood of a black swan occurring before and after the recent financial crisis in most stock markets which is particularly interesting given that the likelihood of a black swan and the underlying volatility level are not correlated. That also suggests that it is worth exploring the possibility of using changes in the frequency of black swans within an early warning system of crises either by itself as a leading indicator or in conjunction with other predictors. Given the poor performance of existing approaches in this literature (see e.g., Christofides, Eicher, & Papageorgiou, 2016; Obstfeld, Shambaugh, & Taylor, 2009, 2010), it is certainly a possibility worth exploring.

If the rise in the frequency of black swans during the 2008 crisis is interpreted as a feature of the arrival process of extraordinary news, then it suggests that during periods of widespread economic turmoil extraordinary news become more frequent. Alternatively, if this is interpreted as a feature of market dynamics in response to the news, it implies a self-fulfilling expectations mechanism but with a twist: when market participants become wary that extreme events may take place, their collective actions increase the likelihood of such extreme events actually taking place hence validating their expectations. The twist? They may be completely wrong as to whether these extreme events are favourable or not.

An important feature of our analysis that is worth noting is that it is based upon data from the benchmark stock market indices of a number of countries. These benchmark indices can also be thought of as dynamically reconfigured portfolios of shares which, at least in principle, have been selected in such a way so as to proxy the respective market portfolio. Consequently, the framework we propose to improve control over the occurrence of black swans, say for pricing out of the money vanilla options, and the inference we draw by using it are also directly relevant to the portfolio and risk management practices of the individual or institutional investor, while its parsimony and minimal assumptions make it readily available as an instrumental device of real-time computerized trading algorithms.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Vasiliki Chatzikonstanti https://orcid.org/0000-0001-8669-1348
Michail Karoglou https://orcid.org/0000-0002-6730-504X

ENDNOTES

[1] We acknowledge that the term “black swan” has been used in other contexts as well which may also imply a rather qualitative interpretation (Taleb, 2007 for example includes an interesting anthology of such uses of the term). It must be underlined,

however, that our focus is the predominant (and quantitative) usage of the term amongst financial markets practitioners.

[2] We have used the term 'swarm' because, apart from the fact that it sounds very similar to the word "swan" and contracts the financial from ornithological jargon, we became aware from colleagues that the more unfamiliar albeit appropriate "wedge" may convey a meaning different from what is intended.

[3] This is similar to the heat waves explanation of the country-specific volatility clustering that Engle et al (1990) proposed.

[4] The launch of exchange-traded derivatives markets or automatic trading built upon the electronic trading platforms as opposed to the traditional forms, such as open outcry or pit trading, can be considered as two clear-cut examples of such changes.

[5] Latvia, which joined rather recently (on the 1 July 2016) is not included.

[6] See for example, Beirlant, Goegebeur, Segers, and Teugels (2005) who discusses dynamic extreme value models to handle clustering of extremes (Chapter 10) or Jondeau, Poon, and Rockinger (2007) who discusses the modelling of volatility jumps and higher moments (Chapters 4 and 5).

[7] Following the practices of engineers other thresholds have also been considered by financial practitioners (most notably the 6-sigma events) but this is the predominant value.

[8] For example, "I've been bitten by a black swan" is a widespread playful metaphor amongst them.

[9] For example, they may erroneously suggest or hide true heteroscedasticity (van Dijk, Franses, & Lucas, 1999), they induce bias even in the case of a single instance to the maximum likelihood estimators (Carnero, Pena, & Ruiz, 2007; Sakata & White, 1998) and they may also bias out-of-sample forecasts (Charles, 2008; Chen & Liu, 1993a; Franses & Ghijsels, 1999; Ledolter, 1989). Moreover, they are inherently associated to "smearing" and "masking" effects (see e.g., Bruce & Martin, 1989), the former referring to the presence of outliers that may bias the diagnostics which results in false identification of other outliers, and the latter being associated to the occurrence of large outliers which prevent the identification of other outliers.

[10] The identification method is apparently associated with the adoption by financial market practitioners of statistical modelling methods devised and used initially for manufacturing processes. This is why black swans are sometimes referred to as *3-sigma* events or 3σ's since this is the boundary point that Walter A. Shewhart, the father of statistical quality control, determined to signal the difference between events that are ordinary and predictable and those that are unusual and unpredictable. However, there is nothing special about that boundary; other boundaries have also been used but only the *6-sigma* may be somewhat noteworthy.

[11] Predictably, the most popular method to pick up the extreme values is named peaks-over-thresholds but the statement is broader.

[12] In our paper, we provide a list of major economic events that could be associated with the identified breaks in each case in Table 3.

[13] For example, they prescribe a more robust window of observations from which the coefficient estimates need to be derived to avoid potentially large biases.

[14] From a somewhat different angle, this distinction can also be viewed as pertinent to the decision maker's problem of detecting if a regime shift has actually taken place as in Massey and Wu (2005) which is effectively about separating "signal" from "noise."

[15] In this respect, it is hardly surprising that works such as Taleb (2007) periodically become the focal topic of heated academic and popular press debate.

[16] Markov-regime-switching extensions of the aforementioned parametric models can be thought of as having been developed in the same spirit. They may be intuitively less appealing (e.g., it is rather hard to convince a market practitioner or academic that groups of observations of vastly different periods belong in the same regime) and, unlike modelling with some break identification method, for most applications practically intractable when considering the possibility of more than a handful of regimes, but they do implement a data-driven identification of a specific number of regimes –which can also be thought of a special form of breaks. And their importance in, say, portfolio decisions has been convincingly demonstrated in several papers (see e.g., Tu, 2010).

[17] Rolling techniques are often adopted by market practitioners but in the finance literature they are generally discussed in the context of technical analysis.

[18] For example, in the pricing of out-of-the-money call and put options.

[19] In other words, the breaks we are referring to may not necessarily correspond to structural changes but they may be an artifice of the (unknown) underlying stochastic process. However, they do signify when the unconditional mean and/or variance changes (either due to a structural change or due to the dynamics of the stochastic process). Therefore, they can constitute a satisfactory proxy for the start and end points of a black swan swarm.

[20] Alternatively, we could have followed the conventional route of the extreme value theory instead by focussing on, say, the 1st and 99th quantiles and examine the sizes of the thresholds or the parameter estimates of the Generalised Pareto distributions. However, in our series there are several segments with less than 500 observations making this approach unsuitable for our purposes even in its simplest variants. More elaborate methods to encompass, say, dynamic clustering of the extreme values are even more data demanding to provide any reasonable estimates. The simulation study included in the tail-index test for stationarity of Kim and Lee (2009) is rather indicative of this issue.

[21] It is worth underlying at this point that we do not need to make any additional assumptions about the properties of the stochastic process within each segment. However, by referring to the results of the relevant literature which we mention before, and in anticipation of the results of the next section, we could confidently expect typically no or very low levels of volatility persistence.

[22] We acknowledge that the notion of "non-parametric" tests in the time series framework is somewhat misleading since the fact that the time ordering inherently implies a modelling structure. Here, we use the notion as is intended in the relevant literature, namely to signify tests that make very weak assumptions about the properties of the underlying stochastic process and hence are valid under a very large set of modelling structures.

[23] Strictly speaking it is conceivable to have a nonlinear process for which our selection of $\tau_1, \tau_2, \ldots, \tau_n$ is not only biased but biased in such a devilish way that the actual presence or absence of different black swan swarms is perfectly masqueraded for all the stock market series we are examining. Our view is that we can reasonably safely assume that this is not the case.

[24] Note however that, as explained later on, we use also tests that do not suffer from size distortions even in the presence of most IGARCH structures.

[25] The order of each element in this process was chosen in order to capture the maximum order of the corresponding element that was identified in all estimated processes. It has therefore been selected *ex post*, and not *ex ante*.

[26] The orders of each term are found according to the maximum best-fit model for all series (i.e., they have been determined *ex-post*, not *ex-ante*).

[27] As seen in Appendix.

[28] A most interesting feature is that the vast majority of the timings of the identified breaks and the respective economic events either coincide or can effortlessly be explained by the all too known lead–lag stock market responses. It is important to bear in mind, however, that for these established financial markets even if we cannot identify a proximate cause for an event it can hardly mean that it is not there.

[29] Such distinction is particularly important for various strands of the finance literature; for example, the frequency of negative black swans is the primary focus of the literature on the negative tail risk.

[30] Although not reported here to conserve space, the negative skewness decreases substantially when we drop from the samples the black swan observations, especially when breaks are taken into account.

[31] In other words, after we count the respective black swans for each country (first, all black swans irrespective of sign, then only the positive black swans that is, those that are above zero, and finally only the negative black swans that is, those that are below zero), we standardise these counts (subtract from each count the arithmetic mean of all counts and then divide this difference by the standard deviation of all counts) and draw the respective histogram for the resulting values.

[32] This means that due to the absence of black swans in some segments the underlying graph may show discontinuities.

[33] From a visual inspection the occurrence of black swans within each swarm seems rather arbitrary, with only a handful of segments showing signs that could be viewed as clusters of black swans.

[34] This is in contrast with the findings of Burnie and de Ridder (2010) who examine the frequency of extreme day returns, defined as the absolute return equal to or greater than 1.5%, on the Stockholm Stock Exchange and who find that the frequency of extreme day returns increased over time. However, they are in line with the results of Ané, Ureche-Rangau, Gambet, and Bouverot (2008) who analyse the Asian stock markets, and find an accumulation of abnormal observations in 1997, 1998 and 2001.

REFERENCES

Akgiray, V., & Booth, G. G. (1988). The stable-law model of stock returns. *Journal of Business & Economic Statistics*, *6*, 51–57.

Ané, T., Ureche-Rangau, L., Gambet, J. B., & Bouverot, J. (2008). Robust outlier detection for Asia-Pacific stock index returns. *Journal of International Financial Markets Institutions and Money*, *18*, 326–343.

Balkema, A. A., & De Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, *2*, 792–804.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2005). *Statistics of extremes: Theory and applications*. Wiley Series in Probability and Statistics. New York, NY: John Wiley & Sons.

Breitung, J., & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, *163*, 71–84.

Bruce, A. G., & Martin, R. D. (1989). Leave-k-out diagnostics for time series. *Journal of the Royal Statistical Society: Series B: Methodological*, *51*, 363–424.

Burnie, D. A., & De Ridder, A. (2010). Far tail or extreme day returns, mutual fund cash flows and investment behaviour. *Applied Financial Economics*, *20*, 1241–1256.

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

Carnero, M. A., Pena, D., & Ruiz, E. (2007). Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis*, *28*, 471–497.

Charles, A. (2008). Forecasting volatility with outliers in GARCH models. *Journal of Forecasting*, *27*, 551–565.

Chen, C., & Liu, L. (1993a). Forecasting time series with outliers. *Journal of Forecasting*, *12*, 13–35.

Christofides, C., Eicher, T. S., & Papageorgiou, C. (2016). Did established early warning signals predict the 2008 crises? *European Economic Review*, *81*, 103–114.

Den Haan, W. J., & Levin, A. (1998). Vector autoregressive covariance matrix estimation. Manuscript, University of California, San Diego.

Dergiades, T., Milas, C., & Panagiotidis, T. (2015). Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers*, *67*, 406–432.

Ehrmann, M., & Fratzscher, M. (2003). Monetary policy announcements and money markets: A transatlantic perspective. *International Finance*, *6*, 309–328.

Franses, P. H., & Ghijsels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, *15*, 1–9.

Fratzscher, M., & Straub, R. (2013). Asset prices, news shocks, and the trade balance. *Journal of Money, Credit and Banking*, *45*, 1211–1251.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, *129*, 121–138.

Jansen, D. W., & De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, *73*, 18–24.

Jondeau, E., Poon, S., & Rockinger, M. (2007). *Financial modelling under non-Gaussian distributions*. London, England: Springer.

Karoglou, M. (2010). Breaking down the non-normality of daily stock returns. *European Journal of Finance*, *16*, 79–95.

Kim, D., & Kon, S. J. (1996). Sequential parameter nonstationarity in stock market returns. *Review of Quantitative Finance and Accounting*, *6*, 103–131.

Kim, D., & Kon, S. J. (1999). Structural change and time dependence in models of stock returns. *Journal of Empirical Finance*, *6*, 283–308.

Kim, M., & Lee, S. (2009). Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli*, *15*, 325–356.

Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, *23*, 1–40.

Liu, H., & Loewenstein, M. (2013). Market crashes, correlated illiquidity, and portfolio choice. *Management Science*, *59*, 715–732.

Massey, C., & Wu, G. (2005). Detecting regime shifts: The causes of under-and overreaction. *Management Science*, *51*, 932–947.

Obstfeld, M., Shambaugh, J., & Taylor, A. (2009). Financial instability, reserves, and central bank swap lines in the panic of 2008. *American Economic Review*, *99*, 480–486.

Obstfeld, M., Shambaugh, J., & Taylor, A. (2010). Financial stability, the trilemma, and international reserves. *American Economic Journal: Macroeconomics*, *2*, 57–94.

Pickands, J., III. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*, 119–131.

Sakata, S., & White, H. (1998). High breakdown point conditional dispersion estimation with application to S&P500 daily returns volatility. *Econometrica*, *66*, 529–567.

Stock, J., & Watson, M. (2005). Implications of dynamic factor models for VAR analysis. Mimeo.

Taleb, N. (2007). *The Black Swan. The impact of the highly improbable*. New York, NY: Random House.

Tu, J. (2010). Is regime switching in stock returns important in portfolio decisions? *Management Science*, *56*, 1198–1215.

van Dijk, D., Franses, P. H., & Lucas, A. (1999). Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics*, *14*, 539–562.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Chatzikonstanti V, Karoglou M. Can black swans be tamed with a flexible mean-variance specification? *Int J Fin Econ*. 2020;1–26. https://doi.org/10.1002/ijfe.2317