

RESEARCH ARTICLE

The Asian Correction Can Be Quantitatively Forecasted Using a Statistical Model of Fusion-Fission Processes

Boon Kin Teh^{1,2*}, Siew Ann Cheong^{1,2}

1 Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Republic of Singapore, **2** Complexity Institute, Nanyang Technological University, Block 2 Innovation Centre, Level 2 Unit 245, 18 Nanyang Drive, Singapore 637723, Republic of Singapore

* S130005@e.ntu.edu.sg



OPEN ACCESS

Citation: Teh BK, Cheong SA (2016) The Asian Correction Can Be Quantitatively Forecasted Using a Statistical Model of Fusion-Fission Processes. PLoS ONE 11(10): e0163842. doi:10.1371/journal.pone.0163842

Editor: Boris Podobnik, University of Rijeka, CROATIA

Received: June 23, 2016

Accepted: September 15, 2016

Published: October 5, 2016

Copyright: © 2016 Teh, Cheong. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: We are not at liberty to share raw tick-by-tick data downloaded from the Thomson-Reuter TickHistory (TRTH) database (<https://tickhistory.thomsonreuters.com>), because of the terms of our subscription. Readers who would like to study the same raw data set that we did should subscribe to the TRTH database. However, we are happy to provide processed data, which includes continuous returns, forecasted crash times corresponding to different t_{Start} and t_{End} , and market risk heat map. These data were uploaded as supplementary document as [S1 Data](#).

Abstract

The Global Financial Crisis of 2007-2008 wiped out US\$37 trillions across global financial markets, this value is equivalent to the combined GDPs of the United States and the European Union in 2014. The defining moment of this crisis was the failure of Lehman Brothers, which precipitated the October 2008 crash and the Asian Correction (March 2009). Had the Federal Reserve seen these crashes coming, they might have bailed out Lehman Brothers, and prevented the crashes altogether. In this paper, we show that some of these market crashes (like the Asian Correction) can be predicted, if we assume that a large number of adaptive traders employing competing trading strategies. As the number of adherents for some strategies grow, others decline in the constantly changing strategy space. When a strategy group grows into a giant component, trader actions become increasingly correlated and this is reflected in the stock price. The fragmentation of this giant component will lead to a market crash. In this paper, we also derived the *mean-field market crash forecast equation* based on a model of fusions and fissions in the trading strategy space. By fitting the continuous returns of 20 stocks traded in Singapore Exchange to the market crash forecast equation, we obtain crash predictions ranging from end October 2008 to mid-February 2009, with early warning four to six months prior to the crashes.

Introduction

Today, the financial markets have become globally interconnected. Market players frequently maintain portfolios over many asset classes as an insurance against idiosyncratic risks. This raises concerns about susceptibility to systemic risks, and consequently the sustainability of the financial systems. This is worrisome, particularly for the central banks, regulators, and financial institutions. The Global Financial Crisis of 2007-2008 wiped out 34 trillion US dollars of value across financial markets around the world [1]. This is equivalent to the combined GDPs of the United States (US\$17.42 trillions [2]) and the European Union (US\$18.45 trillions [3]) in

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

2014. The defining moment of this crisis was the failure of Lehman Brothers in September 2008, which sent shock waves through financial markets around the world in the form of global market crashes namely the October 2008 crash and the Asian Correction (March 2009).

Had the Federal Reserve seen these crashes coming, they might have worked harder to bail out Lehman Brothers, and prevented the crashes altogether. Financial economists largely believe that financial crises are due to exogenous factors, and are hence not predictable. We feel that this opinion reflects more on the limitations of the models chosen by the economic community, rather than the task being fundamentally impossible. For instance, statistically stationary model [4–6] cannot produce booms and busts, whereas time series model with stochastic jumps [7–9] produces unpredictable booms and busts. Beyond statistical modeling, Palmer et al. introduced in 1994 the Santa Fe Stock Market Model [10], and showed that the artificial market can crash in the absence of exogenous shocks. We expect precursory signatures to exist for the endogenous booms and busts in the Santa Fe Stock Market Model, but no one has looked into this systematically. Alternatively, we can also try model-free forecasting of extreme market events using only empirical data [11]. The most successful of these is the *Log Periodic Power Law* (LPPL) method [12, 13], which was originally developed by Didier Sornette in 1996 [14] to forecast earthquakes. As many have likened financial crashes to earthquakes [15–17], Johansen and Sornette fitted the LPPL to stock market indices, and found that the real market crashes times are very close to the finite-time singularities of the fitted LPPL. More importantly, they concluded that 25 out of the 49 crashes identified exhibit LPPL signatures and thus be classified as endogenous crashes [18].

Inspired by a Soup-of-Groups (SOG) model description of statistical fusion-fission processes in fault planes, Cheong et al. were able to forecast the timings, magnitudes, and locations of large Taiwanese earthquakes [19]. Furthermore, in a recent paper [20, 21], we tracked robust clusters of stocks in the Singapore Exchange over 2008 and 2009, and found that these routinely merged to form larger clusters, and also disintegrated into smaller clusters. In particular, one of the clusters grew steadily into a giant cluster as we approached the October 2008 crash. As stock price movement is governed by market participants, this trading strategy clusters picture is similar to the “opinion convergence and divergence” proposed by Lin et al. [22]. Market participants read and react to market signals, during normal period they have diverse opinions, but they behave coherently when there are large price swings. These findings demonstrate that a financial market naturally self-organizes into clusters of strongly-correlated stocks (and trading strategy with similar opinions) that undergo fusions and fissions. Due to the coexistence of fusion-fission processes in both fault planes and trading strategies space, we wanted to duplicate our success with forecasting earthquakes to forecast financial crashes using the SOG model.

The Soup-of-Groups Model

The SOG model was first introduced by Neil Johnson et al. to study the emergent behavior in a vastly different complex systems. It was first applied to model the underlying dynamics of human insurgency [23], where terrorists are assumed to form groups that merge with each other in preparation for attacks, and disintegrate after a successful attack or to avoid detection by security forces. By comparing the model against a database of past terrorist attacks, Johnson et al. found that the terrorist groups are equally effective, and the number of casualties depends only on the group size, which is distributed as a power law with exponent $\alpha = 2.5$. In the follow up paper [24] they suggested that the escalation rate and timing of the fatal attacks from terrorists follow naturally from SOG dynamics. In addition, Johnson et al. also tried to explain

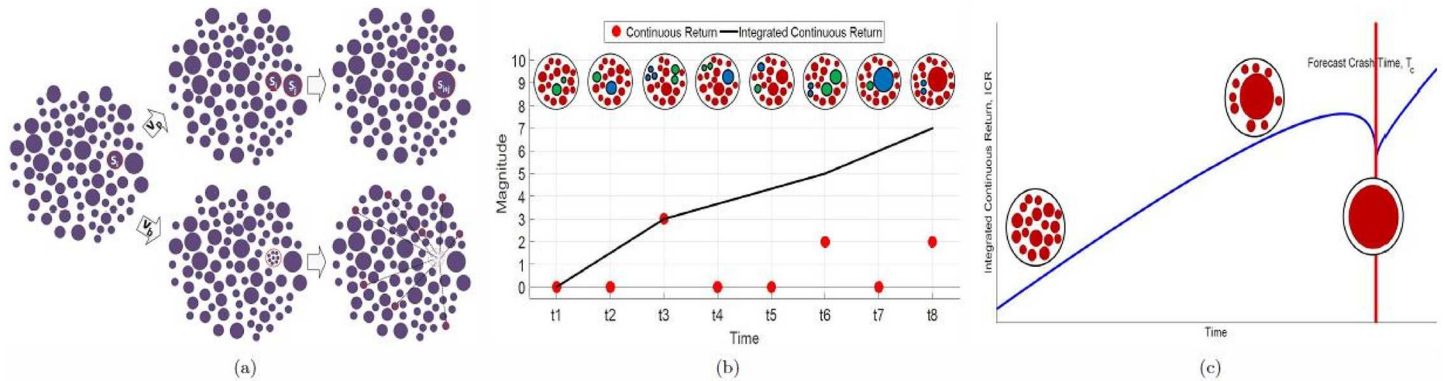


Fig 1. (a) Illustration of Soup-of-Groups (SOG) model, as a system of clusters with various sizes where clusters can undergo fusion at a rate of v_p or fission at a rate of v_b . The upper half represents the fusion process as two clusters with size s_i and s_j are selected and fused into a larger cluster with size $s_i + s_j$. In contrast, during the fission process (lower half) the selected cluster with size s_k will experience complete fragmentation into s_k size-1 clusters. (b) A schematic showing how SOG fusion and fission processes gives rise to Continuous Return (CR) events. Here, the selected cluster is colored green and the resultant clusters after fusion or fission process are colored blue. Each CR is due to a cluster fragmentation with a magnitude proportional to the cluster size. In contrast, a fusion process generates no CR but makes the cluster bigger, and possibly into a giant cluster. In (c) we show the curve generated by Eq (2) that illustrates how the Integrated Continuous Return (ICR) change when a giant cluster is present and growing. Initially, in the absent of giant cluster, the ICR grows linearly due to fragmentation of clusters with an equilibrium distribution of sizes. However, when a giant cluster forms at the expense of other clusters, we start to get only fragmentations of smaller clusters and hence a slow down in ICR growth. Finally at t_c where the giant cluster has exhausted all the resources, and reached the maximum allowed size, it must fragment completely in the next time step.

doi:10.1371/journal.pone.0163842.g001

contagion dynamics on social networks in terms of the SOG model, particularly in the intermediate regime that relates individual behaviors to social group structures [25].

The SOG model, as the name suggests, consists of a “soup” (system) of “groups” (clusters) with various sizes. The essence of SOG dynamics is the fusion-fission processes among clusters. For instance, a cluster of size s_i can merge with a cluster of size s_j to give a cluster with size $s_i + s_j$ at a rate of v_p , or a cluster of size s_k can fragment completely into s_k clusters with size 1 at a rate of v_b , see Fig 1(a) for illustration and refer to the S2 File for more details. In an infinite element equilibrium system, Neil Johnson et. al. discovered that the equilibrium distribution for the SOG model is an *Exponential Truncated Power Law* (ETPL) distribution [26] as

$$N_s = f(S) \sim s^{-\frac{5}{2}} \exp(-S_0 s); \quad S_0 = -\ln \left(\frac{4(v_p + v_b)v_p}{(2v_p + v_b)^2} \right).$$

SOG Forecasting

In finance, one of the most widely used definition for quantifying financial returns is the simple return (and log return). However, the distribution of financial returns have been found to depend on the choice of time horizon, making such returns ambiguous [27]. To get around this ambiguity, we introduce *continuous returns* defined as the continuous price movement in the same direction (refer to the subsection Continuous returns in the Data and Methods section). This definition of an event agrees with that used by da Cruz and Lind [28]. In financial markets, market participants are heterogeneous, so even though they read the same market signals they may arrive at different trading strategies. However, if the market signal is strong, these traders may arrive at the same trading strategies, thereby reinforcing the market signals,

leading to the formation of strategies blocks. When a strategy block is executed, market participants within the strategy block will react in the same way, hence generating a continuous return with a magnitude that depends on the size of the strategy block.

In this study, we focus on data from the Singapore Exchange (SGX) because we are familiar with stocks listed in the SGX. The SGX is also strongly coupled to financial markets around the world, and readily reflects the global market movement. We therefore test the SOG forecasting method for 20 component stocks of the Straits Times Index (STI) over the period January 2006 to December 2011, which is approximately three years before the October 2008 crash and three years after the October 2008 crash (read subsection Data in section [Data and Methods](#)). Since the SOG dynamics is universal, we expect it to be able to do forecasting with the model even when the stock price dynamics is only approximately SOG. To demonstrate this approximate SOG behavior we check the distribution of the continuous returns against the ETPL distribution. For comparison, we also fit the empirical data to the power law (PL) and asymptotic exponential (EXP) distribution using the method developed by Clauset et al. to estimate these parameters (\hat{x}_{min} , $\hat{\alpha}$, and $\hat{\beta}$) [29]. A detail description of the fitting procedure can be found in the [S4 File](#). The fitting results from section *Distribution of Continuous Returns* in [S3 File](#) indicate that the continuous returns for 20 stocks are best fitted to the ETPL, as the PL tends to decay slower than the data, while EXP tends to decay faster than the data. Overall, the average $\hat{\alpha}_{ETPL}$ and $\hat{\beta}_{ETPL}$ across all stocks is 1.82 ± 0.23 and 0.37 ± 0.08 respectively. The reported $\hat{\alpha}_{ETPL}$ is lower when compared with the SOG exponent of $\alpha = 2.5$. Even though we can assume that the continuous return is directly proportional to the size of the cluster, yet the result shows this relationship does not hold. But, if we assume that a cluster of size s produces a continuous return of s^ζ , where $\zeta = 2.5/\alpha$, then the SGX will indeed be a SOG system. We will use this modified scaling to formulate the forecasting equation ([Eq \(2\)](#)).

In the SOG model each continuous return is caused by a fragmentation of the strategy block, therefore a market crash is associated with the fragmentation of a giant cluster. [Fig 1\(b\)](#) illustrates the relation between the continuous return and the SOG fusion and fission process. During the fusion process two clusters merge and form a larger cluster, but generate no continuous return. On the other hand, each fission process will generate a continuous return. In these descriptions, a market crash is due to the fragmentation of a giant cluster (giant strategy block) that rapidly drives the price downwards. Thus if we can track the size and growth of the giant cluster we should be able to deduce the time when the market will crash. This can be done by determining the time when the giant cluster will reaches its maximum size. This time acts as the upper bound time for fragmentation of the giant cluster, thus the forecasted crash time. Unfortunately, we cannot directly observe the giant cluster size in the financial market, since we cannot simultaneously track the opinions and strategies of all market participants. Instead, we infer the giant cluster size based on the continuous returns, which becomes smaller as the giant cluster grows bigger and exhausts the larger clusters around it. This manifests itself as a slowing down in the growth of the integrated continuous returns.

In order to build the forecasting model, the inputs are: the size of the giant cluster $s_G(t)$, the size distribution of the rest of clusters $f(s)$, as well as the relations between the magnitude of the continuous return and the cluster size $CR_t = CR_t(s_t)$. We make several assumptions regarding these inputs. First, we assume that the giant cluster grows linearly with time as $s_G(t) = \gamma t$. Second, instead of using the ETPL distribution, which is difficult to work with, we use the simplified PL distribution $f(s) = As^{-2.5}$. Lastly, we assume that the continuous return is proportional to the cluster size, given as $CR_t \propto s_t^\zeta$ in order for the dynamics of continuous return to be approximately SOG. Based on these assumptions, we can write down the expected integrated

continuous return $\hat{ICR}(0, t)$ from $time = 0$ to $time = t$ as

$$\begin{aligned} \hat{ICR}(0, t) &= \int_0^t dt \int_{s_0}^{s_{max}-s_G(t)} ds A s^\zeta f(s) = \int_0^t dt \int_{s_0}^{s_{max}-s_G(t)} ds A s^{\zeta-2.5} \\ &= \frac{A}{\zeta - 1.5} \left\{ \frac{1}{\gamma(0.5 - \zeta)} \left[(s_{max} - \gamma t)^{\zeta-0.5} - s_{max}^{\zeta-0.5} \right] - s_0^{\zeta-1.5} t \right\}. \end{aligned} \tag{1}$$

Assuming the giant cluster started growing at t_o , Eq (1) becomes

$$\begin{aligned} \hat{ICR}(t_o, t) &= \\ &= \frac{A}{\zeta - 1.5} \left\{ \frac{1}{\gamma(\zeta - 0.5)} \left[(S_{max} - \gamma\{t - t_o\})^{\zeta-0.5} - S_{max}^{\zeta-0.5} \right] - S_0^{\zeta-1.5} \{t - t_o\} \right\} + ICR_o. \end{aligned} \tag{2}$$

Eq (2) is a mean-field forecasting equation that we show in Fig 1(c). From Fig 1(c), we observe that before the giant cluster started growing, $\hat{ICR}(t)$ grows linearly with time. As the giant cluster grows, the growth rate of $\hat{ICR}(t)$ decreases. More importantly, a singularity occurs when the giant cluster reaches the maximum size of $S_G(t_c) = S_{max} = \gamma(t_c - t_o)$. We call this time the *forecasted crash time* (t_c) which is

$$t_c = \frac{S_{max}}{\gamma} + t_o. \tag{3}$$

From the data, we define the empirical *integrated continuous returns* from t_{Start} to t_{End} as

$$ICR(t_{Start}, t_{End}) = \sum_{k=i}^{N; t(N) \leq t_{End}; t(i) \geq t_{Start}} CR_{t(k)}, \tag{4}$$

which is the cumulative sum of the continuous returns. We fit the empirical integrated continuous returns $ICR(t_{Start}, t_{End})$ to our forecasting model $\hat{ICR}(t_{Start}, t_{End})$ (Eq (2)) to obtain the forecasted crash time t_c . For the fitting, we use the non-linear least square method in MATLAB to estimate the parameter set that minimizes the overall residuals. In practice, we must only use data available up to the present moment. Therefore to mimic a real-time forecasting using historical data, we work with dynamic fitting windows starting at t_{Start} and ending at t_{End} , where t_{End} must be before the market crash we want to ‘forecast’. To do this, we fix a t_{Start} then we start with t_{End} one month forward and increase t_{End} 2 weeks at a time. We then obtain t_c as a function of t_{End} for a given t_{Start} , $t_c = t_c(t_{End}|t_{Start})$.

Results

Our objective is to use the SOG model to detect early warnings of market crashes by forecasting the crash times of the STI component stocks. This is a different approach compared to traditional finance and econometric models, wherein sudden changes are assumed to be brought about by exogenous shocks. In the SOG model, a market crash is endogenous, driven by the giant cluster formation and its fragmentation. Within the period that we study, from Jan 2006 to Dec 2011, two global financial crashes left their marks on the SGX. As shown in S Fig 1 in S1 File, the October 2008 crash and the Asian Correction that caused the STI to tumble to its lowest at 1600.29 (27 Oct 2008, October 2008 crash), with a slight recovery after that and further sliding to 1456.29 (9 Mar 2009, Asian Correction). The first crash was larger, thus we select 27 Oct 2008 as our reference date for the market crash.

We fix the start t_{Start} of the fitting window as 15 Dec 2006, which is about two years before the crash. We then slowly expand the fitting window by increasing the t_{End} of the fitting

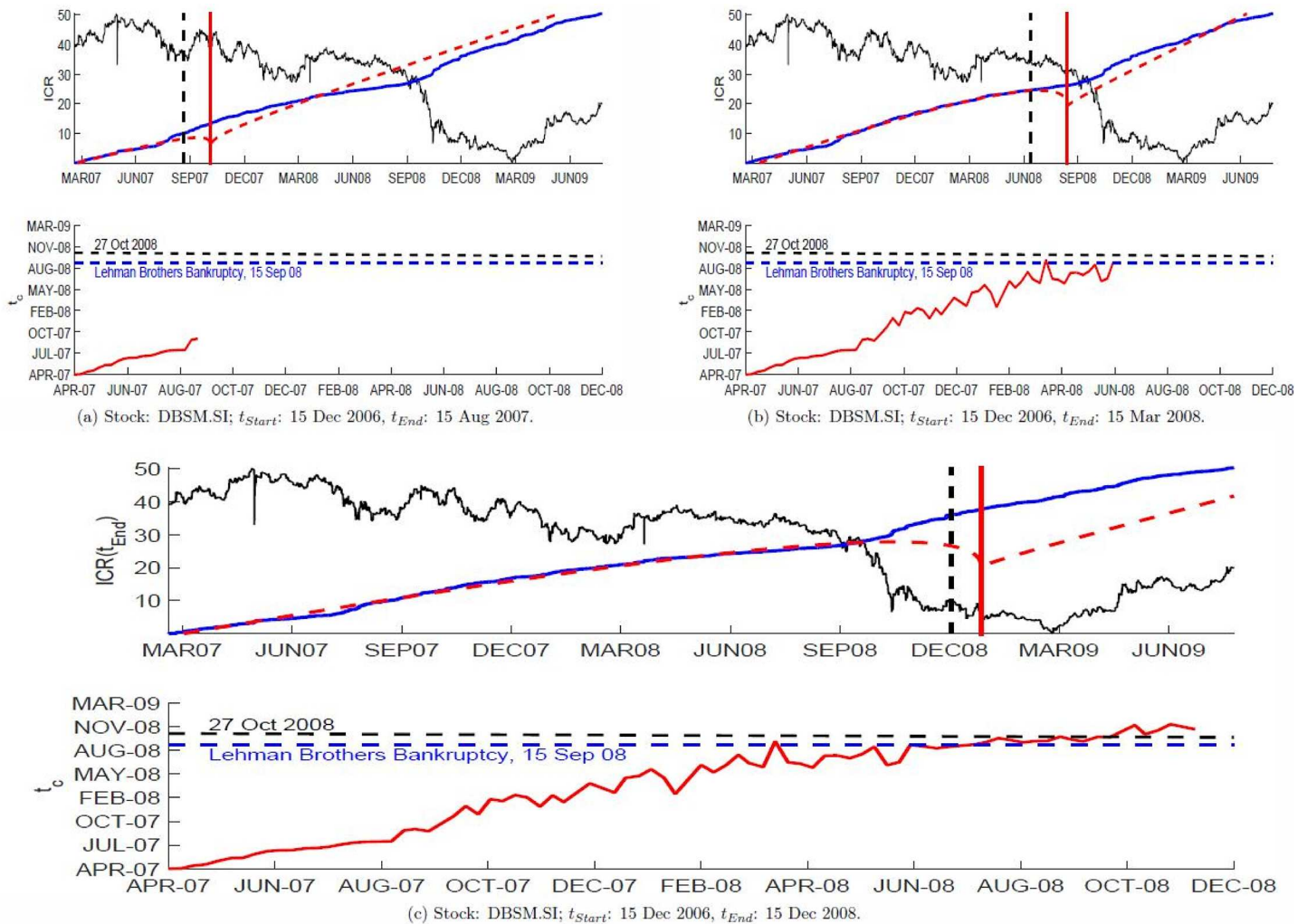


Fig 2. Empirical Integrated Continuous Return ICR fitted to the forecasting model \hat{ICR} (Eq (2)). For illustration purposes, we use DBSM.SI and a t_{Start} of 15 Dec 2006. t_{End} is then slowly increased with an interval of two weeks. The upper half of each sub-figure shows the empirical ICR (blue) and the fitted \hat{ICR} (red dashed-curve), and the rescaled stock price is shown as the background (black). In addition, the black vertical dashed line represents the t_{End} and red vertical line indicates the corresponding t_c . The lower half shows the plot of $t_c(t_{End})$ as a function of t_{End} , where the Lehman Brothers Bankruptcy (15 Sept 2008) is represented by a blue horizontal dashed line and 27 October 2008 is represented by a black horizontal dashed line.

doi:10.1371/journal.pone.0163842.g002

window. Fig 2(a)–2(c) show a series of fits for t_{End} as 15 Aug 2007, 15 Mar 2008, and 15 Dec 2008. In these figures, the black vertical line represents t_{End} and the red vertical line represents the corresponding forecasted crash time t_c . In the lower half of each sub-figure, we plot the corresponding t_c against t_{End} up to t_{End} as 15 Aug 2007, 15 Mar 2008, and 15 Dec 2008 with 15 Dec 2006 as the t_{Start} . From the graph, we can see that when far before the October 2008 crash, t_c is increasing linearly with t_{End} . This seems to suggest that the SGX is always critical, and ready to crash. This agrees with the suggestions that stock markets are self-organized critical (SOC) systems [30, 31], as well as findings from calibrating agent-based models in stock markets [32].

As we approach the October 2008 crash, we see t_c growing at a slower rate and finally stagnating around the actual crash time ($t_{Act} = 27$ Oct 2008). This stagnation occurs a few months

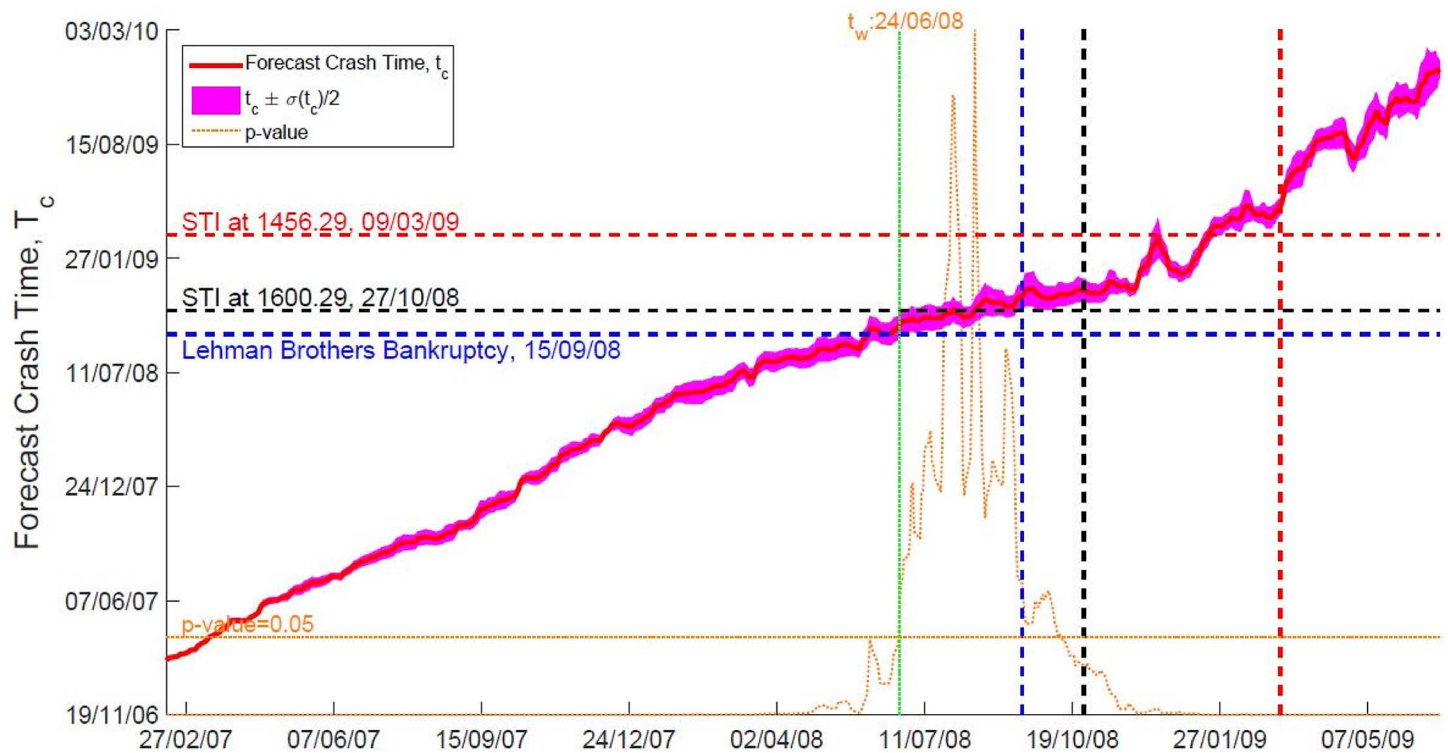


Fig 3. In this figure we show the sensitivity test for stock DBSM.SI, where the blue dashed line represents the date of Lehman Brothers Bankruptcy (15 Sept 2008), black dashed line represents the our benchmark date-27 Oct 2008 when the STI at its lowest due to October 2008 crash, and red dashed line marks the date 9 Mar 2009 that the STI at its lowest due to Asian Correction. We plot the weighted average forecast crash time t_c against t_{End} showing the weighted standard deviation as the spread. We also show the p -value for our null hypothesis $H_0: t_c(t_{End}) = 27$ Oct 2008 cutoff at 95% confidence level (orange curve).

doi:10.1371/journal.pone.0163842.g003

before t_{End} reaches t_{Act} . This is an interesting early warning signature that we must look out for in a system that is already bordering on criticality. This result gives a positive outlook for applying the SOG model to forecast market crashes, but rigorous tests are needed to establish the forecasting power of the SOG model.

Discussion

We perform a sensitivity test on the forecasting results by comparing the forecasted crash time t_c with the October 2008 crash, actual crash time $t_{Act} = 27$ Oct 2008. Given a t_{Start} , t_c is a function of t_{End} as $t_c = t_c(t_{End}|t_{Start})$. We can test the sensitivity of $t_c(t_{End}|t_{Start})$ to t_{End} using 14 t_{Start} 's at the beginning and middle of the months from 1 Nov 2006 to 16 May 2007. We compare the weighted mean of forecasted crash time, $\bar{t}_c(t_{End})$ s with the actual crash time $t_{Act} = 27$ Oct 2008 at 95% confidence level, using the null hypothesis $H_0: \bar{t}_c = t_{Act}$ (see the subsection on Sensitivity analysis in [Data and Methods](#) section). Robust signatures for the market crash forecast can be observed in [Fig 3](#): \bar{t}_c increases linearly with t_{End} when we are far from the market crash, and thereafter stagnates around t_{Act} when t_{End} is closer to the October 2008 crash. For earlier t_{End} 's the predicted \bar{t}_c is statistically different from t_{Act} , whereas for t_{End} close to the actual crash there is no longer statistical difference between \bar{t}_c and t_{Act} .

The t_{End} when t_c first becomes statistically indistinguishable from t_{Act} is the *warning time*, t_w , which in turn acts as the early warning signal for market crashes. In addition, in order to determine the accuracy of the SOG forecasting result, we calculate the average of the weighted mean of forecasted crash time $\langle \bar{t}_c(t_{End}) \rangle$: average of $\bar{t}_c(t_{End})$ for the t_{End} 's between the warning time t_w and the actual crash time t_{Act} (27 Oct 2008). If the $\langle \bar{t}_c(t_{End}) \rangle$ calculated is close to t_{Act} , it means the forecasting result is accurate. The results for all 20 component stocks are listed in the S Table 1 in S5 File, there is a range of four to six months of early warning prior the actual crash. Apart from the component stock DBSM.SI with $\langle \bar{t}_c(t_{End}) \rangle$ falling on 29 Oct 2008, the others predicted crash dates are after t_{Act} , as late as 12 Feb 2009. This is to be expected, as the SOG model predicts the latest possible crash time. However, in reality it is possible for a giant cluster to disintegrate before it reaches its maximum size.

In addition to the latest possible crash date for individual stocks that we are tracking, we can also combine the forecasting results from all stocks into an index to estimate the market-level risk. For example, when the majority of the STI component stocks predict a crash simultaneously on a particular date, the market risk heightens on that particular date. In order to do this cross sectional study, we first count the number of STI component stocks forecasted to crash on a particular day, for various t_{Start} 's and t_{End} 's. Going back to Fig 2, we observe that when we are far from the actual crash, the forecasted crash date increases linearly with t_{End} and thus we get a uniform background count. In contrast, when t_{End} is close to t_{Act} , t_c stagnates leading to a high concentration of forecasted crashes in a narrow range of dates and hence a large number of counts on those dates. We show these counts as a heat map to visualize the market risk, where a date with large counts hence high market risk is shown as a red pixel. In contrast, a blue pixel means background level of counts with low market risk.

On the left of Fig 4, we vary t_{End} over the period Jan 2007 to Jun 2009. For each stock, we accumulate the forecasted crash time t_c into a histogram and these histograms are then represented as a heat map. In principle if we are at time t , we will not know the forecasted crash time for $t_{End} > t$. Thus on the right of Fig 4, we show the average counts for all 20 STI component stocks, but only vary t_{End} from Jan 2007 up to t . We see that there are four red bands for t_c peaking at 25 Jan 2007, 14 Aug 2007, 14 Oct 2007, and 5 Mar 2009. These episodes of heightened market risks are close to the lowest of STI due to the Chinese Correction (27 Feb 2007), begin of the landslide of the Subprime Crisis (17 Aug 2007), and the Asian Correction (9 Mar 2009). However, there is no heightened market risk near the collapse of Lehman Brothers bankruptcy (15 Sep 2008). We believe this is due to the October 2008 crash and the Asian Correction appearing as a single extended market crash to the forecasting model. More interestingly, early warning signals did appear a few months before the October 2008 crash, thus providing us the opportunity to prepare for the financial market crashes.

All in all, these results show that the SOG forecasting model provides early warning in the form of slowing down of the integrated continuous returns. Furthermore the SOG model is also able to indicate the tension level for potential financial crashes via the market risk heat map. Together with other standard critical slowing down indicators like increasing volatility, increasing autocorrelation, red shift in power spectrum, and increasing cross-correlation [33–35], the combination of indicators can act as financial crisis early warning alarm for authorities, such that measures can be taken to prevent or at least soften the impact of financial crashes. Still, there is room for improvement, for example the false negative error for the October 2008 crash compared to Asian Correction. We do not expect that the SOG model to be the best model for describing market fusion-fission dynamics, particularly the complete fragmentation required by the SOG model. We believe partial fragmentation of clusters will be more realistic, and we are working to the test forecasting method based on SOG-like models. In summary, the SOG model forecasting is promising enough to merit further exploration.

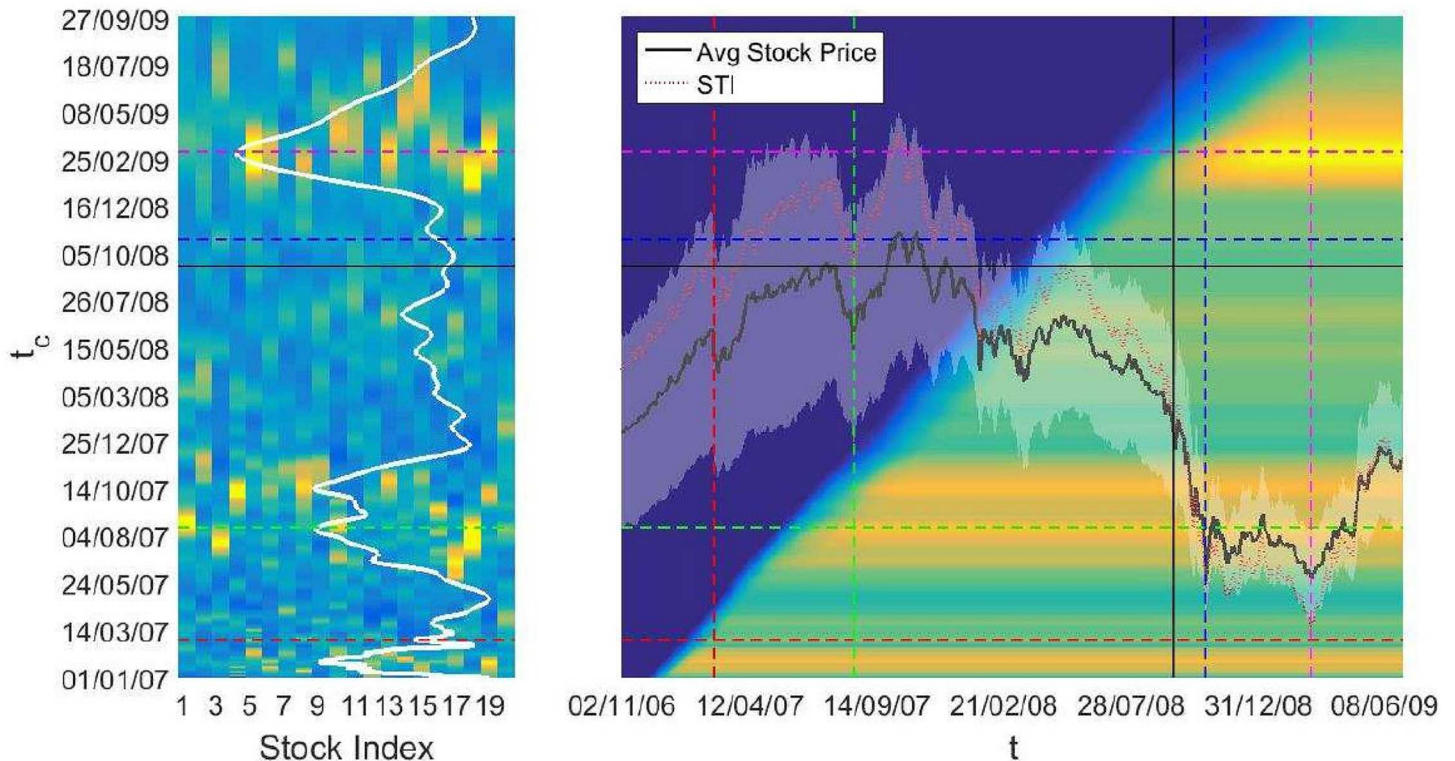


Fig 4. Market risk heat map of the SGX. When a majority of the stocks simultaneously forecast a crash on a particular date, the market risk is high on that date. In this figure, a red pixel indicates a high market risk while the background risk level (low stress) is shown as blue pixels. For both figures, the dash horizontal lines represent the lowest of STI due to: (red) 27 Feb 2007 Chinese Correction, (green) begin of landslide of Subprime Crisis-17 Aug 2007, (blue) Lehman Brothers crisis-9 Oct 2008, (magenta) Asian Correction-9 Mar 2009, and the black solid horizontal line shows the collapse of Lehman Brothers, 15 Sep 2008. [Left] Heat map for all 20 stock with t_{End} running from Jan 2007 to Jun 2009. The white curve represents the sum of risk level for all stocks, which peaks at 25 Jan 2007, 14 Aug 2007, 14 Oct 2007, and 5 Mar 2009. [Right] Sum of all 20 stocks risk level with t_{End} running from Jan 2007 to time t . The mean and standard deviation of daily price for all stocks is shown as the black curve with white band. The increase in risk level can be seen a few months prior to the actual crash time, and thus can act as an early warning for market crashes.

doi:10.1371/journal.pone.0163842.g004

Data and Methods

Data

In this study, we focus on data from the Singapore Exchange (SGX) because we are familiar with stocks listed in the SGX, in contrast to model markets (e.g. the London Stock Exchange (LSE) and the New York Stock Exchange (NYSE)) that are often studied by econophysicists. Although the SGX is an emerging market, it is strongly coupled to financial markets around the world, and readily reflects the global market movement. For instance, the Straits Times Index (STI) is highly correlated with the Dow Jones Index ($\rho \approx 0.91$), and slid more than 45% between August 2008 and October 2008 as the result of the October 2008 crash (see S Fig 1 in S1 File). The STI is made up of the top 30 stocks in SGX, but these components do change from time to time. Thus we only consider 20 component stocks that remain highly traded across January 2006 to December 2011. See S Table 1 in S1 File for the full list of these 20 stocks.

The tick-by-tick data was downloaded from the Thomson Reuters Tick History database (<http://thomsonreuters.com/tick-history>). On an average the number of transactions per stock

is on the order of 10^6 over 1506 trading days. Unlike LSE and NYSE, the SGX is relatively illiquid as the median time interval between transactions ranges from 6 s for the most liquid stock to 47 s for the least liquid stock (refer to S Table 1 in S1 File).

Continuous returns

We define *continuous return* as the continuous price movements in the same direction, and mathematically the continuous return $CR_{t(m)}$ is given as

$$CR_{t(m)} = \frac{|P_{t(m+n)} - P_{t(m)}|}{P_{t(m)}}, \tag{5}$$

that satisfies all conditions simultaneously

$$\begin{aligned} (P_{t(m-1)} - P_{t(m)}) \times (P_{t(m+n)} - P_{t(m+n+1)}) &> 0; \\ (P_{t(m-1)} - P_{t(m)}) \times (P_{t(m)} - P_{t(m+k)}) &< 0; \\ (P_{t(m+n)} - P_{t(m+n+1)}) \times (P_{t(m+n-k)} - P_{t(m+n)}) &< 0, \end{aligned}$$

where $P_{t(i)}$ is the i^{th} transaction price occurs at time $t(i)$, $k \in 1, 2, 3 \dots n$, and at time t_{m+n} the direction of the price movement changes. S Fig 3 in S3 File illustrates a segment of stock price time series, where the continuous return is the fractional price change between two red vertical lines, denoting the start and end of a microtrend [36]. Over the study period, Jan 2006 to Dec 2011 the number of continuous returns on average across stocks is on the order of 10^4 , and the median of microtrend for different stocks ranged from 10 mins to 50 mins (see S Table 1 in S3 File).

Sensitivity analysis

We perform a sensitivity test of the forecasting results by comparing the weighted mean forecasted crash time \bar{t}_c with the actual crash time t_{Act} . Given a t_{Start} , t_c is a function of t_{End} as $t_c = t_c(t_{End}|t_{Start})$. We can test the sensitivity of $t_c(t_{End}|t_{Start})$ to t_{End} , by calculate the weighted mean of t_c at particular t_{End} using various t_{Start} 's. The weighted mean and weighted standard deviation of $t_c(t_{End})$ is then calculated for each respective t_{End} as

$$\begin{aligned} \bar{t}_c(t_{End}) &= \sum_{i=1}^N w(t_{Start}^i | t_{End}) t_c(t_{End} | t_{Start}^i); \\ \sigma^2(t_c(t_{End})) &= \frac{\sum_{i=1}^N w(t_{Start}^i | t_{End}) [t_c(t_{End} | t_{Start}^i) - \bar{t}_c(t_{End})]^2}{N}; \end{aligned} \tag{6}$$

where the weight is defined as

$$w(t_{Start}^i | t_{End}) = \frac{NumDays(t_{Start}^i, t_{End})}{\sum_{i=1}^N NumData(t_{Start}^i, t_{End})}.$$

In Eq (6) the $NumDays(t_1, t_2)$ is the number of trading days between t_1 and t_2 , and weight $w(t_{Start}^i | t_{End})$ is introduced to take into account the different amount of data being used to do the forecast. We compare $\bar{t}_c(t_{End})$ with the t_{Act} at 95% confidence level, using the null

hypothesis $H_o : \bar{t}_c = t_{Act}$, where the t -value and p -value for n observations are calculated as

$$t_{value}(t_{End}) = \frac{|\bar{t}_c(t_{End}) - t_{Act}|}{\sigma(t_c(t_{End}))},$$

$$p_{value}(t_{End}) = 1 - \int_{-\infty}^{t_{value}(t_{End})} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \frac{1}{\sqrt{n\pi}} \frac{1}{(1 + \frac{x^2}{n-1})^{\frac{n}{2}}} dx, \quad (7)$$

where $\Gamma(x)$ is the gamma function. When $p < 0.05$, t_c is significantly different from t_{Act} at 95% confidence level, and we reject H_o . Otherwise, we fail to reject H_o when p_{value} is greater than 0.05.

Supporting Information

S1 File. Supplementary Document: Data, provides the details about data we use in this study.

(PDF)

S2 File. Supplementary Document: Soup-of-Groups Model, provides the details about the Soup-of-Groups model introduced by Johnson et al.

(PDF)

S3 File. Supplementary Document: Continuous Returns, explains the details about Continuous Returns and and shows the continuous return distribution fits with PL, EXP, and ETPL distribution.

(PDF)

S4 File. Supplementary Document: PL, EXP, and ETPL distribution fitting, explains the fitting process for PL, EXP, and ETPL distribution.

(PDF)

S5 File. Supplementary Document: Warning Time and Average Forecasted Crash Time, indicates the forecasting result for the 20 component stocks.

(PDF)

S1 Data. Processed Data. We are not at liberty to share raw tick-by-tick data downloaded from the Thomson-Reuter TickHistory (TRTH) database, because of the terms of our subscription. Readers who would like to study the same raw data set that we did should subscribe to the TRTH database. However, we are happy to provide processed data, which includes continuous returns, forecasted crash times corresponding to different t_{Start} and t_E nd , and market risk heat map. These data were uploaded as supplementary document as S1 Data.

(ZIP)

Acknowledgments

S.A.C and B.K.T acknowledge Lock Yue Chew and Charlie Charoenwong for discussions and constructive comments.

Author Contributions

Conceptualization: SAC.

Formal analysis: BKT.

Investigation: SAC BKT.

Methodology: SAC BKT.

Supervision: SAC.

Validation: SAC.

Visualization: BKT.

Writing – original draft: BKT.

Writing – review & editing: SAC.

References

1. Liu HCK. The Crisis of Wealth Destruction. Economy and Growth, Finance and Wealth, Roosevelt Institute. 2010;.
2. The World Bank. United States; 2014.
3. The World Bank. European Union; 2014.
4. Akgiray V. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *Journal of Business*. 1989; 62(1):55–80. doi: [10.1086/296451](https://doi.org/10.1086/296451)
5. Bollerslev T, Chou RY, Kroner KF. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*. 1992; 52(1):5–59. doi: [10.1016/0304-4076\(92\)90064-X](https://doi.org/10.1016/0304-4076(92)90064-X)
6. Hamilton JD, Susmel R. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*. 1994; 64(1):307–333. doi: [10.1016/0304-4076\(94\)90067-1](https://doi.org/10.1016/0304-4076(94)90067-1)
7. Christie AA. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*. 1982; 10(4):407–432. doi: [10.1016/0304-405X\(82\)90018-6](https://doi.org/10.1016/0304-405X(82)90018-6)
8. Carr P, Geman H, Madan DB, Yor M. Stochastic volatility for Lévy processes. *Mathematical Finance*. 2003; 13(3):345–382. doi: [10.1111/1467-9965.00020](https://doi.org/10.1111/1467-9965.00020)
9. Lee SS, Mykland PA. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*. 2008; 21(6):2535–2563. doi: [10.1093/rfs/hhm056](https://doi.org/10.1093/rfs/hhm056)
10. Palmer RG, Arthur WB, Holland JH, LeBaron B, T aylor P. Artificial economic life: a simple model of a stockmarket. *Physica D: Nonlinear Phenomena*. 1994; 75(1):264–274. doi: [10.1016/0167-2789\(94\)90287-9](https://doi.org/10.1016/0167-2789(94)90287-9)
11. Quax R, Kandhai D, Sloot PM. Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series. *Scientific reports*. 2013; 3. doi: [10.1038/srep01898](https://doi.org/10.1038/srep01898) PMID: [23719567](https://pubmed.ncbi.nlm.nih.gov/23719567/)
12. Johansen A, Sornette D. Financial “anti-bubbles”: Log-periodicity in gold and Nikkei collapses. *International Journal of Modern Physics C*. 1999; 10(04):563–575. doi: [10.1142/S0129183199000437](https://doi.org/10.1142/S0129183199000437)
13. Sornette D, Johansen A. Significance of log-periodic precursors to financial crashes. *Quantitative Finance*. 2001; 1(4):452–471. doi: [10.1088/1469-7688/1/4/305](https://doi.org/10.1088/1469-7688/1/4/305)
14. Saleur H, Sammis C, Sornette D. Discrete scale invariance, complex fractal dimensions, and log-periodic fluctuations in seismicity. *Journal of Geophysical Research-All series*. 1996; 101(BB):17661–17677. doi: [10.1029/96JB00876](https://doi.org/10.1029/96JB00876)
15. Sornette D. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proceedings of the National Academy of Sciences*. 2002; 99(suppl 1):2522–2529. doi: [10.1073/pnas.022581999](https://doi.org/10.1073/pnas.022581999)
16. Lillo F, Mantegna RN. Power-law relaxation in a complex system: Omori law after a financial market crash. *Physical Review E*. 2003; 68(1):016119. doi: [10.1103/PhysRevE.68.016119](https://doi.org/10.1103/PhysRevE.68.016119)
17. Chakrabarti BK, Chatterjee A, Bhattacharyya P. Two-fractal overlap time series: Earthquakes and market crashes. *Pramana*. 2008; 71(2):203–210. doi: [10.1007/s12043-008-0154-5](https://doi.org/10.1007/s12043-008-0154-5)
18. Johansen A, Sornette D. Endogenous versus Exogenous Crashes in Financial Markets. *Social Science Research Network Working Paper Series*. 2002;.
19. Cheong SA, Tan TL, Chen CC, Chang WL, Liu Z, Chew LY, et al. Short-term forecasting of Taiwanese earthquakes using a universal model of fusion-fission processes. *Scientific Reports*. 2014; 4. doi: [10.1038/srep03624](https://doi.org/10.1038/srep03624) PMID: [24406467](https://pubmed.ncbi.nlm.nih.gov/24406467/)
20. Teh BK, Goo YW, Lian TW, Ong WG, Choi WT, Damodaran M, et al. The Chinese Correction of February 2007: How financial hierarchies change in a market crash. *Physica A: Statistical Mechanics and its Applications*. 2015; 424:225–241. doi: [10.1016/j.physa.2015.01.024](https://doi.org/10.1016/j.physa.2015.01.024)

21. Teh BK, Cheong SA. Cluster fusion-fission dynamics in the Singapore stock exchange. *The European Physical Journal B*. 2015; 88(10):1–14. doi: [10.1140/epjb/e2015-60456-y](https://doi.org/10.1140/epjb/e2015-60456-y)
22. Feng L, Li B, Podobnik B, Preis T, Stanley HE. Linking agent-based models and stochastic models of financial markets. *Proceedings of the National Academy of Sciences*. 2012; 109(22):8388–8393. doi: [10.1073/pnas.1205013109](https://doi.org/10.1073/pnas.1205013109)
23. Bohorquez JC, Gourley S, Dixon AR, Spagat M, Johnson NF. Common ecology quantifies human insurgency. *Nature*. 2009; 462(7275):911–914. doi: [10.1038/nature08631](https://doi.org/10.1038/nature08631) PMID: [20016600](https://pubmed.ncbi.nlm.nih.gov/20016600/)
24. Johnson N, Carran S, Botner J, Fontaine K, Laxague N, Nuetzel P, et al. Pattern in escalations in insurgent and terrorist activity. *Science*. 2011; 333(6038):81–84. doi: [10.1126/science.1205068](https://doi.org/10.1126/science.1205068) PMID: [21719677](https://pubmed.ncbi.nlm.nih.gov/21719677/)
25. Zhao Z, Calderón J, Xu C, Zhao G, Fenn D, Sornette D, et al. Effect of social group dynamics on contagion. *Physical Review E*. 2010; 81(5):056107. doi: [10.1103/PhysRevE.81.056107](https://doi.org/10.1103/PhysRevE.81.056107) PMID: [20866297](https://pubmed.ncbi.nlm.nih.gov/20866297/)
26. Johnson NF, Ashkenazi J, Zhao Z, Quiroga L. Equivalent dynamical complexity in a many-body quantum and collective human system. *AIP Advances*. 2011; 1(1):012114. doi: [10.1063/1.3563072](https://doi.org/10.1063/1.3563072)
27. Cont R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*. 2001; 1(2):223–236. doi: [10.1080/713665670](https://doi.org/10.1080/713665670)
28. Da Cruz JP, Lind PG. Self-organized criticality in a network of economic agents with finite consumption. *Physica A: Statistical Mechanics and its Applications*. 2012; 391(4):1445–1452. doi: [10.1016/j.physa.2011.11.012](https://doi.org/10.1016/j.physa.2011.11.012)
29. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM Review*. 2009; 51(4):661–703. doi: [10.1137/070710111](https://doi.org/10.1137/070710111)
30. Scheinkman JA, Woodford M. Self-organized criticality and economic fluctuations. *The American Economic Review*. 1994; 84(2):417–421.
31. Bartolozzi M, Leinweber D, Thomas A. Self-organized criticality and stock market dynamics: an empirical study. *Physica A: Statistical Mechanics and its Applications*. 2005; 350(2):451–465. doi: [10.1016/j.physa.2004.11.061](https://doi.org/10.1016/j.physa.2004.11.061)
32. Lux T, Marchesi M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*. 1999; 397(6719):498–500. doi: [10.1038/17290](https://doi.org/10.1038/17290)
33. Schwert GW. Business cycles, financial crises, and stock volatility. In: *Carnegie-Rochester Conference Series on Public Policy*. vol. 31. Elsevier; 1989. p. 83–125. doi: [10.1016/0167-2231\(89\)90006-7](https://doi.org/10.1016/0167-2231(89)90006-7)
34. Chiang TC, Jeon BN, Li H. Dynamic correlation analysis of financial contagion: Evidence from Asian markets. *Journal of International Money and Finance*. 2007; 26(7):1206–1228. doi: [10.1016/j.jimonfin.2007.06.005](https://doi.org/10.1016/j.jimonfin.2007.06.005)
35. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. *Nature*. 2009; 461(7260):53–59. doi: [10.1038/nature08227](https://doi.org/10.1038/nature08227) PMID: [19727193](https://pubmed.ncbi.nlm.nih.gov/19727193/)
36. Preis T, Schneider JJ, Stanley HE. Switching processes in financial markets. *Proceedings of the National Academy of Sciences*. 2011; 108(19):7674–7678. doi: [10.1073/pnas.1019484108](https://doi.org/10.1073/pnas.1019484108)