THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

A Tale of Two Indexes: Predicting Equity Market Downturns in China

Sébastien Lleo

William T. Ziemba

SRC Discussion Paper No 72

August 2017

Systemic Risk Centre

Discussion Paper Series

ISSN 2054-538X

Abstract
Predicting stock market crashes is a focus of interest for both researchers and practitioners. Several prediction models have been developed, mostly for use on mature financial markets. In this paper, we investigate whether traditional crash predictors, the price-to-earnings ratio, the Cyclically Adjusted Price-to-Earnings ratio and the Bond-Stock Earnings Yield Differential model, predicts crashes for the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Composite Index.

Keywords: equity markets, crashes, China, BSEYD, CAPE.

JEL Classification: G14, G15, G12, G10.

EconLit Subject Descriptors: G140, G150, G120, G100.

This paper is published as part of the Systemic Risk Centre's Discussion Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/K002309/1].

Sébastien Lleo, Finance Department, NEOMA Business School.
William T. Ziemba, University of British Columbia, Vancouver, BC and Systemic Risk Centre, London School of Economics and Political Science.

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

A Tale of Two Indexes: Predicting Equity Market Downturns in China

Sébastien Lleo*and William T. Ziemba†

Abstract

Predicting stock market crashes is a focus of interest for both researchers and practitioners. Several prediction models have been developed, mostly for use on mature financial markets. In this paper, we investigate whether traditional crash predictors, the price-to-earnings ratio, the Cyclically Adjusted Price-to-Earnings ratio and the Bond-Stock Earnings Yield Differential model, predicts crashes for the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Composite Index.

JEL: G14, G15, G12, G10.

EconLit Subject Descriptors: G140, G150, G120, G100.

Keywords: equity markets, crashes, China, BSEYD, CAPE.

*Finance Department, NEOMA Business School, Email: sebastien.lleo@NEOMA-bs.fr

†University of British Columbia, Vancouver, BC and Systemic Risk Centre, London School of Economics

1 Introduction

The Chinese stock market is one of the most interesting equity markets in the world by its size, scope, structure and recency. These features have a deep influence on its behavior and returns, including on the occurrence of rare events, in particular stock market crashes and downturns. In fact, the "2015 Chinese stock market crash" is just the latest in a series of 22 major downturns in a twenty-six years history.

In this paper, we discuss six stylized facts on the return distribution of the Shanghai Stock Exchange Composite Index (SHCOMP) and the Shenzhen Stock Exchange Composite Index (SZECOMP). We explain how equity downturn and crash prediction models work, and how to test their accuracy. The construction process for the signal and hit sequence is crucial to ensure that the crash prediction models produce out of sample predictions free from look-ahead bias. It also eliminates data snooping by setting the parameters *ex ante*, with no possibilities of changing them during the analysis. The construction process removes the effect of autocorrelation, making it possible to test the accuracy of the measures using standard statistical techniques. We also conduct a Monte Carlo study to address small sample bias.

Then, we test whether the price-to-earnings ratio (P/E) based on current earnings, the Bond-Stocks Earnings Yield Differential model (BSEYD) and the Cyclically Adjusted Price-to-Earnings ratio (CAPE), accurately predicts downturns in the SHCOMP and SZECOMP indexes. We find that the logarithm of the P/E has successfully predicted crashes over the entire length of the study (1990-2015 for the SHCOMP and 1991-2016 for the SZECOMP).

During the shorter 9-year period from 2006 to 2015, we find mixed evidence of the predictive ability of the BSEYD. Overall, this study supports the application of crash prediction models to the Chinese market.

The academic literature on bubbles and crashes is well established, starting with studies on bubbles by Blanchard and Watson (1982), Flood et al. (1986), Camerer (1989), Allen and Gorton (1993), Diba and Grossman (1988), Abreu and Brunnermeier (2003) and more recently Corgnet et al. (2015), Andrade et al. (2016) or Sato (2016). A rich literature on predictive models has also emerged. We can classify bubble and crash prediction models in three broad categories, based on the type of methodology and variable used: fundamental models, stochastic models and sentiment-based models.

Fundamental models use fundamental variables such as stock prices, corporate earnings, interest rates, inflation or GNP to forecast crashes. The Bond-Stock Earnings Differential (BSEYD) measure (Ziemba and Schwartz, 1991; Lleo and Ziemba, 2012, 2015b, 2017) is the oldest model in this category, which also includes the CAPE (Lleo and Ziemba, 2017) and the ratio of the market value of all publicly traded stocks to the current level of the GNP (MV/GNP) that Warren Buffett popularized (Buffett and Loomis, 1999, 2001; Lleo and Ziemba, 2015a).

Stochastic models construct a probabilistic representation of the asset prices, either as adiscrete or continuous time stochastic process. Examples include the local martingale model proposed by Jarrow and Protter (Jarrow et al., 2011a; Jarrow, 2012; Jarrow et al., 2011b,c), the disorder detection model proposed by Shiryaev, Zhitlukhin and Ziemba (Shiryaev and Zhitlukhin, 2012a,b; Shiryaev et al., 2014, 2015) and the model proposed by

Gresnigt et al. (2015), which adapts the Epidemic-type Aftershock Sequence model (ETAS) geophysics in Ogata (1988) to the stock market.

Behavioural models look at crashes in relation to market sentiment and behavioral biases. Goetzmann et al. (2016) use surveys of individual and institutional investors, conducted regularly over a 26 year period in the United States, to assess the subjective probability of a market crash and investigate the effect of behavioral biases on the formulation of these subjective probabilities. This research takes its roots in recent efforts to measure investor sentiment on financial markets (Fisher and Statman, 2000, 2003; Baker and Wurgler, 2006) and identify collective biases such as overconfidence and excessive optimism (Barone-Adesi et al., 2013).

In this paper, we focus on the three main fundamental models: the BSEYD, P/E ratio and CAPE, which we will compute daily. We leave aside Warren Buffett's ratio of the market value of all publicly traded stocks to the current level of the GNP (MV/GNP) because this measure cannot be computed more frequently than quarterly.

2 A Brief Overview of the Chinese Stock Market

Mainland China has two main stock exchanges, the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). The Shanghai Stock Exchange is the larger of the two. With an average market capitalization of USD 3.715 billion over the first half of 2016, it is the fourth largest stock market in the world[1]. The modern Shanghai Stock Exchange came into existence on November 26, 1990 and started trading on December 19, 1990. The Shenzhen Stock Exchange was founded on December 1, 1990, and started trading on July 3, 1991. While the largest and most established companies usually trade on the Shanghai Stock Exchange, the Shenzhen Stock Exchange is home to smaller and privately-owned companies.

With an average market capitalization of USD 6.656 billion over the first half of 2016, the Shanghai and Shenzhen Stock Exchanges taken together represent the third largest stock market in the world after the New York Stock Exchange at USD 17.970 billion, and the NASDAQ at USD 6.923 billion, and before 4th place Japan Exchange Group at USD 4.625 billion and fifth place LSE Group at USD 3.598 billion[2].

Chinese companies may list their shares under a variety of schemes, either domestically or abroad. Our study focuses on equity market downturns on the two leading domestic markets: the Shanghai and Shenzhen Stock Exchanges.

3 Six Main Stylized Facts

The SHCOMP and SZECOMP are market capitalization weighted index of shares listed on the SSE and SZSE, respectively. In August 2016, the SHCOMP SZECOMP consisted of the shares of 1,155 and 478 Chinese companies.

We observe and discuss six main stylized facts on the historical distribution of daily log returns on the SHCOMP and SZECOMP. Collectively, these stylized facts indicate that the SHCOMP and SZECOMP behave differently from the mature equity markets in Europe and North America.

3.1 Stylized Fact 1: The return distribution is highly volatile, right skewed with very fat tails

The daily log return on the SHCOMP from December 20, 2017 until June 30, 2016 averaged 0.0541%, with a median return of 0.0693%. The lowest and highest daily returns were respectively -17.91% and +71.92%. Table 1 also gives the corresponding statistics at a weekly and monthly frequency. The returns are highly volatile: the standard deviation of daily returns is 2.40%, equivalent to around 40 times the mean daily return. The distribution of daily returns is positively skewed (skewness = 5.26) with surprisingly fat tails (kurtosis = 149). As a result, the Jarque-Bera statistic is 5,419,808, rejecting normality at any level of significance. The Jarque-Bera statistic also leads to a strong rejection of normality for weekly and monthly data. The aggregational gaussianity, the tendency for the empirical distribution of log-returns to get closer to normality as the time scale Δt over which the returns are calculated increases, is much weaker on the SHCOMP and SZECOMP than on the S&P500 where Cont (2001) initially documented it.

We make similar observations on the SZECOMP. Table 1 shows that over the entire period, the daily log return on the SZE averaged 0.04784%, with a median return of 0.05933%. The lowest and highest daily returns were respectively -23.36% and +27.11%. Here as well, the returns are highly volatile: the standard deviation of daily returns is 2.28%, equivalent to around 50 times the mean daily return. The distribution of daily returns has a mildly positive skewness (skewness = 0.3517) and very fat tails (kurtosis = 17). The Jarque-Bera statistic for the SZECOMP still reaches 52,879. The test leads to a rejection of normality at any level of significance not only for daily data, but also for weekly and monthly data.

[**Place Table 1 here**]

3.2 Stylized Fact 2: The SHCOMP and SZECOMP do not exhibit a strong dependence structure

We turn our attention to the joint behavior of the SHCOMP and SZECOMP during the period from April 4, 1991 to June 30, 2016 (6,170 daily observations). We compute the Pearson linear correlation, Spearman's rho (rank correlation) and Kendal's tau of the daily log returns. While the Pearson linear correlation measures the strength of the linear dependence of two data series, Spearman's rho computes the correlation between data of the same rank, and Kendal's Tau measures the distance between two ranking lists based on pairwise disagreements. Spearman's rho and Kendall's tau are non parametric: they do not require any assumption on the underlying distribution. At 0.6801, 0.7922 and 0.6443 respectively, the Pearson linear correlation, Spearman's rho and Kendall's Tau are all statistically different from 0. However, neither of them is close to 1. In fact, the statistical association between the SHCOMP and the SZECOMP is noticeably weaker than, for example, the association between the S&P500 and the NASDAQ. Over the same period, the two US indices had respective Pearson linear correlation, Spearman's rho and Kendall's Tau of 0.8742, 0.8592 and 0.6884.

3.3 Stylized Fact 3: The tail behavior of the SHCOMP and SZECOMP can be modeled using a Generalized Pareto Distribution

Extreme Value Theory (EVT) is the method of choice to uncover the statistical properties of rare events. We analyze the tail behavior of the SHCOMP and SZECOMP. We refer the reader to Coles (2001) for a concise and clear introduction to EVT and to Embrechts et al. (2011) for a thorough tour of the subject.

Here, we apply EVT to the loss distribution, which we define as the negative of the probability distribution of returns, meaning that if a stock index returns -1.5% on a given day, the associated loss will be 1.5%. We focus on the tail behavior, identified as the loss above a given threshold u, that we will determine during our analysis. Let X be the random variable representing the loss, and let F be its cumulative density function. Then the cumulative density function of the loss in excess of u is:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)},$$

for $0 \leq y \leq x_F - u$, where x_F is the right endpoint of F.

Theorem 3.1 (Pickands-Balkema-de Haan (PBH) (Pickands, 1975; Balkema and de Haan, 1974))**.** *For a large class of distribution functions F, and for*

u large enough, we can approximate the conditional excess distribution $F_u(y)$ by a Generalized Pareto Distribution (GPD) $G_{\xi,\sigma}$, that is:

$$F_u(y) \approx G_{\xi,\sigma}(y), \ \text{where } G_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - e^{-\frac{\xi}{\sigma}} & \text{if } \xi = 0, \end{cases}$$

for y in $[0, x_F - u]$ if $\xi \geq 0$ and $y \in \left[0, -\frac{\sigma}{\xi}\right]$ if $\xi < 0$.

The parameters σ and ξ are respectively the scale and shape parameter of the GPD.

There is no firm rule governing the choice of threshold u. This choice of threshold must achieve a trade-off. If u is to low then the PBH theorem will not apply. If u is too high, then we will have too few observations to estimate the parameters of the GPD accurately. For example, we have 6,242 daily return observations for the SHCOMP, out of which 2,851 correspond to negative returns (i.e. positive loss). We still have 716 observations at a threshold of 2%, and 128 at a threshold of 5% but only 48 at 7%. The situation is similar on the SZECOMP. A popular method to determine u consists in plotting the sample mean excess loss against the threshold u, and picking the threshold u such that the sample mean excess loss is broadly linear for $v \geq u$. Figure 1 displays the excess loss against threshold for both the SHCOMP and SZECOMP. For the SHCOMP, we observe that the sample mean excess loss against the threshold becomes broadly linear in the threshold u starting at about $u = 4\%$. At that level, we still have 211 observations to fit the Generalized Pareto distribution. For the SZECOMP, the post suggests choosing $u = 6\%$, which leaves us with 85 observations to fit the distribution.

[Place Figure 1 here]

Finally, we estimate the scale parameter σ shape parameter ξ of the GPD using maximum likelihood. This estimation is performed against $100y$, or 100 times the loss, in order to improve numerical stability. Table 2 presents the estimated parameters, standard error of estimates as well as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for both indexes.

[Place Table 2 here]

3.4 Stylized Fact 4: Log returns do not exhibit a significant autocorrelation

Figures 2 show that the autocorrelation of daily log returns up to lag 20 are in the interval [-0.03, 0.06]. This suggests that neither indexes exhibits a short-term memory: today's return does not help forecast tomorrow's return. An analysis of the PACF leads to similar conclusions.

[Place Figure 2 here]

3.5 Stylized Fact 5: A Gaussian Hidden Markov Chain provides a good probabilistic description of the evolution of log returns... but we need between five and six states.

Stylized Fact 1 indicates that the distribution of log returns is skewed with fat tails, while Stylized Fact 2 supports the use of a Markov model to describe the probabilistic behavior of the log returns on the SHCOMP and SZECOMP. We look for a simple discrete-time Markov Model able to describe the probabilistic behavior and the evolution of log returns.

A good starting point is to look at Hidden Markov Models (HMMs). HMMs are a useful way to model the behavior of a physical or economic system when we suspect that this behavior is determined by the transition between a finite number of unobservable "regimes" or "states." We refer the reader to the excellent presentation of HMMs in Rabiner (1989) and Rabiner and Juang (1993).

The simplest, and often the best, HMM models are Gaussian Hidden Markov Chains. In these models, the returns in each state are conditionally normally distributed. The parameters of each normal distribution are specific to that state. As the state transitions over time, the returns are drawn from different normal distributions, resulting in an aggregate distribution that bears little resemblance to a normal distribution. Gaussian HMMs are estimated via the Baum-Welch algorithm (Baum et al., 1970), an application of the well-known EM algorithm (see Dempster et al., 1977).

One of the difficulties is to find the optimal number of states for the model. To that end, it is customary to use an information criterion such as the AIC or the BIC to discriminate between model formulations. The optimal model minimize the absolute value of the information criterion. Contrary to the LogLikelihood, the AIC and BIC penalize the model for the number of parameters used. This penalty is stiffer in the BIC than in the AIC.

Tables 3 present the Loglikelihood, AIC and BIC for HMMs with one to seven states, fitted respectively on the SHCOMP and the SZECOMP. We performed the numerical procedure using the *depmixS4* package in R. For the SHCOMP, we find that the optimal model specification, the specification that minimizes the AIC and BIC, is a six-state model, while the optimal model for the SZECOMP is a slightly more parsimonious, but still large, five-state model. By contrast, a two or three-state model usually proves adequate for mature indexes such as the S&P 500.

[Place Table 3 here]

3.6 Stylized Fact 6: Downturns and large market movements occur frequently

The return distribution of the SHCOMP has fat tails, which indicates that extreme events are more likely to occur than a Normal distribution would predict. Here, we focus on the large downward movements that occurred on the SHCOMP and SZECOMP.

Earlier studies, such as Lleo and Ziemba (2015b, 2017), defined an equity market downturn or crash as a decline of at least 10% from peak to trough based on the closing prices for the day, over a period of at most one year (252 trading days). We identify a correction on the day when the daily closing price crosses the 10% threshold. The identification algorithm is as follows:

1. *Identify all the local troughs in the data set.* Today is a local trough if there is no lower closing price within ±30 business days.

2. *Identify the crashes.* Today is a crash identification day if all of the following conditions hold:

    (a) The closing level of the index today is down at least 10% from its highest level within the past year, and the loss was less than 10% yesterday;

    (b) This highest level reached by the index prior to the present crash differs from the highest level corresponding to a previous crash;

    (c) This highest level occurred after the local trough that followed the last crash.

The objective of these rules is to guarantee that the downturns we identify are distinct. Two downturns are not distinct if they occur within the same larger market decline. Although these rules might be argued with, they have the advantage of being unambiguous, robust and easy to apply.

A total of 22 downturns occurred on the SHCOMP between December 19, 1990 and June 30, 2016. On average, the downturns lasted 163 days and had a 27.8% decline in the value of the index. With 22 downturns in 25 years, the SHCOMP had as many downturns as the S&P 500 over the 50 year period from January 31, 1964 to December 31, 2014.

A total of 21 downturns occurred on the SZECOMP between April 3, 1991 and June 30, 2016. On average, the downturns lasted 122 days and had a 26.4% decline in the value of the index. While the number and magnitude of equity market corrections are comparable between both indexes, we observe that downturns tend to last noticeably longer on average on the Shanghai stock Exchange than on the Shenzhen Stock Exchange.

4 Methodology

4.1 Signal Construction

The construction process for the signal and hit sequence is crucial to ensure that the crash prediction models produce out of sample predictions free from look-ahead bias. It also eliminates data snooping by setting the parameters *ex ante*, with no possibilities of changing them when we construct the hit sequence. More importantly, the construction of the hit sequence removes the effect of autocorrelation, making it possible to test the accuracy of the measures using a standard likelihood ratio test.

Equity market crash prediction models such as the BSEYD, the high P/E model or the CAPE generate a signal to indicate that an equity market downturn is likely at a given horizon h. This signal occurs whenever the value of a crash measure crosses a threshold. Given a prediction measure $M(t)$, a crash signal occurs whenever

$$SIGNAL(t) = M(t) - K(t) > 0 \tag{4.1}$$

where $K(t)$ is a time-varying threshold for the signal.

Three parameters define the signal: (i) the choice of measure $M(t)$; (ii) the definition of threshold $K(t)$; and (iii) the specification of a time interval H between the occurrence of the signal and that of an equity market downturn.

We construct the measures using two time-varying thresholds: (i) a dynamic confidence interval based on a Normal distribution; and (ii) a dynamic confidence interval using Cantelli's inequality - see Problem 7.11.9 in Grimmett and Stirzaker (2001) for a statement of the mathematical result, and Lleo and Ziemba (2012, 2017) for applications to crash predictions.

To construct the confidence intervals, we compute the sample mean and standard deviation of the distribution of the measures as a moving average and a rolling horizon standard deviation respectively. Using rolling horizon means and standard deviations has the advantage of providing data consistency. Importantly, this construction only makes use of information known at the time of the calculation. The h-day moving average at time t, denoted by μ_t^h, and the corresponding rolling horizon standard deviation σ_t^h are

$$\mu_t^h = \frac{1}{h}\sum_{i=0}^{h-1} x_{t-i}, \qquad \sigma_t^h = \sqrt{\frac{1}{h-1}\sum_{i=0}^{h-1}(x_{t-i} - \mu_t^h)^2}.$$

We establish the one-tailed confidence interval at the 95% level. This corresponds to 1.645 standard deviations above the mean in the Normal distribution.

We select the one-tailed confidence interval at $\alpha = 95\%$, corresponding to 1.645 standard deviations above the mean in the Normal distribution. This choice is consistent with the crash prediction literature and can be traced to the first published work on the BSEYD Ziemba and Schwartz (1991).

The historical development of statistical inference by Fisher, E. Pearson and Neyman, among others, has contributed to popularizing the choice of $\alpha = 95\%$ for two-tailed tests: R.A. Fisher suggested the use of a two-tailed 5% significance level (see for example pp. 45, 98, 104, 117 in Fisher, 1933; Neyman and Pearson, 1933; Neyman, 1934, 1937).

As an alternative to the normal confidence level, we construct the confidence level using Cantelli's inequality. This inequality relates the probability that the distance between a random variable X and its mean μ exceeds a number $k > 0$ of standard deviations σ to provide a robust confidence interval:

$$P\left[X - \mu \geq k\sigma\right] \leq \frac{1}{1+k^2}.$$

Setting $\beta := \frac{1}{1+k^2}$ yields $P\left[X - \mu \geq \sigma\sqrt{\frac{1}{\beta} - 1}\right] \leq \beta$. Contrary to the normal confidence level, Cantelli's inequality does not require any assumption on the shape of the underlying distribution. It should therefore provide more robust results for fat tailed distributions. The parameter β provides an upper bound for a one-tailed confidence level on any distribution. In our analysis, the horizon for the rolling statistics is h = 252 days. There is no clear rule on how to select β, so we chose $\beta = 25\%$ to produce a slightly higher threshold than the standard confidence interval. In a Normal distribution, we expect 5% of the observations to lie in the right tail, whereas Cantelli's inequality implies that the percentage of outliers in a distribution will be no higher than 25%.

The last parameter we need to specify is the horizon H. Earlier, we defined the crash identification time is the date by which the SHCOMP has declined by at least 10% in the last year (252 trading days). We define the local market peak as the highest level reached by the market index within 252 trading days before the crash. We set the horizon H to a maximum of 252 trading days prior to the crash identification date.

4.2 Signal Indicator and Crash Indicator

Crash prediction models have two components: (1) a signal indicator, which takes the value 1 or 0 depending on whether the measure has crossed the threshold, and (2) a crash indicator, which takes the value 1 when an equity market correction occurs and 0 otherwise. From a probabilistic perspective, these components are Bernoulli random variables, but they exhibit a high degree of autocorrelation, that is, a value of 1 (0) for the crash signal is more likely to be followed by another value of 1 (0) on the next day. This autocorrelation makes it difficult to test the accuracy of the model.

To remove the effect of autocorrelation, we define a signal indicator sequence $S = \{S_t, t = 1, \ldots, T\}$. This sequence records as the signal date the first day in a series of positive signals, and it only counts distinct signal dates. Two signals are distinct if a new signal occurs more than 30 days after the previous signal. The objective is to have enough time between two series of signals to identify them as distinct. The signal indicator S_t takes the value 1 if date t is the starting date of a distinct signal, and 0 otherwise. Thus, the event "a distinct signal starts on day t" is represented as $\{S_t = 1\}$. We express the signal indicator sequence as the vector $s = (S_1, \ldots, S_t, \ldots, S_T)$.

For the crash indicator, we denote by $C_{t,H}$ the indicator function returning 1 if the crash identification date of at least one equity market correction occurs between time t and time $t + H$, and zero otherwise. We identify the vector C_H with the sequence $C_H := \{C_{t,H}, t = 1, \ldots, T - H\}$ and define the vector $c_H := (C_{1,H}, \ldots, C_{t,H}, \ldots C_{T-H,H})$.

The number of correct predictions n is defined as

$$n = \# \{C_{t,H} = 1 | S_t = 1\} = \sum_{t=1}^{T} \mathbf{1}_{\{C_{t,H}=1|S_t=1\}},$$

where $\mathbf{1}_A$ is the indicator function returning 1 if condition A is satisfied, and 0 other wise.The accuracy of the crash prediction model is therefore the conditional probability $P(C_{t,H} = 1|S_t = 1)$ of a crash being identified between time t and time $t + H$, given that we observed a signal at time t. The higher the probability, the more accurate the model.

4.3 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

We use maximum likelihood to estimate the conditional probability $P(C_{t,H} = 1|S_t = 1)$ and to test whether it is significantly higher than a random guess. We obtain a simple analytical solution because the conditional random variable $\{C_{t,H} = 1|S_t = 1\}$ is a Bernoulli trial with probability $p = P(C_{t,H} = 1|S_t = 1)$.

To estimate the probability p, we change the indexing to consider only events along the sequence $\{S_t|S_t = 1, t = 1, \ldots T\}$ and denote by $X := \{X_i, i = 1, \ldots, N\}$

the "hit sequence" where $x_i = 1$ if the ith signal is followed by a crash and 0 otherwise. Here N denotes the total number of signals, that is

$$N = \sum_{t=1}^{T} S_t$$

The sequence X can be expressed in vector notation as $x = (X_1, X_2, \ldots, X_N)$. The empirical probability p is the ratio n/N.

The likelihood function L associated with the observations sequence X is

$$L(p|X) := \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

and the log likelihood function $\mathcal{L}$ is

$$\mathcal{L}(p|X) := \ln L(p|X) = \sum_{i=1}^{N} X_i \ln p + \left(N - \sum_{i=1}^{N} X_i\right) \ln(1-p)$$

This function is maximized for $\hat{p} := \frac{\sum_{i=1}^{N} X_i}{N} = n/N$, so the maximum likelihood estimate of the probability $p = P(C_{t,H}|S_t)$ is the sample proportion of correct predictions.

We apply a likelihood ratio test to test the null hypothesis $H_0 : p = p_0$ against the alternative hypothesis $H_A : p \neq p_0$. The null hypothesis reflects the idea that the probability of a random, uninformed signal correctly predicting crashes is p_0. The probability p_0 is the probability to identify an equity market downturn within 252 days of a randomly selected period. To compute p_0 empirically, we tally the number of days that are at most 252 days before a crash identification date and divide by the total number of days in the sample.

A significant departure above p_0 indicates that the measure we are considering contains some information about future equity market corrections. The likelihood ratio is:

$$\Lambda = \frac{L(p = p_0|X)}{\max_{p\in(0,1)} L(p|X)} = \frac{L(p = p_0|X)}{L(p = \hat{p}|X)}. \tag{4.2}$$

The test statistic $Y := -2\ln\Lambda$ is asymptotically χ^2-distributed with $\nu = 1$ degree of freedom. We reject the null hypothesis $H_0 : p = p_0$ and accept that the model has some predictive power if $Y > c$, where c is the critical value chosen for the test. We perform the test for the three critical values 2.71, 3.84, and 6.63 corresponding respectively to a 90%, 95% and 99% confidence level.

4.4 Monte Carlo Study for Small Sample Bias

A limitation of this likelihood ratio test is that the χ^2 distribution is only valid asymptotically. In our case, the number of correct predictions follows a binomial distribution with an estimated probability of success $\hat{p}$ and N trials. However, "only" 18 downturns occurred during the period considered in this study: the continuous χ^2 distribution might not provide an adequate approximation for this discrete distribution. This difficulty is an example of small sample bias. We use Monte Carlo methods, with $K = 10,000$ paths, to obtain the empirical distribution of test statistics and address this bias.

4.5 Further Robustness Tests

At a first glance, the statistical validity of the model seems to depend crucially on the signal construction, and therefore on two parameters: the confidence level α and the forecasting horizon H. To test the robustness of the models to the choice of parameters, we proposed a method based on optimal parameter choice. We refer the reader to the companion paper, available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2698422` on SSRN, for the full detail. This robustness analysis did not uncover significant weaknesses in the models.

5 The Price-to-Earnings Ratio

5.1 Scope of the Study

Practitioners have used the price-to-earnings (P/E) ratio to gauge the relative valuation of stocks and stock markets since at least the 1930s (for example, Graham and Dodd, 1934, discuss the use of the P/E ratio in securities analysis and valuation).

In this section, we test the predictive ability of the P/E ratio calculated using current earnings. The advantage of this definition for the SHCOMP is that it is available over the entire period from December 19, 1990 to June 30, 2016, a total of 6,243 daily observations. The same is not true for the SZECOMP. earnings and therefore P/E are only available starting July 2, 2001, a total of 3,640 daily observations.

5.2 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

Table 4 shows that the P/E and logarithm of the P/E generated a total of 18 signals (based on normally distributed confidence intervals) and 19 signals (based on Cantelli's inequality) on the SHCOMP. The number of correct predictions across models reaches 16 to 17. The accuracy of the models is in the narrow range from 88.89% to 89.47%. The type of confidence interval - normal distribution or Cantelli's inequality - only have a minor influence on the end result.

Next, we test the accuracy of the prediction on the SHCOMP statistically. To apply the likelihood ratio test, we need to compute the uninformed prior probability p_0 that a day picked at random will precede a crash identification date by 252 days or less. We find that this probability is very high, at $p_0 = 69.57\%$. This finding is consistent with the stylized facts discussed in Section 2. The Likelihood ratio test indicates that both the P/E ratio and the logarithm of the P/E ratio are significant predictors of equity market downturns markets at the 90% confidence level. Moreover, the P/E ratio, computed using a standard confidence interval, and the log P/E ratio, based on Cantelli's inequality, are significant at the 95% confidence level. Thus, we cannot rule out that the P/E and log P/E/ have helped predict equity market downturns over the period.

The P/E and logarithm of the P/E generated a total of 8 to 9 signals, with 7 to 8 correct signals on the the SZECOMP. The accuracy of the models is in a narrow range from 87.50% to 88.89%. Here as well, the type of confidence interval - normal distribution or Cantelli's inequality - only have a minor influence on the end result.

[Place Table 4 here]

5.3 Monte Carlo Study for Small Sample Bias

We continue our analysis with a Monte Carlo test for small sample bias, presented in Table 5. We compute the critical values at the 90%, 95% and 99% confidence level for the empirical distribution. Because we only have a limited number of signals, the distribution is lumpy, making it difficult to obtain meaningful p-values. Still, we find that the Monte Carlo analysis is in broad agreement with our earlier conclusions about significance of the P/E ratio and its logarithm, as both measures are significant at the 90% confidence level. We conclude that small sample bias only has a very small effect on these measures and on their statistical significance.

The uninformed prior probability p_0 that a day picked at random will precede a crash identified date by 252 days or less is 58.49%. The Likelihood ratio test indicates that both P/E ratio measures and the logarithm of the P/E ratio calculated using a standard confidence interval are significant predictors of equity market downturns markets at the 95% confidence. The remaining measure, the logarithm of the P/E ratio calculated with Cantelli's inequality is significant at the 90% confidence level. The results of the Monte Carlo analysis, presented in Table 5, indicate that small sample bias only has a minor effect on the statistical significance of the measures. All the measures are still significant at the 90% confidence level.

[Place Table 5 here]

6 The Cyclically-Adjusted Price-to-Earnings Ratio and the Bond-Stocks Earnings Yield Differential Model

6.1 Scope of the Study

The P/E ratio calculated using current earnings might be overly sensitive to current economic and market conditions. Graham and Dodd (1934) warned against this risk and advocated the use of a P/E ratio based on average earnings over ten years. In their landmark survey, Campbell and Shiller (1988) found that the R^2 of a regression of log returns on the S&P 500 with a 10 year horizon against the log of the price-earnings ratio computed using average earnings over the previous 10 and 30 years equals 0.566 and 0.401 respectively, hinting at a link between average past earning and future stock prices. This later led Shiller to suggest the use of a Cyclically Adjusted Price-to-Earnings ratio (CAPE), or a price-to-earnings ratio using 10-year average earnings, to forecast the evolution of the equity risk premium (Shiller, 2005).

The BSEYD, the second model we test, relates the yield on stocks (measured by the earnings yield, which is also the inverse of the P/E ratio) to that on nominal Government bonds.

$$BSEYD(t) = r(t) - \rho(t) = r(t) - \frac{E(t)}{P(t)}, \tag{6.1}$$

where $\rho(t)$ is the earnings yield at time t and $r(t)$ is the current 10-year government bond yield $r(t)$. The BSEYD was initially developed for the Japanese market in 1988, shortly before the stock market crash of 1990, based on the 1987 stock market in the US (Ziemba and Schwartz, 1991). The BSEYD has since been used successfully on a number of international markets (see the review article Lleo and Ziemba, 2015b), and the 2007-2008 SHCOMP meltdown (Lleo and Ziemba, 2012).

We test the forecasting ability of four measures:

1. **PE0**: P/E ratio based on current earnings. This is the measure we tested in Section 5;
2. **CAPE10**: CAPE, which is a P/E ratio computed using average earnings over the previous 10-years;
3. **BSEYD0**: BSEYD based on current earnings;
4. **BSEYD10**: BSEYD using average earnings over the previous 10-years.

We also test the logarithm of these measures: **logPE0**, **logCAPE10**, **logBSEYD0** and **logBSEYD10**. The logBSEYD is defined as:

$$logBSEYD(t) = \ln \frac{r(t)}{\rho(t)} = \ln r(t) - \ln \frac{E(t)}{P(t)}. \tag{6.2}$$

Because the CAPE10 and BSEYD10 require 10 years of earnings data, and the Bloomberg data series for 10-year government bonds only starts on October 31, 2006, we cannot use the full range of stock market data. The analysis in this section covers the period between October 31, 2006 and September 30, 2015. Over this period, the SHCOMP experienced seven declines of more than 10%, while the SZECOMP had nine.

We omit from the discussion results related to Cantelli's inequality because of space constraints. These results are nearly identical to the results we obtain for measures based on a standard confidence interval.

6.2 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

Table 7 displays results for the eight measures calculated with a confidence interval based on a normal distribution on both stock market indexes.

Looking at the SHCOMP, none of the measures produced more than 5 signals. The CAPE, logCAPE and BSEYD10 generated 3 signals each. The accuracy of the measures reaches a low of 40% for logBSEYD0 and a high of 100% for CAPE10 and logCAPE10. Only five of the eight measures are 75% accurate or better. By comparison, the uninformed prior probability that a day picked at random will precede a crash identification date by 252 days or less is $p_0 = 70.99\%$. Because of the relatively short period and small number of downturns, only CAPE10 and logCAPE10 appear significant. However, these two models only predicted three of the six crashes.

Overall, none of the models perform convincingly on the SHCOMP. The PE0 and logPE0 ratio, which we found to be significant predictors over the entire dataset in the previous section, are not significant over this restricted time period. With a 75% accuracy, they have a small edge over the uniformed prior p_0, but this edged is not significant. What's more, the BSEYD-based models do not perform as well as the P/E-based models. This is a puzzle because the BSEYD model contains additional information that is not in the P/E, namely government bond yields.The BSEYD and logBSEYD models have also been shown to perform better than the P/E ratio and CAPE on the American market (Lleo and Ziemba, 2017).

The situation on the SZECOMP is markedly different: all the measures, but one, have a 100% accuracy on the six or seven signals that they generated. The remaining measure, logBSEYD10, had six correct predictions out of seven signals, which implies a 85.71% accuracy. Although this is much higher than the uniformed prior p_0 at about 67%, the sample is to small for the difference in accuracy to be statistically significant. The discrepancy between the results observed on the SHCOMP and SZECOMP raises a number of questions. Is the difference in accuracy merely statistical, resulting from the small number of equity market downturns in the sample, or does it reveal a divergence in the microstructure of the two indexes? While the results computed in Section 5 for the P/E ratio seem to hint at the former, the latter is also a possibility, especially in light of the second Stylized Fact in Section 3.2.

[Place Table 7 here]

6.3 Monte Carlo Study for Small Sample Bias

The results of the Monte Carlo analysis for small sample bias, presented in table 8 support the conclusions of the asymptotic maximum likelihood test. In the case of the SZECOMP, the Monte Carlo analysis for small bias is not informative because most measure have an infinite test statistic.

[Place Table 8 here]

7 Conclusion And Summary of the Main Results

The Chinese stock market is certainly one of the most interesting and complex equity markets in the world. Its size, scope, structure and the the rapidity of its evolution make it unique. These characteristics inevitably affect its behavior and returns. Although the Shanghai Stock Exchange and the Shenzhen Stock Exchange are among the largest stock exchanges in the world, their behavior is much more volatile than that of more mature equity markets in Europe, and North America. The market is so volatile that the following straddle strategy is widely recommended by brokerage firms: buy at-the-money puts and calls. The idea is that market volatility raises the probability that either the call or the put will move deep in-the-money, making the strategy profitable (Ziemba, 2015).

Overall, the studies in this paper support the application of crash prediction models to the Chinese market, and reveals further research questions both on the behavior of Chinese equity markets, and on crash prediction models.

Our investigation of fundamental crash predictors reveals that the P/E and its logarithm have successfully predicted crashes on both the Shanghai Composite Index and the Shenzhen Composite index over the entire duration of the study. These results are not overly sensitive to changes in the two key parameters of the model: the confidence level α and the forecasting horizon H.

A comparison of the BSEYD, PE and CAPE and their logarithm over a shorter 9-year period, is less conclusive. On the SHCOMP, measures based on the BSEYD do not perform as well as measures based on the P/E and in particular, the CAPE. This is a puzzle because the BSEYD contains more information than the P/E and has been more successful in other markets since 1988. However, all measures perform surprisingly well on the SZECOMP. Two possible explanations for this situation are that (i) the sample is small so any correct or incorrect prediction has a large impact on the accuracy of the measure and its statistical test, and (ii) the market microstructure of the SHCOMP and SZECOMP differ because the Shanghai and Shenzhen stock exchanges were created for two different types of companies: public companies in Shanghai and privately-owned companies in Shenzhen. Both explanations open up avenues for further research.

8 Acknowledgements

The first author gratefully acknowledges support from Région Champagne Ardennes and the European Union through the RiskPerform Grant. We thank the participants to the 34thInternational Conference of the French Finance Association and the 10th International Risk Management Conference for their comments and suggestions.

References

D. Abreu and M. K. Brunnermeier. Bubbles and crashes. *Econometrica*, 71 (1):173–204, 2003.

F. Allen and G. Gorton. Churning bubbles. *Review of Economic Studies*, 60 (4):813–836, 1993.

Eduardo B. Andrade, Terrance Odean, and Shengle Lin. Bubbling with excitement: An experiment. *Review of Finance*, 20(2):447–466, 2016. doi: 10.1093/rof/rfv016. URL `http://rof.oxfordjournals.org/content/20/2/447.abstract`.

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680, 2006.

A. Balkema and L. de Haan. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.

G. Barone-Adesi, L. Mancini, and H. Shefrin. A tale of two investors: Estimating optimism and overconfidence. Technical report, 2013.

L Baum, T Petrie, G Soules, and N Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

O. J. Blanchard and M. W. Watson. Bubbles, rational expectations, and financial markets. In P. Wachtel, editor, *Crisis in the Economic and Financial Structure*, pages 295–315. 1982.

W. Buffett and C. Loomis. Warren Buffett on the stock market. *FORTUNE Magazine*, November 1999.

W. Buffett and C. Loomis. Warren Buffett on the stock market. *FORTUNE Magazine*, December 2001.

C. Camerer. Bubbles and fads in asset prices. *Journal of Economic Survey*, 3(1):3–14, 1989.

J. Y. Campbell and R. J. Shiller. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676, July 1988. Papers and Proceedings of the Forty-Seventh Annual Meeting of the American Finance Association, Chicago, Illinois, December 28-30, 1987.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 2001.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Brice Corgnet, Roberto Hernán-González, Praveen Kujal, and David Porter. The effect of earned versus house money on price bubble formation in experimental asset markets. *Review of Finance*, 19(4):1455–1488, 2015. doi: 10.1093/rof/rfu031. URL `http://rof.oxfordjournals.org/content/19/4/1455.abstract`.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

B. T. Diba and H. I. Grossman. The theory of rational bubbles in stock prices. *Economic Journal*, 98(392):746–754, 1988.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, London, 2011.

K. Fisher and M. Statman. Investor sentiment and stock returns. *Financial Analyst Journal*, pages 16–23, March/April 2000.

K. Fisher and M. Statman. Consumer confidence and stock returns. *Journal of Portfolio Management*, pages 115–127, Fall 2003.

R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh and London, 5 edition, 1933.

R. P. Flood, R. J. Hodrick, and P. Kaplan. An evaluation of recent evidence on stock market bubbles. NBER Working Paper No. w1971, 1986.

W. Goetzmann, D. Kim, and R. Shiller. Crash beliefs from investor surveys. 2016.

B. Graham and D. L. Dodd. *Security Analysis*. Whittlesey House, 1 edition, 1934.

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

G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

R. A. Jarrow. Detecting asset price bubbles. *Journal of Derivatives*, 20(1): 30–34, 2012.

R. A. Jarrow, Y. Kchia, and P. Protter. A real time bubble detection methodology. *Bloomberg Risk Newsletter*, 2011a.

R. A. Jarrow, Y. Kchia, and P. E. Protter. How to detect an asset bubble. *SIAM Journal on Financial Mathematics*, 2:839–865, 2011b.

R. A. Jarrow, Y. Kchia, and P. E. Protter. Is there a bubble in Linkedin's stock price? *Journal of Portfolio Management*, 38(1):125–130, 2011c.

S. Lleo and W. T. Ziemba. Stock market crashes in 2007–2009: were we able to predict them? *Quantitative Finance*, 12(8):1161–1187, 2012.

S. Lleo and W. T. Ziemba. Can Warren Buffett also predict equity market downturns? Technical report, 2015a.

S. Lleo and W. T. Ziemba. Some historical perspectives on the bond-stock yield model for crash prediction around the world. *International Journal of Forecasting*, 31(2):399–425, 2015b.

S. Lleo and W. T. Ziemba. Does the bond-stock earnings yield differential model predict equity market corrections better than high p/e models? *Financial Markets, Institutions & Instruments (in press)*, 2017.

J. Neyman. On the two different aspects of teh representative method. In *Royal Statistical Society*, 1934.

J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A*, 236(767):333–380, 1937.

J. Neyman and E.S. Pearson. The testing of statistical hypotheses in relation to probabilities a priori. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 29, pages 492–510, 1933.

Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83 (401):9–27, 1988.

J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.

L. Rabiner. A tutorial in Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

Lawrence Rabiner and Biing-Hwang Juang. *Fundamentals of Speech Recognition*. Prentice Hall Signal Processing Series. Prentice Hall, 1993.

Y. Sato. Fund tournaments and asset bubbles. *Review of Finance*, 20(4):1383–1426, 2016. doi: 10.1093/rof/rfv044. URL `http://rof.oxfordjournals.org/content/20/4/1383.abstract`.

R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 1 edition, 2005.

A. N. Shiryaev and M. V. Zhitlukhin. Bayesian disorder detection problems on filtered probability spaces. *Theory of Probability and Its Applications*, 57(3), 2012a.

A. N. Shiryaev and M. V. Zhitlukhin. Optimal stopping problems for a brownian motion with a disorder on a finite interval. Technical report, arXiv:1212.3709, 2012b.

A. N. Shiryaev, M. Zitlukhin, and W. T. Ziemba. When to sell Apple and the Nasdaq 100? trading bubbles with a stochastic disorder model. *Journal of Portfolio Management*, 40(2):1–10, Winter 2014.

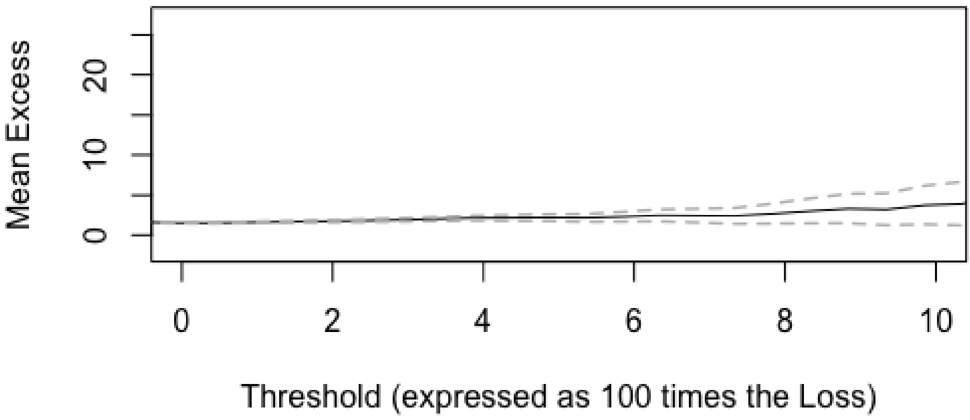
A. N. Shiryaev, M. Zitlukhin, and W. T. Ziemba. Land and stock bubbles, crashes and exit strategies in Japan, circa 1990 and in 2013. *Quantitative Finance (forthcoming)*, 15(9):1449–1469, 2015.

W. T. Ziemba. Presentation. In *Fourth International Conference on Intelligent Finance, Chongqing*, 2015.

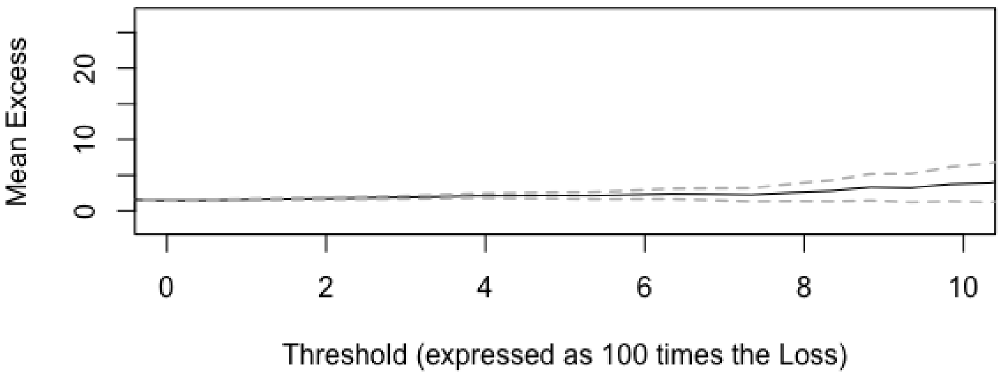
W. T. Ziemba and S. L. Schwartz. *Invest Japan.* Probus, Chicago, 1991.

Descriptive Statistics	SHCOMP			SZECOMP		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Number of observations	6,242	1,318	308	6,235	1,291	302
Mean	0.0541%	0.2497%	1.0326%	0.04784%	0.2345%	1.0644%
Median	0.0693%	0.0652%	0.7122%	0.05933%	0.1938%	0.8864%
Minimum	-17.9051%	-22.6293%	-37.3283%	-23.3607%	-33.5690%	-31.2383%
Maximum	71.9152%	90.0825%	101.9664%	27.2210%	51.9035%	60.9060%
Standard deviation	2.3848%	5.5872%	12.8898%	2.2808%	5.1795%	11.5411%
Variance	0.000569	0.000031	0.000166	0.000520	0.002683	0.013320
Skewness	5.1837	5.3543	2.3414	0.3517	1.2229	0.8724
Kurtosis	148.5003	78.5864	20.7742	17.2496	17.2522	6.6661
Jarque-Bera statistics	5,534,005	320,053	4,336	52,879.47	11,248.32	207.43
(p-value)	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$

Table 1: **Descriptive statistics for daily, weekly and monthly log returns on the SHCOMP and SZECOMP**



(a)

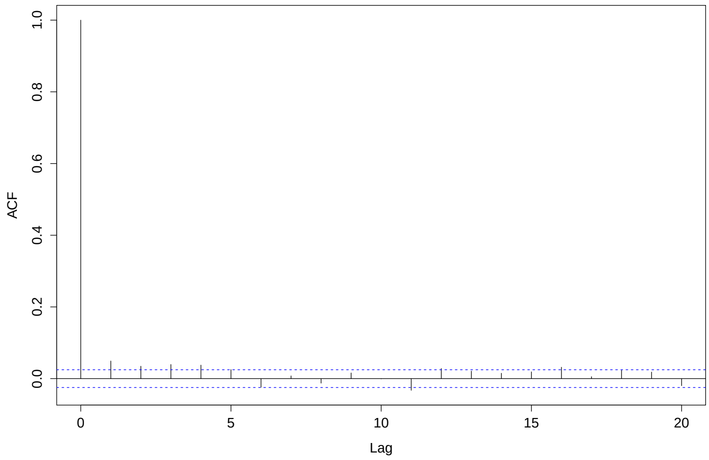


(b)

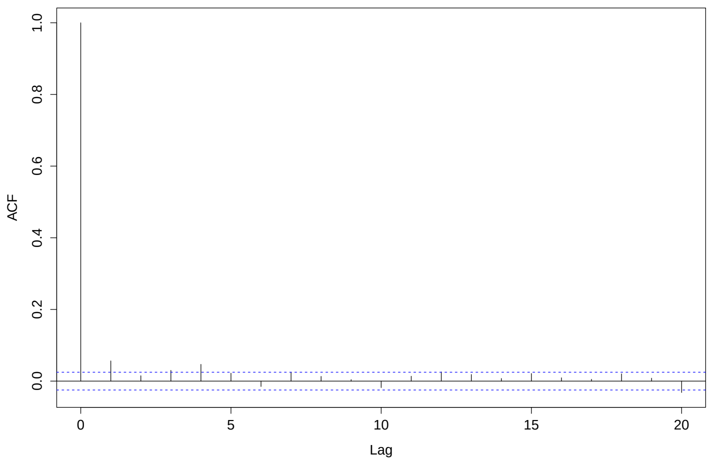
Figure 1: Sample mean excess loss against the threshold for the SHCOMP (a) and SZECOMP (b).

	SHCOMP	SZECOMP
Threshold	4	6
Number of observations	211	85
Scale parameter (standard error)	1.8214 (0.1821)	1.7141 (0.2829)
Shape parameter (standard error)	0.1292 (0.0731)	0.2176 (0.1266)
AIC	734	303
BIC	740	307

Table 2: Parameters of the Generalized Pareto distribution fitted to the tail of the SHCOMP and SZECOMP. The estimation is performed via maximum likelihood against 100× the loss to improve numerical stability.



(a) SHCOMP



(b) SZECOMP

Figure 2: Sample autocorrelation of the daily log returns on the SHCOMP and SZECOMP up to lag 20

	1	2	3	4	5	6	7
				SHCOMP			
LogLikelihood	14,464	16,514	16,827	16,887	16,895	17,183	17,194
AIC	-28,924	-33,013	-33,625	-33,728	-33,723	-34,273	-34,265
BIC	-28,910	-32,966	-33,531	-33,573	-33,494	-33,956	-33,847
Number of parameters	2	7	14	23	34	47	62
				SHCOMP			
LogLikelihood	14,726	16,053	16,225	16,298	16,331	16,346	16,389
AIC	-29,447	-32,091	-32,422	-32,550	-32,593	-32,598	-32,653
BIC	-29,434	-32,044	-32,328	-32,395	-32,364	-32,598	-32,235
Number of parameters	2	7	14	23	34	47	62

Table 3: Hidden Markov Model fitting for the daily log returns on the SHCOMP and SZECOMP

Model	Total number of signals	Number of correct predictions	ML Estimate $\hat{p}$	$L(\hat{p})$	Likelihood ratio Λ	Test statistics $-2\ln\Lambda$	p-value
				SHCOMP			
PE (confidence)	19	17	89.47%	1.67E-03	0.1159	4.3100*	3.79%
PE (Cantelli)	18	16	88.89%	1.88E-03	0.1486	3.8131†	5.09%
logPE (confidence)	18	16	88.89%	1.88E-03	0.1486	3.8131†	5.09%
logPE (Cantelli)	19	17	89.47%	1.67E-03	0.1159	4.31*	3.79%
				SZECOMP			
PE (confidence)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
PE (Cantelli)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
logPE (confidence)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
logPE (Cantelli)	8	7	87.5%	4.91E-02	0.1980	3.2387†	7.19%

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 4: **SHCOMP and SZECOMP: Maximum likelihood estimate and likelihood ratio test for the PE and logPE** The Total Number of SignalS is calculated as the sum of all the entries of the indicator sequence S. The Number of Correct Predictions is the tally of crashes preceded by the signal. It is calculated as the sum of all the entries of the indicator sequence X. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. $L(\hat{p})$ is the likelihood of the crash prediction model, computed using the maximum likelihood estimate $\hat{p}$. The likelihood ratio $\Lambda = \frac{L(p_0|X)}{L(p=\hat{p}|X)}$ is the ratio of the likelihood under the null hypothesis $p = p_0$ to the likelihood using the estimated probability $\hat{p}$. The estimated test statistics, equal to $-2\ln\Lambda$, is asymptotically χ^2-distributed with 1 degree of freedom. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p-value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99.5% level are respectively 3.84, 6.63 and 7.88.

Model	Total number of signals	ML Estimate $\hat{p}$	Critical Value			Test statistics $-2\ln\Lambda(p_0)$
			90% confidence	95% confidence	99% confidence	
	SHCOMP					
PE (confidence)	19	89.47%	2.38	4.31	7.61	$4.3100^{\dagger}$
PE (Cantelli)	18	88.89%	2.38	4.31	7.61	$3.8131^{\dagger}$
logPE (confidence)	18	88.89%	2.99	3.81	6.99	$3.8131^{\dagger}$
logPE (Cantelli)	19	89.47%	2.99	3.81	6.99	4.3100^{*}
	SZECOMP					
PE (confidence)	9	88.89%	2.31	4.06	4.92	$4.0607^{\dagger}$
PE (Cantelli)	9	88.89%	2.31	4.06	4.92	$4.0607^{\dagger}$
logPE (confidence)	9	88.89%	2.31	4.06	8.86	$4.0607^{\dagger}$
logPE (Cantelli)	8	87.50%	2.31	4.06	8.86	$3.2387^{\dagger}$

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 5: **SHCOMP: Monte Carlo likelihood ratio test for the PE and logPE** The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. Colums 4 to 6 report the critical values at the 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in column 7 is equal to $-2\ln\Lambda(p_0) = -2\ln\frac{L(p_0|X)}{L(p=\hat{p}|X)}$ and that in column 9 is $-2\ln\Lambda\left(\frac{1}{2}\right) = -2\ln\frac{L(\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated for both tests are based on the empirical distribution. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance indicated in the test statistics column and the p-value indicated in the table are both based on and empirical distribution generated through Monte-Carlo simulations.

Confidence	0.8	0.85	0.9	0.925	0.95	0.975	0.99
				P/E ratio			
Number of signals	21	21	22	22	19	16	15
Number of correct signals	15	18	18	19	17	15	15
Proportion of correct signals	71.43%	85.71%	81.82%	86.36%	89.47%	93.75%	100%
Test statistics	0.0348	2.9770[†]	1.7190	3.4022[†]	4.3100[*]	5.7847[*]	-
p-value	85.2%	8.45%	18.98%	6.51%	3.79%	1.62%	-
				logP/E ratio			
Number of signals	21	21	21	19	18	14	11
Number of correct signals	15	17	18	17	16	14	11
Proportion of correct signals	71.43%	80.95%	85.71%	89.47%	88.89%	100%	100%
Test statistics	0.0348	1.4050	2.9770[†]	4.3100[*]	3.8131[†]	-	-
p-value	85.2%	23.59%	8.45%	3.79%	5.09%	-	-

[†] significant at the 10% level;
[*] significant at the 5% level;
[**] significant at the 1% level;
[***] significant at the 0.5% level.

Table 6: **SHCOMP and SZECOMP: Accuracy and statistical significance of the P/E ratio and logP/E ratio as a function of the confidence level** α. The numbers presented in this table are based on a forecasting horizon $H = 252$ days. With this choice, the uninformed probability that a random guess would correctly identify an equity market downturn is $p_0 = 67.64\%$ Row 1,2 and 3 respectively report the total number of signals generated by the P/E ratio, the number of correct signals, and the proportion of correct signals computed as the ratio of the number of correct signals to the total number of signals. Rows 4 and 5 respectively report the test statistics and p-value for the P/E ratio. The subsequent rows present the same information for the log P/E ratio.

Signal Model	Total number of signals	Number of correct predictions	ML Estimate $\hat{p}$	$L(\hat{p})$	Likelihood ratio Λ	Test statistics $-2\ln\Lambda$	p-value
				SHCOMP			
BSEYD0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
logBSEYD0	5	2	40.00%	3.46E-02	0.7901	0.4713	49.24%
PE0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
logPE0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
BSEYD10	3	2	66.67%	1.48E-01	0.9228	0.1606	68.86%
logBSEYD10	5	3	60.00%	3.46E-02	0.9778	0.0449	83.23%
CAPE10	3	3	100.00%	-	-	-	-
logCAPE10	3	3	100.00%	-	-	-	-
				SZECOMP			
BSEYD0	6	6	100.00%	-	-	-	-
logBSEYD0	7	7	100.00%	-	-	-	-
PE0	6	6	100.00%	-	-	-	-
logPE0	6	6	100.00%	-	-	-	-
BSEYD10	7	6	85.71%	5.67E-02	0.5266	1.2826	25.74%
logBSEYD10	7	7	100.00%	-	-	-	-
CAPE10	6	6	100.00%	-	-	-	-
logCAPE10	5	5	100.00%	-	-	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 7: **SHCOMP and SZECOMP: Maximum likelihood estimate and likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**The Total Number of Signals is calculated as the sum of all the entries of the indicator sequence S. The Number of Correct Predictions is the tally of crashes preceded by the signal. It is calculated as the sum of all the entries of the indicator sequence X. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. $L(\hat{p})$ is the likelihood of the crash prediction model, computed using the maximum likelihood estimate $\hat{p}$. The likelihood ratio $\Lambda = \frac{L(p_0|X)}{L(p=\hat{p}|X)}$ is the ratio of the likelihood under the null hypothesis $p = p_0$ to the likelihood using the estimated probability $\hat{p}$. The estimated test statistics, equal to $-2\ln\Lambda$, is asymptotically χ^2-distributed with 1 degree of freedom. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p-value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99.5% level are respectively 3.84, 6.63 and 7.88.

Signal Model	Total number of signals	ML Estimate $\hat{p}$	Critical Value			Test statistics $-2\ln\Lambda(p_0)$
			90% confidence	95% confidence	99% confidence	
	SHCOMP					
BSEYD0	4	75%	4.74	4.74	6.44	0.6654
logBSEYD0	5	40%	2.62	5.92	8.05	0.4713
PE0	4	75%	4.74	4.74	6.44	0.6654
logPE0	4	75%	4.74	4.74	6.44	0.6654
BSEYD10	3	66.67%	3.55	4.83	4.83	0.1606
logBSEYD10	5	60%	2.62	5.92	8.05	0.0449
CAPE10	3	100.00%	3.55	4.83	4.83	-
logCAPE10	3	100%	3.55	4.83	4.83	-
	SZECOMP					
BSEYD0	6	100.00%	4.81	4.81	6.48	-
logBSEYD0	5	100.00%	4.31	5.61	5.61	-
PE0	6	100.00%	4.81	4.81	6.48	-
logPE0	6	100.00%	4.81	4.81	6.48	-
BSEYD10	7	85.71%	4.31	5.61	5.61	1.2826
logBSEYD10	7	100.00%	4.31	5.61	5.61	-
CAPE10	6	100.00%	4.81	4.81	6.48	-
logCAPE10	5	100.00%	4.01	4.01	4.66	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 8: **SHCOMP and SZECOMP: Monte Carlo likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm** The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. Colums 4 to 6 report the critical values at the 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in column 7 is equal to $-2\ln\Lambda(p_0) = -2\ln\frac{L(p_0|X)}{L(p=\hat{p}|X)}$ and that in column 9 is $-2\ln\Lambda\left(\frac{1}{2}\right) = -2\ln\frac{L(\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated for both tests are based on the empirical distribution. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance indicated in the test statistics column and the p-value indicated in the table are both based on and empirical distribution generated through Monte-Carlo simulations.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk