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

1 Introduction

Through the summer of 2015, the gyrations of the Shanghai stock exchange captured the headlines of the financial press. In fact, what has been labeled the "2015 Chinese stock market crash" is just the latest in a series of 22 major downturns in the twenty-five years of the Chinese stock market history. Headlines aside, the Chinese stock market is certainly one of the most interesting equity markets in the world by its size, scope, structure and recency. These features have a deep influence on the behavior and returns of the Chinese stock market.

In this paper, we discuss four four stylized facts on the return distribution of the Shanghai Stock Exchange Composite Index (SHCOMP) and the Shenzhen Shenzhen Stock Exchange Composite Index (SZCOMP). Then, we explain how equity downturn and crash prediction models work, and how to test their accuracy. The construction process for the signal and

*The first author gratefully acknowledges the support from Région Champagne Ardennes and the European Union through the RiskPerform grant.

†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

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. Also, 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 provides supporting evidence for the application of crash prediction models to the Chinese market.

The academic literature on bubbles and crashes is well established, starting with the 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 bubble and crash predictions 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). Recently, Callen and Fang (2015) also found evidence that short interest is positively related to one-year ahead stock price crash risk.

Stochastic models construct a probabilistic representation of the asset prices. This representation can be either a discrete or a continuous time stochastic processes. 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 earthquake model of Gresnigt et al. (2015). When it comes to actual implementation, the local martingale model and the disorder detection model share the same starting point: they assume that the evolution of the asset price $S(t)$ can be best described using a geometric Brownian motion:

$$dS(t) = \mu(t, S(t))S(t)dt + \sigma(t, S(t))S(t)dW(t), S(0) = s_0, t \in \mathbb{R}^+$$

where $W(t)$ is a standard Brownian motion on the underlying probability space. However, the two models look at different aspect of the evolution. The disorder detection model detects crashes by looking for a change in regime in the drift μ and volatility σ. The local martingale model detects bubbles by testing whether the volatility σ is a local martingale or a strict martingale. In contrast, the earthquake model uses a jump-diffusion process and has a shorter forecasting horizon: 5 trading days.Gresnigt et al. (2015) implement the Epidemic-type Aftershock Sequence model (ETAS) geophysics model proposed by Ogata (1988) and based on an Hawkes process, a type of inhomogeneous point process.

Behavioural models look at crashes in relation to market sentiment and behavioural 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. They observe that these probabilities are much higher than the actual historical probabilities. To understand this observation, the authors examine a number of factors that influence investor responses and found evidence consistent with an availability bias. 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 main fundamental models: the BSEYD, P/E ratio and CAPE.

2 A Brief Overview of the Chinese Stock Market

Mainland China has two 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 officially came into

[1] Source: The World federation of Exchanges, `http://www.world-exchanges.org/home/index.php/statistics/monthly-reports` retrieved on September 13th, 2016

being on November 26, 1990 and started trading on December 19, 1990. The Shenzhen Stock Exchange was formally founded on December 1, 1990, and it 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 represents 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].

On November 17, 2014, the Chinese government launched the Shanghai-Hong Kong Stock Connect to enable investors in either market to trade shares on the other market. The Hong Kong Exchanges and Clearing is currently the 8th largest exchange in the world with an average market capitalisation of USD 2.932 Billion over the half of 2016[3]. This announcement was followed by the creation of a Shenzhen-Hong Kong link on August 16th, 2016. These initiatives herald a closer integration between securities markets in China and further boosts the rapid development of the Chinese market.

Chinese companies may list their shares under various schemes, either domestically or abroad. Domestically, companies may issue:

- *A-shares*: common stocks denominated in Chinese Reminbi and listed on the Shanghai or Shenzhen stock exchanges.

- *B-shares*: special purpose shares denominated in foreign currencies but listed on the domestic stock exchange. Until 2001, only foreign investors had access to B-shares.

In addition to B-shares, foreign investors interested in the Chinese equity market may buy:

- *H-shares*: shares denominated in Hong Kong Dollars and traded on the Hong Kong Stock Exchange.

- *L-chips, N-chips and S-chips*: shares of companies with significant operations in China, but incorporated respectively in London, New York and Singapore.

[2]Source: *ibid*
[3]Source: *ibid*

- *American Depository Receipts (ADRs)*: an ADR is a negotiable certificate issued by a U.S. bank representing a specified number of shares in a foreign stock traded on an American exchange. As of October 2015, there were about 110 Chinese ADRs listed on American exchanges and another 200 Chinese ADRs on American over-the-counter markets.

The diversity of investment schemes available shows that although the Shanghai and Shenzhen Stock Exchange are a large and crucial part of the Chinese equity market, they do not represent the whole market. For example, there are also *red chips* (shares of companies incorporated outside mainland China but owned or substantially controlled by Chinese state-owned companies) and *P-chips* (shares of companies owned by private individuals and traded outside mainland China, for example on the Hong Kong stock exchange). Our study focuses on equity market downturns on the Shanghai and Shenzhen Stock Exchanges.

3 Four Key Stylized Facts

The Shanghai Stock Exchange Composite Index, or SHCOMP, is the main Chinese stock index. It is a market capitalisation weighted index of all the A-shares and B-shares listed on the SSE. In August 2016, the SHCOMP consisted of the shares of 1,155 Chinese companies.

The Shenzhen Composite Index, or SZCOMP, is a market capitalisation weighted index of all the A-shares and B-shares listed on the SZSE. In August 2016, 478 Chinese companies were listed on the SZSE.

We observe and discuss four key stylized facts on the historical distribution of daily log returns on the SHCOMP and SZECOMP. These stylised facts help understand statistically the behaviour of the market and the frequency of large market movements such as equity market downturns. Undoubtedly, various other aspects of the index are of interest and would warrant a thorough analysis similar to Cont's (2001) analysis of the S&P500, but this is beyond the scope of this paper.

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

Figure 1 displays the evolution of the SHCOMP since its launch on December 19, 1990, as well as the distribution of daily log returns on the index. Figure 2 displays the evolution of the SZEComp since April 3, 1991, as well as the distribution of daily log returns on the index.

Table 1 shows that over the entire period, the daily log return on the SHCOMP averaged 0.0541%, with a median return of 0.0693%. The lowest and highest daily returns were respectively -17.91% and +71.92%. The exhibit 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 very 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 phenomenon in which the empirical distribution of log-returns tends to normality as the time scale Δt over which the returns are calculated increases, is much weaker on the SHCOMP and SZCOMP than on the S&P500 where Cont (2001) initially documented it.

[Place Figure 1 here]

[Place Table 1 here]

We make similar observations on the SZEComp. Table 2 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 fat tails (kurtosis = 17). Although the SZECOMP is much less skewed than the SHCOMP and its tails are fat. 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 Figure 2 here]

[Place Table 2 here]

Next, we turn our attention to the joint behaviour of the SHCOMP and SZECOMP during the period from April 4, 1991 to June 30, 2016 (6,170 daily observations). Figure 3 displays the joint distribution of log returns and a Quantile-Quantile (QQ) plot for the two indexes. Both charts suggest that the two indexes are not perfectly correlated, with a small proportion of outliers in the body of the distribution (relative to the total number of data points), and a larger proportion of outliers in the two tails of the distribution (as evidenced by the QQ plot).

To test this hypothesis, we compute the Pearson linear correlation, Spearman's rho (rank correlation) and Kendal's tau of log returns over the entire period. 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 have the advantage of being non parametric and of not requiring 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 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.

[Place Figure 3 here]

Extreme Value Theory (EVT) is the field of choice to uncover the statistical properties of rare and large events. We analyse the tail behaviour 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 probability distribution of returns, so if the stock market returns -1.5% on a given day, the associated loss will be 1.5%. We focus on the tail behaviour, identified as the loss above a certain 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 $G_{\xi,\sigma}$, that is:*

$$F_u(y) \approx G_{\xi,\sigma}(y)$$

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 or mathematical result to govern the choice of threshold u. Essentially, the choice of thshold 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 Generalized Pareto distribution accurately. Table 3 shows that the number of observations decreases sharply as the threshold increases. 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 SZCOMP.

[Place Table 3 here]

A popular method 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 4 displays the excess loss against threshold for both the SHCOMP and SZCOMP. 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 SZCOMP, the post suggests choosing $u = 6\%$, which leaves us with 85 observations to fit the distribution.

[Place Figure 4 here]

We estimate the scale parameter σ shape parameter ξ of the Generalized Pareto distribution using maximum likelihood. This estimation is performed against $100y$, or 100 times the loss, in order to improve numerical stability. Table 4 presents the estimated parameters, standard error of estimates as well as the AIC and BIC for both indexes.

[Place Table 4 here]

3.2 Stylized Fact 2: Log returns do not exhibit a significant autocorrelation

Figures 5 and 6 show that the autocorrelation of daily log returns is low for both indexes. The autocorrelations do not appear statistically meaningful and the partial autocorrelations up to lag 20 are in the interval [0.03, 0.06]. This suggests that neither indexes tend to have had a short-term memory: hence, today's return does not help forecast tomorrows return.

[Place Figure 5 here]

[Place Figure 6 here]

3.3 Stylized Fact 3: 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 behaviour of the log returns on the SHCOMP and SZECOMP. So, we look for a simple discrete-time Markov Model able to describe the probabilistic behaviour 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 underlying but unobservable "regimes" or "states." We present a short overview of HMMs in Appendix A and 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 normally distributed, but the parameters of each normal distribution are specific to that state. As the state transitions over time, the returns are drawn from different distributions, resulting in an aggregate distribution that bears little resemblance to a normal one. 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 Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) to discriminate between model formulations. The optimal model will minimize the value of teh information criterion, in purchase the AIC or the BIC. 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 the in the AIC.

Tables 5 and 7 present the Loglikelihood, AIC and BIC for a HMM 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.

[Place Table 5 here]

The transition probability matrix P_{SHCOMP} for the SHCOMP is

$$\begin{pmatrix} 7.2689e-01 & 4.3972e-175 & 2.6749e-231 & 2.7311e-01 & 6.2691e-303 & 2.3976e-220 \\ 2.5311e-04 & 9.3881e-01 & 2.2433e-58 & 3.8282e-02 & 2.0163e-05 & 2.2637e-02 \\ 2.5622e-202 & 1.2809e-90 & 8.6320e-01 & 1.4495e-117 & 5.2279e-33 & 1.3680e-01 \\ 8.5411e-03 & 1.1560e-01 & 9.9862e-04 & 8.7043e-01 & 5.1730e-27 & 4.4282e-03 \\ 3.6201e-127 & 2.5515e-02 & 4.1153e-22 & 1.6460e-15 & 5.2982e-01 & 4.4466e-01 \\ 2.3777e-115 & 6.5345e-06 & 1.1122e-02 & 2.8003e-47 & 7.5704e-01 & 2.3184e-01 \end{pmatrix}.$$

The initial probability and the parameters of the normal distribution for each state are given in Table 6.

The transition probability matrix $P_{SZECOMP}$ for the SZECOMP is

$$\begin{pmatrix} 2.9883e-01 & 2.4262e-16 & 1.1294e-27 & 6.7930e-01 & 2.1866e-02 \\ 6.0465e-02 & 9.3154e-01 & 7.9941e-03 & 1.1369e-18 & 5.6279e-14 \\ 4.3118e-16 & 1.7914e-02 & 7.8243e-01 & 3.9601e-62 & 1.9965e-01 \\ 1.4320e-01 & 2.6999e-02 & 3.1915e-03 & 8.1727e-01 & 9.3390e-03 \\ 1.3235e-08 & 2.9150e-03 & 1.8946e-02 & 2.5725e-02 & 9.5241e-01 \end{pmatrix}.$$

The initial probability and the parameters of the normal distribution for each state are given in Table 8.

[Place Table 8 here]

3.4 Stylized Fact 4: 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 $\pm d$ 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; and
   (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.

Table 9 lists the 22 downturns that occurred 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 SHCOMP. With 22 downturns in 25 years, the SHCOMP had as many downturns as the S&P 500 had over the 50 year period from January 31, 1964 to December 31, 2014.

[**Place Table 9 here**]

Table 10 presents the 21 downturns that 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.

[**Place Table 10 here**]

Collectively, these stylized facts indicate that the SHCOMP and SZECOMP behaves differently from the mature equity markets in Europe and North America.

4 Methodology

In order to apply the equity downturn prediction models to the SHCOMP and SZECOMP, we need to examine the inner workings of these models: how they are constructed, how to convert them into a testable model, and how to test the accuracy of their predictions. This is the objective of this section.

4.1 Signal Construction

As discussed above 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* during the signal construction, 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. In addition to the standard likelihood ratio test using the asymptotic χ^2 distribution, we conduct a Monte Carlo study on the empirical distribution to address small sample bias.

Equity market crash prediction models such as the BSEYD (Ziemba and Schwartz, 1991; Lleo and Ziemba, 2012, 2017), the high P/E model (Lleo and Ziemba, 2017), the variations on Warren Buffett's market value-to-GNP measure (Lleo and Ziemba, 2015a), or the continuous time disorder detection model (Shiryaev et al., 2014, 2015) generate a signal to indicate a downturn in the equity market 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 key 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 test the measures using two time-varying thresholds: (1) a dynamic confidence interval based on a Normal distribution; and (ii) a dynamic confidence interval using Cantellis 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 is purely in-sample. 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 $\alpha = 95\%$This corresponds 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).

Looking at the statistical inference literature, $\alpha = 95\%$ is a natural choice for two-tailed tests: R.A. Fisher suggested the use of a two-tailed

5% significance level Fisher (see for example pp. 45, 98, 104, 117 in 1933). Pearson and Neyman insisted on the significance level to be selected *a priori*, before Neyman introduced the idea of a confidence interval (Neyman and Pearson, 1933; Neyman, 1934, 1937) first thought of by Pearson. Although Fisher later clarified that the level of significance need to be selected in relation to the statistical problem under consideration (Fisher, 1955), the 5% significance / 95% confidence has remained in widespread used ever since.

Another way to look at the choice of $\alpha = 95\%$ is in the context of the expert opinion literature; see (Meyer and Booker, 2001; O'Hagan, 2006) for an up-to-date treatment.. Here, we change our frame of reference from a classical frequentist approach, which assumes that that we are sampling repeatedly IID random variables, to subjective probabilities, and more specifically to the personal probability framework introduced by Ramsey, de Finetti and Savage (we refer the reader to the classic book by Savage, 1971). In this framework, the crash prediction model is subjective in nature and akin to an expert opinion. The confidence level α is properly defined as the subjective level of confidence in the our measure's ability to predict "normal" market operations. Any departure above this level would indicate that we are outside of the confidence interval around our measure: a market disruption such as an equity market downturn, is likely to happen. In the expert opinion literature, it is customary to ask for a two-tailed 90% confidence bound, translating into a one-tailed confidence interval. This observation provides another motivation for our selection of α. Still, this discussion emphasises the need to test how our models perform for various choices of α in order to ascertain whether they are robust to a misspecification or a change in confidence level.

As an alternative to the normal confidence level, we also 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 $\alpha = \frac{1}{1+k^2}$ yields $P\left[X - \mu \geq \sigma\sqrt{\frac{1}{\alpha} - 1}\right] \leq \alpha$. 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. Recall that 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.

To conclude this discussion, we revisit the three determinants of the signal: the measure, which e revisit the measure is the stochastic variable we are analysing, two key parameters influencing the signals are:

1. the measure is the stochastic variable that we are studying;
2. the confidence level α and β are the key parameter which determines the threshold K and by extension the number of signals generated;
3. the horizon H is the key parameter which determines the prediction period and influences the accuracy of the measure.

To determine the robustness of the prediction models, we will need to study their sensitivity to the choice of confidence level and the horizon.

4.2 Construction of the Hit Sequence X

Crash prediction models have two components: (1) a signal, which takes the value 1 or 0 depending on whether the measure has crossed the confidence level, 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 $s = (S_1, \ldots, S_t, \ldots, S_T)$. This construction effectively removes the effect of autocorrelation.

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

$$n = \sum_{t=1}^{T} C_{t,H} = \mathbf{1}' c_H.$$

The accuracy of the crash prediction model is 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.

We use maximum likelihood to estimate this probability 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

$$N = \sum_{t=1}^{T} S_t = \mathbf{1}' s$$

where $\mathbf{1}$ is a vector with all entries set to 1 and v' denotes the transpose of vector v. 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.

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

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 $\hat{p} := \frac{\sum_{i=1}^{N} X_i}{N}$ so the maximum likelihood estimate of the probability $p = P(C_{t,H}|S_t)$, is in fact the historical 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. A significant departure above this level indicates that the measure we are considering contains some information about future equity market corrections. The likelihood ratio test 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 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.

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.

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 crashes 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 to obtain the empirical distribution of test statistics and address this bias.

The Monte Carlo algorithm is as follows. Generate $K = 10,000$ paths. For each path $k = 1, \ldots, K$, simulate N Bernoulli random variables with probability p_0 of obtaining a "success."

Denote by $X_k := \left\{X_i^k, i = 1, \ldots, N\right\}$ the realization sequence where $x_i^k = 1$ if the ith Bernoulli variable produces a "success" and 0 otherwise.

Next, compute the maximum likelihood estimate for the probability of success given the realization sequence X_k as $\hat{p} := \frac{\sum_{i=1}^N X_i^k}{N}$, and the test statistic for the path as

$$Y_k = -2 \ln \Lambda_k = -2 \ln \frac{L(p = p_0 | X_k)}{\max_{p \in (0,1)} L(p_k | X_k)} = -2 \ln \frac{L(p = p_0 | X_k)}{L(p = \hat{p}_k | X_k)}.$$

Once all the paths have been simulated, we use all K test statistics $Y_k, k = 1, \ldots, K$ to produce an empirical distributions for the test statistic Y.

From the empirical distribution, we obtain critical values at the 90%, 95% and 99% confidence level, against which we assess the crash prediction test statistic Y. The empirical distribution also enables us to compute a p-value for the crash prediction test statistics. Finally, we compare the results obtained with the empirical distribution to those derived using the asymptotic χ^2 distribution.

4.5 Optimal Parameter Choice and Parameter Robustness

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. The confidence level affects directly the number of signals that the model generates, and indirectly the accuracy of the model. The forecasting horizon influences the number of correct signals, as well as the uninformed probability p_0 used in the significance test, but it does change the number of signals generated. It is easier to produce an accurate forecast if we have a longer horizon to prove us right than a shorter one. In this section, we set up a plan to test the sensitivity or robustness of the model to these two parameters.

First, we compute the optimal value for the confidence level α. We hold the time horizon H constant at 252 days, and seek the range of confidence levels $\alpha \in [0.9, 1]$ that maximizes the empirical accuracy $\hat{p}$:

$$\mathcal{A} = \text{argmax}_{\alpha \in [0.9,1)} \hat{p}(\alpha; H)$$

We are actually interested in the lowest confidence interval for which we $\hat{p} = 100\%$, as well as in the general evolution of the number of predications as the confidence level increases. We expect the accuracy of the measure to increase with the confidence level.

We are also interested in whether the model remains significantly better than a random guess if we chose a confidence level at the lower end of the confidence range. Answering this question will give us an indication on the robustness of the model in relation to a change or misspecification in the confidence level. This approach is an application of the robust likelihood statistics proposed by Lleo and Ziemba (2017) to a case where we test the robustness with respect to a single parameter.

Next, we look for the optimal value for the forecasting horizon H. We hold the confidence level α constant at 95% and look for the range of time horizons H that maximizes the empirical accuracy $\hat{p}$:

$$\mathcal{H} = \text{argmax}_{H \in \{63,126,189,252\}} \hat{p}(H;\alpha)$$

We limit the range of our analysis to up to 252 days after the signal. Note the we cannot test the robustness of the model with respect to a change in forecasting horizon using the robust likelihood statistics proposed by Lleo and Ziemba (2017) because changing the forecasting horizon will affect the uninformed probability p_0.

5 The Price-to-Earnings Ratio

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 analyze 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 12, 1990 to June 30, 2016, a total of 6243 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.

5.1 Shanghai

Table 11 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 Cantellis 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 Cantellis inequality - only have a minor influence on the end result.

Next, we test the accuracy of the prediction 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.

[Place Table 11 here]

We continue our analysis with a Monte Carlo test for small sample bias, presented in Table 12. 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.

[Place Table 12 here]

We follow up with analysis of the sensitivity of the measures to a change of confidence level α and forecasting horizon H. We focus here on measures computed using a standard confidence interval for clarity as we would obtain similar results for measures computed using Cantelli's inequality. Table 13 reports the key statistics of the measure for various confidence levels. Picking a confidence level at the low end of our range, $\alpha = 90\%$, the P/E ratio generates 22 signal while the log P/E produces 21 signals. With an accuracy of 81.82%, the P/E ratio is no longer significant at the 90% confidence level. On the other hand, the log P/E is 85.71% accurate, maintaining itself above the critical value corresponding to a 90% confidence level. Expanding the scope of our investigation outside of the initial [0.9,1) range to consider a broader confidence range of $[0.8, 1)$, we find that the accuracy and significance of the P/E ratio and log P/E ratio broadly increase with the confidence level, while the number of signals decreases monotonically, as expected. In fact, the accuracy of the models reaches 100% at $\alpha = 0.99$ for the P/E ratio and $\alpha = 0.97$ for the log of the P/E, but with only 14 to 15 predictions out of 22 crashes.

The increase in the accuracy and significance is not monotonic because of the limited number of predictions: adding one correct prediction or one incorrect predictions tends to have a noticeable impact on the accuracy of the measure. This makes the transitions lumpy rather than smooth. Still, we observe that both the P/E ratio and the log P/E ratio remain significant at the 90% confidence level in the range $[0.925, 1)$, suggesting that the two measures are not overly sensitive to a small change in the confidence parameter α.

[**Place Table 13 here**]

We conclude our analysis of the P/E and log P/E by investigating the sensitivity of these measures to a change in horizon H. Table 14 reports the key statistics for $H = 63, 126, 189$ and 252 days, corresponding to 3 months, 6 months, 9 months and 1 year. The accuracy of the signals decreases as we we shorten the time horizon, and so does the uninformed probability p_0. Overall, the P/E and log P/E become significant when the horizon reaches 9 months to 1 year, and their test statistics reaches its maximum at 9 months.

[**Place Table 14 here**]

5.2 Shenzhen

Table 15 shows that the P/E and logarithm of the P/E generated a total of 8 to 9 signals, with 7 to 8 correct signals. The accuracy of the models is in the narrow range from 87.50% to 88.89%. Here as well, the type of confidence interval - normal distribution or Cantellis inequality - only have a minor influence on the end result.

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 16, 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 15 here]

[Place Table 16 here]

An analysis of the sensitivity of the measure to a change in the confidence parameter α produces a surprising outcome. Contrary to what we observed with the SHCOMP, the results for the SZECOMP, presented in Table 17, show that the accuracy of the measures, and therefore their statistical significance, declines overall as the α increases. The accuracy of the models decline from 91.67% at $\alpha = 80\%$ to 85.71% at $\alpha = 99\%$. This is enough to push the p-value up from 0.98% to 11.73%. This counterintuitive outcome is a result of the fact that the total number of signals generally decrease, as α increases. This is what we observe here: the models generate 11 to 12 signals at $\alpha = 80\%$ but only 6 to 7 at $\alpha = 99\%$. Since the models are already particularly accurate, an erroneous signal therefore results in a larger loss of accuracy at $\alpha = 99\%$ than at $\alpha = 80\%$.

[Place Table 17 here]

The measures do not exhibit a high sensitivity to a change in the time horizon H. The results of the analysis, summarised in Table 18, show that the models remain significant at the 90% confidence level across all four time horizons: 63 days, 126 days, 189 days and 252 days.

[Place Table 18 here]

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

The drawback of the P/E ratio calculated using current earnings is that it 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) performed a regression of the log returns on the S&P 500 at 1 year, 3 year and 10 year horizons against the log dividend-price ratio, lagged dividend growth rate, average annual earnings over the previous 30 years, and against the average annual earnings over the previous 10 years. They 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. This is higher than the R^2 of regressions against the log dividend-price ratio and lagged dividend growth rate (see Lleo and Ziemba, 2017, for a review of the literature and a discussion of the key results.). This 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 tested 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 tested the logarithm of these measures: **logPE0**, **logCAPE10**, **logBSEYD0** and **logBSEYD10**.

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. Theses results are nearly identical to the results we obtain for measures based on a standard confidence interval.

6.1 Shanghai

Table 19 displays the results for the eight measures, calculated with a confidence interval based on a normal distribution. First, none of the measure produced moe 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. 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 are have an edge over the uniformed prior p_0. But we simply do not have enough crashes and prediction to tilt the statistical scales in their favor: the p-value remains around 40%. The results of the Monte Carlo analysis for small sample bias, presented in table 20, do not give us additional information.

[Place Table 19 here]

[Place Table 20 here]

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 been shown to perform better than the P/E ratio and CAPE on the American market (Lleo and Ziemba, 2017) and has performed well on most international markets (Lleo and Ziemba, 2015b). There are there possible explanations. The first possibility is that the sample we are studying is simply too limited. Seven crash and between three and five predictions is not enough to get reliable statistics, and leaves our conclusions vulnerable to the "law of small numbers". This is undoubtedly a main concern, which the results of our Monte Carlo analysis for small sample bias cannot dispel.

A second explanation relates to the choice of forecasting horizon. If the forecasting BSEYD-based measures generates a signal with at a shorter horizon, than the forecasting horizon $H = 252$, then the measure will appear inaccurate. This is very similar to judging the quality of a camera by first blocking its depth-of-field at 6 meters, but taking a picture of a flower just centimes away. The camera might be excellent, but the picture will appear hopelessly out-of-focus. We will be able to examine this hypothesis later, by analyzing the sensitivity of the measures to a change of horizon.

Another possible hypothesis, which we cannot test directly in the context of this study, is that the market microstructure of the SHCOMP and of the Chinese bond market makes the supply and demand for securities less sensitive to the prevailing government bond rate. If the problem is related to the definition of the interest rate, in particular if the government bond rate is not representative of the financing rate for stock traders and portfolio managers, we can expect to make similar observation on the SZECOMP. On the other hand, if the problem is linked to the microstructure of the SHCOMP, we might see BSEYD-based measures perform relatively better than P/E-based measures on the SZECOMP.

Table 21 reports the results of an analysis of the measures' sensitivity to a change in the confidence parameter α. To the exception of the logBSEYD0, the accuracy of the measures increase as α increases. Five measures out of eight reach a 100% accuracy at $\alpha = 97.5\%$, over two or three signals. The logBSEYD0 and logBSEYD10 remain the worst performing measures. On aggregate the measures behave as expected: their accuracy increases as α increases, but they are not particularly sensitive to our initial choice $\alpha = 0.95$.

[Place Table 21 here]

Finally, we explore the sensitivity of the measures to a change in the forecasting horizon H. The results in Table 22 indicate that BSEYD0, PE0, logPE0, BSEYD10 and logBSEYD10 perform best at $H = 126$, while CAPE10 and logCAPE10 reach 100% accuracy at $H = 126$. In fact, all the measures except logBSEYD0 and logBSEYD10 are significant at the 90% confidence with the choice $H = 126$. We conclude that the measures are sensitive to the forecasting horizon, and that the standard choice $H = 252$ is suboptimal on this dataset. This conclusion comes in support of the second hypothesis we suggested to explain the relatively poor performance of the BSEYD models. It does not, however, fully explain this relative underperformance.

[Place Table 22 here]

6.2 Shenzhen

The situation on the SZECOMP is markedly different: all the measures display a remarkable accuracy. The results in Table 23 show that alls the measures, but one, have a 100% accuracy on the six to 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 Monte Carlo analysis for small bias, reported in Table 24 is not informative in this case, because most measure have an infinite test statistic.

[Place Table 23 here]

[Place Table 24 here]

Overall, the measures are resilient to a change in the accuracy parameter α, as shown in Table 25. The logBSEYD0, PE0, CAPE10 and logCAPE10 maintain a 100% accuracy over the entire range of accuracy parameters. BSEYD0 and logPE0 have a 100% accuracy on the range $[0.85, 0.99]$, while BSEYD10 and logBSEYD10 are 100% accurate over most of the range. None of the measures is less than 83.33% accurate.

[Place Table 25 here]

Seven measures have a 100% accuracy at $H = 189$ and $H = 252$. At a horizon $H = 126$ days, the accuracy of five of the measures is statistically significant at the 95% confidence level. At this horizon, the accuracy of the worst performing measures is 71.43%, far above the prior probability $p_0 = 44\%$. Further reducing the time horizon to $H = 63$ days, reduces the accuracy of the measures. Still, three of the eight measure are statistically significant at the 90% confidence level.

[Place Table 26 here]

7 Conclusion And Summary of the Main Results

The Chinese stock market is certainly one of the most interesting and most 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 Echange 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, this study shows clearly that crash prediction models can be applied directly to the Chinese market, and reveals potential areas for further research both on the behaviors 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 length of the study. These results are relatively robust 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. 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. Exploring this hypothesis is a possible question for future research.

8 Acknowledgements

The first author gratefully acknowledges support from Région Champagne Ardennes and the European Union through the RiskPerform Grant.

A Appendix: An Overview of Hidden Markov Models

A.1 Basic Structure

We start from the assumption that at any point in time t, the financial index, whether the SHCOMP or the SZECOMP, can be in any of N distinct states $S_1, S_2, \ldots, S_n$. Denote by q_t the actual state of the system at time $t = 1, 2, \ldots$, and by $\{q_t = S_i\}$ the event 'being in state i at time t'.

On any day, the index may change state with probability

$$P\left[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \ldots\right]$$

We further assume that the state transitions satisfy the Markov property, which implies that

$$P\left[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \ldots\right] = P\left[q_t = S_j | q_{t-1} = S_i\right]$$

we model this as a discrete (first order) Markov Chain with a transition probability matrix $A = (a_{ij})$, $i, j = 1, \ldots, N$ of the form

$$a_{ij} = P\left[q_t = S_j | q_{t-1} = S_i\right] \tag{A.1}$$

where the state transition coefficients a_{ij} satisfy

$$\begin{aligned} a_{ij} &\geq 0 \\ \sum_{j=1}^{N} a_{ij} &= 1 \end{aligned}$$

Next, denote the initial state probabilities by

$$\pi_i = P\left[q_1 = S_i\right], \qquad i = 1, \ldots, N$$

The observation sequence $O = \{O_1, O_2 \ldots O_T\}$ records the actual states that have occurred from time 1 to time T. We allow $O_1, O_2, \ldots$ to be the vector of returns on the index.

In simple cases, we could read the state directly (as an example, we could think about weather condition outside our window). If we have observed state S_1 at time 1, S_3 at time 2, S_1 at time 3 and S_2 at time 4, then the observation sequence is $O = \{S_1, S_3, S_1, S_2\}$. Often, the current state of the system is not directly observable. In this sense the actual sate of the Markov chain is 'hidden'. As a result, we need to rely on observations to infer the current state of the market. For example, the current state of the a financial market is not directly observable: we need to infer it from the returns we observe.

The theory of HMM was originally built around the idea of discrete *observation symbols* associated with each states. These observation symbols form an alphabet of size M. We denote by V the set of all observation symbols, i.e.

$$V = \{v_1, v_2, \ldots, v_k, \ldots v_M\}$$

Think of V as a set of letters or sounds (music notes, syllables...)

The probability of the observation symbol given that the system is in state j is

$$b_j(k) = P\left[v_k \text{ at } t | q_t = S_j\right], \qquad 1 \leq j \leq N, 1 \leq k \leq M \tag{A.2}$$

The probability distribution of the observation symbol is the matrix $B = (b_j(k))$, $1 \leq j \leq N, 1 \leq k \leq M$.

The idea of a discrete observation set does work for simple coin toss or ball-and-urn experiments as well as for some data processing applications, but it has severe limitations for financial markets where the observation sequence is represented by asset returns.

As a result, we need to change the standard model to allow continuous observation sets and continuous probability distributions.

To that effect, we model the returns in each state as a M-component Gaussian mixture. The mathematical specification of this model is:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} \mathcal{N}\left(\mathbf{O}, \mu_{jm}, \mathbf{\Sigma}_{jm}\right), \quad 1 \leq j \leq N \tag{A.3}$$

where

- $\mathbf{O}$ is a d-dimensional observation vector;
- $\mathcal{N}$ is the Gaussian pdf[4].
- c_{jm} is the mixture coefficient for the j-th state and m-th mixture;
- μ_{jm} is the mean vector for the j-th state and m-th mixture;

[4] Any log-concave or elliptically-symmetric probability would work, although in reality most people will use Gaussian distributions

- $\mathbf{\Sigma}_{jm}$ is the covariance matrix for the j-th state and m-th mixture.

The mixture coefficient c_{jm} satisfies the following constraints:

$$\begin{aligned} c_{jm} &\geq 0 \\ \sum_{m=1}^{M} c_{jm} &= 1 \end{aligned} \tag{A.4}$$

Moreover, for b to be a properly defined pdf we need to have

$$\int_{-\infty}^{+\infty} b_j(x)dx = 1, \quad 1 \leq j \leq N \tag{A.5}$$

When $M = 1$, we revert to the case where the returns in each state are conditionally jointly-Normally distributed.

To sum things up, our HMM model is comprised of:

1. N unobervable states $S_1, S_2, \ldots, S_N$;
2. a $N \times N$ transition probability matrix $A = (a_{ij})$ where

$$a_{ij} = P\left[q_t = S_j | q_{t-1} = S_i\right] \tag{A.6}$$

3. initial state probabilities

$$\pi_i = P\left[q_1 = S_i\right], \qquad i = 1, \ldots, N$$

4. a sequence of d-dimensional vectors $\{\mathbf{O}_t\}_{t=1,\ldots}$ with observation probability given by a M-dimensional Gaussian mixture:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} \mathcal{N}\left(\mathbf{O}, \mu_{jm}, \mathbf{\Sigma}_{jm}\right), \quad 1 \leq j \leq N \tag{A.7}$$

We denote by $B(\mathbf{O})$ the N-dimensional pdf vector.

A.2 The Three (Basic) Problems for HMMs And How to Solve Them

We express the set of model parameters λ as $\lambda = (A, B, \pi)$. Rabiner (1989, p. 261)) summarizes the three 'basic' problems for HMMs:

1. Given an observation sequence $O = O_1 O_2 \ldots O_T$ and a model $\lambda = (A, B, \pi)$, how do we compute efficiently the probability of the observation sequence $P(O|\lambda)$?

2. Given an observation sequence $O = O_1 O_2 \dots O_T$ and the model λ, how do we choose a corresponding state sequence $Q = q_1 q_2 \dots q_T$ which is optimal in some meaningful sense (i.e. best "explains" the observations?

3. How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximise $P(O|\lambda)$?

Solving the first problem requires a forward-backward numerical procedure. The Viterbi algorithm (Vitterbi, 1967) solves the second problem. In terms of structure, the Viterbi algorithm is similar to a forward procedure developed to solve the first problem, but it also includes a maximization at each node and a backtracking step.

The third problem is the most difficult of the three. The standard way of solving it is due to Baum and his coauthors and is known as the Baum-Welch algorithm (Baum et al., 1970, and references within). The Baum-Welch algorithm is in fact a special case of another celebrated algorithm: the EM or Expectation-Maximization algorithm (Dempster et al., 1977).

The Baum-Welch algorithm works by iteratively choosing a set of parameters $\lambda = (A, B, \pi)$ to maximise $P(O|\lambda)$. The iterative reestimation procedure is shown to converge monotonically to a local maximum. Like the EM algorithm, the Baum-Welch algorithm only identifies local maxima. From a practical perspective, this means that it is important to run the algorithm multiple times with different starting values to ensure that the solution obtained is the global maximum and not just a local one.

A.3 Model Selection

One of the difficulties with HMM models is to select the optimal number of states. We cannot use the Likelihood or the Loglikelihood directly, because the likelihood will increase as we increase the number of states. One way of addressing this problem is by selecting the model that optimzes one of the following information criteria:

1. The Akaika information criterion (AIC) is

$$AIC = -2 \ln L + 2p \tag{A.8}$$

   where L denotes the likelihood of the model and p is the number of parameters.

2. Schwartz' Bayesian information criterion (BIC) is

$$SBIC = -2 \ln L + 2p \ln T \tag{A.9}$$

where T is the number of observations.

3. The Hannan-Quinn criterion (HQIC) is

$$HQIC = -2 \ln L + 2p \ln(\ln(T)) \tag{A.10}$$

where T is the number of observations.

All three information criteria maximize the log likelihood of the model penalized by the number of parameters. The key difference between the three criteria relates to the treatment of the number of observations. While the Akaika information criterion ignores the number of observations, the Bayesian information criterion penalises by the log of the number of observations. The Hannan-Quinn cirterion is in between: the penalty is linked to the number of observations, but it is less stiff than the Bayesian information criterion.

In this paper, we consider both the AIC and BIC to determine the optimal number of states.

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.

J.L. Callen and X. Fang. Short interest and stock price crash risk. *Journal of Banking & Finance*, 60:181–194, 2015.

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. A. Fisher. Statistical methods and scientific inference. *Journal of the Royal Statistical Society, Series B*, 17(1):69–78, 1955.

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.

M.A. Meyer and J.M. Booker. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics, 2001.

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.

A. O'Hagan. *Uncertain Judgments: Eliciting Expert's Probabilities*. Wiley, 2006.

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.

Yuki 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`.

L.J. Savage. *The Foundations of Statistics.* Dover, 2 edition, 1971.

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.

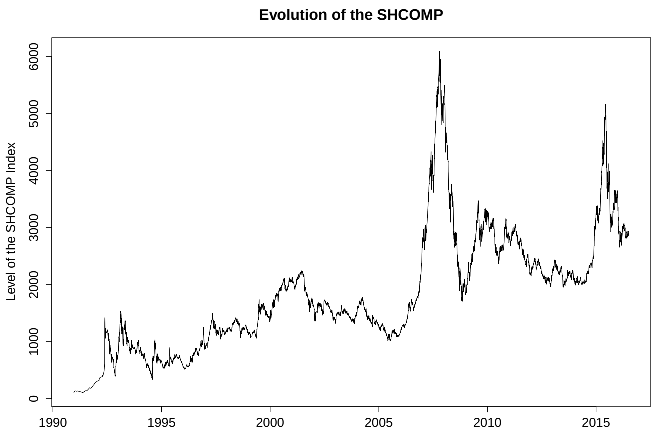
A. N. Shiryaev, M. Zitlukhin, and W. T. Ziemba. When to sell Apple and the Nasdaq 100? trading bubbles with a stochastic disorder model. *Journal of Portfolio Management*, 40(2):1–10, Winter 2014.

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.

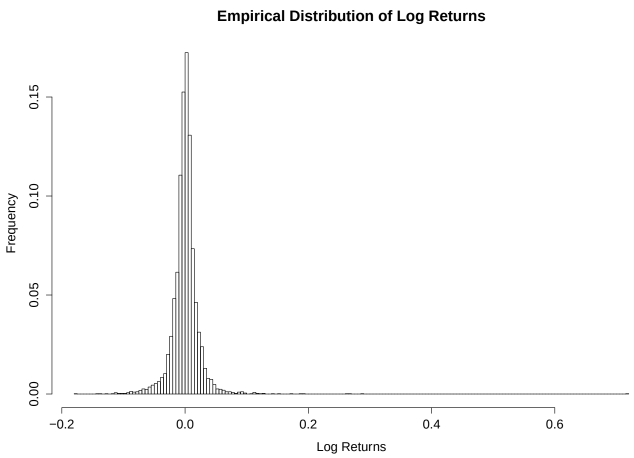
Andrew Vitterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, April 1967.

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.



(a)

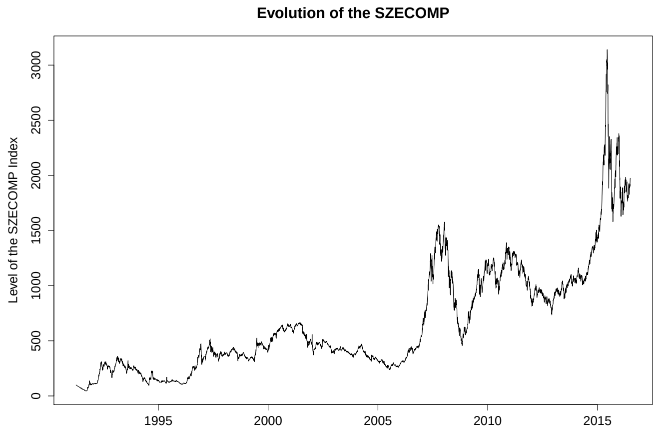


(b)

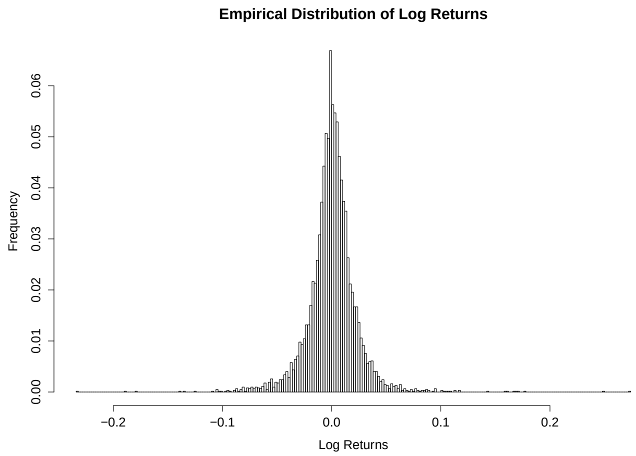
Figure 1: Evolution of the SHCOMP Index and empirical distribution of the daily log return (December 19, 1990 - June 30, 2016).

Descriptive Statistics	Frequency		
	Daily	Weekly	Monthly
Number of observations	6,242	1,318	308
Mean	0.0541%	0.2497%	1.0326%
Median	0.0693%	0.0652%	0.7122%
Minimum	-17.9051%	-22.6293%	-37.3283%
Maximum	71.9152%	90.0825%	101.9664%
Standard deviation	2.3848%	5.5872%	12.8898%
Variance	0.000569	0.000031	0.000166
Skewness	5.1837	5.3543	2.3414
Kurtosis	148.5003	78.5864	20.7742
Jarque-Bera statistics	5,534,005	320,053	4,336
(p-value)	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$

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



(a)

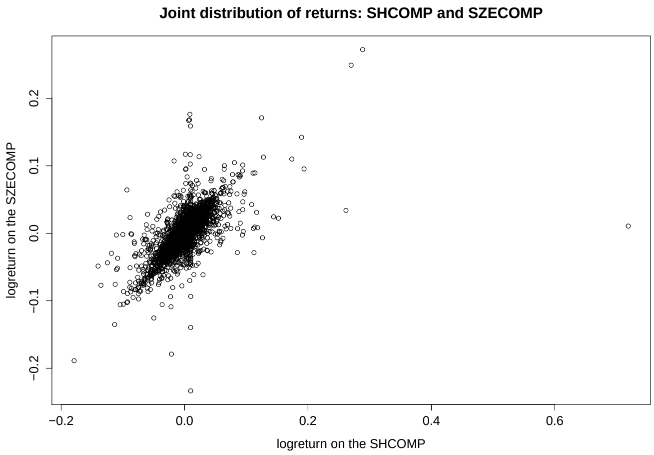


(b)

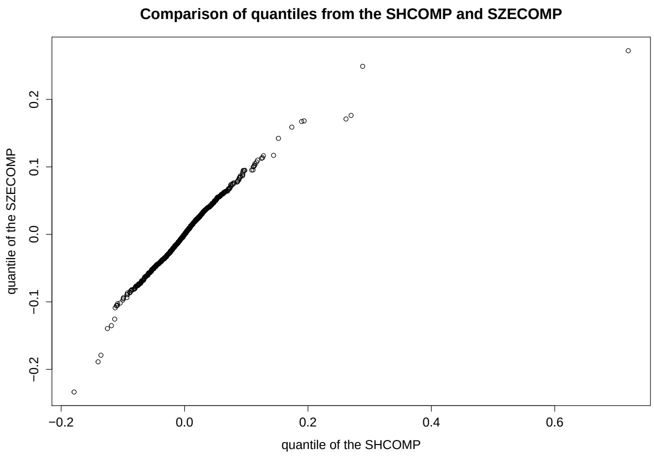
Figure 2: Evolution of the SZECOMP Index and empirical distribution of the daily log return (December 19, 1990 - June 30, 2016).

Descriptive Statistics	Frequency		
	Daily	Weekly	Monthly
Number of observations	6,235	1,291	302
Mean	0.04784%	0.2345%	1.0644%
Median	0.05933%	0.1938%	0.8864%
Minimum	-23.3607%	-33.5690%	-31.2383%
Maximum	27.2210%	51.9035%	60.9060%
Standard deviation	2.2808%	5.1795%	11.5411%
Variance	0.000520	0.002683	0.013320
Skewness	0.3517	1.2229	0.8724
Kurtosis	17.2496	17.2522	6.6661
Jarque-Bera statistics	52,879.47	11,248.32	207.43
(p-value)	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$

Table 2: **Descriptive statistics for daily, weekly and monthly log returns on the SZECOMP**



(a)

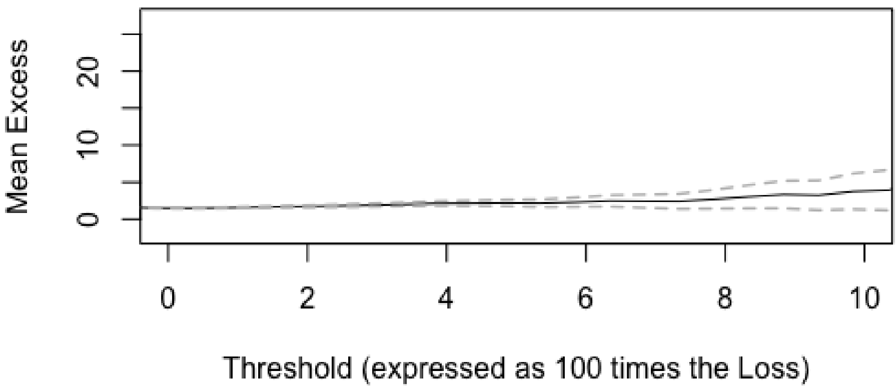


(b)

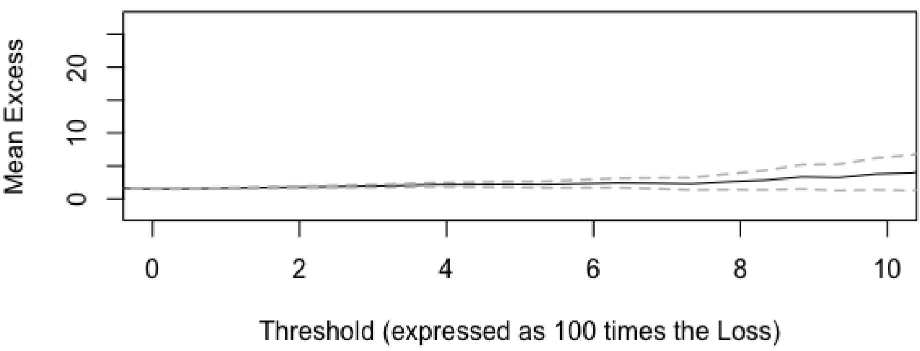
Figure 3: Joint behaviour of the SHCOMP and SZECOMP (April 3, 1991 - June 30, 2016).

	$-\infty$	0%	1%	2%	5%	7%	8%	9%	10%
SHCOMP	6,242	2,851	1,429	716	128	48	30	16	12
SZECOMP	6,235	1,461	734	134	52	32	16	12	

Table 3: Number of observations exceeding a threshold u. Selecting $-\infty$ as a threshold produces the total number of observations for the index.



(a)

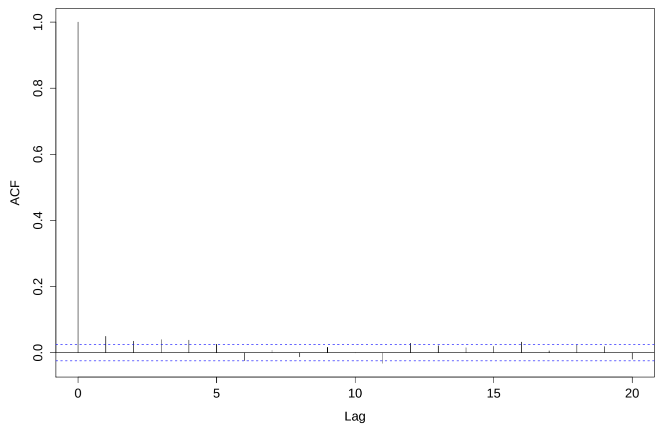


(b)

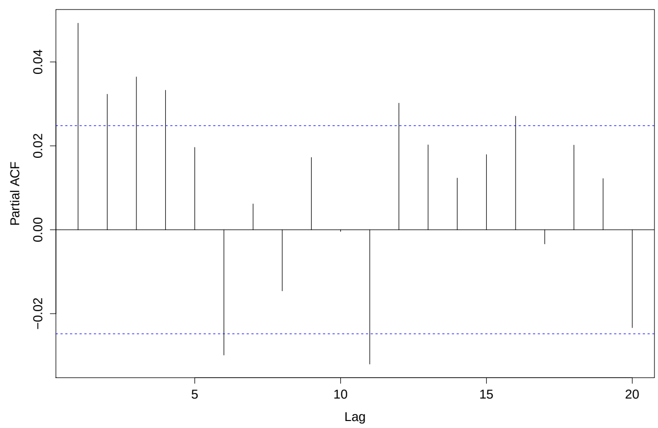
Figure 4: Sample mean excess loss against the threshold for the SHCOMP and SZCOMP.

	SHCOMP	SZCOMP
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	733.56	302.59
BIC	740.26	307.4779

Table 4: Parameters of the Generalized Pareto distribution fitted to the tail of the SHCOMp and SZCOMP. The estimation is performed via maximum likelihood against 100× the loss to improve numerical stability.

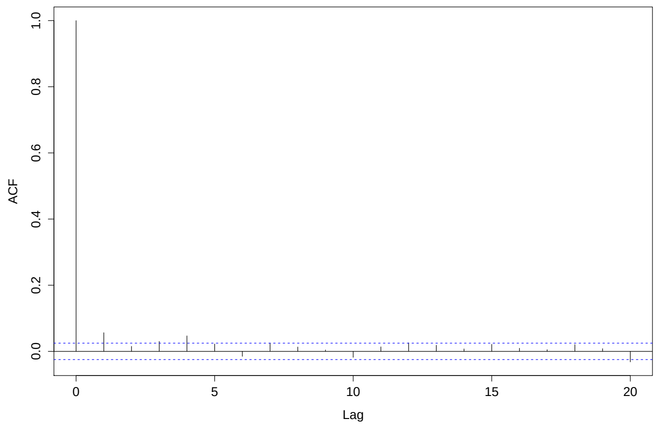


(a) Sample autocorrelation up to lag 20

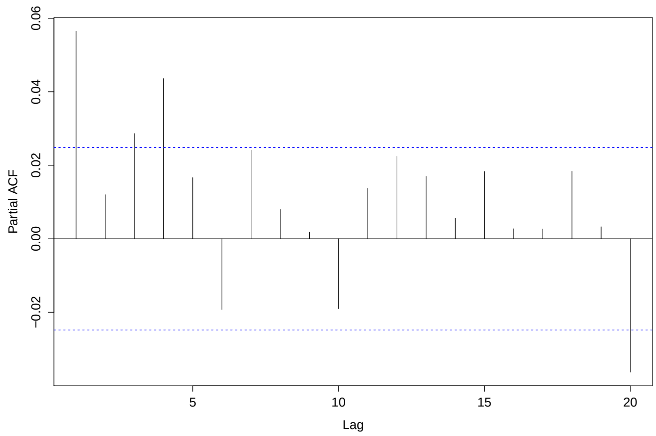


(b) Sample partial autocorrelation up to lag 20

Figure 5: Autocorrelation and partial autocorrelation of the daily log returns on the SHCOMP



(a) Sample autocorrelation up to lag 20



(b) Sample partial autocorrelation up to lag 20

Figure 6: Autocorrelation and partial autocorrelation of the daily log returns on the SZECOMP

Criterion	1	2	3	4	4	6	7
LogLikelihood	14,463.88	16,513.68	16,826.51	16,887.18	16,895.43	17,183.46	17,194.39
AIC	-28,923.76	-33,013.36	-33,625.02	-33,728.36	-33,722.85	-34,272.92	-34,264.78
BIC	-28,910.28	-32,966.18	-33,530.67	-33,573.36	-33,493.73	-33,956.19	-33,846.96
Number of parameters	2	7	14	23	34	47	62

Table 5: Hidden Markov Model fitting for the daily log returns on the SHCOMP

State	Initial Probability	Mean	Standard Deviation
1	0.00	-1.0996%	1.4419%
2	0.00	1.6517%	9.5992%
3	0.00	0.3433%	1.0724%
4	0.00	0.2015%	2.0513%
5	1.00	-0.1354%	0.7037%
6	0.00	0.1464%	3.6882%

Table 6: Initial probability and parameters of the Gaussian distributions for each state of the HMM

Criterion	1	2	3	4	4	6	7
LogLikelihood	14,725.65	16,052.61	16,225.11	16,297.95	16,330.57	16,346.03	16,388.55
AIC	-29,447.29	-32,091.21	-32,422.22	-32,549.90	-32,593.14	-32,598.06	-32,653.1
BIC	-29,433.82	-32,044.05	-32,327.89	-32,394.92	-32,364.05	-32,281.38	-32,235.34
Number of parameters	2	7	14	23	34	47	62

Table 7: Hidden Markov Model fitting for the daily log returns on the SZECOMP

State	Initial Probability	Mean	Standard Deviation
1	0.00	-1.2627%	1.5456%
2	1.00	-0.0750%	0.7600%
3	0.00	0.1433%	7.1826%
4	0.00	0.3734%	1.2108%
5	0.00	0.1057%	2.7757%

Table 8: Initial probability and parameters of the Gaussian distributions for each state of the HMM

	Crash Identification Date	Peak Date	SHCOMP Index at Peak	Trough date	SHCOMP Level at trough	Peak-to-trough decline (%)	Peak-to-trough duration (in days)
1	1992-05-27	1992-05-25	1421.57	1992-11-17	393.52	72.3%	176
2	1993-02-23	1993-02-15	1536.82	1993-03-31	925.91	39.8%	44
3	1994-09-19	1994-09-13	1033.47	1995-02-7	532.49	48.5%	147
4	1996-08-26	1996-07-24	887.6	1996-09-12	757.09	14.7%	50
5	1996-11-6	1996-10-28	1022.86	1996-12-24	865.58	15.4%	57
6	1997-05-16	1997-05-12	1500.4	1997-09-23	1041.97	30.6%	134
7	1998-08-7	1998-06-3	1420	1998-08-17	1070.41	24.6%	75
8	1999-07-1	1999-06-29	1739.21	1999-12-27	1345.35	22.6%	181
9	2000-09-22	2000-08-21	2108.69	2000-09-25	1875.91	11%	35
10	2001-02-21	2001-01-10	2125.62	2001-02-22	1907.26	10.3%	43
11	2001-07-30	2001-06-13	2242.42	2002-01-22	1358.69	39.4%	223
12	2003-04-23	2002-07-8	1732.93	2003-11-18	1316.56	24%	498
13	2004-04-29	2004-04-6	1777.52	2004-09-13	1260.32	29.1%	160
14	2006-08-4	2006-07-11	1745.81	2006-08-7	1547.44	11.4%	27
15	2007-02-2	2007-01-24	2975.13	2007-02-5	2612.54	12.2%	12
16	2007-06-4	2007-05-29	4334.92	2007-07-5	3615.87	16.6%	37
17	2007-11-8	2007-10-16	6092.06	2008-11-4	1706.7	72%	385
18	2009-08-12	2009-08-4	3471.44	2009-08-31	2667.75	23.2%	27
19	2010-10-27	2009-11-23	3338.66	2011-01-25	2677.43	19.8%	428
20	2012-12-27	2012-03-2	2460.69	2013-06-27	1950.01	20.8%	482
21	2014-06-25	2013-09-12	2255.6	2014-06-25	2025.5	10.2%	286
22	2015-06-19	2015-06-12	5166.35	2015-08-26	2927.29	43.3%	75

Table 9: The SHCOMP Index experienced 22 crashes between December 19, 1990 and June 30, 2016.

	Crash Identification Date	Peak Date	SZECOMP Index at Peak	Trough date	SZECOMP Level at trough	Peak-to-trough decline (%)	Peak-to-trough duration (in days)
1	1992-06-3	1992-05-26	312.21	1992-06-16	233.73	25.1%	21
2	1993-03-5	1993-02-22	359.44	1993-07-21	203.91	43.3%	149
3	1996-05-10	1995-05-22	169.66	1996-08-26	152.55	10.1%	462
4	1996-09-10	1996-09-4	274.56	1996-12-24	242.01	11.9%	111
5	1997-05-16	1997-05-12	517.91	1997-09-23	312.73	39.6%	134
6	1998-07-6	1998-06-3	441.04	1998-08-18	317.1	28.1%	76
7	1999-07-1	1999-06-29	525.14	1999-12-27	395.69	24.7%	181
8	2000-09-25	2000-08-21	643.77	2000-09-25	578.76	10.1%	35
9	2001-02-8	2000-11-23	654.37	2001-02-22	568.26	13.2%	91
10	2001-07-30	2001-06-13	664.85	2002-01-22	371.79	44.1%	223
11	2004-04-26	2004-04-7	470.55	2004-09-13	315.17	33%	159
12	2006-08-2	2006-07-12	446.61	2006-08-7	380.26	14.9%	26
13	2007-06-1	2007-05-29	1292.44	2007-07-5	1015.85	21.4%	37
14	2007-10-25	2007-10-9	1551.19	2007-11-28	1219.98	21.4%	50
15	2008-01-22	2008-01-15	1576.5	2008-11-4	456.97	71%	294
16	2009-08-14	2009-08-4	1149.27	2009-09-1	900.53	21.6%	28
17	2009-12-22	2009-12-3	1234.17	2010-07-5	921.34	25.3%	214
18	2010-11-17	2010-11-10	1389.54	2011-01-25	1136.58	18.2%	76
19	2013-06-24	2013-05-30	1043.47	2013-06-25	879.93	15.7%	26
20	2014-03-28	2014-02-17	1160.39	2014-04-28	1007.27	13.2%	70
21	2015-06-19	2015-06-12	3140.66	2015-09-15	1580.26	49.7%	95

Table 10: The SZECOMP Index experienced 21 crashes between March 25, 1992 and June 30, 2016.

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
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	$3.8131^{\dagger}$	5.09%
logPE (confidence)	18	16	88.89%	1.88E-03	0.1486	$3.8131^{\dagger}$	5.09%
logPE (Cantelli)	19	17	89.47%	1.67E-03	0.1159	4.31^{*}	3.79%

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

Table 11: **SHCOMP: Maximum likelihood estimate and likelihood ratio test for the PE and logPE**

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		
PE (confidence)	19	89.47%	2.38	4.31	7.61	$4.3100^{\dagger}$	5.4%
PE (Cantelli)	18	88.89%	2.38	4.31	7.61	$3.8131^{\dagger}$	7.61%
logPE (confidence)	18	88.89%	2.99	3.81	6.99	$3.8131^{\dagger}$	7.84%
logPE (Cantelli)	19	89.47%	2.99	3.81	6.99	4.3100^{*}	3.8%

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

Table 12: **SHCOMP: Monte Carlo likelihood ratio test for the PE and logPE**

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 13: **SHCOMP: 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.

Horizon (days)	63	126	189	252
Uninformed probability p_0	50.99%	59.59%	66.41%	69.57%
		P/E ratio		
Number of correct signals	18	18	18	18
Proportion of correct signals	57.89%	73.68%	89.47%	89.47%
Test statistics	0.3648	1.6561	5.4937*	4.31*
p-value	54.58%	19.81%	1.91%	3.79%
		logP/E ratio		
Number of correct signals	19	19	19	19
Proportion of correct signals	66.67%	77.78%	88.89%	88.89%
Test statistics	1.8093	2.6753	4.904*	3.8131†
p-value	17.86%	10.19%	2.68%	5.09%

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

Table 14: **SHCOMP: Accuracy and statistical significance of the P/E ratio and log P/E ratio as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, both the P/E ratio generated 19 signals, and the log P/E ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.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
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	$3.2387^{\dagger}$	7.19%

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

Table 15: **SZECOMP: Maximum likelihood estimate and likelihood ratio test for the PE and logPE**

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	
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 16: **SZECOMP:Monte Carlo likelihood ratio test for the PE and logPE**

Confidence	0.8	0.85	0.9	0.925	0.95	0.975	0.99
				P/E ratio			
Number of signals	12	12	9	11	9	7	7
Number of correct signals	11	11	8	10	8	6	6
Proportion of correct signals	91.67%	91.67%	88.89%	90.91%	88.89%	85.71%	85.71%
Test statistics	6.6736**	6.6736**	4.0607*	5.7831*	4.0607*	2.4528	2.4528
p-value	0.98%	0.98%	4.39%	1.62%	4.39%	11.73%	11.73%
				logP/E ratio			
Number of signals	11	10	9	10	9	8	6
Number of correct signals	10	9	8	9	8	7	5
Proportion of correct signals	90.91%	90.00%	88.89%	90.00%	88.89%	87.50%	83.33%
Test statistics	5.7831**	4.9107*	4.0607*	4.9107*	4.0607*	3.2387†	1.7150
p-value	1.62%	2.67%	4.39%	2.67%	4.39%	7.19%	19.03%

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

Table 17: **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 $p_0 = 58.49\%$ 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.

Horizon (days)	63	126	189	252
Uninformed probability p_0	19.56%	35.03%	48.79%	58.49%
		P/E ratio		
Number of correct signals	9	9	9	9
Proportion of correct signals	44.44%	66.67%	77.78%	88.89%
Test statistics	$2.8646^{\dagger}$	$3.7184^{\dagger}$	$3.189^{\dagger}$	4.0607^{*}
p-value	9.05%	5.38%	7.41%	4.39%
		logP/E ratio		
Number of correct signals	8	8	8	8
Proportion of correct signals	44.44%	66.67%	77.78%	88.89%
Test statistics	$2.8646^{\dagger}$	$3.7184^{\dagger}$	$3.189^{\dagger}$	4.0607^{*}
p-value	9.05%	5.38%	7.41%	4.39%

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

Table 18: **SZECOMP: Accuracy and statistical significance of the P/E ratio and log P/E ratio as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, both the P/E ratio generated 19 signals, and the log P/E ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.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
BSEYD0	4	3	75%	1.05E-01	0.717	0.6654	41.47%
logBSEYD0	5	2	40%	3.46E-02	0.7901	0.4713	49.24%
PE0	4	3	75%	1.05E-01	0.717	0.6654	41.47%
logPE0	4	3	75%	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%	3.46E-02	0.9778	0.0449	83.23%
CAPE10	3	3	100.00%	-	-	-	-
logCAPE10	3	3	100.00%	-	-	-	-

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

Table 19: **SHCOMP: Maximum likelihood estimate and likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**

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	
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	-

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

Table 20: **SHCOMP: Monte Carlo likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**

	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.8	0.85	0.9	0.925	0.95	0.975	0.99
	BSEYD0							logBSEYD0						
Number of signals	7	6	6	4	4	2	2	6	5	5	5	5	4	4
Number of correct signals	4	3	3	3	3	2	2	3	2	2	2	2	2	2
Proportion of correct signals	57.14%	50%	50%	75%	75%	100%	100%	50%	40%	40%	40%	40%	50%	50%
Test statistics	0.0095	0.0681	0.0681	0.6654	0.6654	-	-	0.0681	0.4713	0.4713	0.4713	0.4713	0.0454	0.0454
p-value	92.22%	79.42%	79.42%	41.47%	41.47%	-	-	79.42%	49.24%	49.24%	49.24%	49.24%	83.13%	83.13%
	PE0							logPE0						
Number of signals	5	5	4	4	4	3	3	5	5	4	4	4	3	2
Number of correct signals	3	4	3	3	3	3	3	3	4	3	3	3	3	2
Proportion of correct signals	60%	80%	75%	75%	75%	100%	100%	60%	80%	75%	75%	75%	100%	100%
Test statistics	0.0449	1.3445	0.6654	0.6654	0.6654	-	-	0.0449	1.3445	0.6654	0.6654	0.6654	-	-
p-value	83.23%	24.62%	41.47%	41.47%	41.47%	-	-	83.23%	24.62%	41.47%	41.47%	41.47%	-	-
	BSEYD10							logBSEYD10						
Number of signals	7	7	6	4	3	4	3	7	4	5	6	5	4	4
Number of correct signals	3	3	2	3	2	3	3	3	3	3	4	3	3	3
Proportion of correct signals	42.86%	42.86%	33.33%	75%	66.67%	75%	100%	42.86%	75%	60%	66.67%	60%	75%	75%
Test statistics	0.436	0.436	1.1741	0.6654	0.1606	0.6654	-	0.436	0.6654	0.0449	0.3212	0.0449	0.6654	0.6654
p-value	50.91%	50.91%	27.86%	41.47%	68.86%	41.47%	-	50.91%	41.47%	83.23%	57.09%	83.23%	41.47%	41.47%
	CAPE10							logCAPE10						
Number of signals	4	3	4	4	3	3	3	4	3	4	3	3	3	2
Number of correct signals	3	3	4	4	3	3	3	3	3	4	3	3	3	2
Proportion of correct signals	75%	100%	100%	100%	100%	100%	100%	75%	100%	100%	100%	100%	100%	100%
Test statistics	0.6654	-	-	-	-	-	-	0.6654	-	-	-	-	-	-
p-value	41.47%	-	-	-	-	-	-	41.47%	-	-	-	-	-	-

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

Table 21: **SHCOMP: Accuracy and statistical significance of the prediction models 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 $p_0 = 55.31\%$ 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.

	63	126	189	252	63	126	189	252
Uninformed probability p_0	22.21%	33.1%	44.2%	55.31%	22.21%	33.1%	44.2%	55.31%
	BSEYD0				logBSEYD0			
Number of signals	4	4	4	4	5	5	5	5
Number of correct signals	2	3	3	3	1	2	2	2
Proportion of correct signals	50%	75%	75%	75%	20%	40%	40%	40%
Test statistics	1.4776	2.9393	1.5663	0.6654	0.0145	0.1043	0.0361	0.4713
p-value	22.41%	8.64%†	21.07%	41.47%	90.41%	74.67%	84.93%	49.24%
	PE0				logPE0			
Number of signals	4	4	4	4	4	4	4	4
Number of correct signals	1	3	3	3	2	3	3	3
Proportion of correct signals	25%	75%	75%	75%	50%	75%	75%	75%
Test statistics	0.0175	2.9393	1.5663	0.6654	1.4776	2.9393	1.5663	0.6654
p-value	89.48%	8.64%†	21.07%	41.47%	22.41%	8.64%†	21.07%	41.47%
	BSEYD10				logBSEYD10			
Number of signals	3	3	3	3	5	5	5	5
Number of correct signals	1	2	2	2	2	3	3	3
Proportion of correct signals	33.33%	66.67%	66.67%	66.67%	40%	60%	60%	60%
Test statistics	0.1947	1.4076	0.6132	0.1606	0.7951	1.5118	0.5019	0.0449
p-value	65.90%	23.55%	43.36%	68.86%	37.26%	21.89%	47.87%	83.23%
	CAPE10				logCAPE10			
Number of signals	3	3	3	3	3	3	3	3
Number of correct signals	2	3	3	3	2	3	3	3
Proportion of correct signals	66.67%	100%	100%	100%	66.67%	100%	100%	100%
Test statistics	2.7014	-	-	-	2.7014	-	-	-
p-value	10.03%	-	-	-	10.03%	-	-	-

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

Table 22: **SHCOMP: Accuracy and statistical significance of the BSEYD and log BSEYD as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, the BSEYD ratio generated 4 signals, and the log BSEYD ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.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
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 23: **SZECOMP: Maximum likelihood estimate and likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**

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	
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 24: **SZECOMP: Monte Carlo likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**

	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.8	0.85	0.9	0.925	0.95	0.975	0.99
	BSEYD0							logBSEYD0						
Number of signals	8	6	7	6	6	5	4	5	6	7	8	7	4	4
Number of correct signals	7	6	7	6	6	5	4	5	6	7	8	7	4	4
Proportion of correct signals	87.5%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Test statistics	1.7971	-	-	-	-	-	-	-	-	-	-	-	-	-
p-value	18.01%	-	-	-	-	-	-	-	-	-	-	-	-	-
	PE0							logPE0						
Number of signals	8	9	6	7	6	6	6	8	8	6	6	6	7	4
Number of correct signals	8	9	6	7	6	6	6	7	8	6	6	6	7	4
Proportion of correct signals	100%	100%	100%	100%	100%	100%	100%	87.5%	100%	100%	100%	100%	100%	100%
Test statistics	-	-	-	-	-	-	-	1.7971	-	-	-	-	-	-
p-value	-	-	-	-	-	-	-	18.01%	-	-	-	-	-	-
	BSEYD10							logBSEYD10						
Number of signals	7	6	5	6	7	4	3	10	9	7	8	7	3	3
Number of correct signals	7	6	5	5	6	4	3	10	8	7	8	7	3	3
Proportion of correct signals	100%	100%	100%	83.33%	85.71%	100%	100%	100%	88.89%	100%	100%	100%	100%	100%
Test statistics	-	-	-	0.8162	1.2826	-	-	-	2.3477	-	-	-	-	-
p-value	-	-	-	36.63%	25.74%	-	-	-	12.55%	-	-	-	-	-
	CAPE10							logCAPE10						
Number of signals	8	9	8	9	6	5	4	8	9	6	6	5	4	3
Number of correct signals	8	9	8	9	6	5	4	8	9	6	6	5	4	3
Proportion of correct signals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Test statistics	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p-value	-	-	-	-	-	-	-	-	-	-	-	-	-	-

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

Table 25: **SZECOMP: Accuracy and statistical significance of the prediction models 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 $p_0 = 66.99\%$ 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.

	63	126	189	252	63	126	189	252
Uninformed probability p_0	24.93%	44.22%	58.47%	66.99%	24.93%	44.22%	58.47%	66.99%
	BSEYD0				logBSEYD0			
Number of signals	6	6	6	6	7	7	7	7
Number of correct signals	3	5	6	6	4	6	7	7
Proportion of correct signals	50%	83.33%	100%	100%	57.14%	85.71%	100%	100%
Test statistics	1.7367	3.9209	-	-	3.2717	5.2181	-	-
p-value	18.76%	4.77%*	-	-	7.05%†	2.24%*	-	-
	PE0				logPE0			
Number of signals	6	6	6	6	6	6	6	6
Number of correct signals	2	5	6	6	2	5	6	6
Proportion of correct signals	33.33%	83.33%	100%	100%	33.33%	83.33%	100%	100%
Test statistics	0.212	3.9209	-	-	0.212	3.9209	-	-
p-value	64.52%	4.77%*	-	-	64.52%	4.77%*	-	-
	BSEYD10				logBSEYD10			
Number of signals	7	7	7	7	7	7	7	7
Number of correct signals	4	5	6	6	4	5	7	7
Proportion of correct signals	57.14%	71.43%	85.71%	85.71%	57.14%	71.43%	100%	100%
Test statistics	3.2717	2.1193	2.4553	1.2826	3.2717	2.1193	-	-
p-value	7.05%†	14.54%	11.71%	25.74%	7.05%†	14.54%	-	-
	CAPE10				logCAPE10			
Number of signals	6	6	6	6	5	5	5	5
Number of correct signals	3	5	6	6	2	4	5	5
Proportion of correct signals	50%	83.33%	100%	100%	40%	80%	100%	100%
Test statistics	1.7367	3.9209	-	-	0.5465	2.6916	-	-
p-value	18.76%	4.77%*	-	-	45.98%	10.09%	-	-

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

Table 26: **SZECOMP: Accuracy and statistical significance of the BSEYD and log BSEYD as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, the BSEYD ratio generated 4 signals, and the log BSEYD ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.The subsequent rows present the same information for the log P/E ratio.