

Does the Bond-Stock Earnings Yield Differential Model Predict Equity Market Corrections Better Than High P/E Models?*

Sébastien Lleo[†] and William T. Ziemba[‡]

Friday 31st October, 2014

*We thank Harry Markowitz for suggesting the idea for this paper. We are also indebted to George Constantinides, Christophe Faugère and Thorsten Hens for useful methodological comments. Finally, we thank the participants to the third F.E.B.S. Conference held at the University of Surrey in June 22, 2014 and participants to the Stochastic Optimization: Models and Algorithms Workshop held as part of the Hausdorff Institute Trimester Program on Stochastic Dynamics in Economics and Finance for their helpful comments. The first author gratefully acknowledges the support from Région Champagne Ardennes and the European Union through the RiskPerform grant.

[†]Finance Department, NEOMA Business School - Campus Reims, 59 rue Pierre Taittinger, 51100 Reims, France, Email: sebastien.lleo@NEOMA-bs.fr

[‡]Alumni Professor of Financial Modeling and Stochastic Optimization (Emeritus), University of British Columbia, Vancouver, BC and Distinguished Visiting Scholar, Systemic Risk Centre, London School of Economics, Email: wtzimi@mac.com

Does the Bond-Stock Earnings Yield Differential Model Predict Equity Market Corrections Better Than High P/E Models?

Abstract

In this paper, we extend the literature on crash prediction models in three main respects. First, we relate explicitly crash prediction measures and asset pricing models. Second, we present a simple, effective statistical significance test for crash prediction models. Finally, we propose a definition and a measure of robustness for crash prediction models. We apply the statistical test and measure the robustness of selected model specifications of the Price-Earnings (P/E) ratio and Bond Stock Earning Yield Differential (BSEYD) measures. This analysis shows that the BSEYD, the logarithmic BSEYD model, and to a lesser extent the P/E ratio, were statistically significant robust predictors of crashes on the US equity market over the period 1964 to 2012.

Keywords: stock market crashes, bond-stock earnings yield mode, Fed model, price-earnings-ratio

JEL classification: G14, G15, G12, G10.

This paper tests statistically the ability of traditional measures to predict equity market crashes using several competing model specifications, proposes a measure of robustness and investigates whether equity market crashes are predictable events. For long-term investors, the ability to forecast market corrections and crashes reduces the risk of unsustainably large losses, and improves risk-adjusted performance (Berge, Consigli, and Ziemba, 2008). The ability to predict crashes should also matter to economists because the S&P500 is a leading indicator of economic growth (Klein and Moore, 1982; Hertzberg and Beckman, 1989; Ozyildirim, McGuckin, and Zarnowitz, 2003). Therefore, predicting equity market crashes may improve money managers' performance and economists' ability to forecast economic downturns.

Markowitz (1952) begins his seminal paper on portfolio selection by stating: “the process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performance of available securities. The second stage starts with the relevant beliefs about future performance and ends with the choice of a portfolio.” While Markowitz has been extremely successful at addressing the second stage, questions related to the predictability of financial markets have spurred major developments in the theory of financial economics.

Fama (1965, 1970) led the research into the short-term behaviour of financial asset prices and introduced the efficient market hypothesis. Considerable efforts have also gone into understanding to what extent long-term returns are predictable. Table 1 motivates this idea empirically by reporting the evo-

lution of the Price-Earnings (P/E) ratio over selected 20-year periods with high annualized returns. In each period the P/E ends 1.6 to 4.7 times higher than it started. Furthermore, there is a 90% correlation between annualized returns and ending P/E ratio. The 1980s witnessed major advances related to the predictability of long-term returns through the research of Flood, Hodrick, and Kaplan (1986), Campbell and Shiller (1998, 1988, 1989), Poterba and Summers (1989), Fama and French (1988a,b, 1992) and Balvers, Wu, and Gilliland (2000) among others. In parallel, Ziemba and Schwartz (1991) published the first crash prediction model, called the bond-stock earnings yield differential model (BSEYD). The focus of the BSEYD is not on forecasting a specific level of performance but on predicting a specific type of rare market event. Despite their diverging objectives, return and crash prediction models share a common starting point: an empirical analysis of historical market data. The works by Shiller (1996, 2005, 2006), the growing popularity of the Fed model and the internet stock crash in 2000-2002 further cemented the relation between return and crash prediction models.

[Place Table 1 here]

Interest in the BSEYD model, the Fed model and the models proposed by Campbell and Shiller has not waned with the advent of newer approaches.

⁰Yardeni (1997) originally used the term “Fed model” to refer to a comment made by Fed Chairman Alan Greenspan during his Humphrey-Hawkins testimony on July 22, 1997.

Koivu, Pennanen, and Ziemba (2005) show that the Fed model has some predictive power with respect to equity prices, earnings and bond yields. Shiller (2006) revisits earlier works with John Campbell (Campbell and Shiller, 1998, 1988, 1989): his model based on real (as opposed to nominal) quantities has an R^2 of 0.566. Shiller observes that low P/E periods appear to lead to higher future stock prices and high P/E periods to lower future stock prices. Consigli, MacLean, Zhao, and Ziemba (2009) propose a stochastic model of equity returns based on an extension of the BSEYD model inclusive of a risk premium in which market corrections are endogenously produced by the bond-stock yield difference. This model accommodates both the case of prolonged yield deviations leading to a long series of small declines in the equity market and that of short-term corrections, which specific to recent speculative bubbles. The authors test the yield differential as a key driver of the market correction process and validate their model with market data. They estimate the current fair value of the S&P500 to predicts stock market crashes and rallies as well as the good times to be in and out of the stock markets. Lleo and Ziemba (2012) show that the BSEYD measure predicts the equity market crashes that occurred in the US, Iceland and China during 2007-2009. Maio (2013) also shows that the “yield gap”, which corresponds to the difference between the earnings yield (or dividend yield) on a stock market index and the long-term yield on Treasury bonds, predicts returns better than other models based on dividend yield.

All three models are empirical: Campbell and Shiller derive a number of

their observations from a linear regression of long-term log returns against the logarithm of the P/E ratio, while the competing asset argument often used to explain the success of the BSEYD model has more to do with pragmatic intuition than economic theory. The first contribution of this paper is to relate these models to asset pricing theory. This is a necessary step if we plan to test their predictive ability in a meaningful way. Lleo and Ziemba (2012) partially address this gap by deriving the relation between the BSEYD measure, the Fed model and the Gordon (1959) Growth model. This paper derives the Campbell-Shiller model from the Gordon Growth model to place it on a firm theoretical footing.

The financial crisis of 2007-2009 and the development of a new generation of probabilistic models have once again put crash prediction models in focus. Among the newer models, Jarrow, Kchia, and Protter (2011a,b,c) and Jarrow (2012) identify a bubble by testing whether the underlying price process is a martingale or a local martingale under the risk neutral measure. The key to this approach is a sophisticated volatility estimation technique. The authors identify bubbles in LinkedIn's stock prices and in gold prices. Shiryaev, Zitlukhin, and Ziemba (2014, 2015) use the discrete time disorder detection model developed by Shiryaev and Zhitlukhin (2012a,b). This optimal stopping model estimates the time when the drift of the price process switches sign, from positive to negative. Shiryaev et al. (2014) show that the disorder model provides good exit points for investors in the NASDAQ 100 circa 2000-2002 and in the Apple computer stock in 2012. Shiryaev

et al. (2015) also extend the analysis to the Japanese stock, land and golf membership markets circa 1990 and in 2013. They show that the disorder model produces clear entry and exit signals. These more recent models differ from the Campbell-Shiller, BSEYD and Fed models in that they rely on a probabilistic representation of the price process rather than on underlying economic variables. This paper focuses on the Campbell Shiller, BSEYD and Fed models.

Traditionally, correction prediction models have been judged on the number or proportion of crashes they succeeded in forecasting, or on their contribution to portfolio performance, rather than tested statistically. For example, Berge et al. (2008) found that the strategy “move to cash when the model indicates a crash and invest in an equity index the rest of the time” provides about twice the final wealth with less variance and a higher Sharpe ratio than a buy-and-hold strategy over the period 1975-2005 and 1980-2005 in five countries considered in the study. On the other hand, long-term prediction models, such as the Campbell-Shiller model, were not intended as a crash prediction model but as a prediction model for long-term returns. As a result, their ability to predict equity market corrections has not been tested.

This paper extends the literature on crash prediction models in three main respects. First, it explicitly relates crash prediction measures, long-term return prediction models and asset pricing models. Second, this paper proposes a simple and effective statistical significance test for crash prediction models (or any event prediction model) and uses it to test the ability of

the Campbell-Shiller and BSEYD models to predict market corrections. An innovation in this paper is that the statistical tests are conducted with respect to events, namely equity market corrections, as opposed to returns. This orientation has an impact on the structure of the results and nature of the statistical tests. Fundamentally, one can give correction prediction models a binary interpretation: a market crash either occurs or not, a correction signal either occurs or not. This means that correction prediction models are nonlinear but simple. Using this binary interpretation, it is possible to conduct the tests efficiently using a likelihood ratio. However, crashes do not occur frequently and we can expect that the number of predictions will be small, which exposes the analysis to small sample bias. To address this problem, this study uses Monte Carlo methods to obtain the empirical distribution of test statistics¹.

Third, this paper proposes a definition and a measure of robustness for event prediction models. It distinguishes between two closely connected views of robustness: *data robustness*, which represents robustness with respect to the dataset, and *model specification robustness*, defined as robustness with respect to the specific parameters of the model. Data robustness addresses the question of the sensitivity of model predictions when the underlying dataset changes. This is the more traditional view of robustness, and it is generally

¹Monte Carlo methods are the method of choice to identify and correct small sample bias. For instance, Campbell and Shiller (1989) used it to confirm the statistical significance of their study of the predictability of long-term returns.

dealt with using out-of-sample studies, or by studying the predictions of a model over two different data sets or time periods. This paper addresses data robustness with an out-of-sample analysis. It also proposes a novel measure of model specification robustness based on a robust likelihood ratio and its associated test statistics. This new measure enables us to test statistically the robustness of the Campbell-Shiller and BSEYD models.

The paper is organized as follows. Section I describes the two fundamental models: the Campbell-Shiller model and the BSEYD model. Section II introduces the dataset and methodology used to identify equity market crashes, generate the time series of crash signals, test statistically the predictive ability of the crash prediction models and analyze their robustness. Section III presents the empirical results and Section IV discusses the results and their implications.

I. A Tale of Two Models: Campbell-Shiller and BSEYD

A. Campbell Shiller Model

Campbell and Shiller (1988) propose a vector-autoregressive model relating the log return on the S&P500 with the log dividend-price ratio, lagged dividend growth rate and average annual earnings over the previous 30 years. Campbell and Shiller start by performing a regression of the log returns on

the S&P 500 at 1 year, 3 year and 10 year horizons against each of these variables and the average annual earnings over the previous 10 years. Campbell and Shiller (1988) find 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 is significant.

Campbell and Shiller define the one-period total return on the stock as

$$h_{1t}^{beg} := \ln \left(\frac{P_{t+1}^{beg} + D_t}{P_t^{beg}} \right),$$

where P_t^{beg} is the price of the stock at the beginning of period t and D_t is the dividend received during period t . Although Campbell and Shiller do not specify a present value or future value rule for the dividend, we will assume that all dividends received through period t are either carried or future valued to the end of period t , so that $D_t = D_t^{end}$. The i period total return on the stock is

$$h_{it}^{beg} := \sum_{j=0}^{i-1} h_{1,t+j}.$$

The regression in Campbell and Shiller (1988) is

$$(1) \quad h_{it}^{beg} = a + b \ln \left(\frac{P_t^{beg}}{E_{t,-n}^{beg}} \right) + \epsilon,$$

where P_t is the level of the S&P 500 index at time t and $\overline{E_{t,-n}^{beg}}$ is the average of the past annual earnings over the last n years, namely

$$\overline{E_{t,-n}^{beg}} = \frac{1}{n} \sum_{i=0}^{n-1} E_{t-i}^{beg}.$$

The R^2 computed by Campbell and Shiller (1988) for $n = 30$ is 0.566, higher than the 0.401 computed for $n = 10$ and higher than the R^2 of regressions against the log dividend-price ratio and lagged dividend growth rate. Hence, the regression based on 30-year average earnings has a higher explanatory power than the regression based on 10 years of earnings. To explain this difference Shiller (1996) suggests that a 10-year average may still be sensitive to changes in the business cycle, whereas a 30-year average should be immune to short term shifts in the business cycle.

Campbell and Shiller do not explain how these empirical results fit with asset pricing theory. To fill this gap and put their observations on a firm theoretical footing, we use the Gordon (1959) growth model to derive an approximate linear relation between the one-period log return on the S&P500 and the earnings yield. Under the assumptions of the Gordon Growth Model, and using standard finance notation, the price of a stock at time t equals

$$P_t^{end} = \frac{D_{t+1} + P_{t+1}^{end}}{1 + k},$$

where where P_t^{end} is the price of the stock at the beginning of period t , and

k is the cost of equity. This relation implies that the holding period return is linear in the current earnings yield:

$$\frac{P_{t+1}^{end} + D_t}{P_t^{end}} = 1 + k - gd\rho_t.$$

where g is the constant growth rate of dividends, d is the dividend payout ratio and $\rho_t^{end} = \frac{E_t^{end}}{P_t^{end}}$. The logarithmic return

$$h_{1t}^{end} = \ln\left(\frac{P_{t+1}^{end} + D_{t+1}}{P_t^{end}}\right) = \ln(1 + k) + \ln\left(1 - \frac{gd}{k}\rho_t^{end}\right).$$

We linearize this expression by performing a first order Taylor expansion around $\rho_t = \bar{\rho}$, where $\bar{\rho}$ is the average long-term earnings yield

$$(2) \quad h_{1t}^{end} \approx a_0 - a_1\rho_t^{end},$$

where $a_0 := \ln(1 + k) + \ln\left(1 - \frac{gd}{1+k}\bar{\rho}\right) + \frac{gd}{1+k-gd\bar{\rho}}$ and $a_1 := \frac{gd}{1+k-gd\bar{\rho}}$.

Provided that the price at the end of period t and at the beginning of period $t + 1$ are equal, that is $P_t^{end} = P_{t+1}^{beg}$, then it follows that $\rho_t^{end} = \rho_{t+1}^{beg}$ and

$$(3) \quad h_{1t}^{beg} = \ln\left(\frac{P_{t+1}^{beg} + D_t}{P_t^{beg}}\right) = \ln\left(\frac{P_t^{end} + D_t}{P_{t-1}^{end}}\right) = h_{1(t-1)}^{end}$$

Then the two models are equivalent, up to a small change in notation.

Equation (2) differs from the Campbell-Shiller regression model (1) in two

important aspects. First, the Campbell-Shiller model regresses log returns against $\ln P/E$ while the Gordon model relates log returns to earning yield (no log). Second, the Campbell-Shiller model uses the P/E ratio whereas the Gordon model is based on the earnings yield. However, the impact of this second difference is minor. The properties of the logarithm imply that

$$(4) \quad \ln(h_{it}) = a + b \ln\left(\frac{P_t}{\bar{E}_{t,-n}}\right) + \epsilon = a - b \ln\left(\frac{\bar{E}_{t,-n}}{P_t}\right) + \epsilon.$$

Hence only the regression slope changes sign. The significance of the model is not affected.

B. The Bond-Stock Earnings Yield Differential (BSEYD) Model

The Bond-Stock Earnings Yield (BSEYD) relates the yield on stocks (measured by the earnings yield) to the yield on nominal Treasury bonds.

$$(5) \quad BSEYD(t) = r(t) - \rho(t) = r(t) - \frac{E(t)}{P(t)},$$

where $\rho(t)$ is the earnings yield at time t and $r(t)$ is the most liquid (10- or 30-year) Treasury bond rate $r(t)$. We have removed the subscript used to identify the timing of the cash flow to clarify the notation.

Similarly to the Campbell-Shiller model, the BSEYD uses price and earnings. The BSEYD also factors in prevailing interest rate levels. We can also

use the Gordon growth model to relate the BSEYD model to the equity risk premium, earnings growth and government bond yield:

$$BSEYD = \frac{1}{d(1+g)} (r_t [d(1+g) - 1] - f_t + g),$$

where d is the payout ratio, g is the earnings growth, r is the yield on a government bond and f is the equity risk premium.

The BSEYD model is closely related to the Fed Model (Estrada, 2006; Weigand and Irons, 2007; Faugère and Van Erlach, 2009; Maio, 2013; Faugère, 2013). In its most popular form, the Fed model states that in equilibrium, the one year forward looking earnings yield of the S&P500 should equal the current yield on a 10-year Treasury Note, that is

$$(6) \quad r(t) - \frac{E_t^f}{P(t)} = 0,$$

where E_t^f is the S&P 500 one year forward looking earnings. When computed using one year forward looking earnings, the BSEYD measures the distance between current market conditions and equilibrium conditions. The Fed Model is a special case of the BSEYD measure, with $r(t) = \frac{E_t^f}{P(t)}$.

II. Data and Crash Prediction Methodology

A. Dataset

In our analysis, we use daily S&P 500 data from January 1, 1950 to December 31, 2012, and earnings and P/E data for the period January 29, 1954 to December 31, 2012. We obtained the data from Bloomberg and cross referenced and checked them with Thomson DataStream. The daily yields on the ten-year Treasury Note for the period January 1, 1962 to December 31, 2012 and the daily yields on seasoned corporate bonds rated Aaa by Moody's for the period January 3, 1983 to December 31, 2012 from the Board of Governors of the Federal Reserve System. Overall, the longest period for which we can test the accuracy of crash prediction models lasts for 51 years, starting January 1, 1962. It comprises 12,846 daily data points. The shortest period for which we can test the accuracy of crash prediction models lasts for 30 years, starting January 3, 1983.

B. Definition of a Crash

We define an equity market crash as a decline of at least 10% in the level of the S&P500 over a time period of at most a year (252 trading days). Table 2 shows the 18 crashes that occurred between January 1, 1962 and December 31, 2012. We observe the October 1987 crash, the Russian currency default and LTCM in 1998, the dot com-Nasdaq 100 collapse in 2000-2002 and the subprime crisis in 2007-2009. The average duration of a crash was 285 days,

and the average decline was 25%.

[Place Table 2 here]

Table 2 gives the crash identification date, defined as the date at which the decline from the previous local peak reaches 10%. With this information, we define a crash indicator sequence $C = \{C_t, t = 1, \dots, T\}$, where C_t takes value 1 if date t is a crash identification date and 0 otherwise, as the vector $c = (C_1, \dots, C_t, \dots, C_T)$. Thus, the event “a crash is identified on day t ” is represented as $\{C_t = 1\}$. This variable plays a leading role in the hypothesis test. We repeat this procedure for each day in our sample. The end result is a binary sequence with “1” in eighteen entries, one for each crash, and “0” everywhere else.

C. Signal Construction

C.1. Definition of Signal

Crash prediction models such as the BSEYD model or the continuous time disorder detection model generate a signal to indicate a downturn in the equity market at a given horizon h . This signal occurs whenever the value of a specific measure (in our case P/E or BSEYD) crosses a threshold.

Given a crash prediction measure $M(t)$, a crash signal occurs whenever

$$(7) \quad \text{SIGNAL}(t) = M(t) - K(t) > 0$$

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 horizon H between the occurrence of the signal and that of the crash.

C.2. The Measure $M(t)$

We test four measures in this paper:

1. *P/E ratio*: We use the P/E as a control variable, as its explanatory power with respect to stock over- and under-performance is well documented (Fama and French, 1992);
2. *log P/E*, computed as the logarithm of the P/E ratio. Campbell and Shiller's regression model has already demonstrated its predictive power on long term returns, but the ability of the logarithm of the P/E ratio to predict crashes has not been tested so far.
3. *BSEYD*, defined in Equation (5) as

$$(8) \quad BSEYD(t) = r(t) - \rho(t) = r(t) - \frac{E(t)}{P(t)}$$

The BSEYD was designed as a crash prediction model but its predictive power has not been tested statistically and it has not been compared with other models.

4. *logBSEYD*, defined by

$$(9) \quad \log BSEYD(t) = \ln \frac{r(t)}{\rho(t)} = \ln r(t) - \ln \frac{E(t)}{P(t)}$$

The log BSEYD is the logarithmic of the BSEYD model. The log BSEYD model has previously been used to predict returns (Koivu et al., 2005) but not crashes.

The first two measures contain information about current prices and earnings while the last two also factor in interest rates. For the interest rates, we use the yield on the ten-year Treasury Note. We use both current earnings and average earnings over the past ten years. This is consistent with the observation that the cyclical nature of earnings matters. Graham and Dodd (1934), for example, suggest using ten years of earnings in stock and company valuation. Although Campbell and Shiller tested their model using average earnings over 30 years, we stop at 10 years because our objective is to predict medium term market downturns rather than long-term market returns. As a result, we do not necessarily want the measures we consider to be immune from the phases of the business cycle.

C.3. The Threshold $K(t)$

We define the threshold as a confidence level based on a rolling horizon sample mean and standard deviation. First, we consider a standard 95% one-tail confidence interval based on a Normal distribution, consistently with earlier works including Ziemba and Schwartz (1991) and Lleo and Ziemba (2012).

A standard confidence level does not provide a robust threshold because the tails of the distribution of the measures may not be approximately Gaussian. To address this problem, we also use Cantelli's inequality, a one-tailed version of Chebyshev's inequality, used to derive a 'worst case' confidence level for heavily skewed or leptokurtic distributions (see Grimmett and Stirzaker (2001), Problem 7.11.9). Cantelli's 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:

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

Setting $\alpha = \frac{1}{1+k^2}$ yeilds

$$P \left[X - \mu \geq \sigma \sqrt{\frac{1}{\alpha} - 1} \right] \leq \alpha.$$

The parameter α provides an upper bound for a one-tailed confidence level on any distribution. We select $\alpha = 25\%$ which produces a slightly higher threshold than the standard confidence interval. In a Normal distribution,

we expect that 5% of the observations lie in the right tail whereas Cantelli's inequality implies that the percentage of outliers in a non Gaussian distribution may reach 25%.

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, as discussed by Lleo and Ziemba (2012). Using rolling horizon means and standard deviations has the advantage of providing data consistency. In particular, rolling horizon mean and standard deviation are not overly sensitive to the starting date of the calculation. Most importantly, this construction addresses the in-sample versus out-sample problem by only using past data and predetermined parameters. Therefore, the h -day moving average at time t , denoted by μ_t^h , and the corresponding rolling horizon standard deviation σ_t^h are

$$\mu_t^h = \frac{1}{h} \sum_{i=0}^{h-1} x_{t-i},$$

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

In this paper, we take $h = 252$. This horizon is short and can be used on most financial markets without requiring a long data history.

C.4. The Horizon H

The last parameter we need to specify is the horizon between the signal and the crash identification date. The crash identification time is the date at which the S&P500 has declined by at least 10% within one 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 choose a two-year horizon (504 trading days) between the signal and the crash identification date. This choice implies that a signal could occur at any time between one trading day prior to the crash identification date (up to 251 days after the associated local market high), and 503 trading days prior to the crash identification date (up to 502 days prior to the associated local market high).

C.5. Conclusion on Signal Constuction

Once we have computed the signals, the last step is to define a signal indicator sequence $S = \{S_t, t = 1, \dots, T\}$. We start by recording the first day in a series of positive signals as the signal date and only count 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 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 $s = (S_1, \dots, S_t, \dots, S_T)$.

In this paper, we test a total 24 different parametrizations of our crash

prediction models. These parameterizations differ in the definition of earnings (current or ten year average), the type of confidence interval (standard or Cantelli) and the type of interest rates used (yield on a ten-year Treasury Note or yield on a seasoned Aaa corporate bond):

1. P/E 1 and $\log P/E$ 1 are computed using current earnings and a standard confidence level;
2. P/E 2 and $\log P/E$ 2 use current earnings and Cantelli's inequality;
3. P/E 3 and $\log P/E$ 3 are calculated with average earnings over ten years and a standard confidence level;
4. P/E 4 and $\log P/E$ 4 are based on average earnings over ten years and Cantelli's inequality.

The BSEYD and $\log BSEYD$ measures are implemented in eight models each:

1. BSEYD 1 and $\log BSEYD$ 1 are computed using current earnings, a standard confidence level and yields on 10-year Treasury Notes;
2. BSEYD 2 and $\log BSEYD$ 2 use current earnings, Cantelli's inequality;
3. BSEYD 3 and $\log BSEYD$ 3 are calculated with average earnings over ten years, a standard confidence level and yields on 10-year Treasury Notes;

4. BSEYD 4 and logBSEYD 4 are based on average earnings over ten years, a Cantelli's inequality and yields on seasoned Aaa corporate bonds;
5. BSEYD 5 and logBSEYD 5 are computed using current earnings, a standard confidence level and yields on seasoned Aaa corporate bonds;
6. BSEYD 6 and logBSEYD 6 use current earnings, Cantelli's inequality and yields on seasoned Aaa corporate bonds;
7. BSEYD 7 and logBSEYD 7 are calculated with average earnings over ten years, a standard confidence level and yields on seasoned Aaa corporate bonds;
8. BSEYD 8 and logBSEYD 8 are based on average earnings over ten years, a Cantelli's inequality and yields on seasoned Aaa corporate bonds.

D. Statistical Test

D.1. Construction of the Hit Sequence X

It is well known that predicting returns is extremely difficult but earlier research suggests that it is possible to predict crashes based on a signal. To test this result, we perform a simple and effective statistical test of significance.

All crash prediction models have effectively two components: (1) a signal which takes value 1 or 0 depending on whether the measure has crossed the

confidence level, and (2) a crash indicator which takes value 1 when an equity market crash has occurred and 0 otherwise. From a probabilistic perspective, these components are Bernoulli random variables.

We denote by $C_{t,H}$ the indicator function returning 1 if at least one crash is identified between time t and time $t + H$. The relation between $C_{t,H}$ and C_t is

$$C_{t,H} := 1 - \prod_{i=1}^H (1 - C_{t+i}).$$

We identify the vector C_H with the sequence $C_H := \{C_{t,H}, t = 1, \dots, T - H\}$ and the vector $c_H := (C_{1,H}, \dots, C_{t,H}, \dots, C_{T-H,H})$.

The accuracy of the crash prediction model is the conditional probability $P(C_{t,H} = 1 | S_t = 1)$ that a crash is 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 draw. We can 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 $\{S_t | S_t = 1, t = 1, \dots, T\}$ and denote by $X := \{X_i, i = 1, \dots, N\}$ the “hit sequence” where $x_i = 1$ if the i th 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 $x = (X_1, X_2, \dots, X_N)$.

D.2. Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$

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 maximised for

$$(10) \quad \hat{p} := \frac{\sum_{i=1}^N X_i}{N}$$

so the maximum likelihood estimate of the probability $p = P(C_{t,H}|S_t)$ is the historical proportion of correct predictions out of all observations.

D.3. Likelihood Ratio Test

We apply a likelihood ratio test to test the null hypothesis $H_0 : p = p_0 := \frac{1}{2}$ against the alternative hypothesis $H_A : p \neq p_0$. The null hypothesis reflects the idea that an arbitrary (uninformed) signal would only correctly predict crashes 50% of the time. A significant departure above this level indicates that the measure we are considering contains some information.

The likelihood ratio test is:

$$(11) \quad \Lambda = \frac{L(p = \frac{1}{2}|X)}{\max_{p \in (0,1)} L(p|X)} = \frac{L(p = \frac{1}{2}|X)}{L(p = \hat{p}|X)}.$$

The statistics $Y := -2 \ln \Lambda$ is asymptotically χ^2 -distributed with $\nu = 1$ degree of freedom. We reject the null hypothesis $H_0 : p = p_0 := \frac{1}{2}$ and accept that the model has some predictive power if $\Lambda > c$, where c is the critical value chosen for the test. We perform the test for the three critical values 3.8415, 6.6349 and 7.8794 corresponding respectively to a 95%, 99% and 99.5% level of confidence level.

A key advantage of this test is that it can be applied to all crash prediction models, whether based on fundamental variables (P/E, BSEYD or Fed model) or on probabilistic models (Jarrow-Protter, Shiryaev-Zitlukhin-Ziembra) and could apply more generally to any model designed to predict a market event.

D.4. Monte Carlo Study for Small Sample Bias

A limitation of the likelihood ratio test is that the χ^2 distribution is only valid asymptotically. In the present case, the number of correct predictions follows a binomial distribution with an estimated probability of success \hat{p} and N trials. Over the past 50 years, 18 crashes occurred and as a result, the total number of signals N should also be low. As a result, the empirical distribution of test statistics will be discrete. The continuous χ^2 distribution might not provide an adequate approximation for this discrete distribution: \hat{p} might appear significantly different from p_0 under a χ^2 distribution but not under the empirical 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 a large number K of paths. For each path $k = 1, \dots, K$, simulate N Bernoulli random variables with probability p_0 of getting a “success.” Denote by $X_k := \{X_i^k, i = 1, \dots, N\}$ the realization sequence where $x_i^k = 1$ if the i th 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

$$(12) \quad \hat{p} := \frac{\sum_{i=1}^N X_i^k}{N},$$

and the test statistics for the path as

$$(13) Y_k = -2 \ln \Lambda_k = -2 \ln \frac{L(p = \frac{1}{2} | X_k)}{\max_{p \in (0,1)} L(p_k | X_k)} = -2 \ln \frac{L(p = \frac{1}{2} | X_k)}{L(p = \hat{p}_k | X_k)}.$$

Once all the paths have been simulated, we use all K test statistics $Y_k, k = 1, \dots, K$ to produce an empirical distributions for the test statistics. From the empirical distribution, we obtain critical values at a 95%, 99% and 99.5% confidence level against which we assess the crash prediction test statistics Y . The empirical distribution also enables us to compute a p -value for the crash prediction test statistics. Finally, we compare the results obtained under the empirical distribution to those derived using the asymptotic χ^2 distribution.

E. Robustness

The question of robustness is critical for all empirical models. In our context, we distinguish two closely connected views of robustness: robustness with respect to the dataset (data robustness) and robustness with respect to the specific parameters of the model (model parameter robustness).

E.1. Addressing Data Robustness: Subperiod Study

Data robustness addresses the question of the sensitivity of model predictions when the underlying dataset changes. This is the more traditional view of robustness, and it is generally dealt with using either an out-of-sample study or by analyzing the behaviour of the model in several subperiods.

To test the robustness of the crash prediction models with respect to a change in the underlying data, we divided the full time period between January 31, 1964 and December 31, 2012 into two subperiods:

1. *Subperiod 1*: January 31, 1964 to December 31, 1981;
2. *Subperiod 2*: January 1, 1982 to December 31, 2012.

Jointly, the two subperiods span the entire dataset. Although the duration of the two subperiods is not equal, with a much longer second subperiod, both subperiods experienced the same number of equity market crashes: nine. This is by far the most important consideration because our statistical tests are based on the “hit sequence” X which only takes signals into account and does not have a time dimension. The duration of the sub period is therefore less important than the number of signals and crashes in each sub period. The last crash in Subperiod 1 occurred on August 25, 1981 while the first crash in Superiod 2 occurred on February 2, 1984. Our construction guarantees that all signals related to the 1984 crash occur in Subperiod 2. No information is lost, and no new bias is introduced in the data or the analysis.

Nonetheless, the choice of December 31, 1981 as a break point is debatable. First, we could have selected December 31, 1987 or December 31, 1988 to ensure that the duration of both subperiods would be of roughly equal duration. The difficulty here is that 11 crashes occurred before December 31, 1987 and only seven after January 1, 1989. With only seven observations, we incur the risk of not having enough data to test our hypothesis. We could

also have chosen December 31, 1982 to get Subperiod 2 to match the dataset available for Aaa corporate yields. This is not practical because a starting date of January 1, 1983 only precedes the next market peak by 10 months and the next crash identification date by less than 14 month. We are well below the 2 years lead time used in our analysis, leaving the possibility that a crash occurring in Subperiod 2 could have been identified in the final year of Subperiod 1.

A final consideration relates to the economic and political context that prevails in both subperiods and that might have an impact on the accuracy of the measures. Subperiod 1 covers a high inflation, high uncertainty period from the mid-1960s to Paul Volcker’s appointment at the head of the Federal Reserve in August 1979. It also includes the suspension of the convertibility of U.S. dollars into gold on August 15, 1971. By contrast, Subperiod 2 is characterized by the so-called “secular” equity Bull market which lasted from 1982 until 2000, and by the various shocks that occurred in the following decade. Ideally, Subperiod 2 should be split into a pre-2000 period and a post-2000 period to reflect the different economic environment. However, we would not have enough data to test the crash prediction models meaningfully.

E.2. Measuring Parameter Robustness: Robust Likelihood Statistics

We propose a novel measure of robustness built on a robust likelihood ratio. The robust likelihood ratio generalizes the (standard) likelihood ratio

to account for the impact of the choice of parameters on the predictions of the model. Compared with the standard likelihood ratio, we need to enlarge the set of model parameters to consider both free parameters, which are the parameters that we seek to estimate based on the model output such as the success probability p , and model specifications, which is the set of parameters that specify the model: type of earnings, threshold calculation, type of bond yields.

Denoting by Ω the enlarged set of parameters, Θ the set of free parameters and \mathcal{M} the set of model specifications, we have $\Omega = (\Theta, \mathcal{M})$. This enlarged formulation leads us to revisit the definition of likelihood function and likelihood ratio. The model specification $\mu \in \mathcal{M}$ directly affects the signal indicator sequence, the total number of signals and the probability of successfully predicting a crash. We denote the signal indicator sequence by $S^\mu := \{S_t^\mu, t = 1, \dots, T\}$ and express the signal indicator sequence as a vector $s^\mu := (S_1^\mu, \dots, S_t^\mu, \dots, S_T^\mu)$. The total number of signals N^μ is

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

and the probability of successfully predicting a crash is $p = P(C_{t,H} = 1 | S_t^\mu = 1)$. As a result the “hit sequence” $X^\mu := \{X_i^\mu, i = 1, \dots, N\}$ where $x_i^\mu = 1$ if the i th signal is followed by a crash and 0 otherwise also depends on the model specification μ .

The likelihood function L associated with the observations sequence X^μ

generated under model specification μ is

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

In our case there is only one free parameter, the probability p , and as a result the set of free parameters Θ is effectively the interval $(0, 1)$. Maximizing the log likelihood function gives us the maximum likelihood estimate of the probability of successfully predicting crashes based on model specification μ

$$(14) \quad \hat{p}^\mu := \frac{\sum_{i=1}^{N^\mu} X_i}{N^\mu}.$$

The standard likelihood ratio under model specification μ and given the null hypothesis $H_0 : p = p_0 := \frac{1}{2}$ is

$$(15) \quad \Lambda^\mu = \frac{L(p = \frac{1}{2}|X, \mu)}{\max_{p \in (0,1)} L(p|X, \mu)} = \frac{L(p = \frac{1}{2}|X, \mu)}{L(p = \hat{p}^\mu|X, \mu)}.$$

We define the robust likelihood ratio $\check{\Lambda}$ given the null hypothesis $H_0 : p = p_0 := \frac{1}{2}$ as the supremum over all possible model specifications of the standard likelihood ratio

$$(16) \quad \check{\Lambda} = \sup_{\mu \in \mathcal{M}} \Lambda^\mu = \sup_{\mu \in \mathcal{M}} \frac{L(p = \frac{1}{2}|X, \mu)}{\max_{p \in (0,1)} L(p|X, \mu)} = \sup_{\mu \in \mathcal{M}} \frac{L(p = \frac{1}{2}|X, \mu)}{L(p = \hat{p}^\mu|X, \mu)}.$$

Higher robust likelihood ratios are indicative of a lower statistical signifi-

cance. The associated robust test statistic is:

$$(17) \quad \check{Y} := \inf_{\mu \in \mathcal{M}} Y^\mu = -2 \ln \sup_{\mu \in \mathcal{M}} \frac{L(p = \frac{1}{2} | X, \mu)}{L(p = \hat{p}^\mu | X, \mu)}.$$

Because the test statistic for the (standard) likelihood ratio under specification μ , Y^μ , is asymptotically χ^2 distributed, the test statistic for robust likelihood ratio \check{Y} is also asymptotically χ^2 distributed.

III. Empirical Results

A. Signal Time Series

We use daily data from February 1, 1964 to December 31, 2012, totalling 12,321 observations, to construct the time series for all the P/E and logP/E models as well as the BSEYD and logBSEYD models based on 10-year Treasury yields. For the BSEYD and logBSEYD models computed with Aaa corporate yields, we use daily data from January 3, 1983 to December 31, 2012, totalling 7,536 observations.

As a preliminary step to computing summary statistics of the joint behaviour of signals and crashes, we construct the crash indicator sequence C and indicator sequence of distinct signals S . We express both sequences as vectors with either 12,321 entries or 7,536 entries, one for each date. The

entry C_t is “1” if a crash is identified on date t , and 0 otherwise while the entry S_t is “1” if date t corresponds to the start date of a new (distinct) signal, and 0 otherwise.

Now we compute the total number of distinct signals N as

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

Table 3, column 2, shows that over the full period (1964-2012), the total number of signals ranges from 32 for the P/E2 and logP/E2 model to 51 for the P/E3 model. P/E and logP/E models using average earnings (P/E3, P/E4, logP/E3 and logP/E4) generate between 25% and 50% more signals than models based on current earnings (P/E1, P/E2, logP/E1 and logP/E2). On the other hand, BSEYD and logBSEYD models based on current earnings generate slightly more signals than models using average earnings. Over the shorter period 1983-2012, the models BSEYD 5 to 8 respectively generated between 16 and 23 signals while the models logBSEYD 5 to 8 yielded between 21 and 24 signals.

The total number of distinct signals only tells a small part of the story. In particular it does not measure directly the accuracy of the crash prediction models. As a first measure of accuracy, we compute the number of correct predictions n , which is defined as a tally of the crashes identified no more

than $H = 504$ days after a positive signal:

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

For models computed over the full period (1964-2012), Table 3, column 3, shows that the number of correct predictions reaches a high of 30 for the logBSEYD 2 and logBSEYD 3 models and a low of 23 for the P/E 2 model. This statistics allows multiples signals for the same crash. The logBSEYD models and the P/E models based on current earnings had between 21 and 23 correct predictions, while the P/E models based on average earnings generated an average of 32 correct predictions, albeit out of a much larger number of signals. The BSEYD and logBSEYD calculated with Treasury yields are fairly consistent with respective averages of 27.75 and 29.75 successful predictions. Over the shorter period (1983-2012), the BSEYD models based on corporate yields produced between 12 and 13 correct predictions while the logBSEYD models generated between 14 and 17 correct predictions.

B. Crash Prediction Test

Bringing together the total number of distinct signals and the number of correct predictions give us a second measure of accuracy: the proportion of correct predictions, defined as the ratio $\frac{n}{N}$, which also corresponds to the Maximum Likelihood Estimate \hat{p} . This measures enables a direct comparison of the accuracy of the various models. It also gives us the maximum likelihood

estimate of the probability $p = P(C_{t,H} = 1|S_t = 1)$ of a correct crash prediction. Table 3, column 4, shows that the accuracy of all the models based on current earnings and Treasury yields (P/E 1, P/E 2, logP/E 1, logP/E 2, BSEYD 1, BSEYD 2, logBSEYD 1 and logBSEYD 2) is consistently above 70%. A change from current earnings to average earnings over 10 years tends to have an opposite effect on P/E-based and BSEYD-based models: the accuracy of the P/E models (P/E3 and P/E4) declines to 64.71% and 68.89% respectively, and that of the logP/E models (logP/E 3 and logP/E4) falls to around 65% and 67.5% respectively. This effect is reversed for the BSEYD and log BSEYD models. The proportion of correct prediction among BSEYD models rises by 6.50% to 7.50% to around 78% when the calculation is based on average earnings. For log BSEYD models the increase of about 4% to 5% to 74%. The BSEYD and logBSEYD models based on corporate yields display similar behaviour albeit over a shorter time period: the accuracy of the models calculated with 10-year average earnings is above 70% whereas the accuracy of models computed using current earnings ranges from 65% and 68%.

Although the proportion of correct signals computed is generally high, the central question is whether it is significantly different from a random draw with probability $p_0 = \frac{1}{2}$. Because the empirical proportion of correct signals corresponds to the maximum likelihood estimate \hat{p} of the accuracy of the signal, we can evaluate its significance using a likelihood ratio test

applied to the test statistics

$$-2 \ln \Lambda = -2 \ln \frac{L(p_0 = \frac{1}{2}|X)}{L(p = \hat{p}|X)}$$

Table 3 reports the likelihood in column 5, followed by the likelihood ratio Λ in column 6 and the estimated test statistics $-2 \ln \Lambda$ in column 7 and the p -value. The estimated test statistic is asymptotically χ^2 -distributed with 1 degree of freedom. The degree of significance and the p -value indicated in the Table 3 are based on this distribution. The critical values at the 95%, 99% and 99.5% confidence level are respectively 3.84, 6.63 and 7.87.

We find that the P/E's probability of correctly predicting a crash is significantly different from $p_0 = \frac{1}{2}$ at a 95% confidence level. The logP/E models based on current earnings are significantly different from a random draw with at least a 99% confidence whereas the logP/E models calculated with average earnings are significant at only 95%. For the BSEYD measure, three out of four models are statistically significant at a 99% confidence level, and BSEYD 2 is not far from the 99% mark. We find that two BSEYD models are even significant at a 99.5% confidence level: BSEYD 3 and BSEYD 4, which are based on average earnings over ten years. Finally, all of the log-BSEYD models are statistically significantly different from $p_0 = \frac{1}{2}$ at a 99% confidence level, with two of these models (logBSEYD3 and logBSEYD4) reaching 99.5% confidence. Over the shorter time period 1983-2012 during which only 9 of the 18 crashes occurred, we find that out of the eight models

computed using corporate yields, only three (BSEYD 7, logBSEYD 7 and logBSEYD 8) are significant at a 95% confidence level. A possible explanation for the low significance is the shorter data history. The data robustness analysis in section D.1 partly resolves this issue and permits a more objective comparison across all models.

[Place Table 3 here]

C. Monte Carlo Study for Small Sample Bias

In addition to performing the likelihood ratio test using the χ^2 distribution, we carry out a likelihood ratio test based on empirical distributions obtained through Monte Carlo simulations. Table 4 reports the total number of signals, calculated as the sum of all the entries of the indicator sequence S and the maximum likelihood estimate, \hat{p} , which is the probability of correctly predicting a crash and is estimated by maximizing the likelihood function of the model. These statistics are identical to those reported in Table 3.

Columns 4 to 6 in Table 4 present the critical values at a 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulations. These values are not uniform across all models as they depend directly on the total number of signals generated by each model. Column 7 reports the test statistics which is equal to $-2 \ln \Lambda = -2 \ln \frac{L(p=\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated is based on the empirical distribution. The last column reports the p -value, defined as the probability

of obtaining a test statistic higher than the one actually observed assuming that the null hypothesis is true, for each of the models under its empirical distribution.

As was the case with the likelihood ratio test using the χ^2 distribution, we find that the P/E's probability of correctly predicting a crash is significantly different from $p_0 = \frac{1}{2}$ at a 95% confidence level. Three out of four log of P/E models are significantly different from a random draw at a 95% confidence but one (logP/E 3) fails the test. For the BSEYD measure, all four models are statistically significant at a 99% confidence level while the two BSEYD models based on average earnings are even significant at a 99.5% level. Finally, all of the logBSEYD models are statistically significantly different from $p_0 = \frac{1}{2}$ at a 99% confidence level. The logBSEYD 3 is even significant at a 99.5% confidence. For the BSEYD and logBSEYD models calculated using corporate yields, the results of the Monte Carlo simulation confirm our earlier observations: three models (BSEYD 7, logBSEYD 7 and logBSEYD 8) are significant at a 95% confidence level and the fourth one (BSEYD 8) would be significant at a 90% confidence level. Overall, the results obtained under the empirical distribution do not differ markedly from those obtained under the asymptotic χ^2 distribution. We conclude that even with only 32 to 51 signals, the χ^2 distribution provides a very good approximation for the empirical distributions generated using Monte Carlo simulations.

[Place Table 4 here]

D. Robustness

D.1. Data Robustness

We test the robustness of the crash prediction models with respect to a change in the underlying data on two subperiods:

1. *Subperiod 1*: January 31, 1964 to December 31, 1981;
2. *Subperiod 2*: January 1, 1982 to December 31, 2012.

Small sample bias is an important consideration here because of the limited amount of crash-related data available for each subperiod. For Subperiod 1, we have access to all statistics for the P/E and logP/E models as well as the BSEYD and logBSEYD models calculated using Treasury yields. For Subperiod 2, we have statistics for all 24 models, with a caveat: data for the BSEYD and logBSEYD models based on corporate yields start on January 1, 1983.

Table 5 reports likelihood statistics for the two subperiods under the asymptotic χ^2 . The Total Number of Signal in columns 2 and 6 is calculated as the sum of all the entries of the indicator sequence S . P/E and logP/E ratio models produced nearly twice as many signals in Subperiod 2 than in Subperiod 1, whereas BSEYD and models generated roughly the same number of signals in both subperiods. The logBSEYD models based on current earnings experienced a small increase in the number of signals, contrary to the logBSEYD models based on average earnings over ten years, for which

the number of signals remains constant. The Maximum Likelihood estimate \hat{p} reported in columns 3 and 7 is the probability of correctly predicting a crash that maximizes the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. Across all models, we observe a sharp decline in the probability \hat{p} from a range of 83% to 95% in Subperiod 1 to a range of 50% to 70% in Subperiod 2. The most accurate models in Subperiod 2 are based on corporate yields: the accuracy of the BSEYD 5 and 6 and logBSEYD 5 and 6 models averages 67% and that of the BSEYD 7 and 8 and logBSEYD 7 and 8 models averages 72%. The decline in accuracy has an impact on the estimated test statistics, equal to $-2 \ln \Lambda = -2 \ln \frac{L(p=\frac{1}{2}|X)}{L(p=\hat{p}|X)}$ and on the test of significance, which is conducted using the asymptotic χ^2 distribution with 1 degree of freedom. Whereas the probability of a correct prediction is statistically significant from 50% at a 95% or 99% confidence level for all models in Subperiod 1, it is only statistically significant at a 95% level for three models (BSEYD 7, logBSEYD 7 and logBSEYD 8) and at a 90% level for a further three models (logP/E 1, BSEYD 8 and logBSEYD 6) in Subperiod 2.

[Place Table 5 here]

Table 6 reports empirical statistics for the two subperiods using an empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in columns 5 and 10 are equal to $-2 \ln \Lambda = -2 \ln \frac{L(p=\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated is based on the empirical distribution.

Conducting the statistical test of significance using the empirical distribution further increase the difference between Subperiod 1 and Subperiod 2. In Subperiod 1, 14 out of 16 prediction probabilities are statistically significant from 50% at a 99% confidence level. The other two (logP/E 1 and logP/E 2) are statistically significant at a 95% confidence level. In Subperiod 2, the prediction probability is statistically significant at a 95% level for three models (BSEYD 7, logBSEYD 7 and logBSEYD 8) and at a 90% level for a further three models (logP/E 1, BSEYD 8 and logBSEYD 6).

[Place Table 6 here]

D.2. Model Parameter Robustness

The last set of results presented in this section address the robustness of the four crash measures (P/E, logP/E, BSEYD and logBSEYD) with respect to a change in parameters. We temporarily leave aside the BSEYD and logBSEYD models based on corporate yields and the question of robustness with respect to a choice of interest rate. The main reason is that the choice of interest rates affects the time period for which data are available (1962 to 2012 for Treasury Notes but only 1983 to 2012 for Aaa corporate bonds), but also the number of crashes over which we can conduct the test. The best approach to this problem is to analyze the robustness of the models to a change in the underlying dataset, as we did previously.

Column 2 in Table 7 reports the robust likelihood ratio $\check{\Lambda}$ for each of the

measures under the null hypothesis $H_0 : p_0 = \frac{1}{2}$. The robust likelihood ratio is calculated using (16) as the highest likelihood ratio across all implementations of the measure. Higher robust likelihood ratios are indicative of a lower statistical significance. Column 3 identifies the specific model, and therefore the set of parameters, corresponding to the likelihood ratio reported in column 2. Columns 4 and 5 refer to tests of significance performed using the asymptotic χ^2 distribution. Column 4 reports the robust test statistics, computed using formula (17), and level of significance for the measure based on the χ^2 distribution. Column 5 gives the associated p -value. Columns 6 and 7 refer to tests of significance performed with the empirical distribution generated by Monte Carlo simulations. The objective here is to correct for small sample bias. Column 6 reports the robust test statistics, computed using formula (17), and level of significance for the measure based on the empirical distribution. Column 7 gives the associated p -value.

The most robust measures are the logBSEYD and the BSEYD measure. The robust test statistics of the logBSEYD measure is significant at 99% based on the asymptotic χ^2 distribution and 95% using the empirical distribution generated by Monte Carlo simulations. The corresponding significance level for the BSEYD measure are respectively 95% and 99%. The P/E ratio also exhibit a high level of robustness: the robust statistic is significant at 95% for both the asymptotic χ^2 and the empirical distribution whereas the logP/E is only significant at 95% under the asymptotic distribution.

[Place Table 7 here]

IV. Discussion

This paper is concerned with the predictability of equity market crashes over a long 60 year horizon and the ability of four traditional measures, the P/E ratio, logP/E, BSEYD and logBSEYD, to predict them. In particular, it seeks to answer the following questions: do these four measures really have a statistically significant ability to predict equity market crashes? Are they robust when the model specifications change? Do equity market crashes exhibit some predictability?

A. Statistical Test of Predictive Power

Our findings show that three measures (P/E, BSEYD and logBSEYD) demonstrate statistically significant crash prediction ability under both asymptotic χ^2 distribution and simulated empirical distribution, regardless of the definition of earnings and of the model used for the threshold $K(t)$. The BSEYD and logBSEYD models based on average earnings are significant at or near the 99.5% level. The P/E measure is statistically significant at a 95% level across all model specifications. On the other hand, the logP/E measure does not appear to have crash prediction ability: it fails to achieve statistical significance at any meaningful level.

These findings support the existing literature and in particular Ziemba and Schwartz (1991) and in Lleo and Ziemba (2012): the BSEYD measure shows an ability to predict crashes. In this paper, we emphasise a much

longer term perspective than previous studies, with a statistical analysis on more than 60 years of daily data. We also performed a comparative analysis of the predictive ability of BSEYD-based and of P/E-based models, which highlights the key role played by interest rates in crash prediction.

We also find that the distribution of test statistics is well behaved. Even with only 32 to 51 signals, the χ^2 distribution provides a very good approximation for the empirical distributions generated using Monte Carlo simulations. This observation is confirmed by the robustness study: the level of significance and p values of the robust test statistics under both distributions are consistent with each other.

The consistency of these findings goes a long way to explaining why the BSEYD, the log BSEYD and to a lesser extent the P/E ratio exhibit a high degree of robustness with respect to model specification under both the asymptotic χ^2 distribution and the simulated empirical distribution. The robust likelihood ratio and test statistics used to measure robustness in this study are in essence the “worst” likelihood ratio and test statistics across all implementations of a single measure. Because all implementations have similar levels of significance, the robust statistics confirms the ranking already established after we examined and tested all of the models.

This observation suggests that the predictive ability of a crash prediction model has more to do with the economic properties of the underlying measure than with the specific parameters corresponding to the model implementation. A possible rationale for the relative success of the BSEYD and log

BSEYD measures is that the earnings yield may be viewed as a rough proxy for the return of equity shares and markets. As a result, these measures provide a coarse approximation for the equity risk premium. In the context of crash prediction, accuracy is less important than evolution: large values for either measure indicate an abnormally low equity risk premium and suggest that a correction back to levels consistent with an historical average should occur. The exact timing of the correction is highly uncertain though.

This explanation is also consistent with the observation that the BSEYD and logBSEYD models based on average earnings are more statistically significant than models based on current earnings. Current earnings point to current conditions whereas average earnings “look through” a larger section of the business cycle and therefore point to deviations from what would appear as “average” conditions through the cycle. The interpretation of the P/E ratio as a relative measure of equity prices explains its predictive power with respect to equity crashes. Periods with abnormally high P/E ratios and prices tend to announce a correction in equity market prices. The difference in significance (and robustness) between the P/E ratio and the BSEYD or logBSEYD measures seems to imply that current levels of interest rates and their relation with equity prices add relevant information.

The lack of predictive power of the log P/E may appear surprising at first, especially in light of the results obtained by Campbell and Shiller (1988, 1989, 1998) and by Shiller (1996, 2006). However, $\log \frac{P}{E} = -\log \frac{E}{P}$ implying that the log P/E ratio equals minus the log earnings yield, a proxy for the return

on equity. Changing the sign of a variable in a regression does not alter the strength of the linear relation between the dependent and the independent variable: it only changes the sign of the slope coefficient. This is what we observe in the works by Campbell and Shiller (1988, 1989, 1998) and by Shiller (1996, 2006). However, crash prediction systems are binary by nature (a crash either occurs or it does not, a signal either occurs or it does not) which means that they are nonlinear. Changing the sign of a variable has a significant impact on the outcome of the model. In the present case, changing the sign destroys the predictive ability of the measure.

B. Robustness

B.1. Addressing Data Robustness: Subperiod Study

All the measures demonstrated a statistically significant ability to predict market crashes over the period 1964-1981. However, the probability of correct prediction and the degree of significance of the models dropped over the period 1981-2012. What could be the cause?

The sharp contrast between the two subperiods that emerges from Tables 5 and 6 blurs out some crucial details. The track record of all measures during the first subperiod is exceptionally good: the lowest accuracy across all 16 models is 83% while the highest is 95%. As a result, even with a limited number of signals (between 9 and 20) and crashes (9 in total), the statistical test will reject the null hypothesis even at a very high confidence

level. When we examine the track record of all 24 models over the second subperiod independently from the first, we see that the accuracy ranges from 52% to 75% with a mean of 63% and a median of 62%. A 60% accuracy hints that these measures may have a predictive value, but it also implies that we need many more crashes to be 95% or 99% confident that this number is statistically different from a 50% chance. In summary, extremes are easier to identify statistically than moderately high (or low) values. This is part of the story explaining the apparent sharp contrast.

The next part of the story lies in the circumstances that led to a reduction in the accuracy of the models. Further research is obviously needed here, but we will venture to give some likely explanations. First, the macro and micro-structure of financial markets is markedly different in 2012 than it was in 1964. As a result, some models that used to work in the 1960s may no longer work in our current environment. We do not believe this argument holds here. The measures we are concerned with were identified in the 1980s (in Subperiod 2) and there is anecdotal evidence, including Shiller (2006), Maio (2013) and Lleo and Ziemba (2012), that they were still producing satisfactory predictions recently. Also, a change in the structure of financial markets would imply a slow decline in the accuracy of the models, rather than an abrupt, highly noticeable drop.

Another possible explanation is that the performance of these measures is affected by the prevailing regime in which financial markets are operating. Regime switching models have attracted significant attention in the last

decade thanks to the pioneering work of Ang and Bekaert (2002). In this scenario, crash prediction models would tend to perform less well in an upward market environment when markets trend up with little volatility, and better in a turbulent state characterized by high volatility. This argument would be consistent with large shifts in the accuracy of the models and with the observation that accuracy deteriorates during the 1982-2012 subperiod which was dominated by a bull market that lasted 18 years. A relation between crash prediction models and market regimes is certainly intriguing and further research is needed on this question. The methodological challenge will be to establish a correspondence between the predictions, which are forward-looking by nature, and the state of the market both at the time the prediction is made and at the time it materializes.

B.2. Measuring Parameter Robustness: Robust Likelihood Statistics

Though crucial, the question of robustness has mostly remained peripheral to the debate on the predictability of market downturns and crashes. This paper addresses the question of robustness directly by distinguishing data robustness from model specification robustness. To quantify the latter, we propose a novel measure: a robust version of the likelihood ratio and of the associated test statistics. Measuring the robustness of a crash prediction model using a robust likelihood ratio has significant advantages. First, the likelihood ratio provides a direct, objective measure of robustness that allows

direct comparison between different model specifications and even different types of models. Second, we can use the robust test statistics to evaluate the significance of the model using a χ^2 distribution. The immediate benefit is that we can equate loss of robustness over a set of model parameters with loss of significance of the prediction. Third, the idea of robust likelihood ratio can be easily extended to account for small sample bias. We can use the Monte Carlo algorithm presented in an earlier section to compute the (standard) likelihood ratio under each specification and then select the lowest one as a robust likelihood ratio.

However, the robust likelihood ratio and test statistics also have an important limitation inherent in all summary measures: because summary measures condense all available information into a single number, they are unable to give a nuanced view of a phenomenon. Having established an objective view of robustness with the robust likelihood ratio, we now discuss some of its subjective facets.

Intuitively, a crash prediction model is robust if its accuracy, reflected in the total number of signals, the number of correct predictions and the proportion of correct predictions, does not change materially with a change in the model specification. The total number of signals, number of correct predictions and proportion of correct predictions reported in columns 1, 2 and 3 of Table 3 suggest that all of the models are impacted by a change in definition of earnings. For the P/E ratio, the total number of signals jumps from 32 to 34 for the models based on current earnings to between 45 and 51

for models using average earnings. The number of correct predictions also soars from about 24 to an average of 32. The result is a 3%-6% decline in the proportion of correct predictions. The total number of signals generated by logP/E models increases from around 32 to an average of 43 for models based on average earnings, whereas the number of correct predictions changes very little. This translates in a full 10% drop in the proportion of correct predictions. This is in sharp contrast with the BSEYD and log BSEYD measures. Going from current earnings to average earnings causes a decrease in the total number of signals while the number of correct predictions either increases slightly or remains constant. The end result is an increase in the proportion of correct predictions by around 7% for the BSEYD measure to an average of 22 and by around 5% for the logBSEYD to an average of 28.5.

The test statistics reported in Tables 3 and 4 confirms this observation for the logP/E, BSEYD and logBSEYD measures. The value of the test statistics for the BSEYD and logBSEYD tends to be markedly higher when the measure is based on average earnings, whereas the converse occurs for the logP/E measure. The test statistics for the P/E do not appear to reflect the change in earnings.

Next we examine the joint effect of a change in the definitions of earnings and threshold. We observe that models based on current earnings seem robust to a change in the definition of the threshold: the total number of signals and the number of correct predictions reported in Table 3 do not differ by more than 2. As a result, the difference in accuracy for measures

based on current earnings never exceeds 1.9%. However, the P/E models and log of P/E models based on average earnings over a ten-year period (P/E 3, P/E 4, logP/E 3 and logP/E4) are more sensitive to the choice of confidence level. The total number of signals in models based on a standard confidence interval (P/E3 and logP/E3) is significantly higher than for models based on Cantelli's inequality (P/E 4 and logP/E 4). The number of correct predictions is also higher for models based on a standard confidence level. As a result, the P/E model's accuracy is higher by 4% when Cantelli's inequality is used. Similarly, the accuracy of the logP/E model is higher by 2.50% when the confidence level is based on Cantelli's inequality. The test statistics reported in Tables 3 and 4 confirm this observation for the P/E. The value of the test statistics for the logP/E models based on average earnings is too low to permit a clear conclusion.

Overall, we find that the models based on current earnings appear robust with respect to the definition of confidence level whereas the P/E models and log of P/E models based on average earnings over a ten-year period (P/E 3, P/E 4, logP/E 3 and logP/E 4) are more sensitive to the choice of confidence level.

C. Predictability of Market Crashes

The fundamental issue addressed in this paper relates to the accuracy of the crash prediction models. We can formulate this from two different but complementary perspectives by asking the following questions:

1. Over a given horizon, what proportion of positive signals predicted actual crashes?
2. What proportion of crashes were preceded by a signal within a given horizon?

Our results and previous discussion show that the proportion of positive signals that predicted actual crashes is statistically strongly significant for both the BSEYD and logBSEYD measures. It is also statistically significant for P/E but not for logP/E. The implication is that the BSEYD and logBSEYD, and to a lesser extent the P/E ratio, have an ability to predict crashes. However, the answer to both questions is not necessarily the same because several positive signals may in fact relate to the same equity market crash.

We now turn to the second question: what proportion of crashes was preceded by a crash? Table 8 provides evidence that the track record of the P/E, BSEYD and logBSEYD measures is excellent: between 15 and 18 of the 18 crashes (83.3% and 100%) were preceded by a positive signal. The log of P/E based on current earnings is the least accurate: only 13 to 14 crashes (72.22% to 77.78%) were preceded by a signal. The logBSEYD measure is the most accurate: all four implementations predict at least 17 out of 18 crashes and two implementations (logBSEYD 2 and logBSEYD 4) predict all 18 crashes. The BSEYD measure is also consistent while the P/E measures are more contrasted. Models based on average earnings over

a ten-year period perform best, reaching 100% accuracy. Models based on current earnings do not perform as well: their accuracy ranges from 83% to 90%.

[Place Table 8 here]

In theory, we could (and should) test the significance of these numbers statistically using the methodology and likelihood ratio test presented in Section D. However, we encounter a difficulty in the formulation of a meaningful null hypothesis for the test. The hypothesis $H_0 : p = p_0 := \frac{1}{2}$, proposed above to reflect the idea that an arbitrary (uninformed) signal should correctly predict crashes only 50% of the time, does not appear appropriate to test the proportion of crashes preceded by a signal. In fact, we do not have a benchmark for the proportion of crashes preceded by an (uninformed) signal. A test of the null hypothesis $H_0 : p = p_0 := \frac{1}{2}$ with a χ^2 distribution with one degree of freedom will lead to an automatic rejection for all models: the proportion of crashes preceded by a signal is statistically significantly above 50%. This is not satisfactory: more research is required to provide a meaningful statistical test addressing the question of the proportion of crashes preceded by a signal.

D. The Inflation Puzzle

Finally, our findings have also revealed an important puzzle: the BSEYD and logBSEYD measures exhibit superior crash prediction power in spite of

ignoring the effect of inflation and of comparing a real quantity, the price earnings ratio, with a nominal quantity, the bond yield².

Our findings appear to contradict the importance generally granted to the effect of inflation. One possible explanation for this puzzle could lie in the “money illusion,” a bias discussed in Behavioural Finance and first described by Irving Fisher (see Thaler, 1997). The “money illusion” bias suggests that portfolio managers either ignore or underweight the impact of inflation over the horizon of the 10-year Treasury yield and concentrate either on the two-year prediction horizon or on the instantaneous decision of shifting from stocks to bonds. Another possible explanation is that the magnitude of the impact of an equity market correction on an investor’s wealth and utility dwarves in many instances the impact that inflation would have over a two-year horizon. Inflation might not be ignored by portfolio managers, but it might simply be regarded as a second order term. In any case, further research is required on this question.

V. Conclusion

This paper is concerned with the predictability of equity market crashes and the ability of four measures traditionally used to forecast equity market downturns: the Price-Earnings (P/E) ratio, the logarithmic Price-Earnings

²Campbell and Vuolteenaho (2004), Asness (2000, 2003) and Ritter and Warr (2002) discuss these issues.

ratio, the BSEYD model and the logarithmic BSEYD Model. Using 49 years of daily data, we tested statistically the ability of these measures to predict equity market crashes under several competing model specifications. The tests were conducted with respect to events, namely equity market crashes, as opposed to returns. This orientation impacts the structure and nature of the statistical tests and the way we define robustness. We constructed a time series of equity market crashes. We also proposed a novel measure of robustness built on a robust likelihood ratio and its associated test statistics.

The statistical tests and robustness analysis suggest that the BSEYD, the logarithmic BSEYD model, and to a lesser extent the P/E ratio, are statistically significant robust predictors of equity market crashes. The logP/E is not statistically significant at any meaningful level. Overall, the BSEYD models show more consistency than the models based on P/E only. For current earnings, the BSEYD models have a similar proportion of correct signals but predict a higher proportion of crashes than P/E models. In addition, using ten-year average earnings markedly improves both the proportion of correct signals and the proportion of crashes detected by BSEYD models. This empirical evidence suggests that although the P/E ratio is by itself a good predictor of equity market crashes, taking instead the difference between the earnings yield and the prevailing 10 year Treasury yield will provide superior predictions.

A. Relation Between the the Gordon Growth Model, the Campbell-Shiller Model and the BSEYD

The Gordon (1959) growth model states that the price of a stock should be equal to the present value of its future dividends growing at a constant rate g , discounted at the cost of equity k :

$$(18) \quad P_t = \frac{D_{t+1}}{k_t - g},$$

where P_t is the price at the end of period t , D_{t+1} is the dividend paid at time $t + 1$. We can break down the dividend at time $t + 1$ into the products of a measure of core (a-cyclical) earnings at time t , the payout ratio d and the growth rate g :

$$(19) \quad P_t = \frac{E_t d (1 + g)}{k - g}.$$

With a bit of algebra

$$k = g + \frac{(1 + g)d}{\gamma_t},$$

where we defined $\gamma(t) = \frac{P_t}{E_t}$. These are nominal rates: we could for example deflate by the PPI to produce real rates.

From equation (19), we obtain the following definition for the earnings yield ρ_t :

$$\rho_t = \frac{E_t}{P_t} = \frac{k - g}{d(1 + g)}.$$

The cost of equity k , is the required return on the equity market, that is

$$k_t = r_t + f_t,$$

where r_t is the yield on a government bond and f is the equity risk premium.

Hence

$$\begin{aligned} BSEYD &= r_t - \rho_t \\ &= \frac{1}{d(1 + g)} (r_t [d(1 + g) - 1] - f_t + g). \end{aligned}$$

References

- Ang, A., and G. Bekaert. 2002. International asset allocation with time-varying correlations. *Review of Financial Studies* 55:1137–1187.
- Asness, C. 2000. Stocks versus bonds: explaining the equity risk premium. *Financial Analysts Journal* pp. 96–113.
- Asness, C. 2003. Fight the Fed model: The relationship between future returns and stock and bond market yields. *Journal of Portfolio Management* .
- Balvers, R., Y. Wu, and E. Gilliland. 2000. Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies. *The Journal of Finance* 55:745–772.
- Berge, K., G. Consigli, and W. T. Ziemba. 2008. The predictive ability of the bond stock earnings yield differential. *Journal of Portfolio Management* pp. 63–80.
- Bertocchi, M., S. Schwartz., and W. Ziemba. 2010. *Optimizing the Aging, Retirement, and Pensions Dilemma*. John Wiley & Sons.
- Campbell, J. Y., and R. J. Shiller. 1988. Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance* 43:661–676. Papers and Proceedings of the Forty-Seventh Annual Meeting of the American Finance Association, Chicago, Illinois, December 28-30, 1987.

- Campbell, J. Y., and R. J. Shiller. 1989. The Dividend Ratio Model and Small Sample Bias: A Monte Carlo Study. *Economic Letters* 29:324–331.
- Campbell, J. Y., and R. J. Shiller. 1998. Valuation Ratios and the Long-Run Stock Market Outlook. *The Journal of Portfolio Management* pp. 11–26.
- Campbell, J. Y., and T. Vuolteenaho. 2004. Inflation illusion and stock prices. *American Economic Review, Papers and Proceedings* 94:19–23.
- Consigli, G., L. C. MacLean, Y. Zhao, and W. T. Ziemba. 2009. The bond-stock yield differential as risk indicator in financial markets. *Journal of Risk* 11:1–22.
- Estrada, J. 2006. The Fed Model: A Note. *Financial Research Letters* 3:14–22.
- Fama, E. 1965. The Behavior of Stock Market Prices. *Journal of Business* 38:34–106.
- Fama, E. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25:383–417.
- Fama, E. F., and K. R. French. 1988a. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22:3–25.

- Fama, E. F., and K. R. French. 1988b. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy* 96:246–73.
- Fama, E. F., and K. R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47:427–465.
- Faugère, C. 2013. The Fear Premium and Daily Comovements of the S&P 500 E/P ratio and Treasury Yields before and during the 2008 Financial Crisis. *Financial Markets, Institutions & Instruments* 22:171–207.
- Faugère, C., and J. Van Erlach. 2009. A Required Yield Theory of Stock Market Valuation and Treasury Yield Determination. *Financial Markets, Institutions & Instruments* 18:27–88.
- Flood, R. P., R. J. Hodrick, and P. Kaplan. 1986. An Evaluation of Recent Evidence on Stock Market Bubbles. Tech. rep., NBER Working Paper No. w1971.
- Gordon, M. J. 1959. Dividends, Earnings and Stock Prices. *Review of Economics and Statistics* 41:99–105.
- Graham, B., and D. L. Dodd. 1934. *Security Analysis*. 1st ed. Whittlesey House.
- Grimmett, G., and D. Stirzaker. 2001. *Probability and Random Processes*. Oxford University Press.

- Hertzberg, M. P., and B. A. Beckman. 1989. Business Cycle Indicators: Revised Composite Indexes. *Business Conditions Digest* 17.
- Jarrow, R. A. 2012. Detecting Asset Price Bubbles. *Journal of Derivatives* 20:30–34.
- Jarrow, R. A., Y. Kchia, and P. Protter. 2011a. A Real Time Bubble Detection Methodology. *Bloomberg Risk Newsletter* .
- Jarrow, R. A., Y. Kchia, and P. E. Protter. 2011b. How to Detect an Asset Bubble. *SIAM Journal on Financial Mathematics* 2:839–865.
- Jarrow, R. A., Y. Kchia, and P. E. Protter. 2011c. Is There a Bubble in LinkedIn’s Stock Price? *Journal of Portfolio Management* 38:125–130.
- Klein, P. A., and G. H. Moore. 1982. The Leading Indicator Approach to Economic Forecasting—Retrospect and Prospect. Tech. rep., NBER.
- Koivu, M., T. Pennanen, and W. T. Ziemba. 2005. Cointegration analysis of the FED model. *Finance Research Letters* 2:248–256.
- Lleo, S., and W. T. Ziemba. 2012. Stock market crashes in 2007–2009: were we able to predict them? *Quantitative Finance* 12:1161–1187.
- Maio, P. 2013. The “Fed Model” and the Predictability of Stock Returns. *Review of Finance* 17:1489–1533.
- Markowitz, H. 1952. Portfolio Selection. *Journal of Finance* 7:77–91.

- Ozyildirim, A., R. H. McGuckin, and V. Zarnowitz. 2003. A More Timely and Useful Index of Leading Indicators. *The Conference Board* .
- Poterba, J. M., and L. H. Summers. 1989. Mean Reversion in Stock Prices: Evidence and Implications. *Journal of Finance* 22:27–59.
- Ritter, J. R., and R. S. Warr. 2002. The decline of inflation and the bull market of 1982-1999. *Journal of Financial and Quantitative Analysis* 37:29–61.
- Shiller, R. J. 1996. Price Earnings Ratios as Forecasters of Returns: The Stock Market Outlook in 1996. Tech. rep., <http://www.econ.yale.edu/~shiller/data/peratio.html>. URL <http://www.econ.yale.edu/~shiller/data/peratio.html>.
- Shiller, R. J. 2005. *Irrational Exuberance*. 2nd ed. Princeton University Press.
- Shiller, R. J. 2006. Irrational Exuberance Revisited. *CFA Institute Conference Proceedings Quarterly* 23:16–25.
- Shiryaev, A. N., and M. V. Zhitlukhin. 2012a. Bayesian disorder detection problems on filtered probability spaces. *Theory of Probability and Its Applications* 57.
- Shiryaev, A. N., and M. V. Zhitlukhin. 2012b. Optimal stopping problems for a Brownian motion with a disorder on a finite interval. Tech. rep., arXiv:1212.3709.

- Shiryaev, A. N., M. Zitlukhin, and W. T. Ziemba. 2014. When to Sell Apple and the Nasdaq 100? Trading Bubbles with a Stochastic Disorder Model. *Journal of Portfolio Management* 40:1–10.
- Shiryaev, A. N., M. Zitlukhin, and W. T. Ziemba. 2015. Land and Stock Bubbles, Crashes and Exit Strategies in Japan, Circa 1990 and in 2013. *Quantitative Finance (forthcoming)* .
- Thaler, R. H. 1997. Irving Fisher: Modern Behavioral Economist. *The American Economic Review* 87:439–441. Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association.
- Weigand, R. A. ., and R. Irons. 2007. The Market P/E Ratio, Earnings Trends, and Stock Return Forecasts. *Journal of Portfolio Management* 33:87–101.
- Yardeni, E. 1997. Tech. rep., US Equity Research, Deutsche Morgan Grenfell.
- Ziemba, W., and S. Schwartz. 1991. *Invest Japan*. Probus, Chicago.

Beginning Year	Ending Year	Annual Rate of Return	Beginning P/E	Ending P/E
1975	1994	9.6%	10.9	20.5
1977	1996	9.7%	11.5	25.9
1942	1961	9.9%	12.2	20.5
1983	2002	10.9%	7.3	25.9
1978	1997	11.9%	10.4	31.0
1981	2000	12.8%	8.8	41.7
1979	1998	12.9%	9.4	36.0
1982	2001	13.0%	8.5	32.1
1980	1999	14.0%	8.9	42.1

Table 1: Evolution of the P/E Over Selected 20-Year Periods With High Annualized Returns. Source: Bertocchi et al. (2010)

Crash Identification Date	Peak Date	S&P Index at Peak	Trough date	S&P Level at trough	Peak-to-trough decline (%)	Peak-to-trough duration (in days)
1	1966-07-26	1966-02-09	1966-10-07	73.20	22.2%	240
2	1969-06-20	1968-11-29	1970-05-26	69.29	36.1%	543
3	1971-08-05	1971-04-28	1971-11-23	90.16	13.9%	209
4	1973-04-30	1973-01-11	1974-04-25	89.57	25.5%	469
5	1975-08-20	1975-07-15	1975-09-16	82.09	14.1%	63
6	1977-05-26	1976-09-21	1978-03-06	86.90	19.4%	531
7	1978-10-27	1978-09-12	1978-11-14	92.49	13.6%	63
8	1980-03-14	1980-02-13	1980-03-27	98.22	17.1%	43
9	1981-08-25	1980-11-28	1982-08-12	102.42	27.1%	622
10	1984-02-22	1983-10-10	1984-07-24	147.82	14.4%	288
11	1987-10-16	1987-08-25	1987-12-04	223.92	33.5%	101
12	1990-08-20	1990-07-16	1990-10-11	295.46	19.9%	87
13	1998-08-28	1998-07-17	1998-08-31	957.28	19.3%	45
14	1999-10-18	1999-07-16	1999-10-18	1254.13	11.6%	94
15	2000-10-12	2000-03-24	2001-09-21	965.80	36.8%	546
16	2008-01-16	2007-10-09	2008-11-20	752.44	51.9%	408
17	2010-05-21	2010-04-23	2010-07-02	1022.58	16.0%	70
18	2011-08-05	2011-04-29	2011-10-03	1099.23	19.4%	157

Table 2: The S&P 500 Index experienced eighteen crashes between January 1, 1962 and December 31, 2012

We define an Equity Market Crash as a decline of at least 10% in the level of the S&P500 over a time period of less than a year (252 trading days). The Crash Identification Date is the date at which the decline from the previous local peak reached 10%. The Peak Date is the date at which the local peak that preceded the crash was reached. The S&P Level at Peak is the closing level of the S&P 500 on the peak date. The Trough Date is the date at which the local trough that followed the crash was reached. The S&P Level at Trough is the closing level of the S&P 500 on the trough date. The Peak-to-Trough Decline is the percentage loss on the S&P 500 from the local peak to the local trough. The Peak-to-Trough duration measures the duration of the decline as the number of calendar days between the local peak and the local trough.

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
P/E 1	34	24	70.59%	1.1345E-09	0.0513	5.9398*	1.48%
P/E 2	32	23	71.88%	5.5528E-09	0.0421	6.3435*	1.18%
P/E 3	51	33	64.71%	4.1666E-15	0.1066	4.4777*	3.43%
P/E 4	45	31	68.89%	7.6461E-13	0.0372	6.5844*	1.03%
logP/E 1	33	20	72.73%	4.0021E-09	0.0291	7.0748**	0.78%
logP/E 2	32	20	75.00%	1.5311E-08	0.0152	8.3720***	0.38%
logP/E 3	46	23	65.22%	1.2379E-13	0.1148	4.3292*	3.75%
logP/E 4	40	21	67.50%	1.1109E-11	0.0819	5.0053*	2.53%
BSEYD 1	39	28	71.79%	8.3989E-11	0.0217	7.6648**	0.56%
BSEYD 2	37	26	70.27%	1.6640E-10	0.0437	6.2597*	1.24%
BSEYD 3	37	29	78.38%	4.0813E-09	0.0018	12.6592***	0.04%
BSEYD 4	36	28	77.78%	5.2270E-09	0.0028	11.7678***	0.06%
BSEYD 5	23	15	65.22%	3.5184E-07	0.3388	2.1646	14.12%
BSEYD 6	21	14	66.67%	1.5663E-06	0.3044	2.3786	12.30%
BSEYD 7	16	12	75.00%	1.2374E-04	0.1233	4.1860*	4.08%
BSEYD 8	17	12	70.59%	3.3682E-05	0.2265	2.9699	8.48%
logBSEYD 1	44	31	70.45%	2.5209E-12	0.0225	7.5842**	0.59%
logBSEYD 2	43	30	69.77%	3.9344E-11	0.0316	6.9080**	0.86%
logBSEYD 3	40	30	75.00%	1.7031E-10	0.0053	10.4650***	0.12%
logBSEYD 4	38	28	73.68%	3.0807E-10	0.0118	8.8778***	0.29%
logBSEYD 5	21	14	66.67%	1.5663E-06	0.3044	2.3786	12.30%
logBSEYD 6	22	15	68.18%	1.0562E-06	0.2257	2.9769	8.45%
logBSEYD 7	24	17	70.83%	5.1080E-07	0.1167	4.2965*	3.82%
logBSEYD 8	24	17	70.83%	5.1080E-07	0.1167	4.2965*	3.82%

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

Table 3: Maximum Likelihood Estimate and Likelihood Ratio Test

The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S . The Number of Correct Predictions is the tally of crashes preceded by the signal. It is calculated as the sum of all the entries of the indicator sequence X . The Maximum Likelihood estimate \hat{p} is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. $L(\hat{p})$ is the likelihood of the crash prediction model, computed using the maximum likelihood estimate \hat{p} . The likelihood ratio $\Lambda = \frac{L(p=\frac{1}{2}|X)}{L(p=\hat{p}|X)}$ is the ratio of the likelihood under the null hypothesis $p_0 = \frac{1}{2}$ to the likelihood using the estimated probability \hat{p} . The estimated test statistics, equal to $-2 \ln \Lambda$, is asymptotically χ^2 -distributed with 1 degree of freedom. The p -value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p -value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99% level are respectively 3.8415, 6.6349 and 7.8794.

Signal Model	Total number of signals	ML estimate \hat{p}	Critical confidence		Critical confidence	Test statistics		Empirical p -value
			95%	99%		$-2 \ln \Lambda$	p -value	
P/E 1	34	70.59%	4.3279	5.9398	7.8353	5.9398*	2.27%	
P/E 2	32	71.88%	4.6119	6.3371	8.3720	6.3435*	2.28%	
P/E 3	51	64.71%	4.4777	7.2520	7.2520	4.4777*	5.00%	
P/E 4	45	68.89%	3.8096	6.5844	6.5844*	6.5844*	1.33%	
logP/E 1	33	72.73%	3.7378	7.0748	7.0748	7.0748*	1.24%	
logP/E 2	32	75.00%	3.1780	6.3371	8.3720	8.3720**	0.68%	
logP/E 3	46	65.22%	4.3292	7.2352	8.9927	4.3292	5.48%	
logP/E 4	40	67.50%	3.6560	6.5826	8.3983	5.0053*	3.91%	
BSEYD 1	39	71.79%	4.4174	5.9206	7.6648	7.6648**	0.79%	
BSEYD 2	37	70.27%	4.4174	5.9206	7.6648	6.2597**	0.99%	
BSEYD 3	37	78.38%	3.3202	6.2597	8.1119	12.6592***	0.05%	
BSEYD 4	36	77.78%	4.0776	7.3660	7.3660	11.7678***	0.11%	
BSEYD 5	23	65.22%	3.6176	7.7998	7.7998	2.1646	14.12%	
BSEYD 6	21	66.67%	3.9849	6.0595	8.6618	2.3786	12.30%	
BSEYD 7	16	75.00%	4.1860	6.7382	6.7382	4.1860*	4.08%	
BSEYD 8	17	70.59%	5.0168	7.7230	7.7230	2.9699	8.48%	
logBSEYD 1	44	70.45%	3.3146	5.9537	7.5842	7.5842**	0.88%	
logBSEYD 2	43	69.77%	3.9924	6.9080	6.9080	6.9080*	1.19%	
logBSEYD 3	40	75.00%	3.6560	6.5826	8.3983	10.4650***	0.18%	
logBSEYD 4	38	73.68%	3.8551	6.9515	8.8778	8.8778**	0.62%	
logBSEYD 5	21	66.67%	3.9849	6.0595	8.6618	2.3786	12.30%	
logBSEYD 6	22	68.18%	4.7166	6.9162	9.6363	2.9769	8.45%	
logBSEYD 7	24	70.83%	4.2965	6.2790	8.7075	4.2965*	3.82%	
logBSEYD 8	24	70.83%	4.2965	6.2790	8.7075	4.2965*	3.82%	

* significant at the 5% level.

** significant at the 1% level.

*** significant at the 0.5% level.

Table 4: Monte Carlo Likelihood Ratio Test

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

Signal Model	Sample 1				Sample 2			
	Total number of signals	ML Estimate \hat{p}	Test statistics $-2 \ln \Lambda$	p -value	Total number of signals	ML Estimate \hat{p}	Test statistics $-2 \ln \Lambda$	p -value
P/E 1	12	83.33%	5.8221*	1.58%	22	63.64%	1.6573	19.80%
P/E 2	9	88.89%	6.1977*	1.28%	23	65.22%	2.1646	14.12%
P/E 3	16	93.75%	14.6994**	0.01%	35	51.43%	0.0286	86.58%
P/E 4	15	93.33%	14.6994**	0.02%	30	56.67%	0.5349	46.45%
logP/E 1	10	90.00%	7.3613**	0.67%	23	65.22%	2.1646	14.12%
logP/E 2	9	88.89%	6.1977*	1.28%	23	69.57%	3.6176†	5.72%
logP/E 3	15	93.33%	13.4465**	0.02%	41	51.61%	0.0323	85.75%
logP/E 4	12	91.67%	9.7515**	0.18%	28	57.14%	0.5734	44.89%
BSEYD 1	19	89.47%	13.5528**	0.02%	20	55.00%	0.2003	65.45%
BSEYD 2	18	83.33%	8.7331**	0.031%	19	57.89%	0.4757	49.04%
BSEYD 3	19	94.74%	18.5043**	0.00%	18	61.11%	0.8964	34.38%
BSEYD 4	18	94.74%	17.2292**	0.00%	18	61.11%	0.8964	34.38%
BSEYD 5					23	65.22%	2.1646	14.12%
BSEYD 6					21	66.67%	2.3786	12.30%
BSEYD 7					16	75.00%	4.1860*	4.08%
BSEYD 8					17	70.59%	2.9699†	8.48%
logBSEYD 1	19	84.21%	9.7654**	0.18%	25	60.00%	1.0068	31.57%
logBSEYD 2	20	85.00%	10.8175**	0.10%	23	56.52%	0.3924	53.10%
logBSEYD 3	20	90.00%	14.7226**	0.01%	20	60.00%	0.8054	36.95%
logBSEYD 4	19	89.47%	13.5528**	0.02%	19	57.89%	0.4757	49.04%
logBSEYD 5					21	66.67%	2.3786	12.30%
logBSEYD 6					22	68.18%	2.9769†	8.45%
logBSEYD 7					24	70.83%	4.2965*	3.82%
logBSEYD 8					24	70.83%	4.2965*	3.82%

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

Table 5: Robustness: Maximum Likelihood Estimate and Likelihood Ratio Test For Each Sample

The table reports likelihood statistics for each of two samples. The first sample covers the period from 1964 until the end of 1981. The second sample covers the period from 1982 until the end of 2012 for model implementations 1 to 4 and the period from 1983 until the end of 2012 for models 5 to 8. The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S . The Maximum Likelihood estimate \hat{p} is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. The estimated test statistics, equal to $-2 \ln \Lambda = -2 \ln \frac{L(p=\frac{1}{2}|X)}{L(p=\hat{p}|X)}$, is asymptotically χ^2 -distributed with 1 degree of freedom. The p -value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p -value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99% level are respectively 3.8415, 6.6349 and 7.8794.

Signal Model	Sample 1				Sample 2					
	Critical Value: 90%	Critical Value: 95%	Critical Value: 99%	Test stat. $-2 \ln \Lambda$	Empirical p -value	Critical Value: 90%	Critical Value: 95%	Critical Value: 99%	Test stat. $-2 \ln \Lambda$	Empirical p -value
P/E 1	3.1395	3.1395	5.8221	5.8221**	4.27%	2.9769	4.7166	6.9162	1.6573	28.08%
P/E 2	2.9419	2.9419	6.1977	6.1977**	3.78%	2.1646	3.6176	7.7998	2.1646	21.17%
P/E 3	2.3059	4.1860	6.7382	14.6994**	0.00%	2.3405	3.5164	6.6414	0.0286	86.58%
P/E 4	3.3970	3.3970	5.7823	13.4465**	0.07%	3.3980	3.3980	6.7939	0.5349	58.57%
logP/E 1	3.8549	3.8549	7.3613	7.3613*	2.09%	2.1646	3.6176	7.7998	2.1646	21.00%
logP/E 2	2.9419	2.9419	6.1977	6.1977*	4.20%	2.1646	3.6176	7.7998	3.6176†	9.13%
logP/E 3	3.3970	3.3970	5.7823	13.4465**	0.00%	2.6509	3.9896	5.6238	0.0323	71.58%
logP/E 4	3.1395	3.1395	5.8221	9.7515**	0.59%	2.3179	3.6515	7.3255	0.5734	56.19%
BSEYD 1	2.6407	4.4389	6.7828	13.5528**	0.09%	3.2913	3.2913	7.7098	0.2003	81.81%
BSEYD 2	2.0388	3.6830	5.8839	8.7331**	0.31%	2.6407	4.4389	6.7828	0.4757	64.66%
BSEYD 3	2.6407	4.4389	6.7828	18.5043**	0.00%	2.0388	3.6830	5.8839	0.8964	47.57%
BSEYD 4	2.0388	3.6830	5.8839	17.2292**	0.00%	2.0388	3.6830	5.8839	0.8964	48.43%
BSEYD 5						3.6176	7.7998	7.7998	2.1646	14.12%
BSEYD 6						3.9849	6.0595	8.6618	2.3786	12.30%
BSEYD 7						4.1860	6.7382	6.7382	4.1860*	4.08%
BSEYD 8						5.0168	7.7230	7.7230	2.9699†	8.48%
logBSEYD 1	2.6407	4.4389	6.7828	9.7654**	0.43%	3.3139	3.3139	7.1034	1.0068	42.60%
logBSEYD 2	3.2913	3.2913	7.7098	10.8175**	0.25%	2.1646	3.6176	7.7998	0.3924	68.01%
logBSEYD 3	3.2913	3.2913	7.7098	14.7226**	0.03%	3.2913	3.2913	7.7098	0.8054	50.06%
logBSEYD 4	2.6407	4.4389	6.7828	13.5528**	0.09%	2.6407	4.4389	6.7828	0.4757	64.94%
logBSEYD 5						3.9849	6.0595	8.6618	2.3786	12.30%
logBSEYD 6						4.7166	6.9162	9.6363	2.9769†	8.45%
logBSEYD 7						4.2965	6.2790	8.7075	4.2965*	3.82%
logBSEYD 8						4.2965	6.2790	8.7075	4.2965*	3.82%

† significant at the 10% level.

* significant at the 5% level.

** significant at the 1% level.

Table 6: Robustness: Monte Carlo Likelihood Ratio Test For Each Sample.

The table reports empirical statistics for each of two samples. The first sample covers the period from 1964 until the end of 1981. The second sample covers the period from 1982 until the end of 2012 for model implementations 1 to 4 and the period from 1983 until the end of 2012 for models 5 to 8. Columns 2 to 4 and 7 to 9 report the critical values at a 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in columns 5 and 10 are equal to $-2 \ln \Lambda = -2 \ln \frac{L(\hat{\theta} = \frac{1}{2} | X)}{L(\hat{\theta} = \hat{\theta} | X)}$. The level of significance indicated are based on the empirical distribution. The p -value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance indicated in the test statistics column and the p -value indicated in the table are both based on an empirical distribution generated through Monte-Carlo simulations.

Crash Prediction Measure	Robust Likelihood Ratio	Model	Asymptotic (χ^2) Robust Test Statistics	Empirical (Monte Carlo) Robust Test Statistics
P/E	0.1066	P/E 3	4.4777* 3.43%	4.4777* 5.00%
logP/E	0.1148	logP/E 3	4.3292* 3.75%	4.3292 5.48%
BSEYD	0.0437	BSEYD 2	6.2597* 1.24%	6.2597** 0.99%
logBSEYD	0.0316	logBSEYD 2	6.9080** 0.86%	6.9080* 1.19%

* significant at the 5% level.

** significant at the 1% level.

*** significant at the 0.5% level.

Table 7: Data Robustness of the Crash Prediction Measures (Models 1 to 4).

Column identifies the four crash prediction measures: P/E, log P/E, BSEYD and logBSEYD. Column 2 reports the robust likelihood ratio under the null hypothesis $H_0 : p_0 = \frac{1}{2}$, calculated as the highest likelihood ratio across all implementations of the measure. A higher robust likelihood ratio is indicative of a lower statistical significance. Column 3 identifies the model corresponding to the likelihood ratio reported in column 2. Columns 4 and 5 refer to test of significance performed using the asymptotic χ^2 distribution. Column 4 reports the robust test statistics and level of significance for the measure based on the χ^2 distribution. Column 5 gives the associated p -value. Columns 6 and 7 refer to test of significance performed with the empirical distribution generated by Monte Carlo simulations. Column 6 reports the robust test statistics and level of significance for the measure based on the empirical distribution. Column 7 gives the associated p -value.

Signal Model	(1)	(2)	(3)	(4)	(5)	(6)
	Total number of signals	Number of crashes preceded by a signal	Proportion of crashes preceded by a signal (%)	Number of correct predictions	Proportion of incorrect predictions (%)	
P/E 1	34	16	88.89%	2	11.11%	
P/E 2	34	15	83.33%	3	16.67%	
P/E 3	51	18	100.00%	0	0.00%	
P/E 4	45	18	100.00%	0	0.00%	
logP/E 1	33	16	88.89%	2	11.11%	
logP/E 2	32	15	83.33%	3	16.67%	
logP/E 3	46	18	100.00%	2	0.00%	
logP/E 4	40	17	94.44%	2	5.56%	
BSEYD 1	39	17	94.44%	1	5.56%	
BSEYD 2	37	16	88.89%	2	11.11%	
BSEYD 3	37	17	94.44%	1	5.56%	
BSEYD 4	36	17	94.44%	1	5.56%	
BSEYD 5	23	8	88.89%	1	11.11%	
BSEYD 6	21	8	88.89%	1	11.11%	
BSEYD 7	16	8	88.89%	1	11.11%	
BSEYD 8	17	8	88.89%	1	11.11%	
logBSEYD 1	44	18	100.00%	0	0.00%	
logBSEYD 2	43	17	94.44%	1	5.56%	
logBSEYD 3	40	18	100.00%	0	0.00%	
logBSEYD 4	38	17	94.44%	1	5.56%	
logBSEYD 5	21	8	88.89%	1	11.11%	
logBSEYD 6	22	8	88.89%	1	11.11%	
logBSEYD 7	24	8	88.89%	1	11.11%	
logBSEYD 8	24	8	88.89%	1	11.11%	

Table 8: Proportion of Crashes Preceded by a Signal.

The Total Number of Signal in Column 2 is a tally of the distinct signals. It is calculated as the sum of all the entries of the indicator sequence S . The Number of Crashes Preceded by a Signal reported in Column 3 is the tally of crashes preceded by the signal. The Proportion of Correct Prediction in Column 4 measures the accuracy of the signal. It is computed as the ratio of column (3) over column (2). The Number of Incorrect Predictions reported in Column 5 counts the number of signals that did not predict a crash. It is calculated as 18, the total number of equity market crashes, minus column (3). Finally, the Proportion of Incorrect Predictions in Column 6 measures the percentage of false positive given by the signal. It is computed as 100% minus column (3).