Predicting Crashes, a Model Comparison

Master Thesis

November 2017 - June 2018

L.F. Kroes

Student number: 374999

Under supervision of: Dr. A.J. Koning

Second assesment by: Dr. X. Leng

Erasmus University Rotterdam

Econonometrics & Management Science - Quantitative Finance

Date final version: 13-06-2018

Abstract

In this paper, the crash prediction performances of the Price Earnings ratio (PE) model and Bond Stock Earnings Yield Differential (BSEYD) model are tested against the crash prediction performance of the Discrete Time Disorder Detection (DTDD) model. The PE and BSEYD models are economic models, depending on economic indicators and backed by asset pricing theory, therefore having a firm theoretical backing. Meanwhile the DTDD model is a purely stochastic model, not depending on any economic data. All models were applied to (subsamples of) the S&P 500 and to the Nasdaq Composite Index. This research showed that, over all, the DTDD model with a model horizon of 1500 trading days was the best performing, most consistent and most robust predictor of the considered models. Moreover, the DTDD model was easier to use due to the lack of dependency on underlying data such as earnings data or treasury notes.

Keywords: stock market crashes, bond-stock earning yield differential, price-earnings-ratio, change point detection, optimal stopping problems

Contents

1 Introduction

Over the last decades, both investors and economists have investigated if equity market crashes are predictable. For investors, the ability to predict crashes would imply high, risk-free returns. While economists are interested in this ability in order to obtain better predictions of economic downturns, since equity markets are leading indicators for economic growth (Klein & Moore (1983), Hertzberg & Beckman (1989)). However, up to what extent are equity market crashes predictable? And if so, what is the best model to use when predicting equity market crashes?

The academic research on the field of return predictability of short- and long term financial asset prices started with the development of the Efficient Mark Hypotheses (EMH) by Malkiel & Fama (1970). Despite this hypothesis, several academics found evidence of predictability of long-term returns (Campbell & Shiller (1988)), which led to the first crash prediction models (the Price Earnings (PE) model, as described in Campbell & Shiller (1988, 1989, 1998)). Some time later, the research on the relation between long-term return prediction and crash prediction models advanced by the introduction of the Bond Stock Earning Yield Differential (BSEYD) model by Ziemba & Schwartz (1992), which proved to deliver good results on forecasting equity market crashes in the US, Iceland and China over the period 2007-2009 (Lleo & Ziemba (2013)). Shiller (1996) further cemented the relation between PE ratios and stock prices after which Yardeni (1997) introduced the Fed model, which is a special case of the BSEYD model. Koivu et al. (2005) showed that the Fed model has predictive power in forecasting changes in the equity prices, earnings and bond yields. Lleo & Ziemba (2017) linked both the PE and the BSEYD model (including the FED model) to the Gordon growth model (Gordon (1959)), hence showing the relationship between the above mentioned models and asset pricing theory. The connection between crash and return prediction model is intuitively easy to explain by the fact that both model types share the same starting point; an empirical analysis of historical data.

Moreover, Lleo & Ziemba (2017) showed that the BSEYD model, FED model and to a lesser extent the PE model are all statistically significant robust predictors of equity market crashes on the S&P 500 over a long period of time. All these models are economic models; they are based on economic indicators and can be linked to asset pricing theory, which offers a meaningful explanation of their predictability. This provides the conclusions of Lleo & Ziemba (2017) with a firm theoretical basis.

Despite the good performance of the PE, BSEYD and FED model in the paper by Lleo & Ziemba (2017), these models received some criticism on their forecasting performance. Shiryaev et al. (2017) showed that the BSEYD model performed (far) less good on other indexes like the NASDAQ 100 (a subset of the NASDAQ Composite index), and all of the economic models appeared to lose a lot of forecasting power in samples when crashes were less frequent (Lleo & Ziemba (2017)).

The research on crash prediction models did not stop with the introduction of these asset pricing theory backed models. The financial crisis of 2008 and a new generation of models based on probability theory

have led to newer crash prediction models. One of these is the Discrete Time Disorder Detection (DTDD) model, developed by Shiryaev & Zhitlukhin (2012) and Zhitlukhin (2014). This is a purely stochastic model, hence not depending on any economic indicator. The DTDD model was originally designed to provide exit points for long positions of stocks and, in short, assumes returns during bull and bear periods of an equity index are both normally distributed with the same variance but with a different drift. Based on a utility function and an initialisation period, the model determines 'changepoints'; the moment an equity market switches from bull to bear. Since the model has to identify the bear market before it gives a signal, all signals the model produces will take place slightly after the changepoint. Thus, the DTDD model which is used in this paper is strictly speaking not a crash *prediction* model, but a (very adequate) crash *monitoring* model[1]. Despite this lag, the model proved to provide good exit points for investors in the NASDAQ100 and the Apple Computer stock in 2010-2012 (Shiryaev et al. (2017)).

Considering these developments in the field of crash prediction, it will be very interesting to pursue the suggestion of Lleo & Ziemba (2017) to determine the difference in crash prediction performance of the different models on a leading economic indexes. This research will review *economic models* (models based on economic indicators; PE and BSEYD) and compare their performance with the more advanced, purely stochastic DTDD model proposed by Shiryaev & Zhitlukhin (2012). Also, the difference in performance over several smaller historical periods can be very interesting, since Lleo & Ziemba (2017) showed that the performance of both economic models was reduced in the most recently used subsample of their data.

In order to obtain a comparison, first crash predictions of the S&P 500 index will be constructed and compared against a benchmark using the test framework proposed in Lleo & Ziemba (2017). For this analysis 51 years of daily data will be used. The test framework is based on crashes (events) rather than returns, this leads to a binary signal on every point in time for every model (crash or no-crash). The binary interpretation makes it possible to test efficiently using a Likelihood Ratio test in order to test the models against a benchmark and compare their performances against each other.

After the construction and analysis of the S&P 500 crash predictions for all models over the complete sample, their performance will be analyzed over two different subsamples: 1964-1981 and 1982-2012, these subsamples will be described further in section 2.5. Lleo & Ziemba (2017) found a decrease in performance of all models when applied to the second sample, hence it will be interesting to determine whether or not the stochastic models are more consistent. Finally, all models will be applied to the NASDAQ Composite Index, since Shiryaev et al. (2017) pointed out a poor performance of the BSEYD model when predicting crashes for a subset of this index.

---

[1]For the sake of simplicity, all crash monitors of the DTDD model will be referred to as 'predictions' throughout this research.

More specifically, this thesis has the following form; first the research of Lleo & Ziemba (2017) will be reproduced, including the reconstruction of the test framework. Next, the DTDD model of Shiryaev & Zhitlukhin (2012) will be constructed, applied to the same dataset and also included in this test framework. Hereafter the performance of all models can be compared. Subsequently, all models will be applied to the NASDAQ and subsamples of the S&P 500, in order to gain more insight in the data robustness and consistency of all models.

1.1 Research Questions

The research in this thesis can effectively be described in the following three research questions:

How does the Discrete Time Disorder (DTDD) Model of Shiryaev & Zhitlukhin (2012) perform relative to models based on economic indicators, like the PE and the BSEYD model?

How is the performance of the DTDD model compared to the PE and BSEYD model concerning different subperiods of the dataset?

How do all proposed models perform when applied to another index?

This paper is organized as follows; in the next section the methodology of this research is described. In the section thereafter the obtained results of the models' crash predictions are reviewed, where after there is a conclusion and discussion section.

2 Methodology

In this section the dataset and all proposed models are described in detail, together with the assumptions which are made. Moreover, the test framework and of the paper of Lleo & Ziemba (2017) will be explained in order to be able to compare the PE and BSEYD model with the DTDD model.

2.1 Data

For this research roughly the same dataset as the paper of Lleo & Ziemba (2017) will be used. That is, daily data from the S&P 500 and ten-year treasury notes for the period of 51 years (January 2, 1962 till December 31, 2012). This comes down to 13,306 observations of daily data. Observations on indexed-averaged earnings are collected at a monthly rate, here the methodology of Lleo & Ziemba (2017) is followed, who interpolate these monthly observations to daily observations since the loss of variance which is caused by this action does not outweigh the increase in prediction power of having a daily instead of a monthly dataset.

In order to give more insights on the performance of the models on other indexes, all models will be applied to the NASDAQ Composite index over the period January 1, 1984 till January 1, 2018 (38 years). This index is chosen since the poor performance of the BSEYD model on this index according to Shiryaev et al. (2017). Furthermore, since the NASDAQ is made up out of mainly information technology companies who are known to be more volatile than other companies, the NASDAQ is almost twice as volatile than the S&P500 index[2]. This will lead to extra insights concerning the predictive robustness of all models.

All datasets are available on Bloomberg and/or Federal Reserve Economic Database.

2.1.1 Crash Definition

The definition of a equity market crash in this research will be: a decline of at least 10% in the level of the index over a time period of at most a year. All crashes will be considered as passed when the decline compared to the year-high is less than 10%. Each crash will have a *peak date*, which equals the point in time from which 10% is lost at the *crash-identification date*, moment the crash is realized. Finally, every crash will have a *thru-date* (the absolute low of the crash).

---

[2]Volatility measured by absolute first difference over the period 1971-2012, adjusted for index size

2.1.2 Detected Crashes

The identified crashes at the sample of the S&P 500 are presented in Table 5. Table 5 shows that 18 crashes occurred during this period at the S&P 500 index, the longest crash lasted 622 days (1980-11 till 1982-08), while the shortest lasted only 43 (1980-02 till 1980-03). The average crash lasted 285 days with an average decline of around 25%. Lleo & Ziemba (2017) find the same crashes over this time period. Analyzing all crashes further, it becomes clear all major economic crashes over this period are included in the table, including the *oil crisis (1973-1975)*, *black Monday (October 1987)*, the *.com bubble* (1999-2000) and the *financial crisis of 2007-2008*. Furthermore, crashes are more present in the first half of the sample than during the second half. Due to this inconsistent appearance of crashes, the sample is divided into two unequal sized samples during the subsample analysis (more on the subsample analysis in section 2.5).

As presented in Table 6, the NASDAQ index has 12 crashes occurring in 30 years of data. The longest crash lasted 810 days and included both the *.com bubble* as the *9/11 terror attacks*. While the shortest crash only lasted 30 days (8-1994). The average crash lasted slightly shorter compared to the S&P 500, namely 159 days. Moreover, crashes seem to appear on a way more consistent basis compared to the S&P 500, with roughly one crash every 3 years.

2.2 The Economic Models (Lleo & Ziemba (2017))

The first category of models are the economic models, which will act as a benchmark for the performance of the DTDD model. A selection of the models presented in Lleo & Ziemba (2017) will be constructed, namely the Price Earnings (PE) model and the Bond Stock Earning Yield Differential (BSEYD) model. The choice for these models is based on the fact that the latter one (BSEYD model) is the best performing model in the paper of Lleo & Ziemba (2017) , while the PE model is very intuitively and its performance concerning stock performance is well documented (Fama & French (1992)).

2.2.1 Price Earnings Model (Campbell & Shiller (1988))

To use the Price Earnings ratio as crash predictor, first the link between PE ratios and stock-index returns needs to be shown. This was done by Campbell & Shiller (1988), who investigated the relationship between the log return and the log dividend-price ratio, lagged dividend growth rate and average annual log real earnings over the previous 10 and 30 years. During this research they analyzed the individual descriptive power of each explanatory variable. In order to do this, they defined P_t the stock price at the beginning of period t and D_t the dividend received over period t. It is assumed that all dividends are future valued to the end of period t ($D_t = D_{end}$). now the 'continuously compounded return' $h_{1,t+1}$

can be defined as:

$$h_{1,t} \equiv log\left(\frac{P_{t+1} + D_t}{P_t}\right) \tag{1}$$

Since the continuously compounded return is log-linearised, the i-th period total return can be obtained by taking the sum over the individual returns.

$$h_{i,t} \equiv \sum_{j=0}^{j=i-1} h_{1,t+j} \tag{2}$$

To obtain insights in the individual explanatory power of a variable X, Campbell & Shiller (1988) regressed the i-th period continuously compounded return on explanatory variable X we have the following form:

$$h_{i,t+1} = \alpha + \beta X_t + \epsilon \tag{3}$$

When reviewing the results of the individual regressions performed in Campbell & Shiller (1988), it becomes clear that the individual regression on the Price / (30 year average earnings) ratio yields the highest R^2 of all individual regressions, namely 0.566. The predictive power of Price/(10 years average earnings) is slightly lower with 0.401. However, in both cases dependency of the continuously compounded return on the PE ratio is clearly shown. Lleo & Ziemba (2017) decided to use only a single explanatory variable as basis for the P/E model, due to the high individual explanatory powers obtained from analysing the individual predictors and the fact that this will keep the model simple. Since this research is focused on medium-term market downturns instead of long-term, the PE ratio with 10 years average earnings will be used as predictive variable on the crashes of the different indexes.

Formally, the Price/(10 year average earnings) explanatory variable comes down to:

$$\overline{E_{t,-10}} = \frac{1}{n}\sum_{i=0}^{n-1} E_{t-i}, \tag{4}$$

with E_t equaling the earnings over period t and n the total number of time periods in the last 10 years.

The regression of Campbell & Shiller is done on the 10 years of data preceding the first data point for which a signal needs to be constructed. The crash prediction measure is constructed as:

$$M_{PE,t+1}(t) = \frac{P_t}{E_{t,-10}} \tag{5}$$

This version of the price earnings ratio is also called the Cyclically Adjusted Price Earnings (CAPE) ratio. Here, P_t equals the daily opening price of the S&P 500. This measure indicates a crash when it exceeds a certain threshold $K(t)$ (see section 2.2.3). Besides the measure defined in equation 5, a logarithmic version will also be considered for completeness:

$$M_{log\_PE,t+1}(t) = log\left(\frac{P_t}{E_{t,-10}}\right) \tag{6}$$

2.2.2 Bond Stock Earnings Yield Differential Model

Next, the BSEYD model of Ziemba & Schwartz (1992) is constructed. This model states that the rate on a ten-year treasury note at time t $(r(t))$ needs to equal the earnings yield at time t $(\rho(t))$. This comes down to:

$$M_{BSEYD,t+1}(t) \equiv r(t) - \rho(t) = r(t) - \frac{E_t^{beg}}{P_t^{beg}}, \tag{7}$$

where E_t^{beg} and P_t^{beg} again are defined as respectively the earnings and stock price at time t. Since the earnings yield is the inversion of the PE model, the BSEYD model can be seen as an version of the PE model where extra information is added in the form of the 10-year T-bill rate yield. Just like the PE model, the model gives a signal crash signal when is passes a certain threshold $K(t)$. Again, a log version of equation 7 will be considered, this logarithmic version is also considered in Koivu et al. (2005) concerning return prediction and yielded good results in Lleo & Ziemba (2017).

$$M_{Log\_BSEYD,t+1}(t) \equiv log\left(\frac{r(t)}{\rho(t)}\right) = log\left(r(t)\right) - log\left(\frac{E_t^{beg}}{P_t^{beg}}\right) \tag{8}$$

2.2.3 Signal Creation of the Economic Models

The PE and BSEYD models will give signals over time when they identify a crash. This dummy series $SIG_i(t)$ can be described as follows:

$$SIG_i(t) = \begin{cases} 1 & \text{if } M_i(t) - K_i(t) > 0, \\ 0 & \text{if } M_i(t) - K_i(t) \leq 0, \end{cases} \tag{9}$$

here, $M_i(t)$ is the measure of model i (described in the sections 2.2.1 and 2.2.2) and $K_i(t)$ is a threshold. The threshold $K_i(t)$ will be defined as the critical value of a one-sided confidence interval of the corresponding model' measure. A normal distribution for these measures will be assumed which leads to a 'standard' confidence level and corresponding critical values. This assumption can be backed using the same empirical analysis as performed at the research of Lleo & Ziemba (2017), who assume a normal distribution of the BSEYD measure despite failing the Jarque-Bera test for normality (see Appendix J for this analysis on the PE measure). In order to also obtain somewhat more robust results and obtain more insights in tail behaviour, a one-tailed version of Chebyshev's inequality will be considered which is known as Cantelli's inequality. This inequality links the probability that the distance between a random variable X and it's mean μ exceeds more than y standard deviations σ. This leads to a confidence interval independent of an underling distribution. Formally this can be noted as:

$$P[X - \mu \geq y\sigma] = \frac{1}{1 + y^2},$$

when substituting $\alpha = 1/(1 + y^2)$ this equation becomes

$$P\left[X - \mu \geq \sigma\sqrt{\frac{1}{\alpha} - 1}\right] \leq \alpha \tag{10}$$

In this equation, the researcher is free of it's choice for the value of y, during this research the suggestion of Lleo and Ziemba (2017) will be used to set α equal to 25%, which corresponds to a value for y of $\sqrt{3}$.

For the construction of these signals rolling window means an averages will be used whereby the rolling window length will be set equal to 1 year (H= 252 trading days).

Finally, the signal series $SIG_i(t)$ will be converted to a distinct signal series. Signals will only be registered in the distinct signal series if there is a period of at least 30 days prior to the signal in which no signals occur.

2.2.4   Asset Pricing Theory

Both proposed economic models seem empirical, to explain the predictive success of these models in a theoretical way, the models will be linked tot the Gordon growth model (Gordon (1959)). A brief version of the derivation, which was first presented in Lleo & Ziemba (2017), is given below. For the full version please see Appendix A . According to Gordon (1959), the price of a stock at time t is defined as:

$$P_t^{end} = \frac{D_{t+1} + P_{t+1}}{1 + k}, \tag{11}$$

with k representing the cost of Equity. This can be rewritten to (see Appendix A for details):

$$h_{i,t} = log\left(\frac{P_t + D_t}{P_{t-1}}\right) = h_{i,(t-1)}, \tag{12}$$

hence the two models are equivalent, except a small change in notation. Thus the continuously compounded return can be rewritten to (a version of) the Price/Earnings ratio when using the definition of the price of a stock.

The BSEYD model can also be related to the Gordon Growth model, since we can rewrite the price of a stock to:

$$\frac{1}{d(1+g)}(r_t[d(1+g) - 1] - f_t + g) = r(t) - \frac{E(t)}{P(t)} = M_{BSEYD,t+1}(t), \tag{13}$$

where g is an assumed growth rate, d is the corresponding discount rate, f_t the equity risk premium and r_t the yield on a government bond. (full derivation can be found in Appendix B). After this rewriting, the price of a stock according to Gordon (1959) is expressed in terms of the yield on a treasury note minus the earnings yield; the BSEYD model.

These derivations show that the economic models are a version of the expression for the price of a stock. Thus, the PE and BSEYD models are no longer empirical but backed by a theoretical framework.

2.3 The Discrete Time Disorder Detection Model (Shiryaev et al. (2012))

In order to be able to describe the evolution of a stock price and detect disorder points in a purely stochastic way, Shiryaev & Zhitlukhin (2012) came up with the Discrete Time Disorder Detection Model. This model describes the progression of stock price/index S at (discrete) times $0, 1, \dots T$. It is assumed that the logarithmic innovations of this process follow an pre-initialised normal distribution with parameters (μ_1, σ_1^2) up to an unknown point in time θ. After this point in time (the *changepoint*), the innovations start to follow a different normal distribution, namely with parameters (μ_2, σ_2^2). This can be seen as the change from a *Bull* to a *Bear* market. The model aims to detects this changepoint θ as soon as it happens, which in our setting will lead to a crash indication signal. A visual representation of the model applied to the stock rally of Apple Inc. is presented in figure 1.
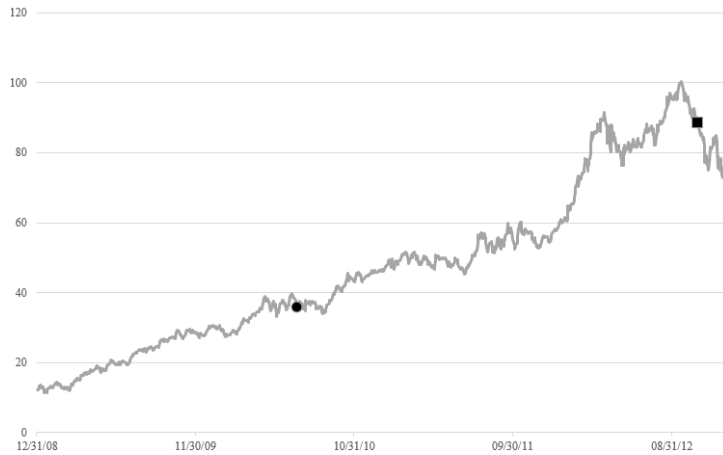
Figure 1: The DTDD model applied on the stock price of Apple inc. over the period January 2009 till December 2012. The round indicator indicates the starting point of the model (30-06-2010) and the square indicator indicates the obtained exit point (08-10-2012). The results presented in this figure are obtained from Shiryaev et al. (2017)

More formally and complete, setup of the model is described as follows:

Define S_t the index price at time t, $S_0 = 1$ and:

$$X_t = log\frac{S_t}{S_{t-1}} = \begin{cases} \mu_1 + \sigma_1^2\eta_t & \text{if } t < \theta, \\ \mu_2 + \sigma_2^2\eta_t & \text{if } t \geq \theta, \end{cases} \tag{14}$$

with $(\mu_1, \mu_2) \in R$ and $(\sigma_1^2, \sigma_2^2) > 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. The values of μ_1 and σ_1 are initialised over an predetermined period, the values of μ_2 and σ_2 are subjective.

To find the changepoint θ, the probability of θ appearing at a certain time t is defined as $p_t = P(\theta = t)$. Furthermore, to prior distribution of p_i up to a certain time T is defined as $G(t) = p_1 + \cdots + p_t$. Subsequently, p_{T+1} and $G(T+1)$ are respectively the chance and sum of chances of θ not appearing in the time frame $[0, \dots, T]$.

The main problem in this setting is finding a moment in time τ which maximizes the utility of an opened

long position at time 0 which needs to be closed before time T. Therefore, a utility measure needs to be constructed which measures the utility of both the bull and the bear model setting. Hereafter, the problem can be seen as a optimal stopping problem with respect to function U_α, which is defined as:

$$U_\alpha(x) = x^\alpha \tag{15}$$

for $\alpha < 0$. At this research, α will be set equal to 0, $U_\alpha(x)$ will be set equal to $U_0(x) = log(x)$ to prevent the model from a constant utility for all periods of time. This utility function is known as the Decreasing Absolute Risk Averse (DARA) utility function, other utility functions were considered at Shiryaev et al. (2015), however they did not yield significantly different results compared to $\alpha = 0$, hence this research refrains from using different values for α. In this section the problem is described using the notation of Zhitlukhin (2014), for simplicity however, all formulas are adjusted to the assumption of α equaling 0.

The crux of this problem is the fact that only $\mathcal{F}_t$ (the filtration of S up to t) can be used to determine this optimum, and not S_{t+1}, S_{t+2}, since these values are unknown at time t. This makes the problem become an optimization which can be described as:

$$V = \sup_{\tau \in M} E[U(S_t)], \tag{16}$$

with M denoting the class of all stopping times $\tau < T$ of $\mathcal{F}_t$.

The values of μ_1 and μ_2 in equation 14 are assumed to be ($\mu_1 > 0$) and ($\mu_2 < 0$). Under these assumptions the sequence of utility $U(S_t)$ increases if the (log) increments are distributed i.i.d. (μ_1, σ_1^2) and decreases if the i.i.d. normal distribution has parameters (μ_2, σ_2^2). Therefore, the parameter θ denotes the moment in time when the indexed value of S_t starts to decrease, or formally speaking; *the drift changes from positive to negative since a bull-market-period is changed to a bear-market-period.*

In order construct this stopping time, the Shiryaev-Robberts Statistic (ψ) will be formulated. The original definition of the Shiryaev-Robberts statistic is very technical since it takes in to account possible different probability measures for both states of the model. In the case of this research however, this statistic comes down to the likelihood-ratio for all points in time, and is defined as follows:

$$\psi_0 = 0, \quad \psi_t = \sum_{u=0}^{t} \frac{dP_t^u}{dP_t^\infty} p_u \quad for \quad t = 1, \ldots, T, \tag{17}$$

here, dP_t^u denotes the behavior of the system when the disorder is not present, and dP_t^∞ denotes the system' behavior when the disorder is present. Knowing both distributions from the estimation of equation 14, dP_t^u/dP_t^∞ can be formulated as:

$$\frac{dP_t^u}{dP_t^\infty} = \left(\frac{\sigma^1}{\sigma^2}\right)^{t-u-1} \exp\left(\sum_{i=u}^{t} \frac{(X_i - \mu_1)^2}{\sigma_1^2} - \frac{(X_i - \mu_2)^2}{\sigma_2^2}\right) \tag{18}$$

13

This can easily be formed in to the following recurrent form:

$$\psi_t = \frac{\sigma^1}{\sigma^2}(p_t + \psi_{t-1})e^{\frac{(X_t-\mu_1)^2}{\sigma_1^2} - \frac{(X_t-\mu_2)^2}{\sigma_2^2}},$$

(19)

where $\psi_0 = 0$.

Next, a time dependent threshold needs to be constructed with respect to ψ_t. Therefore, a family of recurrent functions $V(t,x)$ will be determined based on the parameter values of equation 14. By taking the optimal roots (≥ 0) for every t, one can construct $b(t)^*$, the time dependent threshold for ψ_t.

The family of functions $V(t,x)$ corresponds to the a function V(x) for every point in time which calculates the expected utility for being in both states times the probability of being in the respective states.

The functions of $V(T,x)$ is recurrently computed by:

$$V(t,x) = max\{0, [\mu_2(x + p_{t+1}) + \mu_1(1 - G(t+1))] + f(t,x)\}$$

(20)

This function is bounded at 0 since a it represents a utility, which can not be less than 0. Furthermore, $V(0,x) = G(T+1)$ and function $f_\alpha(t,x)$ is defined as:

$$f(t,x) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{\mathbb{R}} V_\alpha\left(t+1, \frac{\sigma^1}{\sigma^2}(p_{t+1}+x)e^{\frac{(z-\mu_1)^2}{\sigma_1^2} - \frac{(z-\mu_2)^2}{\sigma_2^2}}\right) e^{\frac{(z-\mu_1-\alpha\sigma_1^2)^2}{\sigma_1^2}} dz$$

(21)

Now the final theorem can be formally formulated:

$$\tau^* = \inf\{0 \leq t \leq T : \psi_t \geq b^*(t)\},$$

(22)

where the stopping boundary $b^*(t)$, $t = 0, \ldots, T$ is

$$b^*(t) = \inf\{x \geq 0 : V(t,x) = 0\}$$

(23)

Theorem 22 states that the first moment in time the Shiryaev-Roberts statistic (equation 19) exceeds the time dependent threshold $b^*(t)$. In order to find this function $b^*(t)$ numerically, one first computes the value $V(t,x)$ (equation 20) and find their optimal roots $x \geq 0$.

2.3.1 Proof of the DTDD Model

The problem as a whole can be seen as a Markovian optimal stopping problem with respect to the Shiryaev-Roberts statistic ψ, in the following section, an brief sketch of the proof will be given. For the full version of the proof, please see section 3.3 of Zhitlukhin (2014), while a further intuition of this model is given in Appendix D.

The proof consists of three parts, first it will be shown that the Shiryaev-Roberts statistic ψ_t is a martingale (under the used probability measure). Next, it need to be shown that ψ_t is Markov sequence,

since that makes it possible to apply some general theory on the field of optimal stopping with respect to Markov sequences. Finally, it needs to be proven that $V(t, x)$ is continuous and non-increasing, since this makes it possible to express the final problem as equation 22.

The first step of the proof can be done by proving that the expectation of $\psi_t = 0$ for all t. More formally this comes down to:

$$EU(S_T) = E\left[\sum_{t=1}^{\tau}[\mu_2\psi_t + \mu_1(1 - G(t))]\right],\tag{24}$$

which is proven to equal 0 in Zhitlukhin (2014) section 3.3.

Next, it need to be proven that equation 16 equals a Markov sequence. Using result from the first step of the proof, equation 16 can be written as:

$$V = \sup_{\tau \in M} E\left[\sum_{u=1}^{\tau} F(u, \psi_u)\right]\tag{25}$$

This function can be formed into a family of functions, dependent on x and t:

$$V(t, x) = \sup_{\tau \in M} E\left[\sum_{u=1}^{\tau} F(t + u, \psi_u(t, x))\right]\tag{26}$$

This formula satisfies the conditions for an Markov sequence since it only depends on ψ_t which is a function of ψ_{t-1} and X_t, which is an independent sequence of variables.

Now, the Ward-Bellman equation can be applied to this result, resulting in this alternative notation of equation 20;

$$V(t, x) = max\left\{0, [F(t + 1, \psi_{t,x}) + V(t + 1, \psi_{t,x})]\right\},\tag{27}$$

here, 0 equals the gain from stopping directly, $F(.)$ represents the expected gain from not stopping at time t and $V(.)$ represents the expected gain from all future data points.

Hereafter it is intuitively easy to follow that the optimal stopping time is reached when an investor is indifferent between the left and right side of $V(t, x)$, which more formally can be described as equation 22:

$$\tau^* = \inf\{0 \leq t \leq T : \psi_t \geq b^*(t)\}\tag{28}$$

where the stopping boundary $b^*(t)$, $t = 0, \ldots, T$ is

$$b^*(t) = \inf\{x \geq 0 : V(t, x) = 0\}\tag{29}$$

15

The proof is completed by showing the continuity and non-decreasingness of the $V(t, x)$ function, this is also shown in Zhitlukhin (2014) section 3.3.

2.3.2 Initialisation & Assumptions DTDD Model

In order to use the above described model starting from a certain point in time, one needs to specify certain initialisation values and ratios. First, its assumed that θ is distributed uniformly over the interval $[1, \ldots, T]$. Also, the values of μ_1 and σ_1^2 are initialised using previous data $S_0, \ldots S_{t_0}$, where t_0 equals 100. One is free in the choice on parameters μ_2 and σ_2^2, in this research the assumptions of Shiryaev et al. (2017) will be followed, who state that all variations on $\mu_2 = -\mu_1$ did not led to a significant improvement of results. Hence, μ_2 will be set equal to $-\mu_1$ and σ_1 will equal σ_2 during this research.

The other parameter which needs to be set is the *model horizon T*. This parameter determines the size of the interval on which the model will detect a changepoint. The model will run for 3 different values for T, namely 750 (3 years), 1000 (4 years) and 1500 (6 years). The will assume a changepoint will always be present in the time frame between t and the model horizon T, thus $G(T+1)$ will be assumed to equal 0. Hence, for every point in time t, a changepoint will be detected between $t + 0$ and $t + T$.

Finally, the model is designed to detect only one changepoint over the time period $[t, \ldots, t + T]$, since we want to use the model to predict all crashes over our dataset, we will run the model for every point in time and thus collect a changepoint from the perspective of every point in time.

2.3.3 Crash Signal Creation DTDD Model

In order to convert predicted changepoints for every point in time to crash predictions, all created changepoints need to be *clustered* and need to be converted to *distinct* signals. The clustering is done by making one dummy series $SIG_{DTDD}(t)$ which equals 1 if any point in time predicts a changepoint at time t. Next, this series $SIG_{DTDD}(t)$ is made distinct by eliminating all signals which did not had a signal free period of 30 days preceding it. After these procedures, $SIG_{DTDD}(t)$ is a dummy crash prediction series which can be compared to the crash prediction dummy series of the economic models.

To conclude on the model construction section of the Methodology, an overview of all models used in the analysis is given in Appendix E.

2.4 Test Framework

To be able to compare the crash prediction series of the models described in sections 2.2 and 2.3, a likelihood ratio test framework is used. In order to use this, a first test horizon needs to be set. Moreover,

to analyze the assumption of the underling distribution of the test statistic over smaller samples, a Monte Carlo study will be performed.

2.4.1 Test Horizon

The horizon parameter has to determine whit-in what time frame a $SIG_i(t)$ needs to be followed by a crash in order to be called a good prediction. Since the definition of a crash equals a decline of the equity market index by at least 10% in one year, it is a natural choice to set this *test horizon H* equal to 504 (two trading years) days before the crash identification date. The length of the test horizon leads to the fact that all models can provide more correct signals than there are crashes, since multiple signals can occur within the test horizon of 504 days. The models can identify a crash from 2 years to 1 day before the crash identification date. This definition is again closely linked to existing literature (Lleo & Ziemba (2017)).

2.4.2 Likelihood Ratio Test

To test whether which of the model performs (relatively) the best, maximum likelihood estimator will be used. In order to use this, first the crash predictions signals $SIG_i(t)$ of all models will be turned in to hit rates HR. In order to do this, first dummy statistic $X_{i,t,H}$ is introduced. This statistic equals 1 if a signal was created by model i at time t which is followed by a crash within test horizon H, and 0 otherwise.

$$HR_{i,H} = \frac{C_{i,H}}{N_i}, \quad \text{with} \quad C_{i,H} = \sum_{t=0}^{t=T} X_{i,t,H} \tag{30}$$

here $C_{i,H}$ equals the sum of all correct signals; signals followed by a crash within the test horizon H. N_i equals the total signals created by model i, $N_i = \sum_{t=0}^{t=T} SIG_{i,t}$. In Appendix C is shown that this hit rate equals the maximum likelihood estimator of the historical proportion of correct predictions out of all observations ($HR_{i,H} = P(C_{i,H} = 1|SIG_{i,t} = 1)$).

This Maximum Likelihood estimator of each model is tested against a benchmark of random performance. This research will follow the suggestion of Lleo & Ziemba (2017) that an arbitrary, uninformed prior should correctly predict crashes around 50% of the time ($p = 1/2$). Hereafter, the performances of all models can be compared by comparing their performances against this benchmark.

Now, the likelihood-ratio is constructed as follows:

$$\Lambda = \frac{L(p = \frac{1}{2}|X)}{L(p = \hat{p}_i|X)} \tag{31}$$

where $p = \hat{p}_i$ equals $HR_{i,H}$. Hereafter the test statistic can be calculated: Y = -2log(Λ), with Y asymptotically χ^2-distributed with $v = 1$ degrees of freedom.

2.4.3   Monte Carlo Study for Small Sample Bias

The likelihood ratio test described in section 2.4.2 is distributed χ^2 asymptotically. This can not blindly be assumed in the case of this research, since the total number of (correct) signals is expected to be small[3]. In order to account for this problem, a Monte Carlo simulation will be used to construct an empirical distribution of the test statistic, which can be used to back-test the results obtained from the test proposed in section 2.4.2.

The Monte Carlo simulation to test the null hypothesis $p = p_0 = 1/2$ will be constructed as follows:

First, a large number K of independent paths will be be constructed. For each path, N binomial values X_N^k will be calculated with probability p_0 of being 1. Here, N equals the number of total generated signals when the models were applied to the data. Next, the constructed paths will be converted to hitrates $\hat{p}_k$. Formally, this comes down to:

$$X^k = \{X_i^k, i = 1, \ldots, N\}, \quad \hat{p}_k = \frac{\sum_{i=1}^N X_i^k}{N} \tag{32}$$

Next, the test statistic of every path can be calculated on the same way at at the LR test of section 2.4.2:

$$Y_k = \text{-2 log } (\Lambda_k) = \text{-2 log } \frac{L(p = \frac{1}{2}|X_k)}{L(p = \hat{p}_k|X_k)} \tag{33}$$

After running this algorithm for a large number of paths (this research will use $K = 10,000$), one can sort Y_k in order to obtain an empirical distribution for the test statistic. Using this empirical distribution, all obtained test statistics from the asymptotic Likelihood Ratio test can be back-tested.

2.5   Subsample Analysis

As Lleo & Ziemba (2017) point out in their conclusion, the performance of the Economic models (PE and BSEYD) differ a lot over time. They divided their dataset into two subsamples: one from 1 January 1964 till 31 December 1981 and the second from 1 January 1982 till 31 December 2012. They do not point out a reason for the (large) loss in performance at the second sample compared to the first. In

---

[3]Both this research as Lleo & Ziemba (2017) use 51 years of daily data and found only 18 crashes during this sample period

this research we will reconstruct this subsample analysis and compare the results to the DTDD model applied on these samples in order to gain more insights on this loss in performance.

The intuitive reason behind the action to divide the sample in two unequally sized parts of 18 and 30 years lies in the number of crashes occurring in both samples, since this equals 9 for both samples.

3 Results

In the following section the results of the constructed models are examined. The performance of all considered models on the S&P500 data will be examined in sections 3.1, 3.2 and 3.3. Hereafter, the performance of the models over different subsamples of the S&P 500 data will be described in section 3.4. Finally, the performance of all models on the NASDAQ Composite index will be examined in section 3.5.

3.1 Campbell - Shiller (PE) Models on the S&P 500

In order to construct the PE models as described in Lleo & Ziemba (2017), first the 10 years trailing *Cycally Adjusted Price Earnings* ratio needs to be constructed, as described in section 2.2. This ratio forms the basis of both the *PE* as the *log PE* model, and the rally of this ratio over time is presented in figure F in the Appendix.

Next, the PE and log PE models can be constructed. Examples of the obtained results are presented in the plots figures in 2a and 2b. There, both the PE 1 as the log PE 1 model their model and threshold lines are plotted against each other. One can clearly see both model lines crossing the threshold numerous times, hence generating signals. Furthermore, when comparing these plots with the figures presented in Lleo & Ziemba (2017), it becomes clear they are identical.

The rest of the (log)PE model plots are presented in the Appendix F, who are all again identical to the plots in the appendix of Lleo & Ziemba (2017). The similarities with this paper were in line with the expectations of this research, since identical data sets and methodology are used for this part.

(a) PE 1 model signals and Threshold

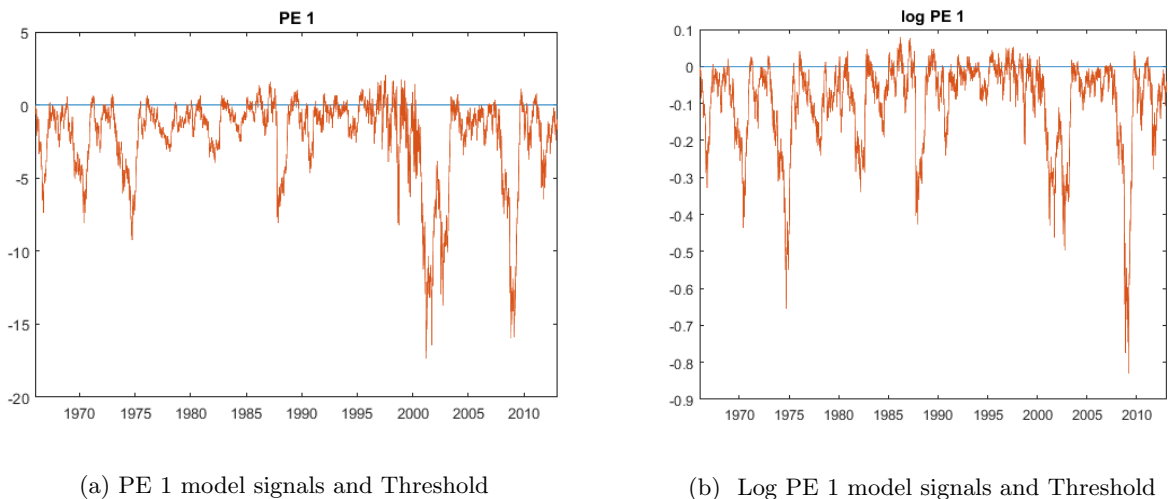(b) Log PE 1 model signals and Threshold

Figure 2: The PE 1 and log PE 1 model' thresholds (blue) and signals (red) over the period 01-01-1964 till 31-12-2012 plotted against each other.

Next, the results of the obtained hit rate values for all PE models are presented in the first 4 lines of Table 2. Here, one can see that all models perform about the same concerning both the amount generated signals as the proportion of correct signals. The total number of signals is slightly higher for the two models based on standard confidence intervals, while hit rates vary from 65% for the PE 1 model up to 69% for the PE 2 model. The suspicion created by the plots presented in the section above can be confirmed by this table; the results obtained exactly match the results of Lleo & Ziemba (2017).

The likelihood ratio test is performed using the obtained hitrates and led to the results which are presented in Table 7 in the Appendix. All models outperform the randomly and uninformed benchmark of $p = 1/2$ on the 95% significance level. This performance coincides with the works of Campbell & Shiller (1988, 1989, 1998), who attribute good forecasting power to the PE model. The log PE model, which has a lower variance by nature compared to the PE model, performs slightly less compared to the PE model, Lleo & Ziemba (2017) argue that this might be caused by the transformation implied by taking the natural logarithm.

The difference in choice of confidence level ((log)PE1 vs (log)PE2 models) is quite substantial (more than 1 percentage point). This confirms the assumption of normality made in the methodology section.

3.2 BSEYD Models on the S&P 500

The results of the BSEYD models applied to the S&P 500 are presented in Figure 6c till 6f and row 5 to 8 of Table 2. The graphs look very similar to the PE and Log PE models. However, an closer look to the obtained hit rates in Table 2 shows an lower amount of generated signals for all models compared to the (log) PE models. The amount of correct predicted crashes is in the same order of magnitude as at the (log) PE models, leading to a higher hit rate for all BSEYD models compared to the models presented in section 3.1.

The likelihood ratio test results are presented in Table 7, all models are significant on the 99% level. Also, all models outperform the (log) PE models. These results are in line with earlier studies by Ziemba et al. (2008), who show that the BSEYD model has a good track record when it comes to predicting crashes in several international markets, including the S&P 500. Just like the PE model, the log version of the BSEYD model performs less good due to the transformation made (Lleo & Ziemba (2017)). Moreover, the gain in predictive power of the BSEYD model compared to the PE model seems to imply that the addition of interest rate level to the model adds relevant information for this index. Finally, the difference in choice of confidence level is smaller compared to the PE models, making the assumption of normality less strong for this measure.

3.3 DTDD Models

To obtain a better understanding on how the DTDD model works, several plots on this model using data around May 15, 2007 are presented in Figures 3a, 3b and 4a. In Figure 3a one can clearly see that for larger values of T, larger values of ψ_t are needed in order to obtain a changepoint at the same point in time t. In Figure 3b the values for ψ_t (blue line) are plotted against the values for $b^*(t)$ (red line) when T is set at 1000 (DTDD 2). The line representing ψ_t breaks the red line around 170 data point after the start date, thus generating a signal.

(a) The vectors of $b^*(t)$ (b) DTDD 2 (T = 1000) on 15-05-2007
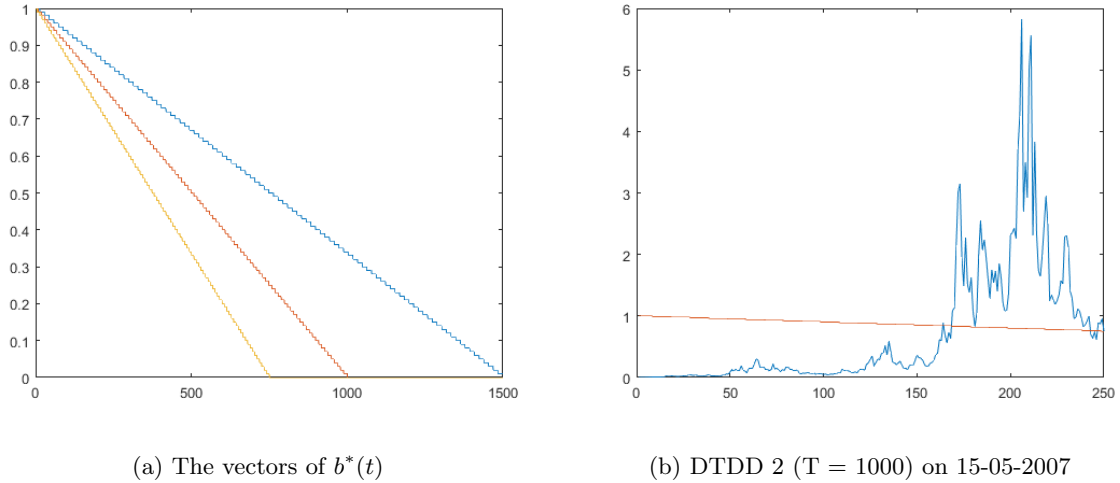
Figure 3: In the (a) section of this figure $b^*(t)$ is plotted for each considered model horizon. The yellow line indicates $b^*(t)$ for $T = 750$, the red line $T = 1000$ and the blue line $T = 1500$. The (b) section shows the DTDD 2 model (T=1000) in full action on datapoint 15-05-2007. The blue line indicates the value for $\psi(t)$ over time, while the red line indicates (the first 350 datapoints of) $b^*(t)$ over time. One can see a signal being generated at $t = 164$.

This example is also visible in figure 4a, were the stock price rally around this date is presented. The starting date for the model is pointed out with the black circle on the stock price process. Preceding this point, the initialisation period is indicated between the two black lines (100 days). Further down the stock price process, the generated changepoint and thus, signal, is presented using the black rectangle, which corresponds to the 8th of January 2008. Finally, the subsequent crash identification date is pointed out using the red line (the 16th of January, 2008; just a few days after the generated signal.)
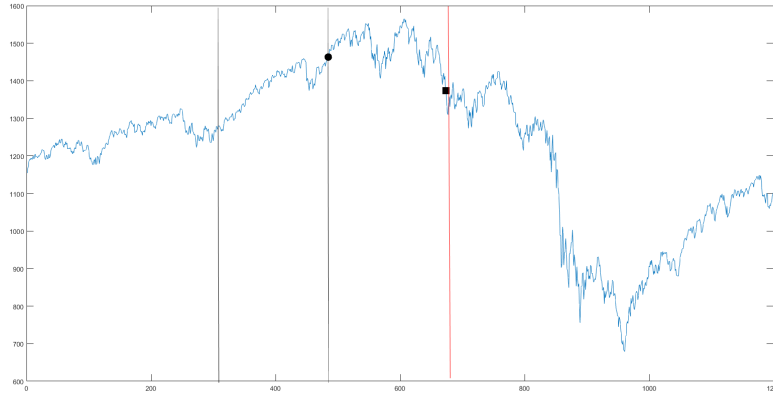
(a) DTDD 2 (T = 1000) on 15-05-2007

Figure 4: The rally of the S&P500 from 13-05-2005 till 19-02-2010 (blue line). The start- and changepoint of the DTDD2 model are respectively indicated with the black circle and square. The black lines indicate the initialisation period preceding the startpoint, while the red line indicates the crisis subsequent to the changepoint.

In order to construct the crash signals of the DTDD models several parameters needed to be estimated, the descriptive statistics of these estimated parameters are presented in the first two columns of Table 1. All parameters were estimated for every point in time starting from $t = 101$ to $t = 13,306 - T$, resulting in a total of over $11,000$ estimations for every parameter. The values for the estimated normal distributions are small, this is caused by the fact that the normal distribution is estimated over the log innovations of the indexes (X_t; equation 14). The values of μ_2 and σ_2 can be derived using the values of μ_1 and σ_1 together with the assumptions made in section 2.3.2.

Using these parameters, the crash signals and corresponding hitrates were constructed and are presented in Table 2 and 7 on the last three lines. The performance of the three different versions of the DTDD model differ quite a lot. First, the DTDD 1 model ($T = 750$) does outperform the uninformed benchmark significantly and beat the PE models on performance, but it does not beat any of the BSEYD models. The DTDD 2 model ($T = 1000$) also outperforms the benchmark, but fails to beat any of the economic models. Finally, the DTDD 3 ($T = 1500$) model is superior to all economic models as well as to the benchmark. The last 3 columns of Table 1 present the descriptive statistics of the estimated values for θ for all model horizons. One can see that the estimated values for θ increase when the model horizon of the DTDD model increases from 750 to 1500.

Table 1: Descriptive Statistics DTDD S&P 500

	μ_1	σ_1	θ (t=750)	θ (t=1000)	θ (t=1500)
Mean	0.0003	0.0088	248	313	422
Min	-0.0038	0.0031	4	4	4
Max	0.0033	0.0324	637	882	1375
Standard deviation	0.0009	0.0038	139	193	298

Concluding on the models' performances on the full sample of the S&P 500, we see the DTDD is the best performing model while all models outperform the benchmark of an uninformed crash prediction model. Furthermore, all models generate more (correct) signals than there are crashes, this is possible since a model can give two distinct signals who both are followed by the same crash within 500 trading days, hence both count as a correct predicted crash. So far, the DTDD model only was used as an exit-timing model rather than a crash prediction model (Shiryaev & Zhitlukhin (2012), Shiryaev et al. (2015)). However, the results in Table 7 confirm the suggestion of Lleo & Ziemba (2017) that the DTDD model might be a good crash prediction model concerning the S&P 500.

Table 11 (in Appendix) presents the p-values using the empirical distributions obtained from the Monte Carlo simulation over the full sample of the S&P 500. All models remain to outperform the benchmark significantly. Thus it can be stated that the χ^2 distribution provides a good approximation for the empirical distribution of the test statistic, even when the total number of signals is relatively small.

Table 2: Signals and Corresponding Hit Rates of all models when applied on the S&P 500

	Total signals	Correct signals	Hit rate	Incorrect signals
P/E 1	51	33	64.71%	18
log P/E 1	46	30	65.22%	16
P/E 2	45	31	68.89%	14
log P/E 2	40	27	67.50%	13
BSEYD 1	37	29	78.38%	8
log BSEYD 1	40	30	75.00%	10
BSEYD 2	36	28	77.78%	8
log BSEYD 2	38	28	73.68%	10
DTDD 1	59	43	72.88%	16
DTDD 2	68	44	64.71%	24
DTDD 3	76	60	78.95%	16

The results of all proposed models concerning the total amount of signals generated (column 1), the amount of correct signals (column 2) and the corresponding hit rate (column 3) when applied to the full sample of the S&P 500 (1-1-1964 till 31-12-2012).

3.4 Sub Period Study

To examine the performance of the different models over different periods of time, a sub period study will be performed. As mentioned in section 2.5, the sample is split at the same point in time as in the research of Lleo & Ziemba (2017), namely at the first end-of-the-year after the the 9th crash, January 1, 1982. Thus the research will examine the sub periods January 1, 1964 - December 31, 1981 and January 1, 1982 - December 31, 2012.

The performance of all models over these periods are presented in Table 3 and 9. The first thing which meets the eye is the enormous drop in performance of all PE models when comparing the first sample with the second. During the first sample, almost all generated signals are correct, while none of the hitrates over the second sample exceeds 60%. Table 9 presents a significantly better performance compared to the benchmark over the first sample for all PE models, but all of them fail to beat the benchmark model during the second sample.

The performance of the BSEYD models is slightly worse during the first sample compared to the PE models, while these models have a slight better performance during the second sample. However, none of the BSEYD models beat the uninformed benchmark over the second sample. Thus it can be stated

that the good performance of both the BSEYD model as the PE model over the full sample is caused solely by a very good performance over the first subsample.

The DTDD model again has a very inconsistent performance over the different model horizons. The DTDD 1 model has a performance comparable to the BSEYD models, thus not outperforming the benchmark at the second sample. The DTDD 2 model is the worst performing model of all models over the second sub sample, with an under 50% hit rate at the second sample. The DTDD 3 model is the most consistent model of all over both samples and the only model outperforming the uninformative benchmark at the second sample. Over the first sample, the DTDD 3 model performs the worst, however it's performance is still significantly better than the benchmark.

The subsample analysis points out an steep decrease for all models at the second sample compared to the first, one of the causes for this phenomenon is probably the low frequency of crashes over this period (only 9 crashes over a period of 30 years), while the models remain to generate signals with more or less the same frequency. This conjecture is further backed by the performance of the DTDD models, where the DTDD 3 model (lowest frequency) performs the worst at the first sample and the best at the second sample.

The empirical values for the sub samples of the S&P 500 are presented in Table 12. The values differ a bit compared to when a χ^2 distribution is assumed, however, the conclusions remain the same; only the DTDD 3 model is significantly outperforming the uninformative benchmark at the second sub sample. Hence the χ^2 distribution again seems a proper approximation of the empirical distribution.

Table 3: Signals and Hit Rates of all models when applied on subsamples of the S&P 500

	Sample 1			Sample 2		
	Signals	**Corr. Signals**	**ML**	**Signals**	**Corr. Signals**	**ML**
P/E 1	16	15	93.75%	35	18	51.43%
log P/E 1	15	14	93.33%	41	21	51.22%
P/E 2	15	14	93.33%	30	17	56.67%
log P/E 2	12	11	91.67%	28	16	57.14%
BSEYD 1	19	18	94.74%	18	11	61.11%
log BSEYD 1	20	18	90.00%	20	12	60.00%
BSEYD 2	18	17	94.44%	18	11	61.11%
log BSEYD 2	19	17	89.47%	19	11	57.89%
DTDD 1	24	22	91.67%	35	21	60.00%
DTDD 2	26	24	92.31%	42	20	47.62%
DTDD 3	38	33	86.84%	40	27	67.50%

The results of all proposed models concerning the total amount of signals generated (column 2 and 5), the amount of correct signals (column 3 and 6) and the corresponding hit rates (column 4 and 7) when applied to subsamples of the S&P 500 (1-1-1964 till 31-12-2012).

3.5 NASDAQ Composite Index Results

In order to gain extra insights on the model' performances all models are applied to the NASDAQ Composite index. The results are presented in Table 4 and 10. If the performances on the NASDAQ are compared to the performances on the S&P 500, a few thinks stand out. The PE models perform better compared to the S&P 500 sample. This might be caused by the increase in the frequency of crashes compared to the S&P 500[4].

The BSEYD models lose all their predictive advantage over the uninformative benchmark and perform very poor when applied to this index. This finding corresponds with the findings of Shiryaev et al. (2017) who also find a poor performance of the BSEYD model on the NASDAQ 100 (a subset of the NASDAQ Composite index).

The DTDD models also perform less good compared to the S&P 500 index, however the DTDD 3 model remains to outperform the uninformative benchmark. Hence, the DTDD 3 model is the only model who beats the benchmark over every (sub)sample of every index, making it the most consistent model of all.

---

[4]crashes per year: S&P 500: 18/51=0.35, NASDAQ: 12/30= 0.4

All these finding are confirmed by the empirical distributions obtained from the Monte Carlo simulations (Table 13). Thus, also for the test statistics of this index it can be stated that the χ^2 distribution with 1 degree of freedom is a proper approximation for the empirical distribution. Moreover, the descriptive statistics of the DTDD model are in line with our expectations (presented in table 8 in the Appendix).

Table 4: Signals and Hit Rates of all models when applied on the NASDAQ Composite Index

	Total signals	Correct signals	Hit Rate	Incorrect signals
PE1	30	24	80.00%	6
Log PE1	29	21	79.31%	8
PE2	33	26	78.79%	7
Log PE2	28	21	75.00%	7
BSEYD1	19	10	52.63%	9
Log BSEYD 1	22	11	50.00%	11
BSEYD2	20	11	55.00%	9
Log BSEYD 2	22	12	54.55%	10
DTDD 1	43	23	53.49%	20
DTDD 2	43	24	55.81%	19
DTDD 3	50	33	66.00%	17

The results of all proposed models concerning the total amount of signals generated (column 1), the amount of correct signals (column 2) and the corresponding hit rate (column 3) when applied to the sample of the NASDAQ Composite index (1-1-1984 till 31-12-2017).

4 Conclusion

In this research, the performances of several crash prediction models were investigated. When the models were applied to a sample of the S&P 500, all models outperformed the benchmark of an uninformative predictor. When looking at the relative performance of the models, the DTDD 3 model performed the best.

Hereafter the models were applied on subsamples of the S&P 500 index, the performance of all models appeared to be inconsistent over time. Over the first sample, all models performed very good with hitrates exceeding 90%. At the second subsample however, performances for all models dropped drastically leading to insignificant outperformance of the benchmark for all models except the DTDD 3 model. Looking at this two subsamples of the S&P 500 it can thus be concluded that the DTDD 3 model is the most consistent model of all applied models.

Finally, all models were applied to the NASDAQ Composite index. This index is known to be more volatile and has a higher crash density. The models performed very differently on this index, the BSEYD models lost all their significant advantage over the benchmark compared to the S&P 500 sample, while the PE models performed very good. Looking at the DTDD models, the DTDD 3 model again outperformed the benchmark, making it the only model outperforming the benchmark at every (sub)sample of every tested index.

Concluding on this research, the DTDD 3 model outperformed the economic models in performance, consistency and robustness. In certain cases, one of the economic models outperformed the DTDD 3 model, but none of the economic models was able to beat the benchmark at every subsample of the S&P 500. Hence it can be stated that, in general, the DTDD 3 model can be seen as superior over the economic models.

Besides the fact that the economic models are outperformed on most of the (sub)samples, this research pointed out another limitation of the economic models. Due to their dependency on explanatory variables such as (price) earnings data and treasury notes, these models become harder to apply compared to the DTDD model. During this research it was (hard, but) possible to find all necessary economic data, but when applying these models to less known indexes like the Dutch or Belgian Stock exchange, this process can become very time consuming or even impossible.

One major side note which need to be made on this conclusion is the lack of economic backing of the DTDD models. Although the model proves to deliver good results on all (sub)samples, it can not be linked to any asset pricing theory, making it a "Black box" and hard to understand for applicants.

4.1 Discussion

During this research, several limitations came up. Firstly, the initial paper on which this research is based (Lleo & Ziemba (2017)), made some assumptions and conclusions which can be questioned. The uninformed benchmark for crash prediction was set to $p = 1/2$, however when analyzing the number of crashes which occurred over the dataset in combination with the test horizon and sample length, the expected value for an uninformed benchmark comes down to $p = 0.70^5$, which would be harder to beat. Since this research is meant to be an extension of the research of Lleo & Ziemba (2017) and we mainly were interested in the relative performance of the models, we have refrained from adjusting this benchmark, but this is something to take in to account when interpreting the absolute performance of the models.

Moreover, during the signal construction of the economic models, it is assumed that all tails are distributed normal, despite failing the Jarque-Bera test for normality. Lleo & Ziemba (2017) justify this decision by performing an empirical analysis on tails of (only) one BSEYD measure and they point out the lack of importance of this subject since the models are compared against each other instead of constructing the absolute best crash prediction. However, if one would pursue the best comparison between models, the tail behaviour needs to be taken in to (further) account.

Furthermore, this research led to some openings for further research. The first subject which would be very interesting to investigate is the performance of all these (economic) models on indexes which could a lesser extend be linked to economic predictors, like the Dutch stock index (AEX). Since this index is depended on the economic climate in the Netherlands, Germany and the US, it is probably harder to predict by economic indicators like the treasury rate than the S&P 500 or NASDAQ, whom mainly are based on the US economy. The DTDD models will probably still be able to come up with good crash predictions, since those are independent of any economic indicators. Hence, applying all considered models to indexes like the AEX might underpin the conclusions of this paper.

Also, the models' performance was checked with respect to true positive and false positive signals. One could think of also analyzing the performance of the models true negative and false negative signals. This will change the scope from predicting economic down turns to (also) predicting periods of economic growth/stability and might lead to different conclusions.

Finally, several assumptions on the side of the DTDD model could be adjusted or further investigated. During this research, a Decreasing Absolute Risk Averse utility function and standard drift setting were assumed, since this was in line with the research of Shiryaev et al. (2017) and Shiryaev et al. (2015), who state that other utility or drift functions do not lead to drastic improvements of performance. Moreover, these setting simplified the formulas of the DTDD model drastically. However, when pursuing

---

[5]This value is calculated by dividing the total amount of crashes (18) by the total length of the sample (51 years) over the test horizon length (2 years); $18/(51/2) = 0.70$

the absolute best performance of the DTDD model one could consider an (overtime) analysis of all assumptions of the model, including the utility function, bull-bear-drift assumption and model horizon T.

References

Campbell, J. Y., & Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance*, *43*(3), 661–676.

Campbell, J. Y., & Shiller, R. J. (1989). The dividend ratio model and small sample bias: a monte carlo study. *Economics letters*, *29*(4), 325–331.

Campbell, J. Y., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, *24*(2), 11–26.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, *38*(1), 34–105.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Gordon, M. J. (1959). Dividends, earnings, and stock prices. *The review of economics and statistics*, 99–105.

Hertzberg, M. P., & Beckman, B. A. (1989). Business cycle indicators: Revised composite indexes. *Business Conditions Digest*, *29*(1).

Klein, P. A., & Moore, G. H. (1983). The leading indicator approach to economic forecasting—retrospect and prospect. *Journal of forecasting*, *2*(2), 119–135.

Koivu, M., Pennanen, T., & Ziemba, W. T. (2005). Cointegration analysis of the fed model. *Finance Research Letters*, *2*(4), 248–259.

Lleo, S., & Ziemba, W. T. (2013). Stock market crashes in 2007–2009: were we able to predict them? In *Managing and measuring risk: Emerging global standards and regulations after the financial crisis* (pp. 457–499). World Scientific.

Lleo, S., & Ziemba, W. T. (2017). Does the bond-stock earnings yield differential model predict equity market corrections better than high p/e models? *Financial Markets, Institutions & Instruments*, *26*(2), 61–123.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383–417.

Shiller, R. J. (1996). Price-earnings ratios as forecasters of returns: The stock market outlook in 1996. *Unpublished manuscript, Yale University*.

Shiryaev, A., & Zhitlukhin, M. (2012). Optimal stopping problems for a brownian motion with a disorder on a finite interval. *arXiv preprint arXiv:1212.3709*.

Shiryaev, A., Zhitlukhin, M., & Ziemba, W. T. (2017). When to sell apple and the nasdaq? trading bubbles with a stochastic disorder model. In *Great investment ideas* (pp. 231–248). World Scientific.

Shiryaev, A., Zhitlukhin, M. V., & Ziemba, W. T. (2015). Land and stock bubbles, crashes and exit strategies in japan circa 1990 and in 2013. *Quantitative Finance*, *15*(9), 1449–1469.

Yardeni, E. (1997). Fed's stock market model finds overvaluation. *Topical study*, *38*.

Zhitlukhin, M. (2014). *Stochastic dynamics of financial markets* (Unpublished doctoral dissertation). University of Manchester.

Zhitlukhin, M., & Shiryaev, A. (2013). Bayesian disorder problems on filtered probability spaces. *Theory of Probability & Its Applications*, *57*(3), 497–511.

Ziemba, W. T., Berge, K., & Consigli, G. (2008). The predictive ability of the bond-stock earnings yield differential in relation to the equity risk premium. *Journal of Portfolio Management*(34, 3), 63–80.

Ziemba, W. T., & Schwartz, S. L. (1992). Invest japan. *Probus, Chicago, IL*, 301–306.

5 Appendix

A Derivation Campbell-Shiller Model from Gordon Growth model

Under the assumptions made in Gordon (1959), the price of a stock at time t equals:

$$P_t^{end} = \frac{D_{t+1} + P_{t+1}^{end}}{1+k} \tag{34}$$

This can be re-written to (g represents growthrate, d the corresponding discountrate and ρ the yield):

$$\frac{P_{t+1}^{end} + D_t}{P_t^{end}} + \frac{(D_{t+1} - Dt)}{P_t^{end}} = 1 + k \tag{35}$$

$$\frac{P_{t+1}^{end} + D_t}{P_t^{end}} = 1 + k - \frac{(D_{t+1} - Dt)}{P_t^{end}} = 1 + k - gd\rho_t \tag{36}$$

Taking log results in:

$$h_{i,t+1}^{beg} = log\left(\frac{P_{t+1}^{end} + D_t}{P_t^{end}}\right) = log((1+k)(1 - \frac{gd}{k}\rho_t)) = log(1+k) + log(1 - \frac{gd}{k}\rho_t) \tag{37}$$

The right hand side of this expression can be liniarised by taking the first order Taylor expansion around $\rho_t = \overline{\rho}$, with $\overline{\rho}$ equals the average long-term earnings yield.

$$h_{1t}^{end} = a_0 - a_1\rho_t^{end} + h.o.t. \approx a_0 - a_1\rho_t^{end}, \tag{38}$$

where $a_0 \equiv ln(1+k) + ln\left(1 - \frac{gd}{1+k}\overline{\rho}\right) + \frac{gd}{1+k-gd\overline{\rho}}$ and $a_1 \equiv \frac{gd}{1+k-gd\overline{\rho}}$

Now, if assumed $P_t^{end} = P_{t+1}^{beg}$ on the left hand side of equation 37, it follows $\rho_t^{end} = \rho_{t+1}^{beg}$, which leads to:

$$h_{i,t}^{beg} = log\left(\frac{P_t^{end} + D_t}{P_{t-1}^{end}}\right) = h_{1(t-1)}^{end} \tag{39}$$

This makes both models equal, despite a small change in notation. These changes are caused by two differences between equation 38 and the Campbell-Shiller regression. The first is that the Gordon Growth Model uses Earings/Price ratio while the PE model uses the inversion of this ratio. Secondly, the explanatory variable in equation 38 is not a logaritmic variable, while the explanatory variable in the Campbell-Shiller model is.

The first difference is not any problem if we consider the properties of a (natural) logarithm:

$$log(h_{i,t}) = a + b \quad log\left(\frac{P_t}{\overline{E_{t,-30}}}\right) + \epsilon = a - b \quad log(\frac{\overline{E_{t,-30}}}{P_t}) + \epsilon, \tag{40}$$

while the second difference is only of relative importance.

B Derivation BSEYD Model from Gordon Growth Model

If equation 34 is inverted, the Gordon 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:

$$P_t = \frac{D_{t+1}}{k_t - g} \tag{41}$$

If this situation is viewed from the perspective of being at time t, the equation becomes:

$$P_t = \frac{E_t d(1+g)}{k_t - g} \tag{42}$$

Which can get re-written to:

$$k = g \frac{d(1+g)}{\gamma_t}, \tag{43}$$

with γ_t equaling P_t/E_t.

The above described equation leads to a 'Gordon' definition of the earnings yield ρ_t:

$$\rho_t = \frac{E_t}{P_t} = \frac{k-g)}{d(1+g)} \tag{44}$$

This finding, together with the definition of the cost of equity k_t ($k_t = r_t + f_t$ where r_t is the yield on a government bond and f_t is the equity risk premium) leads to the BSEYD model in terms of the Gordon Growth model:

$$BSEYD(t) = r(t) - \frac{E(t)}{P(t)} = \frac{1}{d(1+g)}(r_t[d(1+g) - 1] - f_t + g) \tag{45}$$

C Derivation Maximum Likelihood Estimator

The likelihood of this binary problem is defined as follows:

$$L(p|X_{i,H}) \equiv \prod_{t=1}^{N_i} p^{X_{i,t,H}} (1-p)^{(1-X_{i,t,H})}, \tag{46}$$

Equation 46 can be seen as a Bernoulli trial with chance of succes $p_{i,H} = P(C_{i,H} = 1|SIG_{i,t} = 1))$. Taking the log of equation 46 leads to this log likelihood $\mathcal{L}$:

$$\mathcal{L}(p|X_{i,H}) \equiv log(L(p|X_{i,H})) = \sum_{t=1}^{N_i} X_{i,t,H} \quad log(p) + \left(N_i - \sum_{t=1}^{N_i} X_{i,t,H} \right) \quad log(1-p) \tag{47}$$

When maximising this fuction, the Maximium Likelikhood Estimator (MLE) for this problem is obtained:

$$\hat{p_{i,H}} \equiv \frac{\sum_{t=1}^{N_i} X_{i,t,H}}{N_i} = HR_{i,H} \tag{48}$$

Which in the case of this paper represents the estimated proportion of correct predicted equity market crashes by the considered model; the Hitrate, introduced in equation 30).

D Intuition of the DTDD Model

Besides this proof, the intuition behind equation 22 can be explained by expressing the optimal stopping time τ^* in terms of the conditional probability $\pi_t = P(\theta \leq t | \mathcal{F}_t)$ (the change that the changepoint has occurred at time t conditional on the past t point is time). In order to do this, π_t is expressed using Bayes formula:

$$\pi_t = P(\theta \leq t | \mathcal{F}_t) = \frac{\sum_{u=1}^{t} \frac{dP_t^u}{dP_t^\infty} p_u}{\sum_{u=1}^{T+1} \frac{dP_t^u}{dP_t^\infty} p_u} \tag{49}$$

Using equation 17, this leads to the following expression for ψ_t, which implies another definition of π_t:

$$\psi_t = \frac{\pi_t}{1 - \pi_t}(1 - G(t)), \quad \pi_t = \frac{\psi_t}{\psi_t + 1 - G(t)} \tag{50}$$

Now, its possible to substitute these results in to equation 22 and obtain this equation in terms of π_t:

$$\tau_\alpha^* = inf\{0 \leq t \leq T : \pi_t \geq \tilde{b}_\alpha^*\} = \frac{b_\alpha^*}{b_\alpha^* + 1 - G(t)} \tag{51}$$

This representation shows the idea behind equation 22; the model will give a crash signal as soon as it is sufficiently confident that the changepoint in included in the datapoints $[0, \ldots, t]$.

E   Overview of Used Models

Concluding on model construction sections, this research will perform an (sub sample) analysis using the following 11 models:

1. P/E 1: P/E model using the standard confidence interval to construct the signals

2. Log P/E 1: Log P/E model using the standard confidence interval to construct the signals

3. P/E 2: P/E model using the Cantelli's confidence interval to construct the signals

4. Log P/E 2: Log P/E model using the Cantelli's confidence interval to construct the signals

5. BSEYD 1: BSEYD model using the Standard confidence interval to construct the signals

6. Log BSEYD 1: Log BSEYD model using the Standard confidence interval to construct the signals

7. BSEYD 2: BSEYD model using the Cantelli's confidence interval to construct the signals

8. Log BSEYD 2: Log BSEYD model using the Cantelli's confidence interval to construct the signals

9. DTDD 1: DTDD model using a modelhorizon of 750 tradingdays

10. DTDD 2: DTDD model using a modelhorizon of 1000 tradingdays

11. DTDD 3: DTDD model using a modelhorizon of 1500 tradingdays

Table 5: Detected Crashes at the S&P 500 index

	Crash Identification date	Peak Date	Peak index level	Trough date	Trough index level	Total decline loss	Length of crash
1	26/7/1966	9/2/1966	94.06	10/7/1966	73.2	22.2%	240
2	20/06/1969	29/11/1968	108.37	26/05/1970	69.29	36.1%	543
3	5/8/1971	4/28/1971	104.77	23/11/1971	90.16	13.9%	209
4	30/04/1973	11/1/1973	120.24	25/04/1974	89.57	25.5%	469
5	20/08/1975	15/07/1975	95.61	16/09/1975	82.09	14.1%	63
6	26/05/1977	21/09/1976	107.83	6/3/1978	86.9	19.4%	531
7	10/27/1978	12/9/1978	106.99	14/11/1978	92.49	13.6%	63
8	14/03/1980	13/02/1980	118.44	27/03/1980	98.22	17.1%	43
9	25/08/1981	28/11/1980	140.52	12/8/1982	102.42	27.1%	622
10	22/02/1984	10/10/1983	172.65	24/07/1984	147.82	14.4%	288
11	16/10/1987	25/08/1987	336.77	4/12/1987	223.92	33.5%	101
12	20/08/1990	16/07/1990	368.95	11/10/1990	295.46	19.9%	87
13	28/08/1998	17/07/1998	1186.75	31/08/1998	957.28	19.3%	45
14	18/10/1999	16/07/1999	1418.78	18/10/1999	1254.13	11.6%	94
15	12/10/2000	24/03/2000	1527.46	21/09/2001	965.8	36.8%	546
16	16/01/2008	09/10/2007	1565.15	20/11/2008	752.44	51.9%	408
17	21/05/2010	23/04/2010	1217.28	2/7/2010	1022.58	16.0%	70
18	5/8/2011	29/04/2011	1363.61	3/10/2011	1099.23	19.4%	157

All detected crashes at the S&P 500 index over the period 1-1-1964 till 31-12-2012. The first column presents the date the crash was identified, the second and third column present the peak of the crash, while the fourth and fifth column present the through of the crash. The sixth column presents the loss from the peak to the through in percentages, while the last column presents the length of the crash.

Table 6: Detected Crashes at the NASDAQ Composite index

	Crash Identification date	Peak Date	Peak index level	Trough date	Trough index level	Total decline loss	Length of crash
1	9/9/1986	7/4/1986	411.16	1/15/1987	343.67	16.4%	86
2	10/16/1987	8/27/1987	455.26	10/13/1988	291.88	35.9%	250
3	1/22/1990	10/10/1989	485.73	3/26/1990	410.72	15.4%	38
4	8/1/1990	10/10/1989	485.73	4/22/1991	325.44	33.0%	133
5	6/9/1992	2/13/1992	644.92	8/20/1992	547.84	15.1%	38
6	6/20/1994	3/21/1994	803.93	8/8/1994	693.79	13.7%	30
7	8/21/1998	7/21/1998	2014.25	12/23/1998	1419.12	29.5%	53
8	4/3/2000	3/13/2000	5048.62	11/17/2005	1114.11	77.9%	810
9	7/12/2004	1/27/2004	2153.83	11/4/2004	1752.49	18.6%	59
10	1/4/2008	11/1/2007	2859.12	11/12/2010	1268.64	55.6%	438
11	8/4/2011	5/2/2011	2873.54	10/3/2011	2341.84	18.5%	30
12	1/7/2016	7/21/2015	5218.86	4/7/2016	4266.84	18.2%	39

All detected crashes at the S&P 500 index over the period 1-1-1964 till 31-12-2012. The first column presents the date the crash was identified, the second and third column present the peak of the crash, while the fourth and fifth column present the through of the crash. The sixth column presents the loss from the peak to the through in percentages, while the last column presents the length of the crash.

F   Plots of Economic Models
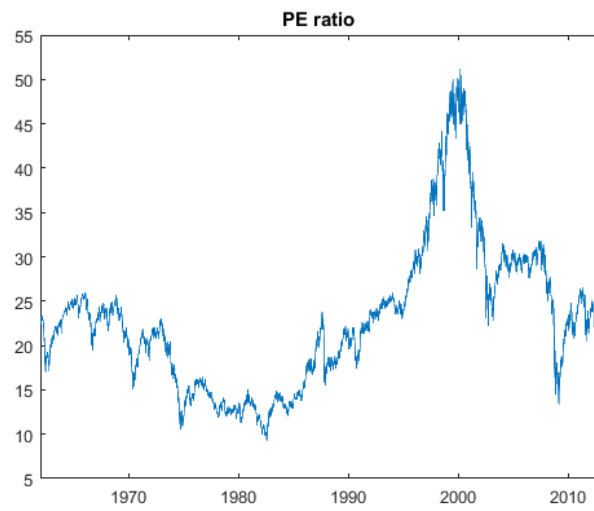
Figure 5: Plot of the 10-year Cycally Adjusted Price Earnings ratio over the period 01-01-1962 till 31-12-2012. The obtained plot exactly matches the plot presented in the paper of Lleo & Ziemba (2017).
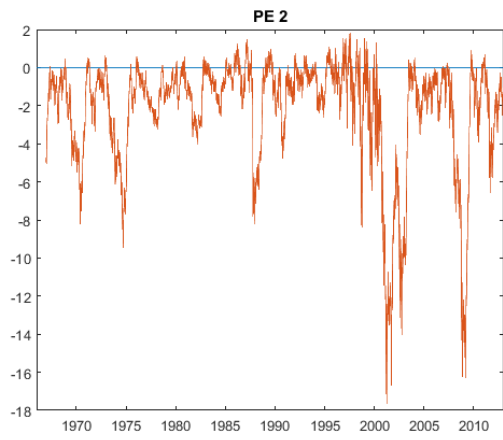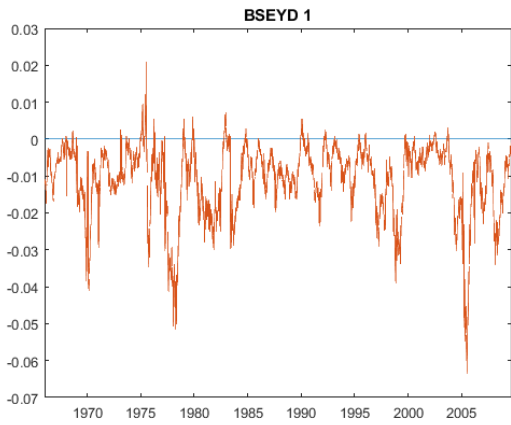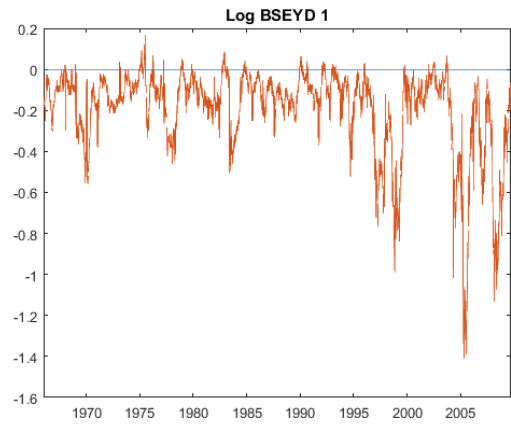
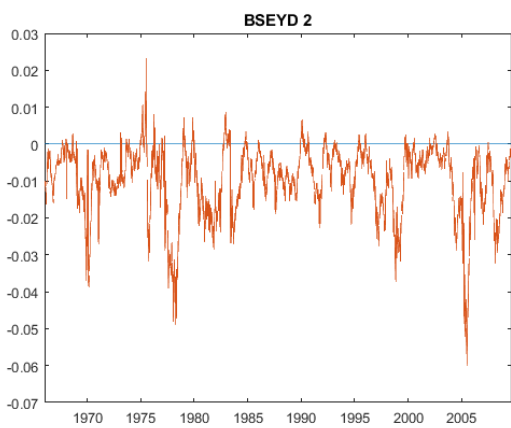(a) PE 2 model signals and Threshold
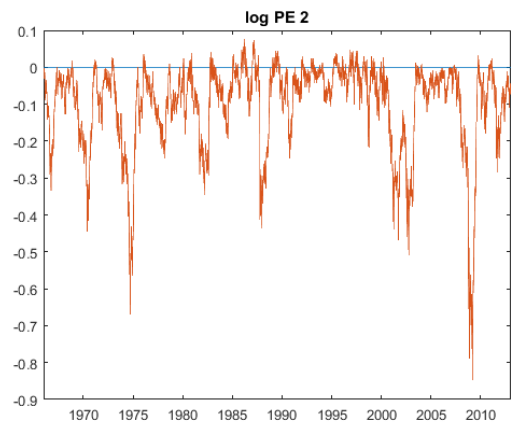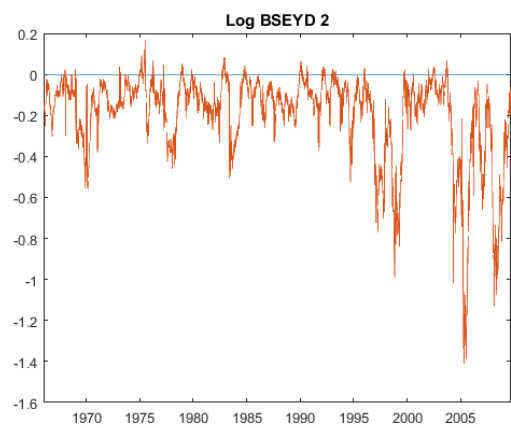
(b) Log PE 2 model signals and Threshold

(c) BSEYD 1 model signals and Threshold

(d) Log BSEYD 1 model signals and Threshold

(e) BSEYD 2 model signals and Threshold

(f) Log BSEYD 2 model signals and Threshold

G Likelihood Ratio Test Results

Table 7: Hitrates and Teststatistics for all models applied to the full sample of the S&P 500

	Total signals	Correct signals	ML estimate	Likelihood	Likelihood ratio	Test statistic	P-value
P/E 1	51	33	64.71%	4.16664E-15	0.10658	4.48*	3.43%
log P/E 1	46	30	65.22%	1.23793E-13	0.11480	4.33*	3.75%
P/E 2	45	31	68.89%	7.64613E-13	0.03717	6.58*	1.03%
log P/E 2	40	27	67.50%	1.11093E-11	0.08187	5.01*	2.53%
BSEYD 1	37	29	78.38%	4.0813E-09	0.00178	12.66***	0.04%
log BSEYD 1	40	30	75.00%	1.70309E-10	0.00534	10.46***	0.12%
BSEYD 2	36	28	77.78%	5.22703E-09	0.00278	11.77***	0.06%
log BSEYD 2	38	28	73.68%	3.08069E-10	0.01181	8.88***	0.29%
DTDD 1	59	43	72.88%	1.05875E-15	0.00164	12.83***	0.03%
DTDD 2	68	44	64.71%	6.70473E-20	0.05053	5.97*	1.45%
DTDD 3	76	60	78.95%	1.03077E-17	0.00000	27.13***	0.00%

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Likelihood ratio test results. Column ML presents the Maximum likelihood estimation, which is this case equals the hitrate. The column likelihood shows the corresponding maximum likelihood when using the parameters of columns 1 to 3. Finally, columns 6 and 7 show the test- and p-value of the obtained likelihood ratio from an Chi-squared distribution with 1 degree of freedom.

H Descriptive Statistics of the Estimated Parameters of the DTDD Model.

Table 8: Descriptive Statistics DTDD Nasdaq

	μ_1	σ_1	θ (t=750)	θ (t=1000)	θ (t=1500)
Mean	0.0003	0.0121	198	241	306
Min	-0.0064	0.0042	3	3	3
Max	0.0059	0.0378	609	828	1275
Standard deviation	0.0015	0.0073	134	179	266

Table 9: Hitrates and Teststatistics for all models applied to subsamples of the S&P 500

| | Sample 1 | | | | | | Sample 2 | | | | | |
	Signals	Corr. Signals	ML est.	LH	LHR	p-value	Signals	Corr. Signals	ML est.	LH	LHR	p-value
P/E 1	16	15	93.75%	0.024	0.00064	0.01%***	35	18	51.43%	2.95226E-11	0.98581	86.58%
log P/E 1	15	14	93.33%	0.025	0.00120	0.02%***	41	21	51.22%	4.60328E-13	0.98788	87.59%
P/E 2	15	14	93.33%	0.025	0.00120	0.02%***	30	17	56.67%	1.21691E-09	0.76532	46.45%
log P/E 2	12	11	91.67%	0.032	0.00763	0.18%***	28	16	57.14%	4.96215E-09	0.75074	44.89%
BSEYD 1	19	18	94.74%	0.020	0.00010	0.00%***	18	11	61.11%	5.97174E-06	0.63879	34.38%
log BSEYD 1	20	18	90.00%	0.002	0.00064	0.01%***	20	12	60.00%	1.42658E-06	0.66851	36.95%
BSEYD 2	18	17	94.44%	0.021	0.00018	0.00%***	18	11	61.11%	5.97174E-06	0.63879	34.38%
log BSEYD 2	19	17	89.47%	0.002	0.00114	0.02%***	19	11	57.89%	2.41947E-06	0.78833	49.04%
DTDD 1	24	22	91.67%	0.001	0.00006	0.00%***	35	21	60.00%	5.88866E-11	0.49424	23.51%
DTDD 2	26	24	92.31%	0.001	0.00002	0.00%***	42	20	47.62%	2.38467E-13	0.95348	75.76%
DTDD 3	38	33	86.84%	0.000	0.00001	0.00%***	40	27	67.50%	1.11093E-11	0.08187	2.53%***

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Likelihood ratio test results of the models applied to subsamples of the S&P 500. Column ML est. presents the Maximum likelihood estimation, which is this case equals the hitrate. The column LH shows the corresponding maximum likelihood while LHR corresponds to Likelihood ratio. Finally, columns 7 and 13 show the test- and p-value of the obtained likelihood ratio from an Chi-squared distribution with 1 degree of freedom.

Table 10: Hitrates and teststatistics for all models applied to the NASDAQ Composite Index

	Total signals	Correct signals	ML estimate	Likelihood	Likelihood ratio	Test statistic	P-value
P/E 1	30	24	80.00%	3.02231E-07	0.00308	11.56***	0.07%
log P/E 1	29	21	72.41%	3.81754E-08	0.04879	6.04*	1.40%
P/E 2	33	26	78.79%	3.92675E-08	0.00296	11.64***	0.06%
log P/E 2	28	21	75.00%	1.45167E-07	0.02566	7.33**	0.68%
BSEYD 1	19	10	52.63%	1.95823E-06	0.97402	0.05	81.85%
log BSEYD 1	22	11	50.00%	2.38419E-07	1.00000	0.00	100.00%
BSEYD 2	20	11	55.00%	1.05415E-06	0.90469	0.20	65.45%
log BSEYD 2	22	12	54.55%	2.61142E-07	0.91299	0.18	66.96%
DTDD 1	43	23	53.49%	1.2624E-13	0.90056	0.21	64.72%
DTDD 2	43	24	55.81%	1.5214E-13	0.74725	0.58	44.53%
DTDD 3	50	33	66.00%	1.20252E-14	0.07386	5.21*	2.24%

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Likelihood ratio test results. Column ML presents the Maximum likelihood estimation, which is this case equals the hitrate. The column likelihood shows the corresponding maximum likelihood when using the parameters of columns 1 to 3. Finally, columns 6 and 7 show the test- and p-value of the obtained likelihood ratio from an Chi-squared distribution with 1 degree of freedom.

I Monte Carlo Study Results

Table 11: Monte Carlo study S&P 500 Likelihood Ratio Test (Full Sample)

	Total number of signals	ML Estimator	CV 95% Emp.	CV 99% Emp.	CV 99.5% Emp.	Test statistic	P-value
P/E 1	51	64.71%	4.4777	7.2520	7.2520	4.48*	2.73%
log P/E 1	46	65.22%	4.3292	7.2352	7.2352	4.33*	2.60%
P/E 2	45	68.89%	3.8096	6.5844	8.2794	6.58*	1.47%
log P/E 2	40	67.50%	3.6560	6.5826	8.3983	5.01*	2.07%
BSEYD 1	37	78.38%	3.3202	6.2597	8.1119	12.66***	0.06%
log BSEYD 1	40	75.00%	3.6560	6.5826	8.3983	10.46***	0.17%
BSEYD 2	36	77.78%	4.0776	7.3660	7.3660	11.77***	0.00%
log BSEYD 2	38	73.68%	3.8551	6.9515	6.9515	8.88***	0.02%
DTDD 1	59	72.88%	3.8557	6.2290	7.6410	12.83***	0.03%
DTDD 2	68	64.71%	3.8002	5.9702	7.2473	5.97**	0.96%
DTDD 3	76	78.95%	4.3039	6.4605	7.7102	27.13***	0.00%

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Empirical distributions obtained from a Monte Carlo simulation using the amount of generated signals as input (column 2). Column ML presents the Maximum likelihood estimation. Columns 4 to 6 present the critical value of the obtained empirical distribution, column 8 presents the empirical p-value of the test statistic.

Table 12: Monte Carlo study S&P 500 Likelihood Ratio Test (Subsamples)

	CV 95%	CV 99%	CV 99.5%	Test stat.	P-val	CV 95%	CV 99%	CV 99.5%	Test stat.	P-val
P/E 1	4.1860	6.7382	6.7382	14.70***	0.00%	3.5164	6.6414	8.6170	0.03	74.39%
log P/E 1	3.3970	5.7823	9.0142	13.45***	0.00%	4.1940	7.2661	7.2661	0.02	52.56%
P/E 2	3.3970	5.7823	9.0142	13.45***	0.00%	3.3980	6.7939	6.7939	0.53	36.16%
log P/E 2	3.1395	5.8221	9.7515	9.75***	0.04%	3.6515	7.3255	7.3255	0.57	35.22%
BSEYD 1	4.4389	6.7828	6.7828	18.50***	0.00%	3.6830	5.8839	8.7331	0.90	23.52%
log BSEYD 1	3.2913	7.7098	7.7098	14.72***	0.03%	3.2913	7.7098	7.7098	0.81	26.38%
BSEYD 2	3.6830	5.8839	8.7331	17.23***	0.01%	3.6830	5.8839	8.7331	0.90	23.41%
log BSEYD 2	4.4389	6.7828	6.7828	13.55***	0.00%	4.4389	6.7828	6.7828	0.48	35.61%
DTDD 1	4.2965	6.2790	8.7075	19.50***	0.00%	3.5164	6.6414	8.6170	1.41	17.21%
DTDD 2	3.9471	5.7541	7.9530	21.94***	0.00%	3.4768	6.2520	7.9697	0.10	87.65%
DTDD 3	3.8551	6.9515	8.8778	23.09***	0.00%	3.6560	6.5826	8.3983	5.01*	3.78%

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Empirical distributions obtained from a Monte Carlo simulation using the amount of generated signals as input. Columns 2 to 5 and 7 to 9 present the critical value of the obtained empirical distribution, column 8 and 11 presents the empirical p-value of the test statistic.

Table 13: Monte Carlo study NASDAQ Likelihood Ratio Test

	Signals	ML Estimator	CV 95% Emp.	CV 99% Emp.	CV 99.5% Emp.	Test statistic	P-value
P/E 1	30	80.00%	3.3980	6.7939	6.7939	11.5***6	0.02%
log P/E 1	29	72.41%	4.2787	6.0404	8.1480	6.04**	0.90%
P/E 2	33	78.79%	3.7378	7.0748	7.0748	11.64***	0.01%
log P/E 2	28	75.00%	3.6515	7.3255	7.3255	7.33***	0.04%
BSEYD 1	19	52.63%	4.4389	6.7828	6.7828	0.05	65.00%
log BSEYD 1	22	50.00%	4.7166	6.9162	6.9162	0.00	100.00%
BSEYD 2	20	55.00%	3.2913	7.7098	7.7098	0.20	52.73%
log BSEYD 2	22	54.55%	4.7166	6.9162	6.9162	0.18	81.25%
DTDD 1	43	53.49%	3.9924	6.9080	6.9080	0.21	54.16%
DTDD 2	43	55.81%	3.9924	6.9080	6.9080	0.58	35.64%
DTDD 3	50	66.00%	3.9729	6.6278	8.2283	5.21*	1.79%

* significant at the 95% level. ** significant at the 99% level. *** significant at the 99.5% level.

The results of all proposed models concerning the Empirical distributions obtained from a Monte Carlo simulation using the amount of generated signals as input (column 2). Column ML presents the Maximum likelihood estimation. Columns 4 to 6 present the critical value of the obtained empirical distribution, column 8 presents the empirical p-value of the test statistic.
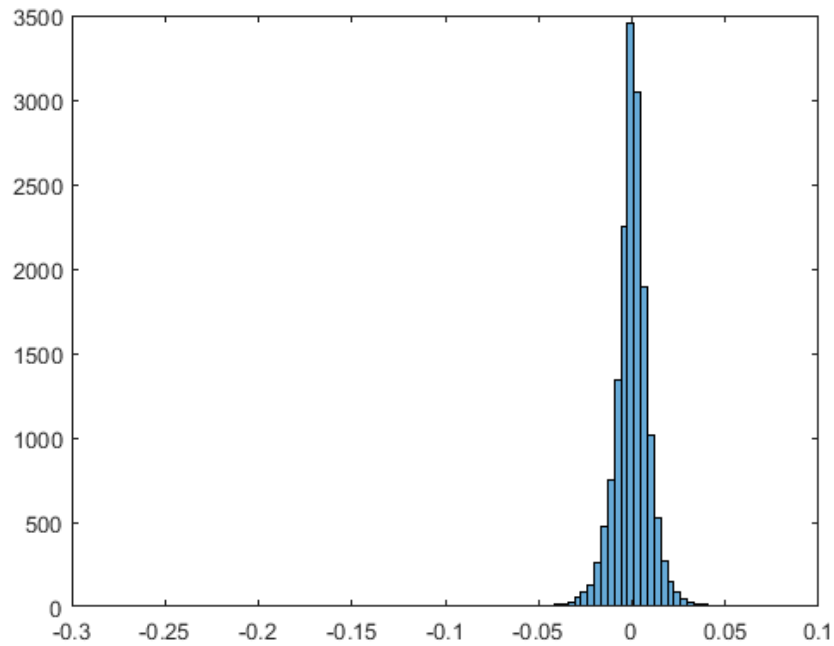
J   Tail Analysis PE model

Figure 7: Histogram of the PE model innovations. The entire dataset fails the JB-test (p=0.000) for normality. At this point the suggestion of Lleo & Ziemba (2017) is followed to zoom in to the tail and empirically check whether it is plausible to assume a Gaussian distribution for the tail behavior (see figure J).
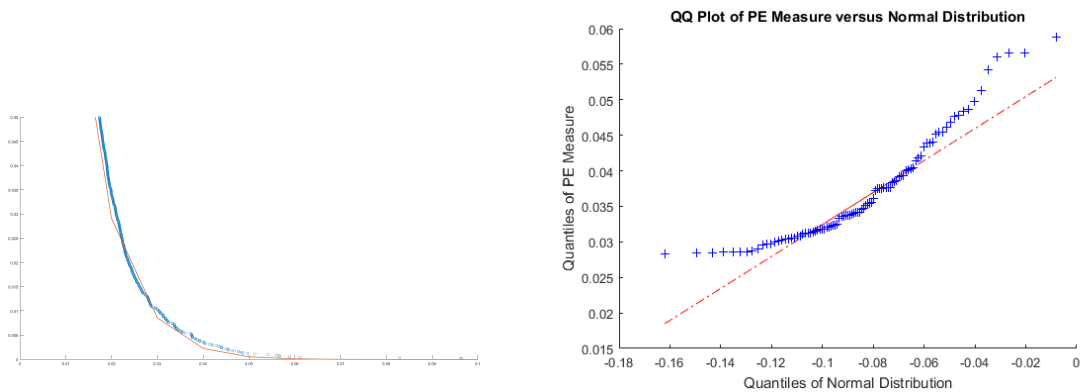

Figure 8: Plot of the Tail of the histogram of the PE model innovations. One can see it is reasonable to assume a Gaussian distribution (parameters $\mu = -0.085$ and $\sigma = 0.03$ used in the example).