A Machine-Learning-Based Early Warning System Boosted by Topological Data Analysis

Devraj Basu[a] and Tieqiang Li[b]

Abstract

We propose a novel early warning system for detecting financial market crashes that utilises the information extracted from the shape of financial market movement. Our system incorporates Topological Data Analysis (TDA), a new set of data analytics techniques specialised in profiling the shape of data, into a more traditional machine learning framework. Incorporating TDA leads to substantial improvements in timely detecting the onset of a sharp market decline. Our framework is both able to generate new features and also unlock more value from existing factors. Our results illustrate the importance of understanding the shape of financial market data and suggest that incorporating TDA into a machine learning framework could be beneficial in a number of financial market settings.

Keywords: Financial crashes, equity markets, machine learning, Topological Data Analysis.

1. Introduction

Developing early warning systems for financial crises has been a topic of considerable interest, particularly since the 2008-2009 financial crisis. Detecting financial crises is important not only for financial market participants but also for policy makers and regulators and hence there is broad interest in such systems. Developing such a system is very challenging, as predicting changes, particularly extreme changes, in financial markets is very difficult. One method for developing an early warning system is to develop a stress indicator which measures the extent of stress in a financial market, based on leading indicators. An ideal stress indicator should be able to both explain the incidence of a crisis based on the identified leading indicators as well as predict the onset of the crisis (Frankel and Saravelos 2012). Such indicators may broadly be divided into three categories. The first is the regression method indicators which tests the statistical significance of different leading indicators in determining the incidence or probability of a financial crisis (Eichengreen, Rose and Wypslosz 1995, Frankel and Rose 1996). The second is the signalling approach in which exceedance of threshold values of predetermined indicators is the crisis indicator (Kaminsky, Lizondo and Reinhart 1996, Gresnig et al 2015). The third and most recent method is the data-driven approach examplified in tools and techniques from machine learning and artificial intelligence, methods to which either determine leading indicator crisis thresholds (Frankel and Wei 2004, Bussiere and Fratszcher 2006) or select the most appropriate indicators (Lin et al 2008, Cevim et al 2014, Chatzis et al 2018).

A financial crisis may be viewed as an abrupt shift in the state of a financial market regarded as a complex dynamical system. Detecting such shifts is particularly difficult due to the inherent noisiness of the financial data as well as potential non-stationarity, which traditional methods of statistical analysis are not always well placed to handle. Novel approaches in data science such as unsupervised feature learning and deep learning (Langkvist, Karlson and Lutfi 2014) have recently been applied in the time series context in order to reduce dimensionality and extract relevant information. One such approach is Topological Data Analysis (TDA), which refers to a combination of statistical,

---

[a] Corresponding author. Strathclyde Business School, University of Strathclyde. Email: devraj.basu@strath.ac.uk

[b] FountainCap Research & Investment. Email: tan.li2006@gmail.com

computational, and topological methods that finds shape-like persistent structures in data. Topological Data Analysis (TDA) has been applied in the context of time series and systems analysis (Gholizadeh and Zadrozny 2018) and in particular in the context of financial market crashes (Gidea and Katz 2018). A key property of Topological Data Analysis is its ability to extract stable features from noisy data and it provides a new econometric tool which appears to complement existing statistical techniques in financial crash detection. A key insight from Gidea and Katz (2018) is the notion that Topological Data Analysis is able to detect the "landscape" of crashes in that the shapes of the financial time series seems to depend on the state of the market, whose structural changes Topological Data Analysis is able to capture (this broad theme is also explored in Huang et al 2017 using manifold learning techniques). However, in order to translate the conceptual benefits of TDA as outlined in Gidea and Katz, a more systematic and purpose-built implementation is required.

In this paper we develop an early warning system for financial crash detection by incorporating Topological Data Analysis into a machine learning framework. This allows us to build a numerical financial stress indicator and also to "unlock" features which conventional models may supress due to either data or human biases, help detect a coming financial crisis. Machine learning has become an increasingly popular framework when it comes to solve data-intensive problems in the financial services sector. The range covers from fraud detections (Bolton and Hand 2002, Cecchini et al 2010, Kundu et al 2008) in the retail space to financial machine learning for trading and quantitative strategies in the institutional space (Lopez de Prado 2018). The power of the machine learning approach, as opposed to the rule-based counterpart, rests in its data-driven nature, objectivity and effectiveness in evaluating performance. In particular, when it comes to deciding whether a given feature or market driver is significant in the system's decision making process, a machine learning approach can score the significance purely based on statistics of the training outcome, hence our approach also illustrate how a machine learning framework objectively evaluate feature innovation like TDA. In addition, the combination of TDA and traditional market descriptors such as momentums and volatilities is naturally fitted into a machine learning framework: TDA summaries are feature-engineered into "bolt-on" descriptors next to the traditional ones, making the whole training testing datasets and learning process identical to the ones without TDA for any chosen machine learning model. We then can not only effectively develop a machine-learning-based early warning system, but also by repeating the same process again bolting on the TDA descriptors, deliver rigorous comparison analysis and present how the shape of data can be value-adding to solve machine-learning problems.

In this introduction section, we highlight four areas of TDA's contribution below.

Firstly, as the two charts illustrated below and in Figure 1 at the end of the article (combining the two indicators in one chart), over the August to December 2008 period the early warning indicators, which is the aggregate probability of a market crash in the next two weeks, look very different with and without TDA features, painted in red and blue coloured bars, respectively. The indicator without TDA features peaks at the end of November 2008 well after the major decline in the S&P and is consistently elevated only from the beginning of November. When TDA features are incorporated the indicator shows a major spike in mid-September about two weeks before a 25% decline in the S&P and continues to remain elevated in the month of October between two and four weeks before a 17% decline in November. The TDA-augmented indicator again shows a spike at the end of November together with the non TDA-augmented indicator, which does not presage a major decline, However the TDA-augmented indicator drops much more sharply than the non-TDA-augmented indicator. The market decline in 1998 also provides a good test case and here a similar pattern is observed. The S&P peaked in the middle of July 1998, and the TDA-augmented crisis indicator correctly forecasts the drop with 3 consecutive spikes of crisis probability before the market declined around 20%. The crisis indicator without TDA peaked near the end of the decline and remained at low levels prior to the decline. Overall these finding indicate that the TDA-augmented indicator is more robust than

indicator built without TDA. This finding is consistent over the longer time periods for the S&P 500 in Figure 6, as well as for the FTSE and the Hang Seng in Figure 5.
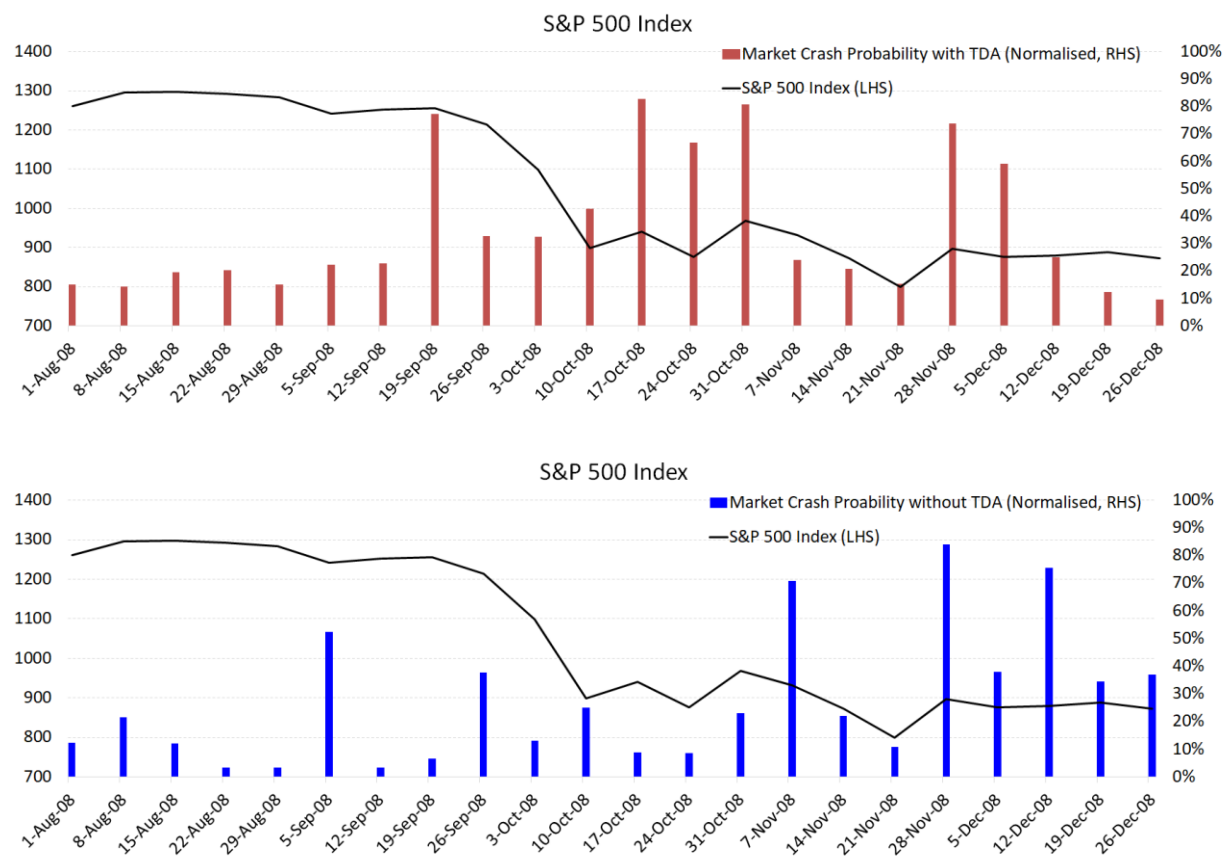
Secondly, another important aspect of the TDA-augmented system is feature selection. There are two aspects to this, the first of which is identification of new features and the second is feature importance. Over both the 1998 and 2008 crash periods the crisis indicator incorporating TDA identifies additional features, over and above those initially identified, which are significant for crash prediction. This suggests that the TDA methodology "unlocks" new features and signals around crashes providing further support for the "landscape" of crashes viewpoint proposed in Gidea and Katz (2018). Feature importance with and without these TDA-augmented features is also completely different. When the TDA-augmented features are omitted the machine learning algorithm weights two or three features, mostly involving short term asset returns, very highly and virtually ignoring most of the others. Once TDA features are added feature weighting is far more uniform and, in particular, longer term returns and term spreads as well as the TDA features all receive higher weightings. Thus the presence of TDA features seems to unlock greater value from existing factors which also indicates that incorporating TDA into a machine learning framework could be beneficial in a number of financial market settings.

Thirdly, while it is important to avoid false positives it is perhaps more important to try to avoid false negatives, as the cost of a false negative which is a crash that was not predicted is likely to be much higher than that of a false positive. Incorporating TDA in this machine learning framework ameliorates the problem of overfitting which is particularly important in the context of being able to predict crashes that would actually happen. This is an advantage of TDA that has been noted in other contexts, namely the ability to exploit weak signals that are nonetheless persistent.

Fourthly, taking a longer term perspective over the 25 year period from 1993-2018 for the S&P, the highest levels for the TDA-augmented crisis indicator were over the 1995-1999 period followed by

2008. This indicates that the late 1990s saw the highest level of stock market uncertainty as proxied by financial market indicators. The relative weight of TDA indicators was lowest over this period with the one week return for the S&P and gold receiving the highest weights. The TDA-augmented feature weighting increased sharply over the 2004-2009 period, declined over the 2010-2012 period and then increased again to its highest overall level while the relative importance of weekly return for the S&P and particularly gold has been much lower since about 1999. These results suggest a pattern of changing linkages within financial markets where TDA is able to detect the shape of changes, echoing the key findings in Gidea and Katz (2018).

Our methodology thus provides a systematic way of incorporating TDA in a standard machine learning framework which greatly facilitates its implementation. Our results indicate that this methodology provides new insights about the shape and landscape of stock market crashes as well as stock market uncertainty. The global nature of TDA allows it to capture different features from traditional machine learning algorithms and overall our findings suggest that this methodology could be a useful addition to the traditional financial risk management toolkit.

The rest of the paper is organised as follows. Our methodology is explained in Section 2 where we first introduce our modelling philosophy which combines TDA with machine learning. In Section 2.1 we describe the machine learning framework while Section 2.2 describes the feature engineering details, in particular, TDA features and how they are incorporated together in the machine learning framework. Our results are discussed in Section 3, where we discuss the performance of the model in crisis periods and also the performance with and without TDA features. Section 4 concludes and outlines directions for further research.

2. Methodology

A typical supervised machine learning modelling process is divided into two phases: the training phase and the testing phase. The training phase is the "in sample" phase, where the model processes considerable amount of data which are, in our case, historical time series of different descriptors of the market and make probabilistic predictions on how likely market will crash in the near future in each round. Then these predictions are compared to the corresponding ground truth to provide feedback, enabling the model to "learn" and improve its predictions until the training session is complete. During the testing phase, each time the model is no longer "learning" but simply making "out of sample" predictions which are then recorded for us to evaluate the performance of the model, after all the testing data are processed and predictions are made.

In such a modelling framework and pipeline, two high level preparations need to be made prior to the actual learning, namely, the data preparation and model specification. Our principle in approaching these tasks is to keep it (a) **simple** and (b) **objective**.

As outlined in the introduction section, as machine learning techniques have become more and more available in solving business problems, one option would have been to employ the most state-of-the-art modelling framework such as deep neural network or try as many models as possible then shopping around, which is similar to the approach in Chatizis et al (2018). This approach does not serve our purpose of examining and promoting TDA as a novel set of machine-learning-compatible toolkits. Instead, the intuitive exposition of how TDA can be applied to financial time series and profiling the crash landscape in Gidea & Katz (2018) inspires us to systematize its use, by taking an objective and measurable approach.

The rest of this section is organised as follow: we first describe the overall model and hyper parameter specification so as to link it to the findings exhibited in the result section; we then move on to the data preparation, in particular, to show that the feature engineering is objective and disciplined.

2.1 Hyper Parameters

The modelling framework we choose for our study is the decision-tree-based XGBoost model popular and effective among many data science competitions[c] (Chen and Guestrin 2016). Meanwhile, its tree-based learning algorithm enjoys a simple logic and explicability as illustrated in the model-learning flow chart below:
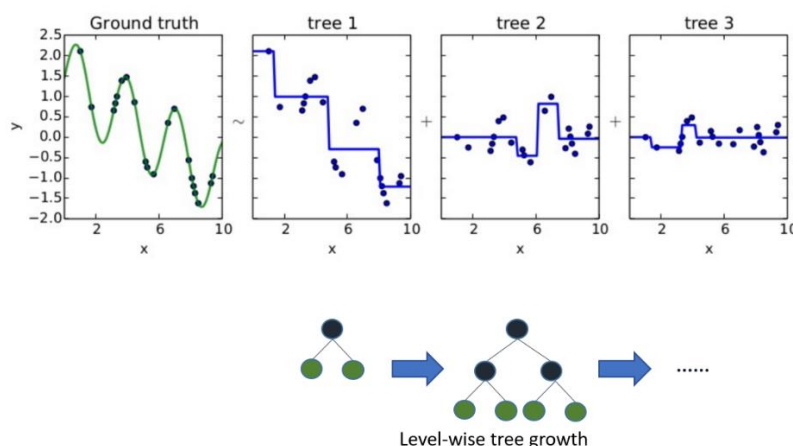
There is a dozen or so hyper parameters to configure and regulate the model-learning process, for the sake of simplicity, we only focus on two key hyper parameters, namely,

1) *max_depth*: the maximum depth the tree can grow, and
2) *nrounds*: the number of iterations performed during the training phase.

The first parameter determines the complexity of the final trained model for prediction; whereas the second parameter determines how fitting to the input dataset the model will be trained as. We exhaust a wide range of combination of the two parameter settings for the three market indices of interest (see the Appendix 1 for this fine-tuning exercise), and finally settle with max_depth=3 or 5 and nrounds = 5 or 10, to minimise the complexity and risk of overfitting and at the same time consistently achieve stable performance (measured by AUC score).

Having determined the model hyper parameters, we next move to the hyper parameter specification for data preparation and feature engineering.

The raw data for our exercise covers 35 years of weekly price time series from major asset types, namely, Equity, Fixed Income, FX and Commodities, so that major market crashes since 1983 are included for the model to learn as widely as different crash scenarios can vary. Specifically, we want to pre-process these raw data into a tabular form, i.e. a matrix: each row is the weekly snapshot of a chosen market, quantitatively described by a number of market features or columns, so that in total, we will have over 1800 datapoints, multi-dimensional, to train and test the model. We will get to the feature engineering subsection to detail the list of indices and the features to construct.

We next define the market crash as over 5% drop of the price of a chosen market index in a time window of 2 weeks. Admittedly, there is a subtle trade-off between data balance and learning the true crash patterns: on the one hand, if we have too few positive samples, i.e. data points of market crash, we will have rather imbalanced dataset to train the model; on the other hand, too low a threshold to define market crash will result in the model learning from noises instead of true crashes. Setting it to be 5% will yield a reasonable portion of positive labels in our learning dataset as shown in the table

---

[c] 17 out of 29 Kaggle competitions were championed by XGBoost based models.
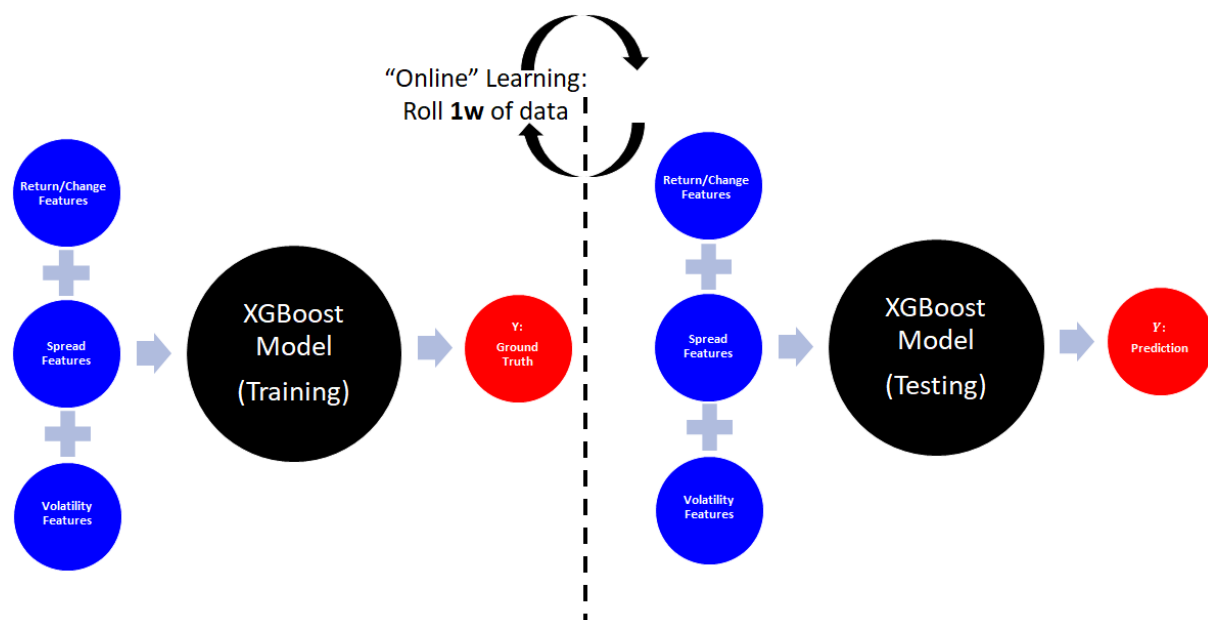
below, and at the same time, a reasonable magnitude that characterises a proper market crash. We have also considered a dynamic definition by number of standard deviations, however, that again would add to the complexity of the exercise and render the outcome and back-testing less intuitive.

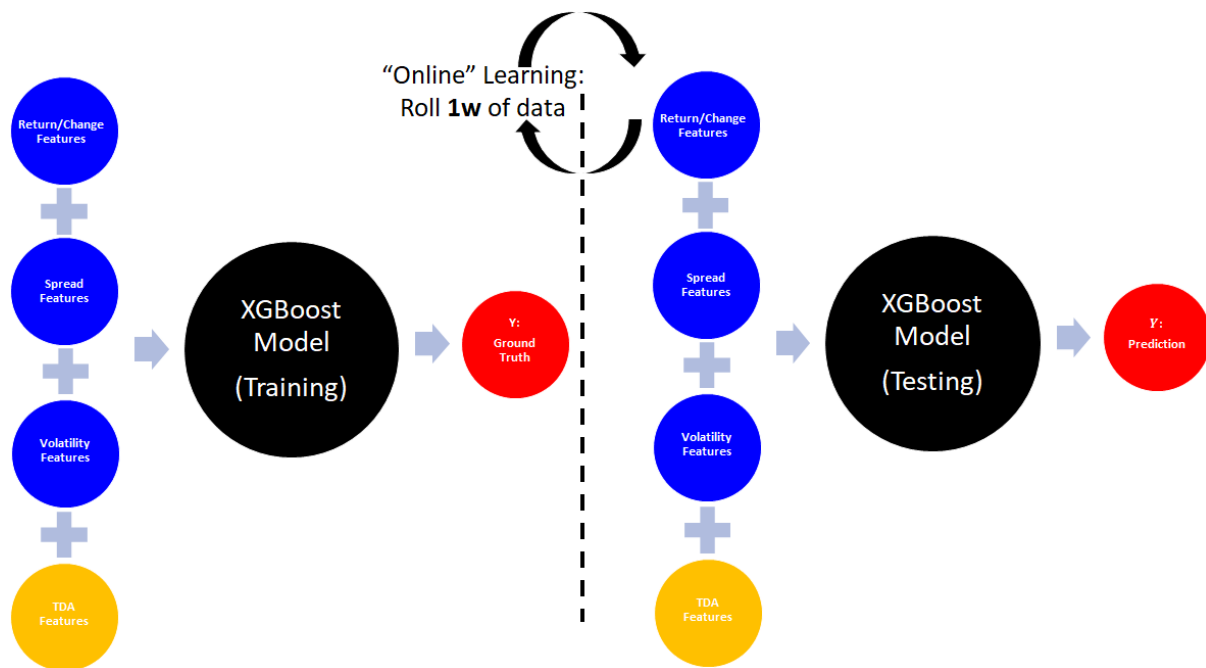|	No. Datapoints (No. weeks)	No. Crashes (No. weeks > 5% drop)	Data Balance (% of crashes over total number of weeks)
S&P 500	1861	72	4%
Hang Seng	1861	190	10%
FTSE 100	1861	74	4%

The next hyper parameter to consider is the width of the time window we would like to describe each weekly market snapshot. In principle, this should be proportional to the forecasting window of the crash indicator we build, and we choose 26 weeks, or 2 quarters as a typical market risk model would have a 26-weeks half-life which is similar to what is done in industry (Barbieri et al 2009). However, we do not do exponential-smoothing but only a simple equally-time-weighted approach when it comes to calculate volatility etc., again for the sake of simplicity and intuitiveness.

Having the above setup, we choose the first 10 years of our dataset as the initial training dataset, i.e. 520 datapoints, and make our first 2-week-window market crash prediction. We then expand the training dataset with one week timestep forward at a time. Every time, we re-train the model and make a new 2-week-window crash prediction, until the dataset is exhausted. We then collect the 1314 "out-of-sample" predictions to evaluate this "online-machine-learning" model against the corresponding ground truth, which yields us the results presented in the preceding section, for both the model scores (summarised in terms of Area under Curve and F1) and back-testing results. The below workflow charts visually illustrate how we obtain our results for models with and without TDA, and at the same time, prelude the data preparation work which we detail next:

Model without TDA:

Model with TDA:

"Online" Learning: Roll **1w** of data

Return/Change Features

Spread Features

Volatility Features

TDA Features

XGBoost Model (Training)

Y: Ground Truth

Return/Change Features

Spread Features

Volatility Features

TDA Features

XGBoost Model (Testing)

Y: Prediction

2.2 Feature Engineering

Governed by the principle of simplicity and objectivity, as visually illustrated above, the list of features we construct is parsimonious but wide enough to cover the following 3 types for the 4 asset classes the raw data was sourced from, in addition to the 4th "bolt-on" type from TDA:

1) Features of returns or changes
2) Features of spreads or differences
3) Features of volatilities
4) Features of TDA or shape of the time series of each time window

Below are the indices we use to represent the corresponding asset classes:

Equity: one from the three, S&P 500, FTSE 100 or Hang Seng

FI: 2 year, 5 year and 10 year treasuries

FX: DXY Dollar Index

Commodities: WTI Crude Oil and Gold

In particular, we take extra caution to only select one equity index to train the model in each market. For instance, for the US market, we only include S&P 500 in our feature engineering process, FTSE 100 and Hang Seng will be disregarded. Due to the nature of supervised learning, using more than one equity indices to train a model and predict one of them is likely to lead to overfitting, even the time segregation from the prediction/testing phase is strictly enforced, due to synchronicities between indices, information can leak: e.g. the crash of S&P on a Friday afternoon might only be felt and translated in HSI the week after. Therefore, to ensure the exercise is optimised to test the value of TDA techniques in enhancing the performance of crash detection, we only allow the relevant index from the same market in which we are detecting crashes.

Overall, we construct 15 features to describe the market on a weekly basis covering the past 35 years from 8 April 1983 to 30 November 2018 inclusive. We now discuss how we customise TDA for our purpose. For detailed exposition of concepts such as persistent homology and persistence landscape,

there are several excellent overviews (see for example Carlsson 2009, Edelsbrunner and Harer 2010 for a general introduction and Perea and Harer 2015 for applications to time series), and we refer to Gidea and Katz (2018) for motivation in applying it to financial markets.

Essentially, we compute 3 TDA features, from the 3 asset classes, namely, the chosen equity market index which we also use to detect crashes, the dollar index (DXY) and the oil price.

We use the same TDA descriptor, persistence landscape to quantitatively summarise the shape of each time series dataset. However, the key customisation are two folds:

1) To compute the persistence landscape, a filtration function is required. We simply use the top-down reverse height function, instead of the ball-radius used in Gidea and Katz;
2) We then further compute the change of the 1-norms of persistence landscape (PL) of the current time period against the one computed from the previous time period. Effectively, this version of the PL norm summarises the evolution of market landscape and regime shift.

Visually, (1) will return a persistence landscape matrix as below:
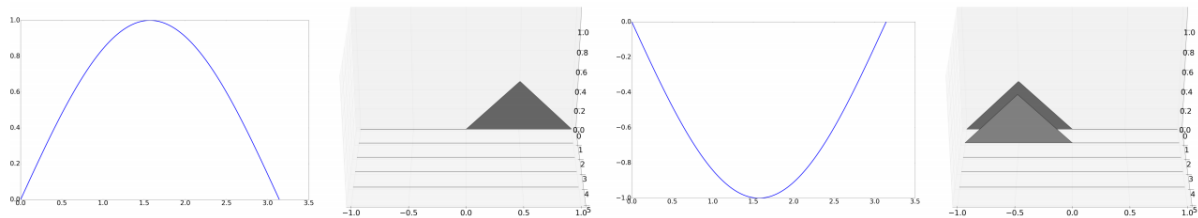
And the properties of TDA and Persistent Homology ensure the result is robust against noises:
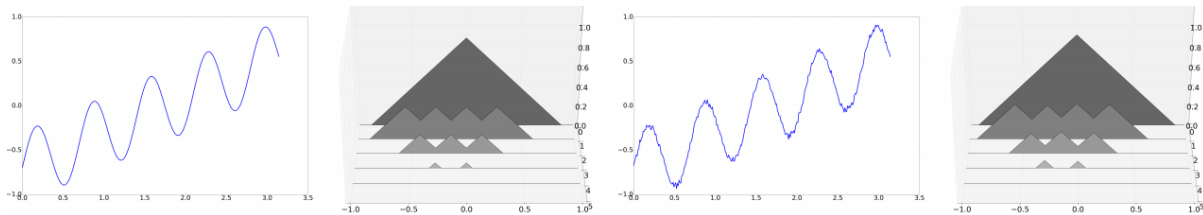
By employing an easy to use decision-tree-based supervised learning model, and preparing the data most economically as above, we hope to achieve the purpose that such methodology not only make our findings understandable but also adaptable for other problem solving where TDA can become an effective performance booster.

3. Results

We have performed analysis to obtain statistics and model performance, as well as drill-in crisis-period-focused examination via the back-testing exercise. We first focus on the performance of our early warning indicator around the 2008 stock market crash and then around the 1998 one as showcases[d]. We then step back to provide the overall statistical summary of the model performances and comparison studies between models with TDA feature bolted on and without.

3.1 Performance of the model around Market Crashes

The period of sharpest decline for the S&P 500 in 2008 was around the period of the collapse of Lehman Brothers. From the middle of September to end of November the S&P declined from 1255 to

---

[d] The scores are normalised using min-max for the entire back-testing window, so it ranges fully from 0-1 as a probabilistic measure.

800, a decline of 36%. It then recovered somewhat ending 2008 at around 900. Within the September-November period there are two episodes of decline, one until the third week of October where it fell 23%, and then over the month of November where it fell another 10%. As we see from the top panel of Figure 1 our TDA-augmented indicator rises sharply in the middle of September just before the first sharp decline, rising sharply from around 0.3 to above 0.8 in the course of a week. Since the beginning of 2008 its previous peak had been around 0.5 in February. It then drops back sharply for the next three weeks, remaining well above its previous level, and then peaks at around 0.85 in the third week of October and remains above 0.7 for the next two weeks, just before the next decline. It then falls sharply for three weeks, and rises sharply for the first two weeks of December where there is no major decline. From the bottom panel of Figure 1 we see that the non TDA indicator shows an increase in the first week of September to around 0.5, which is at the same levels in February, and then continues to decline until it jumps to just under 0.7 in the first week of November 2008. The overall peak is just under 0.9 in the first week of December, well after the sharpest declines have occurred.

The TDA-augmentedaugmented indicator's increase in mid-September was well above anything observed in the course of the year while the non TDA indicator's sharp jump in early September was still at a level seen earlier in the year when no subsequent decline was observed. It thus seems that the TDA-augmented indicator's signal would have received more attention. The peak of the non TDA-augmented indicator in December could be regarded as a false positive as could the sharp increase for the TDA-augmented indicator. However, the low levels of the non TDA-augmented indicator through the rest of September and all of October, when they were consistently below levels in February, could be regarded as a false negative which is a far more serious issue in the context of financial crash prediction.

We next analyse the 1998 situation when the S&P declined around 18% from mid-July to early September, another case of a short, sharp decline followed the debt crisis in Russia early of the year. In this case, as we see from the top panel of Figure 2, the TDA-augmented indicator rises sharply to above 0.9 in the first two weeks of July from levels of around 0.3 in the previous four weeks and then continues to decline, remaining at around 0.3 from August until the end of the year. There is a sharp increase to 0.6 at the end of May, which could be interpreted as a short term false positive. The non TDA indicator as we see from the bottom panel of Figure 2, increases sharply at the beginning of June but remains at a very low level throughout July and then peaks at almost 1.0 at the end of August. Both of these peaks, particularly the second one could be regarded as false positives and the low levels of the indicator in July could certainly be regarded as a false negative. Thus for this particular financial crisis, it is clear that the non TDA-augmented indicator is unable to detect the crash while the TDA-augmented indicator does a much better job.

We next analyse the behaviour of both indicators for two other indices, namely the FTSE 100, the major UK index and the Hang Seng, the main Hong Kong index, both over the 2008 crisis period. The sharpest period of decline for the FTSE was in the first week of October and as we see from Figure 2 the TDA-augmented indicator peaks at exactly this point and then peaks again at the end of October which was the beginning of a less sharp decline. The non TDA-augmented indicator's peak of around 0.7 is during the third and fourth weeks of October near the bottom of the sharp decline. The pattern is similar for the Hang Seng index where the major decline begins around the middle of September. The TDA-augmented indicator jumps around this point, which is in fact the peak of 0.8, and jumps sharply again in mid-October which is the beginning of a shallower decline. There is a third jump at the end of October which is a week before another 10% decline in November, and there are two potential false positives at the end of November. In contrast the non TDA-augmented indicator's peak is much lower at around 0.6 and is at the end of the first major decline in the second week of October and before the second shallower decline. There is a thus a consistent pattern of both performance of the

TDA-augmented indicator in signalling a decline as well in the differences between the performance of the TDA and non TDA-augmented indicators.

The nature of the TDA-augmented indicator is quite different from the non TDA-augmented indicator as can be seen from the features and feature weighting. The TDA-augmented indicator has the potential to unlock and better coordinate extra features based on TDA and so has access to a larger set of features. The feature weightings during September to November 2008 for both indicators are shown in Figure 3 and for the non TDA-augmented indicator we see the dominance of two features, S&P volatility and the previous week's return for gold emerging as the dominant features over this period. The S&P volatility has the highest weighting, between 20% and 30% over this period while that for the lagged gold return is between 15% and 20%. The six month return on the S&P and crude oil also receive weightings of around 10%, with several features receiving no weighting at all. Hence the non TDA indicator seems to weight certain factors quite heavily and some others not at all, suggesting that it tends to optimize on training in sample data quite aggressively which then leaves it susceptible to the problem of overfitting. For both the market declines we focus on, the price of overfitting seems to be both false positives as well as false negatives. In contrast, as we see from Figure 2, the TDA-augmented indicator does not give weightings of over 20% consistently to any of the features, and all features including the TDA features are weighted. In fact only one feature, the S&P volatility consistently receives weightings above 10% over this period. The other features that receive weightings around 10% are the lagged weekly return on the S&P and the six month return on crude oil. The S&P TDA feature receives the highest weighting of the TDA features, around 5%. The TDA-augmented indicator thus does not optimize nearly as aggressively which allows the model training to pick up weaker signals that could contribute to its improved performance relative to the non TDA indicator. Thus incorporating TDA into the machine learning based framework seems to "unlock" a much wider variety of signals and therefore avoid some of the overfitting problems that the non TDA-augmented indicator appears to have. During the 1998 market decline the overall pattern is very similar. For the non TDA-augmented indicator three features, namely crude oil volatility and the lagged weekly return on the S&P as well as gold, all receive weighting above 20% while most of the other features receive zero weighting. In the case of the TDA-augmented indicator only one feature, the volatility of the dollar index, receives weighting of around 20%, while three features of the S&P namely its volatility, lagged one week and six months return all receive weightings of around 10%. All of the other features including the TDA features receive some weighting. Thus while both indicators give high weighting to short and medium term momentum type indicators the TDA indicator also weights other potentially weaker signals which seem to help with crash prediction.

For the FTSE over the 2008 crisis, the non TDA indicator weights a larger number of features than for the S&P with six features receiving weights of around 10% as we see from Figure 4. The TDA-augmented indicator weights four features around 10% with the S&P volatility receiving a weighting close to 20%. The pattern of weightings is somewhat different between the two indicators, with the gold return highly weighted for the non TDA indicator and not at all for the TDA indicator and six month returns more highly weighted for the TDA indicator, for example, suggesting that the TDA indicator tends to focus more on the longer term signals. The pattern for the Hang Seng is more similar to that for the S&P with the non TDA indicator weighting S&P volatility at around 40% and also weighting the lagged one week return on the S&P over 10%. The TDA-augmented indicator also weights S&P volatility highest, but around 25% and weights the other features more uniformly with longer term returns receiving much higher weights than for the non TDA-augmented indicator. At least one TDA feature receives a weighting of around 10% for both the FTSE and the Hang Seng. Thus the pattern of differences in feature weighting with the non TDA indicator weighting short term momentum signals and the TDA indicator weighting longer term, potentially weaker, signals is also visible for these indices.

3.2 Overall Summary of Model with and without TDA features

The "zoom in" focus on the model performances during crisis periods provides us with intuitive appreciation on how TDA features add value. We now "zoom out" to describe the overall performances and comparison studies for models trained with and without TDA features.

By focusing on the model performances during crisis periods, we can see the boosted accuracy of the TDA-augmented models in detecting market crashes when it really is happening, namely, the improvement in reducing false negatives, especially when the bets are high. We aggregate and quantify this observation over the entire back-testing period by the classic F1 score in machine learning practice.

Figure 5 summarises the TDA advantage visually for all three major market indices we have tested. For a given confidence level (a chosen cutoff point in terms of probability, so that a prediction is positive if its probability is higher than this cutoff probability and vice versa), F1 score measures how good a classifier is by taking into account both the precision and recall ratios, i.e. in our case, "how many positive predictions are correct out of all the positive predictions" and "how many positive predictions are correct out of all the actual crashes", respectively. Therefore, by design, F1 score punishes both false negatives and false positives: taking the S&P 500 back-testing for example, when the cutoff is low, most of the predictions will be considered positive but the reality is the opposite, which results in large number of false positives and hence lower precision, but this could be compensated by higher recall, because a lot of the true crashes will be "indiscriminatingly" predicted correctly, and this lower false negatives will reflect in higher recall ratio; on the other hand, when the cutoff is high, most of the predictions might be considered negative and would miss out the true crashes, resulting in high number of false negatives and low recall ratio.

From Figure 5, all models trained without TDA perform poorly when the cutoffs, i.e. the convictions or probabilities of coming market crashes, are beyond the 50% mark, with the F1 scores unforgivingly dropping to near 0%; on the other hand, the models trained with TDA features sustain a better prediction at high cutoffs, with F1 score stabilised at 10%. Admittedly, there is still huge room to improve, but consider the noisiness of financial market data, these results of 10%'s and above are the fair reflection and summary of TDA models' performance during the time when markets are really in crisis mode, as well as, in the same window, the increased model sensitivity when markets are not, i.e. punished by higher crash scores or probability during calm periods.

In the case of the S&P 500, we compare the crash probabilities of models with and without TDA features in Figure 6 and we see that crash probabilities are in general higher for TDA model than for non-TDA model, regardless of market regimes. As a result, the over performance measured by Area Under Curve for TDA models are slightly lower than for those without (also in Figure 5), which can be seen as a "cost" in order to provide more robust crash signals during crisis periods. The Area under Curve measure for both models are broadly of the same magnitude as for the various models in the study by Chatzis et al (2018). A similar pattern of results obtains for the other indices which are shown in Appendix 2.

We next focus on the relative importance of different factors for the S&P 500 over the entire 24 year sample period. We consider three of the most highly weighted features, namely the lagged one week return on the S&P, the volatility of the S&P and the lagged one week return on gold, both with and without incorporating TDA features. In addition we aggregate the weights of the three TDA features and the results are shown in Table 1. A number of stylized features emerge. The weights of the three non-TDA features exhibit considerable time variation across the two halves of the sample period. In the first half the two return based features exhibit much higher average weightings, with the gold return feature having average weightings above 20%, which are higher in the absence of TDA features. However in the second half of the sample the volatility feature dominates and the average weight in the absence of TDA is almost double that in the presence (29% relative to 15%) while the return based features average weighting is below 10%. Thus volatility emerges as the most significant

feature in this early warning system since 2006. The same pattern is visible for the TDA-augmented features with their average weighting more than double over the second period compared with the first (17% relative to 8%). Taken together these suggest a possible shift in the nature of the S&P over the second half of our sample.

4. Conclusion and Discussion

Developing early warning systems for financial crashes is an issue of great interest for market participants, policy makers and regulators, but is very challenging, as predicting changes, particularly extreme changes, in financial markets is very difficult. We propose a novel early warning system for detecting financial market crashes that utilizes the information extracted from the shape of financial market movement. Our system incorporates Topological Data Analysis (TDA), a new set of data analytics techniques specialised in profiling the shape of data, into a more traditional machine learning framework.

Our methodology thus provides a systematic way of incorporating TDA in a standard machine learning framework which greatly facilitates its implementation. Our results indicate that this methodology provides new insights about the shape and landscape of stock market crashes as well as stock market uncertainty. The global nature of TDA allows it to capture different features from traditional machine learning algorithms and overall our findings suggest that this methodology could be a useful addition to the traditional financial risk management toolkit.

As stated and re-emphasised throughout our report above, the simplicity and intended adoptability govern the designs of our exercise and methodology, in order to maximise the value proposition of TDA as a novel toolkit solving problems from noisy financial data. There are, therefore, many interesting directions the research can be extended to: These include (1) machine-learning-based security and fund selection where the framework used here could be used to identify securities for selection; (2) factor investing where TDA could be used to create new factors capturing the periodicity of the market; (3) credit default risk and credit scoring system where a data-driven machine learning framework together with consideration of the shape of the credit dataset, could potentially yield more robust and scalable solution than the traditional Merton type approach; (4) co-ordinated sector and style rotation analysis where TDA could be used to identify cycles.

References

Angelo Barbieri, Kelly Chang, Vladislav Dubikovsky, John Fox, Alexei Gladkevich, Carl Gold, Lisa Goldberg (2009) Modelling Value at Risk with Factors, *MSCI Barra White Paper on Model Insights*.

Richard Bolton and David Hand (2002) Statistical Fraud Detection: A Survey, *Statistical Science,* 17, 235-249.

Gunnar Carlsson (2009) Topology and Data *Bulletin (New Series) of the American Mathematical Society,* 46, 255-308.

Tianqi Chen and Carlos Guestrin (2016) XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ICM International Conference on Knowledge Discovery and Data Mining,* 785-794.

Sotirios P. Chatzis, Vassilis Siakoulis , Anastasios Petropoulos, Evangelos Stavroulakis, Nikos Vlachogiannakis (2018), Forecasting Stock Market Crisis Events Using Deep and Statistical Learning, *Expert Systems With Applications,* 112, 353-371.

Mark Cecchini, Halden Aytug, Gary Koehler and Praveen Pathak (2010), Detecting Management Fraud in Public Companies, *Management Science,* 56, 1146-1160.

Geng Cui, Man Wong and Hon-Kwong Lui (2006), Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming, *Management Science,* 52, 597-612.

Herbert Edelsbrunner and John Harer (2010) Computational Topology: An Introduction, American Mathematical Society, ISBN-13: 978-0821849255

Jeffrey Frenkel and George Saravelos (2010), Are Leading Indicators of Financial Crises Useful for Assessing Country Vulnerability? Evidence from the 2008-09 Global Crisis, NBER Working Paper 16047.

Jeffrey Frenkel and Shan-Jin Wei (2004), Managing Macroeconomic Crises, NBER Working Paper 10907.

Marius Gidea and Yuri Katz (2018) Topological Data Analysis of Financial Time Series:

Landscapes of Crashes, *Physica A: Statistical Mechanics and its Applications,* 491, 820-834.

Shafie Gholizadeh and Wlodek Zadrozny (2018) A Short Survey of Topological Data Analysis in Time Series and Signal Processing, Working Paper Arxiv 1809-10745v2

Yan Huang, Gang Kou and Yi Peng (2017) Nonlinear Manifold Learning for Early Warnings in Financial Markets, *European Journal of Operational Research*, 258,692-702

Justin Lin (2008), The Impact of the Financial Crisis on Developing Countries, *Korean Development Institute Working Paper*.

Francine Gresnigt, Erik Kole, Philip Hans Franses (2015), Interpreting financial market crashes as earthquakes: A new Early Warning System for medium term crashes, *Journal of Banking and Finance,* 56, 123-139.

Martin Langkvist, Lars Karlsson, Amy Loutfi (2014), A Review of Unsupervised Feature Learning and Deep Learning, *Pattern Recognition Letters,* 42, 11-24.

Marcus Lopez de Prado (2018), Advances in Financial Machine Learning, Wiley. **ISBN-13:** 978-1119482086.

Jose Perera and John Harer (2015), Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis, *Foundations of Computational Mathematics,* 15, 799-838.

Cuneyt Sevim, Asil Oztekin, Ozkan Bali, Serkan Gumus and Erkam Guresen (2014), Developing an Early Warning System to Predict Currency Crises, *European Journal of Operational Research,* 237, 1095-1104.

Abhinav Srivastava , Amlan Kundu, Shamik Sural, Arun Majumdar (2008), Credit Card Fraud Detection Using Hidden Markov Model, *IEEE Transactions on Dependable and Secure Computing*, 5, 37-48.

Figures and Table

Figure 1: SP for 2008 and 1998

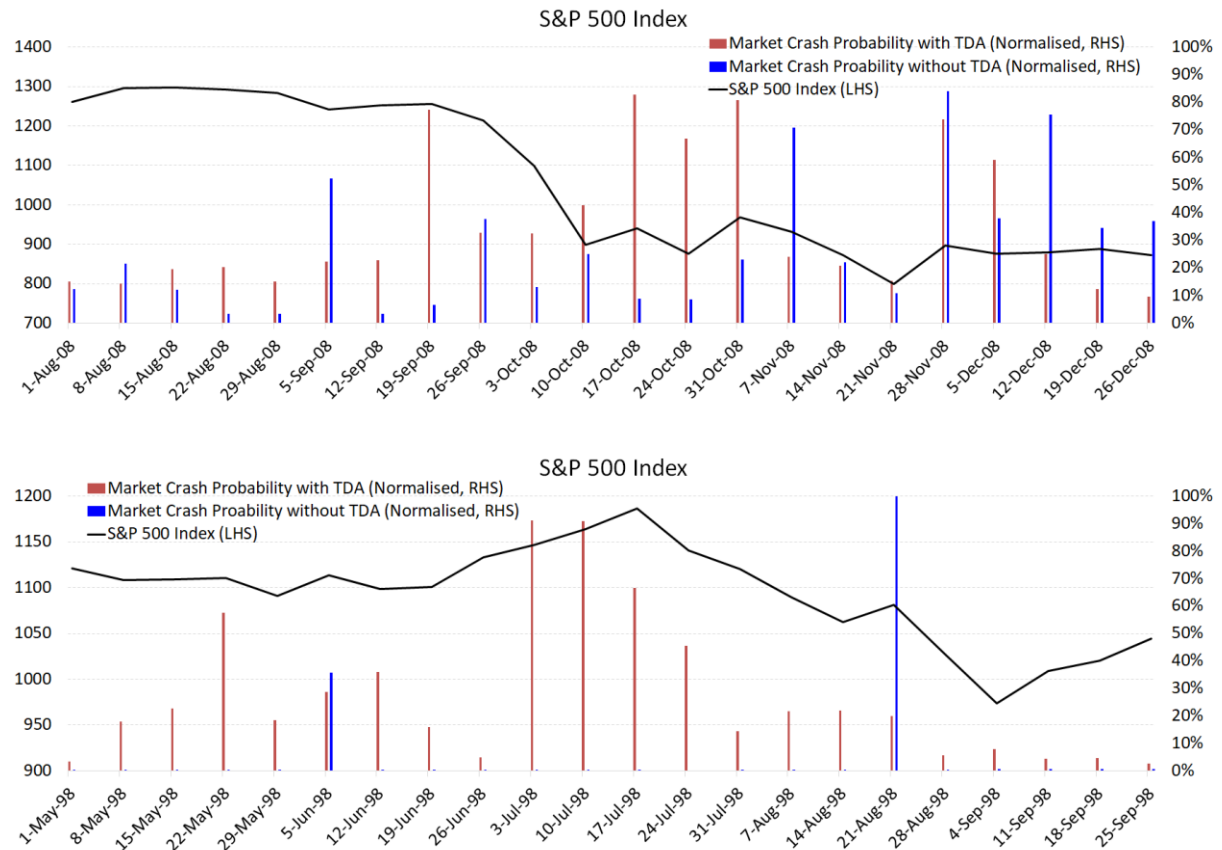**Figure 1:** This figure shows the market crash probabilities for the S&P 500 index using the Early Warning System discussed in Section 2, incorporating TDA features (red bars) and without TDA features (blue bars). The crash probabilities are shown around the period of the 2008 market decline (top panel) and the 1998 market decline (bottom panel). Details of TDA and how it is used in this context are given in Section 2.

Figure 2: FTSE and HSI for 2008

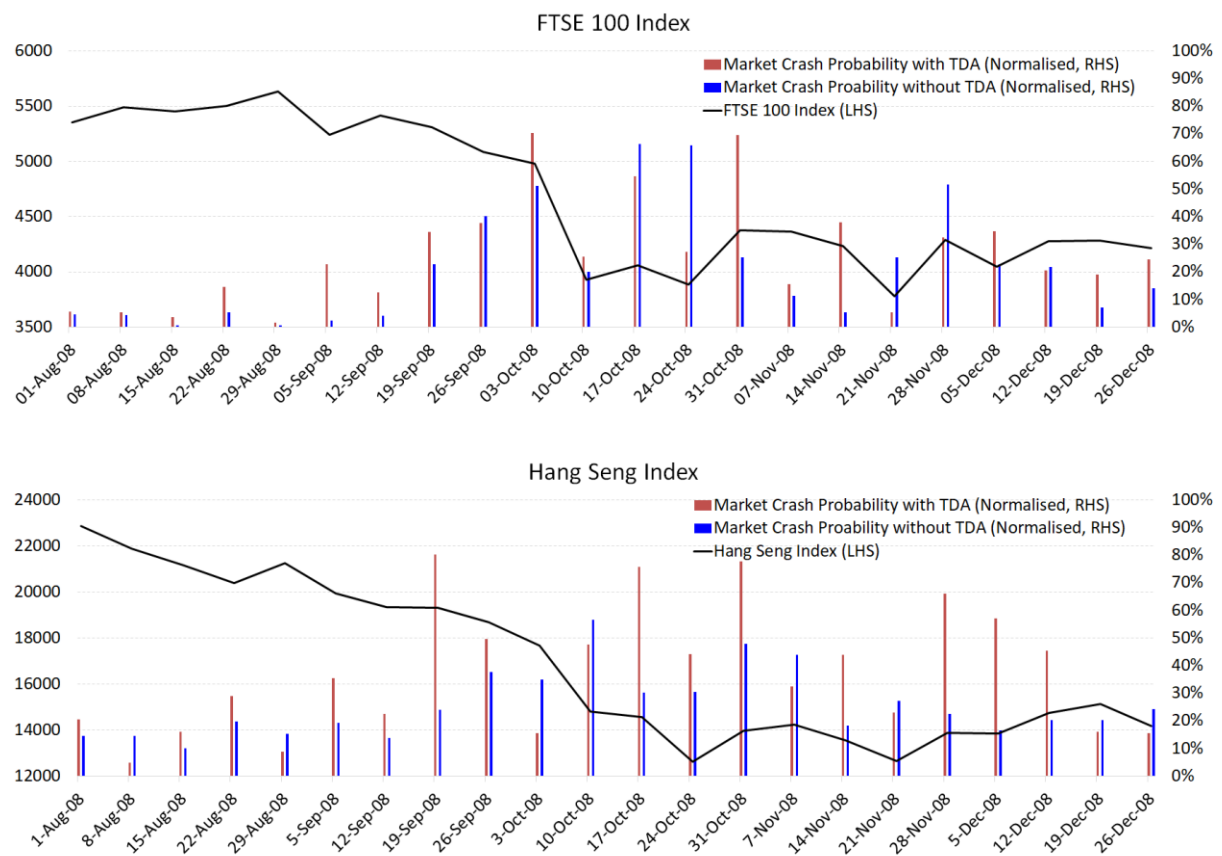**Figure 2:** This figure shows the market crash probabilities for the FTSE 100 (upper panel) and the Hang Seng Index (lower panel) using the Early Warning System discussed in Section 2, incorporating TDA features (red bars) and without TDA features (blue bars). The crash probabilities are shown around the period of the 2008 market. Details of TDA and how it is used in this context are given in Section 2.

Figure 3: Feature weights for SP 2008 and 1998 with and without TDA
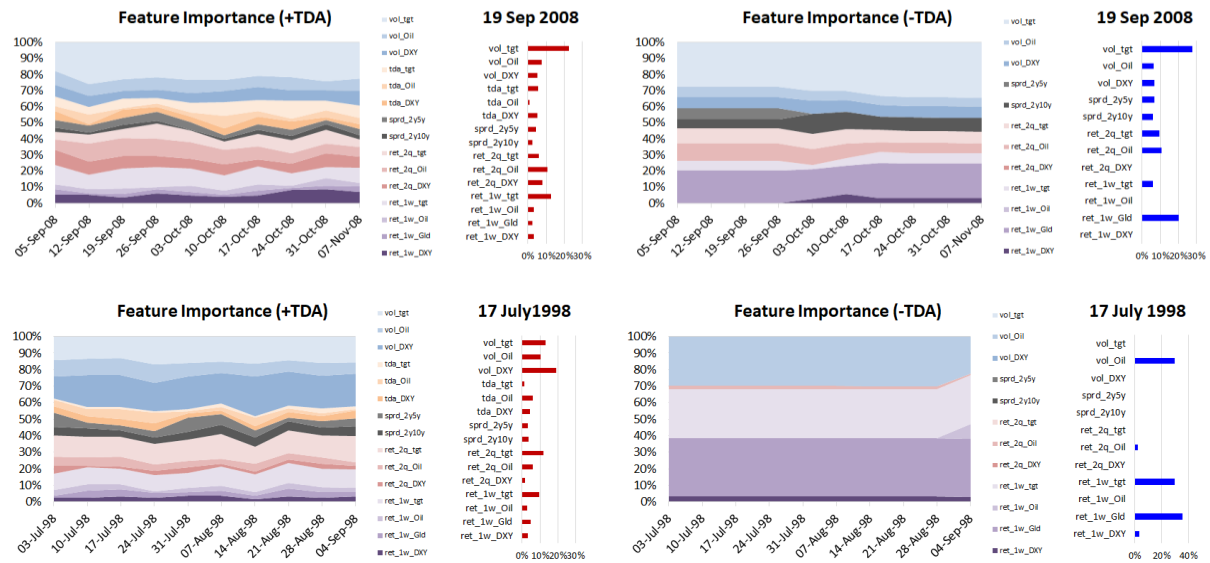
Figure 3: This figure shows the weightings of the various features for the S&P 500 used in our Early Warning System over the period of the 2008 market decline (upper panel) and the 1998 market decline (lower panel). For the top two sub-panels the graph on the left side shows the weightings over the September to November 2008 period while the one on the right is the weightings on a specific date, 19th of September. For the bottom two sub-panels the left side shows the weightings over the July to September 1998 period, while the one on the right shows the weightings on the 17th of July 1998. For both the top and bottom panel the sub-panel on the right incorporates the TDA features while the one on the left omits TDA features. Details of the Early Warning System, feature engineering and weighting with and without TDA are provided in Section 2.

Figure 4: Feature weights for FTSE and HSI 2008 with and without TDA

FTSE 100
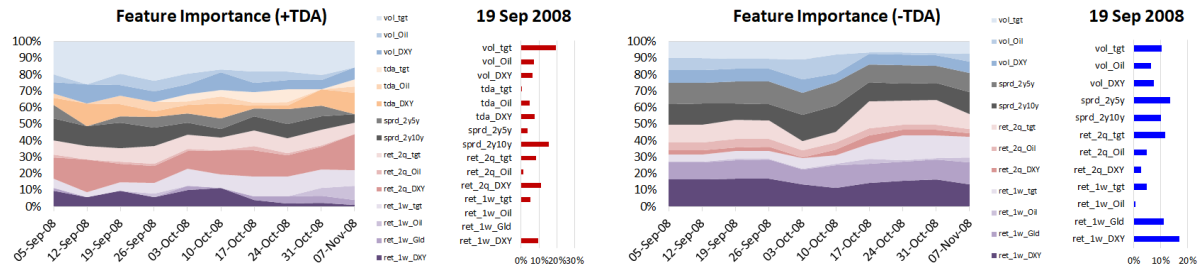
Hang Seng
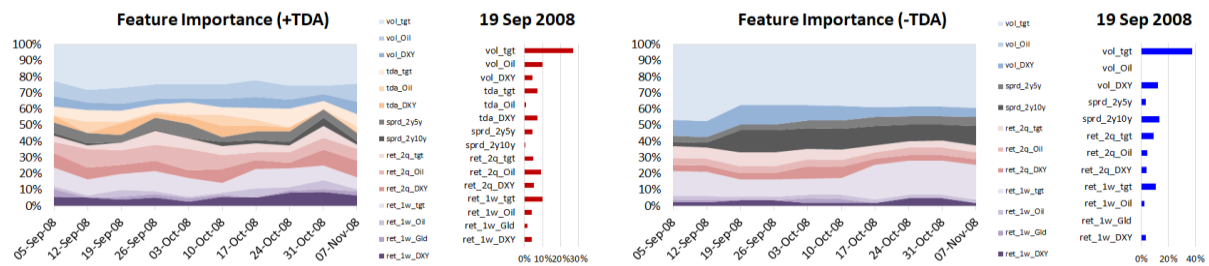
Figure 4: This figure shows the weightings of the various features for the FTSE 100 (upper panel) and Hang Seng Index (lower panel) used in our Early Warning System over the period of the 2008 market decline. For the four sub-panels the graph on the left side shows the weightings over the September to November 2008 period while the one on the right is the weightings on a specific date, 19th of September. For both the top and bottom panel the sub-panel on the right incorporates the TDA features while the one on the left omits TDA features. Details of the Early Warning System, feature engineering and weighting with and without TDA are provided in Section 2.

Figure 5: F1 Score for SP, FTSE, HSI
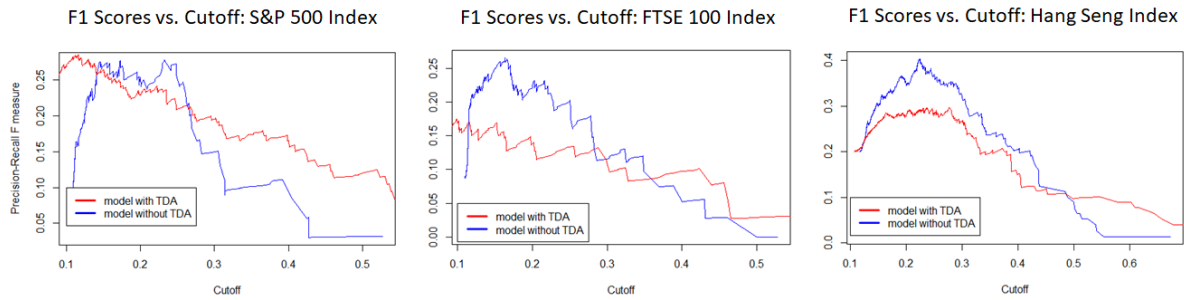
Figure 5: This figure shows the F1 score for our Early Warning System with TDA features (red line) and without TDA features (blue line), as a function of the cut-off probability. These are shown from left to right for the S&P500, the FTSE100 and the Hang Seng Index and the time period is 1993-2018. The F1 score is a measure that considers both false positives and false negatives, which is discussed in more details in Sections 2 and 3. Details of the Early Warning System as well as TDA are given in Section 2.

Figure 6: Crash Probability with and without TDA, AUC for SP
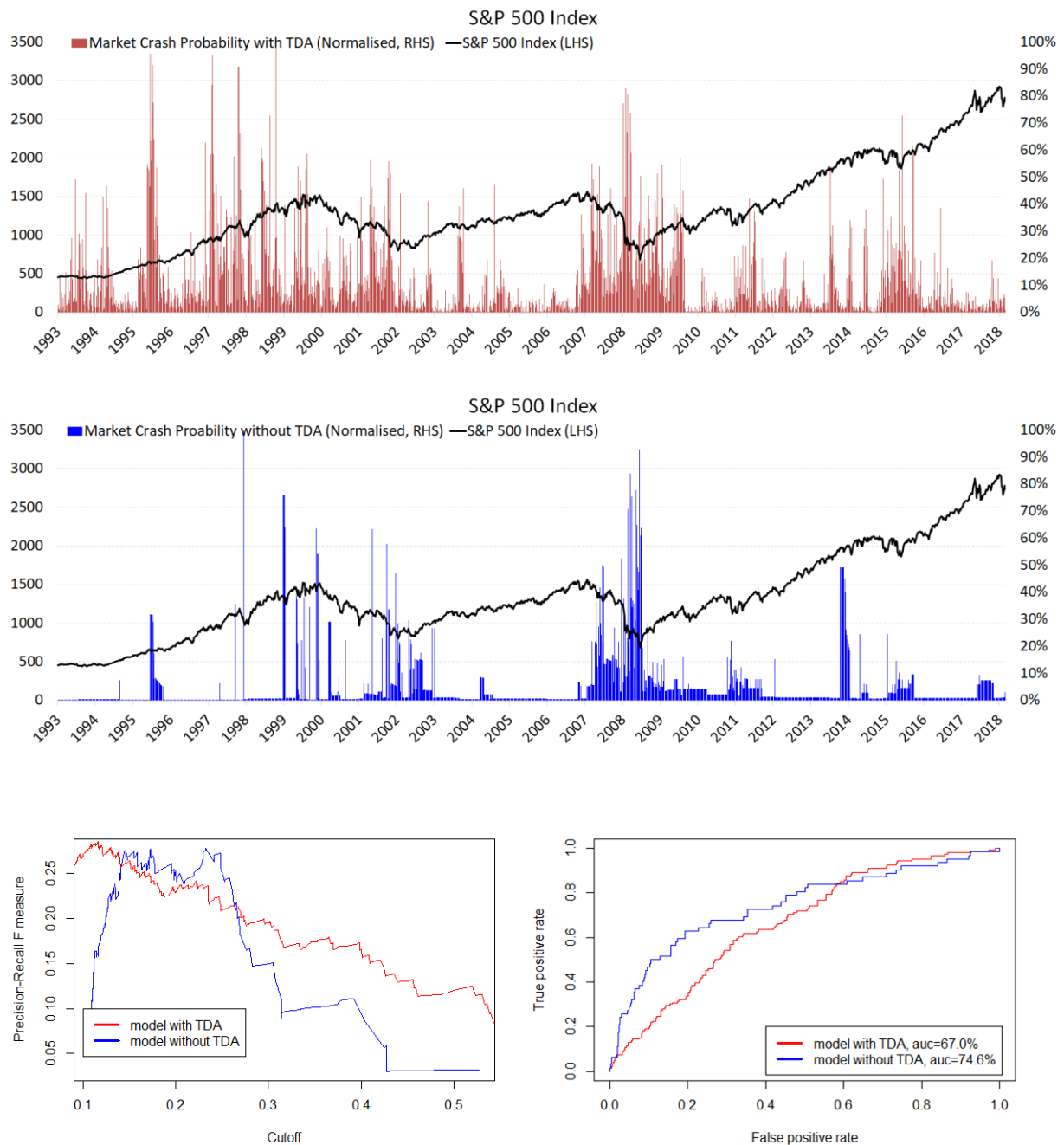
Figure 6: The upper panel of this figures shows the evolution of crash probabilities for the S&P 500 over the 1993-2018 period (red bars) together with the level of the S&P500 (black line), when TDA features are incorporated. The middle panel shows the evolution of the crash probabilities (blue bars) when TDA features are omitted together with the level of the S&P500 (black line) over the 1993-2018 period. The bottom panel shows the both the Precision Recall measure (F1 score) for the S&P500 over the 1993-2018 period as a function of the cutoff probabilities as well as the Area under Curve which measures the area size under the curve plotting true positives versus false positives. The Precision Recall measure is a measure that takes into account of both false positives and false negatives and is discussed in Sections 2 and 3. Details of the crash probabilities as well as TDA are given in Section 2.

Table 1: Selected feature importance over the entire back-testing period by year

	Agg TDA weights	Ret 1w TDA	Ret 1w	Ret 1 w Gold TDA	Ret Gold	Vol tgt TDA	Vol tgt
1994	2.35%	29.62%	28.76%	45.62%	32.69%	0.09%	0.00%
1995	2.78%	28.29%	28.74%	40.90%	32.75%	0.00%	0.00%
1996	6.14%	38.35%	30.32%	40.55%	34.82%	0.00%	0.00%
1997	8.03%	49.37%	31.13%	36.26%	35.82%	0.00%	0.00%
1998	3.62%	0.88%	28.83%	66.00%	34.43%	9.84%	0.00%
1999	3.43%	1.17%	24.72%	7.09%	34.15%	7.84%	0.00%
2000	5.91%	4.34%	11.58%	12.33%	35.13%	3.00%	4.05%
2001	6.02%	2.49%	7.68%	8.01%	27.51%	4.31%	7.20%
2002	6.08%	2.70%	9.11%	12.89%	25.00%	11.43%	7.79%
2003	7.65%	1.37%	9.82%	4.55%	16.32%	17.17%	16.49%
2004	16.26%	2.61%	5.15%	3.26%	14.28%	20.86%	22.14%
2005	14.82%	5.12%	3.16%	2.60%	12.57%	24.35%	26.43%
2006	17.40%	4.86%	3.19%	2.74%	12.45%	22.69%	27.98%
2007	19.38%	5.19%	2.80%	3.80%	9.63%	21.98%	29.87%
2008	16.96%	4.08%	4.48%	6.89%	17.88%	29.11%	27.97%
2009	12.75%	3.31%	4.48%	5.69%	16.19%	28.36%	24.00%
2010	6.47%	3.74%	4.79%	3.36%	16.73%	27.69%	24.46%
2011	5.06%	5.49%	6.56%	1.73%	18.38%	19.25%	25.05%
2012	10.66%	12.29%	8.75%	2.26%	21.49%	12.58%	30.42%
2013	17.29%	11.51%	8.69%	2.39%	21.25%	6.86%	30.75%
2014	20.63%	11.62%	5.40%	3.52%	16.02%	7.97%	31.86%
2015	20.91%	14.23%	4.60%	3.78%	14.92%	8.11%	32.51%
2016	33.15%	12.05%	5.47%	2.20%	16.51%	6.55%	32.66%
2017	18.63%	13.29%	5.51%	1.55%	15.51%	7.46%	31.33%
2018	18.80%	13.89%	5.94%	3.87%	15.63%	6.47%	30.87%

Table 1: This table shows the evolution of feature weightings as the yearly averages of the feature weightings over the 1994-2018 period for the Early Warning System for the S&P 500. The left most column (Agg TDA weights) is the sum of the weightings of the three TDA features described in Section 2. The three other features considered are the lagged 1 week return on the S&P500 (Ret 1w), the lagged one week return on gold (Ret 1w Gold) and the lagged volatility of the S&P 500 (Vol tgt). The columns with TDA show these feature weightings when TDA features are also incorporated while the other ones are the weightings without TDA features incorporated. Details of the Early Warning System, TDA features, and feature weightings are in Section 2.

Appendix 1:

This appendix provides detailed AUC Scores matrix for different combinations of the two key model hyper parameters: Max Depth and Number of Iterations, with and without TDA for the Hang Seng, S&P 500 and the FTSE 100 indices as discussed in Section 2.1.

Hang Seng Index

xTDA

% Max Depth	Number of Iterations									
	5	10	15	20	25	30	35	40	45	50
3	77.2	76.6	75.0	73.3	72.9	71.6	70.2	69.0	68.5	67.5
4	73.5	71.2	69.5	67.7	65.5	65.0	64.4	64.6	64.3	64.5
5	72.8	71.8	70.6	68.9	68.0	67.0	66.6	66.1	65.6	65.5
6	71.2	71.2	70.9	69.0	68.0	66.9	66.3	66.1	66.0	66.0
7	71.6	71.6	71.0	69.6	69.3	68.8	68.6	68.7	68.8	68.3
8	69.0	69.6	68.4	67.6	67.3	67.5	67.5	67.5	67.7	67.4
9	69.7	69.5	68.4	68.8	68.0	67.9	67.8	67.6	67.2	67.2
10	69.2	69.0	67.0	66.3	65.7	66.0	66.0	66.0	66.0	66.2

TDA

% Max Depth	Number of Iterations									
	5	10	15	20	25	30	35	40	45	50
3	77.0	76.7	73.3	72.4	70.9	69.7	68.7	67.4	66.7	66.1
4	74.0	72.7	69.9	69.5	69.1	68.2	66.9	65.2	64.8	63.9
5	72.8	72.2	68.4	66.5	65.6	64.5	63.4	62.6	63.0	63.0
6	71.4	69.2	68.1	66.8	65.5	64.4	64.1	64.0	63.9	63.9
7	71.2	70.6	68.9	67.9	67.8	66.6	66.3	66.5	66.2	66.1
8	71.3	68.8	67.5	67.1	66.3	65.7	65.0	64.5	64.6	64.7
9	70.7	67.8	67.3	66.2	65.4	64.2	65.3	65.3	65.0	64.6
10	72.1	68.9	67.7	66.9	66.5	65.8	65.1	65.0	65.4	65.2

S&P 500 Index

xTDA

% Number of Iterations										
Max Depth	5	10	15	20	25	30	35	40	45	50
3	74.6	77.7	79.3	78.7	78.5	77.9	77.5	77.5	77.0	77.0
4	75.9	78.1	79.6	78.2	78.2	77.5	77.6	77.1	76.9	76.8
5	79.0	78.2	79.5	78.7	78.8	78.7	78.0	77.4	77.3	77.6
6	77.7	79.8	80.4	79.6	78.9	78.8	78.9	79.1	78.6	78.3
7	77.9	79.2	78.3	78.2	78.6	78.5	78.8	78.3	77.9	78.3
8	78.2	80.2	80.4	80.0	79.9	79.6	79.3	79.1	78.7	78.5
9	77.0	79.8	80.0	79.2	79.4	79.2	79.0	78.9	78.5	78.4
10	76.7	79.3	80.4	79.9	79.2	79.2	79.0	78.6	78.5	78.4

TDA

% Number of Iterations										
Max Depth	5	10	15	20	25	30	35	40	45	50
3	67.0	71.6	69.3	68.7	68.4	68.0	67.4	66.9	66.6	66.9
4	72.1	74.5	72.0	71.8	70.2	69.2	69.0	68.7	68.9	69.1
5	70.7	73.9	71.6	71.2	70.0	69.3	69.5	69.1	69.2	69.4
6	70.8	73.3	71.8	71.8	70.6	69.7	69.4	69.5	69.5	69.7
7	69.6	73.1	70.9	71.4	69.7	68.9	68.6	68.6	68.4	68.8
8	69.9	74.0	71.7	72.3	70.6	69.6	69.7	69.8	69.7	69.9
9	69.6	73.2	70.4	71.4	70.0	68.8	69.0	69.3	69.2	69.3
10	70.0	73.4	70.8	71.5	70.4	69.5	69.4	69.5	69.4	69.3

FTSE 100 Index

xTDA

% Number of Iterations										
Max Depth	5	10	15	20	25	30	35	40	45	50
3	73.3	78.2	74.5	72.4	70.1	67.8	65.8	63.8	62.1	61.8
4	73.4	77.1	74.2	72.2	70.2	68.1	67.4	66.2	65.3	64.8
5	75.8	76.1	74.3	72.1	70.2	68.3	67.9	67.8	68.1	68.4
6	75.0	76.9	73.2	71.7	70.0	68.9	68.1	67.8	67.2	67.5
7	74.1	74.4	70.8	70.7	69.9	68.9	68.2	68.0	67.9	68.1
8	73.4	75.1	73.0	73.2	71.5	70.8	71.1	71.1	71.0	70.7
9	74.3	75.7	73.5	72.1	71.4	69.4	69.0	69.2	69.2	69.1
10	74.2	75.8	72.4	72.6	71.7	69.9	70.1	70.1	69.8	70.0

TDA

% Number of Iterations										
Max Depth	5	10	15	20	25	30	35	40	45	50
3	67.0	72.3	69.9	68.8	66.9	66.0	65.0	64.6	64.5	64.3
4	68.3	70.5	69.1	67.7	67.5	67.8	67.4	67.2	67.0	67.1
5	68.2	70.6	68.4	68.9	68.2	67.0	66.8	67.0	66.7	66.3
6	67.6	69.2	68.1	68.4	68.3	67.5	67.0	66.8	66.3	66.0
7	67.4	70.3	68.9	69.5	68.9	68.4	67.9	67.6	67.4	67.3
8	68.0	70.5	69.8	70.2	69.5	69.1	68.5	68.5	68.5	68.3
9	67.6	68.5	67.4	68.3	67.6	67.5	66.9	67.2	66.9	66.6
10	68.2	69.0	68.2	69.0	68.7	68.4	67.9	68.1	67.9	67.6

Appendix 2:

This appendix shows the statistical summaries for Hang Seng Index and FTSE 100 Index, respectively, complementing the results shown in Section 3.2, where we discussed the model performances for the entire market window for S&P 500.