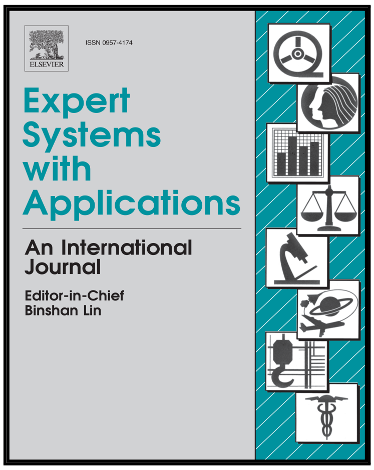
Accepted Manuscript



Forecasting Stock Market Index Daily Direction: a Bayesian Network Approach

Luciana S. Malagrino, Norton T. Roman, Ana M. Monteiro

PII: S0957-4174(18)30185-4
DOI: 10.1016/j.eswa.2018.03.039
Reference: ESWA 11886

To appear in: *Expert Systems With Applications*

Received date: 21 February 2017
Revised date: 26 February 2018
Accepted date: 22 March 2018

Please cite this article as: Luciana S. Malagrino, Norton T. Roman, Ana M. Monteiro, Forecasting Stock Market Index Daily Direction: a Bayesian Network Approach, *Expert Systems With Applications* (2018), doi: 10.1016/j.eswa.2018.03.039

Highlights

- A Bayesian Network is proposed to forecast the São Paulo Exchange main index's closing direction
- Network designed to reflect some intuitive dependencies amongst inter-continental markets
- Simple and intuitive model with results comparable to those of the related literature

Forecasting Stock Market Index Daily Direction: a Bayesian Network Approach

Luciana S. Malagrino[a], Norton T. Roman[a,*], Ana M. Monteiro[b]

[a] *University of São Paulo, São Paulo, Brazil*
[b] *Campo Limpo Paulista Faculty, Campo Limpo Paulista, Brazil*

Abstract

In this work, we investigate the feasibility of Bayesian Networks as a way to verify the extent to which stock market indices from around the globe influence iBOVESPA – the main index at the São Paulo Stock Exchange, Brazil. To do so, index directions were input to a network designed to reflect some intuitive dependencies amongst continental markets, moving through 24 and 48 hour cycles, and outputting iBOVESPA's next day closing direction. Two different network topologies were tested, with different numbers of stock indices used in each test. Best results were obtained with the model that accounts for a single index per continent, up to 24 hours before iBOVESPA's closing time. Mean accuracy with this configuration was around 71% (with almost 78% top accuracy). With results comparable to those of the related literature, our model has the further advantage of being simpler and more tractable for its users. Also, along with the fact that it not only gives the next day closing direction, but also furnishes the set of indices that influence iBovespa the most, the model lends itself both to academic research purposes and as one of the building blocks in more robust decision support systems.

Keywords: Stock direction prediction, Bayesian networks, Machine learning, Applied Artificial Intelligence

*Corresponding author. Address: EACH-USP. Arlindo Béttio, 1000, Ermelino Matarazzo, São Paulo-SP, Brazil – 03828-000. Phone: +55 (11) 2648 0130

Email addresses: `luciana.malagrino@usp.br` (Luciana S. Malagrino), `norton@usp.br` (Norton T. Roman), `anammont@cc.faccamp.br` (Ana M. Monteiro)

2010 MSC: 00-01, 99-00

1. Introduction

Every day, stocks are negotiated in stock markets, where investors decide whether to buy or sell shares of negotiated assets. As a way to measure the relative value of a determined group of stocks, be they all stocks in the market or only a portion of them, market indices are created (Lee & Lee, 2006). As the value of the stocks in this group changes over time, so does its corresponding index. Within this scenario, the ability to predict the closing direction of some stock market index allows investors to devise strategies for trading the stocks comprising that index, thereby increasing their potential for future profit.

In this research, we take an index closing direction to be the direction of the change in the index closing value at a specific day, when compared to its closing value in the day before. Hence, we take the direction to be positive (or high) if the index's closing value in the day before was lower than (or the same as) its current value, indicating that the stocks that build up the index have increased their average closing price. Similarly, a negative (or low) direction would imply a previous closing value higher than the current one, signalling some average loss in the index stock prices.

To date, there is a good number of computational strategies applied to both closing direction and value prediction of stocks and indices. These include the use of Artificial Neural Networks (*e.g.* Guresen et al. (2011)), Fuzzy Systems (*e.g.* Huarng & Yu (2005)), Support Vector Machines (*e.g.* Yeh et al. (2011)) and Decision Trees (*e.g.* Nair et al. (2010)), amongst others. In addition to the use of a specific machine learning technique, current approaches also include models that mix one or more of these techniques with sets of more classical technical indicators (*e.g.* Ticknor (2013); Patel et al. (2015)), such as Moving Average and Relative Strength Index, for example.

Despite the broad coverage in the use of different machine learning techniques, it is noticeable the scarceness of studies reporting the application of

Bayesian Networks in stock market prediction (*e.g.* Zuo & Kita (2012b,a)), even though this approach has already been applied to the modelling of portfolio returns (*e.g.* Shenoy & Shenoy (1999)) and for bankruptcy prediction (*e.g.* Sun & Shenoy (2007)). Moreover, there seems to be a preference, amongst those relying on machine learning approaches, for using training data only from the stock market about which predictions are to be made (*e.g.* Hadavandi et al. (2010); Kara et al. (2011)), thereby ignoring influences from outside sources, with only a few examples of studies being carried out across different markets (*e.g.* Hsieh et al. (2011)).

As a market forecasting methodology, Bayesian Networks have the advantage of not relying on Normal error distributions (Zuo & Kita, 2012a), as do other time-series algorithms used to this end (such as the Auto Regressive and Moving Average models, for example) and which may not accurately describe stock price behaviour. Additionally, they can deal both with continuous and discrete data, which makes them suitable for both price value and direction forecasting. As a learning methodology, such networks have the advantage of giving a human-readable account of dependencies amongst stock markets (when compared to other more complex models, such as Neural Networks, for example), whereby one can readily understand how strong dependencies are by verifying the probability of one market changing given that some other has changed.

This dependency across markets, in turn, and which seems to be out of the scope of much of the extant work on the use of machine learning techniques for stock price forecasting, has been subject of study in the field of Economics, where it has been noticed that some markets present time-varying dependencies (Tam & Tam, 2012; Urquhart & Hudson, 2013; Coronado et al., 2016), which might grow stronger after crisis periods (Puah et al., 2015; Tangpornpaiboon & Puttanapong, 2016). In fact, with the global integration of financial markets, via international trades, common currencies and cross-border investments (Jithendranathan, 2013), capital flows to where it will generate the highest return (Puah et al., 2015; Tangpornpaiboon & Puttanapong, 2016), making these markets more vulnerable to domestic shocks that could spread over the entire dependency network. Even though we are still far from ultimate full integration (Tam & Tam, 2012), to ignore existing dependencies could pose a threat to any forecasting system.

In this article, we seek to help fulfil this gap, by introducing a Bayesian Network model of the Brazilian Stock Exchange market (BM&FBOVESPA)[1], relating its main index (iBOVESPA) to other indices around the world. Founded in 2008, from the merging of the São Paulo Stock Exchange (BOVESPA) and the Brazilian Mercantile and Futures Exchange (BM&F), BM&FBOVESPA has become the main Brazilian company for the management of organized security and derivative markets. Its main index, iBOVESPA[2], is a weighted average of a theoretical portfolio comprising roughly the 50 more actively traded stocks, which was designed to reflect the market's average performance by tracking price changes of its more representative participants.

In our analysis, we accounted for market dependencies by taking a "follow the sun" approach, whereby we hypothesise that, moving along the sequence of closing times in markets around the globe, every closing market influences the next one, thereby simulating capital and news flow over time. We have then experimented two competing models: one which only accounted for closing directions within a 24-hour period; and another using a 48-hour time window. Stock markets were grouped up by continent, and tests were made with sets of one to three market indices per continent. In doing so, our main objective was to test the feasibility of these Bayesian Networks to (i) identify markets and, more specifically, market indices, that influence iBOVESPA the most; and (ii) determine the best time window for this analysis. Also, to the best of our knowledge, this is the first machine learning based study to account for the dependency of stock indices around the world.

Given the conceptual clearness brought by Bayesian Networks, where depen-

[1] http://www.bmfbovespa.com.br/en_us/index.htm

[2] http://www.bmfbovespa.com.br/en_us/products/indices/broad-indices/bovespa-index-ibovespa.htm

dencies are clearly put, and results can be understood in terms of frequencies, our method lends itself to be used with Decision Support Systems for the stock market, whereby a human must decide what to do based on the available information. In this case, a clear account of how the system comes up with its advices might help its users to take or refuse them with more confidence. Also, results from our work could be adapted by researchers to studies in other areas, such as Economics, for instance, furnishing a way to quantify, in terms of probabilities, which indices have more influence on some specific market. Finally, another possibility would be to attach our network to some more complex model, such as a machine learning ensemble for example, or other more classical approaches to the problem, so as to provide traders with a more complete tool set for decision making.

The rest of this article is organised as follows. Section 2 describes our experimental set-up, presenting the network topologies, stock market indices and data used in our experiments. Next, in Section 3, we present the results, in terms of accuracy[3], of the competing models. In this section, we also make a comparison between them, pointing out possible reasons for differences in the models. Section 4, in turn, discusses our results further, giving some possible applications of this approach, along with a theoretical account of our model's strengths and weaknesses. Finally, in Section 5, we compare our results to those of the related literature, while in Section 6 we present our final remarks on this research, along with directions for future improvement.

2. Materials and Methods

The stock market indices used in this research comprise 12 indices worldwide. These are Nasdaq Composite (USA), NYSE Composite (USA), Dow Jones (USA), Merval (Argentina), Nikkei 225 (Japan), Shangai Composite (China), Hang Seng (China), BSE 30 Sensex (India), FTSE 100 (UK), Stockholm Gen-

[3] *I.e.* the proportion of correctly guessed directions (both positive and negative).

eral (Sweden), Dax (Germany) and Cac 40 (France). Table 1 shows the trading period for each market[4], along with their overlapping time with iBOVESPA. Data was collected from 01/06/2005[5] to 05/04/2012 at Yahoo! Finance's website[6], corresponding to a 2,501-day period. Days in which any of the selected markets were closed (due to local holidays, for example) were removed from the data set for all markets, resulting in a total of 1,620 days used in this research for each of the selected indices.

Table 1: Stock markets opening time, ranked by overlapping time with iBOVESPA.

Index	*Country*	Opening Time (UTC)	Closing Time (UTC)	Overlapping Time (h)
iBOVESPA	Brazil	13:00	20:00	7:00
Merval	Argentina	14:00	20:00	6:00
Dax	Germany	07:00	19:00	6:00
Nasdac Comp.	USA	14:30	21:00	5:30
NYSE Comp.	USA	14:30	21:00	5:30
Dow Jones	USA	14:30	21:00	5:30
Cac 40	France	08:00	16:30	3:30
FTSE 100	UK	08:00	16:30	3:30
Stockholm Gen.	Sweden	08:00	16:30	3:30
Nikkei 225	Japan	00:00	06:00	0:00
Shangai Comp.	China	01:30	07:00	0:00
Hang Seng	China	01:30	08:00	0:00
BSE 30 Sensex	India	03:45	10:00	0:00

The date removal procedure was carried out as follows. First, data from the entire 2,501 running days period was collected. We then removed all dates in

[4] `https://en.wikipedia.org/wiki/List_of_stock_exchange_trading_hours`. For a map of opening times, we refer the reader to `https://www.investing.com/tools/market-hours`.

[5] Date format is dd/mm/yyyy.

[6] `https://finance.yahoo.com/`

which at least one of the markets was closed (*i.e.* holidays and weekends). Even though the number of weekends (714 in total, 357 Saturdays and 357 Sundays) was shared by all markets, the number of holidays varied from market to market, reflecting local holiday policies. Hence, we kept the maximum amount of days where we could obtain data from all markets under consideration, which amounted to 1,620 days. For the missing days, no correction was made for price moves, since we did not want to introduce any artificial data into the model. The implications of this decision will be discussed further in Section 3.

Next, we designed our Bayesian Network (see Section 2.1), which was trained and evaluated in the data set using 5-fold cross validation. Within this model, one randomly splits the data in five different subsets (also called folds) of approximately equal size (Kohavi, 1995), which will be held out one at a time for validation/testing purposes. Each time, the network is trained in four of the sets (*i.e.* 80% of the data) and evaluated in the fifth one (the remaining 20%) for its accuracy, that is the number of correct classifications, divided by the number of instances in the test set[7]. This process is repeated exactly five times, with a different test set (and, consequently, different training sets) used each time, in an attempt to reduce variance (Hastie et al., 2009). The network's overall accuracy is taken to be the average of its accuracy across the five test subsets.

To better evaluate the model, we experimented with sets of one, two, and three different indices per continent, along with 24 and 48-hour time windows (that is, looking at data up to 24 and 48 hours in the past). We tried then to predict, from the data within these time windows, iBOVESPA's next day closing direction. To do so, we fit the 24 and 48-hour sliding windows to the data, thereby having a training instance to comprise iBOVESPA's closing direction at a specific date, along with past closing directions for a subset of different markets. These markets were selected according to the experimental set-up, by varying the number of indices per continent, time-window length, and network

[7]It is important to notice that, in this work, we are focusing on directional (binary – up and down) prediction only, instead of index quantification.

topology (which will be made clearer in the next section).

As an example, in the 24-hour model with one index per continent, a training instance would be {iBOVESPA's closing direction at day t; American index's closing direction at t-1; European index's closing direction at t-1; Asian index's closing direction at t-1}. These training instances were built on a daily basis, and then input to the network. From our testing results, we identified the indices with the best and worst performance in predicting iBOVESPA's closing direction, as measured by their accuracy. That allowed us to identify the sets of indices with the highest and lowest influence on iBOVESPA and to evaluate competing models for their accuracy with the best indices. In what follows, we will detail the network topologies used in the experiments.

2.1. Network Topologies

To build our Bayesian Network, we hypothesised that, moving through 24 and 48 hour cycles, and starting from the first stock exchange market to close after BM&FBOVESPA, every closing market influences the next one around the globe, back to iBOVESPA's next day closing value. Under this assumption, each market index closing direction is influenced by those markets that closed before it. For the sake of simplicity, we grouped up markets according to their home continent (*i.e.* Americas, Europe and Asia), since closing time is actually related to their geographical location on the planet, and carried out experiments with one, two and three different indices per continent.

In the first phase of our experiment, we observed a 24-hour time lapse before BM&FBOVESPA's closing time. Figure 1 illustrates this network topology, where SP_{00} means iBOVESPA's closing direction at a specific date, and $Asia_{24}$, $Europe_{24}$ and $America_{24}$ stand for the closing directions of market indices in these continents up to 24 hours before SP_{00}'s closing. Since we account for indices that close up to 24 hours before iBOVESPA's closing, that includes both markets that close before BM&FBOVESPA opened (but still within the 24-hour limit) and those closing within its trading period. In this model, a training instance comprises iBOVESPA's closing direction, along with the closing directions of all of its parent nodes in the network.

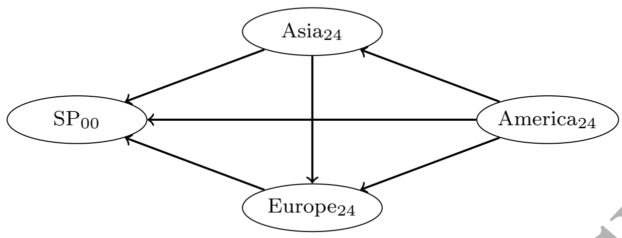


Figure 1: 24 hour network topology for the first phase of the experiment.

In the experiments with more than one index per continent, all continent nodes were replaced by their respective set of indices. In this case, index nodes within the same continent were taken to be independent of each other, whereas dependencies were kept amongst intercontinental nodes (*i.e.* nodes in different time zones). Hence, a dependency between Europe and Asia now corresponds to all indices in Europe depending on all indices in Asia. It is important to notice that we take dependencies to exist in time, and not necessarily on geographical location. That means nodes in the same continent are held independent of each other only at a specific time, but not before it. As shown in Figure 1, even though SP_{00} lies in the American continent, it is independent of other American markets only by the time of the prediction (what would be $\mathrm{America}_{00}$). However, dependencies are taken to exist when we look at these same markets up to 24 hours in the past (*i.e.* $\mathrm{America}_{24}$).

To mathematically model this network, we followed Bayes' theorem, according to which

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)}$$

where h_i is the hypothesis that iBOVESPA belongs to class $C_i = \{high, low\}$, that is iBOVESPA's closing value will be *higher* (h_1) or *lower* (h_2), respectively, when compared to its closing value the day before, and $P(h_i|d)$ is the probability of hypothesis h_i holding given the data set d. In the case of a Bayesian Network, constructed under the assumption of conditional independence between a node and its predecessors in the network, given the node's parents, this equation reduces to

$$P(h_i|d) = P(h_i|\textit{iBOVESPA's parents})$$

where $P(h_i|\textit{iBOVESPA's parents})$ can be estimated by the frequency with which h_i holds true for every combination of values for iBOVESPA's parents. The same method can be used for every node in the network, thereby assigning them probability distributions, according to which one can estimate the probability of observing a specific direction in one node, given the directions observed in its parents. Also, as a way to avoid assigning null probabilities to unseen combinations of parent directions, we have added a Laplace smoothing term to the equation, whereby one adds 1 to the numerator and 2 to the denominator when calculating the frequency used to estimate $P(h_i|\textit{iBOVESPA's parents})$.

In the second phase of the experiment, we repeated this methodology with a 48 hour time window, as shown in Figure 2. Within this model, new nodes were added, to represent closing directions for indices both up to 24 and 48 hours before iBOVESPAS's closing time, roughly corresponding to two spins around the globe towards the past. Under this approach, iBOVESPA's closing direction is taken to be directly influenced only by stock markets closing up to 24 hours before it, while indirectly influenced by those closing up to 48 hours in the past.

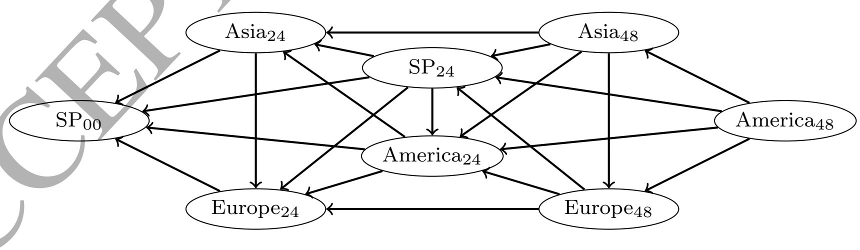


Figure 2: 48 hour network topology for the second phase of the experiment.

The mathematical model for this network, in turn, follows the same expression as that of the 24 hour model, except that, in this case, instead of three parent nodes, iBOVESPA has a fourth one – its closing direction in the day before (SP_{24} in the figure). As such, all value combinations for these four parents must be taken into account when estimating $P(h_i|iBOVESPA\text{'s parents})$. Another difference between models is that we took $iBOVESPA_{24}$ to influence other American markets that close after its closing time, even though they lie within $iBOVESPA_{24}$'s time frame, thereby breaking the independence assumption only in this case. In this topology, we took iBOVESPA out of its continental group as a way to reinforce the assumption that its closing direction in the past plays a major role in its future closing direction, hence accommodating for the possibility of internal dependencies outweighing external ones. Once again, we have experimented with up to three indices per continent.

3. Results

Experiments with the 24-hour topology (Figure 1) and a single index per continent resulted in a mean accuracy of 71.08%. In a total of 64 runs[8], where a different arrangement of indices was tested each run, accuracy ranged from 62.79%, with the Dow Jones, Stockholm General and Nikkei 225 combination, to 77.78%, with NYSE Composite, Cac 40 and Hang Seng. Figure 3 shows accuracy distribution within this range. As can be seen, even though median accuracy was 72.10%, peak values lied around 73%, which was the accuracy obtained by 21 different arrangements in our tests.

Second best accuracy, 75.49%, was obtained with NYSE Composite, FTSE 100 and Nikkei 225. However close, the difference between best and second best accuracies was found to be statistically significant, as determined by an unpaired t-test[9] run on the individual folds ($t_{Welch}(n = 5) = 2.33, p = 0.05$), indicating that, under this model, NYSE Composite, Cac 40 and Hang Seng could be the indices that influence iBovespa's closing direction the most[10]. Ta-

[8] 4^3 – four different combinations per continent, arranged amongst the three continents.

[9] All test results are reported at the 95% confidence level.

[10] Even though a different arrangement of data within folds might lead to a different set of

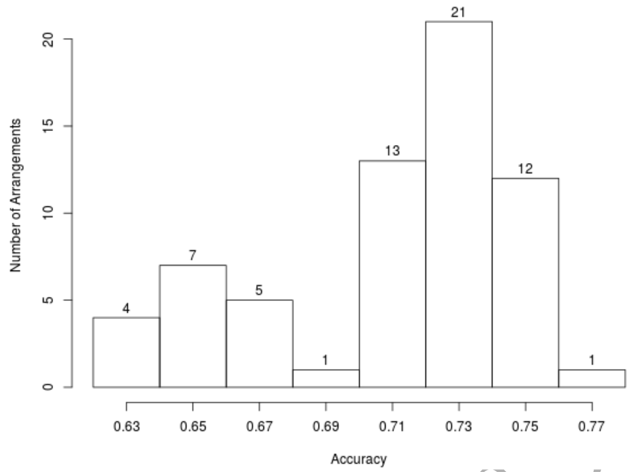


Figure 3: Accuracy distribution in 64 runs for the single index 24h model.

ble 6 (see the Appendix section) shows mean accuracy across the five folds for each market combination in our 24-hour single index per continent model, along with the mean number of hits (*i.e.*, correctly guessed directions[11]) and misses (*i.e.*, incorrectly guessed directions[12]).

Arrangements with two indices per continent led us to a mean accuracy of 70.20% and a 70.32% median value. In a total of 216 runs[13], accuracy ranged from 66.08%, with Dow Jones, Merval, Dax, Cac 40, Nikkei 225 and Shangai Composite, to 74.01%, with NYSE Composite, Merval, Dax, Stockholm General, Shangai Composite and BSE 30 Sensex. Figure 4 shows accuracy distribution within this range. Interestingly, the distribution closely resembles that of a normal curve, with a peak around 70.50%, delivered by 54 arrangements (*i.e.* 25% of the 216).

Second best accuracy (73.58%) was achieved with NYSE Composite, Merval, FTSE 100, Dax, Shangai Composite and BSE 30 Sensex. The difference

indices, so further research on this topic must be done.

[11]Alternatively, the sum of true positives and true negatives.

[12]Alternatively, the sum of false positives and false negatives.

[13]6^3 – six different combinations each continent, arranged amongst the three continents.

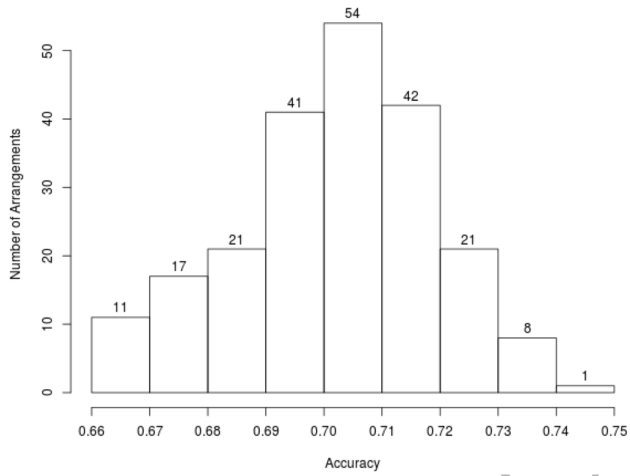


Figure 4: Accuracy distribution in 64 runs for the two-index-24h model.

between best and second best accuracies was, however, not found to be of statistical significance ($t_{Welch}(n = 5) = 0.48, p = 0.65$). Noticeably, three of the indices (Merval, Shangai Composite and Dax) can be found in the arrangement with best, second best and worst accuracy (see Table 2 for a comparison of indices in best, second best and worst accuracy arrangements across models). We understand this to be an indication that these indices are being forced into the model by the two indices per continent demand, thereby introducing noise. This could also explain why the best accuracy arrangement with two indices per continent scored lower than its single index counterpart, a difference found to be of statistical significance ($t_{Welch}(n = 5) = -4.08, p < 0.01$).

Finally, arrangements with three indices per continent resulted in a mean and median accuracy of 56.45%. Through 64 different index arrangements[14], accuracy ranged from 52.13%, with NYSE Composite, Dow Jones, Merval, FTSE 100, Stockholm General, Cac 40, Nikkei 225, Shangai Composite and Hang Seng, to 61.17%, with Nasdaq Composite, NYSE Composite, Merval, FTSE 100, Dax, Cac 40, Nikkei 225, Shangai Composite and Hang Seng. Figure 5

[14] 4^3 – four different combinations each continent, across three continents.

shows accuracy distribution within this range. As with the two indices per continent model, so this distribution resembles the normal curve. This could be another indicative that adding more indices did in fact add random noise, thereby reducing accuracy.

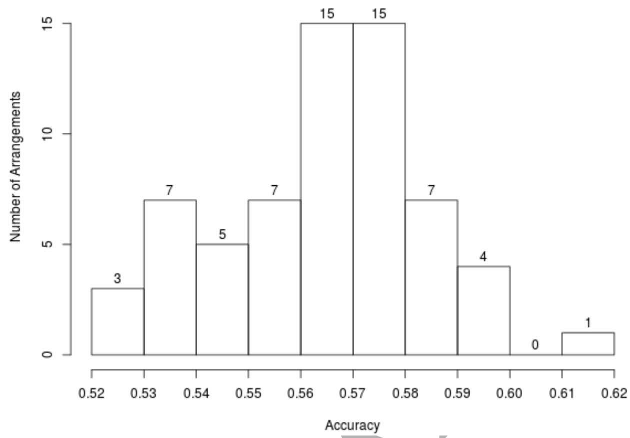


Figure 5: Accuracy distribution in 64 runs for the three-index-24h model.

This argument is further evidenced by the small difference between best and second best accuracies, which was not found to be statistically significant ($t_{Welch}(n = 5) = 2.10, p = 0.08$). In this case, second best accuracy (59.80%) was reached with Nasdaq Composite, NYSE Composite, Dow Jones, FTSE 100, Dax, Cac 40, Nikkei 225, Shangai Composite and Hang Seng. Moreover, from the nine indices in the best accuracy arrangement, six (NYSE Composite, FTSE 100, Cac 40, Nikkei 225, Shangai Composite and Hang Seng) could also be found in the second best and worst accuracy arrangements, a sign that they are but noise in this model.

One puzzling result of the three indices per continent model, as shown in Table 2, is the fact that Nasdaq Composite, which does not occur in best, second best and worst positions in the single and double index models, suddenly comes up as part of the best and second best, but not the worst, scoring arrangements. At first glance, that might indicate an interesting behaviour about that index. However, in looking at the second worst accuracy arrangement (52.95%, with

Table 2: Best, second best and worst accuracy indices with 1, 2 and 3 indices per continent in the 24h model.

Ind./Cont.	*Best*	2^{nd} *Best*	*Worst*
1	NYSE Comp., Cac 40, Hang Seng	NYSE Comp., FTSE 100, Nikkei 225	Dow Jones, Stockholm Gen., Nikkei 225
2	NYSE Comp., Merval, Dax, Stockholm Gen., Shangai Comp., BSE 30 Sensex.	NYSE Comp., Merval, FTSE 100, Dax, Shangai Comp., BSE 30 Sensex.	Dow Jones, Merval, Dax, Cac 40, Nikkei 225, Shangai Comp.
3	Nasdaq Comp., NYSE Comp., Merval, FTSE 100, Dax, Cac 40, Nikkei 225, Shangai Comp., Hang Seng	Nasdaq Comp., NYSE Comp., Dow Jones, FTSE 100, Dax, Cac 40, Nikkei 225, Shangai Comp., Hang Seng	NYSE Comp., Dow Jones, Merval, FTSE 100, Stockholm Gen., Cac 40, Nikkei 225, Shangai Comp., Hang Seng

Nasdaq Composite, Dow Jones, Merval, FTSE 100, Dax, Stockholm General, Nikkei 225, Shangai Composite and BSE 30 Sensex) we see Nasdaq there. Since no difference could be found between worst and second worst arrangements ($t_{Welch}(n = 5) = -0.38, p = 0.71$), we understand this index to be noise as well.

Finally, here too we find the difference between best accuracy with three indices per continent to be statistically relevant, when compared to its counterparts in the two indices ($t_{Welch}(n = 5) = 24.10, p < 0.01$) and single index per continent ($t_{Welch}(n = 5) = -18.74, p < 0.01$) models. Also, another interesting result is that, even though NYSE Composite might be thought of as the most influential index, given its importance worldwide, it happens both in best and worst arrangements in the two last models[15]. This could be an indicative that, however important, this index' influence to iBOVESPA is not strong enough to make the difference when more indices are added to the model.

Moving on to the 48-hour topology (Figure 2), we rerun the experiments using combinations with one, two and three indices per continent. Results with a single index led to a mean accuracy of 68.17%, with a 69,14% median value. Throughout the 4,096 runs of the experiment[16], accuracy ranged from 54.34%, with Dow Jones_{24}, Stockholm General_{24}, BSE 30 Sensex_{24}, Merval_{48}, FTSE 100_{48} and Shangai Composite_{48}, to 75.17%, with Merval_{24}, Stockholm General_{24}, Hang Seng_{24}, Dow Jones_{48}, CAC 40_{48} and Hang Seng_{48}.[17] Figure 6 shows accuracy distribution within this range. As with the single index 24h model, so this distribution presents pronounced peaks around mean and median values.

Second best accuracy was 75.12%, with NYSE Composite_{24}, Dax_{24}, Hang Seng_{24}, Nasdaq Composite_{48}, Dax_{48} and Hang Seng_{48}. The difference between best and second best accuracies was once again not found to be of statistical significance ($t_{Welch}(n = 5) = 7.86, p = 0.96$). As with the 24 hour model, only one index could be found in best and worst accuracy arrangements. In this case, however, it was not Nikkei 225 (second best and worst arrangements in the 24 hour model), but Stockholm General_{24} instead. Finally, differently from the 24 hour topology, NYSE Composite_{24}, which happens in first and second place in that model, comes here as part of the second best accuracy arrangement, even though with no significant difference to the best accuracy setting. We refer the

[15]In the double-index model it comes in the 32^{nd} worst arrangement, which was not found to significantly differ from the worst scoring arrangement ($t_{Welch}(n = 5) = -1.11, p = 0.31$).

[16]4^6 – four different combinations per continent, arranged amongst the three continents, with two spins around the globe.

[17]To better understand these figures, recall that the "24" subscript refers to an index closing value up to 24 hours before iBOVESPA's, while the "48" subscript refers to its value from 24 to 48 hours before iBOVESPA. Also, as shown in Figure 2, iBOVESPA's closing value the day before (*i.e.* SP_{24}) is a constant index in our model, the reason why it is omitted in our results.

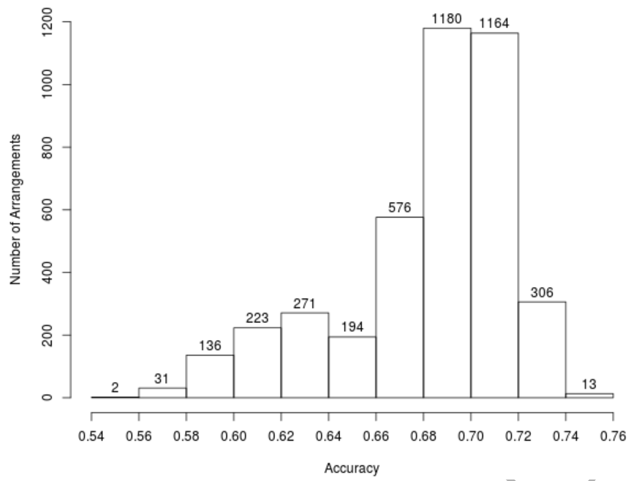


Figure 6: Accuracy distribution in 4,096 runs for the single index 48h model.

interested reader to Table 3 for a comparison of indices in best, second best and worst accuracy arrangements with the 48 hour topology.

In comparing the best accuracy value within this model to those obtained with the 24 hour topology, we see significant differences between this value and results from the single index 24h ($t_{Welch}(n = 5) = -2.39, p = 0.04$) and three indices per continent 24h ($t_{Welch}(n = 5) = 17.83, p < 0.01$) models. Nevertheless, no statistically significant difference could be established between best accuracies in the 48h single index and 24h double index models ($t_{Welch}(n = 5) = 1.41, p = 0.21$). We understand this to be another evidence of the reduction in best accuracy value due to the introduction of more indices into the model, even if they are the same indices, but within a wider time window.

Arrangements with two indices per continent delivered a mean accuracy of 31.04% (30.94% median). In a total of 46,656 runs[18], accuracy ranged from 20.19% with Nasdaq Composite$_{24}$, Merval$_{24}$, Stockholm General$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Composite$_{48}$, Merval$_{48}$, FTSE 100$_{48}$,

[18] 6^6 – six different combinations each continent, arranged amongst the three continents twice around the globe.

Table 3: Best, second best and worst accuracy indices with 1, 2 and 3 indices per continent in the 48h model.

Ind./Cont.	*Best*	2^{nd} *Best*	*Worst*
1	Merval$_{24}$, Stockholm Gen.$_{24}$, Hang Seng$_{24}$, Dow Jones$_{48}$, CAC 40$_{48}$, Hang Seng$_{48}$	NYSE Comp.$_{24}$, Dax$_{24}$, Hang Seng$_{24}$, Nasdaq Comp.$_{48}$, Dax$_{48}$, Hang Seng$_{48}$	Dow Jones$_{24}$, Stockholm Gen.$_{24}$, BSE 30 Sensex$_{24}$, Merval$_{48}$, FTSE 100$_{48}$ and Shangai Comp.$_{48}$
2	NYSE Comp.$_{24}$, Merval$_{24}$, Stockholm Gen.$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, Hang Seng$_{24}$, Nasdaq Comp.$_{48}$, Dow Jones$_{48}$, Dax$_{48}$, Stockholm Gen.$_{48}$, Hang Seng$_{48}$, BSE 30 Sensex$_{48}$	Nasdaq Comp.$_{24}$, NYSE Comp.$_{24}$, Dax$_{24}$, Cac 40$_{24}$, Shangai Comp.$_{24}$, Hang Seng$_{24}$, NYSE Comp.$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Cac 40$_{48}$, Shangai Comp.$_{48}$, Hang Seng$_{48}$	Nasdaq Comp.$_{24}$, Merval$_{24}$, Stockholm Gen.$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Comp.$_{48}$, Merval$_{48}$, FTSE 100$_{48}$, Stockholm Gen.$_{48}$, Nikkei 225$_{48}$, Shangai Comp.$_{48}$
3	Nasdaq Comp.$_{24}$, NYSE Comp.$_{24}$, Dow Jones$_{24}$, FTSE 100$_{24}$, Dax$_{24}$, Cac 40$_{24}$, Shangai Comp.$_{24}$, Hang Seng$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Comp.$_{48}$, NYSE Comp.$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Dax$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Hang Seng$_{48}$, BSE 30 Sensex$_{48}$	Nasdaq Comp.$_{24}$, NYSE Comp.$_{24}$, Dow Jones$_{24}$, FTSE 100$_{24}$, Stockholm Gen.$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, Hang Seng$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Comp.$_{48}$, NYSE Comp.$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Dax$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Hang Seng$_{48}$ and BSE 30 Sensex$_{48}$	Nasdaq Comp.$_{24}$, Dow Jones$_{24}$, Merval$_{24}$, FTSE 100$_{24}$, Dax$_{24}$, Stockholm Gen.$_{24}$, Nikkei 225$_{24}$, Shangai Comp.$_{24}$, Hang Seng$_{24}$, Nasdaq Comp.$_{48}$, NYSE Comp.$_{48}$, Merval$_{48}$, Dax$_{48}$, Stockholm Gen.$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Shangai Comp.$_{48}$, BSE 30 Sensex$_{48}$

Stockholm General$_{48}$, Nikkei 225$_{48}$ and Shangai Composite$_{48}$, to 45.27%, with NYSE Composite$_{24}$, Merval$_{24}$, Stockholm General$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, Hang Seng$_{24}$, Nasdaq Composite$_{48}$, Dow Jones$_{48}$, Dax$_{48}$, Stockholm General$_{48}$, Hang Seng$_{48}$, BSE 30 Sensex$_{48}$. Figure 7 shows accuracy distribution within this range. Once again, we are close to the normal shape, with a clear peak at the 0.31 accuracy, registered by 9,680 different index arrangements.

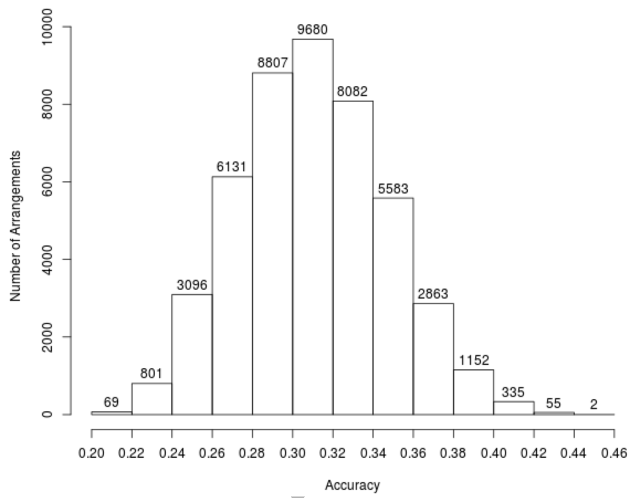


Figure 7: Accuracy distribution in 46,656 runs for the double index 48h model.

Second best accuracy (44.49%) was reached with Nasdaq Composite$_{24}$, NYSE Composite$_{24}$, Dax$_{24}$, Cac 40$_{24}$, Shangai Composite$_{24}$, Hang Seng$_{24}$, NYSE Composite$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Cac 40$_{48}$, Shangai Composite$_{48}$ and Hang Seng$_{48}$. The difference between best and second best accuracies was also not found to be of statistical significance ($t_{Welch}(n = 5) = 0.70, p = 0.50$). Once more, half the indices in the best accuracy arrangement come also in the worst accuracy one. These are Merval$_{24}$, Stockholm General$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, Nasdaq Composite$_{48}$ and Stockholm General$_{48}$. Regarding the difference between the best accuracy arrangement with two indices per continent around 48 hours and other models, we found it to significantly astray from best accuracy arrangements with a single index and the same topology ($t_{Welch}(n = 5) = 31.61, p < 0.01$), and with the 24h single ($t_{Welch}(n = 5) = -31.53, p < 0.01$), double ($t_{Welch}(n = 5) = -38.34, p < 0.01$) and triple ($t_{Welch}(n = 5) = -22.59, p < 0.01$) index models.

Arrangements with three indices per continent resulted in a mean accuracy of 8.21%, with an 8.07% median. Across 4,096 runs of the experiment[19], accuracy ranged from 4.73%, with Nasdaq Composite$_{24}$, Dow Jones$_{24}$, Merval$_{24}$, FTSE 100$_{24}$, Dax$_{24}$, Stockholm General$_{24}$, Nikkei 225$_{24}$, Shangai Composite$_{24}$, Hang Seng$_{24}$, Nasdaq Composite$_{48}$, NYSE Composite$_{48}$, Merval$_{48}$, Dax$_{48}$, Stockholm General$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Shangai Composite$_{48}$ and BSE 30 Sensex$_{48}$, to 14.27%, with Nasdaq Composite$_{24}$, NYSE Composite$_{24}$, Dow Jones$_{24}$, FTSE 100$_{24}$, Dax$_{24}$, Cac 40$_{24}$, Shangai Composite$_{24}$, Hang Seng$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Composite$_{48}$, NYSE Composite$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Dax$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Hang Seng$_{48}$, BSE 30 Sensex$_{48}$. Figure 8 shows accuracy distribution within this range.

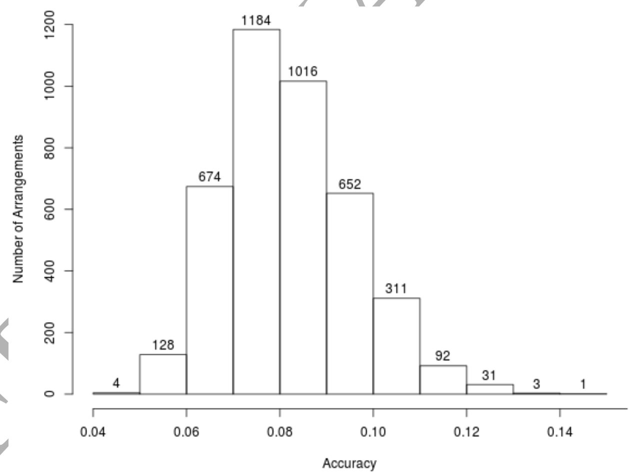


Figure 8: Accuracy distribution in 4,096 runs for the three index 48h model.

It is noticeable in the figure the distribution's skew towards the lower side of

[19]4^6 – four different combinations per continent, arranged amongst the three continents, twice around the globe.

the accuracy range, making this model useless for our prediction. As it seems, in adding more indices to the 48 hour topology we not only added more noise, but had the model systematically fail in its predictions, to the extent that accuracy would increase considerably if we just negate its output. As can be verified in Table 3, 12 out of the 18 indices in this model figure both in best and worst accuracy arrangements. This is exactly the same proportion we found in the 24 hour topology. This great amount of noise being input to the system could be the reason for its failure: it might be the case that it is actually learning some hidden pattern in this noise.

Second best accuracy, in turn, was 13.68%, with Nasdaq Composite$_{24}$, NYSE Composite$_{24}$, Dow Jones$_{24}$, FTSE 100$_{24}$, Stockholm General$_{24}$, Cac 40$_{24}$, Nikkei 225$_{24}$, Hang Seng$_{24}$, BSE 30 Sensex$_{24}$, Nasdaq Composite$_{48}$, NYSE Composite$_{48}$, Dow Jones$_{48}$, FTSE 100$_{48}$, Dax$_{48}$, Cac 40$_{48}$, Nikkei 225$_{48}$, Hang Seng$_{48}$ and BSE 30 Sensex$_{48}$. The difference between best and second best accuracies, as expected, was not statistically significant ($t_{Welch}(n = 5) = 0.89, p = 0.41$). When accounting for best accuracy arrangements only, we found accuracies under this model to significantly differ from those with one ($t_{Welch}(n = 5) = 64.94, p < 0.01$) and two ($t_{Welch}(n = 5) = -35.61, p < 0.01$) indices, around 48 hours, and with one ($t_{Welch}(n = 5) = -62.04, p < 0.01$), two ($t_{Welch}(n = 5) = -80.79, p < 0.01$) and three ($t_{Welch}(n = 5) = -67.69, p < 0.01$) indices in the 24 hour topology.

Finally, when comparing mean, worst and best accuracies across all models (Table 4), we notice a strong correlation between the amount of indices used in each model and its mean accuracy ($pearson = -0.98, t(df = 4) = -8.98, p < 0.01$, at the 95% confidence level[20]), to the extent that as the total number of indices increase, mean accuracy decreases, as shown in Figure 9. Similar correlations were also found between the total amount of indices and worst ($pearson = -0.96, t(df = 4) = -6.86, p < 0.01$) and best ($pearson = -0.98, t(df = 4) = -9.74, p = 0.01$) accuracy values, once again ev-

[20] All pearson coefficient values reported here were obtained at the 95% confidence level.

idencing the introduction of noise as market indices are brought into the model.

Table 4: Mean, worst and best accuracies across models, along with the total amount of indices involved.

	Indexes / Continent	*Total Amount of Indexes*	*Accuracies (%)* *Mean*	*Worst*	*Best*
24h	1	3	71.08	62.79	77.78
	2	6	70.20	66.08	74.01
	3	9	56.45	52.13	61.17
48h	1	7	68,17	54.34	75.17
	2	13	31.04	20.19	45.27
	3	19	8.21	4.73	14.27

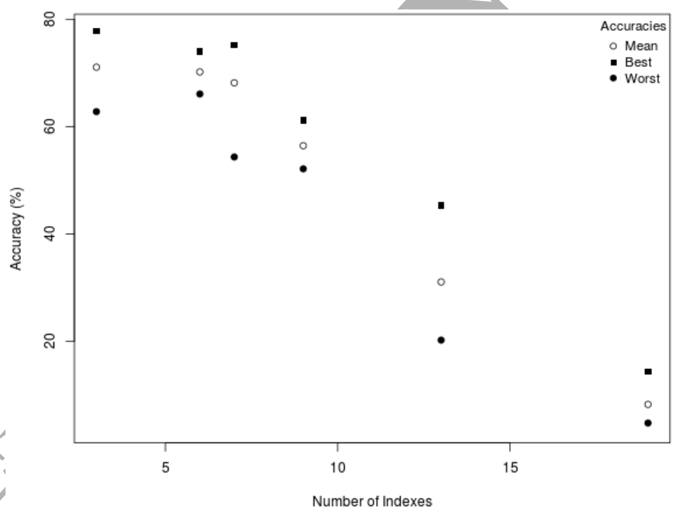


Figure 9: Mean, best and worst accuracies across models.

4. Discussion

The fact that the 48-hour topology performed worse than its 24-hour counterpart (*cf.* Table 4), for each proportion of indices per continent, could be an indicative of the reduction in the strength of other markets' influence on iBOVESPA over time, making these markets behave more like noise as we move away from our prediction date towards the past. This is somewhat in line with current research in Economics (*e.g.* Tam & Tam (2012); Urquhart & Hudson (2013); Coronado et al. (2016)), where such influences were noticed, but also were seem to vary over time. With our model, one can not only identify the existence of these influences, but also quantify them and verify at what rate they decrease with time, up to the point where they cannot be told apart from noise. That could in turn be used to guide investment strategies, with the advantage of being easy to understand, since one can explain the results in terms of probabilities and frequencies.

Another possible use for this model, and that comes directly from the identification of which markets influence iBOVESPA the most, is the reduction of the possible candidates for market contamination during crises. In this case, the most influential markets would be rendered more likely to pass the crisis effects on to the target market (in this case, BM&FBOVESPA). Conditional probabilities of the network could then be used as estimates of the amplitude of this contamination. Naturally, given the changing nature of markets around the world, with economies growing in importance at different paces, one cannot expect those markets that influence some index today to remain the same to play this role in the future. Fortunately, training a new network is just a matter of carrying out a new analysis of frequencies, making this model easily adaptable to changes.

This adaptability, along with the fact that there is no reason to believe this model could not be applied to any other stock market index in the world, given some findings about market dependencies (*e.g.* Puah et al. (2015); Kim & Song (2017)), adds to its robustness. Also, and as already pointed out, its visual and more "human friendly" design has the theoretical advantage of making it more easily adaptable to other circumstances, when compared to more "black-box"-like models, such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), for instance. Like Decision Trees, the underlying structure of our model's reasoning is clear to any human inspector. Unlike such trees, however, their fundamentals are more readily understood, for they are given in terms of frequency of co-occurrences.

On the downside, the rigid topology presented by Bayesian Networks makes them less adaptable to structural changes (*e.g.* when dependencies raise inside continents, thereby breaking the independence assumption). One possible solution to this problem would be to automatically learn the network's structure. That, however, not rarely leads to a Naïve Bayes structure (*e.g.* Zuo & Kita (2012a); Kita et al. (2012)), whereby all nodes are taken to be independent, given the value of the output node. Also, the fact that our model was trained and tested in data comprising part of the sub-prime crisis might have introduced a positive bias towards it, for it has already been reported that dependencies between markets tend to raise in periods of crisis (*e.g.* (Tangpornpaiboon & Puttanapong, 2016)), to the extent that some correlations exist only during such periods. If that is actually the case, it might be that our model was favoured by the crisis period, which in turn reinforced the dependencies that form the basis of the model's structure.

However, when analysing closing values for the indices with best results (*i.e.* indices that build the network with higher accuracy, to wit NYSE Composite, Cac 40 and Hang Seng) in the full dataset period, one sees these dependencies as something not so obvious at first glance, as shown in Figure 10. More importantly, even though there is a steep decline in iBOVESPA during the second half of 2008, with a corresponding incline in the next year, which would in turn make predictions easier, those periods add up to approximately 20% of the entire dataset, something we believe would not turn out as a determinant factor for our accuracy results, specially when one sees a good deal of fluctuations along the period, with a clear inflection point about one third of the way.

Finally, another potential drawback of our model lies in our methodological decision to not compensate for missing days. As pointed out in Section 2, we removed from the data set all dates where at least one of the markets was closed. Even though that has no effect in the actually closed markets, it has some effects on those that remained open. More specifically, for those markets, at those

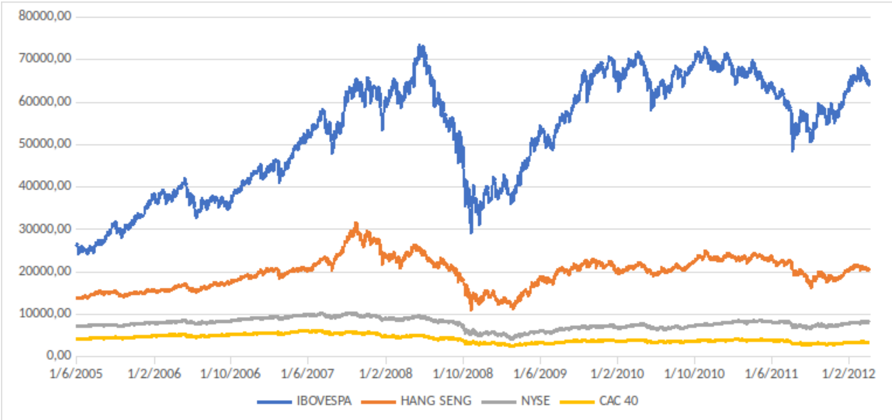


Figure 10: Closing value of the indices in the best accuracy network (24h-single index model), for the full dataset period.

dates, the 24-hour (or 48-hour, depending on the model) window was broken, and data from the day before the removed date was used instead. Although we did so to avoid introducing artificial data (such as by repeating the previous day direction only for the closed markets, for example), we ended up introducing an artificial one-day period of stability to all markets.

The question, however, is to what extent this decision undermined our model. To start with, if we remove all 714 weekend days from the 2,501 running days in the period, we end up with 1,787 days. In total, we removed other 167 holidays (*i.e.*, around 9% of the remaining data set), keeping 1,620 days of data for training and testing. Considering that not all markets would be open at those 167 days (specially in holidays such as Christmas, for example, that are celebrated in many countries), that does not seem to be so high a proportion to make any real difference. Nevertheless, our decision may have actually created a negative bias to our model, since we overlook 24-hour dependencies in favour of their 48-hour counterparts and our results point at the direction of a weakening of dependencies along the time. In this case, we have exchanged a potentially stronger dependency for a weaker surrogate, which in turn would be prejudicial to the model. Even though we do not think this to be an issue, we understand that only future tests may reveal its real extent.

5. Comparison to the Related Literature

Current machine learning based initiatives for predicting stock market indices seem to have a preference for data coming from a single market only (*e.g.* Nair et al. (2010); Bollen et al. (2011); Shen et al. (2011); Kara et al. (2011); Lahmiri (2011); Lahmiri et al. (2013); Patel et al. (2015)). Despite this preference for domestic data, no market is an isolated island in the world economy ocean, and cross-market dependencies do exist (*cf.* Tam & Tam (2012); Urquhart & Hudson (2013); Coronado et al. (2016)). This is the reason why some researchers have moved across their borders when trying to forecast their own market directions. These initiatives either focus on a single market, bringing into the model one different foreign index (*e.g.* Wang (2014)), or try to integrate multiple markets, assuming some dependency amongst them might exist (*e.g.* Zuo & Kita (2012b,a); Kita et al. (2012)).

In our work, we take this later point of view. However, instead of having the algorithm figure out dependencies from a dataset by itself, as is the case with Zuo & Kita (2012b), who use a Bayesian Network built with the K2 algorithm (and which resulted in a very counter-intuitive network, where some past values depend on their future counterparts), we imposed these dependencies, by defining the network topology and then testing the appropriateness of this assumption. As it turned out, the 24h single index model performed better than that of Zuo & Kita (2012b), who report around 61% accuracy (against a 71% mean accuracy in our model). This difference might be even higher, if we bear in mind that our model was trained in data collected partly during the sub-prime crisis (roughly 2007-2008), which would be expected to raise the variance, whereas Zuo & Kita (2012b) dealt with pre-crisis data.

Other reasons for the difference in these results could come both from the network topology and the fact that our model works with short and independent sliding windows for training and testing, as a result of the cross-validation process and the 24-hour time limit (which, in turn, breaks up longer term recurrence relations). This may have diluted any effect due to some long term seasonality. Such an effect, however, could not have been mitigated should we follow the same procedure as many researchers in the related literature (*e.g.* Kara et al. (2011); Zuo & Kita (2012b); Patel et al. (2015)), who usually take the first years of data for training, holding the last years out to test their models. Also, it might be the case that the crisis itself reinforced market dependencies, thereby favouring our model only because we took data from this period. These, however, are hypotheses to be tested in the future.

Moving away from the use of Bayesian Networks, but still trying to accommodate cross-border influences into the model, we find the work by Wang & Choi (2013) and Wang (2014), who add the S&P 500 index as an external factor to their Principal Component Analysis and Support Vector Machine models. Testing their models with two stock indices, their best mean accuracy was 62,80%, still lower than ours. Spanning over 2002 to 2012, their dataset also included the sub-prime crisis, as did ours. The main differences between our research and theirs lie in the applied machine learning techniques and the four-year sliding window they used. Instead of separating the whole data set into two different subsets, one for training and other for testing, they have done so but within a four-year window, which would then be moved along the dataset, thereby reducing the influence of long term seasonalities.

Finally, within single market models we find accuracies ranging from 64% (Lahmiri, 2011), with Support Vector Machines, to slightly over 75% (Kara et al., 2011), with Neural Networks, to around 93% (Lahmiri et al., 2013), with Probabilistic Neural Networks. As can be seen, results from our 24-hour single index model, with its 71% mean and almost 78% maximum accuracies, lie within this range. The main difference between our work and these, however, seems to lie in the simplicity of our model. In this case, we only accounted for data related to closing values (and directions) of indices, whereas much of the extant work (*e.g.* Nair et al. (2010); Kara et al. (2011); Lahmiri (2011); Lahmiri et al.

(2013); Patel et al. (2015)) deals both with values and sets of technical indicators, such as Moving Average and Relative Strength Index, for example, with some of them also relying on the analysis of humour indicators (*e.g.* Lahmiri et al. (2013)) and tweeter messages (*e.g.* Bollen et al. (2011)). It might be the case that in adding more features other than a simple closing direction we could reach better accuracies.

In fact, a side-by-side comparison of related initiatives in terms of accuracy, taking into account the adopted machine learning method and whether researchers also relied on other technical indicators (such as Moving Average, for example) or any other information from outside the market, and which is summarised in Table 5[21], seems to point at this direction, with the exception of the work by Lahmiri et al. (2013) which, despite taking into account technical indicators (other than plain closing, highest, lowest and other related values), did not present high accuracies. From this table, we can see that a good number of machine learning techniques are proposed for market index forecasting. Most popular are Support Vector Machines (SVM in the table) and Artificial Neural Networks (ANN), followed by Probabilistic Neural networks (PNN), Decision Trees (DT), k-Nearest Neighbours (k-NN), Random Forests (RF), Naïve Bayes (NB), Principal Component Analysis (PCA) and Bayesian Networks (BN), with different accuracy results.

If we rank Table 5's results according to their accuracy, we see that our model outperforms all those that do not use any other information apart from the market plain technical data. Notably, the model also outperforms one of the models that rely on external sources or other classical technical indicators (the one by Lahmiri (2011)), also reaching results comparable to those by another such model (by Kara et al. (2011)), in its highest accuracy set-up. This, in turn, is an indication of its suitability for this kind of task, specially when taking into account its already mentioned simplicity. Nevertheless, questions

[21]Results that were not given in terms of accuracy were omitted in this table, since they could not be compared to ours.

Table 5: Comparison of the related literature on stock index value/direction prediction.

Research	*ML Method*	*Extra Information*	*Accuracy*
Nair et al. (2010)	DT	21 Tech. Indicators	90.22%
Bollen et al. (2011)	–	Tweeter feeds	86.70%
Kara et al. (2011)	ANN, SVM	10 Tech. Indicators	75,74% (ANN) 71.52% (SVM)
Lahmiri (2011)	PNN, SVM	12 Tech. Indicators	54.00% (PNN) 64.00% (SVM)
Lahmiri et al. (2013)	k-NN, PNN	Tech. Indicators and Sentiment Measures	93.45% (k-NN) 92.40% (PNN)
Patel et al. (2015)	ANN, SVM, RF, NB	10 Tech. Indicators	86.69% (ANN) 89.33% (SVM) 89.98% (RF) 90.19% (NB)
Wang & Choi (2013)	PCA, ANN, SVM	–	62.80% (PCA) 59.25% (ANN) 58.82 (SVM)
Zuo & Kita (2012b)	BN	–	61.44%

must be raised regarding whether these results are due to the model itself, or are but a by-product of other features, such as the time lapse of the data (which might have favoured the model) and the absence, in our model, of other classical technical indicators (which, in turn, would have limited it).

As it seems, the more information one adds to a model, the better it gets at its predictions. On the other hand, more potentially irrelevant information (*i.e.* noise) will be added. Models that solely rely on raw market data have the advantage of being more adaptable and computationally simple. However, they loose information given by other sources. Those that rely on such information do not suffer from this problem, but at the price of dealing with more noise, which in turn might have them astray from their goal, with the additional disadvantage of being computationally heavier and less adaptable to different circumstances. Relying on data from a single market may reduce the noise, but at the price of assuming market independence. In a ever more integrated world, this does not seem advisable and, even though the best results in Table 5 come from single market models, they also come from those relying on more elaborate data, leading to a confusion of variables. We leave the answers to these questions for future investigation.

Finally, and as a way to stablish some baseline to our research, we have repeated part of our experiment using an Artificial Neural Network (*cf.* Hattori et al. (2015)). In this case, we focused in the 24-hour model with one index per continent, since this was the set-up with best results in this research. We then used Weka[22] to generate a number of Multilayer Perceptron Artificial Neural Networks (MLP-ANN), by varying some of their default parameters in this tool. The same dataset used in our research was input to each new network, whose output should indicate iBOVESPA's next day closing direction. As in our Bayesian Network, the MLP-ANNs were trained and tested using 5-fold cross-validation, and mean accuracy across all folds was measured.

Best results were obtained with two neurons in the hidden layer, 0.2 learning rate, 0.1 momentum and 400 as the maximum number of epochs. With this setting, the network reached a 66.40% top accuracy, with NYSE Composite, Cac 40 and Hang Seng – the same set of indices that resulted in the best accuracy value for our Bayesian Network, in the 24-hour single index per continent model. Interestingly, even though our network has performed better than its MLP-ANN counterpart, delivering a 77.78% accuracy with this same set of indices, it is noticeable that both networks lined up their outputs, agreeing on which indices influence iBOVESPA the most. This is yet another indicative of the robustness of the central idea behind both models: that market dependencies can be used to forecast a specific index closing direction within a 24-hour window.

[22] Waikato Environment for Knowledge Analysis (`http://www.cs.waikato.ac.nz/ml/weka/`)

6. Conclusion

In this work, we set out to identify the applicability of a Bayesian Network structure in the investigation of the extent to which foreign markets influence the main index at the São Paulo Stock Exchange (*i.e.* iBOVESPA), by taking closing directions of stock market indices around the globe as input to the network, so as to try to forecast iBOVESPA's next day direction. Being one of the few to move away from the Neural Network and Support Vector Machine mainstreams, our research is, to the best of our knowledge, the first machine learning based effort to take an "around the globe" approach, that is to base its underlying model on the assumption that markets are interrelated all around the world. We do so through a Bayesian Network that reflects the potential dependencies amongst indices from all continents while the planet rotates. Moreover, this is probably the machine learning effort trying to accommodate the highest number of different markets into a single model.

Since "the central idea to successful stock market prediction is achieving best results using minimum required input data and the least complex stock market model" (Atsalakis & Valavanis, 2009), and with results comparable to those of the related literature, our model moves towards this direction, by presenting the further advantage of being relatively simple and intuitive, when compared to other more "closed" models, such as those that deal with Neural Networks, for example. Also, the model makes no assumptions regarding error distribution, as do some time-series algorithms, thereby generalising over different datasets. Nevertheless, some of the more elaborate models, specially those that account for some traditional technical analysis indicators along with raw technical data, were found to be more accurate than ours, even though there are exceptions to this rule. This is a matter left for future investigation: whether by adding such information we can raise our system's accuracy.

Regarding its usefulness, in the realm of Expert and Intelligent Systems our model brings in a different dimension under which we understand data should be analysed: inter-market dependence. As a rule, current machine learning based models avoid bringing together different markets, with some notable exceptions, as pointed out in Section 5. Related models that deal with market data only, that is those not relying on other classical techniques for market analysis, were outperformed by ours, which is an indicative of this idea's appropriateness. Our model was, however, outperformed by some of those that make use of such techniques along with raw market data. Unfortunately, these were also the same models that did not rely on multiple market information, thereby leading to a confounding of variables. The reasons why these models were more accurate than ours is something that must be addressed in future research.

Still in the field of Expert Systems, we understand our model to serve as one of the building blocks in more elaborate applications. Given its high adaptability, whereby it can be quickly changed to output probabilities instead of closing directions, it could be used as a "foreign market influence" module to more sophisticated decision support systems, for example. Also, and since current research in Economics points towards the actual existence of market inter-dependencies (*e.g.* Coronado et al. (2016); Kim & Song (2017); Tangpornpaiboon & Puttanapong (2016)), which would be stronger after crisis periods, our model comes up as a computationally cheap, human readable alternative, during and after crises, to existing models where inter-market dependencies are ignored.

In the areas of Economics and practical trading, the model could be used to study and forecast market contaminations during crises, for example, by determining, from the identification of the set of markets that influence the target market the most, how likely this target market is to suffer from such contamination. It could also be used as a tool for determining the effects that commercial agreements have on a specific market. In this case, a study of which indices have more influence in that market before and after the agreement and, more specifically, whether this set does actually change, could give an insight on which companies gained or lost international importance with the agreement, indicating how it was perceived by market players and how future agreements might so be perceived. Finally, our network could be modified to model the intuitions market players have on market dependencies. In this case, intuitions would be codified in the network's topology, and an analysis of its accuracy could provide evidence for (or against) them.

Avenues for future improvement, other than those pointed out above, include testing the model with data coming from non-crisis periods. Since part of our dataset comprises the sub-prime crisis, it might be the case that the system was favoured by it. To run it in different periods might clear this doubt and help determine if the same set of indices that influence iBOVESPA in one period would still do so in others. Also, it would be interesting testing this dependency assumption with other Machine Learning techniques, such as Deep Neural Networks, for example, perhaps adding them to an ensemble, if different techniques turn out to be complementary to each other. Another interesting test to our ideas would be to apply them to specific stocks. Even though one might think of an individual stock as a single-company index, the model was tested on general indices only, which comprise dozens of different stocks, and there is no guarantee that these, as a group, would behave as if they were a single company index.

Similarly, and since our model is based on the assumption that markets are interrelated all around the world, it would be interesting testing this same assumption with different topologies. One might, for example, create intra-continental dependencies (as it was somewhat done in the 48-hour model, when we took iBOVESPA out of America in the middle of the network), or account for dependencies on an index, instead of continent, basis, or even to separate indices according to whether they are capitalization or price-weighted[23]. Also, and as a way to verify the robustness of the model, it could be tested with different target-indices (*i.e.* indices whose direction is to be forecast), so as to determine if it can be generalised to other markets. Observing how the model would work in practice, by testing it with different trading strategies, would

[23] *I.e.*, indices whose constituent stocks are weighted according to their market capitalization or price per share, respectively.

give us some interesting insights as well.

Finally, another point we find noteworthy for future research is the methodology adopted for training and testing forecasting systems. We have taken the cross-validation approach, in an attempt to reduce any influence seasonality might have in our results. In that sense, seasonality may pose a problem if the learning algorithm ends up learning features that work only on that specific time lapse. Nevertheless, cross-validation comes at the price of breaking any recurrence relation in the model, which makes it unsuited for most time series analyses. The extent to which this is a valid concern still remains unclear, and is worth investigating.

Appendix

Table 6: Mean amount, across all folds, of correctly (hits) and incorrectly (misses) forecast directions, along with mean accuracy results for the 24h model, with one index per continent.

Index Arrangements	*Hits*	*Misses*	*Accuracy*
Nasdaq Comp., FTSE 100, Nikkei 225	233.6	90.4	0.7210
Nasdaq Comp., Dax, Nikkei 225	236.8	87.2	0.7309
Nasdaq Comp., Stockholm Gen., Nikkei 225	235.6	88.4	0.7272
Nasdaq Comp., Cac 40, Nikkei 225	229.8	94.2	0.7093
Nasdaq Comp., FTSE 100, Shangai Comp.	239.8	84.2	0.7401
Nasdaq Comp., Dax, Shangai Comp.	235.8	88.2	0.7278
Nasdaq Comp., Stockholm Gen., Shangai Comp.	233.6	90.4	0.7210
Nasdaq Comp., Cac 40, Shangai Comp.	238.8	85.2	0.7370
Nasdaq Comp., FTSE 100, Hang Seng	232.4	91.6	0.7173
Nasdaq Comp., Dax, Hang Seng	236.6	87.4	0.7302
Nasdaq Comp., Stockholm Gen., Hang Seng	232.4	91.6	0.7173
Nasdaq Comp., Cac 40, Hang Seng	237.4	86.6	0.7327

Continued on next page

Table 6 – *Continued from previous page*

Index Arrangements	*Hits*	*Misses*	*Accuracy*
Nasdaq Comp., FTSE 100, BSE 30 Sensex	235.0	89.0	0.7253
Nasdaq Comp., Dax, BSE 30 Sensex	236.6	87.4	0.7302
Nasdaq Comp., Stockholm Gen., BSE 30 Sensex	241.2	82.8	0.7444
Nasdaq Comp., Cac 40, BSE 30 Sensex	234.2	89.8	0.7228
NYSE Comp., FTSE 100, Nikkei 225	244.6	79.4	0.7549
NYSE Comp., Dax, Nikkei 225	241.8	82.2	0.7463
NYSE Comp., Stockholm Gen., Nikkei 225	242.6	81.4	0.7488
NYSE Comp., Cac 40, Nikkei 225	237.4	86.6	0.7327
NYSE Comp., FTSE 100, Shangai Comp.	241.0	83.0	0.7438
NYSE Comp., Dax, Shangai Comp.	239.0	85.0	0.7377
NYSE Comp., Stockholm Gen., Shangai Comp.	242.6	81.4	0.7488
NYSE Comp., Cac 40, Shangai Comp.	238.0	86.0	0.7346
NYSE Comp., FTSE 100, Hang Seng	238.0	86.0	0.7346
NYSE Comp., Dax, Hang Seng	241.6	82.4	0.7457
NYSE Comp., Stockholm Gen., Hang Seng	241.2	82.8	0.7444
NYSE Comp., Cac 40, Hang Seng	252.0	72.0	0.7778
NYSE Comp., FTSE 100, BSE 30 Sensex	243.2	80.8	0.7506
NYSE Comp., Dax, BSE 30 Sensex	241.8	82.2	0.7463
NYSE Comp., Stockholm Gen., BSE 30 Sensex	238.0	86.0	0.7346
NYSE Comp., Cac 40, BSE 30 Sensex	240.2	83.8	0.7414
Dow Jones, FTSE 100, Nikkei 225	197.8	116.2	0.6299
Dow Jones, Dax, Nikkei 225	209.8	107.2	0.6618
Dow Jones, Stockholm Gen., Nikkei 225	197.8	117.2	0.6279
Dow Jones, Cac 40, Nikkei 225	208.0	109.0	0.6562
Dow Jones, FTSE 100, Shangai Comp.	214.8	105.2	0.6713
Dow Jones, Dax, Shangai Comp.	210.6	113.4	0.6500
Dow Jones, Stockholm Gen., Shangai Comp.	207.0	115.0	0.6429

Continued on next page

Table 6 – *Continued from previous page*

Index Arrangements	*Hits*	*Misses*	*Accuracy*
Dow Jones, Cac 40, Shangai Comp.	213.8	110.2	0.6599
Dow Jones, FTSE 100, Hang Seng	212.6	111.4	0.6562
Dow Jones, Dax, Hang Seng	212.6	111.4	0.6562
Dow Jones, Stockholm Gen., Hang Seng	206.0	118.0	0.6358
Dow Jones, Cac 40, Hang Seng	215.4	108.6	0.6648
Dow Jones, FTSE 100, BSE 30 Sensex	213.6	104.4	0.6717
Dow Jones, Dax, BSE 30 Sensex	209.2	112.8	0.6497
Dow Jones, Stockholm Gen., BSE 30 Sensex	201.8	118.2	0.6306
Dow Jones, Cac 40, BSE 30 Sensex	217.4	104.6	0.6752
Merval, FTSE 100, Nikkei 225	226.4	97.6	0.6988
Merval, Dax, Nikkei 225	231.8	92.2	0.7154
Merval, Stockholm Gen., Nikkei 225	230.8	93.2	0.7123
Merval, Cac 40, Nikkei 225	230.0	94.0	0.7099
Merval, FTSE 100, Shangai Comp.	238.2	85.8	0.7352
Merval, Dax, Shangai Comp.	233.4	90.6	0.7204
Merval, Stockholm Gen., Shangai Comp.	232.6	91.4	0.7179
Merval, Cac 40, Shangai Comp.	233.0	91.0	0.7191
Merval, FTSE 100, Hang Seng	232.0	92.0	0.7160
Merval, Dax, Hang Seng	238.6	85.4	0.7364
Merval, Stockholm Gen., Hang Seng	236.0	88.0	0.7284
Merval, Cac 40, Hang Seng	230.4	93.6	0.7111
Merval, FTSE 100, BSE 30 Sensex	229.4	94.6	0.7080
Merval, Dax, BSE 30 Sensex	232.6	91.4	0.7179
Merval, Stockholm Gen., BSE 30 Sensex	233.2	90.8	0.7198
Merval, Cac 40, BSE 30 Sensex	235.6	88.4	0.7272

References

References

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – part ii: Soft computing methods. *Expert Systems with Applications*, *36*, 5932–5941.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1–8.

Coronado, S., Rojas, O., Romero-Meza, R., & Venegas-Martínez, F. (2016). A study of co-movements between u.s. and latin american stock markets: A cross-bicorrelations perspective. *DYNA*, *83*, 143–148.

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*, 10389–10397.

Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*, *23*, 800–808.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2nd ed.). Springer.

Hattori, R. H., Roman, N. T., & Monteiro, A. M. (2015). Multi-layer perceptron artificial neural network for forecasting ibovespa's closure trend: indexes that influence ibovespa. In *Proceedings of the 12th International Conference on Information Systems and Technology Management (CONTECSI)* (pp. 1–19). São Paulo, SP, Brazil.

Hsieh, T. J., Hsiao, H. F., & Yeh, W. C. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*, *11*, 2510–2525.

Huarng, K., & Yu, H.-K. (2005). A type 2 fuzzy time series model for stock index forecasting. *Physica A: Statistical Mechanics and its Applications*, *353*, 445–462.

Jithendranathan, T. (2013). Market microstructure in emerging and developed markets. chapter Financial Market Contagion. (pp. 115–133). John Wiley & Sons.

Kara, Y., Boyacioglu, M. A., & Ömer Kaan Baykan (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert Systems with Applications*, *38*, 5311–5319.

Kim, D., & Song, C.-Y. (2017). Bankruptcy of lehman brothers: Determinants of cross-country impacts on stock market volatility. *International Journal of Economics and Financial Issues*, *7*, 210–219.

Kita, E., Zuo, Y., Harada, M., & Mizuno, T. (2012). Application of bayesian network to stock price prediction. *Artificial Intelligence Research*, *1*, 171–184.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence (IJCAI'95)* (pp. 1137–1143). Montreal, Quebec, Canada.

Lahmiri, S. (2011). A comparison of pnn and svm for stock market trend prediction using economic and technical information. *International Journal of Computer Applications*, *29*, 24–30.

Lahmiri, S., Boukadoum, M., & Chartier, S. (2013). Information fusion and s&p500 trend prediction. In *Proceedings of the 2013 ACS International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–7). Ifrane, Morocco.

Lee, C. F., & Lee, A. C. (Eds.) (2006). *Encyclopedia of Finance*. Springer.

Nair, B. B., Mohandas, V. P., & Sakthivel, N. R. (2010). A decision tree- rough set hybrid system for stock market trend prediction. *International Journal of Computer Applications*, *6*.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*, 259–268.

Puah, C.-H., Brahmana, R. K., & Wong, K.-H. (2015). Revisiting stock market integration pre-post subprime mortgage crisis: Insight from bric countries. *Economics and Finance in Indonesia*, *61*, 120–130.

Shen, W., Guo, X., Wu, C., & Wu, D. (2011). Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems*, *24*, 378–385.

Shenoy, C., & Shenoy, P. P. (1999). Bayesian network models of portfolio risk and return. In Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, & A. S. Weigend (Eds.), *Computational Finance 1999* chapter 7. (pp. 87–101). MIT Press.

Sun, L., & Shenoy, P. P. (2007). Using bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research*, *180*, 738–753.

Tam, P. S., & Tam, P. I. (2012). *Rethinking stock market integration: Globalization, valuation and convergence*. Technical Report 2012-052 SFB 649, Humboldt-Universität zu Berlin.

Tangpornpaiboon, S., & Puttanapong, N. (2016). Financial contagion of the global financial crisis from the us to other developed countries. *Journal of Administrative and Business Studies*, *2*, 48–55.

Ticknor, J. L. (2013). A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, *40*, 5501–5506.

Urquhart, A., & Hudson, R. (2013). Efficient or adaptive markets? evidence from major stock markets using very long run historic data. *International Review of Financial Analysis*, *28*, 130–142.

Wang, Y. (2014). Stock price direction prediction by directly using prices data: an empirical study on the kospi and hsi. *International Journal of Business Intelligence and Data Mining*, *9*, 145–160.

Wang, Y., & Choi, I.-C. (2013). Market index and stock price direction prediction using machine learning techniques: An empirical study on the kospi and hsi. arXiv preprint arXiv:1309.7119.

Yeh, C.-Y., Huang, C.-W., & Lee, S.-J. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications*, *38*, 2177–2186.

Zuo, Y., & Kita, E. (2012a). Stock price forecast using bayesian network. *Expert Systems with Applications*, *39*, 6729–6737.

Zuo, Y., & Kita, E. (2012b). Up/down analysis of stock index by using bayesian network. *Engineering Management Research*, *1*, 46–52.