Linkages in international stock markets:

Evidence from a classification procedure[*].

Simon Sosvilla-Rivero

(FEDEA and Universidad Complutense de Madrid)

Pedro N. Rodríguez

(Sungard Trading and Risk Systems--BancWare)

Abstract:

In this paper we propose a new approach to evaluate the predictable components in stock indices using a boosting-based classification technique, and we use this method to examine causality among the three main stock market indices in the world during periods of large positive and negative price changes. The empirical evidence seems to indicate that the Standard & Poors 500 index contains incremental information that is not present in either the FTSE 100 index or the Nikkei 225 index, and that could be used to enhance the predictability of the large positive and negative returns in the three main stock market indices in the world. This in turn would suggest a causality relationship running from the Standard & Poors 500 index to both the FTSE 100 and the Nikkei 225 indices.

JEL Classification Numbers: G15, C32

Keywords: International stock markets, Causality

---

[*] The authors would like to thank the editor and two anonymous referees for helpful comments and suggestions on a previous draft of this paper. This paper was written while Pedro N. Rodriguez was a PhD student at the Universidad Complutense de Madrid and he thanks CONACYT (Mexico) for financial support (Fellowship: 170328). The views expressed here are those of the authors and not necessarily those of the institutions with which they are affiliated.

1. Introduction

Interdependence among national stock markets is an important concern for investors since in recent years the global market has became more integrated as a result of a broad tendency toward liberalisation and deregulation in the money and capital markets of developed as well as developing countries, thereby reducing the opportunities for international diversification. Furthermore, market co-movements can also lead to market contagion as investors incorporate information about price changes in other markets into their trading decisions in an attempt to form a complete information set, carrying the risk that errors in one market may be transmitted elsewhere (see, e. g., King and Wadhwani, 1990).

As traders place orders and take positions simultaneously using different indices, one expects that different indices would be closely linked by both hedging activities and cross-market arbitrage (see, e. g., Karoly and Stulz, 1996). Furthermore, modern technological advancements and computerised trading systems have greatly facilitated the transfer of information from market to market.

One strand of research has examined long-term interdependence in international stock markets using multivariate cointegration techniques [see, Kasa (1992), Corhay, Rad and Urbain (1993), Blackman, Holden and Thomas (1994), Bachman, Choi, Jeon and Kopecky (1996) and Fernández-Serrano and Sosvilla-Rivero (2001, 2003), among others]. The other main area of research has attempted to investigate short-run market relationships through correlation tests [see, e. g., DeFusco *et al.* (1996) and Aggarwall *et al.* (1999)] or price and volatility spillovers using variations of the GARCH approach [see, e. g., Hamo, Masulis and Ng (1990), Koutmos and Booth (1994), Chiang and Chiang (1996) and Fratzscher (2002)].

More recently, there has been a growing interest in the market dynamics and the transmission mechanism driving international market linkages [see, e. g., Eun and Shim (1989), Arshanapalli and Doukas (1993), Masih and Masih (1997) and Connolly and Wang (2002)].

Recent advances in machine learning and statistics have contributed to the appearance of robust classification and evaluation techniques. In particular, these techniques allow us to focus on the predictability of specific price changes (codified into populations). Especially, one would like to discriminate positive from negative returns and/or large positive or negative price changes from the rest of price movements. Doing so allows us to test whether or not lagged returns from a main stock market index provide incremental information concerning the classification of financial movements. Moreover, it permits us to compare and determine which set of explanatory variables provides more information to discriminate index movements.

The purpose of this paper is to investigate the casual relationships between the three major stock markets in the world using daily data covering the period February 1986-June 2004. Our study differs from the previously published papers in several ways. First, while traditional studies focus on markets that are located in neighbouring geographic areas, we examine linkages across global markets. Second, instead of employing short-run market relationships through correlation tests or long-run co-movements, we investigate the relevance of lagged returns for discriminating price changes by using a boosting-based classification technique. Finally, while a growing body of research is devoted to the impact of financial crises on stock market linkages, we examine their relationships both during periods of high, positive and negative returns.

By evaluating and comparing the 'overall' discriminatory accuracy of the boosting-based algorithms (The Gradient Boosting Machine) trained under three different sets of explanatory variables, we find that the S&P 500 does provide incremental information to enhance the predictability of the large price changes in the three main stock market indices in the world. Moreover, this predictive structure behaviour is neither present in the FTSE 100 nor in the Nikkei 225.

The paper is organised as follows. In Section 2, we provide a brief review of the Gradient Boosting Machine classification technique. Section 3 discusses the receiver operating characteristic (ROC) curve and the area under it, both used to evaluate discriminatory accuracy. In Section 4, we report the empirical results. Finally, Section 5 provides some concluding remarks.

[2] We also consider other criteria for identifying large positive price change (such those representing 10 and 30 per cent of the probabilistic distribution of the returns), obtaining similar qualitative results.

2. The Gradient Boosting Machine classification technique

In the function approximation problem, one has a system consisting of a random response variable y and a set of random explanatory variables $\mathbf{x}=\{x_1, x_2,..., x_n\}$. Given a training sample $\{y_i,\mathbf{x}_i\}^N_1$ of known $(y,\mathbf{x})$-values, the goal is to find a function $F^*(\mathbf{x})$ that maps $\mathbf{x}$ to y, such that over the joint distribution of all $(y,\mathbf{x})$-values, the expected value of some loss function is minimised.

Regression and classification problems can be viewed as a task in function approximation. In a classification problem, the goal is to discriminate between two (or more) populations, given a set of explanatory variables. In our particular case, the relevant question is if lagged price changes are capable of providing information for discriminating large positive financial movements.

To answer such question empirically, we must test the out-of-sample discriminatory accuracy of the classifier chosen, in order to understand specific movements. Moreover, this out-of-sample performance should be evaluated with a technique that is invariant to *a-priori* class probabilities and that is independent of the decision threshold (or the cut-off value).

In this paper, we have employed the Gradient Boosting Machine classification technique, which we now proceed to review. Regarding the evaluation of the discriminatory accuracy, in the next section we discuss the receiver operating characteristic (ROC) curve and the area under it.

The underlying idea of boosting is to combine simple "rules" (or tree-based models) to form an ensemble (or weighted committee) such that the performance of the single ensemble member is improved (e.g., "boosted"). Boosting was proposed in the machine learning literature by Freund and Schapire (1997) with the creation of AdaBoost, and has since received much attention.

From the statistician's point of view, the breakthrough took place with the paper by Friedman, Hastie and Tibshirani (2000). They show that AdaBoost is an optimisation method for finding a classifier that minimises a particular exponential loss function, and that this loss function is very similar to the (negative) binomial log-likelihood. For an excellent paper that traces the developments of boosting methodology and its applications to the exponential family and proportional hazards regression models see Ridgeway (1999).

Using the connection between boosting and optimisation, Friedman (2001) proposes the Gradient Boosting Machine (hereafter GBM). Using this technique, function approximation is viewed from the perspective of numerical optimisation in the function space, rather than in the parameter space. The estimation algorithm of the GBM is shown in Chart 1.

[Chart 1, here]

With the purpose of increasing execution speed, approximation accuracy and robustness against over-fitting, we incorporated randomness in the procedure as described in Friedman (2002). At each iteration, a sub-sample consisting of 50 percent of the total observations $\left(\widehat{N}/N = 50\%\right)$ is drawn at random (without replacement) from the training sample. This sub-sample is then used to fit the tree and compute the model output for the current iteration. The loss function employed in all our experiments with the GBM was the Bernoulli function, and the loss function used to generate the "rules" (or tree-based models) was the Gini index.

The GBM has three tweaking parameters: the total number of iterations M, the learning rate (shrinkage parameter v) and the level of interaction among explanatory variables J. It is widely known that fitting the data too well can lead to over-fitting, which degrades the accuracy power on independent data bases. However, M and v do not operate independently: smaller values of v lead to larger values of the "optimal" M.

Empirically, via simulation, Friedman (2001) finds that low values of v (v < 1%) favour better accuracy on test samples. In this paper, we fixed the shrinkage parameter (v) to one percent and allowed 500 boosting iterations (M). Clearly, better results can be obtained if we monitor these parameters in a validation set. Nevertheless, this was not done in order to maximise the number of observations in the test sample. Regarding the level of interaction (J), experience indicates, hitherto, that $4 \le J \le 8$ works well in the context of boosting, with results being fairly insensitive to particular choices in this range (Hastie, Tibshirani, and Friedman (2001). Consequently, we fixed J equal to five in all our experiments. The implementation was carried out in R: Environment for Statistical

Computing and Graphics with the following add-on package: gbm (developed by G. Ridgeway).

3. Assessing the discriminatory accuracy

The receiver operating characteristic (ROC) curve is a well-established method for summarising performances of diagnostic tests [see, Hanley (1999); Zhou, McClish and Obuchowski (2002), among others]. Now, it is widely used for evaluating machine learning algorithms [see, e. g., Weiss and Provost (2003); Provost, Fawcett and Kohavi (1998); Bradley (1998)].

A ROC curve is obtained by plotting Sensitivity versus 1 - Specificity for various cut-off values. The points on a ROC curve are either joined by line segments or smooth curves with the use of non-parametric and parametric procedures, respectively.

The area under a ROC curve (AUC) is equal to the probability that a randomly selected observation from population one scores higher than a randomly selected observation from population two (Hanley and McNeil (1982). Formally, the AUC can be expressed as:

$$\text{AUC} = P(y > x) + \frac{1}{2}P(y = x) \tag{1}$$

where y and x denote the classifier output (class probability) for a randomly selected observation from population one and two, respectively.

In this paper, we make use of the Mann-Whitney-U Statistic, a non-parametric approach to obtain the AUC. The Mann-Whitney-U Statistic is given by:

$$\text{AUC} = \hat{\oplus} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Z_{ij} \tag{2}$$

where $Z_{ij} = I(y_i > x_j) + \frac{1}{2} I(y_i = x_j)$ and $I(\bullet)$ is an indicator function. The variables m and n are the total number of observations for population one and two, respectively.

The variance of the AUC can be obtained from the following:

$$Var(\hat{\oplus}) = \frac{\hat{\oplus}(1-\hat{\oplus}) + (m-1)(Q_1 - \hat{\oplus}^2) + (n-1)(Q_2 - \hat{\oplus}^2)}{mn} \tag{3}$$

and $Q_1 = \hat{\oplus}\big/2-\hat{\oplus};\ Q_2 = 2\hat{\oplus}^2\big/1+\hat{\oplus}.$ Equation 3 is the exponential approximation formula of Hanley and McNeil (1982).

Another important application of the ROC curve is to compare the *overall* discriminatory accuracy of classifiers. The term ‘overall’ is emphasised, since the interest is in the global picture of a test and not in the accuracy performance at a particular cut-off value. See Metz et al. (1998) to understand how different ROC-curve estimates can be compared.

4. Data and empirical results

The data set compromises daily time series on closing values for the three major stock market indices: the FTSE 100 index (United Kingdom), the Nikkei 225 index (Japan), and the Standard & Poors 500 index (United States). The sample period examined spans from June 2, 1986 to June 25, 2004. The data set was obtained from Econwin.

We employed the Gradient Boosting Machine classification technique to extract patterns from the explanatory variables in order to classify large positive and negative price changes, and we used the area under the ROC curve to evaluate and compare the 'overall' discriminatory accuracy of the classifiers. In order to identify large positive price changes, for each index we chose the returns that satisfy the following condition: $prob(\text{return} < (\text{large, positive return})) = 80\%$. In other words, the large positive returns represent 20 per cent of the probabilistic distribution of the returns (or right-tail of the distribution of the returns)[2]. Concerning the large negative price changes, we chose those observations that represented the left-tail of the distributions of the returns (e.g., $prob(\text{return} < (\text{large, negative return})) = 20\%$). Observations that satisfied the condition were codified with 1's, and those that did not were codified as 0's.

Once the dependent (or response) variable was codified, we analysed three sets of exploratory factors per index: lags in Standard & Poors 500 index, lags in the FTSE 100 index , and lags in the Nikkei 225 index. These explanatory variables were incorporated in the GBM algorithm depending on the trading hours. Note that the Japanese stock market is the first to open among the three considered, followed by the UK stock market, and the US stock market is the last one to open[3]. Therefore, when examining each particular stock market, we will successively incorporate the publicly available information on stock prices taking into account these trading hours. For example, to analyse the Japanese stock market, we first take into account only lags in the Nikkei 225 index, then we introduce lags in the FTSE 100 index, and finally we consider also the lags in the Standard & Poors 500 index.

As is well known, keeping a test set additionally to the training set is a general method to estimate the accuracy of a classifier. Therefore, we have split our sample into two sub-samples: the training sample runs from June 2, 1986 to December 31, 2000, while the testing sample covers the period January 3, 2001- June 25, 2004. Usually a fraction of 20-30 per cent of the available data is chosen as the testing set if the size of the test set is larger than 1000. Therefore, our testing sample satisfies the train-and-test procedure (see, e.g., Henery, 1994). Note also that the testing period is long enough to reduce the effects of data snooping.

In the top section of Table 1 (Panel I), we evaluate the discriminatory accuracy of the GBM for the US stock market under different sets of explanatory variables. Test A shows the GBM classifying power when only lags of the Standard & Poors 500 index are used. For the large positive price changes, the discriminatory power is 0.6129, meaning that the probability that a randomly selected observation from the large positive price

[3] The relevant time with respect to Greenwich Mean Time is +9 for the Nikkei 225, 0 for the FTSE 100, and –6 for the Standard and Poors 500.

change cases (returns greater than 0.68%) scores higher than a randomly chosen observation from price changes from returns lower than 0.68% is 61.29 per cent. Notice that when there is no difference between the two populations, the area under the ROC curve will be 0.5 (or coin-toss classifications). The 95 per cent confidence interval for the area can be used to test the hypothesis that the theoretical area is 0.5. Since the confidence interval in this case (0.5758, 0.6482) does not include the 0.5 value, there is evidence that the GBM technique is able to distinguish between the two groups (or populations). Test B reports the results when we use lags of both Standard & Poors 500 and Nikkei 225 indices as explanatory factors, while Test C presents the results when lags of the all three are used. As can be seen, including those lags leads to a small reduction in the discriminatory power of the GBM technique (61.09 and 60.75 per cent, respectively), although the new confidence intervals do not include the 0.5 value either. Concerning the large negative price changes, predictable components were non-existent, since the GBM, trained under the three different sets of explanatory variables, was not able to beat a coin-toss (or random) classification.

In the lower section of Table 1 (Panel II), we report the *z* statistic, testing the difference between areas under the ROC curves, as well as the p-value. As can be seen for the large positive price changes, the *z* statistics obtained when comparing Test A *vs*. Test B and Test A *vs*. Test C are positive, suggesting that Test A has a greater area under the ROC curves than those of Tests B and C. Moreover, the *p*-values suggest that the pair-wise comparisons of all three ROC curves are not significantly different. Therefore, based on this evidence we can conclude that including more explanatory variables does not generate any incremental information to improve the discriminatory power of the GBM technique when identifying the large positive price changes in the Standard & Poors 500 index.

However, when analyzing large negative price changes, none of the explanatory variable sets contained any useful information to discriminate large negative price changes from the rest of price movements. Furthermore, in one of our experiments the ROCKIT software did not converge since the data set was degenerated. In general, degeneracy should be found only in very small datasets and/or in those with many tied values, and in our case it was due to the existence of many tied values.

[Table 1, here]

Regarding the FTSE 100 index, we observe in Panel I of Table 2 that there is a significant improvement in the discriminatory accuracy for the large positive price changes (returns greater than 0.74%) when the lags of the Standard & Poors 500 and the Nikkei 225 indices are introduced in the information set. As can be seen, Test C presents the greatest area under the ROC curve. Moreover, when comparing the areas under the ROC curves for the three sets of explanatory variables, we obtain a negative value for the *z*-statistic, suggesting that Tests B and C have a greater area under the ROC curve than Test A, being also statistically significant at the usual levels. These results, jointly taken with the results when comparing Test B *vs.* Test C, indicate that the Standard & Poors

500 index provides incremental information to enhance the discriminatory power of large positive returns in the FTSE 100 index.

Furthermore, the same pattern emerges in the predictability of large negative price changes, where the Standard & Poors 500 index provide useful information for discriminating the FTSE 100 large negative price changes from the rest of price movements.

[Table 2, here]

Finally, Table 3 reports the results for the Nikkei 225 index. As can be seen in the table, the information provided by the lags in the Nikkei 225 index is not relevant for discriminating large positive price changes (returns greater than 0.82%). In contrast, both the Standard & Poors 500 index and the FTSE 100 index contain relevant information in this context. Nevertheless, the results obtained when comparing Test B vs. Test C suggest that the Standard & Poors 500 index contains incremental information that is not present in the FTSE 100 index and that could be used to increase the discriminatory accuracy of the large positive returns in the Nikkei 225 index. Moreover, the predictability of the large negative price changes is greatly enhanced due to the information extracted from the Standard & Poors 500 index lags.

[Table 3, here]

The traditional method for testing causality in economic time series makes use of the well-known Granger definition of causality (Granger, 1969). Given two variables, *x* and *y*, *x* is said to Granger-cause *y* if the latter can be predicted better by past values of *x* and *y*, rather than by past values of *y* alone. From this definition, the results in Tables 1 to 3 could be taken as empirical evidence of causality running from the Standard & Poors 500 index to both the FTSE 100 and the Nikkei 225 indices.

5. Concluding remarks

In this paper, we extend the previous literature by examining stock market linkages during large positive and negative price changes. The main novelty of the paper lies in its use of the Gradient Boosting Machine classification technique to examine the relevance of lagged returns for discriminating price changes.

The empirical evidence presented in this paper seems to indicate that the Standard & Poors 500 index contains incremental information that is not present either in the FTSE 100 index or in the Nikkei 225 index and that could be used to enhance the forecastability of the large positive and negative returns in the three main stock market indices in the world. This in turn would suggest a causality relationship running from the Standard & Poors 500 index to both the FTSE 100 and the Nikkei 225 indices.

Therefore, our results indicate that the US index plays a prominent role in price leadership across global markets. This is consistent with the fact that the US market has long been the centre of financial transactions as well as the most influential producer of information. Indeed, there is a large literature on the fact that investors tend to react more to news from the US market than from other markets [see, e. g., Eum and Shim (1989), Becker, Finnerty and Friedman (1995), and Masih and Masih (2001)].

Our evidence is also consistent with that presented in Karoly and Stulz (1996), who conclude that large return shocks propagate more internationally than small return shocks.

Our findings support the conclusion coming from agent-based models for financial market which account for behavioural phenomena such as herding, overreaction and the common use of trading rules, that are likely to spread non fundamental price variations across markets (Day and Huang, 1990; Brock, Lakonish and LeBaron, 1992; Connoly and Wang, 2002, Daniel, Hirshleifer and Teoh, 2002, and Hirshleifer and Teoh, 2003). In this sense, King and Wadhwani (1990) argue that trading of stocks in one market per se affects stock prices in other markets, even if the source of the trading is purely noise. This is called the “market-contagion hypothesis”. Karoly and Stulz (1996) also find evidence consistent with this interpretation, pointing out that “[c]ontagion effects results when enthusiasm from stocks in one market brings about enthusiasm for stocks in other markets, regardless of the evolution of market fundamentals”.

References

Arshanapalli, B and Doukas, J., 1993, International stock market linkages: Evidence from pre- and post-October 1987 period, Journal of Banking and Finance 17: 193-208.

Bachman D., Choi J. J., Jeon B. N. and Kopecky, K. J., 1996, Common factors in international stock prices: Evidence from a cointegration study, International Review of Financial Analysis 5: 39-53.

Becker, K. G., Finnerty, J. E., and Friedman, J. E., 1995, Economic news and equity market linkages between the US and UK, Journal of Banking and Finance 19: 1191-1210.

Blackman, S. C., Holden, K. and Thomas, W. A., 1994, Long-term relationships between international share price, Applied Financial Economics 4: 297-304.

Bradley, A. P. , 1998, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition 30: 1145-1159.

Brock, W., Lakonish, J. and LeBaron, B., 1992, Simple technical rules and the stochastic properties of stock returns, Journal of Finance 47: 1731-1764.

Chiang, T. C. and Chiang, J. J., 1996, Dynamic analysis of stock return volatility in an integrated international capital market, Review of Quantitative Finance and Accounting 6: 5-17.

Connoly, R. A. and Wang, F. A., 2002, International equity market co-movements: Economic fundamentals or contagion?, Pacific-Basin Finance Journal111: 23-43.

Corhay, A. A., Rad, T. and Urbain, J., 1993, Common stochastic trends in European stock markets, Economic Letters 42: 385-390.

Daniel, K., Hirshleifer, D. and Teoh, S. H., 2002, Investor psychology in capital Markets: Evidence and policy implications, Journal of Monetary Economics 49: 139-209.

Day, R. H. and Huang, W., 1990, Bulls, bears and the market sheep, Journal of Economic Behavior and Organization 14: 299-330.

Eum, C. S. and Shim, S., 1989, International transmission of stock market movements, Journal of Finance and Quantitative Analysis 24: 241-256.

Fernández-Serrano, J. L. and Sosvilla-Rivero, S., 2001, Modelling evolving long-run relationships: The linkages between stock markets in Asia, Japan and the World Economy 13: 145-160.

Fernández-Serrano, J. L. and Sosvilla-Rivero, S., 2003, Modelling the linkages between US and Latin American stock markets, Applied Economics 35: 1423-1434.

Fratzscher, M., 2002, Financial market integration in Europe: On the effects of EMU on stock markets, International Journal of Finance and Economics 7: 165-194

Freund, Y., and Shapire, R. E., 1997, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and Systems Sciences 55: 119-139.

Friedman, J. H., 2001, Greedy function approximation: A gradient boosting machine, Annals of Statistics 29: 1189-1232.

Friedman, J. H., 2002, Stochastic gradient boosting, Computational Statistics and Data Analysis 38: 367-378.

Friedman, J. H., Hastie, T. and Tibshirani, R. 2000, Additive logistic regression: A statistical view of boosting, Annals of Statistics 28: 337-374.

Granger, C. W. J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 24-36.

Hamo, Y., Masulis, R. and Ng, V., 1990, Correlations, price changes and volatility across international stock markets, Review of Financial Studies 3: 281-307.

Hanley, J. A., 1999, Receiver operating characteristic (ROC) curves, in: P. Armitage and T. Colton, Encyclopedia of Biostatistics (J. Wiley, New York), 3738-3745.

Henery, R.J., 1994, Methods for comparison, in D. Michie, D.J. Spiegelhalter, and C.C. Taylor, Machine Learning, Neural and Statistical Classification (Ellis Horwood, Hemel Hempstead, UK), 107-124.

Hirshleifer, D and Teoh, S. H., 2003, Herding and cascading in capital markets: A review and synthesis, European Financial Management 9: 25-66.

Karoly, G. A. and Stulz, R. M., 1996, Why do markets move together? An investigation of U.S.-Japan stock return comovements, Journal of Finance 51: 951-986.

Kasa, K., 1992, Common stochastic trends in international stock markets, Journal of Monetary Economics 29: 95-124.

King, A. M. and Wadhwani, S., 1990, Transmission of volatility between stock markets, Review of Financial Studies 3: 5-33.

Koutmos, G. and Booth, G., 1994, Asymmetric volatility transmission in international stock markets, Journal of international Money and Finance 10: 231-251.

Masih, R. and Masih, A. M. , 1997, Dynamic linkages and the propagation of stock market fluctuations driving major international stock markets: An analysis of the pre- and post-crash eras, Quarterly Review of economics and Finance 37: 859-885.

Masih, R. and Masih, A. M., 2001, Long and short term dynamic causal correlation amongst international stock markets, Journal of International Money and Finance 20: 563-587.

Metz, C.E., Herman, B.A. and Roe, C.A., 1998, Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets, Medical Decision Making 18: 110-121.

Provost, F., Fawcett, T. and Kohavi, R., 1998, The case against accuracy estimation for comparing induction algorithms, in: Jude W. Shavlik, Proceedings of the Fifteenth International Conference on Machine Learning (Morgan Kaufmann, San Francisco, CA), 445-453.

Ridgeway, G., 1999, The state of boosting, Computing Science and Statistics 31: 172-181.

Weiss, G. M. and Provost, F., 2003, Learning when training data are costly: The effect of class distribution on tree induction, Journal of Artificial Intelligence Research 19: 315-354.

Zhou, X.-H., McClish, D- K. and Obuchowski, N. A., 2002, Statistical Methods in Diagnostic Medicine (John Wiley, New York).

Chart 1. Friedman's (Stochastic) Gradient Boosting Machine Algorithm

---

Initialise $\hat{F}_0(\mathbf{x}) = \arg\min \sum_{i=1}^{N} L(y_i, \rho).$

For $m = 1,\dots, M$ do:

1. Given a training sample, $\{y_i, \mathbf{x}_i\}_1^N$, perform a random permutation of the

   integers $\{1,\dots, N\}$ and obtain a sub-sample ($\hat{N} < N$)

   $\{\pi(i)\}_1^N = \text{rand\_perm}\{\text{i}\}_1^N \rightarrow \{y_{\pi(i)}, x_{\pi(i)}\}_1^{\hat{N}}$

2. Compute the negative gradient as the working response/output

   $$g_{\pi(i)m} = -\left.\frac{\partial L(y_{\pi(i)}, F(\mathbf{x}_{\pi(i)}))}{\partial F(\mathbf{x}_{\pi(i)})}\right|_{F(\mathbf{x}_{\pi(i)})=\hat{F}_{m-1}(\mathbf{x}_{\pi(i)})}$$

3. Fit a classification tree, $T(\mathbf{x}_{\pi(i)};\Theta_m)$, predicting $g_{\pi(i)m}$ from the

explanatory variables $\mathbf{x}$.

4. Choose a gradient descent step size as

   $$\rho = \arg\min_{\rho} \sum_{i=1}^{\hat{N}} L(y_{\pi(i)}, \hat{F}_{m-1}(\mathbf{x}_{\pi(i)}) + \rho T(\mathbf{x}_{\pi(i)};\Theta_m))$$

5. Given a learning rate *v*, update the estimate of $F_m(\mathbf{x})$, as

   $$\hat{F}_m(\mathbf{x}) \leftarrow \hat{F}_{m-1}(\mathbf{x}) + v\rho T(\mathbf{x}_{\pi(i)};\Theta_m) \qquad)$$

End For

Table 1. Results for the S&P 500

Summary of movement codification:
S&P 500 Large positive movements: returns > 0.68%
S&P 500 Large negative movements: returns < -0.52%

I.- Assessing the discriminatory power.

A. Using S&P$_{t-1}$ to S&P$_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6120 ± 0.0181	0.5273 ± 0.0180

Test sample: 01/03/2000-06/25/2004

B. Using Nikkei$_{t-1}$ to *Nikkei*$_{t-24}$ and S&P$_{t-1}$ to S&P$_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6109 ± 0.0181	0.5180 ± 0.0180

Test sample: 01/03/2000-06/25/2004

C. Using FTSE$_{t-1}$ to FTSE$_{t-24}$, Nikei$_{t-1}$ to Nikkei$_{t-24}$ and S&P$_{t-1}$ to S&P$_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6075 ± 0.0182	0.5187 ± 0.0180

Test sample: 01/03/2000-06/25/2004

II.- Comparing the 'overall' discriminatory power (z-statistic, one-tailed p-value).

	Large Positive Movements	Large Negative Movements
Panel A vs. Panel B:	(0.0416, 0.4834)	(0.7637, 0.2225)
Panel A vs. Panel C:	(0.5185, 0.3021)	Procedure does not converge
Panel B vs. Panel C:	(0.5302, 0.2980)	(-0.3624, 0.3585)

Table 2. Results for the FTSE 100

Summary of movement codification:
FTSE 100 Large positive movements: returns > 0.74%
FTSE 100 Large negative movements: returns < -0.64%

I.- Assessing the discriminatory power.

A. Using $FTSE_{t-1}$ *to* $FTSE_{t-24}$ *as explanatory variables*

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6169 ± 0.0186	0.5537 ± 0.0185
Test sample: 01/03/2000-06/25/2004		

B. Using $FTSE_{t-1}$ *to* $FTSE_{t-24}$ *and* $S\&Pi_{t-1}$ *to* $S\&P_{t-24}$ *as explanatory variables*

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6643 ± 0.0175	0.6206 ± 0.0176
Test sample: 01/03/2000-06/25/2004		

C. Using $FTSE_{t-1}$ *to* $FTSE_{t-24}$, $Nikei_{t-1}$ *to* $Nikkei_{t-24}$ *and* $S\&P_{t-1}$ *to* $S\&P_{t-24}$ *as explanatory variables*

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6702 ± 0.0174	0.6115 ± 0.0177
Test sample: 01/03/2000-06/25/2004		

II.- Comparing the 'overall' discriminatory power (z-statistic, one-tailed p-value).

	Large Positive Movements	Large Negative Movements
Panel A vs. Panel B:	(-3.1261, 0.0009)	(-4.7505, 0.0000)
Panel A vs. Panel C:	(-3.6474, 0.0001)	(-4.1322, 0.0000)
Panel B vs. Panel C:	(-1.3775, 0.0842)	(1.2333, 0.1088)

Table 3. Results for the Nikkei 225

Summary of movement codification:
Nikkei 225 Large positive movements: returns > 0.82%
Nikkei 225 Large negative movements: returns < -0.82%

I.- Assessing the discriminatory power.

A. Using $Nikkei_{t-1}$ to $Nikkei_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.5350 ± 0.0190	0.5345 ± 0.0188

Test sample: 01/03/2000-06/25/2004

B. Using $Nikkei_{t-1}$ to $Nikkei_{t-24}$ and $FTSE_{t-1}$ to $FTSE_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.5872 ± 0.0184	0.6033 ± 0.0179

Test sample: 01/03/2000-06/25/2004

C. Using $FTSE_{t-1}$ to $FTSE_{t-24}$, $Nikei_{t-1}$ to $Nikkei_{t-24}$ and $S\&P_{t-1}$ to $S\&P_{t-24}$ as explanatory variables

	Large Positive Movements	Large Negative Movements
Discriminatory accuracy:	0.6379 ± 0.0175	0.6635 ± 0.0168

Test sample: 01/03/2000-06/25/2004

II.- Comparing the 'overall' discriminatory power (z-statistic, one-tailed p-value).

	Large Positive Movements	Large Negative Movements
Panel A vs. Panel B:	(-4.1795, 0.0000)	(-4.3122, 0.0000)
Panel A vs. Panel C:	(-6.0609, 0.0000)	(-6.3780, 0.0000)
Panel B vs. Panel C:	(-4.0159, 0.0000)	(-4.1185, 0.0000)