Algorithmic sign prediction and covariate selection across eleven international stock markets

Markku Karhunen[1]

Discipline of Economics, Department of Political and Economic Studies, University of Helsinki, Arkadiankatu 7, 00014 Finland

A R T I C L E   I N F O

A B S T R A C T

I investigate whether an expert system can be used for profitable long-term asset management. The trading strategy of the expert system needs to be based on market predictions. To this end, I generate binary predictions of the market returns by using statistical and machine-learning algorithms. The methods used include logistic regressions, regularized logistic regressions and similarity-based classification. I test the methods in a contemporary data set involving data from eleven developed markets. Both statistical and economic significance of the results are considered. As an ensemble, the results seem to indicate that there is some degree of mild predictability in the stock markets. Some of the results obtained are highly significant in the economic sense, featuring annualized excess returns of 3.1% (France), 2.9% (Netherlands) and 0.8% (United States). However, statistically significant results are seldom found. Consequently, the results do not completely invalidate the efficient-market hypothesis.

1. Introduction

This paper examines whether it is possible to use expert systems for long-term asset management in the stock markets. The decision rule of such expert systems is trivially simple: Invest in stocks if the stock market is likely to rise and invest in the money market if the stock market is likely to decline. However, to this end, one needs predictions of the market movements. According to mainstream opinion in economics, it is impossible to predict the stock markets, as that would generate an arbitrage opportunity. This view is known as the efficient-market hypothesis (EMH; e.g. Fama, 1991). However, there are other schools of thought. For example, the adaptive-markets hypothesis of Lo (2004) states that individuals use simple heuristics to trade in the stock markets, and consequently, they are not completely rational. This seems to contradict EMH. Moreover, there are theoretical constructions within the discipline of neoclassical economics (e.g. Singleton, 2006, Chapter 9) which show that there can be some degree of predictability in the stock markets, even if the assumptions of EMH are in force. Thus, it is a question of obvious empirical interest if the markets can be predicted or not.

The empirical evidence regarding stock market predictability is mixed. In an influential paper, Welch and Goyal (2008) refuted previous reports of market predictability. The argument was that most

authors hitherto had investigated in-sample correlations and the models had no out-of-sample predictive power. Even the in-sample correlations were often lost when the models were updated by new data. Thus, the results could be refuted as statistical artefacts. However, others have challenged the findings of Welch and Goyal (2008). For example, Chevapatrakul (2013) has produced significant out-of-sample predictions for the UK stock market. Similarly, Skabar (2013) and Fiévet and Sornette (2018) have published significant results regarding daily data from the US market. Thus, the debate is ongoing.

In this paper, I use contemporary statistical and machine-learning methods to generate out-of-sample predictions in 11 developed stock markets. The methods considered involve ordinary least squares, logistic regressions, regularized regressions (e.g. Tibshirani, 1996; Zou, 2006) and similarity-based classification (Skabar, 2013). Some authors have reported it to be easier to give a binary prediction of profit or loss than to give an estimate of the expected return (e.g. Leung, Daouk, & Chen, 2000; Nyberg, 2011; Nyberg & Pönkä, 2016). At any rate, it is such sign predictions that the expert system ultimately needs to manage the investment. Thus, I have chosen sign prediction as the objective of this study. I use a combination of statistical tests and trading simulations to assess the potential of the expert system to perform profitable asset management. The rest of this paper is organized as follows. Chapter 2 surveys related work. (The lessons learned from previous work to a large degree guide the modelling choices made in this paper.) Chapter 3 introduces the material and methods. The results are presented in Chapter 4. These are divided in

---

E-mail address: markku.karhunen@gmail.com
[1] Present address: Medical Bioinformatics Centre, Turku Centre for Biotechnology, Tykistökatu 6A, 20520 Turku, Finland

two main categories: Main results (Chapter 4.1) and results obtained from sensitivity analyses (Chapter 4.2). Chapter 5 concludes and presents directions for future work.

2. Related work

There is a large body of literature regarding stock market prediction, both in-sample and out-of-sample. Consequently, it is possible to give only a few references to recent work in this paper. As noted in the Introduction, Welch and Goyal (2008) refuted predictability in the stock markets. Some authors (Fiévet & Sornette, 2018; Lanne, Meitz, & Saikkonen, 2013) have pointed out that predictability in the stock markets is rather non-linear than linear, and consequently, most previous authors have been using methods ill-suited for the problem at hand. On the other hand, many authors have also found linear predictability in the stock markets (e.g. Chevapatrakul, 2013; Nyberg & Pönkä, 2016; Pönkä, 2016). Consequently, I use a combination of linear and non-linear methods in this study. (Linearity here is to be understood in context of the effects of covariates. If the effects of all covariates on the probability of profit are monotone, a model is said to be 'linear'.)

A number of factors are believed to affect the stock markets, thus allowing for profitable prediction. These involve market volatility (Chevapatrakul, 2013), oil price (Gupta & Wohar, 2017; Liu, Ma, & Wang, 2015; Pönkä, 2016) and the lags of the US stock market return (Narayan, Phan, & Narayan, 2018; Nyberg & Pönkä, 2016). Dividend yield, interest rate, industrial production growth and exchange rate growth have also been suggested as potential predictors (Ang & Bekaert, 2007; Rapach, Strauss, & Zhou, 2013), among others. The factors affecting the stock market are believed to be country-specific (Hadhri & Ftiti, 2017). Thus, one would like to have a data set sufficiently rich in covariates to attempt stock market prediction. One such data set is offered by the monthly data of Rapach et al (2013), also analyzed by Nyberg and Pönkä (2016) and Pönkä (2016). These data are also adopted for this paper.

Regarding sign prediction in the stock markets, any method suited for binary classification may be used. A natural starting point are the logit and probit regressions (e.g. Leung et al., 2000). In the machine-learning literature, decision trees are well known, and they have also been applied to stock market prediction (e.g. Fiévet & Sornette, 2018). Same applies to artificial neural networks (Zhong & Enke, 2017a,b) and linear discriminant analysis (Leung et al., 2000). More exotic methods used in this domain include fuzzy robust principal component analysis (Zhong & Enke, 2017a), copulas (Anatolyev & Gospodinov, 2010), empirical mode decomposition (Pan & Hu, 2016) and similarity-based classification (Skabar, 2013) which is also used in this paper.

Previous literature offers a number of suggestions regarding the methods used for testing the expert system. Firstly, Welch and Goyal (2008) stress the importance of out-of-sample testing. Secondly, trading simulations carried out by using historical data are central to the credibility of the algorithm (Nyberg, 2011). Ideally, the trading strategy obtained from a predictive model is robust towards substantial trading costs. It may also be desirable to test the sensitivity of the results towards other assumptions (cf. Narayan et al., 2018). However, strategies may often appear economically profitable, even if the predictions are not statistically significant (Nyberg, 2011). Consequently, it is advisable to test the binary predictions by contrasting them against the reality, e.g. by using the Pesaran–Timmermann test (2009). As a result of these considerations, this paper features trading simulations, statistical significance tests and excessive sensitivity analyses. All of these are performed out of sample, by using rolling windows. Additionally, the predictive accuracy of the methods is compared and a ranking of the methods is attempted. This comparison is based

on the model-confidence set algorithm of Hansen, Lunde, and Nason (2011) which tests all the models against each other.

3. Material and methods

Let us consider a binary variable y_t and a vector of continuous and binary covariates $\mathbf{x}_{t-1}$. The subscript $t-1$ refers to the fact that the covariates are observed in the previous period. The problem is to predict y_t from $\mathbf{x}_{t-1}$. In this paper, y_t is a monthly indicator of profit or loss and $\mathbf{x}_{t-1}$ involves macroeconomic variables. I discuss two types of predictions, binary ($\hat{y}_{mt} = 0, 1$) and continuous ($p_{mt} \in \mathbf{R}$) where $m = 1, \ldots, 9$ denotes the predictive model. The continuous predictions are called probability scores and are model-based estimates of $E(y_t|\mathbf{x}_{t-1})$. The binary predictions are calculated by truncating the probability scores.

Perhaps the most basic predictive method is ordinary least squares (OLS), also known as the linear probability model in this type of setting (Cameron & Trivedi, 2005, Chapter 14). In certain cases, it is possible that OLS predicts $p_{mt} < 0$ or $p_{mt} > 1$, but even if this occurs, OLS may give better binary predictions than other, more sophisticated models. Consequently, OLS is used as a baseline model in this study.

3.1. Logistic regression

Logistic regression is a type of generalized linear model (McCullagh & Nelder, 1989). It is used extensively in many fields of science to model binary data. It is based on the logistic link function

$$\Lambda(x) = \frac{e^x}{e^x + 1}, \ x \in \mathbf{R}. \tag{1}$$

Using this link function, logistic regression can be defined as

$$(y_t|\mathbf{x}_{t-1}) \sim B\big(\Lambda\big(\beta_0 + \boldsymbol{\beta}'\mathbf{x}_{t-1}\big)\big) \tag{2}$$

where B denotes a Bernoulli distribution. This implies the following log-likelihood function

$$
\begin{aligned}
\ell\big(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}\big) = &\sum_{t=1}^{T} y_t \log \Lambda\big(\beta_0 + \boldsymbol{\beta}'\mathbf{x}_{t-1}\big) \\
&+ \sum_{t=1}^{T} (1 - y_t) \log\big(1 - \Lambda\big(\beta_0 + \boldsymbol{\beta}'\mathbf{x}_{t-1}\big)\big)
\end{aligned} \tag{3}
$$

where $\mathbf{X}$ is the matrix of covariates, $\boldsymbol{\beta}$ is the vector of their regression coefficients and β_0 is an intercept. Another popular choice is to use the probit link function in place of Λ. This yields the probit regression model. In this paper, logit regression is taken as a starting point, in line with Anatolyev and Gospodinov (2010) and Chevapatrakul (2013). With a suitable scaling of regression coefficients, the two link functions are virtually indistinguishable (Cameron & Trivedi, 2005, Chapter 14).

Logistic regression if fitted by maximizing (3) over β_0 and $\boldsymbol{\beta}$. If all covariates are included in $\mathbf{X}$, this yields a full model which is prone to overfitting. To this end, one usually performs some sort of model choice. In this paper, I use logistic regression in combination with Akaike's information criterion (AIC; e.g. Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978). I perform stepwise search to minimize AIC and BIC. The idea of this algorithm is that it starts with the full model and then removes and adds covariates one by one, until it reaches a local minimum of AIC or BIC. To summarize, there are three variants of logistic regression in this paper: full model (henceforth, FM) and models chosen by AIC and BIC.

3.2. Regularized regression

In regularized regression, one maximizes a regularized log-likelihood function in place of the usual one. I modify logistic regression (3) into regularized logit regression as

$$\ell^*(\boldsymbol{y}|\mathbf{X}, \boldsymbol{\beta}) = \sum_{t=1}^{T} y_t \log \Lambda\left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)$$
$$+ \sum_{t=1}^{T} (1 - y_t) \log\left(1 - \Lambda\left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)\right) - \lambda P(\boldsymbol{\beta}) \quad (4)$$

where P is some convex function known as the penalty and $\lambda > 0$ determines the strength of the penalty. The choice of P determines the type of regularized regression. I use regularization to avoid overfitting. Thus, it can be seen as an alternative approach to AIC and BIC. I consider four variants of the penalty function $P(\boldsymbol{\beta})$. Three of these are commonplace in literature and have been used in a wide variety of applications across different disciplines. These three are as follows:

L1-norm, giving Lasso regression (Tibshirani, 1996),

$$P(\boldsymbol{\beta}) = \sum_{j=1}^{p} |\beta_j|, \quad (5)$$

a mixture of L1-norm and L2-norm, giving Elastic Net (henceforth, EN; Zou & Hastie, 2005),

$$P(\boldsymbol{\beta}) = \sum_{j=1}^{p} \left(\frac{1-\alpha}{2}\beta_j^2 + \alpha|\beta_j|\right), \ \alpha \in [0, 1], \quad (6)$$

and L2-norm, giving ridge regression (Hoerl & Kennard, 1970),

$$P(\boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j^2. \quad (7)$$

Lasso and EN with $\alpha > 0$ perform variable choice, i.e. they tend to set less important coefficients to zero. Ridge merely offers a shrinkage estimate of the full model (i.e. an estimate such that $\hat{\beta}_j^2 < \hat{\beta}_j^{(ML)2}$ for all j), but it is presented in this paper for completeness.

Additionally, a fourth type of regularized regression is considered. This is the adaptive Lasso of Zou (henceforth, AL; Zou, 2006). It has the penalty function

$$P(\boldsymbol{\beta}) = \sum_{j=1}^{p} \frac{|\beta_j|}{|b_j|^{\gamma}} \quad (8)$$

where b_j is some initial, statistically consistent estimate of β_j and $\gamma > 0$ is a hyper-parameter to be freely chosen. It has been shown that AL has Oracle properties in generalized linear models, including logistic regression. The Oracle properties imply that the estimator is consistent and asymptotically normal, while the coefficients of false predictors tend to be exactly zero. More formally,

$$\lim_{n \to \infty} \mathbf{P}\left(\hat{\beta}_j = 0\right) = 1 \quad (9)$$

for all j such that $\beta_j = 0$. This property is very useful in model choice. For a proof, see Zou (2006).

In regularized regression, λ, α and γ are hyper-parameters and do not have true values as such. In line with this, the type of b_j in (8) is not determined by theory. In EN, I set $\alpha = 0.5$ to represent a compromise between Lasso and ridge. In AL, I set $\gamma = 0.5$, as large values of γ tend to cause numerical problems (personal observation). In AL, I set

$$b_j = \hat{\beta}_j^{(ML)}, \quad (10)$$

i.e. I use the maximum-likelihood estimate from full model. Finally, I use leave-one-out cross validation to choose the value of λ. This

makes Lasso, EN, ridge and AL deterministic algorithms, eliminating random noise. In cross validation, I define the loss function as deviance

$$D(\boldsymbol{y}|\boldsymbol{\beta}) = -2 \sum_{k=1}^{10} \sum_{u \in A_k} \left[y_u \log \Lambda\left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_{u-1}\right)\right.$$
$$\left. + (1 - y_u) \log\left(1 - \Lambda\left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_{u-1}\right)\right)\right] \quad (11)$$

where A_k is the test data on cross validation fold k. Deviance measures the discrepancy between observation and prediction. I fit Lasso, EN, ridge and AL by using the R package glmnet (Friedman, Hastie, & Tibshirani, 2010).

3.3. Similarity-based classification

This method is a variant of Skabar's (2013) similarity-based classification (SiBC). SiBC is a novel machine-learning method which is fast to calculate and robust to non-linearity. Here, I briefly explain the theory of SiBC, noting the modifications I have made.

Let A_ω denote the set where $y_t = \omega \in \{0, 1\}$. SiBC characterizes A_0 and A_1 in terms of their typical $\boldsymbol{x}_k$ values and compares the $\boldsymbol{x}_t$ value of a new observation to these characterizations. In the notation below, it is assumed that t and u are two arbitrary points in time, such that $\boldsymbol{x}_t$ and $\boldsymbol{x}_u$ belong to the same set A_ω. The same steps (12)–(18) are performed for both A_0 and A_1. SiBC characterizes A_0 and A_1 as follows.

Firstly, let us define the distance of $\boldsymbol{x}_t$ and $\boldsymbol{x}_u$ as

$$d^2(\boldsymbol{x}_t, \boldsymbol{x}_u) = (\boldsymbol{x}_t - \boldsymbol{x}_u)'\mathbf{B}^{-1}(\boldsymbol{x}_t - \boldsymbol{x}_u). \quad (12)$$

I define $\mathbf{B}$ as the sample covariance of $\boldsymbol{x}_t$ within class A_ω. In the rare cases where $\mathbf{B}$ is not invertible, I use the Moore-Penrose pseudoinverse in place of $\mathbf{B}^{-1}$ (pseudoinverse in R). Subsequently, the similarity of $\boldsymbol{x}_t$ and $\boldsymbol{x}_u$ is calculated as

$$w(\boldsymbol{x}_t, \boldsymbol{x}_u) = \exp\left\{-\frac{d^2(\boldsymbol{x}_t, \boldsymbol{x}_u)}{2\sigma^2}\right\}. \quad (13)$$

Skabar (2013) discusses many choices of σ, as this parameter dictates how fast similarity falls with distance. I define

$$\sigma^2 = \frac{1}{n^2} \sum_{t \in A_\omega} \sum_{u \in A_\omega} d^2(\boldsymbol{x}_t, \boldsymbol{x}_u). \quad (14)$$

to avoid lengthy bootstrap schemes.

Subsequently, a graph Laplacian $\mathbf{S}$ is calculated for each class as

$$\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \quad (15)$$

where the entries of matrix $\mathbf{D}$ are given by

$$d_{ii} = \sum_{j=1}^{n} w_{ij}, \quad d_{ij} = 0, i \neq j \quad (16)$$

and $\mathbf{W}$ is given by (13), i.e. $w_{ij} = w(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then, $\mathbf{C}$ is calculated as the dominant eigenvector of the equation

$$\mathbf{SC} = \mathbf{C}. \quad (17)$$

$\mathbf{C}$ is called the centrality vector and serves to measure the centrality of each observation within its own class (Skabar, 2013). Notably, $\mathbf{B}$, σ, $\mathbf{W}$, $\mathbf{S}$, $\mathbf{D}$ and $\mathbf{C}$ are class-specific, even though this is not explicit in the notation.

Finally, 12–(17) are used to classify a new observation, denoted here as $\boldsymbol{x}_t$. The likelihood of $\boldsymbol{x}_t$ falling in class ω is defined as

$$p(\boldsymbol{x}_t|\omega) = n_\omega C_\omega(\boldsymbol{x}_t), \ \omega = 0, 1 \quad (18)$$

in SiBC (Skabar, 2013). Above, n_ω is the sample size of class ω and $C_\omega(\boldsymbol{x}_t)$ is defined as

$$C_\omega(\boldsymbol{x}_t) = \sum_{i \in A_\omega} \left(\frac{w_{ti}c_i}{\sqrt{\sum_{j \in A_\omega} w_{tj} \sum_{k \in A_\omega} w_{jk}}}\right) \quad (19)$$

where w_{ti} is obtained by applying (13) to $\boldsymbol{x}_t$ and $\boldsymbol{x}_i$ where $i \in A_\omega$. In (19), c_i and w_{jk} are the entries of matrices $\mathbf{C}$ and $\mathbf{W}$, respectively.

Above, (17) specifies a network-based pattern of likelihood for a new observation $\boldsymbol{x}_t$. The question remains, how to translate this into the probability score p_{mt}. In this paper, I define

$$p_{mt} = \frac{p(\boldsymbol{x}_t|1)}{p(\boldsymbol{x}_t|0) + p(\boldsymbol{x}_t|1)}, \qquad (20)$$

i.e. I use the relative likelihood of $y_t = 1$. Skabar (2013) uses a Bayesian formulation, but this is equivalent to the present choice, if an uninformative prior is used.

3.4. Testing the predictions

3.4.1. Statistical significance

For testing statistical significance, I truncate the probability scores into binary predictions as $\hat{y}_{mt} = 1\{p_{mt} > 0.5\}$ for each model m. Then, I contrast them with the reality (y_t) and calculate the Pesaran–Timmermann test (PT test, 2009). To illustrate the results from PT test, I calculate a summary statistic

$$S_{mt} = sign\left(cor(\hat{y}_{mt}, y_t)\right)\sqrt{\chi_{mt}} \qquad (21)$$

where χ_{mt} is PT test statistic for model m. Under the null of no association, S_{mt} has an asymptotical standard normal distribution (Pesaran & Timmermann, 2009) and is thus easy to interpret.

3.4.2. Model comparison

I perform model comparison by calculating model confidence sets (MCSs) by the R package MCS (Catania & Bernardi, 2015). An MCS can be understood as a set where i) the best model is included and ii) none of the models differ significantly from each other (Hansen et al., 2011). I measure model performance by the quadratic loss function

$$L_{mt} = (y_t - p_{mt})^2 \qquad (22)$$

and assess significance at 0.05 level by using the $T_{\mathcal{R},\mathcal{M}}$ approach (Hansen et al., 2011). In the other approach of Hansen et al. (2011), property ii) is modified so that none of the models differ significantly from the mean performance of the MCS. However, I find the best model a more natural point of comparison.

3.4.3. Trading simulations

I perform trading simulations according to the following algorithm. These simulations demonstrate what would be obtained if an expert system was let to manage investments in the real world.

1. Truncate the probability scores p_{mt} into binary predictions as $\hat{y}_{mt} = 1\{p_{mt} > 0.5\}$.
2. Assume $\hat{y}_{m0} = 0$, indicating that the investor starts with cash.
3. At each time step $t = 1, 2, \ldots,$ if $\hat{y}_{mt} = 1$, invest all money in stocks. If $\hat{y}_{m(t-1)} = 0$, a trading cost of $\zeta_s = 0.50\%$ is incurred.
4. At each time step $t = 1, 2, \ldots,$ if $\hat{y}_{mt} = 0$, invest all money in the risk-free rate. If $\hat{y}_{m(t-1)} = 1$, a trading cost of $\zeta_b = 0.10\%$ is incurred.

The stock return is taken to be the return of the stock market index obtained from data, whereas the risk-free rate is taken to be the three-month interest rate. Parameters ζ_s and ζ_b measure the trading costs, and their values follow the convention in the field (Chevapatrakul, 2013; Nyberg, 2011). However, the values of ζ_s and ζ_b are doubled in the sensitivity analysis, as trading success is likely to be sensitive towards these parameters.

For each model m, a resepective portfolio is obtained. I compare the historical performance of this portfolio to that of the buy-and-hold (BH) strategy. Portfolio m is called mean-variance dominating, if the mean return is higher and the volatility is lower than

for BH. According to standard financial theory, any risk averse investor should favour such portfolios over the BH portfolio. In this analysis, I concern monthly figures, as mean-variance analysis is most relevant in short time scales where the portfolio returns are approximately normal (as opposed to e.g. log-normal).

3.5. Empirical data

I use the same data as Rapach et al. (2013) because these data involve a number of macroeconomic covariates. The data comprise nine variables: Excess stock market return (denoted z_t in the sequel), real oil price, three-month interest rate, dividend yield, CPI inflation, term spread, the growth of industrial production, the US recession indicator, and the growth rate of real exchange rate (for the non-US markets). Some of these covariates have been found highly significant predictors in the previous literature (see Chapter 2). The data have been downloaded from the homepage of David Rapach (http://sites.slu.edu/rapachde/home/research) and they concern the time period 1980/03 - 2010/12. The countries concerned are Australia (AUS), Canada (CAN), France (FRA), Germany (GER), Italy (ITA), Japan (JPN), the Netherlands (NLD), Sweden (SWE), Switzerland (SUI), the United Kingdom (UK) and the United States (US). Industrial production was missing from AUS, ITA and SUI.

I use the lags of the nine variables to predict the binary excess return ($y_t = 1\{z_t > 0\}$). I also consider the lags of the binary returns, and for the non-US data, the lagged US excess return. As some of the variables are highly non-stationary in some of the markets, I take first differences of all predictor variables. The total sample size is $T = 370$, of which one observation is lost due to differentiation. The size of the rolling windows is taken to be 185 and the models are tested by calculating 184 predictions. Thus, the period used for out-of-sample testing is 1995/08 - 2010/11.

4. Results

4.1. Main results

Table 1 presents results from PT test where $\hat{y}_{mt}$ and y_t are compared to each other. This analysis tests the predictive power in the simplest, most intuitive way: Can you tell the correct label or not? Significant associations can be found in four cases: by using SiBC in Canada, by using Lasso or EN in Netherlands, and by using Lasso in Sweden. The point estimate of success ratio appears fairly high (around 0.60) also in a few other cases. However, this is not to be interpreted as significant, because an investor could get more than 50% of the sign returns right merely by guessing that the index is always rising. In fact, the more uneven the distribution of y_t, the easier it is to guess correctly. Consequently, a rigorous statistical test has been used. Furthermore, in Table 1, the summary statistic S_{mt} is sometimes negative, implying a negative correlation of prediction and reality (significant only for SiBC in CAN).

Table 2 presents results from the model comparison exercise. The results are fairly even. Most methods are present in all model confidence sets (AIC, BIC, Lasso, EN, ridge, AL and SiBC). OLS and FM are present in 10 out of 11 MCSs. Based on these results, it is impossible to say which method is the best. Consequently, it is of interest to try variants of the MCS procedure, as has been done in the sensitivity analyses (see Appendix). The results of this exercise are discussed in Chapter 4.2.

Table 3 presents results from the trading simulations. Best results can be found by using ridge in FRA which brings excess returns of 24 monthly basis points (24 mbp), and by using Lasso in NLD (22 mbp). Annualized, this would mean 3.1% excess returns

Table 1

Results of PT test. The format of the numbers is such that the first figure is success ratio, the figure in parenthesis is S_{mt}, and the figure in square brackets is the P value. S_{mt} is a summary statistic calculated out of PT test statistic and it has an asymptotic standard normal distribution. The sign of S_{mt} tells the sign of correlation between prediction and reality. NA indicates cases where PT test cannot be calculated, because the model always predicts either profit or loss. The significance codes are: * significant at 0.05 level, ** significant at 0.01 level, *** significant at 0.001 level.

	AUS	CAN	FRA	GER	ITA	JPN	NLD	SWE	SUI	UK	US
OLS	0.57	0.59	0.6	0.54	0.51	0.51	0.59	0.59	0.6	0.55	0.56
	(−0.63)	(1.3)	(1.3)	(0.15)	(0.39)	(0.12)	(1)	(1.4)	(−0.019)	(−0.28)	(0.57)
	[0.53]	[0.21]	[0.18]	[0.88]	[0.7]	[0.9]	[0.31]	[0.16]	[0.98]	[0.78]	[0.57]
FM	0.57	0.58	0.59	0.54	0.5	0.52	0.59 (0.84)	0.6	0.6	0.54	0.57
	(−0.53)	(1.1)	(1.2)	(0.18)	(0.22)	(0.47)	[0.4]	(1.7)	(−0.16)	(−0.44)	(0.98)
	[0.59]	[0.26]	[0.24]	[0.86]	[0.82]	[0.64]		[0.083]	[0.88]	[0.66]	[0.33]
AIC	0.59	0.58	0.55	0.54	0.52	0.49	0.57 (0.32)	0.6	0.6	0.54	0.53
	(−0.026)	(0.79)	(0.33)	(0.32)	(0.81)	(−0.63)	[0.75]	(1.8)	(−0.29)	(−0.68)	(−0.81)
	[0.98]	[0.43]	[0.74]	[0.75]	[0.42]	[0.53]		[0.071]	[0.77]	[0.49]	[0.42]
BIC	0.62	0.58	0.61 (1.1)	0.57	0.52	0.47	0.61 (0.93)	0.6	0.62	0.59	0.54
	(0)	(0.69)	[0.27]	(0.58)	(0.56)	(−0.24)	[0.35]	(1.5)	(−0.83)	(−0.88)	(−1.8)
	[NA]	[0.49]		[0.56]	[0.57]	[0.81]		[0.13]	[0.41]	[0.38]	[0.065]
Lasso	0.61	0.57	0.61	0.57	0.53	0.51	0.64 (2.3)	0.61	0.64	0.6	0.56
	(−0.085)	(−0.57)	(0.69)	(1.1)	(1.1)	(0.083)	[0.022] *	(2.1)	(0.56)	(1.2)	(−1.1)
	[0.93]	[0.57]	[0.49]	[0.27]	[0.29]	[0.93]		[0.04] *	[0.58]	[0.22]	[0.26]
EN	0.61	0.56	0.61	0.56	0.53	0.51	0.64 (2.7)	0.61	0.63	0.6	0.57
	(−0.085)	(−0.82)	(0.44)	(0.65)	(1.2)	(0.22)	[0.0079] **	(1.8)	(0.23)	(1.2)	(−0.87)
	[0.93]	[0.41]	[0.66]	[0.51]	[0.25]	[0.83]		[0.073]	[0.82]	[0.24]	[0.39]
ridge	0.62	0.61	0.63	0.59	0.49	0.49	0.61 (0.13)	0.6	0.63	0.59	0.59
	(0)	(0.71)	(1.4)	(1.3)	(−0.28)	(−0.81)	[0.9]	(1.6)	(−0.053)	(0)	(0.63)
	[NA]	[0.48]	[0.15]	[0.19]	[0.78]	[0.42]		[0.12]	[0.96]	[NA]	[0.53]
AL	0.61	0.58	0.6	0.58	0.52	0.54	0.61 (0.77)	0.6	0.61	0.58	0.57
	(0.85)	(0.1)	(0.28)	(1.5)	(0.93)	(1.2)	[0.44]	(1.7)	(−0.7)	(−0.98)	(−0.24)
	[0.39]	[0.92]	[0.78]	[0.14]	[0.35]	[0.24]		[0.087]	[0.48]	[0.33]	[0.81]
SiBC	0.62	0.54	0.61	0.58	0.44	0.48	0.6 (−1.1)	0.58	0.65	0.6	0.61
	(0.21)	(−2.5)	(0.56)	(0.71)	(−0.41)	(−0.72)	[0.28]	(−0.048)	(1.5)	(1.4)	(1.6)
	[0.83]	[0.014] *	[0.58]	[0.48]	[0.68]	[0.47]		[0.96]	[0.14]	[0.16]	[0.099]

Table 2

Model confidence sets (MCSs) based on mean squared error. The figures in parenthesis are the mean squared errors. The most accurate method for each market is indicated by an asterisk, but the models within an MCS cannot be statistically distinguished from each other.

AUS	OLS (0.26), FM (0.26), AIC (0.25), BIC (0.24), Lasso (0.24), EN (0.24)*, ridge (0.24), AL (0.24), SiBC (0.24)
CAN	OLS (0.25), FM (0.25), AIC (0.24), BIC (0.25), Lasso (0.25), EN (0.25), ridge (0.24)*, AL (0.24), SiBC (0.25)
FRA	OLS (0.26), FM (0.25), AIC (0.26), BIC (0.25), Lasso (0.24), EN (0.24)*, AL (0.25), SiBC (0.24)
DEU	AIC (0.26), BIC (0.26), Lasso (0.26), EN (0.26), ridge (0.25)*, AL (0.25), SiBC (0.25)
ITA	OLS (0.27), FM (0.27), AIC (0.27), BIC (0.26), Lasso (0.25), EN (0.25)*, ridge (0.25), AL (0.26), SiBC (0.26)
JPN	OLS (0.27), FM (0.27), AIC (0.26), BIC (0.26), Lasso (0.25), EN (0.25), ridge (0.25), AL (0.25), SiBC (0.25)*
NLD	OLS (0.25), FM (0.25), AIC (0.25), BIC (0.24), Lasso (0.23)*, EN (0.23), ridge (0.24), AL (0.24), SiBC (0.24)
SWE	OLS (0.25), FM (0.24), AIC (0.24), BIC (0.25), Lasso (0.24), EN (0.24), ridge (0.24)*, AL (0.24), SiBC (0.25)
CHE	OLS (0.24), FM (0.24), AIC (0.23), BIC (0.24), Lasso (0.24), EN (0.24)*, ridge (0.24), AL (0.23), SiBC (0.24)
UK	OLS (0.26), FM (0.26), AIC (0.26), BIC (0.25), Lasso (0.24), EN (0.24)*, ridge (0.24), AL (0.25), SiBC (0.24)
US	OLS (0.26), FM (0.26), AIC (0.26), BIC (0.26), Lasso (0.25), EN (0.25), ridge (0.25), AL (0.25), SiBC (0.25)*

for FRA, and 2.9% excess returns for NLD. However, a more typical example is perhaps OLS in US which has 6 mbp, yielding 0.77% annualized excess returns. Based on the trading simulations, AL and SiBC seem to be promising methods, as they can generate excess returns in most of the markets concerned (in six and seven markets, respectively). However, statistical significance (Table 1) coincides with economic significance only in three cases: For Lasso and EN in NLD, and for Lasso in SWE. It is a common result in literature that statistical significance and results from trading simulations do not coincide (e.g. Nyberg, 2011). Given the fact that the stock markets are notoriously difficult to forecast, the predictive power of algorithms is bound to be small. Consequently, the present sample size ($T = 184$ used for testing the predictions) may be too small to detect any significant effects which may partly explain the negative results of Table 1.

Fig. 1 illustrates the trading simulations in four markets. Ridge in FRA beats consistently the index. Yet, it was found statistically insignificant in Table 1. EN in ITA is neither economically nor statistically significant. Yet, it offers a relatively smooth performance as a result of decreased volatility. Lasso in NLD and SWE is economically and statistically significant in both markets. Yet, in NLD, it first tracks the index and then for a while underperforms, with

the economic gains occurring only towards the end of the sampling period. This illustrates how appealing summary statistics may be obtained, even if the results are not consistent in time (cf. Welch & Goyal, 2008; Liu et al., 2015).

4.2. Robustness

This Chapter presents a summary of the results obtained in sensitivity analyses. These analyses are explained in full detail in Appendix. The names of the analyses here refer to Appendix, such that the first part of the name indicates the primary analysis, e.g. *PT-1* is a perturbation of the Pesaran-Timmermann test. Analyses *X-1*, *X-2* and *X-3* cover all of the analyses presented in the main matter.

The results are somewhat sensitive towards the choice of statistical test (*PT-1*). Positive results of PT test and Fisher's exact test only coincide for Lasso and EN in NLD. On the other hand, both tests produce null results in most cases. Relaxing the significance threshold of PT test does not much improve the results (*PT-2*). As expected, changing the loss function in the MCS algorithm (*MCS-1*) changes the ranking of the methods, but the results are still

Table 3

Results of the trading simulations. The format of the numbers is such that the first figure is the mean return, the figure in parenthesis is the volatility and the figure in square brackets is the Sharpe's ratio. Portfolios which mean-variance dominate the buy-and-hold portfolio (BH) are indicated by an asterisk. All figures are monthly percentage points.

	AUS	CAN	FRA	GER	ITA	JPN	NLD	SWE	SUI	UK	US
BH	0.41	0.61	0.58	0.53	0.4	−0.0021	0.53	0.83	0.61	0.29	0.43
	(3.8)	(4.7)	(5.6)	(6.2)	(6.4)	(5.2)	(5.9)	(6.4)	(4.7)	(4.2)	(4.7)
	[0.11]	[0.13]	[0.1]	[0.085]	[0.064]	[−0.00031]	[0.089]	[0.13]	[0.13]	[0.068]	[0.093]
OLS	0.14	0.54	0.58	0.34	0.13	0.0028	0.65	0.73	0.52	0.23	0.49
	(3.5)	(3.3)	(3.9)	(5)	(4.2)	(3.3)	(5.1)	(5.3)	(4.3)	(3.3)	(3.9)
	[0.04]	[0.17]	[0.15] *	[0.07]	[0.031]	[0.00098] *	[0.13] *	[0.14]	[0.12]	[0.07]	[0.13] *
FM	0.12	0.52	0.53	0.32	0.1	0.033	0.64	0.79	0.46	0.21	0.51
	(3.5)	(3.3)	(3.9)	(4.9)	(4.2)	(3.3)	(5.1)	(5.3)	(4.3)	(3.3)	(3.9)
	[0.036]	[0.16]	[0.14]	[0.065]	[0.026]	[0.01] *	[0.12] *	[0.15]	[0.11]	[0.063]	[0.13] *
AIC	0.25	0.58	0.32	0.44	0.31	−0.061	0.68	0.84	0.44	0.099	0.31
	(3.6)	(3.5)	(4.2)	(5.1)	(3.9)	(3.1)	(4.9)	(5.3)	(4.4)	(3.8)	(3.9)
	[0.068]	[0.17]	[0.076]	[0.087]	[0.08]	[−0.019]	[0.14] *	[0.16] *	[0.1]	[0.027]	[0.08]
BIC	0.41	0.66	0.57	0.41	0.29	−0.18	0.59	0.79	0.5	0.27	0.31
	(3.8)	(3.8)	(5)	(5.7)	(4.8)	(3.5)	(5.6)	(5.7)	(4.6)	(4.2)	(4.5)
	[0.11]	[0.17] *	[0.11]	[0.071]	[0.062]	[−0.052]	[0.11] *	[0.14]	[0.11]	[0.064]	[0.068]
Lasso	0.35	0.5	0.65	0.51	0.27	−0.037	0.75	0.95	0.65	0.36	0.38
	(3.8)	(4.2)	(5)	(5.5)	(4.1)	(3)	(5.6)	(5.8)	(4.6)	(4.1)	(4.3)
	[0.093]	[0.12]	[0.13] *	[0.093]	[0.068]	[−0.012]	[0.13] *	[0.16] *	[0.14] *	[0.088] *	[0.089]
EN	0.35	0.49	0.63	0.47	0.36	−0.032	0.69	0.87	0.63	0.31	0.4
	(3.8)	(4.3)	(5)	(5.5)	(4)	(3)	(5.7)	(5.9)	(4.6)	(4.2)	(4.3)
	[0.093]	[0.12]	[0.13] *	[0.086]	[0.091]	[−0.011]	[0.12] *	[0.15] *	[0.14] *	[0.073] *	[0.092]
ridge	0.41	0.76	0.81	0.6	0.028	−0.0078	0.48	0.82	0.63	0.29	0.49
	(3.8)	(4.1)	(4.7)	(5.8)	(4.2)	(3.3)	(5.8)	(5.9)	(4.7)	(4.2)	(4.4)
	[0.11]	[0.18] *	[0.17] *	[0.1] *	[0.0078]	[−0.0022]	[0.082]	[0.14]	[0.14] *	[0.068]	[0.11] *
AL	0.35	0.62	0.67	0.55	0.33	0.17	0.49	0.98	0.56	0.24	0.44
	(3.8)	(4.1)	(4.7)	(5.4)	(4.3)	(2.9)	(5.8)	(5.4)	(4.4)	(4)	(4.3)
	[0.094]	[0.15] *	[0.14] *	[0.1] *	[0.077]	[0.059] *	[0.085]	[0.18] *	[0.13]	[0.059]	[0.1] *
SiBC	0.41	0.38	0.63	0.54	−0.34	0.029	0.51	0.78	0.76	0.3	0.57
	(3.8)	(4.5)	(5.2)	(5.6)	(3.4)	(3.5)	(5.9)	(6.4)	(4.5)	(4.1)	(4.2)
	[0.11] *	[0.085]	[0.12] *	[0.096] *	[−0.097]	[0.0084] *	[0.086]	[0.12]	[0.17] *	[0.075] *	[0.13] *

very even. Likewise, relaxing the significance threshold (*MCS-2* and *MCS-3*) does not much help ranking the methods.

The results are surprisingly robust towards increasing the trading costs (*Trading-1*). For example, in Netherlands, mean-variance dominating portfolios can still be constructed by using six methods. Lasso can still be used to obtain economically significant results in five markets and the best annual excess returns still exceed 2%. Eliminating the trading costs (*Trading-2*) greatly improves the economic significance of the results. If the costs are eliminated, most methods can be used to construct mean-variance dominating portfolios in the majority of markets concerned. The use of soft boundaries (i.e. acting on basis of p_{mt}, not $\hat{y}_{mt}$) gives mixed results, improving the performance in some markets and impairing it in others (*Trading-3*).

Using GARCH-based volatility (Engle & Kroner, 1995) as a covariate (*X-1*) gives mixed results and changes the pattern of statistical significance. This is contrary to earlier findings (Chevapatrakul, 2013). Some of the most interesting results are found in *X-2* where y_t is re-defined as $y_t = 1\{z_t > -0.02\}$, i.e. as a large downturn not occurring. The results are generally better than for the original response variable $y_t = 1\{z_t > 0\}$. This is reflected both in the PT test and in trading simulations. Some of the PT test results are highly significant, and statistically significant predictions can now be generated in the majority of markets. In line with this, mean-variance dominating portfolios can now be generated in all markets. The best annual excess returns exceed 4% in this analysis. Thus, predicting large downturns seems as a profitable direction for future research. However, some of the models always predict $y_t = 1$. This can be seen as a limitation, because in these cases, the models cannot do any better than BH. This behaviour seems to be more common for ridge and SiBC than for other methods.
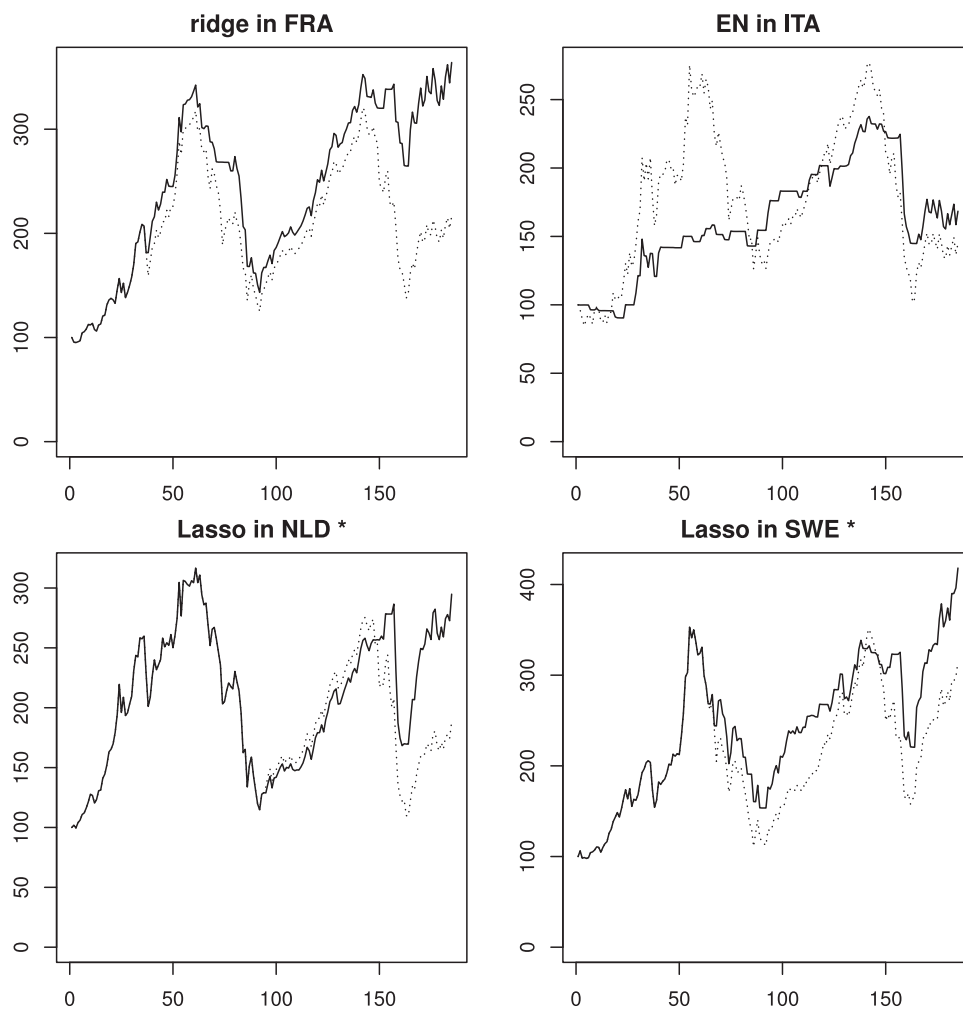
In *X-3*, random forests (Breiman, 2001) and artificial neural networks (e.g. Ripley, 1996) are included in the suite of models. Ran-

dom forests generate economically and statistically significant results for the UK market (9 mbp and $P = 0.002$). Neural networks do not produce any statistically significant results and underperform in the model comparison. This may partly result from the fact that these two methods (random forests and neural networks) are used here with default parameter values to save computation time. Otherwise, the inclusion of new methods does not change the results.

5. Conclusions

This paper demonstrates how algorithmic methods and macroeconomic covariates can be used to construct an expert system to perform independent trading in the stock markets. Some of the results are qualitatively appealing, but statistically significant results are seldom found. Moreover, for a large number of statistical tests, positive results are bound to occur randomly. In Table 1, there are four significant associations, about as many as would be expected at random (4.95). In the trading test (Table 3), there are 39 mean-variance dominating portfolios, significantly less would be expected if the portfolios were equally likely to be mean-variance dominating or not ($P = 0.02$). However, if the trading costs are removed (Appendix), there are 61 dominating portfolios, more than would be expected at random ($P = 0.008$).

Interesting results are found by using Lasso and Elastic Net in Netherlands. These results are both statistically and economically significant (Tables 1 and 3, respectively) and they are robust both towards the choice of statistical test and increasing the trading costs (Appendix). Moreover, very interesting results are found if the target of the analysis is shifted, i.e. if the algorithms are used to predict large downturns (more than 2% monthly) in place of sign return. In that case, statistically significant predictions can be generated in most markets and economically significant results can be obtained in all markets. This result gives hope that there is use for

Fig. 1. Within each panel, the solid line represents the model-based portfolio, with the initial investment scaled as 100. The dotted line represents the value of the total return index (BH strategy in the text). The investment horizon in this figure is 1995/08 - 2010/11 and the unit of x axis is months. * refers to statistical significance at 0.05 level (see Results). The annualized mean returns of the model-based portfolios are 10.2%, 4.4%, 9.4% and 12.0% in FRA, ITA, NLD and SWE, respectively. These may be compared to the annualized mean returns of the BH strategy (7.1%, 5.0%, 6.5% and 10.4% in FRA, ITA, NLD and SWE, respectively).

expert systems in long-term asset management. The question is, what to predict and by using which data.

Based on these considerations, the following directions for future research may be given:

1. Development of standardized, high-quality data sets which involve meaningful covariates. Then, these data sets may be used as a basis for benchmarking different methods and comparing them to each other. The data set of Rapach et al. (2013), also used in this paper, is emerging as one such data set, but it is starting to be rather old, with the latest observations coming from 2010.

2. Use of higher-frequency, e.g. daily data. While daily stock market data are easily available, the covariates seldom are. Most macroeconomic time series are updated less frequently than financial data, i.e. monthly, quarterly or even annually.

3. More thorough analysis of the large downturns. In the sensitivity analyses of this paper, it was found that large downturns are easier to predict than the sign return. It would be of interest to see, if this result holds more generally, or whether it is a peculiar feature of these data.

4. Analysis of volatility. Volatility (i.e. standard deviation of the instantaneous return) to a large degree determines the frequency of large downturns. In the econometric literature, it is customary to model volatility by using GARCH-type models (Engle &

Kroner, 1995). These models have some predictive power, but it is possible that other methods could give more accurate predictions. For example, one might use the machine-learning methods presented in this paper to predict the volatility.

As an ensemble, the results presented in this paper seem to indicate that there is some degree of mild predictability in the stock markets. This conclusion is based on the large number of significant results found in the sensitivity analyses, i.e. when the trading costs are removed or when the large downturns are predicted. However, it is more difficult to say when and where this predictability occurs and which method should be used to exploit it. An exception to this rule are the results found for Netherlands in the main analysis. Consequently, the results do not serve to invalidate EMH. It is conceivable that the struggle between the proponents of EMH and other schools of thought (e.g. Lo, 2004) will go on for the foreseeable future. One can only hope that the results from meticulous data analyses and applications of expert systems help to inform this discussion.

Declaration of interest

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eswa.2018.07.061.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Anatolyev, S., & Gospodinov, N. (2010). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics, 28*, 232–245.

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies, 20*, 651–707.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. New York, NY: Cambridge University Press.

Catania, L., & Bernardi, M.. *MCS: Model confidence set procedure* R package version 0.1.3 https://CRAN.R-project.org/package=MCS.

Chevapatrakul, T. (2013). Return sign forecasts based on conditional risk: Evidence from the UK stock market index. *Journal of Banking & Finance, 37*, 2342–2353.

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory, 11*, 122–150.

Fiévet, L., & Sornette, D. (2018). Decision trees unearth return sign predictability in the S&P 500. *Quantitative Finance*. doi:10.1080/14697688.2018.1441535.

Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance, 46*, 1575–1617.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*, 1–22.

Gupta, R., & Wohar, M. (2017). Forecasting oil and stock returns with a Qual VAR using over 150 years off data. *Energy Economics, 62*, 181–186.

Hadhri, S., & Ftiti, Z. (2017). Stock market predictability in emerging markets: Does the choice of predictors and models matter across countries? *Research in International Business and Finance, 42*, 39–60.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*, 453–497.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*, 55–67.

Lanne, M., Meitz, M., & Saikkonen, P. (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics, 13*, 682–705.

Leung, M. T., Daouk, H., & Chen, A-S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting, 16*, 173–190.

Liu, L., Ma, F., & Wang, Y. (2015). Forecasting excess stock returns with crude oil market data. *Energy Economics, 48*, 316–324.

Lo, A. W. (2004). The adaptive market hypothesis. *The Journal of Portfolio Management, 30*, 15–29.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd Ed.). Boca Raton, FL: Chapman & Hall/CRC.

Narayan, P. K., Phan, D., & Narayan, S. (2018). Technology-investing countries and stock return predictability. *Emerging Markets Review* in press.

Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting, 27*, 561–578.

Nyberg, H., & Pönkä, H. (2016). International sign predictability of stock returns: The role of the United States. *Economic Modelling, 58*, 323–338.

Pan, Y., & Hu, Y. (2016). Empirical mode decomposition-based models for predicting direction of stock index movement. *International Journal of Data Science, 1*, 205–226.

Pesaran, M. H., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association, 485*, 325–337.

Pönkä, H. (2016). Real oil prices and the international sign predictability of stock returns. *Finance Research Letters, 17*, 79–87.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *Journal of Finance, 68*, 1633–1662.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge, UK: Cambridge University Press.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Singleton, K. J. (2006). *Empirical dynamic asset pricing*. Princeton, NJ: Princeton University Press.

Skabar, A. (2013). Direction-of-change financial time series forecasting using a similarity-based classification model. *Journal of Forecasting, 32*, 409–422.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267–288.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies, 21*, 1455–1508.

Zhong, X., & Enke, D. (2017a). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications, 67*, 126–139.

Zhong, X., & Enke, D. (2017b). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. *Neurocomputing, 267*, 152–168.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association, 101*, 1418–1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B, 67*, 301–320.