# Predicting Euro Stock Markets

Ioannis Praggidis(✉), Vasilios Plakandaras, and Eirini Karapistoli

Department of Economics, Democritus University of Thrace,
Panepistimioupoli, 69100 Komotini, Greece
{gpragkid,vplakand}@econ.duth.gr, ikarapis@duth.gr
http://www.econ.duth.gr/index.en.shtml

**Abstract.** Forecasting exercises are mostly concentrated on the point estimation of future realizations of stock returns. In this paper we try to forecast the direction of the Eurostoxx 50. Under a Dynamic Probit framework we test whether subsequent sign reversals can be accurately forecasted. To this end, we make use of industrial portfolios constructed in the spirit of Fama and French. Furthermore, we augment the forecasting models with macroeconomic variables. Finally, we construct a new sentiment index based on the news for Oil prices. Results show, that the out-of-sample forecasting accuracy approximates 80%.

**Keywords:** Eurostoxx 50 · Portfolio industries · Dynamic probit models · Uncertainty

## 1 Introduction

The efficient market hypothesis Fama [9] postulates that publicly available information cannot be used to predict stock returns. Nevertheless, there are numerous empirical studies that reach to a contradicting outcome. Among others Chen [6], Nyberg [19,20], Driesprong *et al.* [8] and Hong *et al.* [14] provide significant evidence that macroeconomic variables, industry portfolios and oil prices can predict future movements in stock markets. Nevertheless, the vast majority of the existing literature studies the U.S. stock market, providing little empirical results to forecasting European stock indexes.

In this paper we contribute to the recent literature by examining the ability to forecast the Eurostoxx 50 index. Interestingly Eurostoxx 50 is an under investigated index although refers to the financial conditions in the Eurozone. In order to do so we first construct the 38 industrial portfolio using almost 4500 stocks traded in European stock exchange markets in the vein of Fama and French [10]. According to Hong *et al.* [14], the gradual information diffusion hypothesis explains why information extracted from specific sectors can act as leading indicators for the market index. The basic idea is that certain investors, such as those that specialize in trading the broad market index, receive information originating from particular industries such as commercial real estate or commodities like metals only with a lag. In contrast, investors that specialize in a specific market cannot detect the trends of the entire stock market and


A. Satsiou et al. (Eds.): IFIN and ISEM 2016, LNCS 10078, pp. 81–97, 2016.
DOI: 10.1007/978-3-319-50237-3_3

thus are unable to transmit efficiently their specialized knowledge of the sector to the entire market. Combining the information from different sources (industries) we expect to be able to grasp their informational content in forecasting the Eurostoxx 50 index.

In order to measure the forecasting ability of our models to the market efficiency assumption, we use the Random Walk model as a benchmark. The use of the RW model as a forecasting benchmark is also a contemporaneous testing of the Efficient Market Hypothesis (EMH). Proposed by Eugene Fama [9], states that the determination of prices in an efficient market follows a random walk and thus it is impossible to create a forecasting model that achieves sustainable positive returns on the long-run. The EMH is usually presented in three forms; the weak, the semi-strong and the strong form of efficiency. We have a weak-form efficient market when historic prices of the variable in question cannot forecast the future ones, as the generating mechanism of prices follows a RW. Thus, autoregressive models, in these cases, have no forecasting power and the best forecast about next period's price is today's price. Semi-strong efficiency imposes more strict assumptions in that all historic prices and all publicly available information is already reflected in current asset prices and thus they cannot be used successfully in forecasting. Finally, the strong EMH builds on the semi-strong case adding all private information and thus making impossible to forecast successfully the future evolution of an asset's price.

Furthermore, Driesprong *et al.* [8] points that changes in oil prices predict stock market returns worldwide finding significant predictability in both developed and emerging markets. Following Driesprong *et al.* [8] we use a new developed proxy for sentiment in oil market. We test the predictability of a sentiment index as sentiment measure acts as a leading indicator for oil prices [24].

The remainder of this paper is organized as follows. In Sect. 2, we outline the recent advances in the area of market sentiment analysis. In Sect. 3, we present the methodology we followed in our market predictive sentiment analysis and in Sect. 4, we describe the textual and market data that are used as input in our system. In Sect. 5, we provide a detailed description of the probit model and its specifications, while empirical results from our analysis are provided in Sect. 6. Finally, Sect. 7 provides concluding remarks and points towards future research avenues.

## 2 News Media Sentiment and the Stock Market

The recent years, there has been a significant amount of work on using textual resources from the Web (such as financial news articles, online reviews, blogs, Twitter, etc.) to predict the stock market as well as financial and economic variables of interest [27]. Much of this work has relied on some form of *sentiment analysis* to represent the text [17]. Sentiment analysis targets open issues in various fields (including politics, psychology, finance, and society), because sentiments' understanding can largely impact interactions, policies, and decision-making. Due to its importance, this form of textual information processing has become a growing part of the empirical finance research.

Typically, the two major sentiment analysis tasks studied today are *subjectivity* detection and *polarity* detection. The most common approach to deal with these tasks is either to train a machine learning classifier on a labeled corpus (supervised learning) and apply the learned model on the desired test set, or use a predefined dictionary consisting of words that are annotated with their semantic orientation value (polarity and strength). The lexicon-based approaches are based on the assumption that the polarity of a given text is the sum of the semantic orientation of each word or phrase contained in it (unsupervised learning). Limited work has also been conducted on hybrid methodologies as well as on ontology-supported approaches.

### 2.1 Lexicon-Based Approaches

The lexicon-based approach is an unsupervised technique that extracts the sentiment of a text using lexicons (dictionaries) that may be either domain-specific (such as finance, politics, psychology, etc.) or domain-independent [29]. Such dictionaries consist of words that are annotated with their semantic orientation value (polarity and strength); usually a positive or negative value is assigned to each word (for example, the score for the word "good" is 0.65 and the score for the word "bad" is 0.85).

Dictionaries for lexicon-based approaches can be created either manually [25,28], or automatically, using seed words to expand the list of words [13,29,30]. Much of the lexicon-based research has focused on using adjectives as indicators regarding the semantic orientation of the text [13,15,26,31]. According to this, a list of adjectives and corresponding semantic orientation values is compiled into a dictionary. Then, for any given text, all adjectives are extracted and annotated according to their semantic orientation value using the dictionary scores. The semantic orientation scores are in turn aggregated into a single score for the text. However, lexicon-based approaches suffer from their absolute dependence on lexicons, which are often characterized by word shortage or inappropriate assignment of semantic orientation values [18].

### 2.2 Machine Learning Approaches

The text classification approach is essentially a supervised classification task, which involves building classifiers from labelled instances of texts or sentences. The introduction of the machine learning approach in sentiment analysis originates from Bo Pang *et al.* [21]. This approach is also referred to as a statistical or machine learning approach. The majority of the statistical text classification research relies on Nave Bayes, Support Vector Machines (SVM), and Max entropy classifiers trained on a particular data set using features such as unigrams or bigrams, and with or without part-of-speech (POS) labels [21,23].

Classifiers built using the supervised methods may achieve high accuracy when detecting the polarity of a text [1,3,4]. However, although such classifiers perform very well in the domain that they are trained on, their performance degrades when the same classifier is used in a different domain [2]. Moreover, utilizing an individual classifier might result in poor sentiment detection, since the performance of each classifier varies significantly, when for instance someone is using different features or weight measures. Therefore, for overcoming the deficiencies of each classifier and to proceed with a more robust and successful sentiment detection process, ensemble classifiers are created, which represent a combination of multiple classifiers Xia *et al.* [32].

### 2.3 Hybrid Approaches

It became apparent that the lexicon-based approaches suffer from their absolute dependence on lexicons, which are often characterized by word shortage or inappropriate assignment of semantic orientation values. While, machine-learning approaches overcome these limitations, their need for a large volume of training data also lessens their advantage [12].

Accordingly, the hybrid approach targets at solving these limitations by combining the aforementioned approaches. An exemplar use case scenario for a hybrid approach is to follow a two-step process; i.e., to initially generate a set of training data by automatically identifying the texts' semantic orientation score (using a lexicon-based approach), and then to proceed with the classification (using a machine learning approach) that is independent from the lexicons' limitations [11]. As Prabowo and Thelwall [22] point out, hybrid approaches enhance the stability and accuracy of existing methodologies while exploring the strong characteristics of each of them.

### 2.4 Ontology-Based Approaches

Recently, semantic web-based sentiment analysis that relies on the usage of ontologies has started to attract attention. Ontologies are not a classification method in itself, but they can be used to enhance existing classification methods. Zhou and Chaovalit [33] were among the first to use ontologies for this purpose. They incorporated an ontology to a supervised machine learning technique and a basic term counting method. In their work, as in most other works, the ontology contained semantic features and was used only to identify these features inside the text. Cadilhac *et al.* [5] on the other hand, they did not use the ontology only as a taxonomy by taking into account the "is a relation" between concepts. Instead, they presented an approach that uses other relations between concepts for creating summaries of opinions and computing the SO scores.

### 2.5 Our Approach

In this paper, we tackle the problem of polarity detection in financial news headlines. Using a simple Java client API (with no third party libraries), we query the Virtuoso SPARQL endpoint of PoolParty[1]; a Thesaurus Management Tool (TMT) for the Semantic Web, and fetch 14205 date-stamped news headlines

[1] https://www.poolparty.biz.

related to the Eurostoxx 50 index. We then create two different representations for the gathered textual resources (one based on machine learning and one on lexicon-based learning) and train one classifier for each one of them.

## 3 Methodology

In the current work, the relevant market-predictive sentiment analysis process is considered to have four major phases, namely: information retrieval, pre-processing, model creation and prediction, and evaluation (see Fig. 1). Information retrieval is the activity of obtaining information resources relevant to a financial instrument from a collection of online information sources. Our search was based on full-text indexing and was restricted to the digital archive of the `investing.com` website. All those articles containing one of the following terms: 'oil prices', 'crude oil', 'brent', 'future prices', 'WTI' or 'OPEC' (including their variants) were stored in PoolParty.

After acquiring the textual resources, we process the input data so that it can be fed into our models. Natural language processing (NLP) of document texts includes among others part-of-speech (POS) tagging, chunking and named entity recognition. We can also apply tokenisation, sentence splitting, and morphological analysis in order to further clean the text [7]. Having transformed the unstructured text into a representative format that is structured, it can now be processed by the machine. Two model representations of the news headlines are created (one model based on text-based learning and one on lexicon-based learning). For the text-based representation, we used the binary, term frequency (tf) and $tf \cdot idf$ n-gram models. We set $n = 1, 2$ resulting into six representations in total. Stop-word removal and stemming were ignored. For the lexicon-based representation, we used the Harvard dictionary[2]. Instead of assigning the majority class label on each news headline, we counted the sum of nouns, verbs, adjectives, and adverbs as indicated by [15]. In the presence of negation, the polarity value of the term was inverted. We used these four features to learn the lexicon-based model, and we compare our results with the simple counting method of the lexicon itself.

Regarding the model creation and prediction phase, the proposed system is using Multinomial Naive Bayes (MNB) and Support Vector Machines (SVM) as provided by Weka[3]. The MNB model is a popular method for text classification due to its computational efficiency and relatively good predictive performance. MNB computes the posterior probability of a class, based on the multinomial distribution of the words in the document. The model works with the bag-of-words (BOWs) feature extraction which ignores the position of the word in the document. Finally, it uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label. On the other hand, the main principle of SVMs is to determine linear separators in the search space which can best separate the different classes. Text data are ideally suited for SVM classification

[2] http://www.wjh.harvard.edu/~inquirer/.
[3] http://www.cs.waikato.ac.nz/ml/weka/.

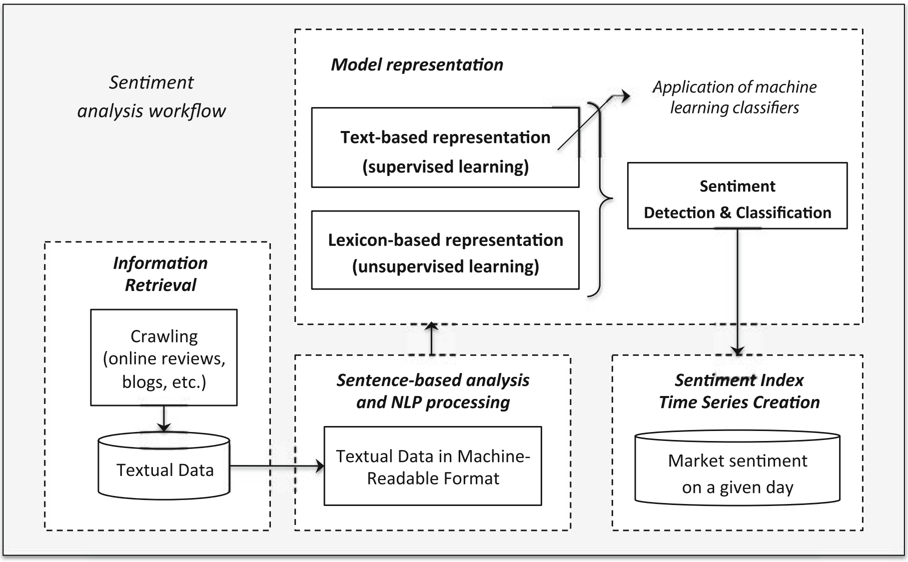


**Fig. 1.** The proposed market-predictive sentiment analysis process.

because of the sparse nature of text, in which few features are irrelevant, but they tend to be correlated with one another and generally organized into linearly separable categories. We chose MNB to be applied on the text-based representations and SVM for the lexicon-based representation due to its capability of dealing with double-valued attributes.

For the evaluation, we randomized the test dataset and used a 70%–30% split for the purpose of training and evaluation respectively. Table 1 presents the accuracy for all examined text-based representation models. As expected, bigrams outperformed unigrams. Several differences were also observed moving from binary to the $tf \cdot idf$ weighting scheme. Table 2 illustrates the results obtained by SVM compared to the "counting" method using the Harvard dictionary alone, revealing the superiority of the SVM approach. Our findings are also consistent with the claims supporting that adjectives carry more sentimental weight.

Figure 2 illustrates the monthly aggregated sentiment polarity index. First, we extract the daily market sentiment polarity index with respect to a specific financial instrument or indicator (the oil prices in our case) based on the sentiment analyses of news headlines obtained during the last 24 h. The aggregation of the sentiment to the month level is done by aggregating the sentiment of the daily textual resources that were tagged positive and negative in the previous step. Our empirical results show that the MNB model produces statistically significant results, while the lexicon based method does not produce any significant results. Granger causality test also indicates a statistically significant causation stemming from oil sentiment index to oil prices while the opposite is

**Table 1.** Accuracy achieved for the text-based model

| Representation | Binary | | Term freq. | | tf-idf | |
|---|---|---|---|---|---|---|
| ngrams | 1 | 2 | 1 | 2 | 1 | 2 |
| Accuracy (%) | 74.89 | **80.06** | 75.31 | **81.92** | 76.87 | **82.24** |

**Table 2.** Accuracy achieved by SVM applied on the lexicon-based model

| | Adverb | Noun | Verb | Adjective |
|---|---|---|---|---|
| Accuracy (%) | 53.4 | 54.9 | 58.8 | **66.93** |

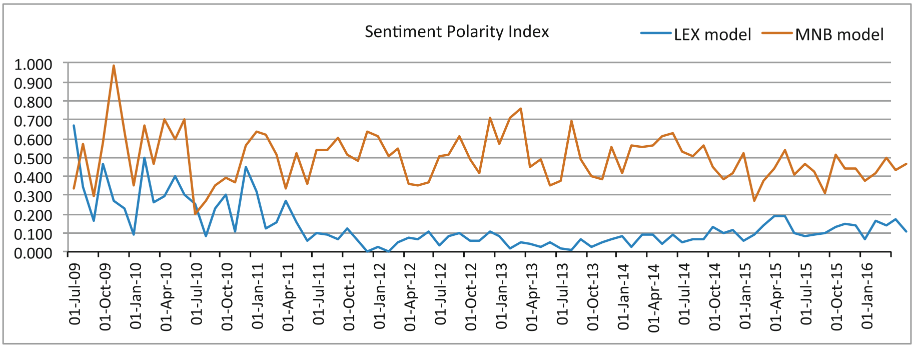


**Fig. 2.** Monthly aggregated sentiment polarity index.

rejected. Overall, we find that the sentiment is an important consideration when explaining crude oil prices using data from online textual resources.

## 4 Datasets

The propose system is taking two sources of information as input, namely, *textual data* from online resources (financial websites), and *market data*.

### 4.1 Textual Data

The study covers the period 21 July 2009 to 06 May 2016. Our sentiment analysis relies on daily content from `investing.com`; a leading financial platform offering real-time quotes, streaming charts, financial news, technical analysis and more. We searched the digital archive of `investing.com` to obtain articles containing one of the following terms: 'oil prices', 'crude oil', 'brent' or 'future prices' (including their variants). In other words, for an article to meet our criteria, it must contain at least one term pertaining to the Eurostoxx index. The final sample contains 14205 news headlines together with their associated body text, which are fetched by querying the Virtuoso SPARQL endpoint of PoolParty using a simple web crawling algorithm. The sentiment polarity index is published each day and is calculated automatically based on the sentiment analyses of news headlines obtained during the last 24 h.

### 4.2 Market Data

In order to construct the industry portfolios, we compile 4442 daily stock prices from Yahoo finance! for the period 2008 to 2016, traded in 15 European countries. The countries and their number of stocks are reported in Table 3[4].

**Table 3.** Number of stocks per country used

| No | Country | Number of stocks |
|---|---|---|
| 1 | Austria | 70 |
| 2 | Belgium | 136 |
| 3 | Denmark | 148 |
| 4 | Finland | 131 |
| 5 | France | 697 |
| 6 | Germany | 703 |
| 7 | Greece | 167 |
| 8 | Iceland | 17 |
| 9 | Italy | 293 |
| 10 | Netherlands | 98 |
| 11 | Norway | 175 |
| 12 | Portugal | 45 |
| 13 | Spain | 144 |
| 14 | Turkey | 404 |
| 15 | United Kingdom | 1210 |
| | | Total: 4442 |

We take first logarithmic differences in order to compute daily returns, and we convert daily to monthly returns using monthly averages. The selection of the monthly forecasting horizon of the Eurostoxx 50 index is based on data availability. Following the classification system of Fama and French [10] in 38 industry portfolios, we construct 38 equally weighted industry portfolios, depicted in Table 4. We end up with 20 monthly portfolios, according to the activity of each listed company and 99 monthly observations.

As stated in the Introduction section, the scope of this exercise is to forecast the future direction of the Eurostoxx 50 index. After compiling daily values

[4] The selected countries and the number of stocks are upon availability of the data.

**Table 4.** Industry portfolios

| | Industry code | Industry description (SIC codes) | Number of stocks |
|---|---|---|---|
| 1 | Agric | Agriculture, forestry, and fishing (0100–0999) | 31 |
| 2 | Mines | Mining (1000–1299) | 74 |
| 3 | Oil | Oil and Gas Extraction (1300–1399) | 102 |
| 4 | Stone | Non-metalic Minerals Except Fuels (1400–1499) | 0 |
| 5 | Cnstr | Construction (1500–1799) | 129 |
| 6 | Food | Food and Kindred Products (2000–2099) | 156 |
| 7 | Smoke | Tobacco Products (2100–2199) | 3 |
| 8 | Txtls | Textile Mill Products (2200–2299) | 0 |
| 9 | Apprl | Apparel and other Textile Products (2300–2399) | 8 |
| 10 | Wood | Wood Lumber and Wood Products (2400–2499) | 0 |
| 11 | Chair | Furniture and Fixtures (2500–2599) | 33 |
| 12 | Paper | Paper and Allied Products (2600–2661) | 27 |
| 13 | Print | Printing and Publishing (2700–2799) | 86 |
| 14 | Chems | Chemicals and Allied Products (2800–2899) | 261 |
| 15 | Ptrlm | Petroleum and Coal Products (2900–2999) | 0 |
| 16 | Rubbr | Rubber and Miscellaneous Plastics Products (3000–3099) | 5 |
| 17 | Lethr | Leather and Leather Products (3100–3199) | 0 |
| 18 | Glass | Stone, Clay and Glass Products (3200–3299) | 0 |
| 19 | Metal | Primary Metal Industries (3300–3399) | 36 |
| 20 | MtlPr | Fabricated Metal Products (3400–3499) | 0 |
| 21 | Machn | Machinery, Except Electrical (3500–3599) | 189 |
| 22 | Elctr | Electrical and Electronic Equipment (3600–3699) | 307 |
| 23 | Cars | Transportation Equipment (3700–3799) | 66 |
| 24 | Instr | Instruments and Related Products (3800–3879) | 38 |
| 25 | Manuf | Miscellaneous Manufacturing Industries (3900–3999) | 0 |
| 26 | Trans | Transportation (4000–4799) | 81 |
| 27 | Phone | Telephone and Telegraph Communication (4800–4829) | 139 |
| 28 | TV | Radio and Television Broadcasting (4830–4899) | 32 |
| 29 | Utils | Electric, Gas, and Water Supply (4900–4949) | 167 |
| 30 | Garbg | Sanitary Services (4950–4959) | 33 |
| 31 | Steam | Steam Supply (4960–4969) | 0 |
| 32 | Water | Irrigation Systems (4970–4979) | 0 |
| 33 | Whlsl | Wholesale (5000–5199) | 227 |
| 34 | Retail | Retail Stores (5200–5999) | 376 |
| 35 | Money | Finance, Insurance, and Real Estate (6000–6999) | 894 |
| 36 | Srvc | Services (7000–8999) | 885 |
| 37 | Govt | Public Administration (9000–9999) | 32 |
| 38 | Other | Everything else | 8 |

of the index from Yahoo finance!, we take monthly averages of the index. The classification of the stock market in bear and bull market is achieved directly by comparing the successive values of the index, constructing a binary index of ones and zeros, to be fed in the forecasting model. Unlike Nyberg [20], we do not use filer algorithms in order to discern bear and bull markets, since this category of filters tends to detect prolonged periods of market states (15 months at minimum) that cannot be used effectively in active index trading.

## 5 Probit Specification

### 5.1 The Model

The majority of previous empirical studies that considers different regimes in stock market exploits Markov switching models. Although the specific category of models has the advantage of providing state-dependent inferences, the main drawback is that it is based on an unobservable Markov switching process that cannot be explicitly described. In our study we consider a binary response model that predicts the future direction of the index as a binary time series. Considering the binary state variable $s_t$, the value 1 states a bull and 0 a bear market as follows:

$$s_t = \begin{cases} 1 & \textit{bull} \text{ market at time t,} \\ 2 & \textit{bear} \text{ market at time t,} \end{cases} \tag{1}$$

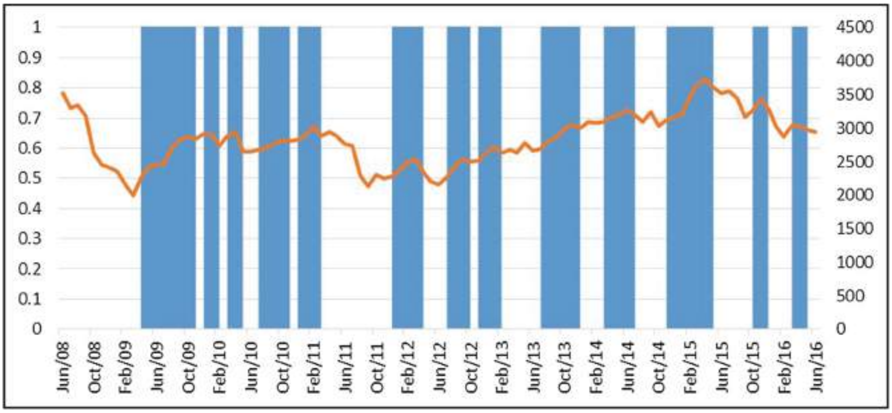


**Fig. 3.** Bear vs bullish periods of the Eurostoxx 50 index. Note: the continuous line reports actual values of the index, while the shaded areas are bullish periods.

for $t = 1, 2, 3, .., T$ the range of the monthly observations. Denoting the conditional expectation $E_{t-1}(s_t|\Omega_{t-1})$ in the information set $\Omega_{t-1}$ at time $t-1$, the conditional probability at time $t$ that the market is in a bull state is:

$$p_t = E_{t-1}(s_t|\Omega_{t-1}) = P_{t-1}(s_t = 1) = \Phi(\pi_t) \tag{2}$$

Where $\pi_t$ is a linear combination of variables and $\Phi(\bullet)$ is the normal cumulative distribution function. Naturally, the conditional probability of the bull market is the complement of the bear market probability $P_{t-1}(s_t = 0) = 1 - p_t$. In order to predict the linear function $\pi_t$ we consider different static and dynamic models. Starting from the univariate probit model [6]

$$\pi_t = \omega + \chi'_{t-h}\beta \tag{3}$$

where $\omega$ is a constant, $\beta$ is the coefficients vector and $\chi'_{t-h}$ a matrix of predictive regressors. The index h denotes the forecasting horizon. The popular static model is extended by adding lags of the dependent variable $s_t$ resulting in the autoregressive static model

$$\pi_t = \omega + \alpha(s_{t-1})\chi'_{t-h}\beta \tag{4}$$

or by adding lags of the dependent variable $\pi_t$, leading to the dynamic model proposed by Kauppi and Saikkonen [16]:

$$\pi_t = \omega + \delta\pi_{t-1} + \chi'_{t-h}\beta \tag{5}$$

By recursive substitution, Eq. 5 can be seen as an infinite order static Eq. 4 where the whole history of the values of the predictive variables included has an effect $\chi_{t-h}$ on the conditional probability. Thus, if the longer history of explanatory variables included in $\chi_{t-h}$ are useful to predict the future market status, the autoregressive Eq. 5 may offer a parsimonious way to specify the predictive model. A natural extension would be the dynamic autoregressive model:

$$\pi_t = \omega + \alpha(s_{t-1}) + \delta\pi_{t-1} + \chi'_{t-h}\beta \tag{6}$$

## 6 Preliminary Results

### 6.1 Empirical Results

We forecast the future direction of the Eurostoxx 50 for 1 to 12 months ahead based on the static models Eqs. 3 and 4 and the dynamic models Eqs. 5 and 6. After forecasting the value of $\pi_t$ we compute the probability $\Phi(\pi_t)$ from the normal cumulative distribution function. We assign a threshold in the probability of belonging to a bull or bear market. A popular selection is 0.5, but as we observe from Fig. 3, there are more bear than bullish periods. Since the mean value of $s_t$ is 0.3544 we also compute bear and bull periods according to the threshold 0.35. Each time we keep the last 18 observations from the training phase of the model in order to measure the out-of-sample forecasting accuracy.

As a first step, we forecast $\pi_t$ considering only the portfolio of one industry at a time as a regressor, in order to observe the potential contribution of each industry in forecasting the market. In Table 5, we report the in and out-of sample forecasting accuracy for the most accurate models. The selection of the

**Table 5.** Results of the estimated four models

| | Static | | | | AR static | | | | Dynamic | | | | AR dynamic | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| h | $R^2$ | In | Out | Sector | $R^2$ | In | Out | Sector | $R^2$ | In | Out | Sector | $R^2$ | In | Out | Sector |
| 1 | 0.22 | 0.73 | 0.78 | TV | 0.22 | 0.83 | 0.78 | Smoke | 0.24 | 0.72 | 0.72 | TV | 0.29 | 0.74 | 0.61 | TV |
| 2 | 0.00 | 0.65 | 0.61 | Smoke | 0.13 | 0.70 | 0.61 | Apprl | 0.01 | 0.64 | 0.61 | Oil | 0.19 | 0.70 | 0.56 | Apprl |
| 3 | 0.00 | 0.66 | 0.67 | Phone | 0.08 | 0.72 | 0.67 | Phone | 0.00 | 0.66 | 0.67 | Phone | 0.14 | 0.71 | 0.67 | Cnstr |
| 4 | –0.01 | 0.68 | 0.67 | Garbg | 0.09 | 0.72 | 0.72 | Smoke | 0.02 | 0.67 | 0.50 | Paper | 0.17 | 0.71 | 0.56 | Garbg |
| 5 | –0.02 | 0.62 | 0.67 | Rubbr | 0.08 | 0.70 | 0.67 | Garbg | 0.05 | 0.66 | 0.44 | Utils | 0.16 | 0.74 | 0.67 | Rubbr |
| 6 | –0.01 | 0.64 | 0.72 | Garbg | 0.08 | 0.71 | 0.67 | Smoke | 0.04 | 0.64 | 0.56 | Agric | 0.14 | 0.71 | 0.61 | Smoke |
| 7 | 0.02 | 0.68 | 0.78 | TV | 0.10 | 0.71 | 0.78 | TV | 0.05 | 0.63 | 0.50 | Agric | 0.15 | 0.72 | 0.61 | Apprl |
| 8 | 0.01 | 0.65 | 0.61 | Agric | 0.07 | 0.72 | 0.72 | Agric | 0.04 | 0.61 | 0.50 | Agric | 0.13 | 0.69 | 0.56 | Garbg |
| 9 | 0.01 | 0.63 | 0.39 | Oil | 0.07 | 0.70 | 0.72 | Rubbr | 0.03 | 0.59 | 0.44 | Agric | 0.15 | 0.67 | 0.67 | Smoke |
| 10 | 0.04 | 0.59 | 0.56 | Agric | 0.09 | 0.68 | 0.72 | Agric | 0.04 | 0.61 | 0.50 | Agric | 0.15 | 0.70 | 0.67 | Govt |
| 11 | 0.02 | 0.63 | 0.56 | Agric | 0.08 | 0.68 | 0.67 | Smoke | 0.02 | 0.63 | 0.56 | Agric | 0.13 | 0.71 | 0.67 | Agric |
| 12 | 0.00 | 0.61 | 0.67 | Instr | 0.07 | 0.72 | 0.72 | Instr | 0.00 | 0.60 | 0.67 | Instr | 0.13 | 0.72 | 0.67 | TV |
| | Static | | | | AR static | | | | Dynamic | | | | AR dynamic | | | |
| h | R2 | In | Out | Sector | R2 | In | Out | Sector | R2 | In | Out | Sector | R2 | In | Out | Sector |
| 1 | 0.22 | 0.69 | 0.72 | TV | 0.22 | 0.73 | 0.78 | Smoke | 0.24 | 0.71 | 0.72 | TV | 0.29 | 0.76 | 0.72 | TV |
| 2 | 0.00 | 0.52 | 0.61 | Smoke | 0.13 | 0.68 | 0.61 | Apprl | 0.01 | 0.52 | 0.50 | Oil | 0.19 | 0.71 | 0.67 | Apprl |
| 3 | 0.00 | 0.50 | 0.39 | Phone | 0.08 | 0.72 | 0.72 | Phone | 0.00 | 0.46 | 0.39 | Phone | 0.14 | 0.72 | 0.67 | Cnstr |
| 4 | –0.01 | 0.47 | 0.28 | Garbg | 0.09 | 0.71 | 0.72 | Smoke | 0.02 | 0.57 | 0.28 | Paper | 0.17 | 0.72 | 0.61 | Garbg |
| 5 | –0.02 | 0.51 | 0.39 | Rubbr | 0.08 | 0.72 | 0.72 | Garbg | 0.05 | 0.51 | 0.28 | Utils | 0.16 | 0.73 | 0.67 | Rubbr |
| 6 | –0.01 | 0.60 | 0.44 | Garbg | 0.08 | 0.67 | 0.67 | Smoke | 0.04 | 0.52 | 0.28 | Agric | 0.14 | 0.67 | 0.67 | Smoke |
| 7 | 0.02 | 0.50 | 0.44 | TV | 0.10 | 0.74 | 0.72 | TV | 0.05 | 0.54 | 0.28 | Agric | 0.15 | 0.71 | 0.67 | Apprl |
| 8 | 0.01 | 0.54 | 0.39 | Agric | 0.07 | 0.70 | 0.67 | Agric | 0.04 | 0.62 | 0.28 | Agric | 0.13 | 0.66 | 0.56 | Garbg |
| 9 | 0.01 | 0.53 | 0.33 | Oil | 0.07 | 0.64 | 0.72 | Rubbr | 0.03 | 0.59 | 0.28 | Agric | 0.15 | 0.73 | 0.67 | Smoke |
| 10 | 0.04 | 0.61 | 0.44 | Agric | 0.09 | 0.64 | 0.61 | Agric | 0.04 | 0.61 | 0.39 | Agric | 0.15 | 0.64 | 0.61 | Govt |
| 11 | 0.02 | 0.57 | 0.44 | Agric | 0.08 | 0.68 | 0.67 | Smoke | 0.02 | 0.54 | 0.33 | Agric | 0.13 | 0.71 | 0.67 | Agric |
| 12 | 0.00 | 0.51 | 0.44 | Instr | 0.07 | 0.70 | 0.72 | Instr | 0.00 | 0.54 | 0.33 | Instr | 0.13 | 0.67 | 0.67 | TV |

**Table 6.** Best fitted results for each forecasting period

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| In-sample accuracy | | | | | | | | | | | | |
| RW | 0.73 | 0.53 | 0.49 | 0.49 | 0.50 | 0.49 | 0.53 | 0.54 | 0.49 | 0.49 | 0.48 | 0.55 |
| Model | 0.83 | 0.71 | 0.72 | 0.72 | 0.74 | 0.67 | 0.71 | 0.72 | 0.70 | 0.68 | 0.71 | 0.72 |
| Out-of-sample accuracy | | | | | | | | | | | | |
| RW | 0.71 | 0.67 | 0.60 | 0.50 | 0.62 | 0.42 | 0.60 | 0.60 | 0.56 | 0.50 | 0.29 | 0.50 |
| Model | 0.78 | 0.67 | 0.72 | 0.72 | 0.67 | 0.67 | 0.78 | 0.72 | 0.72 | 0.72 | 0.67 | 0.72 |
| Sector | Smoke | Apprl | Phone | Smoke | Rubbr | Smoke | TV | Agric | Rubbr | Agric | Agric | Instr |

most accurate model among the available portfolios is based on the maximum McFadden R square of the training models.

In Table 6, we depict only the best results for each forecasting horizon, as well as the forecasting accuracy of the benchmark RW model. As we observe from Table 6, the considered forecasting models exhibit similar or outperform by large the RW in all forecasting horizons when it comes to out-of-sample accuracy. Another interesting conclusion is that the agricultural industry outperforms all other sectors in the longer forecasting horizons. Thus, we reject even the weak

**Table 7.** Results from all 20 available industry portfolios

| | Static | | AR static | | Dynamic | | AR dynamic | |
|---|---|---|---|---|---|---|---|---|
| In sample | | | | | | | | |
| h | In | Out | In | Out | In | Out | In | Out |
| 1 | 1.00 | 0.78 | 1.00 | 0.67 | 1.00 | 0.67 | 1.00 | 0.67 |
| 2 | 0.79 | 0.44 | 0.82 | 0.61 | 0.91 | 0.50 | 0.88 | 0.50 |
| 3 | 0.79 | 0.61 | 0.82 | 0.56 | 0.89 | 0.67 | 0.92 | 0.61 |
| 4 | 0.91 | 0.33 | 0.93 | 0.50 | 0.93 | 0.44 | 1.00 | 0.50 |
| 5 | 0.77 | 0.44 | 0.96 | 0.44 | 1.00 | 0.44 | 0.97 | 0.39 |
| 6 | 0.89 | 0.44 | 1.00 | 0.50 | 0.99 | 0.44 | 1.00 | 0.44 |
| 7 | 0.79 | 0.56 | 0.82 | 0.61 | 0.90 | 0.56 | 0.92 | 0.56 |
| 8 | 0.83 | 0.44 | 0.83 | 0.50 | 0.94 | 0.50 | 0.99 | 0.39 |
| 9 | 0.89 | 0.50 | 0.96 | 0.56 | 0.90 | 0.50 | 1.00 | 0.56 |
| 10 | 0.81 | 0.33 | 0.84 | 0.39 | 0.94 | 0.44 | 0.91 | 0.39 |
| 11 | 0.88 | 0.56 | 0.88 | 0.56 | 0.88 | 0.44 | 0.88 | 0.61 |
| 12 | 0.97 | 0.50 | 0.99 | 0.61 | 0.94 | 0.39 | 0.94 | 0.50 |
| Out of sample | | | | | | | | |
| h | In | Out | In | Out | In | Out | In | Out |
| 1 | 1.00 | 0.78 | 1.00 | 0.67 | 1.00 | 0.67 | 1.00 | 0.67 |
| 2 | 0.74 | 0.39 | 0.81 | 0.61 | 0.91 | 0.39 | 0.88 | 0.50 |
| 3 | 0.76 | 0.56 | 0.79 | 0.56 | 0.87 | 0.67 | 0.91 | 0.61 |
| 4 | 0.92 | 0.39 | 0.93 | 0.50 | 0.92 | 0.44 | 1.00 | 0.50 |
| 5 | 0.76 | 0.44 | 0.92 | 0.44 | 1.00 | 0.44 | 0.96 | 0.39 |
| 6 | 0.82 | 0.44 | 1.00 | 0.50 | 0.97 | 0.44 | 1.00 | 0.44 |
| 7 | 0.75 | 0.61 | 0.82 | 0.61 | 0.89 | 0.56 | 0.93 | 0.56 |
| 8 | 0.79 | 0.44 | 0.82 | 0.50 | 0.92 | 0.50 | 0.96 | 0.39 |
| 9 | 0.91 | 0.50 | 0.96 | 0.56 | 0.91 | 0.50 | 1.00 | 0.56 |
| 10 | 0.87 | 0.44 | 0.84 | 0.50 | 0.96 | 0.39 | 0.91 | 0.33 |
| 11 | 0.84 | 0.50 | 0.85 | 0.56 | 0.82 | 0.39 | 0.88 | 0.56 |
| 12 | 0.91 | 0.50 | 0.99 | 0.61 | 0.94 | 0.39 | 0.90 | 0.50 |

form of market efficiency. In Table 7, we report the results from training models using all the available 20 industry portfolios as regressors.

In Table 8, we concentrate on the forecasting accuracy of the most accurate model per forecasting horizon, vis a vis with the RW model. As it can be seen from this Table, the forecasting ability of the forecasting models that incorporate the entire dataset is smaller than the use of specific industry portfolios. More specifically, when we focus on the out-of-sample accuracy, the accuracy of the models is smaller than the random selection (random guess) of the future

**Table 8.** Forecasting accuracy of the most accurate model per forecasting horizon.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| In-sample accuracy | | | | | | | | | | | | |
| RW | 0.73 | 0.53 | 0.49 | 0.49 | 0.50 | 0.49 | 0.53 | 0.54 | 0.49 | 0.49 | 0.48 | 0.55 |
| Model | 1.00 | 0.82 | 0.89 | 0.93 | 0.94 | 0.96 | 0.82 | 0.94 | 0.96 | 0.94 | 0.88 | 0.99 |
| Out-of-sample accuracy | | | | | | | | | | | | |
| RW | 0.71 | 0.67 | 0.60 | 0.50 | 0.62 | 0.42 | 0.60 | 0.60 | 0.56 | 0.50 | 0.29 | 0.50 |
| Model | 0.78 | 0.61 | 0.67 | 0.5 | 0.50 | 0.56 | 0.61 | 0.50 | 0.56 | 0.44 | 0.61 | 0.61 |

**Table 9.** Forecasting power of oil sentiment index

| Oil sentiment index | | | |
|---|---|---|---|
| 2010M10–2016M03 | | | |
| Horizon | In | Out | Model |
| 1 | 0.68 | 0.55 | AR static |
| 2 | 0.61 | 0.60 | AR dynamic |
| 3 | 0.60 | 0.55 | AR static |
| 4 | 0.60 | 0.55 | AR static |
| 5 | 0.66 | 0.60 | AR dynamic |
| 6 | 0.60 | 0.55 | AR static |
| 7 | 0.67 | 0.60 | AR dynamic |
| 8 | 0.68 | 0.60 | AR dynamic |
| 9 | 0.70 | 0.60 | AR dynamic |
| 10 | 0.69 | 0.60 | AR dynamic |
| 11 | 0.71 | 0.60 | AR dynamic |
| 12 | 0.65 | 0.60 | AR dynamic |

direction of the market. This fact could be attributed to the introduction of information to the model that is irrelevant to forecast (noise).

### 6.2 Oil's Sentiment Predictive Power

In this section we present the results of the oil sentiment index as a predictor of the Eurostoxx 50. As previously we estimate four different Probit models and in Table 9 we present the results. We present here results of the oil sentiment index produced by using Multinomial Naive Bayes and SVM system. The other two methods do not produce any statistically significant results,

Results above show that the oil sentiment index has predictive power on the Eurostoxx 50. It is interesting that almost for any estimated forecast period the AR Dynamic model produces the best results. Using the oil sentiment index we can forecast in 60% accuracy the future sign reversals of Eurostoxx 50.

## 7 Conclusions

In this paper, we forecast the future direction of the Eurostoxx 50 index using monthly data for various horizons. In doing so, we develop static and dynamic probit models, while we construct industry portfolios from a plethora of stocks traded in different stock markets. Our empirical findings suggest that the dynamic models do not improve the forecasting ability of the models upon the use of static version of the model, while certain sectors of the stock market outperform the RW model, leading to profitable trading strategies. In future work, we intend to expand our sample to 2004 for both stock prices and oil sentiment. It is also crucial to estimate the predictive power of specific macroeconomic variables as well as combinations of industry portfolios and macroeconomic variables. Another interesting aspect will be to estimate whether industry portfolios from the core countries of the Eurozone have different predictive power than industry portfolios from the periphery countries. With regard to sentiment analysis, we intend to apply several accuracy-boosting machine learning techniques to the sentiment classification problem.

**Acknowledgements.** This work has been supported by the EU HORIZON 2020 project PROFIT (Contract no: 687895).

## References

1. Alistair, K., Diana, I.: Sentiment classification of movie and product reviews using contextual valence shifters. In: Proceedings of FINEXIN (2005)
2. Aue, A., Gamon, M.: Customizing sentiment classifiers to new domains: a case study. In: Proceedings of Recent Advances in Natural Language Processing (RANLP), pp. 1–2 (2005)
3. Bartlett, J., Albright, R.: Coming to a theater near you! Sentiment classification techniques using SAS text miner. In: SAS Global Forum 2008 (2008)
4. Boiy, E., Hens, P., Deschacht, K., Moens, M.F.: Automatic sentiment analysis in on-line text. In: ELPUB, pp. 349–360 (2007)
5. Cadilhac, A., Benamara, F., Aussenac-Gilles, N.: Ontolexical resources for feature based opinion mining: a case-study. In: 23rd International Conference on Computational Linguistics, p. 77. Citeseer (2010)
6. Chen, S.S.: Predicting the bear stock market: macroeconomic variables as leading indicators. J. Bank. Finance **33**(2), 211–223 (2009)
7. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: Gate: a framework and graphical development environment for robust NLP tools and applications. In: Proceedings of 40th Anniversary Meeting of the Association for Computational Linguistics (ACL) (2002)
8. Driesprong, G., Jacobsen, B., Maat, B.: Striking oil: another puzzle? J. Financ. Econ. **89**(2), 307–327 (2008)
9. Fama, E.: The behavior of stock-market prices. J. Bus. **38**(1), 34–105 (1965)
10. Fama, E., French, K.: Industry costs of equity. J. Financ. Econ. **43**(2), 153–193 (1997)

11. Ghiassi, M., Skinner, J., Zimbra, D.: Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert Syst. Appl. **40**(16), 6266–6282 (2013)
12. Gonçalves, P., Araújo, M., Benevenuto, F., Cha, M.: Comparing and combining sentiment analysis methods. In: Proceedings of the first ACM Conference on Online Social Networks, pp. 27–38. ACM (2013)
13. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics, pp. 174–181. Association for Computational Linguistics (1997)
14. Hong, H., Torous, W., Valkanov, R.: Do industries lead stock markets? J. Financ. Econ. **83**(2), 367–396 (2007)
15. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM (2004)
16. Kauppi, H., Saikkonen, P.: Predicting us recessions with dynamic binary response models. Rev. Econ. Stat. **90**(4), 777–791 (2008)
17. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
18. Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., Ngo, D.C.L.: Text mining for market prediction: a systematic review. Expert Syst. Appl. **41**(16), 7653–7670 (2014)
19. Nyberg, H.: Forecasting the direction of the us stock market with dynamic binary probit models. Int. J. Forecast. **27**(2), 561–578 (2011)
20. Nyberg, H.: Predicting bear and bull stock markets with dynamic binary time series models. J. Bank. Finance **37**(9), 3351–3363 (2013)
21. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
22. Prabowo, R., Thelwall, M.: Sentiment analysis: a combined approach. J. Inf. **3**(2), 143–157 (2009)
23. Salvetti, F., Reichenbach, C., Lewis, S.: Opinion polarity identification of movie reviews. In: Shanahan, J.G., Qu, Y., Wiebe, J. (eds.) Computing Attitude and Affect in Text: Theory and Applications, pp. 303–316. Springer, Netherlands (2006)
24. Schmeling, M.: Investor sentiment and stock returns: some international evidence. J. Empirical Finance **16**(3), 394–408 (2009)
25. Stone, P., Dunphy, D.C., Smith, M.S., Ogilvie, D.: The general inquirer: a computer approach to content analysis. J. Reg. Sci. **8**(1), 113–116 (1968)
26. Taboada, M., Anthony, C., Voll, K.: Methods for creating semantic orientation dictionaries. In: Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), pp. 427–432 (2006)
27. Tetlock, P.C.: Giving content to investor sentiment: the role of media in the stock market. J. Finance **62**(3), 1139–1168 (2007)
28. Tong Richard, M.: An operational system for detecting and tracking opinions in online discussions. In: Working Notes of the ACM SIGIR Workshop on Operational Text Classification, New York, pp. 1–6 (2001)
29. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)

30. Turney, P.D., Littman, M.L.: Measuring praise and criticism: inference of semantic orientation from association. ACM Trans. Inf. Syst. (TOIS) **21**(4), 315–346 (2003)
31. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354. Association for Computational Linguistics (2005)
32. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. **181**(6), 1138–1152 (2011)
33. Zhou, L., Chaovalit, P.: Ontology-supported polarity mining. J. Am. Soc. Inf. Sci. Technol. **59**(1), 98–110 (2008)