Learning Algorithms, and current developments – Banking at the Crossroads

Essays on Machine learning / Applied Data Analytics, Asset Encumbrance / Bail-in, Sustainability and Resource Availability

Dissertation zur Erlangung des akademischen Grades
Doctor rerum politicarum
an der Fakultät Wirtschaftswissenschaften Technische Universität Dortmund

Dipl. Wi-Ing. Joachim Erhardt
London, December 2017

Acknowledgments

This PhD thesis was supported by *Landesbank Baden-Württemberg*. I am grateful to Professor Dr. Peter N. Posch for his ongoing encouragement and valued advice. I would like to express thanks to my team, particularly to Dr. Rainer Dürr, Tarek Hard (PhD) and David West for their support and modelling work over the years. I am thankful for having had the opportunity to supervise the master theses of Mark Wolters, Johannes Luebbers, Bernd Schumacher and Patrick Kosa. Last but not least, I would like to thank my wife and especially our daughter for their mostly unwavering patience.

Contents

List of Tables

List of Figures

1. Introduction

Banking at the crossroads[1] investigates themes that are currently pertinent for the success of banks. There are four major topics, all with substantial effects on long-term profitability. The thesis is structured accordingly:

1. Machine learning => Artificial Intelligence (AI) and possible efficiency gains;
2. Asset Encumbrance => Banking regulation, and the consequences on a micro level;
3. Ethics and Sustainability => 'Economic return' as well as 'social return'; and
4. Impact on Resource Availability => Macroeconomic properties of finance.

Chapter 2 focuses on new data technology. We[2] describe how an experimental set-up utilising machine learning algorithms can affect real change. We introduce and debate the different elements of choice when training a model. The process starts with a proper definition of the respective problem, carries on with the preparation of data and continues with how to transform the observations into features – the model input. The 'learning process' involves picking 'the best' model, selecting reasonable model parameters and establishing a suitable training and testing routine. The challenge is to fit the model adequately. 'Under-fitting' leads to underperformance and 'overfitting' to a model that is overly sensitive to data noise. Model bias and variance needs to be balanced. We ponder specific model items that purportedly minimises the problematic, i.e. randomisation, shrinkage and complexity in objective functions *etc.* Increased speed in computing and powerful data collection and storage capabilities allow for 'application in real'. We demonstrate the 'power of machine learning' and its capabilities. We apply learning models to a diverse set of 'cases'. We start by predicting market trends. In addition we show how we could optimise investment portfolios by adopting these results, thus creating an 'enhanced Markowitz' model approach. Lastly, we show how data analytics could help to achieve better customer coverage, *inter alia* providing justification for 'appropriate inventory levels' under the German Steagall act. This is done, as a novel 'first', based on data from institutional customers.[3] We explain in the context of customers how to affect change in a methodical manner. The new tool set offers the possibility to test modifications of how business is done in a control group regulated

---

[1] Erasmus quoted in the 1500s from the Greek in Thornes's *Elegies* (ca. 600 B.C.): "I stand at the crossroads."

[2] Chapters 3 and 5 are at the time of publication already 'published in Journals' as a collaborative effort (see also collaboration in Chapter 2). For consistency the remaining thesis is written in plural.

[3] Not as usual data from retail customers.

environment. The model predictions help to identify the most promising alterations. For further progress new kind of data (e.g. touch data) has to be collected. Striving for new data will revolutionise why and how we do business.

In chapter 3 we demonstrate the impact of regulation on finance by looking at the new bail-in rules as a typical example. In the aftermath of the financial crisis, legislators significantly increased the regulatory burden placed on financial institutions. The main purpose was to increase the resilience of the financial system in addition to preventing future tax payer bailouts. Reliance on short-term funding of long-dated assets, in combination with intertwining financing relationships between banks was exposed as a weakness of the system. Hence a strict framework targeting excessive liquidity transformation and reduction of interdependence was established. Ratios like LCR (Liquidity Coverage Ratio) and NSFR (Net Stable Funding Ratio) are monitored and regulated. Moreover, funding is treated differently depending on its source. For example, funding from retail or corporate investors is 'privileged' over funding from financial institutions. As a result banks are now focused on diversifying their funding sources. Besides funding from central banks through various liquidity programs, secured or collateralised funding (e.g. covered bonds and repos/total return swaps) has gained prominence. Thus the effect on asset encumbrance within the legal bail-in framework needs to be monitored. We analyse overall funding costs of European banks and estimate 'optimal levels of asset encumbrance' from a banks perspective. We present how and which banks can optimise their funding strategies. Based on our model, we estimate by how much the banks in our sample can increase secured financing to decrease their overall costs of funding. We accept that there is no universal optimum. Targets vary with perspective. For example, the perfect ratio from a senior unsecured investor's angle is different from that of a subordinate investor or from that of a bank's treasurer.

In chapter 4 we debate the concept of "Sustainability" and "Ethics" as a stricter framework than that of legality. Environmental exploitation, disarmament, peace, and social issues like 'socially responsible employment', are assessed under Sustainability guidelines including aspects of corporate "Governance". Principles of Responsible Investments, as conveyed by the UN (PRI, 2010), are seen as ethical minimum standards globally. We investigate the fund industry's claims that following the Sustainability framework in asset management creates 'alpha' – long-term superior results. The fund industry contends that high environmental and social standards act as a filter against bad investment decisions and, as such, create a positive selection filter. Accordingly, well-managed companies follow a long-term oriented strategy rather than unsustainable short-term profits. They avoid excessive risks and are rarely hit by punitive penalties. It appears that a positive economic return for the investor – so-called 'alpha' – can be accomplished. We report on the theoretical foundation and the empirical studies of this asset class. In addition to a literature review, we interview Sustainability experts within the industry. As a second step we research the claim that one can increase 'social return' with 'Environmental, Social and Governance' (ESG), achieving a discernible benefit for society overall. We conclude the chapter by exploring "Green Bonds". The Green Bond market has grown by 92% in a year, up to 81bn USD in 2016. This is despite a minor spread advantage compared to 'non-green' alternatives with similar economic risk and returns. Finally, recommendations are made for future development, requiring new legal and regulatory guidelines.

In chapter 5 we evaluate the impact on resource availability through provision of finance – an example of how banking activity supports the 'greater good' of the economy. Most industrial nations are reliant on a secure supply of raw materials, but typically do not possess sufficient primary resources themselves. This is a situation widely accepted by their respective governments which have instigated a variety of programs to secure availability of raw materials. We explore the influence of financing conditions on the availability of base metals. Using fixed effects regression on international trade and banking data, we find a consistent negative relationship between the financing costs and imports of base metals after allowing for prices and country risk. These results indicate that resource availability with respect to base metals is increased with a reduction in financing costs for market participants. The degree of increase differs across the base metals, where copper sees the highest increase of 3.3 tons against a decrease of short-term financing costs of one basis point. Furthermore, the effect varies across countries with the EU member states being highly dependent on imports in these materials. We consider – at a firm's level – the funding requirement during the import process and the relative sensitivity of market participants to financing costs.

Finally, we summarise our findings, illustrating each topic's significance and connotation from a bank's perspective. There has rarely been a more decisive period for the financial sector. Consolidation and innovation will gather strength. Success and survival depends on who is able to adapt fast enough.

2. Efficiency Impact through Data Analytics

2.1. Introduction

The way business is done changes with the arrival of "Predictive Analytics".[4] Machine learning already disrupts many traditional business models. Finance and banking are no exception. In the following section we explore the essential basics, theories and aspects of the practical application of machine learning. We establish an experimental framework for training and testing machine learning techniques on vast amounts of data. The set-up can be used for a broad range of topics and remains largely constant. We concentrate on 'classification' algorithms using Random Forest and Gradient Boosting (see *inter alia* Friedman *et al.*, 2009). Features take central stage. By working on different cases we demonstrate the versatility of the learning set-up. The market applications are trend prediction and prediction within a portfolio context. As an unrelated topic we adapt the approach to analyse institutional customer behaviour in a financial markets environment. We work with public data from Bloomberg as well as anonymised proprietary data in the customer case. Advanced features are engineered by adapting invariant shape analysis on a single market time series. Following the 'labelled landmarks routine' we design socioeconomic or behavioural features based on a cyclical market idea. Elliott Waves theory (EW) assumes interference between dynamic and correcting tendencies. Market trends are stimulated through psychological effects on a macro level. Last but not least, we look at "Spill-over" effects between markets. Here we work with multiple time series and apply a concept that stipulates 'Causality'. There are markets that 'lead' and those that 'follow'. We show results based on statistics as well as investment performance. The results show that there is significant 'data structure' within and across markets. Thereafter we utilise the results to optimise a market portfolio, 'enhancing' the traditional Markowitz approach. With strong trends towards 'passive investing' combined with 'robotic-advisory' this is currently the topic with the highest potential impact. The last application is modelling customer behaviour in fixed income. 'When?' and 'what?' are questions of interest. It is important for any trading business to prioritise customer coverage. Given limited capacity, we want to find out which customers should be covered; i.e. who are the customers most likely to trade this week. This is an interesting question especially for customers that do not trade that frequently. The goal

---

[4] Predictive Analytics is an independent discipline within Data Analytics.

is to become more efficient in targeting customer needs and to become more efficient overall. We want to be selective in what we do and how we do it.

2.2. Machine Learning – Elements of Choice

Machine learning is a term first introduced by Arthur Lee Samuel in 1959. Machine learning enables computers “to learn without being explicitly programmed" (Rose, 2017). Computing power and data storage were for a long time limiting factors. As the marginal costs of computing and storage are approaching zero, cases for machine learning have risen exponentially (see Cetinsoy *et al.*, 2016). Learning algorithms are used *inter alia* in driver-less cars, for improving traffic flow, in border and terror security (facial or body movement recognition), and in agriculture and health (see Chen *et al.*, 2016). Their use is highly disruptive in social media and consumer profiling. In finance, learning algorithms are used in high-frequency trading (HFT). HFT exploits ‘in size’ short-term market ‘arbitrage’, executing high probability trades (Arifovic *et al.*, 2016). In some parts of the literature HF is ‘associated’ with an increase in market volatility (Kirilenko *et al.*, 2016), not necessarily on average but amplifying extreme events. Learning Algorithms are also examined in the context of Market Making (Dixon, 2016), risk-management (Ranjan Das, 2016; Lessmann *et al.*, 2015), fraud detection, consumer credit (Sculley *et al.*, 2014) and portfolio management (Li and Hoi, 2012; Li *et al.*, 2015; Kom Samo and Vervuurt, 2016).

2.2.1. Unsupervised *versus* Supervised learning

First, we want to differentiate between “Supervised” and “Unsupervised” learning. Unsupervised learning starts with the data, looking for structural information, and for any relationship within the data. No problem gets formulated, no response is predefined or ‘labelled’. Unsupervised learning draws conclusions, but it “obtains neither supervised target outputs, nor rewards from its environment” (Ghahramani, 2004, p.3). A typical approach is to ‘cluster’ data by finding patterns. Data is organised into groups, where items in one group are similar to data in the same group and dissimilar to data in others. A popular method in Unsupervised learning is to cluster data by way of k-means. Data samples are partitioned around a centre (the ‘centroids’) in an iterative process, i.e. by optimising the sum of squared error (Stanford Course, 2014). Other techniques are dimensionality reduction, recommender systems and deep learning (Guo *et al.*, 2016). The major disadvantage of Unsupervised learning is that there is no problem specification, thus a lack of direction. The advantage is that the algorithm may find patterns not previously considered.

Supervised learning, on the other hand, starts with a ‘predefined’ problem. The first step is to describe what needs to be solved. Next, we need to define a set of response variables: Labelled responses support the answer to the problem. Supervised learning is typically grouped into regression and classification[5] problems. We assess in this setting ‘how well does a response support the solution?’ and ‘how strong is the relationship between the data and the response?’. Supervised learning is foremost an iterative process. Classifications are continuously adjusted, often in incremental steps. Supervised learning algorithms achieve a high level of accuracy by combining weak classifiers (‘base learners’) into strong classifiers with low generalisation and prediction error rates (see Schapire *et al.*, 1998). An overview of the different machine learning algorithms is shown in Figure 1.

[5] The response is a ‘class’ or ‘classification’.

Figure 1: Mind-map for various types of machine learning algorithms[6]

2.2.2. Data Preparation

"Data Preparation" stands at the beginning of any data analysis. Inaccurate data or misleading labelling of data can lead to massive deterioration of the algorithm's ability to extract useful information. Private data is often unaudited and it can take a significant amount of time to make the data ready for analysis. With respect to customer data, for example, counterparty names evolve, merge and segment over time. In addition, a single counterparty can contemporaneously take multiple name variations. As large data sets disallow manual mapping, automated similarity techniques must eventually be deployed. Other issues relate to missing data. With regard to public data (e.g. Bloomberg data), one still needs to deal with

[6] See Brownlee (2016).

time-differences, holidays and weekends. Cut-off times need to be regulated with care. In short, Data Preparation is the process of getting raw data into useful input for the algorithms. According to Brownlee (2016) it can be broken down into several steps.

2.2.2.1. Selection and Enriching Data

The learning algorithm operates on a given set of observations or instances. If necessary, we have to look for additional data sources – external and internal data. More and more data becomes available. However, the goal is to produce 'quality data input', not 'quantity data input'. It takes time to find the 'right kind of data' and to catalogue it properly. Sometimes missing data needs to be imputed or simulated. Redundant information 'should be' eliminated. In general it is advisable to take notes on why data was included or excluded. Selecting and enriching data is a specialist area in itself; this is why we need more data scientists. In our view data science will become its own profession.

2.2.2.2. Pre-processing

"Pre-processing" data is the 'shaping' of data into its 'workable form'. Pre-processing includes formatting, cleaning, editing, sampling and scraping data. Corrupt or inaccurate data is removed, replaced or modified. In the case that the data set becomes too big,[7] exploring and prototyping can be done on a smaller sub sample.

2.2.2.3. Transformation

"Transformation" includes scaling, attribute decomposition and aggregation. Data often contains different quantities or scales. It is generally advisable to set the data to the same scale, e.g. between [0, 1]. If features are becoming too complex, it may be better to split them into their constituent parts. On the other hand sometimes we get superior results after aggregating a large feature set into a smaller feature set, 'unifying' individual features.

2.2.3. Features as Model Input

Learning algorithms frame data. The aim is to match a learned function as closely as possible to a hypothetical function. F(Input) would describe perfectly the reality between the Input and the Output: *Output = F(Input)*

[7] This is in particular relevant if i.e. running times are becoming 'unreasonable' or in the case of computational and memory overload.

As we do not know the perfect function, we start with comparatively ‘weak learners’ and improve on them. Supervised learning is about finding powerful ‘mechanics’ between the data and the targeted output. Mechanics are different ideas or concepts of how data potentially relates to a response. Their representation is called a feature or attribute. Depending on the intended model there is no need for them to be independent. This is different to linear regression. Creating ‘the right kind of feature’ is a central step. "Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models..." (Brownlee, 2016), thus transforming raw data into a sophisticated model input. Typically, we represent through features relationships we believe in. But we also have the liberty to work with contradictory concepts for testing purpose. The information gain through features should be continuously assessed. Features that do not create benefit should be taken out. However, we should be mindful that we always start with weak learners. Much iteration is needed to improve them over time.

2.2.3.1. Feature Creation and Engineering

In ‘Neglected machine learning ideas’, Locklin (2017) states: "Feature engineering is another topic which doesn't seem to merit any review papers or books [...], but it is absolutely vital […]. Much of the success of machine learning is actually success in engineering features that a learner can understand." Enhanced features allow for simpler models, creating superior results. Feature creation is a ‘hands-on’ process. It requires expertise in the area *in which one tries to find a solution*.[8] Automatic techniques such as k-means clustering or deep learning algorithms may become an interesting possibility; however, in our opinion they need further development (see more on this point in section 2.3.4.2.).

We follow best practice in our feature engineering procedure:

1. Brainstorming – to find possible ideas for different mechanics, coming from market practise or economic literature;
2. Clustering the ideas and transforming them into mechanics; and
3. Feature creation, implementing and testing several versions.

[8] ‘To find a solution’ is meant in the sense: ‘improving to what is there’.

2.2.3.2. Feature Selection / Extraction

Feature extraction is the selection process for the final input parameters. After the ‘creation of multiple feature sets’ we need to filter for the relevant ones. ‘Feature selection / extraction’ removes irrelevant and redundant features. As such it is different to dimensionality reduction methods like Principal Component Analysis (PCA[9]) or Singular Value Decomposition (SVD). PCA and SVD techniques do not reduce the feature set *per se*, but rather create new, simpler parameters or combinations thereof. However, all of these methods are aiming at complexity reduction and model simplification.

According to Guyon and Elisseeff (2003) "the objectives of Variable Selection are three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data." In general, we differentiate between filter and wrapper on the one hand and embedded methods on the other. ‘Filtering’ is usually score based. We rank each feature with regard to the dependant variable. Possible methods are a Chi squared test or correlation coefficient scores. ‘Wrapping’ extends the idea of ‘Filtering’. Wrapping is about evaluating and comparing different feature combinations. More relevant for our implementations are embedded methods. As, for example, in tree-based algorithms the trees are grown step- or rather ‘stage-wise’. The stage-wise procedure helps to identify features with the highest contribution to the accuracy rate – while growing the trees. We use regularisation methods, which penalise models of higher complexity by introducing additional optimisation constraints. Models with fewer coefficients are prioritised as long as their performance is comparable.

2.2.4. Causality – determining Cause and Event

We start by explaining the difference between correlation and causation. *Correlation* describes the relationship between two or more variables.[10] *Causation* indicates that one event is the result of another event. Causation allows the differentiation between what leads and what follows. In machine learning establishing “Causality” is not required, but it may be

[9] See *inter alia* Jolliffe (2002). *Excursus*: PCA was invented by Karl Pearson in 1901. It “uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components…” , see: http://www.jstor.org/topic/principal-components-analysis/?refreqid=excelsior%3Ac40fba7d17b2d27126b654afe6171a01

[10] For correlation based feature selection see Hall (1999).

prolific. Machine learning can operate in principal with (stable) affiliations. Causality is an add-on. Causal features are, by definition, relevant features. This is true even for multi-causal or multi-directional relationships. The number of input variables can be reduced. Unnecessary noise is cancelled out. The model becomes less complex. It is easier to understand and better to interpret. Further optimisations can be more specific; there are fewer 'degrees of freedom'. Faster computation and less memory 'consumption' are welcome side effects. The 'iterative model evaluation process' becomes more efficient.

How can we determine whether a connection is causal? Based on literature research there are currently two main concepts. One is tried and tested; the other is in the experimental stage.
Proposed in 1969 the "Granger Causality" test determines whether a time series 'supports the prediction' of another time series (see Granger, 1969). It establishes a 'predictive causality'. If a signal 'Granger causes' another signal then past values of variable 1 contains information that could be used to predict variable 2 beyond the information contained in the former values of variable 2. The Granger cause is usually determined by a series of t-tests and F-tests, demonstrating that variable 1 provides statistically significant information about variable 2.[11]
Additive noise models are more recent and are based on a simple perception: "If one event influences another, then the random noise in the causing event will be reflected in the affected event".[12] Additive noise models (see Mooij *et al.*, 2009, Peters *et al.*, 2014) allow for 'nonlinear causal discovery'. The assumption is that if the relationship between two correlated variables is causal then it cannot be symmetrical. In any data there will be noise from various causes. We can deduct that the pattern of noise in the cause will be different to the pattern of noise in the effect. One has an effect on the other but not vice versa. Observations should reflect this. It is this nonlinearity which allows the determination of the direction of cause and effect. Going forward this seems to be a very promising concept. However, Granger Causality is computationally simpler and works well in an economic context (see *inter alia* Lütkepohl, 2011).[13]

---

[11] "X is said to Granger-cause Y if Y can be better predicted using the histories of both X and Y than it can by using the history of Y alone." (Giles, 2011).
[12] Mooji, J.M., see quote on Quartz web-page, CAUSE AND EFFECT, Mathematicians have finally figured out how to tell correlation from causation, online at: https://qz.com/316826/mathematicians-have-finally-figured-out-how-to-tell-correlation-from-causation/.
[13] For a more extensive discussion on Granger Causality we can refer to Kosa (2015), see also FN 53.

2.2.5. Ensembles' Choice

In this section we will discuss the current choices in learning algorithms. There are a fair number of Unsupervised and Supervised methods such as neural nets morphing into deep learning[14] (with nets on several layers), k-nearest neighbours (KNN) algorithms, Support Vector Machines (SVM), Classification and Regression Trees (CART), and Generalised Additive Models (GAM). Comparisons can be based on predictive power and accuracy, computational scalability, training time, robustness to outliers, missing or irrelevant data, number of parameters, *etc*. (see *inter alia* Hastie and Tibshirani, 1986, Caruana and Niculescu-Mizil, 2006, MATLAB, 2017). In our paper, we focus on classification techniques, specifically Random Forest and Gradient Boosting machines.

Both methods are based on fitted classifiers or classification trees. Classification trees are able to handle big data-sets, and quantitative and qualitative predictors. Redundant variables are ignored. Missing observations can be addressed through surrogate splits. The ability to interpret depends on the tree size. Small trees are easy to interpret, and large trees are difficult. Each 'terminal leaf' provides a 'class probability estimate'. In advanced tree algorithms we examine the variables at each split to find the best points for making the split. A balance has to be found between the ability to deal with complexity, variance and bias. Instead of using big trees to solve a complex problem, it is often better to average many smaller trees or to work with distinct additions. Doing so keeps the variance in check. This is true if the trees are 'sufficiently different'. Training with identical data creates similar trees. A way to 'shake' the data and to create a range of distinct trees is to sub-sample the training data.

We need to strike a balance between overfitting and underperformance. The more a model is optimised, the better a tree is fitted to the data. With noisy data, we run the risk that the model is fitted to the noise, rather than to the underlying structural information. At some point the variance of the model increases; and the model starts to over-fit. Yet if the model is not optimised, then the outcome is a weak learner with a high bias. As a consequence the model underperforms.

---

[14] See e.g. Smith and Topin (2016).

2.2.5.1. Random Forest

The Random Forest algorithm, introduced by Breiman in 1999, 'bootstraps' with replacement data sub-samples and grows numerous parallel trees (see Breiman, 2001). Simple majority votes across independent trees lead to the final classification/prediction. The trees should not become too shallow otherwise early biased optimums are locked in, based on very few weak base learners. Hence the trees need to be reasonably complex and deep. Generally, the number of trees should be high. As each tree has the same weight, increasing the quantity of trees does not alter the bias. It does not over-fit. The benefit for each additional tree 'levels out' – meaning it stops being beneficial but does not harm.

To achieve a small error rate, the trees need to be as uncorrelated as possible; the less correlated the trees are, the lower the variance. To increase the randomness, not all features are used at each 'nod' to determine the best split. Only a random sample of m < p is drawn, usually $m = \sqrt{p}$, where p is the number of features. Two to five features are standard; the smaller the number, the less correlated the trees. The number of 'm' is thus an important tuning factor for Random Forest. As with parallel trees, there is no way to reduce bias; complex problems can only be addressed with deep trees. As a consequence, we have to accept a high variance figure initially. By averaging across a higher quantity of trees we reduce the variance subsequently. In parts of our experiment, we employ the Random Forest classifier as implemented by Diaz-Uriarte and Alvarez de Andres (2006). It allows us to eliminate the least important variables in an iterative process. The method is implemented in 'R open source'.[15]

Highly correlated features are fairly problematic for tree-based learning algorithms such as Random Forest. They make optimal class boundary searches difficult. To differentiate between correlated and less correlated features we can apply PCA to the feature set as a pre-process step (see Wold *et al*., 1987). PCA reduces the variable universe, in particular by choosing the less correlated features. As such, applying PCA can make sense when using Random Forest. On the other hand, we find that PCA becomes counterproductive when using XGBoost (see below), as Gradient Boosting models have less issue with bias in general. We

[15] For a more general analysis of Random Forest models, see *inter alia* Biau (2012).

believe that the relative underperformance when applying PCA is due to implicit information loss.

2.2.5.2. Gradient Boosting

Boosting, introduced by Freund and Shapire in 1996 and generalised by Friedman in 1999[16], expands trees in an additive manner. Trees are not grown in 'parallel', as in Random Forest. Trees are grown in an iterative process, building a tree extension on top of an existing tree. Through the minimisation of cost functions, the algorithm grows incrementally into its complex functional form. Boosting can work with different loss functions like regression, logistic regression, resistant regression, K-class classification and risk modelling. The data is 'weighted' towards areas where the current trees are deficient and where the prediction accuracy is sub-optimal. It grows new trees specifically in areas missed by the past model. The classification is done by 'weighted majority vote' – with the weight adjusted for errors of previous trees. The weighting de-correlates the trees. The process according to AdaBoost (Freund and Shapire, 1996), is:

a) Initialise the observation weights $w_i = 1/N$, $i = 1,2,\ldots,N$
b) For $m = 1$ *to* M repeat the following:
    a. Fit a classifier to the training data using weights w_i.
    b. Compute the weighted error of the newest tree.
    c. Re-weight the weights basis on a log-normal calculation.

Gradient Boosting (Friedman, 1999) expands trees in a sequential process by fitting tree additions to so-called 'residuals'. Residuals are the 'gradient' of the cost or loss function, which is minimised. The gradient is approximated by a new tree and becomes an additive expansion of the former model. The trees are added using coefficients. The parameters are fitted in a 'forward stage-wise' fashion. The parameters of the next tree targeting the gradient are optimised, whilst holding fixed the parameters of the old model. Stage-wise fitting slows potential overfitting. Overfitting happens when the expected loss made on training data reduces beyond an optimum level. Continuing causes the population-expected loss to increase after the optimal point, which makes it counterproductive. In cases where the population-

[16] See Friedman (2001).

expected loss is not known, it becomes difficult to realise when the process is beyond the optimal point. We can only take precautions. Regularisation methods, for example, are such precautionary measures. They are designed to constrain the fitting process. A natural regularisation parameter limits the number of components. Another measure is to shrink the supplement. A ‘shrinkage parameter’ controls the learning rate. Instead of adding a new tree extension with the full learning rate to the model, we shrink the addition (by multiplying with a factor of < 1). Hence, we only change the model by a small amount. Again, we expand the tree to the new residuals. The residuals stay relatively bigger. The expansion results in an elevated number of smaller trees. This slows the minimisation of the loss function. Thus, by scaling the learning rate we create more steps. According to Friedman, the effect becomes more pronounced when shrinkage is applied with each iteration (incremental shrinkage). This seems to create a bigger effect than applying it only once, as a proportion to the entire model (global shrinkage). Simulation studies show that there is a trade-off between the optimal number of components M and the shrinkage factor. In general small learning rates (< 0.1) create higher optimal values for M (Friedman, 2001) and lead to better generalisation error rates (Friedman, 2001). Another ‘restraining mechanism’ against overfitting is a LASSO regularisation. Proposed by Friedman and Popescu (2004), it shrinks some of the coefficients to zero whilst adding a tree. Other restraining mechanisms are ‘influence trimming’ or ‘post-processing selection’. Many ensembles build on thousands of small trees. Post-processing selects a subset of trees and combines them efficiently. A potential method is to use regression to weight trees. As a result similar trees are left out, which reduces the number of trees and simplifies the model. The process is quicker as it uses less computational power.

In the following section we introduce the basic elements of a general gradient descent boosting paradigm in detail[17]. Predictive learning is ultimately an estimation problem. We can describe it as an optimisation process of a function with a response (random) output and an explanatory (random) input - based on a set of ‘training’ observations:

$$F^*(\boldsymbol{x}): \boldsymbol{x} \rightarrow y$$

with $F^* = argmin_{\text{F}}\, E_{y,\boldsymbol{x}}[L(y, F(\boldsymbol{x}))]$ the differential to F.

[17] The following mathematical presentation follows discussion with Tarek Hard.

The function $F(x)$ predicts observations by fitting base learners. 'Fitting' ensues, for example, through regression or a classification tree $h(x; a_m)$. Parameters are amongst others splitting variables, split locations and the terminal node responses. The parameterisation is based on:

$$F(\boldsymbol{x}; \{\beta_m, a_m\}_1^M) = \sum_{m=1}^M \beta_m \, h(\boldsymbol{x}; a_m)$$

with

β_m being the expansion coefficient and

a_m the parameterisation for each step of the algorithm m.

The algorithm minimises the expected value of the loss function $L(y,\ F(\boldsymbol{x}))$ over the joint distribution of all observations. Subject to 'regression or classification' typically the loss functions are chosen amid a range of fairly standard loss criterions like 'squared-error', 'absolute error', 'negative binomial log-likelihood' (Friedman *et al.*, 2000) or exponential loss (Freund and Schapire, 1996) *etc.* The base learning components β_m and $h(x; a)$ are added 'stage-wise sequential'. At each iteration m a base learner $h(x; a_m)$ is fitted to the negative gradient of the loss function $L(y,\ F(\boldsymbol{x}))$. The function $\beta_m h(x; a_m)$ can be interpreted as the 'best greedy step' toward estimate of F^*. The approximation $F_{m-1}(x)$ as predictor function is updated at each input data point to $F_m(\boldsymbol{x})$.[18] The process as a whole is described as a 'greedy forward, gradient descent, stage-wise additive' learning.

The optimal response γ at iteration *m* for a particular leaf *j* (of a tree with *J*-terminal nodes) that captures the response $h(x; a_m)$ and any scaling factor, is given as:

$$\gamma_{jm} = argmin_{\,\gamma} \sum_{x_i \epsilon R_{jm}} L(y_i,\ F_{m-1}(\boldsymbol{x}_i) + \gamma)$$

with

R_{jm} regions of terminal nodes in iteration *m*.

[18] Although the finite nature of the Training Data restricts gradient calculations to individual data points, the response $h(x; a_m)$ is a "sufficient" replacement to the unconstrained theoretical equivalent of the negative gradient $-\mathrm{g}_m(\mathrm{x}_i)$ in the functional space. The response is considered "most correlated" to $-\mathrm{g}_m(\mathrm{x}_i)$ (Friedman, 2001).

$$-\mathrm{g}_\mathrm{m}(\mathrm{x}_\mathrm{i}) = -\left[\frac{\partial \mathrm{L}(\mathrm{y}_\mathrm{i}, \mathrm{F}(\mathbf{x}))}{\partial \mathrm{F}(\mathbf{x})}\right]_{\mathrm{F(x)=F_{m-1}(x)}}$$

For a *binary* classification Friedman *et al.* (2000) suggest the following negative binomial log-likelihood as loss function:

$$L(y,F) = \ log(1 + exp(-2yF(x))) \qquad y\,\epsilon\{-1,1\}$$

$$F(x) = \frac{1}{2}\ Log\left[\frac{Pr(y = 1, \boldsymbol{x})}{Pr(y = -1, \boldsymbol{x})}\right]$$

The additive expansion of $\mathrm{F_m}(\mathbf{x})$ is updated that:

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \textstyle\sum_{j=1}^{J} \gamma_m\, 1(\boldsymbol{x} \in R_{jm})$$

$$\gamma_{jm} = \textstyle\sum_{x\in R_{jm}} \tilde{y}_i / \sum_{x\in R_{jm}} |\tilde{y}_i|\ (2 - |\tilde{y}_i|) \qquad (1)$$

We use as pseudo-response the derivative $\tilde{y}$ at the data point *i*:

$$\tilde{y}_i = -\left[\frac{\partial\, L\big(y_i, F(\boldsymbol{x_i})\big)}{\partial F(\boldsymbol{x_i})}\right]_{F(\boldsymbol{x_i})=F_{m-1}(x)} = 2y_i/(1 + exp(2y_i F_{m-1}(\boldsymbol{x_i})))$$

To update $F_m(\boldsymbol{x})$ at iteration m we apply a recursive algorithm. The final approximation $F_m(\boldsymbol{x})$ is converted to the probability for belonging to class +1 or -1:

$$p_+(\boldsymbol{x}) = \widehat{Pr}(y = 1 \mid \boldsymbol{x}) = \frac{1}{(1 + e^{-2F_M(\boldsymbol{x})})}$$

$$p_-(\boldsymbol{x}) = \widehat{Pr}(y = -1 \mid \boldsymbol{x}) = \frac{1}{(1 + e^{2F_M(\boldsymbol{x})})}$$

For *multi-class* classifications (k-class problems) the framework is extended and the loss function is:

$$L(\{y_k, F_k(\boldsymbol{x})\}_1^K) = -\sum_{k=1}^{K} y_k\, log\, p_k\,(\boldsymbol{x})$$

$y_k = 1\ (for\ class = k) \in \{0,1\}$ and $p_k(\boldsymbol{x}) = Pr(y_k = 1 \mid \boldsymbol{x})$.

K-trees are produced to predict the corresponding current residuals at iteration m as pseudo-responses:

$$F_{km}(\boldsymbol{x}) = F_{k,m-1}(\boldsymbol{x}) + \sum\nolimits_{j=1}^{J} \gamma_{jm}\, 1(\boldsymbol{x} \in R_{jkm})$$

$$\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} \tilde{y}_{ik}}{\sum_{x_i \in R_{jkm}} |\tilde{y}_{ik}|\,(1-|\tilde{y}_{ik}|)}$$

The point derivative is:

$$\tilde{y}_i = -\left[\frac{\partial\, L(\{y_{il},\, F_l(\boldsymbol{x_i})\}_{l=1}^{K})}{\partial F(\boldsymbol{x_i})}\right]_{\{F_l(x)=F_{l,m-1}(x)\}_1^K} = y_{ik} - p_{k,\ m-1}(\boldsymbol{x_i})$$

Thus the probability for class *k* is:

$$p_k(\boldsymbol{x}) = exp\big(F_k(\boldsymbol{x})\big) / \sum\nolimits_{l=1}^{K} exp(F_l(\boldsymbol{x}))$$

2.2.5.3. XGBoost – enhanced Gradient Boosting implementation

"XGBoost" – 'eXtreme Gradient Boosting' for Supervised learning - is a scalable machine learning system for tree boosting. It is available in R as open source (package by Chen and Tong He). It is computationally quite powerful as it employs parallel and distributed computing. According to Chen (2014) and Chen & Guestrin (2016) XGBoost allows for handling sparse data by applying a theoretically justified 'weighted quantile sketch procedure'. It improves further on the regularised objective, following the first and second order gradient boosting method from Friedman *et al.* (2000). XGBoost introduces as an additional regularisation a complexity driven term Ω (see below). Summed up over k trees it is added into the usual differentiable convex loss function. Complexity gets penalised. Model adaptions that are complex by nature are only accepted if there is a superior performance / information gain. As a consequence the model selects predictive but mostly simple functions.[19]

[19] All these functions are intended to reduce overfitting. Whether they are ultimately effective in balancing the bias - variance predicament is a priori difficult to infer. Conclusive empirical studies are not / not yet available.

A loss function $L(.)$ complemented by a complexity term expands into the objective function which is minimised on the m^{th} step (see Chen, 2014):[20]

$$Obj = L(.) + \sum_{k=1}^{K} \Omega(.)$$

$$\Omega(f_m) = \pi T + \frac{1}{2}\tau \sum_{j=1}^{T} \gamma_{oj}^2$$

π and τ are constants and γ_o is the response at each leaf j.

Following certain assumptions the objective function turns into:

$$Obj = -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \tau} + \pi T$$

The optimised response of each leaf is:

$$\gamma_{oj} = -\frac{G_j}{H_j + \tau}$$

While Gj is the sum of the first order derivatives, Hj of the second order derivatives of the loss function at each point of the observations.[21] The algorithm ranks and enumerates possible tree structures. The optimum is achieved by maximising the information gain at the splitting points of each node:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\tau} + \frac{G_R^2}{H_R+\tau} - \frac{(G_L+G_R)^2}{H_L+H_R+\tau}\right] - C$$

Subscripts L and R represent the left and right 'daughters' of a split and C conveys the 'complexity cost' added through the additional leaf.

Besides incremental shrinkage, XGBoost offers another interesting anti-overfitting technique, so far solely used in Random Forest. It allows for column or feature subsampling at each

[20] Some symbols are changed to avoid duplication with symbols already used in the text.

[21] *Obiter dictum*: For logistic classification Gj is given by the numerator of Equation (1) and likewise Hj by the denominator.

splitting point. According to user feedback, this appears to be very promising. Unfortunately, it is not yet available in open source, only as part of commercial software packages.[22] Besides a computational demanding 'exact greedy algorithm', which enumerates over all possible splits and all features, XGBoost also entails a less demanding 'approximation algorithm'. Candidate splitting points are proposed based on a quantile strategy using percentiles of feature distribution to bucket the continuous features. It determines the best solution by dissecting the aggregated statistics among the proposals. There is just one catch. Percentiles are distributed evenly across the data. This is appropriate for equally weighted data, but not for unequally weighted data. To find appropriate candidate splitting points, XGBoost uses an approximation algorithm called 'weighted quantile sketch algorithm'.

2.2.6. Tuning Hyper-parameters

Tuning parameters are ultimately optimisation constraints in the learning process. We have to distinguish between standard model parameters and hyper-parameters. Standard model parameters are determined by 'training on data', so that they reconstruct inputs well. Hyper-parameters, on the other hand, are either 'predefined'[23] or determined 'by inspection'. They are 'higher-level' properties such as a maximum level of complexity or a pre-set learning rate. Their general purpose is to limit overfitting. Typical examples for hyper-parameter are:

1. Number of trees or the leaves of a tree;
2. Depth of the trees;
3. Number of components M;
4. Number of classes in a k-means classification;
5. Learning or shrinkage rate – incremental, ideally < 0.1;
6. Complexity term factor for the cost function;
7. Number of randomised value subsamples; and
8. Number of randomised feature subsamples.

[22] User feedback on this feature is quite positive. It is implemented in several commercial software packages. There is a similar function 'colsample_bytree' in R but unfortunately not properly documented; so we did not use it.

[23] The levels either come from the AI community and are generally accepted levels that have worked in the past with other problems or they come from literature research.

The traditional way for optimising hyper-parameters is to conduct a parameter sweep. The so-called grid search is simply an examination of a manually-specified subset of the hyper-parameter space. Since the space may include real-valued or unbounded values, we often need to define limits or boundaries before performing the grid search. We take the outputs with the highest score in the validation process. "Cross Validation" is routinely used to estimate the generalisation. Grid search is unfortunately a parallel procedure, meaning that it has to be done for each one separately. The reason is that hyper-parameters are mostly independent of each other.[24]

Grid search is often an expensive method as a multitude of labels need to be examined. Less exhaustive search methodologies include, for example, a 'random search' and a 'Gradient-based optimisation'. Instead of searching over all labels, we search in a random search only over a randomised sub-sample of parameter settings. Restricting the search often makes the process more effective, particularly in high-dimensional spaces. There are many Hyper-parameters that do not significantly affect the outcome. For some learning algorithms, it is possible to calculate the gradient of the Hyper-parameters. Whenever this is possible we can use the gradient descent for optimisation purposes.

2.2.7. Training and Testing Routine

In this section we describe the 'training and testing routine' in more detail. It is impossible to assess beyond doubt whether a model functions in the future. Observations in the future are not known. The next best thing is to examine whether it would have worked in the past. To do this, we test with data that was not used for training. We cannot allow any 'future slippage' – information slipping from the future into the past, such as the modelling process. This is the reason for splitting observations into one set for training and one for testing. Training data can be tweaked, sampled or randomised. Test data on the other hand is for a single purpose only – to evaluate the 'found' model. We cannot be over-sensitive. The sample split needs to happen at the beginning. In an ideal world, the people training the

[24] This is sometimes referred to as the 'curse of dimensionality'.

model should not have access to test data. They would not even see it. Nor would they be allowed to test different models repeatedly.[25]

Training can be done in various ways. We already mentioned Cross Validation, also called left-out or hold-out validation. Hold-out data is simply a small subset of training data – up to 10% – taken by chance. Its purpose is to tune the modelling, setting 'reasonable' Hyper-parameter levels. After use the hold-out data is simply added back to the training set. Cross Validation sets are randomly sampled subsets of training data. The basic protocol is N-fold Cross Validation – dividing n-times and training n-times, each time with a different hold-out. An alternative to Cross Validation is "Walk Forward" analysis ('walking through time'). The model is trained with a time window of earlier observations and evaluated with data thereafter. The testing has to be out of sample. With each training step, the data simply walks forward. Former 'out of sample' data is now 'in sample' data. The modelling starts anew. It is possible to compare the performance of the algorithm over the different pre-test periods. We can assess for stability/robustness and for abnormalities. It is a good proxy for how the model will perform operationally in real-time. The standard routine is to replace the elder observations when walking forward. The training set stays constant in size. Alternatively, the new data is added. Longer-term data structures can be modelled. The disadvantage is that when information loss through time decay is fast, new information gets diluted. Old and new observations receive the same importance and are equally weighted. A counter-measure is to sample the data exponentially. We call this 'stratified sampling on an exponentially distributed sub-set of instances'. This technique puts more emphasize on recent observations. As a side effect it can also standardise the modelling on different sized observation sets.

Walk Forward analysis is used not only with testing, but also with running thereafter. As Walk Forward is done continuously, it solves a general predicament. We prefer long track-records, but we like to train on the most current information. The status of the world may change. There may be structural fractures. We test the model without prolonged black-out periods in training. With Walk Forward we continue to update the model while running it in real-time.

---

[25] Otherwise we end up as it is done in many customer presentations. 'Back-testing' strategies are proposed - often based on 'stop-loss'. The strategy outperforms the benchmark. The strategy works 'in hindsight'. Sometimes there is some 'track-record' – realised by running various strategies in parallel. Only the 'successful' ones make it to the presentation. The strategy does not work in real-time.

2.2.8. Dependent Variable – Prediction as Model Output

Learning algorithms frame data aiming to match a learned function as closely as possible to a hypothetical function (F) that would describe perfectly the reality between the input and the output. The variation of the output is studied as we change input or model parameters. We determine the effects that features have on output across the different model variations. Dependent variables represent the output. Hence it is important that the dependent variable matches the solution we are seeking as closely as possible.

2.2.9. Measuring Success

Comparing the performance of different classifiers is critical. When we control for quality we need to establish whether the results are stable and not just subject to chance. It can be done either by statistical tests (Garcia and Herrera, 2008) or by assessing similarities and differences (Jurman *et al.*, 2012). A "Confusion Matrix" is typically the basis for an assessment. The Confusion Matrix classifies the predictions into true or false answers. For a binomial Confusion Matrix there are four prediction categories TP (True Positives), TN (True Negatives), FP (False Positives) and FN (False Negatives).

		Predicted	**Class**
		Positive	Negative
Correct	Positive	TP	FN
Class	Negative	FP	TN

Figure 2: Binary Confusion Matrix

The Confusion Matrix is a helpful tool. It becomes complex in a multi-class extension. If we want to compare different classifiers, we need measures that condense the information in the Confusion Matrix. We list below some of the broadly used metrics. The ultimate goal is to summarise the Confusion Matrix into a single value to simplify the comparison process.

Typical statistical measures for a binary Confusion Matrix are:[26]

Accuracy (ACC) $= \frac{TP+TN}{TP+FN+TN+FP} = \frac{|trues|}{|positives|+|negatives|}$

Recall (REC)[27] $= \frac{TP}{TP+FN} = \frac{TP}{|positives|}$

Precision (PREC)[28] $= \frac{TP}{TP+FP} = \frac{TP}{|predicted\ as\ positive|}$

Matthews Correlation Coefficient (MCC) $= \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$

2.2.10. Summary – Elements of Choice

At the end of this section, we list once more the various factors in the context of training a machine learning algorithm. All elements in our list can influence the outcome. The checklist does not purport to be complete. It is intended as a practical starting point. Degrees of freedom in the learning process are:[29]

1. Defining the problem;
2. Opting for underlying data, adding complementary data sources, enriching data, selecting data;
3. Model selection out of SVMs, neuronal networks, deep learning, Random Forest, Gradient Boosting, *etc.*;
4. Opting for the best implementation with regard to bias, variance and overfitting, e.g. XGBoost in R as an additive Gradient Boosting model implementation with several randomisation features;
5. Sub-sampling, bagging or boosting techniques can be used to reduce potential overfitting in the learning algorithm;
6. Devising variations for the dependent variable;
7. Selecting benchmarks and other performance measures;
8. Deciding on the response, regression or classification, and the number of classifications;
9. Eventually refining the response of choice using additional constraints;

[26] The selection is based on the general acceptance in the field (Kosa, 2015, is opting for a broader spectrum of metrics).

[27] Recall is also called the True Positive Rate.

[28] Precision is also called the Positive Predictive Value.

[29] To be clear - not all 'degrees of freedom' are available in each of the model implementations. Choosing a specific implementation is already a decision in this context. It will allow or disallow some options – notwithstanding that most implementations allow for alterations.

10. Picking class limits, and calculating class cut-off values;
11. Re-balancing unbalanced observations by targeting equal probability for each classification;
12. Creating a range of features by designing relevant and diverse mechanics, transforming them into features, testing variations or different implementations;
13. Determining Causality to recognise leading variables for selected following variables;
14. Feature selection and extraction – selecting fewer features but a more relevant subset, using variable importance measures like PCA or Singular Value Decomposition;
15. Grid search, sub-sampling for Cross Validation purposes; the goal is to narrow down possible parameter values by holding steady all but one to search for local optima;
16. Identifying the loss function best suited to the problem, e.g. least squares, logarithmic loss, exponential loss; they differ *inter alia* in how heavily false predictions with high probability are penalised;
17. As an alternative to re-balancing unbalanced observations add a penalty factor to the loss function which gives higher focus to rarer events;
18. Randomisation – use of randomisation in the algorithmic process of growing trees, this has implications for reducing bias and overfitting;
19. One possibility is value or row sampling (motivated by Breiman, 1999, and used in Stochastic Gradient Boosting by Friedman, 2001) – equally or exponentially distributed;
20. Another possibility is column sampling, varying the feature universe per tree or per node;
21. Defining max depth per tree (low -> increasing bias, high -> helps with bias, but increasing complexity; to a degree necessary to solve difficult problems, increasing variance and possible interdependence);
22. Limiting the number of trees or the number of nodes – Tree size is an important parameter; stumps often do best; tree depth controls the interaction of trees;
23. Pruning – downsizing the tree algorithm by assessing which parts of the tree are redundant or only marginally effective and cutting them out;
24. Different optimisation procedures, such as stage-wise optimisation; they keep the good part of the tree fixed and only improve on the part where there is difficulty;
25. The weighting in a stage-wise optimisation, needs to be normalised to add up to 1. Eventually grid search to find reasonable weights for improving the areas of underperformance by improving on the error rates;

26. Limiting greedy boosting by applying a shrinkage factor [0, 1], either off-the-shelf value or grid search for establishing a reasonable level. The small shrinkage factor slows the growth of the tree, so growth is in increments, which helps with overfitting;
27. Regularisation – choice between LASSO and other regression methods; penalising complex tree additions, preventing them if the information gain is not substantial;
28. Different training and testing routines – for example, Walk Forward with (1) expanding subset of training data or (2) constant rolling window;
29. Instead of taking the data 1:1, the training set can be sampled either equally or through stratified exponential sampling – interesting when information loses value fast;
30. Controlling for quality – simple performance measures are Error Rate and Accuracy Rate; more balanced measures across all classifications are the Matthew's Correlation Coefficient and the Confusion Entropy; and
31. Instead of fixed hurdle rates perform ranking on a relative basis, e.g. 25% best true predictions, or refine further by leaving out high probability predictions.

2.3. Application – Market Trends[30]

Recent regulatory initiatives (such as the German Steagall act) have significantly restricted leeway when it comes to managing own / speculative trading positions. Despite primary focus on customer business we still need to risk manage positions arising from customer trading:

1. Primary markets;
2. Positions bought from customers – trading on Secondary markets;
3. Building Trading Inventory in anticipation of future demand.

In this paper all methods regarding forecasting market trends are centred on residuals, i.e. the errors between the value predicted and the actual value. We start with singular time series. There are numerous methods to cluster time series (e.g. Liao, 2005, Keogh *et al.*, 2003 or Fu, 2011). Predicting market trends is based on an assumption that patterns recur over time. Observations are grouped by means of shapes. Statistical shape analysis is applied in areas, where recurring shapes and forms are studied, e.g. biology, medicine, image analysis and archaeology (Dryden and Mardia 1998). A generalisation across a number of attributes aids the classification of new observations.

Another idea is to make classifications according to socioeconomic configurations. EW has a strong fellowship among market practitioners. The issue from an academic prospective is that wave labels are 'subject to subjective' interpretation. Hence, efforts are made to objectivise the rules. As a result, the framework becomes more rigid and loses some flexibility.

Finally, we look at data structure between asset classes. This last approach draws together elements of three topics – Spill-over effects, hybrid models and anomaly detection. We use Granger Causality to differentiate between "Leading" and "Following Markets". We assess whether Leading Markets which (Granger) cause a Following Market produce residuals that can be used to forecast such a Following Market[31].

[30] The application was made possible through and performed in cooperation with LBBW, as part of part of an author-led bank project, providing the necessary resources and infrastructure to run the extensive analytics.

[31] *Obiter dictum*: It is common to traditional hybrid models to work with features comprising time series residuals for forecasting purposes. However a point of difference for this paper is that we consider solely residuals of Leading Markets.

We summarise in Figure 3 the step sequence of the applied modelling procedure discussed in the previous section:

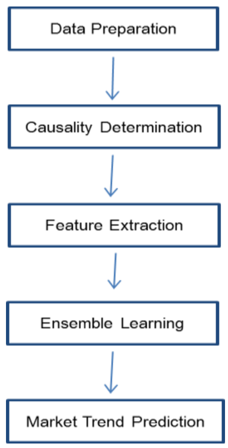


Figure 3: ‘Step-wise’ modelling procedure

2.3.1. Data and Data Preparation[32]

We use Bloomberg market data as raw data for this section, looking at daily opening and closing prices. We are using data on 300 asset classes. All our training is done on closing values. The time period covered is 18 years from January 1999 up to December 2016. The validation period is 2003 up to 2016. As long as we analyse a singular time series this is sufficient. When looking across different asset classes we need to ascertain, that time stamps of the different markets do not overlap. An extensive process was carried out to guard against data leakage from the future. One of the consequences though is information loss with regard to the timeliness of data.

2.3.2. Problem and Response

2.3.2.1. Specifying the problem

There are alternative ways to specify trend prediction of time series. We started with the ambition of simply predicting ‘whether the market would go up or down’ the next day. From an academic viewpoint a daily close of business day[33] (COB) analysis is sufficient. For real

[32] The input data was provided / prepared by the bank.

[33] The COB prices are taken for the calculations. COB is always meant in this paper as ‘COB to COB the following day’.

life applications different aspects come in to play. It depends on the intended purpose. Fund Management, for example, enters a position typically over a longer period of time. Thus, the problem can be solved using COB data. On the other hand there is short-term risk taking in a bank. Traders usually have a short time horizon. They are customarily measured based on daily profit and loss. Most positions are taken intra-day. Over-night positions, especially if there is no 24 hour hand-over process across different time zones, are risky and risk limits are often restrictive. Regulatory aspects have to be considered, too. Hence, the trading case is generally an intra-day or (at best) an 'open-to-close' problem. Therefore, depending on the intended purpose we should work with open, close or intra-day market data to minimise data friction.

2.3.2.2. Dependent Variable

The dependent variable is the data property or attribute we want to forecast, the response. In machine learning the choice is commonly between regression and classification. We keep it simple and start from the basic idea to differentiate between stable, rising and falling markets. To this end, the dependant variable reflects three conditions: Positive, Negative and Stable.[34] The multi-class response variable is therefore [+1, 0, -1]. Such classifications are calculated for each business day.

2.3.2.3. Classifications

Picking the right classifiers is important. Fixed hurdle rates for daily returns are highly sensitive to the accuracy level of the predictions. It is less ambitious to sort the trading days in a relative order. Minor miscalculations matter less. It is more stable and easier to optimise. To avoid unbalanced classifications we calculated the class cut-off points using a routine, which equalises the class dimensions. Thus no absolute hurdle rates regulate the classifications. Instead we 'bootstrap' the cut-off rates each time based on the top and lowest one third of the past data.

---

[34] The implemented classification (see below) is an approximation of the basic idea, as it reflects the relative order in thirds. For the most part the relative order will get us close to the original idea of positive, neutral and negative markets; but it is not same. The reason for doing it this way, is in keeping the classes balanced.

The *modus operandi* for Application (1) for each observation/prediction is:

1. We calculate every historical daily return before this date;
2. We draw at random 105 returns (with replacement);
3. We sort them in ascending order, with the 35^{th} and the 70^{th} number as initial approximation for the 33% and 66% quantile;
4. We repeat a 100.000 times steps 2 and 3; and
5. We average over all 33% and 66% quantiles using the results as class cut-off points / the upper respective lower class boarders.

Obiter dictum: For the customer coverage section we utilise later an alternative method - penalising for the difference by manipulating the loss function.[35]

2.3.3. Measuring and Benchmarks

Quality control is central to data analytics. The Confusion Matrix is again the basis for the assessment, classifying the predictions into true (T) and false (F) answers.[36]

2.3.3.1. Matthew's Correlation Coefficient[37]

The disadvantage of most of the statistics is the exclusive focus of each metric on one class specification. To control for quality overall we need to evaluate the classifiers by summarising the individual metrics into one single metric. In general, there are two conditions – the metric should make a distinction between different misclassification distributions and should work on unbalanced data. The current research suggests two different measures for multi-class Confusion Matrices – the Confusion Entropy (see Wei *et al.*, 2010) and a generalised Matthew's Correlation Coefficient (MCC) (see Gorodkin, 2004).

---

[35] We introduce a multiplication factor in the objective function for certain outcomes.

[36] For simplification and didactical reasons we could start with 'not discriminating between how false an answer is'. In this case we can extend the binary classification matrix into a simplified three-class matrix. Instead of TN, FN, TP and FP we differentiate between 6 prediction results T+1N (True +1 Negatives), T0N (True 0 Negatives), T-1N (True -1 Negatives), F+1P (False +1 Positives), F0P (False 0 Positives) and F-1P (False -1 Positives). Accordingly the typical measures based on the confusion matrix change from binary notations to:

$$\text{Accuracy (ACC)} = \frac{(T+1P)+(T0P)+(T-1P)+(T+1N)+(T0N)+(T-1N)}{|positives| + |negatives|} \qquad \text{etc…}$$

[37] For some elements in this section see Kosa (2015).

They are comparable but not identical. Confusion Entropy discriminates more between similar Confusion Matrices. Overall MCC seems to be the better compromise amid discrimination, consistency and coherence (Jurman *et al.*, 2012). Additionally we can control with the Chi-squared test statistic whether the predictions are better than random guesses (Baldi *et al.*, 2000). For two classes the MCC is related to the Chi-squared test statistic by:

$$|MCC| = \sqrt{x^2/_N}$$

We extend the multi-class classification problem and, following Jurman *et al.* (2012, p.2), define the MCC in terms of the Confusion Matrix as the ratio:

$$MCC = \frac{\sum_{k,l,m=1}^{N} C_{kk}C_{ml} - C_{lk}C_{km}}{\sqrt{\sum_{k=1}^{N}(\sum_{l=1}^{N} C_{lk})\,(\sum_{\substack{f,g=1\\ f\neq k}}^{N} C_{gf})}\ \sqrt{\sum_{k=1}^{N}(\sum_{l=1}^{N} C_{kl})\,(\sum_{\substack{f,g=1\\ f\neq k}}^{N} C_{fg})}}$$

with regard to a classification problem on S samples $S = \{si : 1 \leq i \leq S\}$ and N classes $\{1,..., N\}$. We define two functions tc, pc: $S \rightarrow \{1,..., N\}$ that represent for each sample s its true class tc(s) and its predicted class pc(s).

The corresponding Confusion Matrix is the square matrix $C \in M(N \times N, N)$. The ij^{th} entry Cij is the quantity of true class i instances assigned to class *j* by the classifier:

$$Cij = |\{s \in S: tc(s) = i \text{ and } pc(s) = j\}|$$

MCC has a range [-1, 1]. A perfect classification will score $MCC = 1$, where an extreme misclassification registers as $MCC = -1$. The random roll of a dice tallies as $MCC = 0$. We assume as 'Null hypothesis' independence between the true and the predicted outcomes (class labels), with the Chi-squared/Pearson's test statistic being approximately Chi-squared distributed. The significance level for rejection of the Null hypothesis is set to 0.05, corresponding to a Chi-squared/Pearson's test statistic above 9.488. We use daily COB values for the calculations.[38]

[38] When applying Granger Causality across different markets in different time zones we unfortunately have a 'time problem'. For calculating MCC we use consistently COB values, even in cases, where there is a limited 'overlap' between Following and Leading markets.

2.3.3.2. Performance *versus* naïve benchmarks

Another way to control for quality is to measure the performance against a naïve benchmark. With stock indices we compare the performance relative to a naïve outright long position in the index. The assessment is done on an absolute return basis and calculating the "Sharpe Ratio". The result of the Sharpe Ratio depends *inter alia* on the actual investment strategy.[39] The performance for all prediction models is calculated based on daily 'Long only positions for classes 0 and +1'.[40]

2.3.3.3. Performance *versus* the most simple model – the 'Single- asset classic Features' model

In addition we will run the algorithms on two simple attribute variables – return and sigma. Analyses 'on the effect of returns in the past on returns in the future' or 'the effect of standard deviation in the past on return in the future' are manifold and go way back.[41] We want to test in particular whether enhanced features out-perform the "Classic Features". We control for periods of different length L – periods of up to 20 days (see Table 1).

Table 1: Classic Feature set

	N	P	L	maxlvl
return(L)			1 2 3 5 20	
sigma(L)			5 20	

For the US markets, for example, all Leading markets have closed before the US market closes => no issue there. But for the DAX the US markets close is after the DAX market close => issue! We still use COB, as the 'overlap in time' is small compared to the 24h overall time window we predict. (When calculating the investment strategies we do this different. There we use instead OOB / 'open-to-open' market values.)

[39] The Sharpe Ratio can vary significantly, see i.e. section 2.3.7.4.

[40] Alternative strategies are i.e.: 'Long only for class +1', 'Long only for classes 0 and +1' and 'Long-Short for classes +1 and -1'. We use daily COB prices or respectively OOB in case of time overlap between markets – see each Footnote when relevant.

[41] See i.e. Chan *et al.* (1996). Moskowitz et al. (2012) report significant 'momentum' for various asset classes on a monthly basis with time lags of one to twelve months.

2.3.4. Invariant Shape Analysis using Landmarks[42]

2.3.4.1. Shape Analysis and Labelled Landmarks

The concept of Shape Analysis is to discover differences between observations and to sort them into separate classes. The idea was expanded by Lele and Richtsmeier in 2001. Form is defined as a "… characteristic that remains invariant under any rotation, translation and reflection…" (p. 73). As it is impossible to compare every element of geometric evidence about objects and their shapes, the information needs to be reduced and specified. The characteristics are called landmarks and are labelled for the purpose of differentiation. Thus, Landmarks are combinations of distinguishable points that describe relevant information within a string of data on a reduced basis. They condense information and are expressed as vectors. Landmarks can be extricated in multiple variations. Our goal is to create a different enough feature-set so that the learning algorithm can extract from the various reductions the informational essence of the underlying data.

2.3.4.2. The Feature Set[43]

We work based on a feature set of five[44] attributes, and variations thereof. Landmarks are instigated by using standard concepts like MPP(N,P), maxmin(L), maxminend(L), equi(N,L) and maxYdir(maxlvl,L). To test for sensitivity we vary the values for N, P and L. For details see Table 2.

L is the time interval, the number of days preceding the 'feature date'. Parameter N conveys how many landmarks, e.g. local minima and maxima, are required to produce the feature. P is the minimum price difference in percentage terms (5%, 10% …) for regulating the feature – the minimum return hurdle.[45] The parameters L, N and P are not pertinent for each landmark feature. Table 2 displays for each feature the relevant parameters and the parameter range.

---

[42] This section makes use of concepts initially considered by the author, the project team and Bernd Schumacher during student - supervisory / bank project discussions as per Schumacher (2014). Concepts (subset thereof) were subsequently expanded upon / modified and developed into model input Features within the author-lead bank project.

[43] Features are based / modified / expanded upon a previous bank project (Schumacher, 2014), see FN 42.

[44] We dropped one additional Unsupervised learning feature after initial tests. See below '*obiter dictum*'.

[45] We assume the optimal P to be larger for high volatility in the underlying asset or when we look for stronger, longer-lasting trends.

Each feature is described as a feature vector. The vectors are used as input variables for the model. In the following section we describe the chosen Landmark features in more detail.

Table 2: The Landmark Feature set

	N	**P**	**L**	**maxlvl**
MPP(N,P)	5	0.01		
	5	0.02		
	5	0.03		
	5	0.05		
maxmin(L)			20	
maxminend(L)			20	
equi(N,L)	2		20	
	3		20	
	5		20	
	10		20	
maxYdir(L, maxlvl)			20	1
			20	2
MPP(N,P)	5	0.01		
	5	0.02		
	5	0.03		
	5	0.05		
maxmin(L)			20	

The Minimal Percentage Principle (MPP) is a technique that automatically detects trends. It characterises time series by identifying local extremes. Each day is labelled as an up or down day, depending on whether the day is a part of an 'up-trend' or 'down-trend' between local extremes. 'Up' and 'down' in this sense is different to whether the respective return on a day is positive or negative. It is possible to have a negative day in an 'up-trend' and a positive day in a 'down-trend'. The concept is derived from the earlier Minimal Distance Percentage Principle by Perng *et al.* (2000). The rules pertaining to *MPP(N,P)* consist of four elements – (1) alternating Up and Down trends, (2) trends between local minima and maxima, (3) the 'ratio of change' has to be above a minimum percentage hurdle P, and (4) trends have to be 'maximal expanded' to the next local extreme.

Maxmin(L) and *maxminend(L)* are calculated based on the maximum, the minimum and the end of a time-let with length L number of days. The idea is that the position of a maximum and a minimum relative to each other or relative to the end point entails information about an upcoming trend.

Attribute *equi(N,L)* describes the equidistant N number of points within length L that split each time-let into parts of equivalent length (see equidistant sampling by Fu, 2011). Relevant information may be gathered by assessing prior (equidistant) return levels of the last L days – the immediate period before.

The last feature *maxYdir(maxlvl,L)* is based on the concept of Perceptually Important Point Compression by Chung *et al.* (2001). The idea is that it may be important to examine the pattern of relative outliers with respect to the smaller linear trends over the previous days. For that purpose time series are divided based on splitting points. The new splitting landmarks are calculated as the point of maximal relative delta (in y-direction / return-wise) to the linear trend-line between two landmarks. In the first iteration these are the start (landmark 1) and the end points (landmark 2). We add the splitting point as landmark 3. We repeat the split for the left (landmarks 1 and 3) and for the right side (landmarks 3 and 2), each time adding landmarks. The repetitive process continues for a predefined number of iterations - *maxlvl* times. For feature creation we run the iteration between 1 and 3 times.

Obiter dictum on a k-means Clustering algorithm: Additionally we experiment with 'Unsupervised Learning features being front-loaded' into a Supervised Learning algorithm. For this purpose we cluster observations by 'Partitioning around Medoids' (PAM), a method conceived by Kaufman and Rousseeuw (1987). We select certain patterns and depict them as vectors. The vectors are grouped by identifying a predefined number of representative incidents by means of PAM.[46] The characterisations - the representations of each cluster - are termed "Medoids". Each observation is clustered as a product of its minimal distance to one of the Medoids. The cluster assignment in return is taken as a feature for the Supervised

[46] To ascertain a 'reasonable' input for the number of clusters we operate the PAM algorithm with different numbers – trial and error. The optimum is chosen based on the Silhouette Coefficient statistic, as described by Kaufman and Roussew, 1990 on p.87.

learning algorithm. The initial results showed no relevance. As a consequence we dropped this feature.

2.3.5. Technical and Behavioural Analysis[47]

2.3.5.1. Background – Technical Analysis

Socioeconomics as a market concept has a psychological dimension. The ideas are based on emotional human behaviour. Markets are places where humans interact with humans. The interaction between humans may not be rational, rather based on emotions. Periods of fear alternate with periods of greed. People buy when others have bought before and sell when others are selling. The socio economic concept is not in sync with the ratio of a *homo economicus*. A lot of economic research is being done on the rational human being (i.e. Fama, 1970, 1989), less is being done on social economic effects. Behavioural biases and asymmetries in information affect markets (Mandelbrot and Hudson, 2004, Shiller, 1981, 2003, 2005, Orlitzky, 2013). Mass or macro level psychology is difficult to assess. Strategies are often subjective. Market practitioners often apply “Technical Analysis”. Indicators from Technical Analysis are manifold and often subject to interpretation. In our analysis we work with ‘objectivised’ unique identifiers.

2.3.5.2. Fractals and Fibonacci Numbers

Central concepts in Technical Analysis are ‘Fractals’ and ‘Fibonacci numbers’. Fractals are widely known, as we find them repeatedly in nature. Technical Analysis attempts to relate to the ‘underlying laws in nature’. Defining fractals in strict mathematical terms is challenging (Mandelbrot, 1987). For our purpose we define fractals as ‘shapes that are similar in form on different dimensional levels’. Fibonacci sequences can also be found in nature, for instance, the heads of sunflowers, pine cones, animal horns, shell spirals *etc*. It was first commented on in 1202 by an Italian mathematician Leonardo Pisano (Fibonacci).[48] The sequence is calculated by adding the previous two sequence numbers, leading to 0, 1, 1, 2, 3, 5, 8, 13, 21…. The Golden Ratio is $(F_{n+1}/F_n) = 1.618$; variations are e.g. 0.618 (F_n/F_{n+1}), 0.382

[47] This section makes use of concepts initially considered by the author, the project team and Mark Wolters during student - supervisory / bank project discussions as per Wolters (2014). Concepts (subset thereof) were subsequently expanded upon / modified and developed into model input Features within the author-lead bank project.

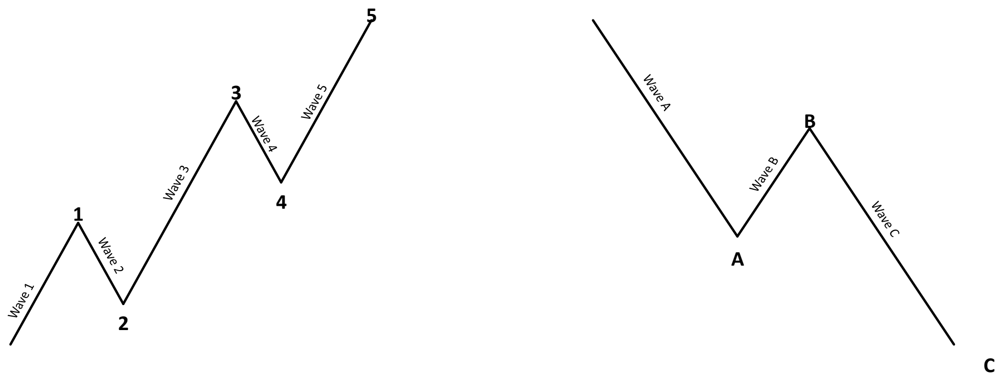
[48] The ancient Greeks must have been aware of it before this, as they used it in the context of the ‘Golden Ratio’ or ‘Divine Proportion’ when designing buildings. See Encyclopaedia Britannica.

(F_n/F_{n+2}) or its inverse 2.618. Fibonacci ratios are central to Technical Analysis, in particular to Elliott Wave theory (EW).

2.3.5.3. Elliott Wave Theory

The belief in EW is that markets rise and fall in a 'similar' and 'ever repetitive' fashion. Progressive periods alternate with corrective periods (Elliott, 1993). The concept incorporates implicitly the concept of Fractals, as waves of different order continuously overlap. The limited number of patterns resemble each other by definition when analysing the different 'degrees' of waves. The underlying assumption is that the laws of nature are reflected in mass psychology and thus markets. Fibonacci ratios are in general important for the correct classification of EW patterns.

EW as a concept is based on stock market studies, conducted by Ralph Nelson Elliott in the 1930th. He described market moves through a limited number of distinguishable patterns, which "are repetitive in form but not necessarily in time or amplitude." (Frost and Prechter, 2005, p.19). The most comprehensive work on EW is probably "Elliott Wave Principle: Key to Market Behaviour", published in 1978 by A. J. Frost and Robert R. Prechter (2005 / 1978). The wave principle reflects mass psychology, the progress of mass emotions. History repeats itself but not in an identical way; nevertheless according to Frost and Prechter (2005) markets do tend to follow a similar path each time. After a local extreme is reached, the trend starts to correct. If the wave develops in the same direction as the 'wave of one higher degree' it usually takes the shape of a five wave dynamic pattern, sometimes even with possible extensions. If it goes against the direction 'of one higher degree' it develops typically only three waves. Figures 4 + 5 illustrate the idealised basic pattern of 'dynamic' (see 1 to 5) and 'corrective' waves (see A to C).



(a) Dynamic wave (b) Corrective wave

Figure 4: Structure of dynamic and corrective waves[49]

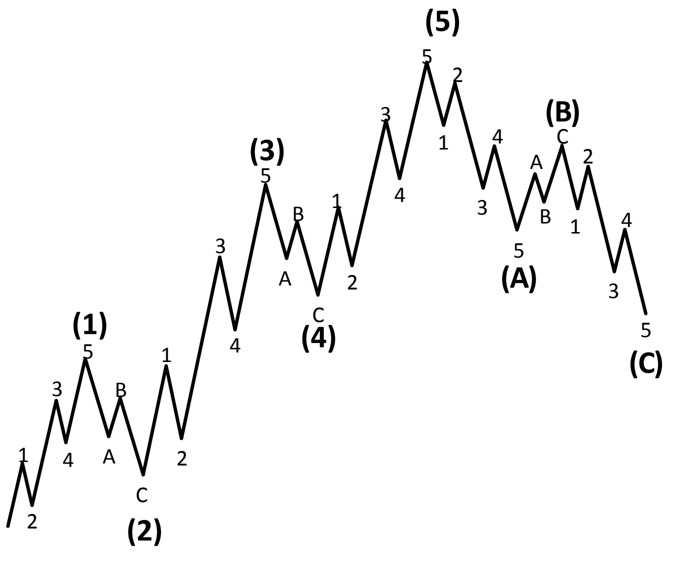


Figure 5: The Basic Pattern

The distinction between patterns is often difficult to make, as waves overlay. EW therefore defines further rules (strict) and guidelines (loose[50]). They do not in itself allow for a unique classification. The differentiation in degrees, and thus the classification to a specific pattern,

[49] Figures 4 + 5 are summary versions of the basic patterns shown in Frost and Prechter (2005) or by Elliott Wave International (see *e.g.* the 'Capsule Summary of the Wave Principle section' of "Global Market Perspective" November 2017 issue, p.55). Figures extracted from Wolters (2014), p.9 and p.10.

[50] One of the guidelines states that the more significant local extremes are, the more pronounced the countertrend will be.

is done in combination with Fibonacci numbers. Trend-/Reversal-target-areas are often predicted based on Fibonacci ratios. There are several patterns. Rankings in probability allow for categorising the wave structure. Unfortunately they are rather loosely defined. For operational reasons we introduce a set of specifying assumptions.[51] We further implement Fibonacci Multiples in fixed order and apply predefined probabilities to each pattern. We thus meddle with the pure EW conceptual framework. Details can be found in Appendix 8.1.1 and 8.1.2.

2.3.5.4. Features

The "EW-Features" we select are based on the idea of overlapping waves of different dimensions or degrees, alternating dynamic and corrective patterns. We calculate possible wave-structures on a 1500 data points rolling basis according to the operationalised EW rules.[52] We assess the probability of possible reversal areas based on Fibonacci relations. The calculations are done for different wave degrees. The odds are aggregated to 'a factor weighted average'. The reversal area of a wave with a higher degree is by definition the reversal area of lower degree waves. The more 'estimates' of different dimensions fall within a close target range, the more likely it is that the range will be a 'reversal area' (see e.g. Reversal Time / Areas Indicators in Appendix 8.1.4.).

All socio-economic indicators are calculated over N iterations and are expressed in vector format. The values for N are between 10 and 30. Waves of different degrees are extracted from alternating time horizons. As the maximum time window is 1500 days – we start with 1500 and curtail the time thereafter. We use a shortening factor of SF = 25%. The iteration stops when the analysed time intervals are getting too short. At max = 1500 and SF =25% we stop at N = 19. To summarise - a series of 'up-trends' and 'down-trends' constitute a wave. Usually a series of five is dynamic; a series of three is corrective. Waves of different degrees cover different lengths in time. Short time horizons produce only a few local extremes of low order; whereas long periods produce many low order maxima and minima, but few of higher order.

---

[51] We don't allow for i.e. side-ways patterns as they 'cover no ground'. In addition we establish an order of preference for wave patterns in up- and down-trends. The order of rules and guidelines becomes strict, instead of loose. In addition we correct for and reduce complexity.

[52] The 'operationalised EW features' and therefore Figures 6 - 8 are based / modified / partially expanded upon a previous bank project (Wolters, 2014), see FN 47.

‘Coverage Indicators’ display the number of finished up or down waves within time. The ‘Number of Waves Indicator’ is the result of subtracting the number of down-waves from the number of finished up-wave, expressed in percentage terms. The countertrend is supposed to be more pronounced after significant local extremes. EW-Features used in our modelling are listed in Appendix 8.1.4. To visualise the EW indicators we show in Figure 6 - 8 graphic examples.

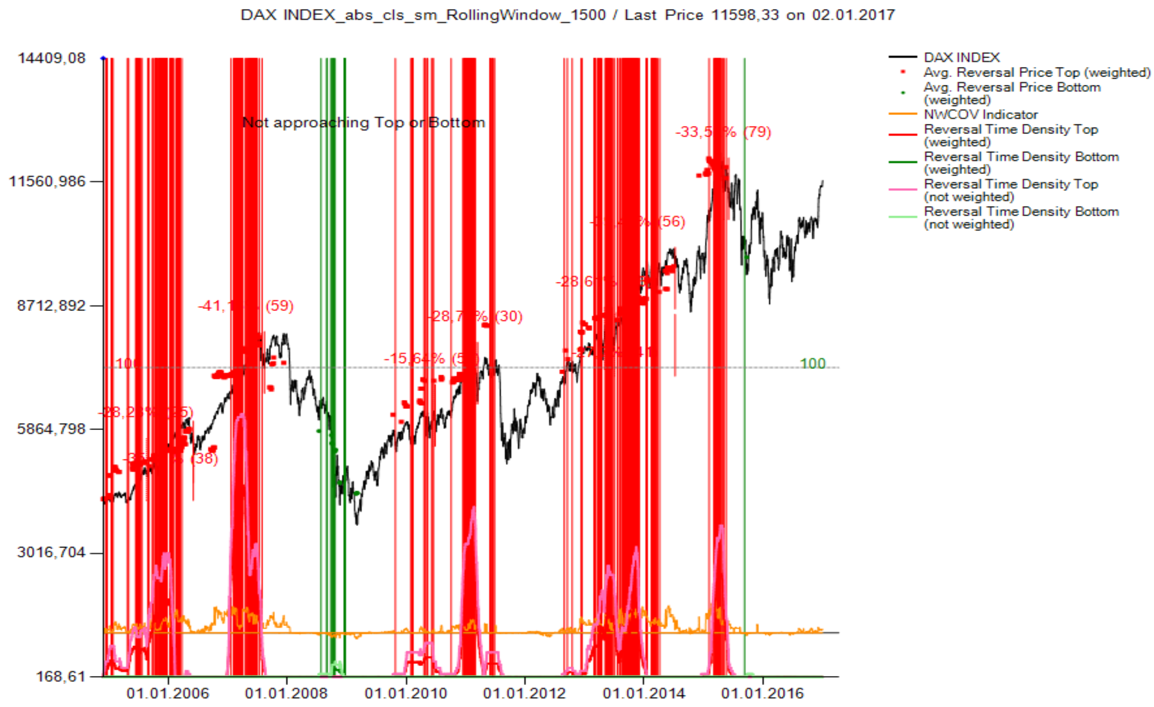


Figure 6: EW-Indicators DAX Index

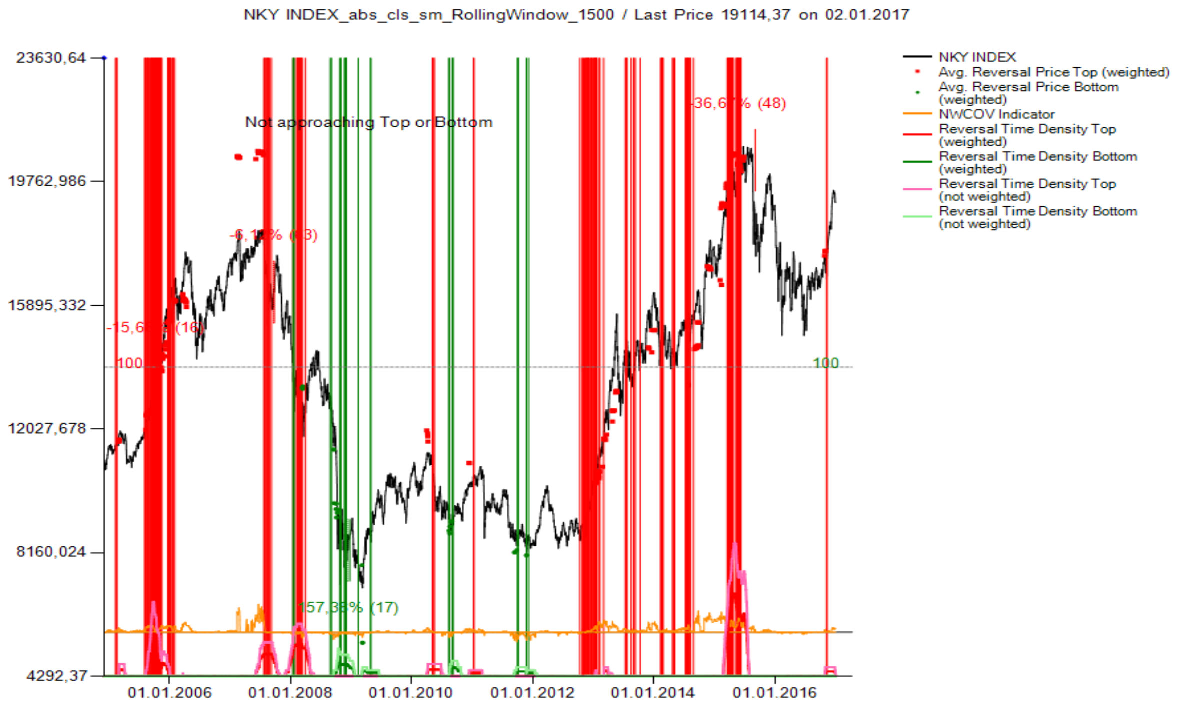


Figure 7: EW-Indicators NKY Index

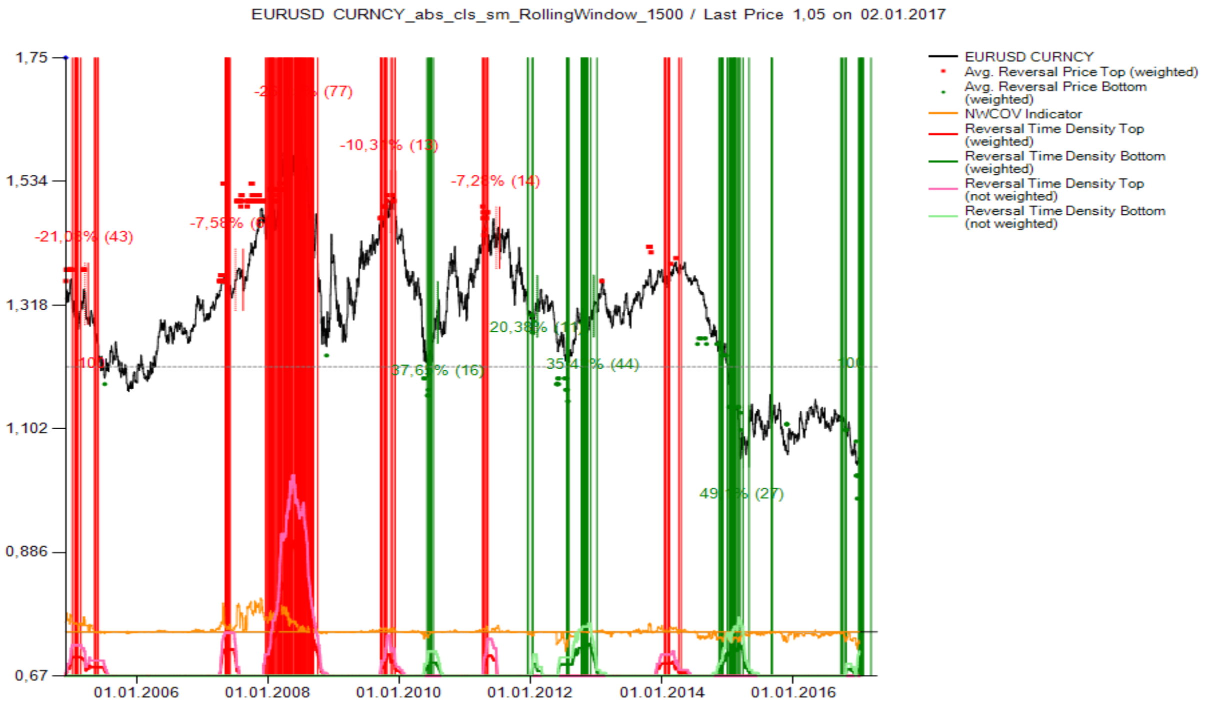


Figure 8: EW-Indicators EUR/USD

2.3.6. Leading and Following Markets[53]

This approach draws together three topics (1) Spill-over effects between markets, (2) hybrid models and (3) anomaly detection:

(1) Spill-overs occur when fluctuations in the price of an asset causes changes in the prices of other assets (IMF, 2017). Globalisation has strengthened Spill-over effects across countries (see *inter alia* Bessler and Yang, 2003). As driving factors behind this advance are seen more frequent and faster information transmission and fewer frictions to international trade and investments. The degree of financial integration between markets seems to matter more than their relative proportion. Historically USA and Japan were highly aligned (Eun and Shim, 1989, Sok-Gee and Karim 2010, Sidek *et al.*, 2011), more recently China has a strong impact on other emerging markets (IMF, *2017*). Spill-overs are strongest within sectors; and more so if the respective markets/sectors are dependent on external leverage. To quantify cause and effect many empirical studies utilise autoregressive models (Eun and Shim, 1989) as a metric, e.g. Granger causality (Nikmanesh *et al.*, 2014, Gurgul and Lach, 2009). Markets that spill over to other markets – show unidirectional Granger causality. We call them Leading Markets *versus* Following Markets.

(2) Traditional time series models often assume linearity between the variables (usually following certain Pre-processing operations such as differencing *etc*.). However, the links appear to be more complex. In the nineties a lot of research was done utilising non-linear models such as neural networks, SVMs or genetic programming. Soon hybrid models combined traditional linear models such as GARCH, ARIMA, and SARIMA with the non-linear methods (Khashei and Bijari, 2010). The data input for non-linear models is the historic residuals[54] generated by the linear model calibration. Thus the hybrid model infers further information from residuals which traditionally would have been discarded as white noise.

(3) Residual analysis is also applied in anomaly detection, the identification of instances that do not conform to prior patterns (Cheboli, 2010). Prediction based detection again makes use

[53] This section makes use of concepts initially considered by the author, the project team and Patrick Kosa during student - supervisory / bank project discussions as per Kosa (2015). Concepts (subset thereof) were subsequently expanded upon / modified and developed into model input Features within the author-lead bank project.

[54] By 'historic' we mean the observed residuals for past data points that contribute to current parameterisation of the time series model.

of autoregressive models to monitor for outliers. Anomalies are registered as observations that do not fit the confidence interval - the usual variance - of the model. Applications include network attacks (Yaacob, 2010), network traffic management (Moayedi, 2008), environmental sensing (Hill, 2010) and motor failure prediction. Residuals contribute to the detection of anomalies. They produce a warning signal that the current observation of the time series has deviated significantly from the expected value given its history.

While hybrid models consider residuals as a predictive feature for forecasting, anomaly detection utilises residuals to signal a change in behaviour. We aim to combine the different themes. We intend to exploit the signalling properties of current residuals generated by calibrating various time series models to Leading Markets. Leading and Following Markets are deduced by Granger Causality. On the other hand, we use residuals as Pre-process inputs for the modelling[55]. Following this approach we are broadening the data space considerably. We use information from 300 markets globally.

2.3.6.1. Dependent Variable and Data Preparation

We designate several major stock indices across Asia, Europe and USA as Following Markets. The stock indices are the markets we aim to forecast. We use the same classifications [-1, 0, +1]. To establish which markets are the Leading Markets for the respective Following Market we test 300 individual stocks, indices, FX and commodities for Granger causality (see Appendix 8.2.1.). The Granger test is ultimately a pre-test on training observations. The analysis was carried out on publicly available data comprising daily prices/returns of such prices. In the event that data was not available in a Leading market, for example, due to different holiday calendars of various markets, then data from the preceding business day was used in its place. Whilst more sophisticated techniques such as surrogate variables (Friedman *et al.*, 2009 p.311) could be used to manage missing incidents, the adopted approach is simple and provides a sufficiently accurate proxy. The approach is consistent for chronologically comparable time series.

[55] *Obiter dictum*: Common to traditional hybrid models we build a predictive feature set comprising time series residuals for forecasting purposes, however a point of difference for this paper is that it considers just residuals of various Leading Markets, whereas traditional hybrid models consider residuals of the same market being forecast in addition to the other linear forecast components described above.

2.3.6.2. Granger Causality

As discussed earlier, there are different concepts on how to test for causality. We looked at Granger and, as an alternative, the Additive Noise model. Because of its simplicity and its acceptance in financial empiric studies we have selected Granger Causality as our method of choice. Financial time series are innately non-stationary (see Granger, 1981, Engle and Granger, 1987). Ordinarily we would use differencing procedures to obtain a stationary process. To minimise potential information loss we employ instead a procedure proposed by Toda and Yamamoto (1995). Its advantage is that it can be used in non-stationary cases without transformation via differences – as long as only one of the process variables is non-stationary.[56]

2.3.6.3. Features

On asset classes showing Granger causality we calculate residuals that result from standard models, such as ARIMA, ARIMAX and GARCH. Later we add the previously used attributes: Classic Features, labelled Landmarks and EW-Features. Due to computational restrictions we use simplified Landmarks – 'Landmarks light'.

We calibrate each of the models to each of the designated Leading Markets for each Price time series. Table 3 summarises the Time Series Models and corresponding definition of the residuals.

Table 3: List of Residuals and 'Landmark light' Features[57]

Model	"Pre-Process Feature Set"	Implementation Reference
Vector Autoregressive		Lütkepohl (2011)
Vector Error Correction Model (VEC(p))		Johansen (1991)
Autoregressive Integrated Moving Average Model (ARIMA)	Residual= $\hat{y}_{t-1} - y_{t-1}$	Pascual *et al.* (2004), Box et al. (2013)
Autoregressive Integrated Moving Average Model with explanatory variable (ARIMAX)		Box *et al.* (2013)

[56] See Kosa (2015), with ref. to FN 53.

[57] Residuals and 'Landmarks light' are based on a previous bank project (Kosa, 2015), with ref. to FN 53.

Generalized Autoregressive Conditional Heteroscedasticity Model (GARCH 2,2)	$\varepsilon_{t-1} = \frac{y_{t-1}}{\sigma_{t-1}}$	Engle (1982), Bollerslev (1986)
Block Extrema	$\delta_i^{j,\ \max} := X_i^j - \max_{1 \le l \le p} X_{i-l}^j$ $\delta_i^{j,\ \min} := X_i^j - \min_{1 \le l \le p} X_{i-l}^j$	
Maximum Gain & Loss	$L_i^j := \min\left(-\max_{0 \le l_1 < l_2 \le p}\left(X_{i-l_2}^j - X_{i-l_1}^j\right),\ 0\right)$ $G_i^j := \max\left(-\min_{0 \le l_1 < l_2 \le p}\left(X_{i-l_2}^j - X_{i-l_1}^j\right),\ 0\right)$	

2.3.7. Results and Discussion[58]

To discuss the models we run various models on six stock indices, two in Asia, two in Europe and two in the USA.[59] The focus is on the major economies in each region. As a start we show a typical example with strong result, evaluating the sample model based on the three measures explained in section 2.3.3. In Table 5 we show the collected statistics to give a general overview. Following that, we single out three singular cases of interest. For further details please see Appendix 8.2.2.

2.3.7.1. Example case

Multi-asset calculation on the Nikkei Index with Granger Causality pre-selection and ‘Classic Features‘:

Table 4: Confusion Matrix – NKY,
Multi-asset with Granger Causality and Classic Features.

		Predicted Class		
		-1	0	1
Correct Class	-1	518	323	183
	0	266	620	355
	1	143	381	632

The Confusion Matrix for next day predictions is calculated on daily closing prices of the Nikkei. The matrix comprises a high number of TPs. The strong prediction performance of this sample model shows in:

[58] Calculations and charts were performed as part of the bank project, see also FN 30.
[59] The selection being: Nikkei, HSI, DAX, UKX, SPX and INDU indices.

a) The statistical prediction measures are:

ACC	= 0.52	MCC	= 0.27
REC (-1, 0, +1)	= 0.56/0.47/0.54	Chi-squared	= 612.59
PREC (-1, 0, +1)	= 0.51/0.50/0.55		

b) The performance measures we compare a sample investment strategy based on the model predictions with a long only position in the Nikkei (benchmark).[60] As the sample strategy we decided on long only for classes ‘0‘ and ‘+1‘. The return is calculated on a ‘open-to-open‘ daily (OOB)[61] basis.

- ⇨ The performance improves significantly from a Sharpe Ratio of 0.25 (benchmark) to a Sharpe Ratio of 1.11 (model). Similarly the Maximum Drawdown decreases from 11185 (benchmark) to 5398 (model).

c) Finally, we compare the investment strategy performance based on the enhanced feature set with the investment strategy performance based on the most simple ‘Single-asset Classic Features‘[62]:

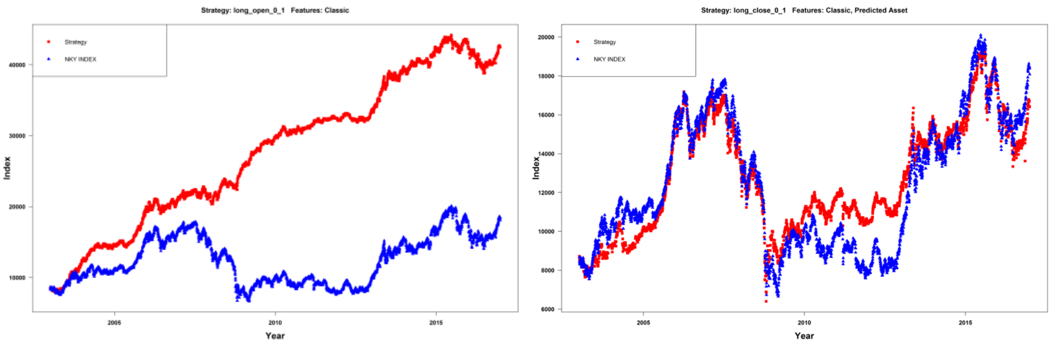


Figure 9a, b: Investment Performance – ‘Granger & Classic’ *versus* ‘Classic’ Features.
Left is ‘XGBoost, multi-asset with Granger & Classic Features’. Right is ‘XGBoost, single-asset with Classic Features’. Model (Red) versus benchmark NKY Index (Blue).

---

[60] For simplicity and to make the comparison to the benchmark (without trading costs) from a performance perspective analogous we don’t correct for bid-offer costs in the ‘model investment strategy’.

[61] ‘Opening of business day’ prices *versus* ‘Close of business day’ prices. OOB is in this paper always meant as ‘OOB to OOB the following day’.

[62] See section 2.2.9.: Return and sigma.

2.3.7.2. Overview - Table

Table 5: Prediction statistics – Overview across models for NKY, DAX, and SPX Index.

				NKY[63]				
Time series	Granger	Feature		ACC	PREC	REC	MCC	Sharpe R[64]
single-asset		Classic	-1	0.38	0.27	0.32	0.07	0.24
			0		0.46	0.45		
			1		0.40	0.37		
single-asset		Landmarks	-1	0.36	0.29	0.30	0.04	0.19
			0		0.42	0.41		
			1		0.35	0.35		
single-asset		EW	-1	0.37	0.22	0.33	0.05	0.10
			0		0.51	0.39		
			1		0.35	0.36		
multi-asset	x	Classic	-1	0.52	0.51	0.56	0.27	1.11
			0		0.50	0.47		
			1		0.55	0.54		
multi-asset	x	Landmarks	-1	0.46	0.48	0.48	0.19	1.01
			0		0.42	0.43		
			1		0.47	0.47		
multi-asset	x	EW	-1	0.39	0.34	0.43	0.08	0.01
			0		0.40	0.40		
			1		0.43	0.36		
multi-asset	x	Residuals	-1	0.44	0.50	0.42	0.17	0.63
			0		0.32	0.43		
			1		0.54	0.48		
multi-asset	x	Res + LM	-1	0.45	0.50	0.43	0.18	0.86
			0		0.32	0.44		
			1		0.55	0.48		

[63] The Sharpe Ratio for the benchmark is ca. 0.25.

[64] The result for the 'model Sharpe Ratio' depends *inter alia* on the actual investment strategy implementation. The Sharpe Ratio can vary substantially, see i.e. section 2.3.7.4 with the strong outperformance of the strategy 'long_only_+1' *vs.* 'long_only_0_+1'.

Table 5 (cont.): Prediction statistics – Overview across models for NKY, DAX, and SPX Index.

				DAX[65]				
Time series	Granger	Feature		ACC	PREC	REC	MCC	Sharpe R
single-asset		Classic	-1	0.39	0.33	0.33	0.08	0.44
			0		0.58	0.43		
			1		0.24	0.37		
single-asset		Landmarks	-1	0.37	0.40	0.32	0.06	0.17
			0		0.52	0.42		
			1		0.19	0.38		
single-asset		EW	-1	0.38	0.26	0.38	0.07	0.23
			0		0.56	0.41		
			1		0.32	0.35		
multi-asset	x	Classic	-1	0.41	0.36	0.39	0.11	0.39
			0		0.53	0.43		
			1		0.33	0.40		
multi-asset	x	Landmarks	-1	0.38	0.39	0.35	0.07	0.13
			0		0.47	0.42		
			1		0.29	0.36		
multi-asset	x	EW	-1	0.38	0.32	0.35	0.06	0.21
			0		0.54	0.39		
			1		0.26	0.39		
multi-asset	x	Residuals	-1	0.41	0.37	0.39	0.11	0.56
			0		0.45	0.42		
			1		0.39	0.41		
multi-asset	x	Res + LM	-1	0.42	0.39	0.40	0.12	0.45
			0		0.45	0.43		
			1		0.41	0.41		

[65] The Sharpe Ratio for the benchmark is ca. 0.49.

Table 5 (cont.): Prediction statistics – Overview across models for NKY, DAX, and SPX Index.

				SPX[66]				
Time series	Granger	Feature		ACC	PREC	REC	MCC	Sharpe R
single-asset		Classic	-1	0.42	0.30	0.33	0.12	0.34
			0		0.55	0.48		
			1		0.37	0.41		
single-asset		Landmarks	-1	0.41	0.33	0.33	0.12	0.49
			0		0.44	0.49		
			1		0.45	0.41		
single-asset		EW	-1	0.42	0.14	0.34	0.12	0.31
			0		0.55	0.46		
			1		0.53	0.40		
multi-asset	x	Classic	-1	0.41	0.28	0.34	0.10	0.51
			0		0.56	0.46		
			1		0.35	0.38		
multi-asset	x	Landmarks	-1	0.40	0.33	0.33	0.09	0.34
			0		0.50	0.45		
			1		0.35	0.39		
multi-asset	x	EW	-1	0.41	0.15	0.34	0.11	0.27
			0		0.57	0.44		
			1		0.48	0.39		
multi-asset	x	Residuals	-1	0.38	0.32	0.33	0.06	0.45
			0		0.48	0.41		
			1		0.32	0.37		
multi-asset	x	Res + LM	-1	0.40	0.31	0.35	0.09	0.47
			0		0.53	0.44		
			1		0.33	0.38		

[66] The Sharpe Ratio for the benchmark is ca. 0.4.

2.3.7.3. Special case – ‘Outstanding’ performance

Multi-asset predictions based on XGBoost with Granger pre-selection and ‘Residuals & Landmark light’ Features, calculated on the UKX index, United Kingdom and the DAX index:

Table 6: Confusion Matrix – UKX,
Multi-asset with Granger Causality and ‘Residuals & Landmark light’ Features.

		Predicted Class		
		-1	0	1
Correct Class	-1	493	319	312
	0	395	549	328
	1	301	354	487

Table 7: Confusion Matrix – DAX,
Multi-asset with Granger Causality and ‘Residuals & Landmark light’ Features.

		Predicted Class		
		-1	0	1
Correct Class	-1	434	380	301
	0	350	583	367
	1	291	389	467

a) The statistical prediction measures are for UKX (upper) *versus* DAX (lower):
-> *similar, slightly weaker for DAX.*

ACC	= 0.43	MCC	= 0.15
REC (-1, 0, +1)	= 0.41/0.45/0.43	Chi-squared	= 161.72
PREC (-1, 0, +1)	= 0.44/0.43/0.43		

ACC	= 0.42	MCC	= 0.12
REC (-1, 0, +1)	= 0.40/0.43/0.41	Chi-squared	= 109.09
PREC (-1, 0, +1)	= 0.39/0.45/0.41		

b) Comparing the strategy performances:[67]

- ⇨ For the UKX: The Sharpe Ratio improves from 0.23 (benchmark) to 1.78 (model). The Maximum Drawdown decreases from 3320 (benchmark) to 923 (model).
- ⇨ For the DAX: The Sharpe Ratio improves from 0.41 (benchmark) to 0.45 (model) only. The Maximum Drawdown decreases from 4425 (benchmark) to 4035 (model).

c) Comparing model 'outperformance' *versus* benchmark over a period of time:

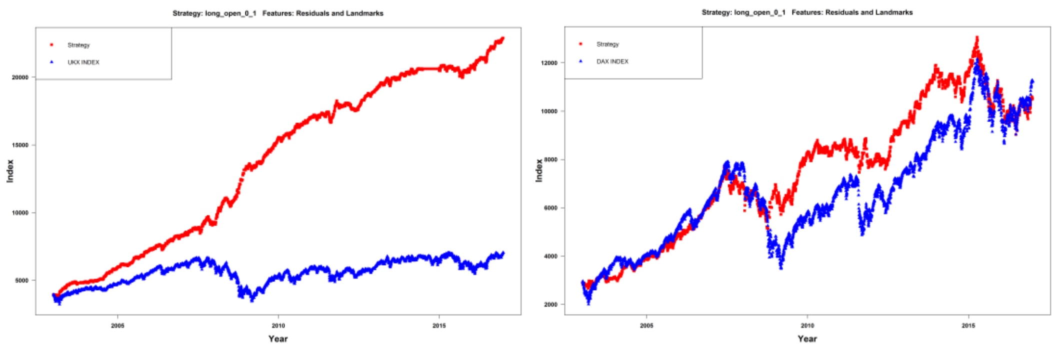


Figure 10a, 10b: Investment Performance – UKX (Right) versus DAX (Left).
'XGBoost, multi-asset with Granger Causality and Residuals & 'Landmarks light' Features.
Model (Red) versus benchmark indices (Blue).

- ⇨ ASTOUNDING OUTPERFORMANCE of the UKX model in comparison to the DAX model, despite the model predictions statistics not being substantially different.
- ⇨ The lesson we finally learned: When preparing the data we were aware, that there is a time-overlap issue regarding multi-asset calculations (e.g. with US indices - Leading Markets based on Granger Causality - closing after the UKX index or DAX index close). We adjusted for this effect by calculating the performance based on OOB prices. The alteration affected the DAX results as usual, but not the UKX results. After several iterations with Bloomberg it became clear, that Bloomberg uses – only in the case of the UKX index - 'closing prices' as next day 'opening prices'.[68]

- ⇨ Using data from 3rd parties is problematic. On a bigger scale it needs sophisticated validation tools. Beside other techniques we could use e.g. similar prediction models as such a 'quality validation tool' set.

---

[67] Once more the investment strategy is long only for 'classes 0 and +1', The return is calculated on a OOB basis.

[68] This is different for the DAX index, where Bloomberg uses the proper opening prices. A justification could not be provided.

⇨ It gives an indication of ‘overnight’ performance relative to the performance ‘over trading hours’. A lot of important company or market news are purposefully released outside trading hours. The aggregated effect is astonishing, even when considering the time ratio of 2:1.[69]

2.3.7.4. Special case - Comparing different model investment strategies

Single time series predictions, based on XGBoost with ‘Landmark’ Features, and calculated on the SPX index, United States:

Table 8: Confusion Matrix – SPX, Single-asset with Landmark Features.

		Predicted Class		
		-1	0	1
Correct Class	-1	357	309	400
	0	358	549	347
	1	356	268	514

The Confusion Matrix for ‘next day predictions‘ is calculated on daily closing prices of the S&P index. The matrix comprises a positive, but rather unremarkable number of TPs. The prediction performance is of slightly superior quality:

a) The statistical prediction measures are:

ACC	= 0.41	MCC	= 0.12
REC (-1, 0, +1)	= 0.33/0.49/0.41	Chi-squared	= 136.18
PREC (-1, 0, +1)	= 0.33/0.44/0.45		

b) Comparing the investment strategy performance based on an altered investment strategy.

We calculate the performance based on long only just for class ‘+1‘, and compare the results with the usual strategy based on long only for classes ‘0‘ and ‘+1‘. The return is calculated on

[69] 16 hours ‘overnight’ *vs.* 8 ‘trading hours’.

a COB basis. The Sharpe Ratio varies between 0.43 (benchmark), 0.49 (model [0, +1]), and 0.88 (model [+1]).

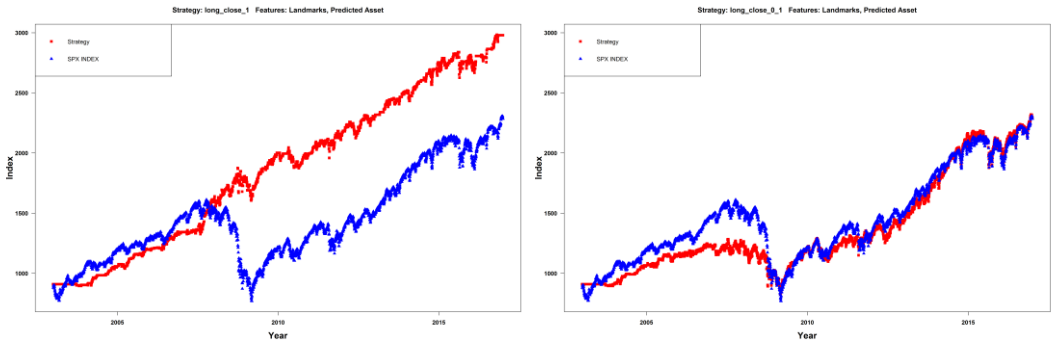


Figure 11a, 11b: Investment Performance – comparing strategies ‘long_only_+1’ and ‘long_only_0_+1’. Left is ‘XGBoost, Single-asset with Landmarks and strategy ‘+1‘. Right is ‘XGBoost, Single-asset with Landmarks and strategy ‘0’ and ‘+1‘. Model (Red) versus benchmark SPX Index (Blue)

The strong investment strategy performance compared to the only slightly superior statistical prediction measures is astounding. The massive outperformance compared to our standard investment strategy based on long only for classes ‘0‘ and ‘+1‘ is even more so. After the ‘lesson learned’ on the UKX index we investigated the data quality. We went through several iterations with Bloomberg, but could not identify any data irregularities.

Obiter dictum: Let’s define for the SPX index a simple trading rule: Whenever the 2-day return, calculated on ‘COB - 1 day’, is negative we put on a long only position for one day (COB). This gives us the following performance for the time period 2003 - 2016, see Figure 12b.

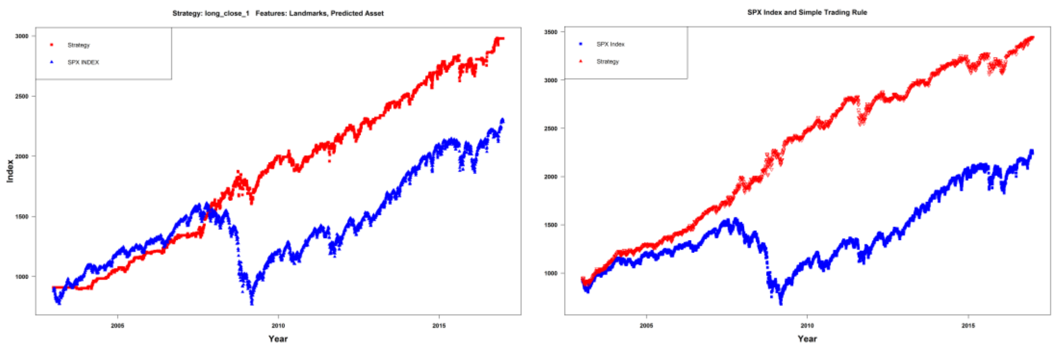


Figure 12a, 12b: Investment Performance Figure 11a (Left) to a Simple Trading Rule (Right) Model / Trading Rule (Red) *versus* SPX Index (Blue).

The performance of the trading rule looks astonishingly similar to the results before.[70] The learning algorithm seems to successfully pick up on a pattern like this.

2.3.7.5. Special Case - Model comparison – XGBoost *versus* Random Forest

Comparing Multi-asset with **XGBoost** model, Granger Causality and 'Residuals' and Multi-asset with **Random Forest** model, Granger Causality and 'Residuals':

Table 9: Confusion Matrix – DAX with XGBoost, Multi-asset with Granger Causality and Residuals.

		Predicted Class		
		-1	0	1
Correct Class	-1	418	397	300
	0	336	586	348
	1	278	423	446

Table 10: Confusion Matrix – DAX with Random Forest, Multi-asset with Granger Causality and Residuals.

		Predicted Class		
		-1	0	1
Correct Class	-1	398	440	273
	0	335	686	278
	1	298	469	374

a) The statistical prediction measures for the **XGBoost** model are:

ACC	= 0.41	MCC	= 0.11
REC (-1, 0, +1)	= 0.39/0.42/0.41	Chi-squared	= 88.50
PREC (-1, 0, +1)	= 0.37/0.45/0.39		

The statistical prediction measures for the **Random Forest** model are:

ACC	= 0.41	MCC	= 0.11
REC (-1, 0, +1)	= 0.39/0.43/0.40	Chi-squared	= 85.92
PREC (-1, 0, +1)	= 0.36/0.53/0.53		

[70] This amazing fact highlights why data quality is so important, especially when data is not controlled by oneself.

b) Performance comparison:[71] The Sharpe Ratio changes from 0.41 (benchmark) to 0.56 (XGBoost) and 0.39 (Random Forest). The Maximum Drawdown reduces from 4425 (benchmark) to 3609 (XGBoost) and 2863 (Random Forest).

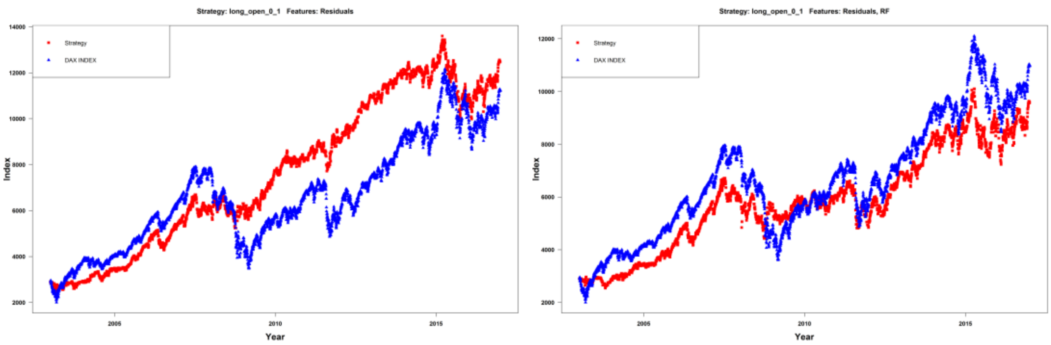


Figure 13a, 13b: Investment Performance – comparing XGBoost and Random Forest. Left is 'XGBoost, multi-asset with Granger Causality and Residuals'. Right is 'Random Forest, multi-asset with Granger Causality and Residuals'. Model (Red) *versus* benchmark DAX (Blue)

⇨ There are generally differences in outcome depending on which algorithm is being used, as demonstrated in this example. While the prediction quality is similar, see for example Accuracy, Precision or MCC, we can see greater differences in the Performance, see e.g. Sharpe Ratio. In this case XGBoost results in a superior performance compared to Random Forest. The variation may be less or the opposite in other circumstances.[72]

[71] We use the usual strategy of 'long only_0_+1'.

[72] We had, for instance, one result on the Nikkei where the RF gave better results than XGBoost. Comments from Kaggle competitions suggest that XGBoost is 'at large' the superior model (Kaggle, 2013). We got the impression from our results that on the question 'which model is superior', the answer depends *inter alia* on 'how noisy' the underlying data is. To substantiate any claim, further research on a broad / representative basis would be necessary.

2.3.8. Concluding Remarks Regarding the Results

1. In general, the learning algorithms *add value*. We see positive MCC values and overall a superior investment performance.
2. The models appear to be stable and *robust*.
3. The *results vary with different features*. Thus we have validated our claim that features influence the results other things being equal.
4. The prediction results from the Single time series learning models are *overall positive*, but *at large inferior* to the multi-asset learning models. This is not surprising, as we feed additional market information into the models. By applying complexity reduction methods – i.e. Granger Causality tests - we limit the negative ramifications of information overload (model intricacy).
5. Landmark results are generally lesser quality than the results based on the Classic Features, at least for DAX and Nikkei. This holds true for Single time series as well as multi-asset calculations. We see some positive effect when adding 'Landmarks light' to Residuals.
6. It is notable that class '0' predictions have higher 'TP values' than categories '-1' and '+1'. This can be seen across Single time series and multi-asset calculations for all indices (with the exception for the Nikkei multi-asset results). The reason may be that many features, e.g. return and sigma, are calculated over multi-day periods. The learning algorithm seems to pick up on the fact that markets start to become volatile. As a result we see a drop of class '0'categorisations around periods of high volatility (see Figure 49 as an example calculation in Appendix 8.2.2.6.).[73] Let's call this effect for the lack of a better word 'volatility clustering'.
7. The learning effect through EW patterns seems generally low.
8. The HSI index results are surprising when comparing them to Nikkei, see Appendix 8.2.2.2. While MCC values are similar, the strategy performance is substantially poorer.[74] From a macroeconomic perspective this may be explained by the specific characteristics of the Hong Kong stock exchange. China was not, and still is not, a fully open market economy. Anomalies are, for example, A- and B-shares, market transfer restrictions

[73] Thus the models seem to perform well in regard that it should not be class '0', but less so whether it should be class '+1' or class '-1'.

[74] Therefore we repeated the data controls, especially checking again for time issues on the opening and closing prices. No issues were identified. So it may just be the case that the model performs well on the less volatile days and much worse on the volatile ones.

regarding the Renminbi, *etc*. Existing transfer restrictions have eased over time. Interestingly, it looks that the investment strategy performs better after 2010.

9. The outperformance through time-overlap as shown in the case of the UKX index[75] is extraordinary. The effect is unexpectedly pronounced.
10. The outperformance in the case of the SPX index through change in strategy from 'long_0_1' to 'long_1'[76] is astonishing.[77]
11. We compared Random Forest with XGBoost on an exemplary basis, running calculations on the DAX index. While MCC results are identical, the True Positives for the categories '-1' and '+1' are better with XGBoost, while the TPs for class '0' are better with RF (see TP values in the different Confusion Matrices). This explains the relative better strategy performance for the XGBoost implementation.[78]
12. We carried out another inspection[79], running a multi-asset feature set *with* and *without* the Granger pre-filter. While the MCC was better 'with' than 'without', the Granger Causality test did not advance the investment performance as such.[80]
13. The Granger pre-filter has a far bigger reduction effect for the SPX Index (reduction to 16 Leading Markets out of 300) than the DAX index (99 Leading Markets) and the NKY index (214 Leading Markets), see Appendix 8.2.1.

Obiter dictum: The model results are robust enough to show reliably any anomalies. They either lead to interesting properties or hint at hidden data inconsistencies.[81] Both are important elements for advancement. We expand on this point further after the customer coverage section.

[75] Bloomberg takes the 'closing prices' as 'opening prices' for the following day.

[76] Instead of going 'long' for classes '0' and '+1', the strategy changes to going 'long' only for '+1' predictions. It would be interesting to see the effect for 'long_1_short_-1', meaning going 'long' for the '+1' category and 'short' for the '-1' category. There are many more variations for optimising a potential investment strategy. Augmenting around learning algorithms is an open-ended endeavour.

[77] Despite all that is been done so far, this would vindicate further inquiry going forward, ideally by checking with i.e. data from Reuters or by using real-time data with different cut-off times per day.

[78] It can be different in other cases where there is, for example, less 'noisy' data (see Nikkei).

[79] 'Inspection by example', see Appendix 8.2.2.1. For conclusive impact studies to discern reliably between effects coming from certain variations we would need to employ specific 'causality / impact measures' as part of the modelling processes.

[80] This is true apart from a minor improvement on the Maximum Drawdown.

[81] This is the reason why 'Data scientists' and 'data quality control algorithms' will become imperative.

2.3.9. *Excursus*: Extension idea – designing automated Portfolio Advice[82]

An extension idea on trend prediction could be to exploit the results for optimising the portfolio composition in the context of a given risk-return profile. If successful this could become a first step on the way to an optimised robotic-advisory tool. As theoretical justification for our 'enhanced' portfolio approach we refer to Merton (1987) who demonstrates that, with incomplete and asymmetric information, the assumption of an efficient portfolio based on mere diversification is invalid. Portfolios benefit from more and better information. Asset management is a multi-dimensional problem as it looks for efficient portfolios over many assets at different points in time. An investment advisor / fund manager has to select from a given set of products the optimal portfolio combination. The optimum depends on the market environment and on predefined risk categorisations. Customer classification takes centre stage in modern regulation and potential litigation[83]. Customers' knowledge and expertise, as well as their risk-bearing-capacity, need to be regularly assessed. An automated process could minimise operational and legal risks. As long the modelling is done impartially it should go a long way to prevent discrimination.

The Traditional Markowitz model

The Harry Markowitz model works on the presumption that the past is a reflection of the near term future. The model is also known as the mean-variance model as it optimises the portfolio composition out of an asset universe on the basis of expected return (mean) and standard deviation (variance).[84] For that purpose an objective function consisting of several parameters – e.g. return, volatility and a pairwise correlation matrix - is calibrated over a given number of assets. The model is based on specific assumptions, i.e. rational investor, risk adversity, consumption preference and a convex and increasing utility function of an investor *etc.* A risk of 0 produces a portfolio with minimal volatility but low return. A high risk number means the portfolio will be more volatile but the expected return is greater.

[82] Concept developed in discussion between the author and the bank project team.

[83] Litigation *per se* comes about after the fact. The judicial system evaluates whether a customer understood or was able to understand the risks. Furthermore it assesses whether the transaction was appropriate in itself. The judicial findings often affect the regulation. Some products and markets are effectively now closed for certain categories of customers, i.e. IRS for German municipalities. Rules around the topic are complex and a declared intention is just one of many considerations.

[84] The Markowitz portfolio allocation model functions in parts as a filter, selecting the assets that would have constituted the most efficient portfolio during a certain period in the past. In addition it calculates the optimal 'dose' per asset.

Computed over different risk parameters the model extracts the efficient frontier (the efficient portfolios under mean-variance importance).

The Enhanced Markowitz Model

We describe in the following an idea how a "Traditional Markowitz model" could become an "Enhanced Markowitz model", simply by restricting the input parameters based on our market prediction results in the prior sections. Instead of feeding every allowed asset class as population into the Markowitz model, we feed only assets with positive or the more positive[85] prediction values for the specific point in time. We 'extend' by 'limitation'. We select the ones that the XGBoost model believes to have a neutral or positive return on that day. We take the learning model with the best predictive value for each asset following our procedure in the prior sections. We apply the prediction result 'by day and by asset'. If the classification on the day for the asset is [0, 1] it is in. If it is [-1] it is out. The underlying asset sub-sample for the Markowitz model reduces by the number of assets with '-1' classification on that day. Depending on the day we may end up with very few assets in the sub-sample. We are aware that this may bias the results considerably. Studies have shown that the Markowitz model portfolio usually performs best with 30 to 50 underlying assets. As diversification matters in a portfolio context we may by far undershoot the optimum. This is a disadvantage that the enhanced model has *versus* the traditional model. To restrain this negative effect we could limit the pre-filter function to a minimum number of assets, ranked according to their classification probability of being '-1'.

To assess the performance of the model on a more general level we could generate multiple strategies by sub-sampling the initial asset portfolio. Each of the strategies would have varying asset composition and risk levels. For example:

1. A minimum of e.g. five asset sub-samples are randomly selected and checked for uniqueness.[86]
2. Twenty-one risk weightings are produced, ranging from 0 to 1 in 0.05 increment.

[85] To be exact – the assets with a prediction value of the top two classes which are classified as "0" and "+1". The values in the classes are not necessarily positive in absolute value terms.

[86] Our intention is to increase the data set over which we can evaluate the two Markowitz models. Unfortunately due to the computational limitations we have to accept that the portfolios are similar by nature.

3. This ensues in 21 risk weighting vectors, which reflect the respective weighting for each asset in the sub-sample. The weighting factors by asset are optimised over 10-day-moving average values for return and pair-wise correlation with respect to the specific risk parameter. The optimal weighting for an asset ranges between 0% and max 20%. We do not allow for short positions in an asset.
4. We apply those 21 vector weightings to all 5 sub-sample groups, creating 105 unique strategies and their respective statistics: 105 Sharpe Ratios, 105 portfolio returns and 105 portfolio variances.

To compare the Enhanced model with the Traditional model we could accumulate the daily portfolio returns over certain time windows, for example, 100 days.

$$return\ (i) = \sum_{n=1}^{100} daily\ return(n)$$

If we deduct the accumulated returns of the Traditional model from the Enhanced model we receive the out-performance value for $i = 1, \ldots, 105$:

$$y_i = return(i)_{Enhanced\ Markovitz} - return(i)_{Traditional\ Markovitz}$$

We could implement the same over the Sharpe Ratio and portfolio variance. For each statistic this results in 105 observations, from which we can perform a t-test. If the results differ significantly from 0 we conclude that one model is better than the other. If the number is positive we can be confident that the Enhanced model is superior to the Traditional one – especially if we consider the disadvantage due to the suboptimal number of underlying assets. Based on the assumption of y_i being independent and identically distributed we can apply the central limit theorem, calculating the following test statistic:

$$Z = \sqrt{n}\frac{\bar{y}}{S} \sim N(0,1)$$

with *n* the number of observations, $\bar{y}$ the sample mean of y_i, *S* the sample standard deviation of y_i and *Z* following a N(0,1) Gaussian distribution with mean 0 and variance 1. The goal is to show that both portfolios – the enhanced and the traditional – are different with 90%, 95% or even 99% significance level.

We tested the described experimental set-up on an exemplary basis with the five[87] global stock indices, which we used in the prior section. From a diversification perspective this is less than the minimum. Nevertheless, the outcome is surprisingly robust. The Enhanced model does outperform the Traditional model substantially. Obviously the Enhanced model aggregates implicitly the prediction outperformance of the single prediction models in the Portfolio context. The results are shown in Appendix 8.2.3. As the modelling is computational-wise highly intensive, we show at this stage just the score for a single test calculation.[88] Nevertheless, already the over-simplified model calculation (with only five underlying indices) confirms that further investigation is warranted. Furthermore, the sample calculation validates the positive test results of trend prediction section overall. The research would benefit from real-time data during the day. Any data inconsistencies with regard to data could be properly explored.

We wanted to demonstrate in this section that there are other applications conceivable for our developed tool-set then just trading. The finance industry would most likely be very interested[89] in a functioning 'fully-automated Investment Management' set-up.

---

[87] We excluded the UKX index due to the identified data / time-overlap issue.

[88] The annualised return from Enhanced *versus* Traditional Markowitz portfolio increases from 1.02 % to 6.78%. The Sharpe Ratio from 0.01 to 0.33, which is impressive based on just five underlying indices. For more please see the Appendix. Calculations were performed as part of the bank project, using a Markowitz bank model coded / run by Vineet Gupta.

[89] Portfolio theory is such a broad and important area of research and highly relevant for the financial industry, particularly in the current market environment. Fired up by the 'cheap money' resulting from central bank policy, trillions have flown into asset management. It is noticeable that investments shift from Active Fund Management (with high fees) to Passive Fund Management (with low fees). Portfolio allocation based on quantitative evaluation is predominately linked to passive investing. The human element is mostly oversight and fine-tuning. Hence, once a pool of quantitative strategies is developed the day to day operations are relatively straightforward. The actual run costs are low.

2.4. Application – Institutional Customer Behaviour[90]

Improving on customer coverage can be highly disruptive for any industry. Amazon and social networks are famous for judging potential client behaviour and thus their interest. It is not just memorisation of what buyers bought or looked at in the past. The powerful algorithms in play are making *ex ante* deductions from patterns or network connections centred on proprietary data and expanded by public or externally acquired data sources. They affect targeted changes and analyse the effects in customer behaviour. Tests take place real-time on sub-samples of their clientele – in an experimental set-up and after it is proven to work in a controlled environment it is rolled out to all customers. Up to now this is focused mostly on retail - in particular in finance due to the massive volumes of data available. We want to demonstrate that it can also be done in institutional business. The potential for disruption is exceptionally high for transparent (and soon fully automated) businesses like bond or equity trading, primary and secondary, IR derivatives, commodities and FX.

The final goal is to build a model and ultimately an experimental framework where we can optimise (1) sales efforts, (2) inventory and (3) customer behaviour to changes in price and contact. To do this we would need to analyse internal and external behaviour via a feed-back loop. To produce statistically conclusive results, we would mean to establish ‘a random control group experimental set-up’. The effect of the proposed changes could be observed by measuring how they affect the behaviour in the groups that received the changes verses those who have not. On Sales (1) the experimental line-up may look like the following: Every sales person is given a recommendation of ‘whom to talk to and what about’. One group receives the model’s suggestion, and a control group a randomised set of recommendations, and a 3rd group are given none. The groups are occasionally switched to account for externalities. In (2) the inventory case two different bond lists could be send to customers – one containing additional bonds predicted to be of interest by the model and another without. Again we would alter the target and control groups over time. It is important to experiment in a control group fashion and assess what works. This is the primary principal. Modern A/B testing frameworks would allow us to conduct on-going experimentation in an efficient manner, with minimal disruption to daily operating business.

---

[90] The application was made possible through and performed in cooperation with LBBW, as part of an author-led bank project, providing the necessary resources, the proprietary novel whole-sale banking data and the infrastructure to run the extensive analytics.

Besides optimising customer business for long-term relationships and sustainable profitability we see yet another reason for predicting customer behaviour. Under the German Steagall act unrestricted risk taking is not allowed. Banks need to justify short-term risk positions. Risk needs to be in relation to actual or potential customer demand. Therefore, it may be beneficial to optimise the trading inventory composition not only for profitability but also for regulatory reasons. Justification for making a trade would be objective and systematic. The German Steagall act is subject to interpretation and individual judgement. A trader takes personal responsibility when executing a trade. An automated process based on data analytics could free the trader and simplify the auditing and supervision process. It would require a documented model and could be done *ex ante*. In this paper we focus on building the prediction model as a prerequisite for the experimental set-up.

2.4.1. Problem and Response

Assessing institutional customer behaviour is a novel case for learning algorithms. Empirical studies can be found on consumer data, however in the case of bigger institutions (the institutional case), there is no public data available. We use anonymised data from a bank's proprietary database. The problem to solve is first to model 'what do customers do?' and second 'how do customers react when something is done?'. We split the task into two parts. Regarding 'what do customers do?' it is necessary to answer 'when do they do things?' and 'what they do when they do something?' separately. With the data available our goal in this section is to predict whether a customer will trade within a week and what is traded, conditional on a trade happening. To determine the best customer set customers are ranked according to their probability to trade. Binary True or False classifications can be derived based on the optimisation of a probability boundary. The intent is to prioritise on probability, i.e. to identify which are the most worthwhile customers to call. Our ambition is to optimise the utilisation of given resources and to steer the Sales effort accordingly. We need to work out the most profitable 'sweet spot' of which customers to cover when. Humans are inherently biased. They use heuristics and experience to determine whom to call and what to sell them. They like to call people they like, talk about things they know or do what they believe to be successful with. Using this model we aim to remove these biases. Our goal is to develop an objective assessment instead. To do this we need to calculate the probability of a class of trades being executed in a week.

2.4.2. Data and Data Preparation[91]

The raw data was taken in an anonymised manner from a customer database. The data relates to big institutional clients and companies. All personal data was excluded before the start of the analysis. The database includes all the trades and some orders. The information originates from in-house trading systems and iMarket. Missed trades, trades lost against competitors, are put in manually by Sales staff. The data ranges from 2010 to 2016. The data is split into a five year training set and a two year test set.

As we linked different data sets, not all the data we used was audited and we had to employ data cleaning operations. Issues were *inter alia* missing data points, incomplete information about trades, no customer reference as well as customer details changing over time. Some customers merged; other got divided over time. There were issues around unique identification criteria. Some customers had different names or name variations. Subsidiaries were not consistently linked to the correct parent companies. Given the large data set of 1mn entries we applied similarity techniques. Methods were, amongst others, 'dynamic time warping' to determine similarity of trading behaviour over time, 'Levenschtein distance' to measure similarity of name, and 'Euclidean distance' to analyse similarity of profile. Approximately 1000 trades were discarded as there was not sufficient evidence to link them to specific customers. We only consider trades to be those that were completed or we received a firm order for.

2.4.3. Features

To design features we follow methods from Supervised and Unsupervised learning. One idea is to determine the mechanics of what drives customer action and then represent the mechanics as something tangible, e.g. the rebalancing cycle of a fund. This could potentially be linked to specific states of the world or market patterns. As an example we link by trade the level of Itraxx relative to the yield on IBoxx EUR, or to the S&P equity volatility *etc.* Alternatively we use algorithmically selected clustering mechanisms resulting from unintuitive black boxes (Kaggle, 2013). In other words we run Unsupervised clustering techniques to identify patterns. We divide customers by assessing their similarity - a technique by Hartigan (1979). We analyse, for example, their trading activity, using a 'k

[91] The input data was provided / prepared by the bank.

Medoids Clustering' algorithm. We accumulate the results by customer. This functions as feature input to the learning algorithm. As a customer reference we include a categorical variable. Otherwise we end up with over several thousand different categories, significantly increasing the computation time. Instead we categorise customers as being similar based on a set of exponentially weighted moving averages. Typical characteristics are: Time between trades, average size of trade, and percentage of financials/corporates/sovereigns traded by trade count *etc.* We cluster the customers around k-Medoids - using the elbow criterion to select an optimal number of clusters. We average the sum of the squares within a cluster and measure it against the number of clusters. We minimise the ratio and round it to the nearest integer. The result is the optimal number of clusters – in our case six. We run the clustering process on a yearly basis to adapt for changing customer behaviour over time, assigning the new 'cluster membership'. New entrants are assigned to the 'Null category' and included in the new clustering after a minimum of nine months data. The clusters become model features along with additional cluster based modifications and expansions such as inter-cluster trade leadership. For an overview of the most significant features see Appendix 8.3.2.

For a given number of clusters k and customers n we assign the customer to its cluster by minimising the objective function J in an iterative fashion. To ensure that we achieve the clustering closest to the global optimum we use a hundred thousand seeds. We select the one with the highest value of J when optimising:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

$x_i^{(j)}$ is part of cluster j (with centroid c_j) and quantifies the customer attribute used to measure the distance to the centroid $(x_i^{(j)} - c_j)$ in the Euclidian space. We minimise for an increasing number of clusters k (k = 1, 2, 3 …). Typically this results in a marginal decrease in inter-member variance J. The marginal increase in clusters fails to improve intra-cluster similarity. Updating the clustering yearly ensures that a change in customer behaviour is properly considered in the underlying customer segmentation.

2.4.4. Training and Testing

When the underlying data varies in time we cannot use Cross Validation to check for overfitting and consistency. Cross Validation would use data from the future to model and predict events in the past. This forces us to use Walk Forward Validation. The technique sets a retraining frequency for the model and walks forward through time on this increment and calculating predictions for the future observations. We start with two years of data and a retraining with weekly frequency. The process does not result in future information leaking into the past model. We choose a weekly routine as it fits to the customer coverage problem we want to improve on.[92] The training period is not held constant; instead we expand the set iteratively to the full length of the Training data. During Training we dissect the results, change the Hyper-parameter, apply different sets of features and potentially even use several ensemble combinations *etc.* However, during Testing we run the chosen optimal algorithm without interference - in Walk Forward mode. The discipline ensures that overfitting is kept to a minimum.

Obiter dictum: It is known that customer behaviour shifts over time and with markets. Thus, we assume time variance with respect to our customer trade data. To assess the severity we perform some preliminary tests and check 'through rolling windows'. With 'stratified sampling on exponentially distributed sub-sets of instances' we could control for 'recency' effects. However, the outcomes are such that they don't justify the additional complexity.

2.4.5. Results and Discussion[93]

In the following section we show the results based on XGBoost - the upgraded gradient boosting implementation. Initial analysis was also carried out based on the Random Forest algorithm. Accuracy and Precision were better-quality with XGBoost compared to Random Forest, so we continue with XGBoost.[94]

The ratio between traded and not traded (*per* customer *per* week) is on average 1 to 9. To correct for imbalanced classes we employ a logistic loss evaluation metric. Log loss doesn't

---

[92] The standard routine for customer calls in a financial markets environment depends on their trading frequency. As the sweet spot for coverage optimisation is anyway the infrequent trading customers a weekly routine makes on average sense.

[93] Calculations and charts were performed as part of the bank project, see FN 90.

[94] The initial test results are such in line with the outcome in the prior section and with recent competitions (Kaggle, 2013).

easily support ‘high probability estimates’. ‘False’ predictions with ‘high probability’ are stressed exponentially, thus heavily penalised. This is true for FN (False Negatives) and FP (False Positives) in the Confusion Matrix. The exponential effect is so strong that ‘it does superimpose’. The relative rarity of one against the other loses relevance. As second corrective we introduce a penalty factor for FN predictions. From an economic perspective we mostly care about an instance where the model predicts that there will be ‘no trade’ but there is ‘a trade’ - or at least ‘an inquiry’ potentially leading to a trade. In this instance the bank misses a money making opportunity – the margin of the trade if executed. Less relevant economically is a False Positive where the prediction is that there will be ‘a trade’ but ‘no trade’ is taking place. Here, the missed opportunity costs are that of a telephone call or missing out on another trade the bank would otherwise have made (low probability situation). In the first instance (FN) we multiply the log loss function by factor 3, in the second (FP) by 1.

First, we just predict ‘whether a customer will trade in a week’. Thus, we need to assign binary labels $\hat{y}_w$ ($\hat{y}_w \in \{0,1\}$) to customers. As our model provides probabilities for its classifications we need to transform the outcomes. Customers with a probability of trading in a week [> threshold %] are therefore labelled ‘1’ and those with [< threshold %] are labelled ‘0’. To benchmark the model we define a simple “Null1” hypothesis. Historically, most customers do not trade on a weekly basis[95]. Thus, we use the assumption ‘that a customer does not trade in a given week’ as Null1. To validate the model performance we calculate the Classification Accuracy for the model and Null1 (Base Rate). The Classification Accuracy is given as:

$$Clasification\ Accuracy = \frac{\sum_{i=1}^{N} 1_{[F(x)=y]}}{N}$$

for a given week w and total number of customers N.

[95] On average across the frequently and sporadically trading customer - the chance over time is 80% that a customer does not trade within a week.

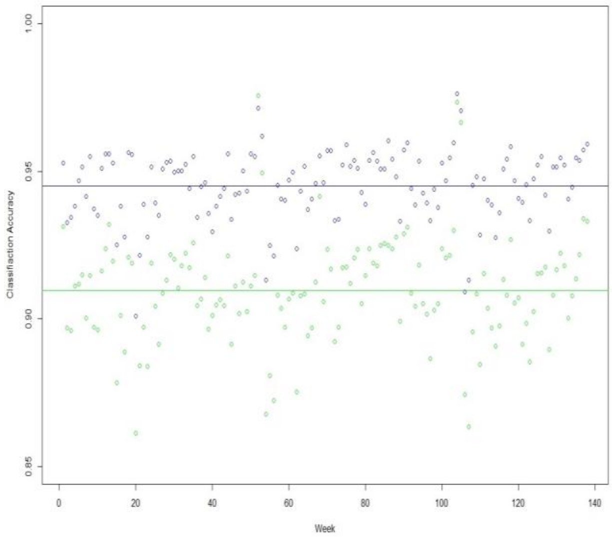


Figure 14: Comparison between Model Classification Accuracy (Blue) and Base Rate (Green)

Figure 14 illustrates the comparison between Model Accuracy and Base Rate (Null1). Obviously, the model predictions are superior. The outcome though has limited significance, as we just predict that infrequent traders will not trade while frequent traders will.[96]

Next, we calculate with our model 'the probability p_w of any customer executing in a given week' ($p_w \in \mathbb{R}\ [0,1]$). We limit this experiment on customers with high trading probability. Figure 15 exemplifies the activity level of customers ranked in the upper quantile. We show the 'actual proportion of trading activity' for customers grouped according to their relative activity. For instance, the top 5% active customers account for 48% of activity. Customers in the top 25 percentile account for 90% of activity. This is valuable output for any recommender system (whom to contact). Let's presume the bank defines its trade coverage target to be 85%. This means that only 20% of customers need to be actively covered.

[96] We could have done this similarly by basing our prediction on the historic mean 'of customers trading'. As such the model run counts just a 'plausibility test'.

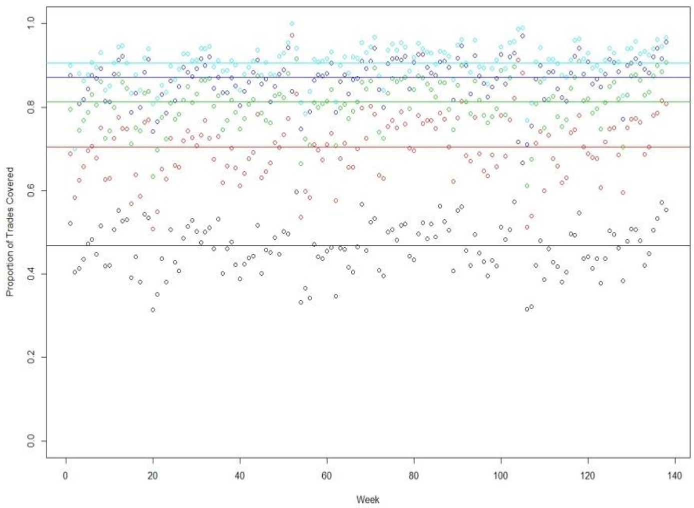


Figure 15: Proportion of Trades covered by customer ranking.
Top 5% (black), 10% (red), 15% (green), 20% (blue) and 25% Percentiles (turquoise).

To assess the model we relate the results to a less naive baseline. The “Null2” hypothesis presumes that customers are trading according to their trading pattern in the past. Therefore the probability of a customer trading in a given week according to the new Null is their historic mean, their average trading rate in the past.[97] The calculation of the historic mean at any point in time is done based on the test data using an ‘extending rolling window’.[98]

As our model gives trading probabilities, it allows us to fine-tune the probability cut-off for its classification.[99] We can choose the optimal ‘minimum probability level’ for predicting a trade. Below we show Accuracy, Re-call (the True Positive Rate) and Precision[100] by probability cut-off, with the model consistently outperforming the Null2:

[97] Average trading rate = (# weeks traded) / (# weeks absolute), starting with their first trade during the period.

[98] Hence, if a customer becomes less active during the test period his ‘trading mean’ will decrease over time. We adjust for changing behaviour over time. This effect is more pronounced at the beginning than at the end of the test period. In the rare case that a customer doesn’t exist anymore, changed the name identifier or merged with another entity we remove him from the test population when rolling forward.

[99] This would be not possible if the model provides just a [0, 1] classification.

[100] Accuracy $= \frac{TP+TN}{(TP+FN)+(TN+FP)}$, Recall $= \frac{TP}{TP+FN}$, and Precision $= \frac{TP}{TP+FP}$ (see section 2.2.9.).

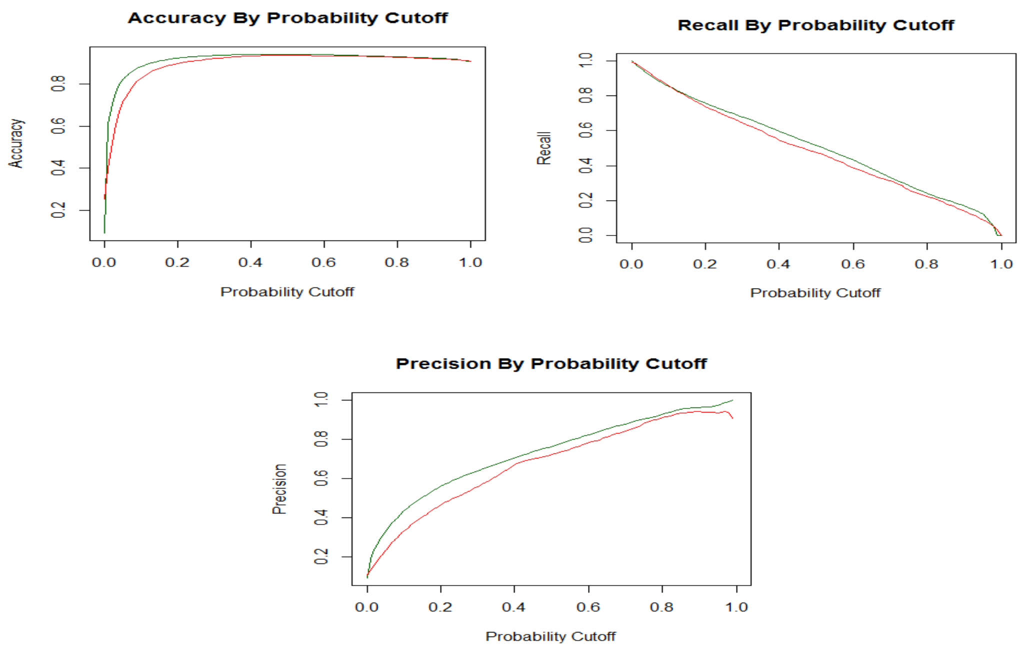


Figure 16a, 16b, 16c: Statistical Measures by Probability Cut-off.
Model (Green) vs. Null2 (Red).

We appraise the best probability cut-off point via grid search. For the optimisation we assume an equally distributed profit-to-cost ratio by trade of 3:1. In other words, for each successful prediction (TP) we earn '+3'. We need to consider the costs correspondingly. We assume costs of '-1' for each time a trade gets predicted (TP + FP).[101] Let's interpret these as costs for covering the customer. To identify the optimal cut-off probability we take the maximum 'Net Payoff' of the optimisation function shown in Figure 17.

[101] Each 'TP' trade results in our simplified optimisation as +2 (= 3 – 1). Each 'FP' trade produces -1 for our optimisation. Hence, the 'optimisation' assumes that all trades done by customers are either similar in size and profit margin, or 'equally distributed' across all instances.

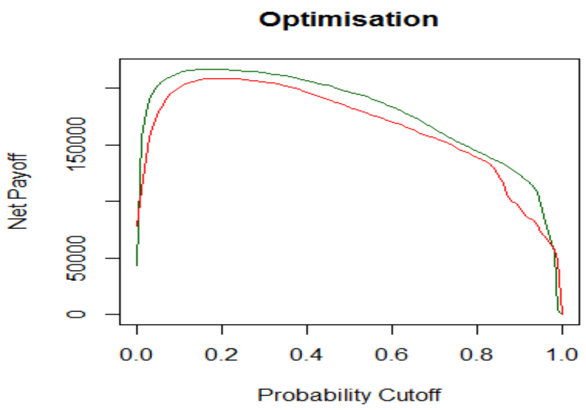
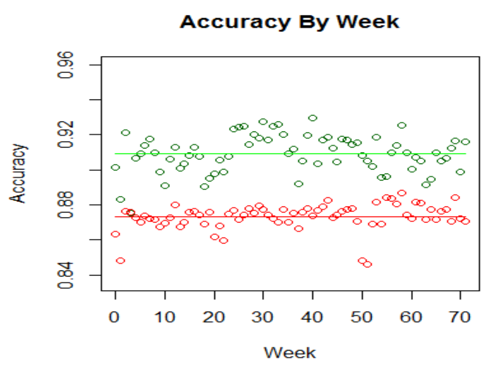
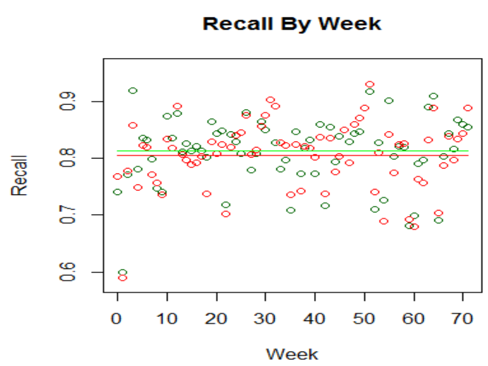


Figure 17: Net Payoff by Probability Cut-off.
Model (Green) vs. Null2 (Red).

The optimisation peaks at around 15% probability. Thus we categorise going forward all predictions with trading probability ≥15% as '+1', and with probability <15% as '0'. This gives us the following Accuracy, Re-call and Precision results. The model results (green) are shown in comparison to Null2 (red):





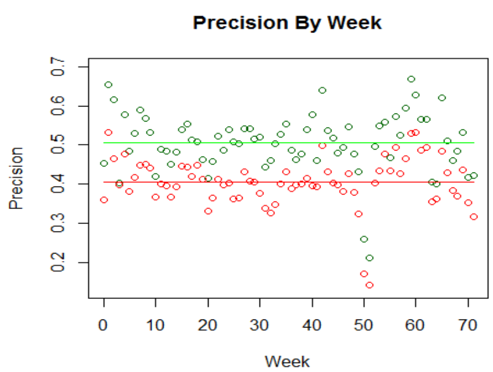


Figure 18a, 18b, 18c: Statistical Measures for 15% Prob. Cut-off.
Model (Green) vs. Null2 (Red).

In Table 11 we group the customers according to their trading frequency over the test period. Customers in Group [0-5], for example, traded in the 2015/16 period a maximum of 5 times p.a. (the least frequent traders), while customers in the [45-52] group traded nearly every week (the most frequent traders). The model improves over Null2 particularly for customers with low trading frequency. However, this is also true for the sweet spot from an economic perspective, the customers trading [10-25] times a year.

Table 11: Comparison of Prediction Statistics according to Customer Activity Ranking. Model (Green) vs. Null2 (Red)

GROUP	m_ACC	N2_ACC	delta_ACC	m_REC	N2_REC	delta_REC	m_PRE	N2_PRE	delta_PRE
0-5	97.85%	94.41%	3.64%	7.84%	8.08%	-2.92%	8.13%	2.09%	288.35%
5-10	67.62%	60.56%	11.67%	36.97%	37.67%	-1.87%	16.70%	13.61%	22.69%
10-15	49.17%	42.17%	16.61%	67.83%	64.57%	5.06%	26.51%	23.08%	14.84%
15-20	43.43%	39.44%	10.11%	85.59%	82.34%	3.95%	35.56%	33.45%	6.30%
20-25	47.02%	44.79%	4.96%	95.96%	93.45%	2.68%	45.03%	43.84%	2.72%
25-30	52.59%	52.17%	0.81%	97.68%	96.03%	1.72%	52.61%	52.43%	0.35%
30-35	63.17%	60.73%	4.02%	97.10%	93.28%	4.09%	63.50%	62.54%	1.55%
35-40	72.80%	72.00%	1.11%	99.90%	99.86%	0.05%	72.63%	72.07%	0.78%
40-45	81.67%	81.67%	0.00%	100.00%	100.00%	0.00%	81.67%	81.67%	0.00%
45-52	95.15%	95.11%	0.04%	99.98%	100.00%	-0.02%	95.16%	95.11%	0.06%

Finally, we focus on the question of 'how much a customer would buy if he traded'. We need to fit the model to the exact configuration of the customer's portfolio – each time a trade occurs. We postulate that the number of bonds traded by a customer follows the Poisson random variable $(N_k(t) - 1) \sim Poisson\left(\mu_k t_{k,w}\right)$. The model predicts whenever a customer is active he trades on average $(\mu_k t_{k,w} + 1)$ number of bonds. $\ln(\mu_k t_{k,w}+1)$ is the working response for an OLS regression with the predictor function (see equation below) reflecting the estimate. In Figure 19 we relate 'the number of bonds predicted to trade' against 'the actual number traded'[102].

[102] The analysis was based on 'buy & sell' trades and not just 'sell trades' due to time constraints.

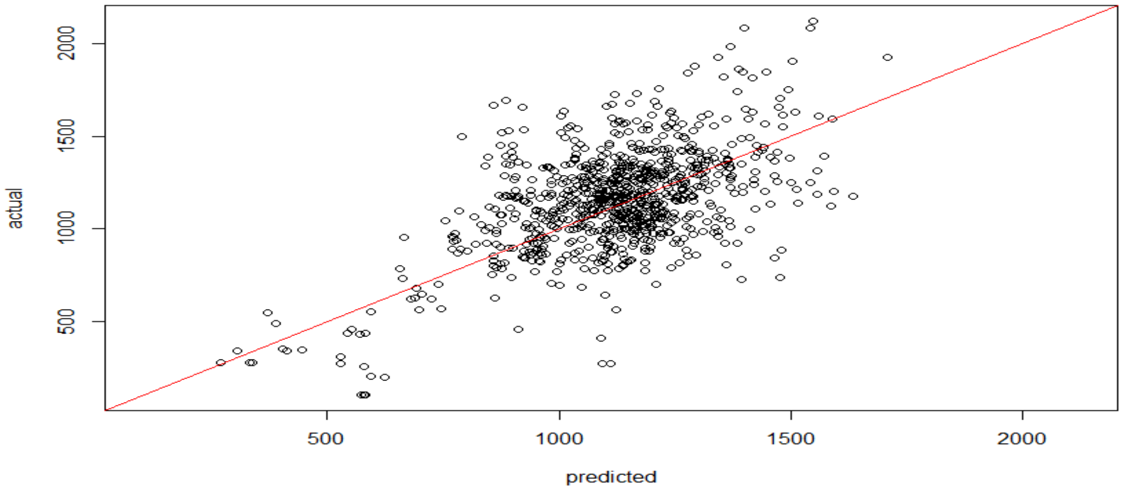


Figure 19: Comparison between the Predicted number of Trades and Actual number of trades in a given week.

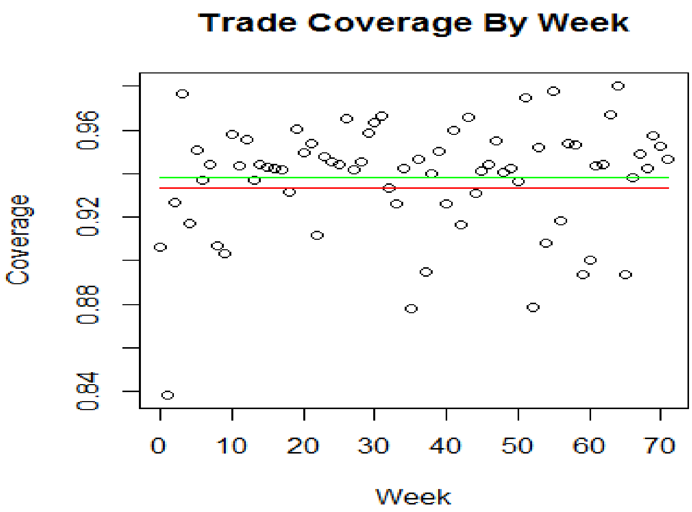


Figure 20: Comparing Trade Coverage results.
Model (Green) vs. Null2 (Red).

To model the exact portfolio configuration for each trade incident we differentiate the bonds according to issuer type and maturity profile into nine sector categories = {S1, S2, S3, F1, F2, F3, C1, C2, C3} where:

S = sovereign;	1 = short-term bonds (0-2 years to maturity);
F = financial;	2 = medium-term bonds (2-5 y-t-m);
C = corporate;	3 = long-term bonds (5+ y-t-m).

XGBoost evaluates the probability of classification ($\hat{q}_{S1}(w), \ldots\ldots \hat{q}_{C3}(w) \in [0,1]$) for each sector. The measure to quantify accuracy is a multi-class log loss function. We aggregate across all instances where a customer trades to the log $\hat{q}$ of the relevant category[103].

$$Accurancy = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} 1_{(y_i=j)} log(\,\hat{q}_j)$$

n is the number of instances a customer traded, m the number of sectors j and $1_{(y_i=j)}$ an indicator function returning 1 if the sector describes the traded bond and 0 otherwise.

"Null3" reflects past activity. We calculate across the training data 'the aggregate number of bonds traded per category' - expressed as the proportion of total bonds traded – and use the results as the 'predicted probability for a sector'. Figure 21 illustrates the improved accuracy over the Null3 hypothesis across the entire test data. We show in Figure 21 an optimisation score; thus lower is better. The model is again superior to simply aggregating over historic activity, as it gives a weighting to the feature set and by determining what a customer will trade.

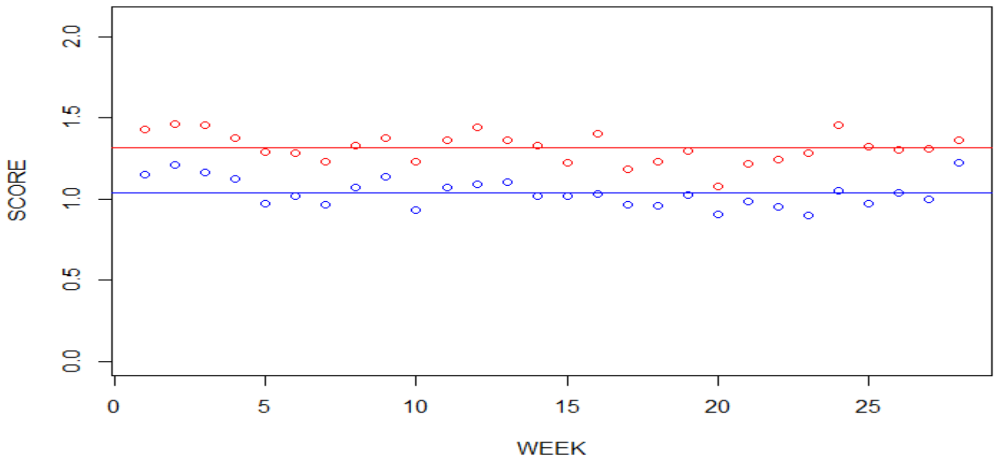


Figure 21: Comparison between Fitted model (Blue) and Null3 (Red).

[103] Trade data includes both buy and sell orders from the customer.

Obiter dictum:

i. Independent of sales units, the aim is to cover the sweet-spot of the customer clientele, those who can be swayed by coverage or by inventory.[104] The longer-term objective is to assess targeted changes, which affect behaviour; thus optimising for profitability and customer satisfaction.
ii. Data discipline has to be enforced and properly supported by integrated customer relationship management (CRM) systems. Ultimately, specific requirements, when working with touch-data, will dictate the structure of the experiment.
iii. The ensemble choice could become more relaxed if the justification can be removed why a model should be trusted.[105] Enhanced ensembles / model combinations could be tested, with XGBoost just as a sub-model within a highly specialised model ensemble.

2.4.6. Concluding Remarks Regarding the Results

When analysing institutional customer data, we achieved advanced insights into customer activity. We were able to forecast 'when' and 'what' customers are going to trade. The quality of the results was superior over several predefined Null hypotheses. Nevertheless, we expect that the results could be improved upon by broadening the underlying data. Additional information content could provide the learning algorithm with further signals. Currently, we rely almost entirely on transactional information from the proprietary data-base. We believe that if we were to include information on rates, credit, FX, equity *etc.*, we could link the predictions to overall market activity. Likewise, intelligence on customers' requirements, assets under management, the regulatory framework, 'wallet' size, investment strategy and so on could add value. Further information could be gained by touch data: records about human interaction, such as customer contact, invitations, attendance at work-shops, on-going services and requests. All this information needs to be supervised from a compliance and data protection perspective. Having more of this data should give enhanced results. Service and product offerings could be better targeted with a higher degree of automation. However, big amounts of data from various sources require strict data handling protocols and high process

[104] Some customers only trade when there are specific assets on the inventory list.

[105] We investigated at the start different base ensembles but we limited our project – for acceptance and traceability reasons - to the current 'standard' models in literature and praxis. XGBoost is the most powerful of those but not *per se* the optimal solution overall.

diligence. The good news is that XGBoost, due to its approximation mechanism, is computationally efficient and can enumerate more complex tree structures. Therefore, additional features should be manageable. By making the results transparent, an organisation is able to influence the behaviour of its staff. It can organise 'customer coverage' efficiently, and perhaps even more effectively. Improved understanding of customers and their needs will help to find better solutions. When in production, an institution has a tool-box for 'trial and error' testing. It can arrange placebo controlled social experiments[106]. Paired with a powerful CRM system, an organisation can analyse activities, responses, work and information flows. An institution can test with limited friction whether changes work. It can compare outcomes with previous expectations. With data analytics comes the understanding 'how to affect behavioural changes'.

2.5. Summary and Conclusion - the Use of Learning Algorithms in Finance

- ⇨ What did we do?

We started with an overview on learning algorithms. We explained in detail the theoretical foundation of the models and their requirements. We discussed the pros and cons of different models. In particular, we compared the performance of Random Forest and Gradient Boosting models. We described typical problems when applying the models, and solutions for overcoming these challenges. Parts of the paper can be taken as a practitioners guide on how to implement the technology. We reviewed data gathering, data preparation and selection of the model. We described, based on examples, how to calibrate model Hyper-parameters. We emphasised the importance of defining the problem, and following that, how to derive the appropriate dependent variable. We looked at how to monitor progress and how to measure success. A greater part of the paper was dedicated to how to create an enhanced feature set. According to our findings this is a very important step when applying learning algorithms. Our aim was to exhibit the theoretical and practical implications of Artificial Intelligence in finance - ultimately to demonstrate the substantial disruption potential to whole-sale banking. Hence, not just retail banking will have to change radically with the new

[106] How this should be done, see Kahneman (2011).

technological possibilities. The ‘proof of concept’ incorporated the actual implementation for various tasks. We experimented with different settings across distinct cases: trend prediction/portfolio optimisation and customer coverage. For trend prediction, we constructed features from various financial market theories. We tested the results and compared them against benchmarks based on statistical measures, but also performance-based measures. We extended the use of the market predictors to portfolio optimisation. We described how the predictor models of the previous section could be applied as pre-selection process to a traditional Markowitz portfolio optimisation. A sample calculation was provided. Finally, we employed the technology in a customer coverage context. We did it as a novel first on institutional data. We predicted when and what customers will trade. The thought behind this is that by analysing customer activity we could improve on customer coverage, therefore optimising daily business operations.

⇨ What did we find out?

We summarise the results from sections 2.3.8., 2.3.9., 2.4.5. and draw higher level conclusions[107]. We link the numbering to 2.3.8. for cross-reference purposes and continue thereafter. We found that:

1. Learning algorithms *add value*;
2. The models are *consistent* and *robust*;
3. The *results vary with different features*;
4. *Single* time series have *less information value* than *multi-asset* learning models;
5. Certain features have less information value than others;
6. *Volatility clustering* in market behaviour is evident and influences the model predictions;
7. Features based on *Behavioural finance* (as implemented in our setup) had less relevance;
8. We found *transformation mechanisms* between markets:
    a) Discrepancy between prediction statistics and investment performance for the HSI.
    b) Performance improvement from Single asset analysis to Multi asset analysis is highest for the NKY. For DAX and SPX the performance difference between Single asset and Multi asset is less;

[107] Some ‘conclusions’ are at this stage rather ‘speculative’; however the findings as such are consistent. They warrant in our opinion further investigation.

- ⇨ The existence of Spill-over effects between markets seems conclusive;
- ⇨ China is less connected with the world. China is not yet a fully open market economy; the results improved after 2010;
- ⇨ Japan is internationally intertwined with a strong macroeconomic interdependence;
- ⇨ Europe and USA are leading markets;

9. The outperformance through *time-overlap* is pronounced;
   - ⇨ Market performance 'overnight' has a major impact relative to 'intra-day';
10. Choosing the correct *investment strategy* is highly *significant*;
    - ⇨ It is not enough to know what the market will do. It is imperative to know how to operate under uncertainty;
11. Different learning algorithms produce different results;
    - ⇨ The choice of *model matters*;
12. *Complexity reduction* to reduce information overload / data noise improves the outcomes;
13. Reduction through *Granger varies across markets*. Granger pre-filter shows only 16 Leading Markets for SPX (out of 300), 99 Leading Markets for DAX and 214 Leading Markets for NKY;
    - ⇨ Dependence on Globalisation is highest for Japan, less so for Europe and very little for the USA;
    - ⇨ The USA still leads the world;
14. The portfolio section confirms in principle that Learning algorithms *add value* and that models are *robust*;
    - ⇨ Information gain through learning algorithms seems *consistent*;
15. The customer coverage section confirms, too, that Learning algorithms *add value* and that the models are *robust*;
    - ⇨ Information gain through learning algorithms seems *consistent*;
16. With regard to customer coverage we showed *superior results* compared to different Null hypotheses;
17. We used *different Nulls*, making them more sophisticated. Results were consistent over all changes to the Null.
18. Frequent traders have a consistently high percentage of trade activity across longer time periods; e.g. when we controlled for variance the results did not change much.
    - ⇨ *Customer behaviour* is fairly *constant* over time.

All our machine learning applications show that the models and their results are robust.[108] The results are steady. The evidence is consistent throughout. The 'dependability' is even such that we are able to identify hidden anomalies and data inconsistencies. If we conclude that the 'robustness and model consistency is genuine' it allows for postulating an interesting "Converse Practise". Let's call the underlying assumption "Genuine Model Robustness". So far we worked strictly transitive. We think about a 'mechanic' and implement the mechanic as feature into the modelling process. We do this to improve on the model. Let's assume now that we found the 'final model' and that the model is stable and robust. The predictor model at that stage has fitted the tree stage-wise to the underlying structural information in the data. As we took precautions not to over-fit, we can be confident that the model is at this stage a high quality statistical representation of the true relationship between the feature and the predictor results. Based on Genuine Model Robustness we are now able to endorse 'conversely' the mechanic. If we find that the model improves by adding a specific feature, all other things equal, we can conclude that evidence is found supporting the mechanic first-hand. We could use the model improvement as measure for the underlying relationship/mechanic. Features are anyway often ranked subject to their influence on the model. So-called 'importance scores' can be applied to each feature. The importance scores are usually used for identifying redundant features (Huynh-Thu *et al.*, 2012). We postulate that based on Genuine Model Robustness feature importance scores together with 'the isolated improvement score in prediction quality[109]' could be used as a measure for statistical significance of the underlying mechanic as such. The argument of using predictor models in a 'Converse Practice' broadens the utilisation of learning algorithms for application in future academic field studies.[110] It illustrates a point we were making earlier that Predictive Analytics is a sub-part of Data Analytics in general.[111]

---

[108] This is meant *inter alia* in the sense of consistency but also being 'reproducible', all things equal.

[109] The score could be based on i.e. ACC, PREC or MCC, combined with Chi-squared statistics.

[110] In consequence we could generalise some of our conclusions above. I.e. The results show that there is 'structural information' in markets and that 'timeliness' of information matters. These findings support the idea of Spill-over effects and that it takes time for markets to interpret information and adjust pricing. This in itself is a finding of macroeconomic relevance.

[111] Further research is needed to substantiate the underlying Genuine Model Robustness assumption. This could be done, for example, by cross-checking results with other empiric studies on a representative scale.

⇨ What is our ultimate goal?

The ultimate goal is to use these models in real life within a control-group verified experimental framework, analysing day-to-day operations. For example, we could monitor changes in feature effects (i.e. predictor statistics and/or importance scores) and thus the underlying mechanics on-going on a larger scale with hundreds of features.[112] Depending on data it can be used in the context of risk and credit management, fraud detection, identification of data corruption, asset management, asset-liability management and customer coverage; eventually even in the context of macroeconomic research studies. The areas of application are manifold. *Fail quickly and often.*

[112] For instance, we could use tick-data with constant recalibration across many markets - technically possible already.

3. Bail-In and Asset Encumbrance: Implications for Banks' Asset Liability Management[113]

3.1. Introduction

The recent financial crisis laid bare the weakness of banks' funding strategies which are reliant upon short-term wholesale funding. During the 2007/08 crisis counterparty risk among financial institutions rose dramatically resulting in a virtual standstill of interbank lending. To avoid a liquidity crisis, governments stepped in assuring bank depositors and creditors by guarantees on interbank lending (Demirgüç-Kunt and Huizinga, 2010). In the aftermath investors behavior changed and the demand for secured products such as covered bonds increased. In the US, the Dodd-Frank Act was put into legislation. A critical reflection of this statute and the 'too big to fail' issue is given by Karmel (2014). He recommends employing size standards and activity restrictions for big banks. Also in the EU, new regulatory requirements, e.g. the Basel III were proposed and put into legislation. Due to these new regulations, secured funding sources have become increasingly important for both banks and investors (Houben *et al.*, 2013). This resulted in a higher demand for high quality assets that can be used as collateral for secured funding instruments. The new bail-in resolution framework aims to pre-empt bank crises and resolve any financial institution insolvency in an orderly manner (EU Council, 2013). As a consequence, certain bail-in liabilities like senior unsecured and subordinated bonds will replace the public subsidy and the risk for tax payers to pay for the failure of banks will be minimised.

Instruments commonly used for the funding of a bank are equity, subordinated capital, the interbank market, retail deposits and market funding. They are characterised by their maturity, priority in the case of bankruptcy and collateralisation. Maturity is typically split into short-term liabilities maturing within one year, e.g. repos, and longer-term liabilities, e.g. covered bonds and senior unsecured funding. Unsecured elements are not backed by a pool of cover assets and therefore, in the event of insolvency of the issuing bank, their creditors receive payments from the insolvency estate according to their creditor status. A senior unsecured investor has priority over subordinated bondholders. Secured elements are backed by a certain pool of assets, referred to as collateralisation. By pledging or encumbering assets

[113] This chapter is published in Journal of Banking Regulation, April 2017, Volume 18, Issue 2, pp. 149 – 162, see Erhardt, Luebbers, Posch (2017).

as collateral to secured creditors they become unavailable for meeting claims of unsecured creditors in the case of insolvency. Hence, the higher the level of secured funding or the higher the level of asset encumbrance the greater the subordination risk for senior unsecured bond holders, and by that the lower their recovery values in the case of default of the issuing bank (Houben *et al.*, 2013). On the other hand, assuming secured funding to be cheaper than unsecured funding, a higher amount of secured funding lowers the overall costs of funding of the bank.

In this chapter we propose an answer to the question of how higher levels of asset encumbrance affects the funding policy of banks. The trade-off between optimality for banks and senior unsecured investors will be explored using a sample of major European banks. We investigate funding strategies of European banks and estimate their overall level of funding costs at their current level of asset encumbrance. The main goal is to figure out if it would be possible for those banks to reduce their overall level of funding costs by increasing the amount of secured funding. Looking at the effects on senior unsecured investors we estimate the optimal level of asset encumbrance with respect to their recovery rates. On the one hand, theory already indicates different funding strategies looking at specific funding instruments in greater detail. For instance, according to Calomiris (1999), protection and guarantee schemes of bank debts enable banks to take higher risks. To overcome this issue banks should hold a minimal portion of unprotected subordinated debt, which could be used to cover parts of possible insolvency costs. If banks take excessive risk or manage their assets poorly they will not be able to sell their subordinated debt at all or just at an increased cost level. As a consequence, subordinated debt holders might act as a monitor of the bank. In the new bail-in framework this effect extends to senior unsecured and non-deposit funding. We will thus model the effect of decreasing recovery values of unsecured bond investors on the overall level of banks' funding costs.

Houben *et al.* (2013) give a detailed analysis of current levels of asset encumbrance among 60 European credit institutions and illustrate possible implications for policy and the financial system. Based on their estimates, the average amount of secured funding plus deposits (if 100% of deposits is excluded from bail-in) among those European banks is around 69.5% of funded assets. Because of higher counterparty risk amongst investors and regulatory changes there even is an increasing demand for collateralised funding, particularly in Europe. Implications for the markets and the financial stability are shortage of high quality assets usable as coverage funds and lower recovery values for senior unsecured investors in the case of insolvency of the bank. From a political point of view, it is important to increase transparency about the extent to which bank assets are encumbered. Also, to restrict the risks of rising asset encumbrance, prudent limits might constrain the growth of secured funding. Hence, the overall question remains what an optimal level of asset encumbrance from the perspective of a bank itself, a senior unsecured investor or the public sector in general might be. We introduce an approach on how such a limit under the new regulatory environment can be estimated and show how banks can optimize their funding strategies. We find that for all banks in our sample the amount of secured funding can be increased and the overall amount of funding costs decreased. The optimal level of asset encumbrance for a senior unsecured investor is on a lower level than for the corresponding bank. The remainder of this paper is organized as follows: In section 3.2. we describe the data and the distribution of current funding strategies of our bank-sample. Our simulation setup is laid out in section 3.3., while section 3.4. applies this framework to current balance sheet data of our bank sample and it is here where we derive optimal levels of asset encumbrance as related to the overall level of funding costs and quantify the effect of the new regulation on banks and investors. The final section in chapter 3 summarizes our conclusions.

3.2. Data Description

Our simulation is based on a cross section of balance sheet data from major European banks. We obtain data from Bloomberg and directly from the banks' annual reports. As reference point we chose 2013's full year results. Our sample consists of the 17 largest banks from Bloomberg's major bank index covering eight European countries, cf Table 12 for an overview. According to the European Central Bank report on the comprehensive assessment from October 2014 for the preparation of the assuming banking supervision tasks in November 2014, none of the banks in our sample was among the 25 banks which showed capital shortfalls or attracted negative attention (ECB, 2014).

The liability side of each bank divides into traditional sources of bank funding like equity, deposits, repos, interbank funding, long-term debt securities like covered bonds and senior unsecured funding, subordinated capital and other long-term liabilities, like reserves for insurance companies, trading securities and other borrowings. Since the total share of covered bonds and senior secured funding could not be obtained from the annual reports directly, we look at all outstanding bonds of each bank. Whereas elements like financial liabilities measured at fair value, liabilities for tax and provisions or derivatives are not included in our simulation. Neglecting the role of derivatives has likely the relative largest impact on the costs for bail-in-able debt. If derivatives would be considered as bail-in-able debt this might have several effects. Firstly, to incorporate the possibility of being bailed-in, it might lead to price distortions of existing derivative contracts. Secondly, once derivatives would have been bailed-in, meaning compulsory dissolutions of existing contracts, banks have to bear extra costs when entering into new contracts in the course of restoration. As there are no regulatory standards on how to treat derivatives in the case of a bail-in yet in place, we cannot estimate the effect of excluding derivatives on senior unsecured bondholders separately. However, our general framework allows for such an extension once the regulatory framework is setup. In the case of insolvency, we assume these to be financed with the corresponding parts on the asset side. Table 12 shows the amount of different funding sources as percentage of funded assets of our banking sample.

Table 12 shows that deposits in general play an important role in all funding strategies considered ranging from 31% for Societe Generale up to 63% for Standard Chartered. On average, the deposit funding is around 45% of funded assets. Considering the bail-in mechanism, the equity and the subordinated capital vary for different banks. For example Svenska Handelsbanken’s equity and subordinated debt components comprise only 5.4% of funded assets. In contrast, Royal Bank of Scotland’s funding is based on 11.7% and the average is around 7.8%. At around 8%, the average interbank funding is only slightly above the equity and subordinated capital. Furthermore, also debt securities like senior unsecured and covered bonds seem to be highly used funding sources especially for banks in northern Europe like Nordea Bank (33%), Svenska Handelsbanken (48%), Swedbank (42%).

Table 12: Funding composition of the bank sample.
This is according to balance sheet data at 31 December 2013 as percentage share of funded assets.

Bank	Total Equity	Subord Capital	Interbank	Sec Sold w/ Repo	Debt Securities	Deposits	Other LT Liab
BNP Paribas	6%	1%	6%	1%	12%	37%	36%
Soc. Gen	5%	1%	9%	2%	13%	31%	39%
Credit Agri	4%	2%	12%	1%	11%	37%	33%
Barclays PLC	6%	2%	6%	20%	9%	43%	13%
Lloyds Bank	6%	4%	2%	0%	11%	56%	21%
Royal B of Sc	8%	3%	5%	12%	10%	58%	3%
Stan Chart	8%	3%	7%	0%	11%	63%	8%
Dt. Bank	5%	1%	6%	1%	13%	50%	24%
Commerzbank	6%	3%	17%	11%	14%	49%	1%
Intesa Sanp	7%	0%	9%	3%	24%	33%	24%
Unicredit SPA	7%	0%	14%	9%	21%	46%	3%
ING Bank	5%	2%	3%	0%	12%	46%	32%
Santander	8%	2%	7%	4%	17%	56%	6%
Nordea Bank	5%	1%	11%	2%	33%	34%	14%
Svenska Han.	5%	1%	7%	0%	48%	34%	4%
Swedbank AB	6%	1%	7%	1%	42%	35%	8%
UBS AG	7%	0%	2%	2%	12%	57%	19%
Average	**6%**	**2%**	**8%**	**4%**	**19%**	**45%**	**16%**

To estimate the effect of secured funding on senior unsecured investors, a prudent measure of asset encumbrance needs to be defined. According to Houben *et al.* (2013), there are three possible calculation methods. First, one could calculate the ratio of unencumbered assets to unsecured liabilities. This would indicate the amount of assets available for unsecured investors in the case of bank failure. The two other measures which tend to be more accepted by market participants and easier to apply are the liability-side and the asset-side approaches. The liability-side approach is defined as the proportion of secured borrowing in banks' liabilities. Since this method considers neither the amount of over-collateralisation nor the derivative-related liabilities, it will undervalue the level of asset encumbrance. But if repo-based funding plays an important role in the bank's funding strategy, this ratio might even overestimate the level of asset encumbrance if offsetting reverse repo transactions are not taken into account. The third approach is based on the proportion of pledged balance sheet assets. The main problem of this approach is the lack of comparable data available on pledged assets.

In order to measure the level of asset encumbrance we use a combination of the measures. To estimate the share of longer term secured funding, we take data of all outstanding secured bonds and the amount of repo-based funding from each bank. This enables us to cover the liability-side approach. To cover the asset-side approach, we apply a 2% haircut (cf. Enthofer, 2013) on repos and an average amount of over-collateralisation for covered bonds of 13% based on Moody's European Covered Bonds Monitoring Overview (Moody’s, 2012), see Table 13. Note that there is no data available for derivative transactions, margin calls and the net repo effect.

Table 13: Simple average over-collateralisation levels for covered bonds. This is consistent with Aaa when issuer rated A2 by country in 2012, based on Moody’s European Covered Bonds Monitoring Overview (Moody’s, 2012).

Average over-collateralisation levels for different countries	
Austria	21%
Denmark	10%
Finland	9%
France	12%
Germany	12%
Netherlands	15%
Norway	10%
Sweden	12%
United Kingdom	19%

In the case of insolvency, assets which are not used to redeem the claims of covered bondholders go back to the insolvency estate of the bank. However, since the amount of assets going back to the insolvency estate is unknown a priori, we assume the whole amount of over-collateralisation not to be available for senior unsecured investors in the event of insolvency. Hence, the asset encumbrance ratio (AE ratio) is calculated as

$$AE\ ratio_i = \frac{Encumbered\ Assets_i}{Funded\ Assets_i}$$

for each bank i. The results for our sample are shown in Table 14, where the average level of asset encumbrance is around 15%. The bank with the highest level is Swedbank with a proportion of 39.3% of funded assets. Whereas Standard Chartered issues the smallest amount of covered funding at around 0.3%.

Table 14: The level of asset encumbrance (AE) and deposits.
This is calculated for each bank as percentage of funded assets.

	AE	**Deposits**
BNP Paribas	5.1	37
Soc. Generale SA	7.6	31
Credit Agricole	3.3	37
Barclays PLC	23.7	43
Lloyds Bank PLC	7.1	56
Royal Bank of Sc.	14.0	58
Standard Chartered	0.3	63
Deutsche Bank	5.8	50
Commerzbank	21.0	49
Intesa Sanpaolo	8.4	33
Unicredit SPA	16.7	46
ING Bank	5.6	46
Banco Santander	15.0	56
Nordea Bank	28.3	34
Svenska Han.	35.1	34
Swedbank AB	39.3	35
UBS AG	5.4	57

To consider the depositor preference against senior unsecured funding Houben *et al.* (2013) assume that all retail deposits have a preferred status. A critical consideration of the depositor guarantee scheme is given by Ayadi and Lastra (2010) and Kleftouri (2015) who examine the effectiveness and design of deposit protection systems. Generally, deposits do not encumber assets as they are not secured by collateral. But due to the bail-in mechanism, depositor preference schemes change the priority of the liability side of a bank (EU Council, 2013). This in turn leads to a subordination risk for unsecured investors. Assuming all deposits to be preferred over unsecured investors, we have to add the amount of assets which are encumbered and the corresponding amount of deposits which are also not available for unsecured investors in the case of bank insolvency. As we can see from Table 14, a combination of the level of asset encumbrance and the deposit funding is highest for Swedbank, Royal Bank of Scotland and Banco Santander. But considering the recapitalisation case of Bank of Cyprus in summer 2013, large customer deposits were also converted into equity. To mimic such events we vary the share of deposits excluded from the bail-in for each bank.

To estimate and optimise the overall level of funding costs in relation to the level of asset encumbrance, we need to estimate various funding curves for each bank. Looking at different creditor status we estimate funding curves for subordinated, senior unsecured and covered funding. Using those curves we can calculate an overall or blended funding curve (WACF) as described in the following section. Hull *et al.* (2004) propose the CDS par spread to be equal to the simple bond Z spread under simplifying assumptions. In pre-crisis times, this relationship more or less holds as demonstrated by Bai and Collin-Dufresne (2011). But after the financial crisis, mainly due to liquidity premia in both the bond and CDS market, the CDS basis (CDS spread - bond credit spread) widened (Kenyon and Stamm, 2012). To reflect these changes we obtain simple bond Z spreads for each kind of funding and each bank from Bloomberg as of December 31, 2013. However, to get reasonable estimations of different spreads we only focus on investment grade bullet and soft bullet bonds, denominated in Euro with fixed coupon payments and a nominal greater or equal to EUR 500 million. To estimate missing values we compare banks with similar senior unsecured Z spreads like combinations of Standard Chartered, UBS, ING Bank, Nordea, Svenska and Swedbank. Since we could not obtain prudent data for the one year covered and one year subordinated bond spreads we look at the relation of the one to one-to-five year senior unsecured spread and apply these relations to the respective one-to-five year spread of the covered and subordinated bond spreads. Table 15 shows the results of our funding curve approximation.

Table 15: Approximated funding curves based on simple Z spreads.
This is for senior unsecured, covered and subordinated funding instruments
for different maturity buckets as of December 31, 2013.

Bank	**Senior Spreads [Bps]**			**Covered Spreads [Bps]**			**Subord Spreads [Bps]**		
	< 1 y	**1-5 y**	**> 5 y**	**< 1 y**	**1-5 y**	**> 5 y**	**< 1 y**	**1-5 y**	**> 5 y**
BNP Paribas	32.3	42.5	64.0	-5.5	-3.1	46.5	131.1	172.5	190.8
Soc. Generale SA	22.7	47.7	65.5	1.5	3.3	24.0	73.6	154.7	190.3
Credit Agricole	55.6	100.1	132.2	-0.8	-0.5	25.5	127.8	230.1	236.0
Barclays PLC	43.5	61.7	77.6	-2.1	-1.3	9.4	114.1	162.0	218.0
Llyods Bank PLC	33.0	43.4	110.9	-1.3	-0.8	36.3	90.2	118.8	210.0
Royal Bank of Sc.	59.4	84.1	184.2	0.1	0.1	43.4	127.1	179.9	212.1
Standard Chartered	40.8	53.8	77.6	12.0	6.8	32.9	99.1	130.4	202.1
Deutsche Bank	23.4	29.0	45.7	5.3	6.3	23.4	101.3	125.7	128.0
Commerzbank	70.5	92.7	116.9	14.7	19.4	70.0	260.6	343.1	387.5
Intesa Sanpaolo	108.1	134.3	198.6	49.3	61.2	99.8	224.5	278.9	296.7
Unicredit SPA	147.7	194.5	323.6	43.8	57.7	98.2	167.2	220.1	394.4
ING Bank	37.2	49.0	81.4	4.7	6.2	34.5	90.3	118.8	199.6
Banco Santander	143.5	188.9	281.8	47.2	62.1	92.6	184.9	243.3	286.9
Nordea Bank	29.2	38.4	69.1	-7.5	-4.2	17.0	63.8	84.0	138.8
Svenska Han.	24.4	32.2	56.9	-2.7	-3.5	14.0	53.4	70.3	114.5
Swedbank AB	35.9	31.6	91.5	-0.3	-0.1	51.1	78.6	69.2	151.6
UBS AG	31.9	42.0	81.5	-7.2	-4.1	8.7	69.8	91.8	135.1
Average	**55.2**	**74.5**	**121.1**	**8.9**	**12.1**	**42.8**	**121.0**	**164.3**	**217.2**

Looking at the one-to-five year senior unsecured bond spread, we can see the lowest spread level for Deutsche Bank, Swedbank, Svenska Handelsbanken and Nordea Bank with 29, 31.6, 32.2, and 38.4 basis points respectively. Comparing Table 14 and Table 15 we cannot see a clear link between a high level of asset encumbrance and deposits and higher senior unsecured spreads. For example Svenska Handelsbanken, which has the highest combined level of asset encumbrance and deposits, indicates one of the lowest senior unsecured spreads. Whereas Banco Santander with the second highest level of asset encumbrance and deposits shows one of the highest senior unsecured spreads. On the other hand, Intesa

Sanpaolo and Credit Agricole show very low levels of asset encumbrance and deposit compared to the rest of our sample but are combined with some of the highest one-to-five year senior unsecured spreads.

3.3. Simulation Setup

The bail-in framework favours a creditor funded recapitalisation. In the event of a bank failure, certain bail-in liabilities replace the public subsidy therefore reducing the risk of taxpayers paying for the failure of the bank. The order of bailing-in those liabilities follows the order of their ranking in national insolvency laws, see for example Figure 22.

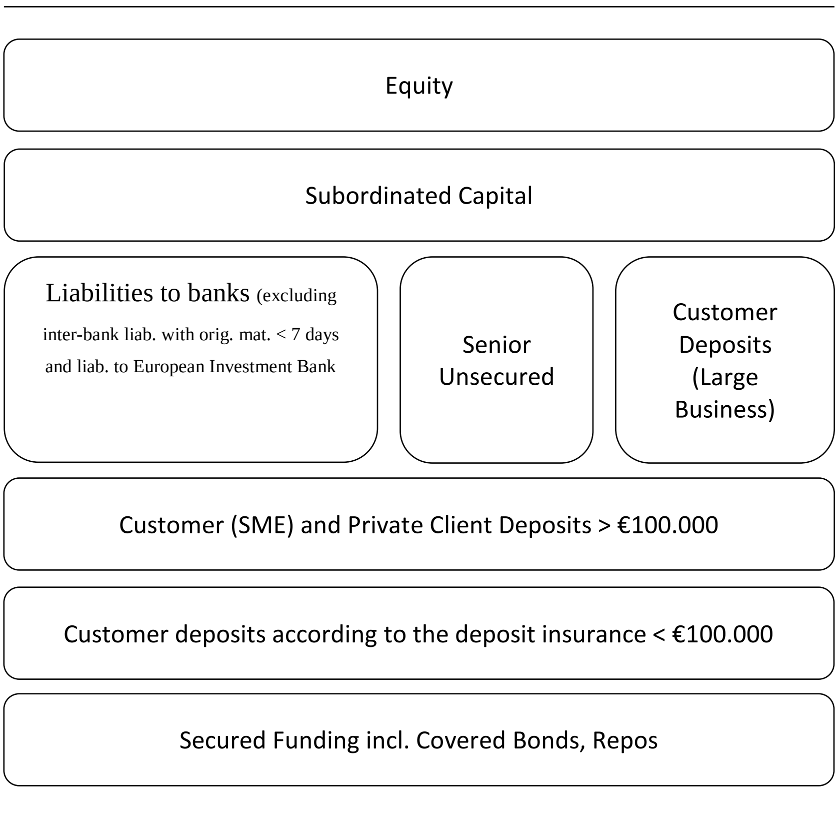


Figure 22: Ranking of liabilities according to bail-in rules (see EU Council, 2013).

Excluded liabilities are covered deposits, secured liabilities including covered bonds, liabilities to employees of failing institutions, such as fixed salary and pension benefits, commercial claims relating to goods and services critical for the daily functioning of the institution, liabilities arising from a participation in payment systems which have a remaining maturity of less than seven days and inter-bank liabilities with an original maturity of less than seven days (EU Council, 2013). We use the following prescribed order in which different bank creditors are subject to a bail-in. In different asset loss scenarios, first the equity of the bank is used to cover those losses. If this is not sufficient, the bail-in mechanism is activated and after the subordinated bondholders also senior unsecured investors and large deposits have to bear the costs.

Based on a Modigliani-Miller world, the composition of corporate financing - and by that a switching between secured and unsecured financing - would have no substantive effects on the costs of capital, unless the underlying market unveils imperfections. The depositor preference scheme as proposed by the bail-in framework might be one example for such imperfection. In fact, even the deposits are one kind of unsecured funding; most of them typically do not react to a variation of secured funding (Houben *et al.*, 2013). In order to derive benefits from this effect, minimising the overall costs of funding by increasing the amount of secured funding products becomes the objective of the bank. Hence, to set the funding costs relative to the level of asset encumbrance, we calculate an overall weighted average cost of funding (WACF), also referred to as a blended funds transfer pricing-rate (Choudhry, 2012). This approach is often used in the treasury department of banks, when they use an average cost of funding applied to all deals (Pallavicini, 2011). The calculation of the WACF is based on the funding spreads given in Table 15. It is calculated as the discount rate which equates the present value of future values of each loan to the original amount borrowed. In order to calculate the overall costs of funding; we divide liabilities into three maturity buckets. We assume demand deposits and 50% of saving and time deposits to be in the first bucket, while the rest is in the second maturity bucket. Equity, repos and short-term borrowings are short-term liabilities, while the subordinated capital and the long-term debt are allocated along the maturity profile according to the maturity distribution of all outstanding bonds of each bank. Following Choudhry (2012), not more than 20% of the inter-bank liabilities should have an original maturity of less than seven days. They are therefore excluded from a bail-in as are all secured funding sources and a variable amount of all deposits, reflecting customer deposits of less than Euro 100.000, customer (SME) and private client deposits.

As long as the recovery rate for the senior unsecured investor is 100%, its spread is equal to the unsecured spread obtained for each bank. Since the senior unsecured investor prefers additional secured funding only up to a certain amount, we estimate what happens as soon as his recovery rate decreases as follows:

We consider the relationship between the recovery rate and the CDS spread by using the annualised probability of default (PD)[114]:

$$PD = \frac{(annualized\ CDS\ spread)}{1 - recovery\ rate}$$

In a worst case scenario for a senior unsecured investor his recovery rate goes to zero in the case of the bank's default. In this case, the probability of default for both, the subordinated bondholder and the senior unsecured investor are the same (Glionna *et al.*, 2012). Based on the PD formula above, this leads to:

$$\frac{(annualized\ subordinated\ CDS\ spread)}{1 - recovery\ rate} = \frac{(annualized\ unsecured\ CDS\ spread)}{1 - recovery\ rate}$$

On this account, and using the assumption that both recovery rates are zero, the subordinated CDS spread is a reasonable upper bound for the senior unsecured CDS spread.

We apply this relation and take the estimated subordinated bond spread from Table 15 as an upper value for the senior unsecured funding curve. As a last step, the development of the senior spread in-between those two bounds needs to be estimated. Assuming a compression between the spreads payable for senior and subordinated bonds, the only question that remains is how fast they both converge to each other, assuming a decreasing recovery rate for senior unsecured investors. First, an increasing amount of secured funding and therefore a higher level of asset encumbrance leads to a lower recovery rate and therefore a higher loss given default (LGD) at some point ($LGD = 1 - recovery\ rate$). Since secured funding instruments require higher quality assets in their cover pool, the amount of assets available for senior unsecured investors is decreased, and the quality of the remaining assets is worse than

[114] See Grossmann (2011).

the assets used for the coverage funds of the covered bonds. To incorporate the effect of having poorer quality assets, we use a square root relationship between the LGD and the corresponding spread for senior unsecured investors as an upper evaluation level:

$$senior\ spread_{LGD} = (sub\ spread_{Dec13} - senior\ spread_{Dec13}) \times \sqrt{LGD} + senior\ spread_{Dec13}$$

The $senior\ spread_{Dec13}$ and $sub\ spread_{Dec13}$ refer to approximated funding curves of each bank as of December 31, 2013, see Table 15.

Our simulation setup proceeds as follows. First, we calculate the costs for different asset loss scenarios as a percentage share of assets multiplied by the amount of funded assets. Secondly, we increase this amount by the amount of over-collateralisation for secured funding, as described in section 3.2.. We then compare this sum with the available equity and subordinated capital of each bank. If these are not sufficient, we reduce the recovery values for senior unsecured investors to cover the additional losses. Finally, we estimate the effect of the reduction in recovery values on the overall funding costs, as explained above. To find the optimal level of asset encumbrance and by that find the optimal level of secured funding for each bank we base each simulation on the current funding strategy of each bank. We simulate low levels of asset encumbrance and increase this level using additional covered bonds. To change the level of asset encumbrance, we replace senior unsecured funding by secured with matching maturity. With a view to the bail-in framework, we consider senior bonds, deposits from large entities and also interbank funding products with an original maturity of more than seven days as senior unsecured funding instruments.

To include the post global financial crisis changes regarding the explicit requirement to have adequate bail-in-able debt, we incorporate a total loss-absorbing capacity (TLAC) ratio for the banks considered. According to the Financial Stability Board (2014), we assume the banks to hold at least 20% of its risk weighted assets (RWA) in form of regulatory capital and long-term unsecured debt. Since we focus on the optimal funding strategy and do not explicitly model the equity of the bank, we assume the equity to be fixed. Hence, the compliance with the TLAC ratio constitutes an upper limit when replacing the senior unsecured by secured funding. For each level of asset encumbrance we calculate the WACF. This enables us to determine the optimal or cheapest level of funding from the bank's perspective. Consider a risk averse senior unsecured investor in the case of insolvency of the issuing bank, he would like to get back the whole amount of his investments. Thus, we assume the optimal level of secured funding from a risk-averse senior unsecured investor's perspective to be the maximal level of asset encumbrance that still corresponds to an expected recovery rate of 100%. This optimal level is lower than the optimal level for the bank's treasurer. For the treasurer it might be beneficial to further increase the amount of secured funding until the increasing cost factor of unsecured funding dominates the other cost advantages due to increasing secured funding products.

3.4. Effect of Asset Encumbrance on Costs

An increase in secured funding might not only be beneficial from a bank's perspective because of regulatory issues like the net stable funding ratio in the Basel III framework but also because of a high demand for secured products on the investor side. To examine the effect of the bail-in we model three different asset loss scenarios of 2%, 10% and 20% of funded assets. Looking at the recapitalisation of Bank of Cyprus and also based on our estimations it seems to be prudent only to exclude 80% of the deposits of each bank from the bail-in mechanism. The remaining 20% of deposits are therefore treated equally to the senior unsecured funding instruments.

The results of the three different asset loss scenarios can be seen in Table 16, Table 17, and Table 18. They show the percentage share of covered bonds to funded assets (Covered), the overall weighted average cost of funding (WACF) for each bank, and the recovery rate for a senior unsecured investor (RR) in the case of the corresponding asset loss scenario and bank's default. Furthermore, the theoretically possible savings (WACF savings) for each bank are shown. In each scenario, these values are calculated for the current funding composition of each bank, the optimal funding strategy for each bank and the optimal structure from the perspective of a senior unsecured investor.

Table 16: Shown are the percentage share of covered bonds (Covered) based on a 2% asset loos scenario. We are using an average covered bond over-collateralisation of 13%, the total weighted average cost of funding (WACF) as bond Z spread, and the recovery rate (RR) for a senior unsecured investor from the perspective of the current balance sheet of each bank, an optimal funding strategy of each bank and from the senior unsecured investor's perspective. It is based on a 2% asset loss scenario, a concave relation between the LGD of senior unsecured investors and their spreads, and on 80% deposits excluded from bail-in.

Bank	**Current Balance Sheet**			**Optimum Bank**				**Optimum Senior**		
	Covered	**WACF**	**RR**	**Covered**	**WACF**	**RR**	**WACF Savings**	**Covered**	**WACF**	**RR**
	[%]	**[Bps]**		**[%]**	**[Bps]**		**[%]**	**[%]**	**[Bps]**	
BNP Paribas	3.7	53.8	1	11.2	52.7	1	2.0	11.2	52.7	1
SOC Gen	4.5	53.6	1	12.0	52.4	1	2.0	12.0	52.4	1
Credit Agricole	2.1	113.9	1	7.6	111.5	1	2.0	7.6	111.5	1
Barclays PLC	2.7	58.3	1	7.8	56.8	1	3.0	7.8	56.8	1
Llyods Bank PLC	6.3	75.3	1	9.6	74.5	1	1.0	9.6	74.5	1
Royal Bank of Sc	1.7	100.3	1	6.7	97.2	1	3.0	6.7	97.2	1
Stan. Chartered	0.0	70.0	1	7.7	67.9	1	3.0	7.7	67.9	1
Deutsche Bank	3.8	35.7	1	11.8	34.9	1	2.0	11.8	34.9	1
Commerzbank	9.0	91.2	1	11.2	90.4	1	1.0	11.2	90.4	1
Intesa Sanpaolo	4.9	160.0	1	24.1	153.3	1	4.0	24.1	153.3	1
Unicredit SPA	6.4	193.2	1	20.4	178.9	1	7.0	20.4	178.9	1
ING Bank	4.9	63.6	1	12.0	62.4	1	2.0	12.0	62.4	1
Banco Santander	9.5	173.8	1	15.1	169.9	1	2.0	15.1	169.9	1
Nordea Bank	23.4	34.3	1	31.3	32.4	1	5.0	31.3	32.4	1
Svenska Handelsb	30.7	25.4	0.97	44.1	21.5	0.95	15.0	24.5	26.2	1
Swedbank AB	33.5	29.7	1	41.9	28.5	1	4.0	41.9	28.5	1
UBS AG	2.9	61.4	1	10.8	59.4	1	3.0	10.8	59.4	1
Average	**8.8**	**82.0**	**1**	**16.8**	**79.1**	**1**	**3.7**	**15.6**	**79.4**	**1**

As we can see from Table 16 in a 2% asset loss scenario banks with a higher amount of covered bonds show the lowest overall funding costs. But looking at Deutsche Bank, Societe Generale and BNP Paribas, also banks with relatively small portions of covered bonds and low levels of asset encumbrance (see Table 14) can fund themselves at a low level of around 35.7, 53.6 and 53.8 basis points indicating rather short-term funding strategies. On the other hand, for banks with a relatively high average funding level of around 100 to 193 basis points the share of covered bonds varies from 1.7% - 9.5% corresponding to a level of asset encumbrance between 3.3% - 16.7%. With the exception of Svenska Handelsbanken, senior unsecured investors would have recovery values of 100% based on a 2% asset loss scenario. Furthermore, based on this scenario, all banks sampled should increase their amount of secured funding to lower their overall costs. An average increase of covered bond issuance of 8% would lower the average cost of funding by around 3.7%. The greatest impact of an increased amount of covered bonds can be seen for Svenska Handelsbanken who might lower their WACF by 15% or 4 basis points, respectively. But in the case of Svenska Handelsbanken, the recovery values for senior unsecured investors would be further decreased. Hence, from their perspective, Svenska Handelsbanken should rather decrease their current amount of covered bonds by 6% which would lead to an increase of WACF of around 0.8 basis points.

Table 17 illustrates a 10% asset loss scenario. Except for Royal Bank of Scotland and Standard Chartered we can see reduced recovery rates for senior unsecured investors of 86% on average. The two other British banks Barclays and Lloyds Bank also show recovery values above the average of our bank sample. Besides the British, the two Italian banks Intesa Sanpaolo and Unicredit and the Swiss bank UBS show high recovery values for senior unsecured investors of 98%, 93%, and 98%. Whereas the Nordic banks Nordea Bank, Svenska Handelsbanken, and Swedbank which show the highest amount of covered funding also show the lowest recovery values for senior unsecured investors of only 74%, 70%, and 64%. All others show an average recovery rate of 84%. The average WACF for all banks is increased by 6 basis points compared to the WACF based on a 2% asset loss scenario. Nevertheless, it would be possible for all banks considered to further reduce their increased WACF by 5% on average. Again, the greatest reduction of funding costs would be possible for Svenska Handelsbanken (18% or 6 basis points). In order to reduce their WACF, the banks need to increase their level of covered bond funding by around 8% on average.

With the exception Royal Bank of Scotland and Standard Chartered this would further reduce the recovery rates for senior unsecured investors. To maximise their recovery rates the banks considered should issue only around 1% covered bonds of funded assets on average. Since senior unsecured funding is more expensive than secured, this in turn would increase the WACF from the banks' perspective on average by around 10 basis points.

Table 17: Shown are the percentage share of covered bonds (Covered) based on a 10% asset loss scenario. We are using an average covered bond over-collateralisation of 13%, the total weighted average cost of funding (WACF) as bond Z spread, and the recovery rate (RR) for a senior unsecured investor from the perspective of the current balance sheet of each bank, an optimal funding strategy of each bank and from the senior unsecured investor's perspective. It is based on a 10% asset loss scenario, a concave relation between the LGD of senior unsecured investors and their spreads, and on 80% deposits excluded from bail-in.

Bank	**Current Balance Sheet**			**Optimum Bank**				**Optimum Senior**		
	Covered	**WACF**	**RR**	**Covered**	**WACF**	**RR**	**WACF Savings**	**Covered**	**WACF**	**RR**
	[%]	**[Bps]**		**[%]**	**[Bps]**		**[%]**	**[%]**	**[Bps]**	
BNP Paribas	3.7	62.9	0.84	11.2	60.2	0.76	4.0	0.1	65.8	0.89
SOC Gen	4.5	61.6	0.79	12.0	59.0	0.68	4.0	0.2	68.9	0.86
Credit Agricole	2.1	124.4	0.83	7.6	120.7	0.78	3.0	0.1	130.1	0.85
Barclays	2.7	64.0	0.90	7.8	61.6	0.86	4.0	0.1	67.8	0.97
Llyods Bank	6.3	78.5	0.92	9.6	77.4	0.91	1.0	0.3	86.0	0.98
Royal Bank Sc	1.7	100.3	1	6.7	97.2	1	3.0	6.7	97.2	1
Stan Chartered	0.0	70.0	1	7.7	67.9	1	3.0	7.7	67.9	1
Deutsche Bank	3.8	44.5	0.81	11.8	41.9	0.70	6.0	0.2	47.0	0.86
Commerzbank	9.0	114.0	0.90	11.2	111.4	0.89	2.0	0.4	133.5	0.97
Intesa Sanp	4.9	165.6	0.98	24.1	157.2	0.95	5.0	0.2	167.1	1
Unicredit SPA	6.4	197.1	0.93	20.4	182.0	0.88	8.0	0.3	227.0	0.97
ING Bank	4.9	69.2	0.79	12.0	66.8	0.66	3.0	0.2	76.4	0.86
Santander	9.5	176.8	0.92	15.1	172.5	0.90	2.0	0.4	208.3	0.98
Nordea Bank	23.4	41.4	0.74	31.3	38.3	0.62	7.0	0.9	58.7	0.93
Svenska Hand	30.7	32.3	0.70	44.1	26.6	0.46	18.0	1.2	48.3	0.92
Swedbank AB	33.5	34.1	0.64	41.9	31.9	0.39	6.0	1.3	55.1	0.94
UBS AG	2.9	62.3	0.98	10.8	60.2	0.97	3.0	0.7	63.9	1
Average	**8.8**	**88.2**	**0.86**	**16.8**	**84.3**	**0.79**	**5.0**	**1.2**	**98.2**	**0.94**

The results of a 20% asset loss scenario are shown in Table 18. We can see an average recovery rate for senior unsecured investors of 44%. In this scenario even Royal Bank of Scotland, Standard Chartered, and Intesa Sanpaolo show values of only 62%, 70%, and 69%. But with only 16% we can see the lowest recovery rates for Swedbank. Due to the further decreased recovery values of senior unsecured investors we calculated an average increase of the WACF of 15 basis points to a 2% asset loss scenario.

Table 18: Shown are the percentage share of covered bonds (Covered) based on a 20% asset loss scenario. We are using an average covered bond over-collateralisation of 13%, the total weighted average cost of funding (WACF) as bond Z spread, and the recovery rate (RR) for a senior unsecured investor from the perspective of the current balance sheet of each bank, an optimal funding strategy of each bank and from the senior unsecured investor's perspective. It is based on a 20% asset loss scenario, a concave relation between the LGD of senior unsecured investors and their spreads, and on 80% deposits excluded from bail-in.

Bank	Current Balance Sheet			Optimum Bank				Optimum Senior			Government Intervention		
	Covered [%]	WACF [Bps]	RR	Covered [%]	WACF [Bps]	RR	WACF Savings [%]	Covered [%]	WACF [Bps]	RR	Covered [%]	WACF [Bps]	RR
BNP Par.	3.7	72.5	0.37	11.2	67.9	0.04	6.0	0.1	77.6	0.50			
SOC Gen	4.5	68.2	0.34	11.7	64.6	0.02	5.0	0.2	79.1	0.51	12.0	64.4	0
Cr Agric	2.1	133.0	0.44	7.6	128.5	0.29	3.0	0.1	140.3	0.51			
Barclays	2.7	72.4	0.38	7.8	69.0	0.16	5.0	0.1	77.8	0.73			
Llyods	6.3	85.2	0.33	9.6	83.4	0.16	2.0	0.3	100.0	0.54			
RBS	1.7	113.0	0.62	6.7	108.4	0.51	4.0	0.1	120.0	0.77			
Stan Cha	0.0	89.4	0.70	7.7	84.7	0.59	5.0	0.0	89.4	0.70			
Dt Bank	3.8	51.8	0.37	11.8	47.8	0.02	8.0	0.2	56.1	0.50			
Coba	9.0	140.5	0.54	11.2	137.7	0.50	2.0	0.4	178.9	0.75			
Intesa San	4.9	182.0	0.69	24.1	168.1	0.29	8.0	0.2	191.6	0.76			
Unicredit	6.4	203.0	0.63	20.4	186.3	0.35	8.0	0.3	235.1	0.77			
ING Bank	4.9	74.1	0.26	9.7	71.9	0.01	3.0	0.2	84.9	0.44	10.0	71.7	0
Santander	9.5	181.6	0.50	15.1	176.7	0.35	3.0	0.4	215.8	0.72			
Nordea	23.4	45.4	0.35	31.3	41.8	0.06	8.0	0.9	67.0	0.73			
Svenska	30.7	36.8	0.37	41.4	31.4	0.01	15.0	1.2	55.5	0.75	41.9	31.1	0
Swedbank	33.5	36.3	0.16	36.9	35.3	0.00	3.0	1.3	62.5	0.76	37.2	35.1	0
UBS	2.90	68.2	0.45	10.8	64.7	0.05	5.0	0.1	72.4	0.58			
Average	**8.8**	**97.2**	**0.44**	**16.2**	**92.2**	**0.20**	**5.0**	**0.4**	**112.0**	**0.65**	**25.3**	**50.6**	**0**

For Commerzbank, for example, we can see the largest increase of the WACF of around 49 basis points. Based on Table 15 this increase is due to the large spreads Commerzbank has to pay for subordinated bonds compared to the senior unsecured spread. Due to the relationship between the senior unsecured spread, their LGD and the subordinated spread as explained above, we can see a strong impact on the WACF. For banks with lower subordinated spreads and smaller differences between the senior unsecured and subordinated spread, the increase in WACF would be moderate. For Banco Santander, Swedbank, and UBS we see an increase of the WACF of only 8, 7, and 7 basis points, respectively. Furthermore, comparing the Nordic Banks, the small difference of an increased WACF between Nordea Bank, Svenska Handelsbanken (11 basis points), and Swedbank (7 basis points) might be due to the difference in the amount of senior unsecured funding, which is lowest for Swedbank with 21% of funded assets compared to 26% and 30% for Nordea and Svenska Handelsbanken respectively.

Besides these effects, it would still be possible for all banks to reduce their increased level of WACF by around 5 basis points on average by increasing their amount of covered bonds by around 7.4%. For Intesa Sanpaolo, Unicredit, and Deutsche Bank the cost advantage would even be around 14, 17, and 4 basis points or 8%, respectively, increasing their amount of covered bonds by 19%, 14% and 8%. At this optimal level from the banks' perspective the recovery rates for senior unsecured investors would only be 20% on average. We can see the lowest values for Swedbank (0%), ING Bank, Svenska Handelsbanken with 1%, Societe Generale (2%), BNP Paribas (4%), and Nordea Bank with 6%. Hence, those would also be the first banks, the government or the covered bond investors would have to step in to cover additional losses of the bank. As also indicated in the 10% asset loss scenario, senior unsecured investors would prefer an average share of covered bonds of only 0.4% of funded assets to increase their recovery values.

To evaluate the robustness of our results we review the sensitivity of our results to changes of certain variables in our model. First, if we assume 100% of the deposits to be excluded from the bail-in, the government would have to step in for almost all banks of our sample. Furthermore, it would have been not possible for most of the banks considered to increase their amount of covered bonds or even decrease their WACF based on a 20% asset loss scenario. Second, we also found the effect of higher or lower levels of over-collateralisation for covered bonds on the total WACF to be rather small. For example, in a 10% asset loss scenario it only causes average deviations of our estimated total WACF of $\pm$ 1% in the case of 3% or 25% over-collateralisation. The larger the share of covered bonds in the funding strategy of the corresponding bank, the greater is the impact of a lower covered bond over-collateralisation. Third, we also found the effect of higher or lower levels of over-collateralisation for covered bonds on the total WACF to be rather small, i.e. +/- 1% in the case of 3% or 25% over-collateralisation. The larger the share of covered bonds in the funding strategy of the corresponding bank, the greater is the impact of a lower covered bond over-collateralisation.

In general, considering a 2% asset loss scenario, we find the bail-in framework neither to affect the overall funding costs of the corresponding bank nor the costs of senior unsecured debt, except for Svenska Handelsbanken (see Table 16). This is in line with Cœuré (2013), who referred to an econometric study undertaken at the ECB which suggests the announcement of the bail-in to have limited impact on the cost of senior unsecured debt. But incorporating higher asset loss scenarios, the impact of decreased recovery values for senior unsecured investors cannot be ignored as shown in Table 17 and Table 18. Furthermore, from the perspective of a senior unsecured investor it is difficult to estimate his potential risk in terms of lower recovery values in the case of the bank's default. It is not an easy task to get information of all outstanding covered bonds or the level of asset encumbrance of different banks.

3.5. Conclusion

This chapter analyses the impact of the bail-in mechanism proposed by the European Union. This regulation led to an increasing level of asset encumbrance on banks funding strategies. From the banks' perspective we found an overall increase of the weighted average cost of funding incorporating the monitoring effect of senior unsecured investors. Nevertheless, we show that banks can still lower their overall costs of funding when increasing their amount of covered bonds on average by around 17% of funded assets. However, when considering asset losses above 20% or an amount of covered funding above the optimum, the senior unsecured instruments might not be sufficient to cover the losses of the bank which could also affect the spreads of covered bonds.

4. Sustainability, Green Bonds and Ethics

4.1. Introduction

Sustainability - 'Environmental, Social and Governance' - was introduced as a concept forty years ago, starting in the USA and Britain.[115] Initially Sustainability may have been just about environment. It has later adopted social responsibility and matters of corporate governance. As 'movement' it requires volume, widely accepted standards, consistency and a well-defined framework. Various attempts at creating a framework were initiated. Attempts come from different ideological backgrounds. Depending on whether they come from 'Green policies', human rights or religious groups *etc.* the goals can be diverse. In Germany, for example, the churches, particularly the protestant church, were early drivers of the sustainable thought process. Thus their criteria have become the gold standard in the German market.[116] An unambiguous consensus of what Sustainability truly means globally does not exist. In our paper we refer to the following definition: Sustainable investment "applies a set of investment screens to select or exclude assets based on ecological, social, corporate Governance, or ethical criteria, and often engages in the local communities and in shareholder activism to further corporate strategies towards the above aims" (Renneboog et al., 2008, p.1723). Despite the differences[117] – in particular and in principle – the market has acknowledged 10 principles of the UN-global-compact as the minimum standard. Serious violations of the Principles of Responsible Investments (PRI) lead under Sustainability regimes to plain divestments. To provide a perspective for its relevance - roughly 50% of the 60 trillion USD in assets under management globally are managed by PRI signatories (PRI, 2015). Assets under PRI management grew from 6.5 trillion USD in 2006 to 34 trillion USD in 2014. There is a strong trend and the market share for sustainable management is increasing over the years (Eurosif, 2012, 2016). By 2016, 58 stock exchanges with a market share of 70% are committed to Sustainability, requiring, for instance, from companies specific ESG information to be listed (PRI, 2016). However, the topic of Sustainable asset management is complex and multi-layered. The practices diverge. Some institutional investors exclude companies missing certain ESG minimum standards (exclusion or negative approach). Others actively engage in sustainable growth prospects (inclusion or positive

[115] Actually 'ethical' investments appeared in the USA as early as the 1920s by exclusion of companies considered dealing in 'immoral activities' like alcohol, tobacco *etc*. (see Revelli and Viviani, 2013). On the specific topic of Ethics in Banking today see e.g. Villa (2015).

[116] Other guidelines exist, i.e. from DSGV (Deutscher Sparkassen und Giro Verband) or Forum NG.

[117] The biggest differences are in the area of Corporate Governance.

approach). They engross themselves in the active management of companies – calling for ESG initiatives either publicly or in private. Some file shareholder proposals or start public relations and marketing activities (see Busch et al., 2015; Busch and Koelbel, 2014, Kurtz, 2008).

Sustainability decisions, meaning the actual *assignment* of 'what is sustainable and what is not?',[118] are usually externalised, similar to credit decisions. ESG differentiation on a firm level is not easy. Companies often engage in special or activist projects, but not consistently across all their activities. Therefore often disparate facts are condensed to a single Sustainability rating, published by Sustainability Rating Agencies like Oekom Research or KLD Research & Analytics, Inc. (KLD) who produces MSCI's ESG Research. Their assessment considers a broad range of criteria on a detailed level. By way of example, the corporate rating of Oekom covers two dimensions with three sub-categories. The Social dimension is segregated into 'Staff and Suppliers', 'Society and Product Responsibility' and 'Corporate Governance and Business Ethics'. The three sub-categories for the Environmental Rating are 'Environmental Management', 'Products and Services' and 'Eco-efficiency' (Oekom, 2014). The information is condensed to a twelve-tier rating scale, ranging from D- (very poor performance) to A+ (excellent performance).

The MSCI's ESG Research Rating on the other hand assigns a nine-point scale from AAA to C instead, after reviewing > 500 data points and scoring > 100 indicators.[119] MSCI's rating evaluates companies based on five categories:

1. Environment – how a company manages its environmental challenges;
2. Community and Society – e.g. treatment of local population, handling of human rights and philanthropic activities;
3. Employees and Supply Chain – how a company deals with employees, contractors and suppliers, in particular labour-management relations, anti-discrimination policies and practices, furthermore how a company deals with employee safety and labour rights of workers *throughout* the supply chain;

---

[118] See here also as a further reference the Darmstadt Definition of Sustainable Investments and their discussion by Busch *et al.* (2015).

[119] Scores and ratings are not normalised across individual industries. That means certain industry sectors may have i.e. no AAA or no C ratings. See MSCI (2011).

4. Customers – safety and quality of products, marketing practices and its record of regulatory or anti-competitive controversies; and
5. Governance and Ethics – investor relations and management practices, any Sustainability reporting, whether the board is really accountable and a company's policies and practices on 'Business Ethics'.

As such the process of Sustainability ratings copies Credit ratings – with one exception. Credit ratings are paid by the issuer. Sustainability ratings are paid for by the investor.[120] Sustainability Agencies engage actively with the companies they rate. They require confirmations and documentation. However, they do not work for the companies they rate, unlike credit rating agencies. They do not consult them, as accounting firms often do. Thus there is a high degree of independence.[121] The public feels less unease about conflicts of interest. Following the introduction we consult the multitude of empirical research on this topic to assess whether ESG has a justification as such. We distinguish the discussion between economic return (creating positive performance for the investor, as claimed by the fund industry) and social return (that it creates a benefit for the society as a whole).[122] We start with the economic return discussion.

4.2. Business Case for Investing under ESG – Economic Return

First, there are reputational benefits. In a world where brand reputation is a valued currency, to be protected at all costs, elements of Ethics, reputation and credibility is of high importance. Investing responsibly creates a positive image and is, as such, a valid argument in any business case. Secondly, we appraise performance related aspects. We call the economic impact of investing under Sustainability constraints the 'ESG alpha' – positive or negative. To assess the ESG alpha we employ the following three strategies:

1. We evaluate the theoretical foundation of an ESG alpha. Theories in support of ESG alpha is for instance the 'Stakeholder Theory' or the 'Theory of information effect' (see

[120] Credit Ratings are paid by the company that is rated while Sustainability Ratings are paid by the party that utilises the rating for its investment decisions.

[121] On the role of Sustainability agencies and their influence on corporate social performance see Slager and Chapple (2015).

[122] The academic standard is not about, whether it is intuitively justifiable or whether it should be done regardless from a 'conscience' perspective. The evidence should be assignable and conclusive …*obiter dictum:* We all agree anyway from an ethical perspective.

Kurtz, 2002, 2005, Clark *et al.*, 2015). Bauer et al. (2005, 2006) point out that better understanding and information leads to a learning effect that counterbalances additional costs over time. A part of the literature argues that ESG management reduces idiosyncratic portfolio risk and volatility. Early opponents like Milton Friedman (1970, p.1) argue that "business as a whole cannot be said to have responsibilities". As it is not their money, managers have responsibilities solely towards their principals. He argues that in a society of 'Capitalism and Freedom' (his book's title) managers' main responsibility is to their owners. Whatever lowers return and profit is against their primary objective as agents (see also Orlitzky 2015). Nowadays this sounds very archaic. Still the line of thought continues, criticising the inherent costliness and the reduction of operative cash flows. Taken to the extreme it puts the entire organisation at risk and undermines the long-term legitimate business objective (see *inter alia* Devinney, 2009 or Nohria, Piper, Gurtler, 2006). Others base their arguments on modern portfolio theory.[123] The reduction of eligible assets leads to diversification costs and therefore *per se* to a lower efficient frontier. In addition, limiting the investment spectrum creates an unwanted sector bias in the portfolio selection (see Le Maux and Le Saout, 2004). Merton (1987) ascertained that the notion of an efficient portfolio just based on diversification is not valid if information is incomplete or asymmetrically distributed. Portfolios benefit from more and better information.[124] Thus an asset manager's expertise and how he processes information is important. Orlitzky (2013) and Busch et al. (2015) argue from a 'Behavioural finance' perspective that information about ESG can lead to erratic price effects. ESG information distorts the market and creates unnecessary noise. The volatility gets inflated by information disadvantage between the different market participants. Normal investors cannot differentiate between 'genuine' and 'disingenuous' commitments to Sustainability, as there are no binding accounting and reporting standards. As such ESG commitments are highly subjective to short-term manipulation and an organisation's rhetoric. This is despite the structured rating process undertaken by Sustainable Rating Agencies. The rating does depend on information provided

[123] See Markowitz (1952).
[124] See reference also in section 2.3.9.

voluntarily. With regard to Sustainability the information distributed is unfortunately heterogeneous - much more so than for Credit ratings[125].

⇨ *Conclusion on theoretical foundation: inconclusive.*

2. We assess research on first-level empirical studies which compares the relative performance of ESG asset management to unrestricted asset management or market benchmarks (see *inter alia* Revelli and Viviani, 2013, 2015). The results in primary studies vary with time frame, the investment strategy, portfolio constraints, the specific ESG criteria and whether the chosen benchmarks are at all appropriate or representative on a broader scale. Adjustments to balance bias in the data are sometimes made, sometimes not. Overall first-level research on ESG alpha struggles with a clear message.

   ⇨ *Conclusion on first-level empirical studies: inconclusive.*

3. Last, we consider meta-studies that integrate research findings across empirical primary studies. Meta-analysis employs various statistical methods to deal with the above issues. They are called artefacts[126] in this context. We refer to Schmidt and Hunter (2015) how to correct 'error and bias in research findings'. Revelli and Viviani (2013, 2015) conduct their meta-analysis based on 85 studies and 190 experiments. They correct for publication bias, which means that interesting and statistically significant results are more likely to be submitted. They verify the absence of selection bias by controlling for overrepresentation, excluding duplicates or studies with quality issues from a methodological perspective. The primary studies are screened based on moderators for quality and methodology of study[127]. The moderators are characterised by their model impact – negative, neutral and positive. After consideration of the moderators' impact they find *"no apparent link between [ESG] and financial performance"* (Revelli and Viviani, 2013, p.113) – with 40 positive, 80 neutral and 41 negative impacts. The "… performance clearly depends on the

---

[125] Credit rating agencies rely in major parts on internationally defined accounting and legal standards or generally accepted financial ratios or cash flow calculations like EBITDA, ROE, various capital ratios, operating cash flows, *etc*.

[126] Artefacts are for instance sampling error, error of measurement, dichotomisation, range variations of independent and dependent variables, imperfect construction validity, computational errors, data errors, bias in sample correlation *etc*. (see Schmidt and Hunter, 2015, part I).

[127] These are: Financial performance measure, observation period, sample size, type of research, type of markets i.e. European, US, EM *etc*., data comparison method, investment family.

methodological choices … [and] the ability of [ESG] fund managers to generate performance" (Revelli and Viviani, 2015, p.14).

Friede et al. (2015) equally surveyed existing studies; again they examined a large number of review and primary studies. First, they simply count the primary studies with positive, negative and non-significant results. Secondly they aggregate the review studies to a second-order meta-analysis. For this purpose 60 vote count and meta-analysis studies got reviewed with an underlying studies number - adjusted for overlaps - of 2200 plus. The share of *positive findings* is with 48% 'vote-count' and 63% 'meta-analysis' *versus negative findings* of 7% and 8% quite remarkable. The question remains whether the result is conclusive - as much as anyone would like to support their main conclusion: "the orientation toward long-term responsible investing should be important for all kinds of rational investors in order to fulfil their fiduciary duties and may better align investors' interests with the broader objectives of society" (Friede et al., 2015, p.227). In the footnotes they state that the statistical explanatory power of vote studies are low as they may come to biased conclusions and "potentially ... overestimate nonsignificant results" (p.227). The logic of aggregating inconclusive individual studies to make deductions on a higher level is challenging. Nevertheless, as the study covers a wide range of asset classes, regions and strategies we find some aspects noteworthy. In particular the correlation between ESG and financial performance seems to be stronger in North America and Emerging Markets. G for better 'Governance' may be relevant. For equities the correlation is weak. Intuitively this makes sense as there are many more drivers than ESG for equity performance. Equity investment strategies are less designed towards managing tail-risk than is the case with credit performance – limited up-side *versus* theoretically unlimited down-side.

⇨ *Conclusion on meta-studies: inconclusive.*

⇨ There is no definite evidence for a positive economic return, outside of a clear reputational benefit for the investor.[128] But in turn there is none for a negative economic return either.[129]

---

[128] The reason in our view is, that investment performance depends on too many factors and it is difficult to discern which ones were responsible for how much.

4.3. Conclusions on ESG - and Social Return

In discussions with Sustainability experts[130] they confirmed that proof of ESG alpha may be difficult; but that there are *valid arguments for employing the ESG framework anyway*. Anecdotal evidence suggests that credit analysis benefits from including ESG. Examples like BP, Siemens and Deutsche Bank[131] with a prior low score on Governance and, in the case of BP, on Environmental standards do substantiate the claim. These hazardous events are often not accidental, rather a train-wreck in the making. Analysing corporate fines and settlements McGregor and Stanley (2014) find that banks have paid >100bn USD in fines since the financial crisis[132] (G) and according to Almashat *et al.* (2010) pharmaceutical companies paid >30bn USD (E). Ignorance of (ethical) standards and absence of consistent and consequent sanctioning mechanisms are strong indicators of cultural deficiencies. The wrong company culture leads up to disastrous consequences long-term. Bauer et al. (2010) conclude that, for example, human capital management and good employment practices and policies (S and G) lower firm-specific risk. Stronger employee relations diminish cash flow uncertainty and idiosyncratic stock volatility. This in turn is 'supposedly' reflected in a lower firm-specific default risk and therefore in the credit spread of corporate bonds or the firms overall financing costs of a company. Chava (2011) states, companies with environmental concerns and hazardous environmental risks (E) have less institutional ownership. Their cost of capital increases. Also fewer banks are willing to participate in a loan syndicate which again rises their financing costs[133].

In our view, ESG standards help to gain a holistic understanding of companies as investment targets. Companies should not just be assessed on a micro level. ESG can help with identifying cliff risks. Cliff risks, and 'major opportunities', come from trend changes – see,

---

[129] If there were it would have been possible to devise a 'cost-benefit curve' in the sense of how much does a favourable reputation or a calm conscience cost in economic return.

[130] We reference to Asset Managers and 'Sustainability panels'.

[131] BP's 'Deep Water Horizon' costs are >20bn, Siemens corruption settlements and penalties >2bn and Dt.Bank diverse settlements and penalties adding up to double-digit bn costs.

[132] I.e. JP Morgan 13bn in 2013, BNP 9bn in 2014, CS 3bn in 2014, HSBC 2bn in 2012, DB >10bn *etc*. for misleading product information, violation of sanction rules, helping US citizens with tax violations, failing to maintain effective anti-money laundering processes, Libor and FX rigging, …

[133] This is confirmed by Bauer and Hann (2010). Their analysis shows also that in the context of environmental concerns there is a premium on the cost of debt financing and lower credit ratings. ESG information affects the pricing of bonds and loans for corporates.

for instance, the reduction of carbon prints.[134] The trend to reduce carbon emissions is likely to have a hugely negative impact on the long-term profit outlook for coal, oil and gas companies. Even car manufacturers may have relied for too long on combustion engine technology (see *inter alia* Oekom, 2017). The reduction of carbon emissions will be on the other hand positive for companies that focus on renewable themes – see Tesla's high market capitalization of 55bn USD relative to expected 2017 profits of 860mio USD (EBITDA, adj.).[135] Most companies have 'positives' and 'negatives' in their product portfolio, enabling them to manage the transitory period. This applies particularly to the big industrial heavy weights or the Food & Beverage industry. The Automobile industry is a "prime example of transformational change" (Oekom, 2017, p.7f.). BMW, Bosch and Siemens, as well as Eon and RWE follow a double strategy - they position themselves for the future without giving up yet on the past. "Such changing conditions in business environments can affect business risk, profitability, and ultimately firms' competitive advantage" (Busch et al., 2014, p.10). As it concerns tail-risk - unlikely events but with big potential economic impact – it is particularly important for credit. However, ESG analysis is not fool-proof. ESG missed, for instance, 'Diesel gate'. Prior to the scandal, Volkswagen was rated by Oekom as 'best in class'. Relative comparison within a sector is at best problematic, even if it is complemented by a minimum level on an absolute scale. We point in this context to the controversial business practises of whole industries[136] (Oekom, 2014, 2017). We accept though that 'best in class' is a necessary conceptual extension. Modern portfolio theory stresses two elements: Diversification and expertise. ESG helps with the comprehensive understanding of investment targets. It increases expertise. But ESG also reduces the investment universe by 40% to 60% depending on the concrete criteria and the applied minimum rating. Thus to allow for some diversification the 'best in class' approach is a necessary compromise. It is required under risk-return considerations and caps diversification costs in the context of modern portfolio theory.

---

[134] Binding agreements are i.e. Paris concord 2015 or Sweden targeting "Zero carbon emissions" by 2030/35.

[135] With the BF EV/EBITDA of 1499% (see Bloomberg Equity Relative Valuation by 01.08.2017). BF EV/EBITDA is the blended Forward of the ratio between Equity Valuation and Profitability. As a comparison Daimler, BMW and Volkswagen with market capitalisation of 50-66bn EUR are at -25%, 131% and -42%.

[136] I.e. Oil & Gas, Consumable Fuels industries (ca. 50-60%) or Metals & Mining (39%) with severe controversies toward principles of UN Global Compact; Oil & Gas (30%) and Construction (15%) toward Corruption; Textiles & Apparel (21%) and Electronics (14%) toward labour rights; Oil & Gas and Metals & Mining (ca. 30-40%) severe controversies toward environment (see Oekom, 2017, p.20-26).

However, independent of economics there are other drivers to be considered. The EU/ECB requires *pension funds* going forward to evaluate Sustainability risks in their portfolio and to provide transparency.[137] The EU Member states are obliged to draft the new regulation into national law within 18 months. Transparency will force investment guidelines to adjust and to develop. Yet, Busch *et al.* (2015, p.1) point out: “ESG data must become more trustworthy.” ‘How ESG is measured and appraised’ should be further enhanced.[138] Information asymmetries should be reduced. It requires improved data collection processes and an independent third party control. Information provided by companies’ needs verification.[139] It needs greater transparency with regard to the screening techniques applied. Investors have “to understand and trust ESG-related data and ratings. Adequate managerial and investor competencies and knowledge are a (…) prerequisite” (Busch *et al.*, 2015, p.15) to grow PRI market penetration. The goal is to grow the 19% for asset owners, to match the 63% for investment managers[140]. Reorientation towards a long-term paradigm and a perspective towards new opportunities become vital. Asset classes like real estate or project finance are promising extensions for ‘green = good’ (see Eichholtz et al., 2010, 2016, Coulson *et al. 2017*). Regulation can support the trend.

Now let’s look at the social return, the positive ESG effect for the society as a whole. The World Wide Fund for Nature (WWF et al., 2010) states that if we do not improve towards a closed-loop economy, humanity will require two Earths by 2030 (-> ecological perimeters). Decreasing biodiversity[141], a growing human population, rising water levels, exhaustion of natural resources[142] and air pollution in cities with detrimental effects on human health are ‘cliff risks for humanity’ (see also WWF, 2016). To what extent can financial investments under ESG actually advance, for example, ecology? Busch *et al.* (2015) come to the conclusion that the transformation effects are weak. The link between a decision to invest in a company and the decision of the company’s management to actually engage in sustainable

---

[137] **https://www.ecb.europa.eu/press/pr/date/2017/html/ecb.pr170726_1.en.html**, see also European Insurance and Occupational Pensions Authority (EIOPA) and European Political Strategy Centre Strategic Notes (2017).

[138] See Orlitzky (2013).

[139] According to the PRI report on progress (2015) only 13% of the PRI signatories have confirmed “that their submissions had been assured by a third-party provider” (p.6).

[140] See PRI report on progress (2015) (p.9).

[141] A drop of 58% in animal population over the last 40 years

[142] The number is i.e. minus 239mio hectares for natural forest.

business practises is indirect and loose.[143] Despite a trend toward ESG in asset management they did not observe a shift toward Sustainability in operating businesses. In our interpretation, this is due to other investors stepping in when ESG investors are stepping out. Large companies have usually sufficient alternative investors and so far ESG oriented investors are just minority owners.[144] This may change with an increasing market share. ESG investors would have a stronger stake in management selection and the setting of management incentives. In the meantime success stories like Tesla or the recent upsurge in Green Bonds may pave the way for more social return.

4.4. Green Bond Solution

Green Bonds are to date one of the fastest growing markets in fixed income. Green Bonds differ from regular bonds by label only. The Green Bond label signifies the "commitment to exclusively use the funds raised to finance or re-finance 'green' projects, assets or business activities"[145] (see also ICMA, 2017). Thus Green Bonds create a direct link between investors and projects deploying sustainable business practices. Together with new technology they increasingly become a game changer for the world, ecologically speaking - accelerating social return. The market is growing exponentially with 81bn USD Green Bonds issued by 2016, up 92% within a year (Kidney, 2017). A growing number of issuers are responsible for the doubling in market size, with 90+ issuers by 2016. Issuers come from 24 countries. More and more banks are joining the trend. The largest Green Bond to date - with 4.3bn USD - was issued in 2016 for instance by a Chinese bank, the Bank of Communications. Standardisation in this new market segment is gaining pace. By 2015, four Green Bond indices have been established.[146] Unfortunately each index uses different thresholds and eligibility criteria, e.g. currency, size, rating and second-party opinions (see OECD, 2017).

---

[143] An example of such rather loose connection is the link between the Sustainability rating process and corporate social performance; there seems to be a link through exclusion threats, signalling, and engagement (Slager and Chapple, 2015).

[144] Respectively they stay even below relevant blocking minorities.

[145] See OECD (2017).

[146] The four indices are: Bank of America Merrill Lynch Green Bond Index, Barclays MSCI Green Bond Index, S&P Green Bond Index and Green Project Bond Index and Solar-active Green Bond Index.

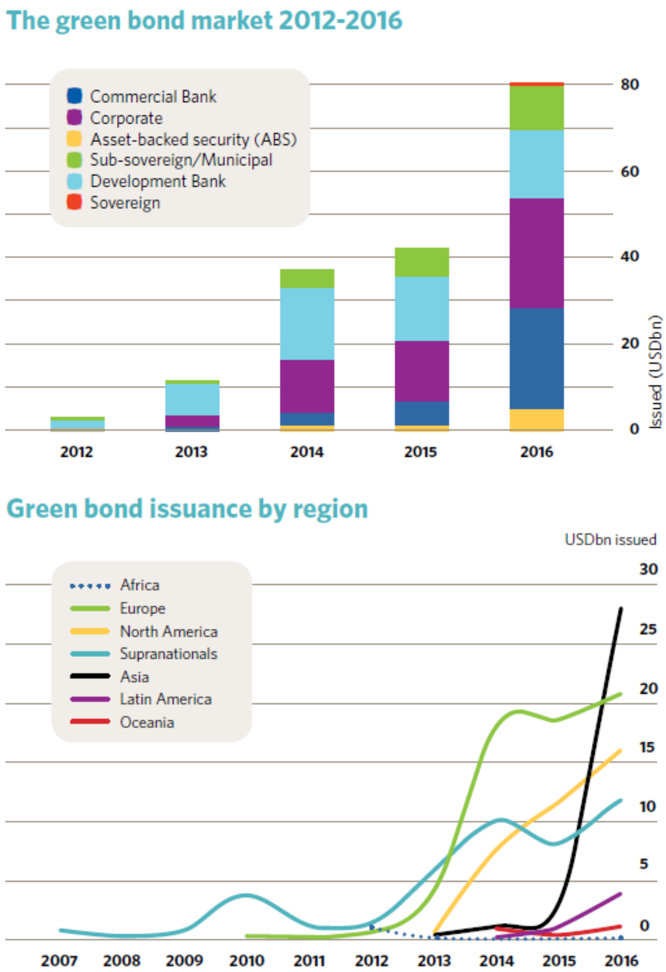


Figure 23a, 23b: The Green Bond market and issuance by issuer category and region in USD bn.[147]

Issuer motivation of Green Bonds seems primarily driven by promotion and green credentials rather than accessing funding at lower costs. Green Bonds are typically issued with spreads in-line with other non-green bonds from the same issuer. In a market study covering 62 Green Bonds over 15 months in 2016/17 'Climate Bonds Initiative' found that some bonds priced inside, some on and some outside their own credit curves (Mancuso, 2017). 70% tightened "within seven days after announcement date". The bonds were on average 3 times oversubscribed (yet, this is not dissimilar to vanilla bonds in the current market environment). Over a period of 28 days Green Bonds have outperformed according to 'Climate Bonds Initiative' on an absolute and on a relative basis (compared to corresponding credit indices). In the secondary market Green Bonds are generally low marked-to-market volatile, as they are usually acquired by long-term investors. Whether all this indicates that Green Bonds are

[147] See Kidney (2017).

still under-priced at issuance, as ‘Climate Bonds’ suggests, and why that should be is not clear. The Green Bond performance may not be too different to the performance of the primary bond market as a whole. The bond markets are flushed with central bank money. However, under-pricing could be a sign that the market needs to develop further. Certain issuers, for example, KfW, have just instituted detailed green programs. New transparency rules regulate whether a specific ‘Green project’ is included or not. Other issuers will follow suit. First signs of ‘Green washing’ are countered by requesting a ‘Second party opinion’, provided by Sustainability Agencies. However, this comes with an issue as the Agencies are remunerated by the issuer. This creates future conflict of interest and a potential weakening of their independence.

Another issue is still the legal nature of the cover pool register. At present, most issuers of Green Bonds are high quality and, as such, safe and low yielding. As issuers become more diverse in quality the market may become more challenging. Ring-fencing the collateral to protect the Green Bonds against issuer default would allow lower credit quality issuers to enter the market. Most projects covering Green Bonds are yet just earmarked for the pool. A developing market with better funding levels will require regulation and a thought-through structure for administering the cover pool. Ring-fencing, comparable to Covered bonds, will promote the market segment by lowering funding costs. It will provide an economic benefit for issuers, compensating for the additional effort of running a register. A funding advantage will make the collateral, the financing of green practises and projects, more economical. This in turn will strengthen the ecological effect and therefore the social return.

5. The Impact of Commodity Finance on Resource Availability[148]

5.1. Introduction

The availability of natural resources is distributed heterogeneously across the earth. Most of the developed countries face a situation in which they demand the majority of resources but do not possess primary access to them. In recent years, the rise of China or Brazil leads to a sharpening of this situation for the advanced economies. Most industrial nations are reliant on a secure supply of raw materials; a situation widely accepted by their respective governments which have instigated a variety of programs to secure availability. Given the disparate geographical locations of resource deposits it is not surprising that international trade plays a central role in the supply and availability of commodities to industrialised nations. Global trade, however, is sensitive to transaction costs. Hummels (2007) illustrates how a reduction in historic transport costs accounts for increased levels of trade. Evans and Harrigan (2005) show how international trade is impacted by distance and production costs, and Deardorff (2014) and Schmidt-Eisenlohr (2013) highlight how the extent and pattern of inter-country trade has been influenced by financing cost.

In this article, we refine the assessment of global trade to base metals; a resource primarily imported by industrialised nations. Importing a scarce resource increases its availability to the domestic market. Scarcity occurs if the resource neither is produced domestically in sufficient quantity nor can be imported to meet internal demand. The availability of imports thus directly impacts the availability to consumers with respect to the resource. We expect factors that impact resource imports to likewise be factors that impact resource availability to a domestic consumer and model the availability of base metals by analysing country-specific import data and the financial environment of that country. Whilst a reasonable stock will act as a buffer in future shortages, such as the oil depots in major countries, we are more interested in the effect of a change in worldwide commodity flows during normal times. We assumed that a fixed amount of any commodity is produced worldwide in a certain year; then we explore the effect of financing conditions on the flow of material to this country.

[148] This chapter is published in Applied Economics Letters, 2014 by Posch, Erhardt, Hard (2014).

5.2. Empirical Analysis

We expect financing costs to negatively influence the relative amount of goods imported. We use data on import and exports from UNComtrade on the six nonferrous metals traded at the London Metal Exchange (LME). The LME price is accepted as a worldwide benchmark and local prices are traded with a fixed spread to the LME price. Exploring the question how financial conditions of a country influence the resource availability in this country we use the following model:

$$\Delta Import_{i,t}^{j} = \alpha + \beta_1 \Delta SovCDS_t^j + \beta_2 \Delta\, Price_{i,t} + \gamma_1 STFinancing_t^j + \gamma_2 LTFinancing_t^j + \beta_3 \Delta GDP_t^j$$

where *j* denotes the country and *i* is the resource we are exploring. The import data are on yearly basis and we average variables with higher frequency. We use data for 18 industrialised countries of which most do not have access to primary resources. Sovereign CDS (SovCDS) markets provide a good estimate of country risk, cf. Kalteier and Posch (2013), and we expect the riskiness of a country to negatively impact imports. Since we control for the financing conditions of the country, the Sovereign CDS will only capture the remaining country risk, such as political risks, freedom to conduct business *etc.* Using prices for each commodity is essential to capture the change in import and thus the change of resource availability in a country due to higher market price and changes in the exchange rate. We use the 3-month forward contract which is the most liquidly traded contract on the LME. The price effect on imports can be lightly positive or negative. For countries with a low value-added on the imported raw materials we expect a negative influence as the producers are unlikely to pass on higher raw material prices to their clients. For countries using imports to produce high-value-added products, we would expect to see no or even a positive price impact on imports.

For the short-term (ST) financing conditions we use the country's 3-year government bond, and for the long-term (LT) the difference of the 10-year and the 5-year government bond in each country. The latter definition allows controlling for the steepness of the term structure curve and captures the change of imports due to rising long-term financing costs. The financing conditions of commodity user within the country will include risk and liquidity premia that will be captured by the constant and error term in the regression. We expect both variables to negatively influence the imports of resources. The reason being that a restriction of credit primarily restricts the financing available to agents enacting global trade. This is particularly the case when considering the financing requirements of intermediaries who purchase product as principal using a leveraged capital structure. Furthermore a higher interest rate implies higher opportunity costs for providers of credit. Conducting robustness checks on other aggregation measures (median, volatility weighted *etc*.) does not yield qualitatively different results.

We control for each country's change in productivity, measured by the GDP. Increases in productivity will *ceteris paribus* increase demand for raw materials. We estimate the model using pooled and fixed effects regression with cluster and heteroscedasticity robust standard errors. The results are given in Table 19.

The results show the effects across all countries, including producing countries like Turkey, Indonesia or Malaysia, as well as highly industrialised countries like the US or Germany with little primary resource access. Notable in the regression is the strong explanatory power of the model; for a regression in changes the R2 indicates a good fit. While the sovereign CDS is insignificant for any metal, it still adds to the explanatory power of the model. The remaining variables are the driving factors of imports worldwide. Here, the price change shows the expected mixed results, for Zinc we have a strong positive price impact of 45.294 kg more imports for a price change of 1 USD, while for Tin we have a negative -2.271 kg/USD. The other metals do not show significant impacts however with a negative tendency confirming our hypotheses.

The financing variables are significant for all but lead to and show a negative sign. The highest coefficient is given for Copper and Aluminium, for the former one basis point increase in financing costs reduces imports by 3353 kg, for Aluminium by 2025 kg. The same holds for the steepness of the LT-curve: the steeper these curve the lower the imports.

Table 19: Fixed Effects Regression of Import Changes, worldwide
The Table shows the results of the OLS regression described above. Coefficients are accompanied with a T statistic (in parentheses) and a *,**,*** indicator of statistical significance on the 10%, 5% and 1% level respectively. A goodness of fit R^2 measure is given for each metal for N observations.

	Zinc	Tin	Nickel	Lead	Copper	Aluminium
Sov_CDS	-151	-5.467	-357.856	-27.376	-1.580	-6.78
	(-1.47)	(-0.36)	(-1.16)	(-0.71)	(-1.14)	(-0.93)
Price	45.294**	-2.271*	-1.26	-5.89	-121.98	-60.91
	(-2.11)	(-1.73)	(-0.42)	(-1.14)	(-1.18)	(-1.04)
ST Finance	-213.545**	-82.070*	-686.562*	-61.804	-3353.985**	-2025.675***
	(-1.98)	(-1.90)	(-1.89)	(-1.39)	(-1.98)	(-2.45)
LT Finance	-286.850**	-95.319*	-897.960*	-7.85E+01	-4433.022**	-2415.207**
	(-2.14)	(-1.91)	(-1.92)	(-1.30)	(-2.03)	(-2.35)
GDP	-80.338**	-14.2	-137.811*	-13.7	-525.895*	-329.824*
	(-2.30)	(-1.70)	(-1.86)	(-0.83)	(-1.78)	(-1.89)
Const	117**	59516.516**	3.40*	43009.084*	2.23**	1.16**
	(-2.19)	(-2.57)	(-1.81)	(-1.89)	(-2.4)	(-2.79)
N	119	119	119	119	119	119
R^2	0.107	0.052	0.047	0.021	0.081	0.129

The negative relationship of metal imports to finance costs is qualitatively consistent with Schmidt-Eisenlohr (2013) who shows a 1% higher financing cost in a country is associated with 2.3% lower imports of goods by that country. While this first analysis is for all countries in our sample, we now turn to the European case, repeating the model's regression for European countries only, cf. Table 20.

We find the general tendency of the world model but with interesting deviations: Zinc prices do not influence the imports. The findings from the financing variables however still hold with all metals having a negative relationship for both short- and long-term finance costs.

Table 20: Fixed Effects Regression of Import Changes, European Union
The Table shows the results of the OLS regression described above. Coefficients are accompanied with a T statistic (in parentheses) and a *,**,*** indicator of statistical significance on the 10%, 5% and 1% level respectively. A goodness of fit R^2 measure is given for each metal for N observations.

	Zinc	Tin	Nickel	Lead	Copper	Aluminium
Sov_CDS	-53.727**	-5.607	-132.689	-13.806	-319.961	-251.849
	(-2.27)	(-0.48)	(-0.93)	(-0.45)	(-1.00)	(-0.50)
Price	60.711**	-0.507	2.646	1.656	-4.761	35.136
	(-2.55)	(-1.43)	(-1.01)	(-0.23)	(-0.15)	(-0.42)
ST Finance	-244.645**	-73.061	-704.288	-118.473	-2094.226**	-2349.119
	(-2.14)	(-1.29)	(-1.42)	(-1.32)	(-1.73)	(-1.66)
LT Finance	-281.459**	-81.473	-812.963	-131.522	-2475.908*	-2551.243*
	(-2.31)	(-1.29)	(-1.46)	(-1.28)	(-1.90)	(-1.66)
GDP	-101.490**	-15.254	-202.19*	-24.782	-395.009	-458.669
	(-2.33)	(-1.07)	(-1.70)	(-1.05)	(-1.36)	(-1.29)
Const	1.04E+05**	35253.555	2.86E+05	56887.082	9.87E+05*	1.03E+06*
	(-2.28)	(-1.47)	(-1.36)	(-1.49)	(-1.92)	(-1.75)
N	58	58	58	58	58	58
R^2	0.251	0.088	0.083	0.089	0.227	0.196

5.3. Conclusion

For the purposes of interpreting why funding costs impact availability on a firm level, we consider Amiti and Weinstein (2011) who estimate the average transit time for US imports to be 2 months which includes administrative periods prior to shipping and post landing. Their implication being that firms engaged in international trade are likely to be more reliant than domestic firms on working-capital financing to cover the costs of goods that have been produced but not yet delivered. We elaborate on the above-mentioned reliance to consider the financing cost for participants in the base metal market during the dead period, a time when the commodity cannot contribute to economic activity and produces no return in terms of dividends or yields. As per the exporting firms mentioned by Amiti and Weinstein (2011), financing costs in this regard would impact to the return on equity of the firms in question. The involvement of an intermediary in the import process enables a third party to bear the funding burden for the dead period rather than exporter producers or importer consumers. Intermediaries supply customers by effecting a physical arbitrage, e.g. they purchase product as principal from a producer then store, transport or refine it using the infrastructure at their disposal, and finally make delivery to industrial consumers in the agreed quantity, grade and time frame. When considered in the arbitrage framework, intermediary funding costs are one of several frictions that determine transaction profitability and by consequence whether the transaction will be executed in the first place. The impact of this cost is especially felt by intermediaries whose business model requires a leveraged capital structure. Here, short-term external finance (often collateralised with the product under transaction and self-liquidating in nature) is critical to effect the arbitrage in a profitable manner. Irrespective of the parties involved in the importation of base metals, supply is sensitive to funding costs borne by market participants. In other words, with all else equal, the marginal transaction is more likely to be executed when funding costs are reduced.

Our results allow for the role financing costs play with respect to the availability of base metals to countries. Key is the consistent negative relationship between the financing costs and imports of base metals after allowing for prices and country risk. Under the assumption that marginal imports of a scarce resource increase the marginal availability of such resource to the domestic market, these results indicate that resource availability with respect to base metals is increased with a reduction in financing costs to market participants. We interpret this at a firm level by considering, in a qualitative manner, the funding requirement during the import process and the relative sensitivity of market participants to financing costs.

6. Concluding Summary

The financial world is facing a period of transformation. The ecological and social challenges on the one hand and new technology to generate innovative solutions on the other, put the whole global system in flux. Due to their multiplier effect in money creation, financial institutions are an important transmission constituent for the ecological and technological revolution. In this paper, we have investigated four prime topics as sample cases for the necessary re-engineering of the financial business model. With central banks and start-ups entering the traditional business areas of financial institutions, society may question at some point the justification for protecting the business model of banks by regulation and expensive entry hurdles. Besides profitability, stability and employment of people, financial institutions have to account for purpose and efficiency. The biggest challenge banks face is the coming AI revolution in their business model. To learn AI's potential, we experimented with machine learning algorithms, focusing on different areas of a bank's business. We started with researching relevant methodological elements and their theoretical foundation. As a secondary objective, we wanted to demonstrate that employing the new technology can yield (in parts substantial) improvements; some of them leaving room for further research on a wider scale. However, ultimately, we wanted to analyse what are specific success factors in employing learning algorithms. We did this in a methodical manner by breaking the process into discernible segments. We researched and systematically analysed the relevance of certain modelling parameters and processes. We found that an intelligently enhanced feature set is central to getting superior results – after choosing a sensible ensemble.[149] Depending on the features, in the market trend prediction section we accomplished a substantial outperformance *versus* Null based on precision, accuracy and MCC statistics. Several models achieved a sweeping improvement in Sharpe Ratio and aggregated return over time compared to a naïve benchmark. Single market features were generally less successful. Features across markets, after employing Granger causality, had superior performance. Picking up on Spill-over effects between markets seems to be decisive for a good model performance. To generalise these findings, we would need to repeat the studies on a wider scale with real-time data, making several cuts during the day.[150] The customer coverage analysis is a novel first in

[149] XGBoost, as the 'newest' Gradient Boosting implementation, outperformed the more traditional Random Forest tree algorithm in our experimental set-up. In addition, setting sensible Hyper-parameters is also relevant, but can be simply done by grid search or literature checks.

[150] This would also help to check and correct possible data inconsistencies.

relation to institutional data. We achieved useful insights into customer activity. We were able to forecast 'when' and 'what' customers will trade, outdoing increasingly sophisticated Null hypotheses. On the customer data alone, we achieved once more robust precision, accuracy and MCC statistics, with a notable optimisation gain over Null. However, we believe that results can be improved by broadening the data. Market data and particularly new touch data add valuable information for the model to analyse. Nonetheless, the main gain in economic terms will come by using learning algorithms in a life environment. By analysing day-to-day operations, an institution can monitor, control-group verified, how changes alter behaviour. It is an ideal setting for 'fail quickly and often', which is often a prerequisite to break static behavioural patterns. As the model results were consistent and robust across all applications we proposed the assumption of a 'Genuine Model Robustness' (GMB). Based on GMB we could use importance and improvement scores, i.e. the delta in MCC and Chi-squared values, as a measure for statistical significance of the underlying feature mechanic *retro versa*. The 'Converse Practice' could broaden the utilisation of learning algorithms for application in academic field studies. Our findings leave room for further research by means of generalising the findings on a wider scale, e.g. using real-time data with equally distributed cut-off points.

Following the AI revolution in finance we looked into regulation as a further crossroads topic for banks. As reaction to the financial crisis, legislators and regulators proposed multiple new regulatory instruments to curb the risk of failing banks. These regulatory restrictions come with associated costs. We assessed the impact of the new bail-in mechanism in a European context.[151] Applying a new simulation framework, we found that the new EU bail-in regime should increase the average cost of funding across major European banks by up to 49 basis points. The increase in funding costs reflects the higher risk for senior unsecured investors. As a counter-measure banks can lower their overall funding costs by issuing a higher percentage of covered bonds or other collateralised products (repos *etc*.). Based on our sample, this could produce a funding cost reduction in the order of magnitude of 17 basis

[151] See Erhardt, Luebbers and Posch (2017).

points. In turn, it raises the level of asset encumbrance,[152] as banks' funding strategies diversify away from expensive, unsecured funding.[153]

As third topic, we studied the role of conscience and its economic and social return in the context of Sustainability. 'Investing in a socially responsible way' as a theme affects the society on multiple levels. We completed the review on three levels – theoretical foundation, empirical primary studies and meta-studies. We evaluated whether there is a business case for investing under ESG, as the fund industry often claims that there is an ESG alpha – in the sense that there is a positive economic return from following social conscience. We first looked at the topic from a purely economic perspective. We found that the supposed evidence is muddled and inconclusive. Secondly, we looked at the social return; the positive impact on society as a whole, for instance, in ecological terms. We found once more a rather weak substantiation. The link between investing under ESG and actually furthering sustainable practices in the industry is frail. Evidence in support of cause and effect is merely anecdotal. Companies are layered and complex structures and ESG concepts like 'best in class' try to deal with this complexity. Yet, it does not change the fact that often only parts of a company are ESG oriented; others are not. Hence, the categorisations are often high level and somewhat arbitrary. Despite all these shortcomings, we found evidence that social, market and regulatory pressures will drive the sustainable momentum regardless. A prominent example is the Green Bond market. The market has grown exponentially on a global scale, particularly in China. With Green Bonds, the market has found a more efficient transmission mechanism. Green conscience finally has a chance to interlink with economics directly, as the cover pool needs to be made up from genuinely green projects. However, as the Green Bond market is expected to move down the credit spectrum, the legal framework around ring-fencing the collateral in the event of an issuer default needs specification.[154] Regulating eligibility criteria may be required as well. Done properly, Green Bonds can become a game changer in the fight against depletion of ecological reserves.

[152] Increase of covered bond funding by up to 17%.

[153] As such it is a potentially unwanted secondary impact of the 'bail-in' regulation, which initiated further rules and regulatory adjustments, e.g. minimum ratios for senior, unsecured funding.

[154] This could broaden the concept of 'alternative covered bonds', however increasing further the asset encumbrance problematic on a bank's balance sheet.

Finally, we explored the stimulus effect of finance on the real economy: its macroeconomic impact. We showed that there is a statistically significant positive influence on resource availability through cheaper and reliable financing. Lowering financing costs for resources can be accomplished through guarantee structures with protection from agencies (such as Hermes) or through collateralisation. The collateral may well mean commodities as physical collateral. The discussion in this case is exemplary for how reliable financing supports society as a whole. Developed countries typically do not possess enough primary resources to cover their own demand. They rely on a secure supply of raw materials from developing countries, provided by companies with high leverage and low credit ratings. Hence, the issue of affordable and reliable working-capital financing during the transit time of several months is an essential friction cost for companies in this space.[155] Marginal transactions will be executed – or not – depending on the short-term costs. We used fixed effects regression on international trade data and found a consistent negative relationship between the cost of financing and the availability of resources with regard to base metals. Resource availability generally increased with lower costs, though the effect varies depending on the base metal and the respective country. We concluded the study by interpreting the result from a company's perspective. The results confirm that resource availability can be improved by reducing financing costs at a firm level. A case can be made for 'import-export guaranteed' or 'collateralised' financing – being in a society's interest.

All four topics are of major significance for banks going forward. But we believe that the technological revolution entails the 'primary disruption potential'. The financial world has always been concerned with the dissemination and analysis of information. The speed in which complex and big data is handled has changed the playing field. The rapid propagation of information,[156] real-time pricing (in nanoseconds), regulatory transparency rules[157], risk aversion, and the threat of punitive legal settlements reinforce the trend to automated work flows. Automating manual processes increases not just efficiency, but is equally improving process security. Individual mistakes are avoided. There is the expectation that machines are easier to supervise as behavioural guidelines can be coded in. Transaction costs are

---

[155] This is independent of them being primary producer or intermediary.

[156] One of the 'transformers' are social networks with billions of real-time users..

[157] On the one hand, regulation, with its eagerness for transparency and data downloads (regarding stress tests, real-time data around execution for transactions, automated trading, etc.), enforces the technological revolution. On the other hand regulation stalls technological advancement, as it also erects high entry hurdles for new competitors.

minimised. Decisions can be automated, practically externalised. The traditional organisation in banks will be altered. The low interest environment and shrinking product margins through productivity gains will force financial institutions into continuous optimisation. Nonetheless, banks need to be specific about 'what they stand for'. Banks need to articulate where they are in regard to 'social return matters' of the broader public. The support of 'Green projects' and 'Resource Availability' are long-term commitments. An opportunistic approach doesn't provide the required stringency for decades. New *alternative covered bond* legislation[158] could sustenance the long-term orientation toward such alternative topics. The positive effect of such legislation can be seen in the mortgage space. A distinct legal framework around covered bonds sustains even specialised banks.[159] Constructive choices are to be made.

[158] Alternative topics could be: Green covered bonds, Commodity covered bonds and Corporate covered bonds.

[159] They are called *Pfandbrief Banken* in Germany and their funding is mostly based on *Pfandbriefe*, which go back more than 200 years. The rules for the cover pool are precise and ring-fencing of assets in case of an issuer default are specified. The existence of *Pfandbrief Banken* for decades validates the claim that asset encumbrance concerns can be ultimately governed.

7. Literature

References

1. Almashat, S., Preston, C., Waterman, T., Wolfe, S. (2010), Rapidly increasing criminal and civil monetary penalties against the pharmaceutical industry: 1991 to 2010, Washington D.C.: Public Citizen's Health Research Group; 2010
2. Amiti, M., Weinstein, D. E. (2011), Exports and financial shocks, The Quarterly Journal of Economics, 126: 1841 – 77, doi:10.1093/qje/qjr033
3. Arifovic, J., Chiarella, C., He, X., Wej, L. (2016), High frequency trading and learning in a dynamic limit order market, International Conference on Computing in Economics and Finance, Bordeaux, Available online at: https://opus.lib.uts.edu.au/handle/10453/85524
4. Ayadi, R., Lastra, R. M. (2010), Proposals for reforming deposit guarantee schemes in Europe, Journal of Banking Regulation, 11(3): 210 – 222
5. Bai, J., Collin-Dufresne, P. (2011), The determinants of the CDS-bond basis during the financial crisis of 2007-2009, Working Paper, Available online at: http://arno.uvt.nl/show.cgi?fid=127965
6. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., Nielsen, H. (2000), Assessing the accuracy of prediction algorithms for classification: An overview, Bioinformatics, 16(5): 412 – 424
7. Bauer, R., Derwall, J., Hann, D. (2010), Employee Relations and Credit Risk, Working Paper, Maastricht University, European Centre for Corporate Engagement
8. Bauer, R., Hann, D. (2010), Corporate Environmental Management and Credit Risk, Working Paper, Maastricht University, European Centre for Corporate Engagement
9. Bauer, R., Koedjk, K., Otten, R. (2005), International evidence on ethical mutual fund performance and investment style, Journal of Banking & Finance, 29, 1751 – 1767
10. Bauer, R., Otten, R., Tourani Rad, A. (2006), Ethical investing in Australia: is there a financial penalty?, Pacific-Basin Finance Journal, 14: 1, 33 – 48
11. Bessler, D. A., Yang, J. (2003), The structure of interdependence in international stock markets, Journal of international money and finance, 22(2): 261 – 287
12. Biau, G. (2012), Analysis of a random forests model, The Journal of Machine Learning Research, 13(1): 1063 – 1095

13 Bollerslev, T. (1986), Generalized autoregressive conditional heteroscedasticity, Journal of econometrics, 31(3): 307 – 327

14 Box G., Jenkins, G., Reinsel, G. (2013), Time Series Analysis, 4th ed., doi:10.1002/9781118619193

15 Breiman, L. (2001), Random forests, Machine Learning, 45(1), 5 – 32

16 Breiman, L., Cutler , A. (2003), Manual Setting Up, Using, And Understanding Random Forests, V4.0., Available online at: https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf

17 Brownlee, J. (2016), Machine learning mastery, email course, Available online at: https://machinelearningmastery.com/author/jasonb/

18 Busch, T., Koelbel, J. (2014), How culture moderates the effect of coporate sustainablity on firm risk, Working Paper, Univerisity of Hamburg / ETH Zuerich

19 Busch, T., Bauer, R., Orlitzky, M. (2015), Sustainable Development and Financial Markets: Old Paths and New Avenues, Business & Society, 1 - 27, doi:10.1177/0007650315570701

20 Calomiris, C. W. (1999), Building an incentive-compatible safety net, Journal of Banking & Finance, 23(10), 1499 - 1519

21 Caruana, R., Niculescu-Mizil, A. (2006), An Empirical Comparison of Supervised Learning Algorithms, Available online at: https://www.cs.cornell.edu/~caruana/ctp/ct.papers/caruana.icml06.pdf

22 Cetinsoy, A., Martin, F. J., Ortega, J. A., Petersen, P. (2016), The Past, Present, and Future of Machine Learning APIs, In Proceedings of The 2nd International Conference on Predictive APIs and Apps, 43 - 49

23 Chan, L., Jegadeesh, N., Lakonishok, J. (1996), Momentum strategies, Journal of Finance, 1681 - 1714

24 Chava, S. (2011/2014), Environmental Externalities and Cost of Capital, Working Paper, Georgia Institute of Technology (published also in Management Science 2014)

25 Cheboli, D. (2010), Anomaly Detection of Time Series, University of Minnesota, Available online at: https://conservancy.umn.edu/bitstream/handle/11299/92985/?sequence=1

26 Chen, D., Bellamy, R. K., Malkin, P. K., Erickson, T. (2016), Diagnostic visualization for non-expert machine learning practitioners: A design study, In Visual Languages and Human-Centric Computing (VL/HCC), 2016 IEEE Symposium on, 87 – 95, IEEE

27 Chen, T. (2014), Introduction to Boosted Trees, Presentation to the University of Washington in October 2014, Available online at: https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf

28 Chen, T., Guestrin, C. (2016), XG Boost: A Scalable Tree Boosting System, March, 1 - 13

29 Choudhry, M. (2012), The principles of Banking, 1st ed., John Wiley & Sons, Singapore

30 Chung, F., Fu, T., Ng, V., Luk, R. (2001), Pattern discovery from stock time series using self-organizing maps, Workshop Notes on KDD2001 Workshop on Temporal Data Mining, 27 - 37

31 Clark, G., Feiner, A., Viehs, M. (2015), From the Stockholder to the Stakeholder: How Sustainability Can Drive Financial Outperformance, doi:10.2139/ssm.2508281

32 Cœuré, B. (2013), The implications of bail-in rules for bank activity and stability. Opening speech at the conference on Financing the recovery after the crisis- the roles of bank profitability, stability and regulation, Bocconi University, Milan, Vol. 30, 2013

33 Coulson, N.E., Lipscomb, C.A., Wang, Y. (2017), Energy Effeciency and the Future of Real-Estate, doi:10.1057/978-1-137-57446-6

34 Crane, A., Matten, D. (2016), Business Ethics, Managing Corporate Citizenship and Sustainability in the Age of Globalization, Fourth Edition, Oxford University Press

35 Deardorff, A. V. (2014), Local comparative advantage: trade costs and the pattern of trade, International Journal of Economic Theory, 10, 9 – 35

36 Demirgüç-Kunt, A., Huizinga, H. (2010), Bank activity and funding strategies: The impact on risk and returns. Journal of Financial Economics, 98(3), 626 – 650

37 Devinney, T. (2009), Is the Socially Responsible Corporation a Myth? The Good, the Bad, and the Ugly of Corporate Social Responsibility, Available online at: https://aom.org/uploadedFiles/Publications/AMP/May09CorpSocialRespbyDevinney.pdf

38 Diaz-Uriarte, R., Alvarez de Andres, S. (2006), Gene selection and classification of microarray data using random forest, BMC Informatics, 2006, 7:3, Available online at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-3

39 Dixon, M. (2016), High Frequency Market Making with Machine Learning, November, 1 – 17

40 Dryden, I., Mardia, K. (1998), Statistical shape analysis, Wiley series in probability and statistics, Wiley, Chichester

41 ECB (2014), Aggregate Report on the Comprehensive Assessment, Technical report, European Central Bank, Frankfurt am Main

42 Eichholtz, P., Kok, N., Quigley, J.M. (2010), Doing Well by Doing Good? Green Office Buildings, American Economic Review 100 (5): 2492 – 2509, doi:10.1257/aer.100.5.2492

43 Eichholtz, P., Kok, N., Quigley, J.M. (2016), Ecological Responsiveness and Corporate Real Estate, Business and Society 55 (3): 330 - 360

44 Elliott, R.N. (1993), [Prechter, R. Ed.], Elliott's Masterworks, 3rd ed., Gainesville, Georgia: New Classics Library (1993)

45 Engle, R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica: Journal of the Econometric Society, 987 – 1007

46 Engle, R. F., Granger, C. (1987), Co-integration and error correction: representation, estimation, and testing, Econometrica: Journal of the Econometric Society, 251 – 276

47 Enthofer, H. (2013), Treasurer's Handbook - Financial Markets: A Practitioner's Guide, 2nd ed., Linde Verlag, Vienna

48 Erhardt, J., Luebbers, J., Posch, P. (2017), Bail-in and asset encumbrance - Implications for banks' asset liability management, Journal of Banking Regulation, April 2017, Volume 18, Issue 2, 149 – 162, also Available online at SSRN (2015): https://dx.doi.org/10.2139/ssrn.2545196

49 EU Council (2013), Council agrees position on bank resolution, Council of the European Union, Technical Report 11228/13, Economic and Financial Affairs Council, Brussels

50 Eun, C., Shim, S. (1989), International Transmission of Stock Market Movements, The Journal of Financial and Quantitative Analysis, Vol. 24, No. 2 (Jun. 1989), 241 – 256, doi:10.2307/2330774

51 European Political Strategy Centre Strategic Notes (2017), Available online at: https://ec.europa.eu/epsc/sites/epsc/files/strategic_note_issue_25.pdf

52 Eurosif (2012), European SRI Study 2012, Available online at: http://www.eurosif.org/wp-content/uploads/2014/05/eurosif-sri-study_low-res-v1.1.pdf

53 Eurosif (2016), European SRI Study 2016, Available online at: http://www.eurosif.org/wp-content/uploads/2016/11/SRI-study-2016-HR.pdf

54 Evans, C. L., Harrigan, J. (2005), Distance, time, and specialization: lean retailing in general equilibrium, American Economic Review, 95, 292 – 313, doi:10.1257/0002828053828590

55 Fama, E. (1970), Efficient Capital Markets: A Review of Theory and Empirical Work, The Journal of Finance 25 (2): 383 – 417, doi:10.1111/j.1540-6261.1970.tb00518.x

56 Fama, E. (1989), Market efficiency, long-term returns and behavioural finance, Journal of Financial Economics, 49, 283 – 306

57 Financial Stability Board (2014), Adequacy of loss-absorbing capacity of global systemically important banks in resolution. Consultative Document

58 Freund, Y., Schapire, R.E. (1996), Experiments with a New Boosting Algorithm, In Machine Learning: Proceedings of the Thirteenth International Conference, 1996, 1 - 9

59 Friede, G., Busch, T., Bassen, A. (2015), ESG and financial performance: aggregated evidence from more than 2000 empirical studies, Journal of Sustainable Finance & Investment, 5:4, 210 - 233, doi:10.1080/20430795.2015.1118917

60 Friedman, J.H. (2001), Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29, 5 (2001), 1189 – 1232, Available online at: https://projecteuclid.org/euclid.aos/1013203451

61 Friedman, J.H., Hastie, T., Tibshirani, R. (2000), Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors, In Ann. Statist.28, 2, 337 – 407

62 Friedman, J.H., Hastie, T., Tibshirani, R. (2009), The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., New York, NY, USA

63 Friedman, J.H., Popescu, B.E. (2004), Gradient directed regularization for linear regression and classification, Stanford University, Department of Statistics, Technical report, Available online at: http://statweb.stanford.edu/~jhf/ftp/pathlite.pdf

64 Friedman, J.H., Popescu, B.E. (2004), Predictive Learning via Rule Ensembles, Available online at: http://statweb.stanford.edu/~jhf/ftp/RuleFit.pdf

65 Friedman, M. (1970), The Social Responsibility of Business is to Increase Its Profits, The New York Times Magazine, 33, doi:10.1007/978-3-540-70818-6_14

66 Frost, A., Prechter, R. (2005 / 1978), Elliott Waves principles, 10th ed., Gainesville, Georgia: New Classics Library (2005), originally published 1978, revised 1998

67 Frost, A., Prechter, R. (2005), Elliott Waves, Fibonacci and Statistics, 2005, see www.elliottwave.com/books

68 Fu, T. (2011), A review on time series data mining, Eng. Appl. of AI 24(1), 164 - 181

69 Garcia, S., Herrera, F. (2008), An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons, In Journal of Machine Learning Research 9 (2008) 2677 - 2694

70 Ghahramani, Z. (2004), Unsupervised learning, Gatsby Computational Neuroscience Unit, University College London, UK, Available online at: http://www.inf.ed.ac.uk/teaching/courses/pmr/docs/ul.pdf

71 Giles, D. (2011), Testing for granger causality, Available online at: http://davegiles.blogspot.co.uk/2011/04/testing-for-granger-causality.html

72 Glionna, J., Hernandez, M., Monteleone, B., Pigott, C., Hedge, K., Crivelli, M. (2012), The implications of 0% recovery, Credit Research Barclays Capital, Technical report, New York

73 Gorodkin, J. (2004), Comparing two k-category assignments by a k-category correlation coefficient. Computational biology and chemistry, 28(5): 367 – 374

74 Granger, C. W. (1969), Investigating causal relations by econometric models and cross-spectral methods, Econometrica: Journal of the Econometric Society, 424 – 438

75 Granger, C. W. (1981), Some properties of time series data and their use in econometric model specification, Journal of econometrics, 16(1): 121 – 130

76 Grossmann, R. (2011), CDS Spreads and Default Risk - a leading indicator?, New York: Fitch Ratings, Technical report

77 Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M. S. (2016), Deep learning for visual understanding: A review, Neurocomputing, 187, 27 - 48

78 Gurgul, H., Lach L. (2009), Linear versus nonlinear causality for dax companies, EconPapers, Operations Research and Decisions, 2009, vol. 3, 27 - 46, Available online at: http://econpapers.repec.org/article/wutjournl/v_3a3_3ay_3a2009_3ap_3a27-46.htm

79 Guyon, I., Elisseeff, A. (2003), An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003), 1157 - 1182

80 Hall, M. A. (1999), Correlation- based Feature Selection for Machine Learning, In Department of Computer Science, Thesis Doctor of Philosophy, University of Waikato, New Zealand, April

81 Hartigan, J. A., Wong, M. A. (1979), Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society, Series C. 28 (1): 100 – 108

82 Hastie, T., Tibshirani, R. (1986), Generalized Additive Models, Statist. Sci. 1 (1986), no. 3, 297 – 310, doi:10.1214/ss/1177013604, Available online at: https://projecteuclid.org/euclid.ss/1177013604

83 Hill (2010), Anomaly detection in streaming environmental sensor data: A data-driven modelling approach, Available online at: http://www.sciencedirect.com/science/article/pii/S1364815209002321

84 Houben, A., Committee on the Global Financial System (2013), Asset encumbrance, financial reform and the demand for collateral assets, Technical Report no. 49, Bank for International Settlements

85 Hull, J., Predescu, M., White, A. (2004), The relationship between credit default swap spreads, bond yields, and credit rating announcements, Journal of Banking & Finance, 28(11), 2789 - 2811

86 Hummels, D. (2007), Transportation costs and international trade in the second era of globalization, The Journal of Economic Perspectives, 21, 131 – 154, doi:10.1257/jep.21.3.131

87 Huynh-Thu, V., Saeys, Y., Wehenkel, L., Geurts, P. (2012), Statistical interpretation of machine learning-based feature importance scores for biomarker discovery, In Bioinformatics, Volume 28, Issue 13, July 2012, 1766 – 1774, Available online at: https://doi.org/10.1093/bioinformatics/bts238

88 ICMA (2017), Available online at: https://www.icmagroup.org/assets/documents/Regulatory/Green-Bonds/GreenBondsBrochure-JUNE2017.pdf

89 IMF (2017), THE GROWING IMPORTANCE OF FINANCIAL SPILLOVERS FROM EMERGING MARKET ECONOMIES, Available online at: http://www.imf.org/external/pubs/ft/gfsr/2016/01/pdf/c2.pdf

90 James, G., Witten, D., Hastie, T., Tibshirani, R. (2013), An introduction to statistical learning, Springer, Available online at: http://www.springer.com/gb/book/9781461471370

91 Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, Econometrica: Journal of the Econometric Society, 1551 – 1580

92 Jolliffe, I. (2002), Principal component analysis, 2nd ed., Wiley Online Library, at: http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf

93 Jurman, G., Riccadonna, S., Furlanello, C. (2012), A comparison of mcc and cen error measures in multi-class prediction, Available online at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0041882

94 Kaggle (2013), Solar Energy Prediction Contest. How is Solar Flux Recorded?, Available online at: https://www.kaggle.com/forums/f/278/ams-2013-2014-solar-energy-prediction-contest/t/5072/how-is-solar-flux-recorded?forumMessageId=27026#post27026

95 Kahneman, D. (2011), Thinking, Fast and Slow, ISBN 9780141099570

96 Kalteier, E., Posch, P. N. (2013), Sovereign asset values and implications for the credit market, Review of Financial Economics, 22, 53 – 60, doi:10.1016/j. rfe.2013.02.001

97 Karmel, R. S. (2014), A law professor's perspective on 'too big to fail', Journal of Banking Regulation, 15(3-4), 227 - 234

98 Kaufman, L., Rousseeuw, P. (1987), Clustering by Means of Medoids, Reports of the Faculty of Mathematics and Informatics

99 Kaufman, L., Rousseeuw, P. (1990), Finding Groups in Data: An Introduction to Cluster Analysis, Wiley Series in Probability and Statistics, Wiley-Interscience

100 Kenyon, C., Stamm, R. (2012), Discounting, LIBOR, CVA and Funding, Palgrave Macmillan, London

101 Keogh, E., Lin, J., Truppel, W. (2003), Clustering of time series subsequences is meaningless: Implications for previous and future research, in ICDM, 115 - 122

102 Khashei, M., Bijari, M., (2010), An artificial neural network (p, d, q) model for time series forecasting, Available online at: https://pdfs.semanticscholar.org/f97f/fbe8b0835640166e1d83507ebcf8d76debf2.pdf

103 Kidney, S. (2017), Climate bonds, 2016 highlights, Available online at: https://www.climatebonds.net/2017/01/climate-bonds-2016-highlights-big-issuers-big-numbers-trends-count-and-2017-forecast

104 Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T. (2016), The Flash Crash: High Frequency Trading in an Electronic Market. In Journal of Finance, Forthcoming, July. 1 - 43

105 Kleftouri, N. (2015), Deposit protection system's design and effectiveness, Journal of

Banking Regulation, 16(1), 51 - 63

106 Kom Samo, Y.-L., Vervuurt, A. (2016), Stochastic Portfolio Theory: A Machine Learning Perspective, Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence, 1 - 9

107 Kosa, P. (2015), Financial Time Series Trend Prediction: Statistical Learning and Granger Causality, LBBW project / Master thesis TU Dortmund in 2015 (Ref. Prof. Posch/Engelen), available on request.

108 Kurtz, L. (2002), Studies in the field of SRI investing, Working Paper

109 Kurtz, L. (2005), Answers to four questions, Journal of Investing, 14(3), 125 - 139

110 Kurtz, L. (2008), Socially responsible investment and shareholder activism. In Crane, McWilliams, Matten, Moon & Siegel (Eds), The Oxford handbook of corporate social responsibility, 249 – 280, Oxford

111 Le Maux, J., Le Saout, E. (2004), The performance of sustainability indexes, Finance India, 18, 737 – 750

112 Lele, S., Richtsmeier, J. (2001), An invariant approach to statistical analysis of shapes, Chapman & Hall/CRC, Boca Raton, Fla, USA

113 Lessmann, S., Baesens, B., Seow, H.-V., Thomas, L.-C. (2015), Benchmarking state–of-the-art classification algorithms for credit scoring: An update of research. November, Volume 247, 1, 124 - 136

114 Li, B., Hoi, S. (2012), Online Portfolio Selection: A Survey, In ACM Computing Surveys, December, V, N, Article A

115 Li, B., Yu, J., Zhang, J., Ke, B. (2015), Detecting Accounting Frauds in Publicly Traded U.S. Firms: A Machine Learning Approach. In JMLR: Workshop and Conference Proceedings, 45, 173 - 188

116 Liao, T. (2005), Clustering of time series data – a survey, Pattern Recognition 38, 1857 - 1874

117 Locklin, S. (2017), Neglected machine learning ideas, Available online at: https://scottlocklin.wordpress.com/2014/07/22/neglected-machine-learning-ideas/

118 Lütkepohl, H. (2011), Vector autoregressive models, EUI Working Papers, ECO 2011/30, ISSN 1725-6704, European University Institute, Florence, Springer.

119 Mancuso, M. (2017), Green Bond Pricing in the Primary Market - Is There a Greenium?, Green Bond Pricing in the Primary Market: January 2016 - March 2017, Climate Bonds Initiative, Available online at:

https://www.climatebonds.net/2017/09/green-bond-pricing-primary-market-there-greenium-latest-analysis-climate-bonds

120 Mandelbrot, B. (1987), Fractals, Encyclopedia of Physical Science of Physical Science and Technology 5, 579 – 593, Academic Press, New York

121 Mandelbrot, B., Hudson, R. (2004), The (mis)behaviour of markets: A fractal view of financial turbulence, New York, NY: Basic Books

122 Markowitz, H. (1952), Portfolio Selection, Journal of Finance 7 (1): 77 – 91, doi:10.2307/2329297

123 MATLABP (2017), Machine Learning with MATLAB, Applying Supervised Learning, Available online at: https://uk.mathworks.com/content/dam/mathworks/tag-team/Objects/i/90221_80827v00_machine_learning_section4_ebook_v03.pdf

124 McGregor, R., Stanley, A. (2014), Banks pay out $100bn in US fines, Financial Times, Available online at: https://www.ft.com/content/802ae15c-9b50-11e3-946b-00144feab7de

125 Merton, R. (1987), A Simple Model of Capital Market Equilibrium with Incomplete Information, The Journal of Finance, 42, 3, 1540 - 6261doi:10.1111/j.1540-6261.1987.tb04565.x, Available online at: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1987.tb04565.x/full

126 Moayedi (2008), Arima model for network traffic prediction and anomaly detection, Available online at: http://ieeexplore.ieee.org/abstract/document/4631947/

127 Moody's (2012), Special Comment: Moody's European Covered Bonds Monitoring Overview: Q2 2012, Technical report, Moody's Investor Service, London

128 Mooij, J.M., Hoyer, P.O., Janzing, D., Schoelkopf, B. (2009), Nonlinear causal discovery with additive noise models, Advances in neural information processing systems (NIPS*2008), 689 - 696

129 Moskowitz, T., Ooi, Y., Pedersen, L. (2012), Time series momentum, Journal of Financial Economics, 104(2), 228 – 250

130 MSCI (2011), Available online at: https://www.msci.com/eqb/methodology/meth_docs/MSCI_KLD_400_Index_May2011.pdf

131 MSCI ESG Rating (2017), Available online at: https://www.msci.com/eqb/methodology/meth_docs/MSCI_KLD_400_Index_May2011.pdf

132 Nikmanesh, L., Mohd Nor, A., Sarmidi, T., Janor, H. (2014), Return and Volatility

Spillovers Between the US, Japanese and Malaysian Stock Markets, Jurnal Pengurusan 41(2014) 101 – 111, Availabel online at: http://journalarticle.ukm.my/8373/1/7770-20204-1-SM.pdf

133 Nohria, N., Piper, T., Gurtler, B. (2006), Malden Mills (A), Harvard Business School Case 404-072, December 2003. (Revised August 2006.), Available online at: http://www.hbs.edu/faculty/Pages/item.aspx?num=30633

134 OECD (2017), Available online at: https://www.oecd.org/environment/cc/Green%20bonds%20PP%20%5Bf3%5D%20%5Blr%5D.pdf

135 Oekom (2014), Corporate Responsibility Review 2014 – compact version, Taking stock of sustainability performance in corporate management, Oekom Research

136 Oekom (2017), Corporate Responsibility Review 2017, Global Transformation Process – Are Companies Already On The Right Track?, Oekom Research

137 Orlitzky, M. (2013), Corporate Social Responsibility, Noise, and Stock Market Volatility, Academy of Management Perspectives 27 (3): 238 – 254, doi:10.5464/amp.2012.0097

138 Orlitzky, M. (2015), The politics of corporate social responsibility or: why Milton Friedman has been right all along, Annals in Social Responsibility, 1, 5 – 29, doi:10.1108/ASR-06-2015-0004

139 Pallavicini, A., Perini, D., Brigo, D. (2011), Funding Valuation Adjustment: a consistent framework including CVA, DVA, collateral, netting rules and re-hypothecation, Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1969114

140 Pascual, L., Romo, J., Ruiz, E. (2004), Bootstrap predictive inference for ARIMA processes, Journal of Time Series Analysis 25: 449 - 465

141 Perng, C., Wang, H., Zhang, S., Parker, D. (2000), Landmarks: a new model for similarity-based pattern querying in time series databases, in Proceedings of the 16th International Conference on Data Engineering, ICDE ‘00, IEEE Computer Society, 33 - 42

142 Peters, J., Mooij, J.M., Janzing, D., Schoelkopf, B. (2014), Causal discovery with continuous additive noise models, Journal of Machine Learning Research 15 (2014) 2009-2053 Submitted 7/13; Revised 4/14; Published 6/14

143 Posch, P., Erhardt, J., Hard, T. (2014), The impact of commodity finance on resource availability, Applied Economics Letters, doi:10.1080/13504851.2014.955166

144 PRI (2010), Principles of Responsible Investment Association & United Nations Environment Programme Finance Initiative, Universal ownership: Why environmental externalities matter to institutional investors, London, England

145 PRI (2015), The Principles of Responsible Investment – Report on Progress 2015

146 PRI (2016), The Principles of Responsible Investment – Report on Progress 2016

147 Ranjan Das, S. (2016), Big Data's Big Muscle, In Finance and Development, September, 26 - 28

148 Renneboog, L., Ter Horst, J., Zhang, C. (2008), Socially responsible investments: institutional aspects, performance and investor behaviour, Journal of Banking & Finance, 32, 1723 - 1742

149 Revelli, C., Viviani (2013), The Link Between SRI and Financial Performance: Effects and Moderators, Management International 17 (2): 105 – 122, doi:10.7202/1015403ar

150 Revelli, C., Viviani (2015), Financial Performance of Socially Responsible Investing (SRI): What Have We Learned? A Meta-Analysis, Business Ethics: A European Review 24 (2): 158 – 185, doi:10.1111/beer.12076

151 Rose, M. (2017), An enterprise guide to big data in cloud computing, Available online at: http://searchcloudcomputing.techtarget.com/essentialguide/An-enterprise-guide-to-big-data-in-cloud-computing

152 Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.S. (1998), Boosting the margin: a new explanation for the effectiveness of voting methods, The Annals of Statistics, 1998, 26, 5, 1651 - 1686

153 Schmidt, F., Hunter, J. (2015), Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, 3rd ed., Newbury Park: Sage

154 Schmidt-Eisenlohr, T. (2013), Towards a theory of trade finance, Journal of International Economics, 91, 96 – 112. doi:10.1016/j.jinteco.2013.04.005

155 Schumacher, B. (2014), Financial Time Series Trend Prediction - An Invariant Shape Analysis Adaption, LBBW project / Master thesis University Ulm in 2014 (Ref. Erhardt/Prof. Posch), available on request.

156 Sculley, D., Phillips, T., Ebner, D., Chaudhary, V., Young, M. (2014), Machine learning: The high-interest credit card of technical debt, In SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop), Available online at: http://www.eecs.tufts.edu/~dsculley/papers/technical-debt.pdf

157 Shiller, R. (1981), Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?, The American Economic Review, Vol. 71, No. 3, 1981

158 Shiller, R. (2003), From Efficient Markets Theory to Behavioral Finance, Journal of Economic Perspectives, Vol. 17, No. 1, 83 - 104, 2003

159 Shiller, R. (2005), Irrational exuberance, 2nd ed., Princeton, NJ: Princeton University Press

160 Sidek, M., Abdul-Rahman, A., Zahirah, N. (2011), Spill-over effect of US Sub-Prime Crisis on ASEAN-5 stock markets, International Review of Business Research Papers

161 Slager, R., Chapple, W. (2015), Carrot and stick? The role of financial market intermediaries in corporate social performance, Business & Society

162 Smith, L. N., Topin, N. (2016), Deep convolutional neural network design patterns. arXiv preprint arXiv:1611.00847, Submitted to the International Conference on Learning Representations (ICLR) and rejected, 2017, Available online at: https://pdfs.semanticscholar.org/8863/9a6e21a8a8989e6d25e44119a90ba0b27628.pdf

163 Sok-Gee, Ch., Karim, M. (2010), Bank Efficiency and Non-Performing Loans: Evidence from Malaysia and Singapore, doi:10.18267/j.pep.367, Available online at: *https://ideas.repec.org/a/prg/jnlpep/v2010y2010i2id367p118-132.html#cites*

164 Stanford Course (2014), Available online at: *https://www.coursera.org/learn/machine-learning/lecture/olRZo/unsupervised-learning*

165 Toda, H. Y., Yamamoto, T. (1995), Statistical inference in vector autoregressions with possibly integrated processes, Journal of econometrics, 66(1): 225 – 250

166 Villa, J. (2015), Views of Banking Ethics, In: Ethics in Banking, Palgrave Macmillan Studies in Banking and Financial Institutions, Palgrave Macmillan, London

167 Wei, J., Yuan, X., Hu, Q., Wang, S. (2010), A novel measure for evaluating classifiers, Expert Systems with Applications, 37 (2010), 3799 – 3809, Available online at: *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.708.1668&rep=rep1&type=pdf*

168 Wold, S., Esbensen, K., Geladi, P. (1987), Principal component analysis, Chemometrics and Intelligent Laboratory Systems, 2, 37 – 52

169 Wolters, M. (2014), Trading Supervision: Anticipation of Major Market Reversals - Algorithmic Implementation Using Elliott Wave Theory, LBBW project / Master thesis University Ulm in 2013/14 (Ref. Erhardt/Prof. Posch), available on request.

170 WWF (2010), Zoological Society of London & Global Footprint Network, Living Planet Report 2010: Biodiversity, biocapacity and development, London, England: World Wide Fund for Nature

171 WWF (2016), Zoological Society of London & Global Footprint Network, Living Planet Report 2016, London, England: World Wide Fund for Nature
172 Yaacob, (2010), ARIMA Based Network Anomaly Detection, Available online at: *http://ieeexplore.ieee.org/abstract/document/5437603/*

8. Appendix

8.1. Appendix on Elliott Wave Theory[160]

8.1.1. Reversal prices and assigned probabilities

Table 21: Calculation of reversal prices based on Fibonacci multiples.
Subscript numbers denote the price at the respective position of the wave.
Formulas are based on Frost and Prechter (2005).

Wave Pattern	Reversal Price RP^1	Reversal Price RP^2	Reversal Price RP^3
Impulse	$P_4 + 0.618 \cdot (P_4 - P_0)$	$P_3 + 0.618 \cdot (P_3 - P_0)$	
Diagonal Triangle	$P_4 + 0.618 \cdot (P_4 - P_0)$	$P_3 + 0.618 \cdot (P_3 - P_0)$	
Wave 1 Extension	$P_6 + 0.618 \cdot (P_6 - P_0)$	$P_5 + 0.618 \cdot (P_5 - P_0)$	
Wave 3 Extension	$P_8 + 1.0 \cdot (P_1 - P_0)$		
Wave 5 Extension	$P_4 + 1.618 \cdot (P_4 - P_0)$	$P_3 + 1.618 \cdot (P_3 - P_0)$	
Zigzag	$P_2 + 0.618 \cdot (P_1 - P_0)$	$P_2 + 1.0 \cdot (P_1 - P_0)$	$P_2 + 1.618 \cdot (P_1 - P_0)$
Flat	$P_2 + 1 \cdot (P_1 - P_0)$	$P_2 + 1.618 \cdot (P_1 - P_0)$	
Contr. Triangle	$P_4 + 0.618 \cdot (P_3 - P_2)$		
Expand. Triangle	$P_4 + 1.618 \cdot (P_3 - P_2)$		

Table 22: Assigned probabilities for alternative multiples,
as introduced in Table 21

Wave Pattern	Prob. p_1 for RP^1	Prob. p_2 for RP^2	Prob. p_3 for RP^3
Impulse	0.5	0.5	-
Diagonal Triangle	0.5	0.5	-
Wave 1 Extension	0.5	0.5	-
Wave 3 Extension	1	-	-
Wave 5 Extension	0.5	0.5	-
Zigzag	0.5	0.25	0.25
Flat	0.75	0.25	-
Contracting Triangle	0.5	0.5	-
Expanding Triangle	0.5	0.5	-

[160] This Appendix section is extracted from / summarising a bank project (Wolters, 2014), see also FN 47.

8.1.2. Simplifying Assumptions

The applied EW Algorithm implements the following simplifying Assumptions:

1. Non-usage of sideways patterns: Wave patterns can be associated with an up or down trend. However, combinations are horizontal in character. They cover less price movement than other patterns and are therefore less useful in predicting the end of a trend. Hence they are not implemented.
2. Arbitrary starting points of waves: According to EW every wave consists of sub waves. Vice versa every found wave must be the sub wave of a larger wave. Its starting and endpoint are therefore restricted to the pivot points given by that larger wave. Intertwining waves result in a consistent wave count. However, the developed algorithm is only restricted to local minima and maxima as starting or ending points.
3. Weakening of the degree concept: A major top or bottom is the ending point of several waves of different degrees. To account for wave degrees, the length of a subsequently checked interval of price data is determined by the length of the last sub wave of the preceding wave. However, the Elliott Wave Algorithm successively shortens the analysed intervals by a constant factor *SF*. Therefore, found waves of successive intervals are of different lengths but not necessarily of different degrees. The considerable advantage is that an erroneous classification on the largest degree of trend does not influence the possible interpretations of smaller degrees.
4. Waves with a downtrend are corrective waves: In general dynamic waves support the direction of the trend of one larger degree, while corrective waves retrace it in parts. This is simplified by assuming that a downturn is more likely to unfold as a corrective wave and an upturn as a dynamic wave - as long as the largest trend observable is up. This is the case for most analysed indices. If the largest trend is a downtrend the roles of corrective and motive waves have to be inverted.

5. Order of preference for wave patterns are predefined to avoid ambiguity:

Table 23: Order of preference for wave patterns in 'up- and down-trends'.

Preference	Uptrend	Downtrend
1	Impulse	Zigzag
2	Extension	Flat
3	Diagonal Triangle	Contracting Triangle
4	Zigzag	Expanding Triangle
5	Flat	Double Zigzag
6	Contracting Triangle	Triple Zigzag
7	Expanding Triangle	Impulse
8	Double Zigzag	Extension
9	Triple Zigzag	Diagonal Triangle

6. Linear extrapolation of wave patterns: Once a wave pattern is finished, its alternative reversal price based on Fibonacci multiples can be calculated. Wave patterns are assumed to unfold linearly, maintaining the slope of the sub wave as the waves consist of straight line segments.
7. Exclusion of complex rules/guidelines: The manifold complex rules and guidelines of EW are simply not implemented into the analysis. Wave patterns are assessed separately and independently.

8.1.3. Outline of the EW Algorithm

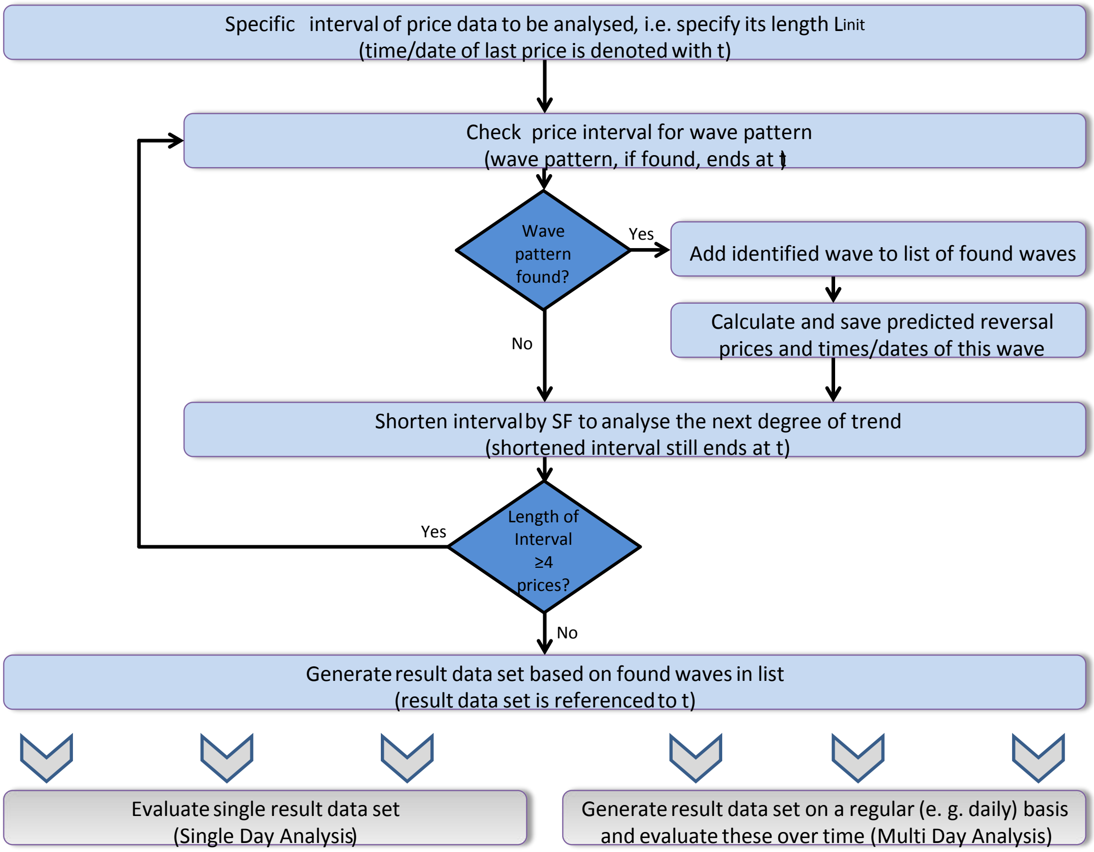


Figure 24: Outline of the EW Algorithm.
An interval of price data is continuously shortened and checked for wave patterns. The waves generate a result data set. Result data sets can be evaluated individually (Single Day Analysis) or be compared over time (Multi Day Analysis).

8.1.4. List of EW Features

Table 24: EW-Features used as modelling set

File	Labelling
RollingWindow_1500_Coverage_Dwn	Coverage Indicator
RollingWindow_1500_Coverage_Up	Coverage Indicator
RollingWindow_1500_NrOfWaves	Number of Waves Indicator
RollingWindow_1500_NrOfWaves_x_Coverage_Dwn	Number of Waves x Coverage Indicator
RollingWindow_1500_NrOfWaves_x_Coverage_Up	Number of Waves x Coverage Indicator
RollingWindow_1500_ReversalDensityBottom_ NotWeighted	Reversal Time Density Indicator
RollingWindow_1500_ ReversalDensityBottom_ Weighted	Reversal Time Density Indicator
RollingWindow_1500_ ReversalDensityTop_NotWeighted	Reversal Time Density Indicator
RollingWindow_1500_ ReversalDensityTop_Weighted	Reversal Time Density Indicator

Table 25 (cont.): EW-Features used as modelling set

File	Labelling
RollingWindow_1500_ReversalPriceW_Dwn	Average predicted reversal price indicator / Average reversal price and time (weighted)
RollingWindow_1500_ReversalPriceW_Up	Average predicted reversal price indicator / Average reversal price and time (weighted)
RollingWindow_1500_ReversalPriceUnw_Dwn	Average predicted reversal price indicator / Average reversal price and time (unweighted)
RollingWindow_1500_ReversalPriceUnw_Up	Average predicted reversal price indicator / Average reversal price and time (unweighted)
RollingWindow_1500_SurpassedReversals_Dwn	Surpassed reversals indicator
RollingWindow_1500_SurpassedReversals_Up	Surpassed reversals indicator
RollingWindow_1500_TmDistncToRvrsl_Dwn	Average predicted reversal time indicator
RollingWindow_1500_TmDistncToRvrsl_Up	Average predicted reversal time indicator

8.2. Appendix on Trend Prediction

8.2.1. Granger List

Table 26: List of Leading Markets – out of 300 assets - based on Granger Causality test: With 214 items for NKY, 99 for DAX and only 16 for SPX Index

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
RR_LN EQUITY_abs_cls_sm	AAPL US EQUITY_abs_cls_sm	5108 JP EQUITY_abs_cls_sm
6501 JP EQUITY_abs_cls_sm	ADBE US EQUITY_abs_cls_sm	7751 JP EQUITY_abs_cls_sm
6502 JP EQUITY_abs_cls_sm	ADP US EQUITY_abs_cls_sm	9766 JP EQUITY_abs_cls_sm
6773 JP EQUITY_abs_cls_sm	AIG US EQUITY_abs_cls_sm	BET INDEX_abs_cls_sm
7201 JP EQUITY_abs_cls_sm	ALV GR EQUITY_abs_cls_sm	CEC GR EQUITY_abs_cls_sm
7269 JP EQUITY_abs_cls_sm	ALXN US EQUITY_abs_cls_sm	EBAY US EQUITY_abs_cls_sm
7751 JP EQUITY_abs_cls_sm	AMAT US EQUITY_abs_cls_sm	FIE GR EQUITY_abs_cls_sm
8058 JP EQUITY_abs_cls_sm	AUDEUR CURNCY_abs_cls_sm	GS US EQUITY_abs_cls_sm
AABA US Equity_abs_cls_sm	AXP US EQUITY_abs_cls_sm	HNR1 GR EQUITY_abs_cls_sm
AAPL US EQUITY_abs_cls_sm	BA US EQUITY_abs_cls_sm	IIA AV EQUITY_abs_cls_sm
ABBN VX EQUITY_abs_cls_sm	BAC US EQUITY_abs_cls_sm	KRN GR EQUITY_abs_cls_sm
AC FP EQUITY_abs_cls_sm	BAS GR EQUITY_abs_cls_sm	MMM US EQUITY_abs_cls_sm
ADBE US EQUITY_abs_cls_sm	BAYN GR EQUITY_abs_cls_sm	RHM GR EQUITY_abs_cls_sm
ADEN VX EQUITY_abs_cls_sm	BEI GR EQUITY_abs_cls_sm	VIV FP EQUITY_abs_cls_sm
ADP US EQUITY_abs_cls_sm	BIIB US EQUITY_abs_cls_sm	VIX INDEX_abs_cls_sm
ADS GR EQUITY_abs_cls_sm	BMW GR EQUITY_abs_cls_sm	WFC US EQUITY_abs_cls_sm
AI FP EQUITY_abs_cls_sm	BMY US EQUITY_abs_cls_sm	
ALO FP EQUITY_abs_cls_sm	BOSS GR EQUITY_abs_cls_sm	
ALU FP EQUITY_abs_cls_sm	C US EQUITY_abs_cls_sm	
ALV GR EQUITY_abs_cls_sm	CAT US EQUITY_abs_cls_sm	
ALXN US EQUITY_abs_cls_sm	CCMP INDEX_abs_cls_sm	
AMAT US EQUITY_abs_cls_sm	CELG US EQUITY_abs_cls_sm	
AMZN US EQUITY_abs_cls_sm	CMCSA US EQUITY_abs_cls_sm	
ASML NA EQUITY_abs_cls_sm	CON GR EQUITY_abs_cls_sm	
AXP US EQUITY_abs_cls_sm	COST US EQUITY_abs_cls_sm	
BA US EQUITY_abs_cls_sm	CSCO US EQUITY_abs_cls_sm	
BARC LN EQUITY_abs_cls_sm	CTSH US EQUITY_abs_cls_sm	
BAS GR EQUITY_abs_cls_sm	CVS US EQUITY_abs_cls_sm	

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
BAYN GR EQUITY_abs_cls_sm	CVX US EQUITY_abs_cls_sm	
BBVA SM EQUITY_abs_cls_sm	DAI GR EQUITY_abs_cls_sm	
BCOMCL INDEX_abs_cls_sm	DD US EQUITY_abs_cls_sm	
BCOMCO INDEX_abs_cls_sm	DIS US EQUITY_abs_cls_sm	
BCOMNG INDEX_abs_cls_sm	DTE GR EQUITY_abs_cls_sm	
BEI GR EQUITY_abs_cls_sm	DUE GR EQUITY_abs_cls_sm	
BET INDEX_abs_cls_sm	EBS AV EQUITY_abs_cls_sm	
BIIB US EQUITY_abs_cls_sm	EOAN GR EQUITY_abs_cls_sm	
BLT LN EQUITY_abs_cls_sm	ESRX US EQUITY_abs_cls_sm	
BMW GR EQUITY_abs_cls_sm	FOX US EQUITY_abs_cls_sm	
BMY US EQUITY_abs_cls_sm	GBPEUR CURNCY_abs_cls_sm	
BNP FP EQUITY_abs_cls_sm	GBPUSD CURNCY_abs_cls_sm	
BOSS GR EQUITY_abs_cls_sm	GE US EQUITY_abs_cls_sm	
BUX INDEX_abs_cls_sm	GILD US EQUITY_abs_cls_sm	
CA FP EQUITY_abs_cls_sm	GS US EQUITY_abs_cls_sm	
CAC INDEX_abs_cls_sm	HD US EQUITY_abs_cls_sm	
CAI AV EQUITY_abs_cls_sm	HKDGBP CURNCY_abs_cls_sm	
CAP FP EQUITY_abs_cls_sm	HNR1 GR EQUITY_abs_cls_sm	
CAT US EQUITY_abs_cls_sm	IBOV INDEX_abs_cls_sm	
CCMP INDEX_abs_cls_sm	INDU INDEX_abs_cls_sm	
CEC GR EQUITY_abs_cls_sm	INTC US EQUITY_abs_cls_sm	
CELG US EQUITY_abs_cls_sm	INTU US EQUITY_abs_cls_sm	
CFR VX EQUITY_abs_cls_sm	JNJ US EQUITY_abs_cls_sm	
CLS1 GR EQUITY_abs_cls_sm	JPM US EQUITY_abs_cls_sm	
CMCSAUS EQUITY_abs_cls_sm	KO US EQUITY_abs_cls_sm	
CON GR EQUITY_abs_cls_sm	LIN GR EQUITY_abs_cls_sm	
COST US EQUITY_abs_cls_sm	MCD US EQUITY_abs_cls_sm	
CS FP EQUITY_abs_cls_sm	MEXBOL INDEX_abs_cls_sm	
CSCO US EQUITY_abs_cls_sm	ML FP EQUITY_abs_cls_sm	
CSGN VX EQUITY_abs_cls_sm	MMM US EQUITY_abs_cls_sm	
CTSH US EQUITY_abs_cls_sm	MO US EQUITY_abs_cls_sm	
CVS US EQUITY_abs_cls_sm	MRK US EQUITY_abs_cls_sm	
CVX US EQUITY_abs_cls_sm	MSFT US EQUITY_abs_cls_sm	
DAI GR EQUITY_abs_cls_sm	MUV2 GR EQUITY_abs_cls_sm	
DAX INDEX_abs_cls_sm	MYL US EQUITY_abs_cls_sm	

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
DAXK Index_abs_cls_sm	NKE US EQUITY_abs_cls_sm	
DBK GR EQUITY_abs_cls_sm	NYA INDEX_abs_cls_sm	
DD US EQUITY_abs_cls_sm	ORCL US EQUITY_abs_cls_sm	
DG FP EQUITY_abs_cls_sm	PCLN US EQUITY_abs_cls_sm	
DIS US EQUITY_abs_cls_sm	PEP US EQUITY_abs_cls_sm	
DTE GR EQUITY_abs_cls_sm	PFE US EQUITY_abs_cls_sm	
DUE GR EQUITY_abs_cls_sm	PSM GR EQUITY_abs_cls_sm	
EBAY US EQUITY_abs_cls_sm	QCOM US EQUITY_abs_cls_sm	
EBS AV EQUITY_abs_cls_sm	REGN US EQUITY_abs_cls_sm	
EI FP EQUITY_abs_cls_sm	ROG VX EQUITY_abs_cls_sm	
EN FP EQUITY_abs_cls_sm	RUBCAD CURNCY_abs_cls_sm	
ENEL IM EQUITY_abs_cls_sm	RUBCHF CURNCY_abs_cls_sm	
ENI IM EQUITY_abs_cls_sm	RUBEUR CURNCY_abs_cls_sm	
ESRX US EQUITY_abs_cls_sm	RUBHKD CURNCY_abs_cls_sm	
FIE GR EQUITY_abs_cls_sm	RUBJPY CURNCY_abs_cls_sm	
FME GR EQUITY_abs_cls_sm	RUBUSD CURNCY_abs_cls_sm	
FOX US EQUITY_abs_cls_sm	SAP GR EQUITY_abs_cls_sm	
FPE GR EQUITY_abs_cls_sm	SBUX US EQUITY_abs_cls_sm	
FR FP EQUITY_abs_cls_sm	SIE GR EQUITY_abs_cls_sm	
FRE GR EQUITY_abs_cls_sm	SLB US EQUITY_abs_cls_sm	
FTSEMIB INDEX_abs_cls_sm	SPX INDEX_abs_cls_sm	
G IM EQUITY_abs_cls_sm	SZG GR EQUITY_abs_cls_sm	
G1A GR EQUITY_abs_cls_sm	T US EQUITY_abs_cls_sm	
GBF GR EQUITY_abs_cls_sm	TKA GR EQUITY_abs_cls_sm	
GE US EQUITY_abs_cls_sm	TRV US EQUITY_abs_cls_sm	
GEBN VX EQUITY_abs_cls_sm	TXN US EQUITY_abs_cls_sm	
GIL GR EQUITY_abs_cls_sm	UTX US EQUITY_abs_cls_sm	
GILD US EQUITY_abs_cls_sm	VIX INDEX_abs_cls_sm	
GLE FP EQUITY_abs_cls_sm	VOW3 GR EQUITY_abs_cls_sm	
GS US EQUITY_abs_cls_sm	VRTX US EQUITY_abs_cls_sm	
GWI1 GR EQUITY_abs_cls_sm	VZ US EQUITY_abs_cls_sm	
HD US EQUITY_abs_cls_sm	WBA US EQUITY_abs_cls_sm	
HEI GR EQUITY_abs_cls_sm	WFC US EQUITY_abs_cls_sm	
HEN3 GR EQUITY_abs_cls_sm	WMT US EQUITY_abs_cls_sm	
HNR1 GR EQUITY_abs_cls_sm	XOM US EQUITY_abs_cls_sm	

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
HOT GR EQUITY_abs_cls_sm	AABA US Equity_abs_cls_sm	
HSI INDEX_abs_cls_sm		
IBEX INDEX_abs_cls_sm		
IBM US EQUITY_abs_cls_sm		
IBOV INDEX_abs_cls_sm		
IIA AV EQUITY_abs_cls_sm		
INDU INDEX_abs_cls_sm		
INGA NA EQUITY_abs_cls_sm		
INTC US EQUITY_abs_cls_sm		
INTU US EQUITY_abs_cls_sm		
IPSA INDEX_abs_cls_sm		
ISP IM EQUITY_abs_cls_sm		
JCI INDEX_abs_cls_sm		
JNJ US EQUITY_abs_cls_sm		
JPM US EQUITY_abs_cls_sm		
KER FP EQUITY_abs_cls_sm		
KO US EQUITY_abs_cls_sm		
KRN GR EQUITY_abs_cls_sm		
KU2 GR EQUITY_abs_cls_sm		
LEO GR EQUITY_abs_cls_sm		
LHA GR EQUITY_abs_cls_sm		
LHN VX EQUITY_abs_cls_sm		
LIN GR EQUITY_abs_cls_sm		
LLOY LN EQUITY_abs_cls_sm		
LNZ AV EQUITY_abs_cls_sm		
MAN GR EQUITY_abs_cls_sm		
MC FP EQUITY_abs_cls_sm		
MCD US EQUITY_abs_cls_sm		
MDAX INDEX_abs_cls_sm		
MERVAL INDEX_abs_cls_sm		
MEXBOL INDEX_abs_cls_sm		
ML FP EQUITY_abs_cls_sm		
MMM US EQUITY_abs_cls_sm		
MO US EQUITY_abs_cls_sm		
MRK GR EQUITY_abs_cls_sm		

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
MRK US EQUITY_abs_cls_sm		
MSFT US EQUITY_abs_cls_sm		
MUV2 GR EQUITY_abs_cls_sm		
MYL US EQUITY_abs_cls_sm		
NDA GR EQUITY_abs_cls_sm		
NESN VX EQUITY_abs_cls_sm		
NIFTY INDEX_abs_cls_sm		
NKE US EQUITY_abs_cls_sm		
NOKIA FH EQUITY_abs_cls_sm		
NOVN VX EQUITY_abs_cls_sm		
NYA INDEX_abs_cls_sm		
OMV AV EQUITY_abs_cls_sm		
OR FP EQUITY_abs_cls_sm		
ORCL US EQUITY_abs_cls_sm		
PCLN US EQUITY_abs_cls_sm		
PEP US EQUITY_abs_cls_sm		
PFE US EQUITY_abs_cls_sm		
PG US EQUITY_abs_cls_sm		
PSM GR EQUITY_abs_cls_sm		
PUB FP EQUITY_abs_cls_sm		
QCOM US EQUITY_abs_cls_sm		
REGN US EQUITY_abs_cls_sm		
REP SM EQUITY_abs_cls_sm		
RHI AV EQUITY_abs_cls_sm		
RHK GR EQUITY_abs_cls_sm		
RHM GR EQUITY_abs_cls_sm		
RI FP EQUITY_abs_cls_sm		
RNO FP EQUITY_abs_cls_sm		
ROG VX EQUITY_abs_cls_sm		
RRTL GR EQUITY_abs_cls_sm		
RWE GR EQUITY_abs_cls_sm		
SAF FP EQUITY_abs_cls_sm		
SAN FP EQUITY_abs_cls_sm		
SAN SM EQUITY_abs_cls_sm		
SAP GR EQUITY_abs_cls_sm		

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
SAZ GR EQUITY_abs_cls_sm		
SBO AV EQUITY_abs_cls_sm		
SBUX US EQUITY_abs_cls_sm		
SCMN VX EQUITY_abs_cls_sm		
SDF GR EQUITY_abs_cls_sm		
SGO FP EQUITY_abs_cls_sm		
SGSN VX EQUITY_abs_cls_sm		
SHCOMP INDEX_abs_cls_sm		
SIE GR EQUITY_abs_cls_sm		
SKY LN EQUITY_abs_cls_sm		
SLB US EQUITY_abs_cls_sm		
SPR GR EQUITY_abs_cls_sm		
SPX INDEX_abs_cls_sm		
SREN VX EQUITY_abs_cls_sm		
SU FP EQUITY_abs_cls_sm		
SX5E INDEX_abs_cls_sm		
SZG GR EQUITY_abs_cls_sm		
SZU GR EQUITY_abs_cls_sm		
TEC FP EQUITY_abs_cls_sm		
TKA GR EQUITY_abs_cls_sm		
TOP40 INDEX_abs_cls_sm		
TRV US EQUITY_abs_cls_sm		
TXN US EQUITY_abs_cls_sm		
UCG IM EQUITY_abs_cls_sm		
UHR VX EQUITY_abs_cls_sm		
UKX INDEX_abs_cls_sm		
UNA NA EQUITY_abs_cls_sm		
UNH US EQUITY_abs_cls_sm		
UTX US EQUITY_abs_cls_sm		
VER AV EQUITY_abs_cls_sm		
VIG AV EQUITY_abs_cls_sm		
VIV FP EQUITY_abs_cls_sm		
VIX INDEX_abs_cls_sm		
VOD LN EQUITY_abs_cls_sm		
VOE AV EQUITY_abs_cls_sm		

Granger Causality list NKY Index	Granger Causality list DAX Index	Granger Causality list SPX Index
VOW3 GR EQUITY_abs_cls_sm		
VRTX US EQUITY_abs_cls_sm		
VZ US EQUITY_abs_cls_sm		
WBA US EQUITY_abs_cls_sm		
WFC US EQUITY_abs_cls_sm		
WIE AV EQUITY_abs_cls_sm		
WIG INDEX_abs_cls_sm		
WMT US EQUITY_abs_cls_sm		
XOM US EQUITY_abs_cls_sm		
XU100 INDEX_abs_cls_sm		
ZURN VX EQUITY_abs_cls_sm		

8.2.2. XGBoost Model Results[161]

We show the different modelling results for 6 global stock market indices – NKY, HSI, DAX, UKX, SPX and INDU. We start with the Nikkei index in Japan.

A technical remark: We calculate the performance on the maximal possible timeframe for each learning algorithm. We need different pre-periods for separate feature sets. Thus the timeframe varies slightly between distinct model performance calculations. Hence, the statistics for the indices fluctuate to some extent – see, for example, Sharpe Ratio and Maximum Drawdown[162].

8.2.2.1. NKY Index, Japan

(1) Single time series calculation with the 'Classic Features' set: For this model the MCC is calculated as MCC = 0.07 with Chi-squared = 61.51. Below we show the Confusion Matrix and the performance of the formalised investment strategy compared to a long only benchmark in the underlying index. The coefficients and the formalised investment strategy are calculated on a COB basis.[163] For the investment strategy we assume an outright long position for every '0' and '1' classification. In our results we do not correct for bid-offer costs.

Table 27: Confusion Matrix NKY - Single time series with 'Classic Features'.

		Predicted Class		
		-1	0	1
Correct Class	-1	279	337	408
	0	290	572	380
	1	313	376	466

[161] Calculations and charts were performed as part of the bank project, see also FN 30.

[162] The benchmark is evaluated each time on the days we use for calculating the values for the learning algorithms. There is an analogous effect of lesser degree (-> negligible) in regard to the MCC calculations when time frames differ.

[163] The features are calculated based on a single time series. Thus it is not necessary to differentiate between different closing times in different markets / regions.

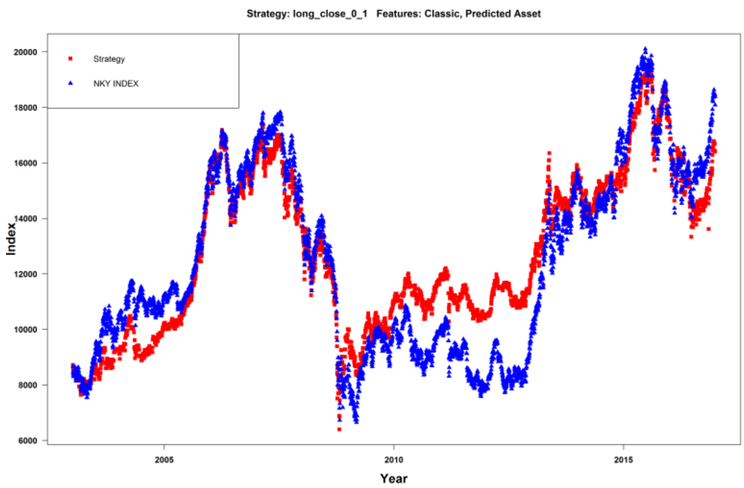


Figure 25: Investment Performance NKY - Single time series with ‘Classic Features’.
Model (Red) versus benchmark (Blue).

(2) Single time series calculation with ‘Landmarks’: MCC result with MCC = 0.04 and Chi-squared = 23.16 -> slightly positive, but no improvement.

Table 28: Confusion Matrix NKY – Single time series with ‘Landmarks’.

		Predicted Class		
		-1	0	1
Correct Class	-1	295	337	371
	0	335	518	372
	1	346	399	399

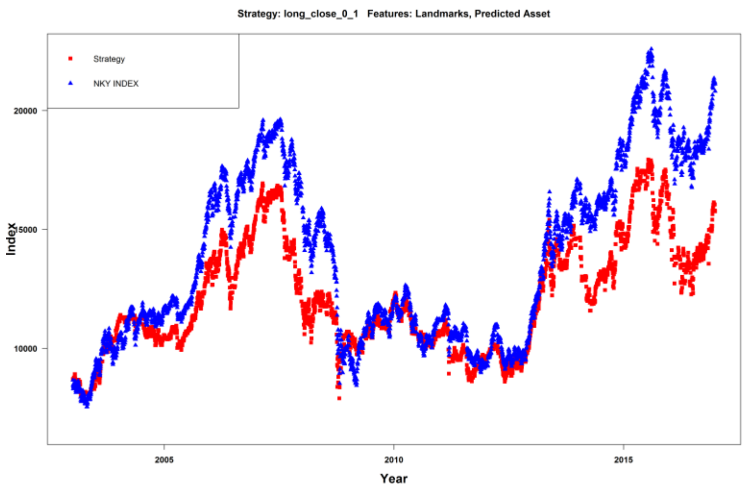


Figure 26: Investment Performance NKY - Single time series with ‘Landmarks’.
Model (Red) versus benchmark (Blue).

(3) Single time series calculation with ‘EW-Features‘: MCC result with MCC = 0.05 and Chi-squared with 22.30 -> slightly positive. The strategy looks less volatile.

Table 29: Confusion Matrix NKY – Single time series with ‘EW Features’.

		Predicted Class		
		-1	0	1
Correct Class	-1	156	306	235
	0	129	383	232
	1	186	304	265

Figure 27: Investment Performance NKY - Single time series with ‘EW-Features‘.
Model (Red) versus benchmark (Blue).

(4) Multi-asset calculation with Granger Causality and ‘Classic Features‘: see section 2.3.7.1.

(5) Multi-asset calculation with Granger Causality and ‘Landmarks‘: With MCC = 0.19 and Chi-squared = 276.65.

Table 30a, 30b: Confusion Matrix & Performance Statistics NKY – Multi-asset with Granger and ‘Landmarks’.

		Predicted Class		
		-1	0	1
Correct Class	-1	447	271	206
	0	286	476	359
	1	195	365	500

	Maximum Drawdown	Sharpe Ratio
Index	8974	0.40
long 0_1	2917	1.01

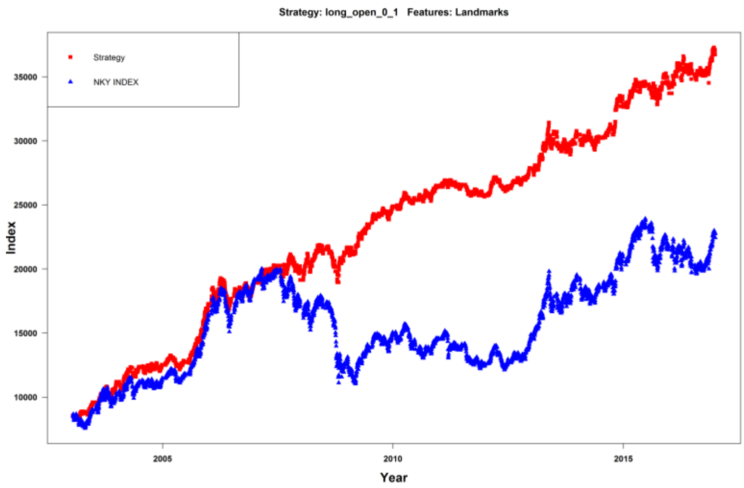


Figure 28: Investment Performance NKY - Multi-asset with Granger and 'Landmarks'. Model (Red) versus benchmark (Blue).

⇨ Quick take: Again massive improvement over benchmark and the Single time series models, but not necessarily over the combination of Granger & Classic Features. The Sharpe Ratio is with 1.01 slightly worse, but the Maximum Drawdown is better. The Confusion Matrix shows again a fairly high number of TPs for all three categories. Overall the strategy looks less profitable, but also less risky.

(6) Multi-asset calculation with Granger Causality and 'EW-Features': With MCC = 0.08 and Chi-squared = 52.72

Table 31a, 31b: Confusion Matrix & Performance Statistics NKY – Multi-asset with Granger and 'EW-Features'.

		Predicted Class		
		-1	0	1
Correct Class	-1	235	183	279
	0	145	297	302
	1	172	256	327

	Maximum Drawdown	Sharpe Ratio
Index	7465	0.15
long 0_1	6349	0.01

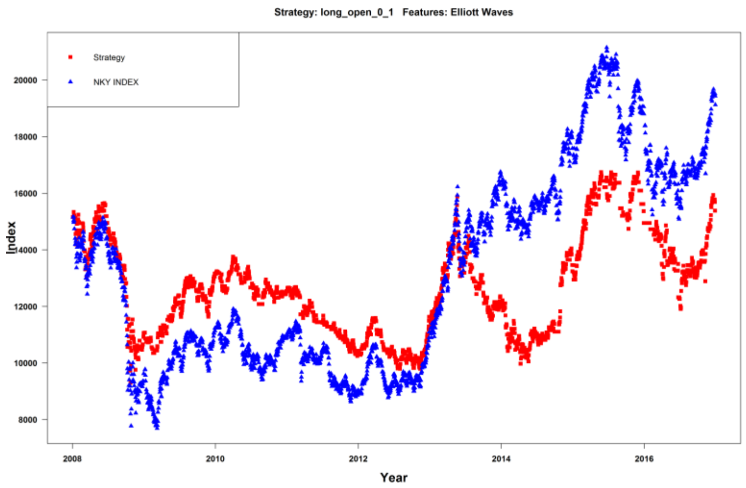


Figure 29: Investment Performance NKY - Multi-asset with Granger and 'EW-Features'. Model (Red) versus benchmark (Blue).

(7) a) Multi-asset calculation **with** Granger Causality and 'Residuals'[164]: MCC = 0.17 and Chi-squared = 257.09 -> excellent, but less so than Granger & Classic Features or Granger & Landmarks.

Table 32a, 32b: Confusion Matrix & Performance Statistics NKY – Multi-asset with Granger and 'Residuals'.

		Predicted Class		
		-1	0	1
Correct Class	-1	517	263	263
	0	445	399	409
	1	258	268	610

	Maximum Drawdown	Sharpe Ratio
Index	11210	0.28
long 0_1	3178	0.63

[164] VEC(p), ARIMA, AIMAX and GARCH 2,2. For more see section 2.3.6.3.

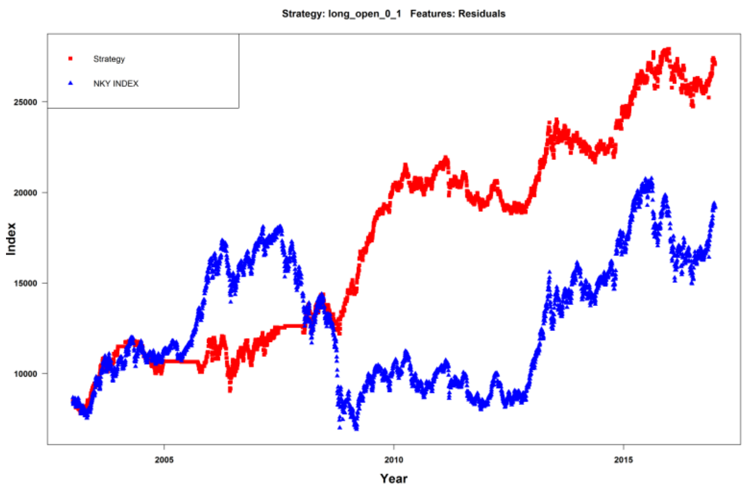


Figure 30: Investment Performance NKY - Multi-asset with Granger and ‘Residuals‘. Model (Red) versus benchmark (Blue).

(7) b) Multi-asset calculation with ‘Residuals‘, but **without** Granger Causality test:

MCC = 0.15 and Chi-squared = 171.89

Table 33a, 33b: Confusion Matrix & Performance Statistics NKY – Multi-asset without Granger and ‘Residuals’.

		Predicted Class		
		-1	0	1
Correct Class	-1	443	209	373
	0	386	332	529
	1	246	222	692

	Maximum Drawdown	Sharpe Ratio
Index	11210	0.27
long 0_1	3779	0.63

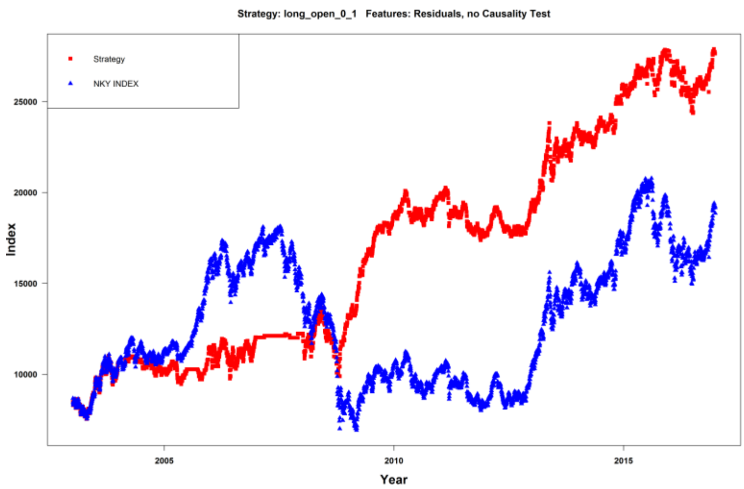


Figure 31: Investment Performance NKY - Multi-asset without Granger and 'Residuals'. Model (Red) versus benchmark (Blue).

Quick take: Surprisingly close to Granger & Residuals. Nevertheless we see a positive impact when using Granger Causality pre-tests. The major influence seems to come straight from the 300 assets (Spill-over effects). There is some limited effect coming from the complexity reduction through the Granger pre-filter.

(8) Multi-asset calculation with Granger Causality, 'Residuals' and 'Landmarks light'[165]: With MCC = 0.18 and Chi-squared = 268.51-> small benefit in comparison to without Landmarks.

Table 34a, 34b: Confusion Matrix & Performance Statistics NKY – Multi-asset with Granger and 'Residuals' & 'Landmarks light'.

		Predicted Class		
		-1	0	1
Correct Class	-1	511	253	261
	0	423	396	428
	1	264	259	637

[165] See again 5.3.6.3. The Landmarks light are just BlockExtrema and Maximm Gain & Loss. The full Landmarks Feature set does currently 'overtax' our storage capacities.

	Maximum Drawdown	Sharpe Ratio
Index	11210	0.27
long 0_1	3031	0.86

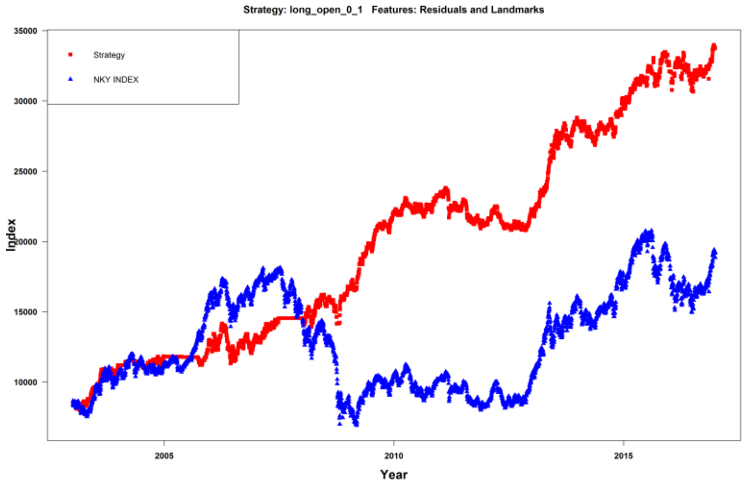


Figure 32: Investment Performance NKY - Multi-asset with Granger and ‘Residuals‘ & ‘Landmarks light‘. Model (Red) versus benchmark (Blue).

⇨ Final conclusion for the NKY Index: The best three strategies are multi-asset strategies (4), (5) and (9), Granger Causality in combination with Classical Features, Landmarks and Residuals & Landmarks light. The ‘multi-asset effect’ is not surprising when considering that Japan is the most open society from a macroeconomic sense with strong dependencies to other international markets.

8.2.2.2. HSI Index, Hong Kong

To get an impression for the second Asian index it is enough to focus on a selected few and compare them to the results for the Nikkei. We start straightaway with the Multi-asset models.

(1) Multi-asset calculation with Granger Causality and ‘Residuals‘: MCC is with 0.19 and Chi-squared = 307.50 - as we would expect it (judging from the Nikkei) - similar or even slightly better than the Nikkei (MCC = 0.17 and Chi-squared = 257.09). The Confusion Matrix looks in line with the two coefficients, with fairly good classification results for all TPs. But the investment strategy, calculated again on an OOB basis, shows total

underperformance (see also Sharpe Ratio and Maximum Drawdown). This is somewhat surprising. Therefore we repeated all the data controls, especially checking for time issues on the opening and closing prices. No issues were identified. So it may just be the case that the model performs well on the less volatile days and much worse on the volatile ones. From a macroeconomic perspective this may be explained by the specific characteristics of the Hong Kong stock exchange. The multi-asset calculations rely on Spill-over effects from other international markets. China was not, and still is not, a fully open market economy with certain anomalies (A- and B-shares, market transfer restrictions regarding the Renminbi, *etc.*). Existing transfer restrictions have eased only over time - if one looks at the examined time period. Likewise the Chinese people are just recently allowed to transfer assets from the mainland. All this are potential explanations for the effect shown below, especially as it looks that the investment strategy performs by far better after 2010.

⇨ Interesting results, opening up room for further inquiries, even from a macroeconomic perspective.

Obiter dictum: The fact that we can pick-up on these kind of inconsistencies after running through the model calculations and comparing the results to similar settings is one of the key advantages of our experimental setting for future use. Inconsistencies, for lack of a better word, are starting points for further inquiries and with it comes the potential to identify remarkable anomalies - of course under the pre-condition that the modelling can be trusted.

Table 35a, 35b: Confusion Matrix & Performance Statistics HSI – Multi-asset with Granger and 'Residuals'.

		Predicted Class		
		-1	0	1
Correct Class	-1	473	387	153
	0	405	705	254
	1	227	419	432

HSI	Maximum Drawdown	Sharpe Ratio
Index	20629	0.19
long 0_1	30123	-0.09

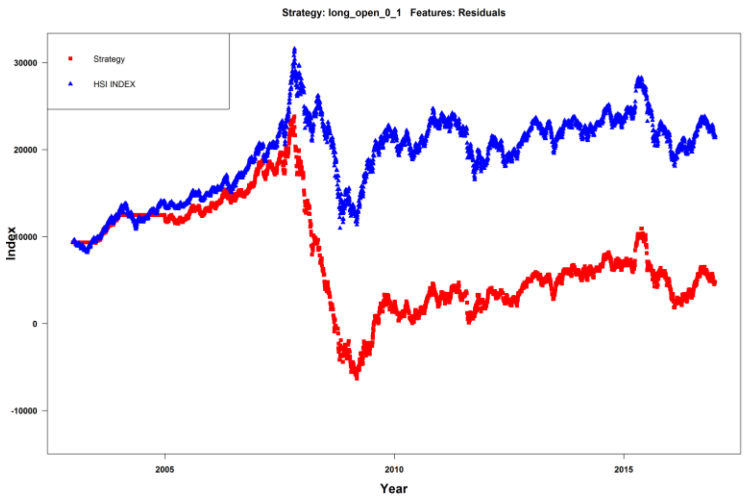


Figure 33: Investment Performance HSI - Multi-asset with Granger and 'Residuals'. Model (Red) versus benchmark (Blue).

(2) Multi-asset calculation with Granger Causality, 'Residuals' and 'Landmarks light'[166]: MCC = 0.19 and Chi-squared = 288.73, compared to Nikkei with MCC = 0.18 and Chi-squared = 268.51 -> same overall picture, see comments (1).

Table 36a, 36b: Confusion Matrix & Performance Statistics HSI – Multi-asset with Granger and 'Residuals' & 'Landmarks light'.

		Predicted Class		
		-1	0	1
Correct Class	-1	464	399	150
	0	378	730	256
	1	223	444	411

	Maximum Drawdown	Sharpe Ratio
Index	20629	0.19
long 0_1	23274	-0.08

[166] See again 5.3.6.3. The Landmarks light are just BlockExtrema and Maximm Gain & Loss. The full Landmarks Feature set does currently 'overtax' our storage capacities.

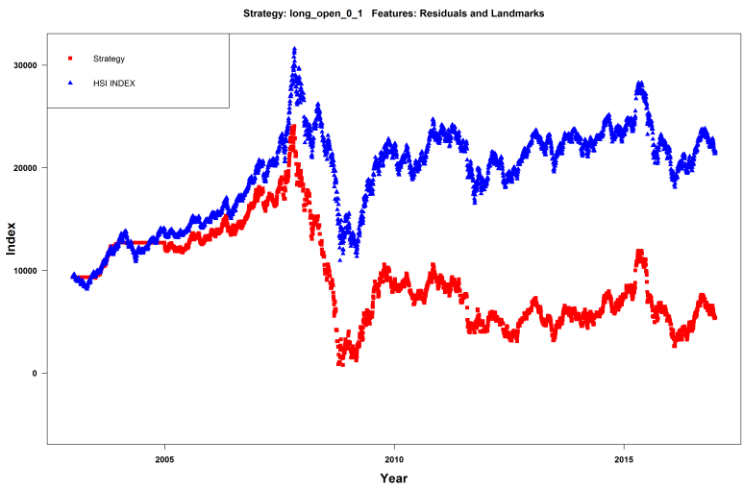


Figure 34: Investment Performance HSI - Multi-asset with Granger and 'Residuals' & 'Landmarks light'. Model (Red) versus benchmark (Blue).

8.2.2.3. DAX Index, Germany

The DAX index is the first European index and the strongest economy in Europe, representing the EURO zone as a whole.

(1) Single time series calculation with 'Classic Features': MCC = 0.08 and Chi-squared = 76.01 -> slight improvement over the benchmark, in line with Nikkei.

Table 37: Confusion Matrix DAX - Single time series with 'Classic Features'.

		Predicted Class		
		-1	0	1
Correct Class	-1	363	481	267
	0	357	748	194
	1	375	491	275

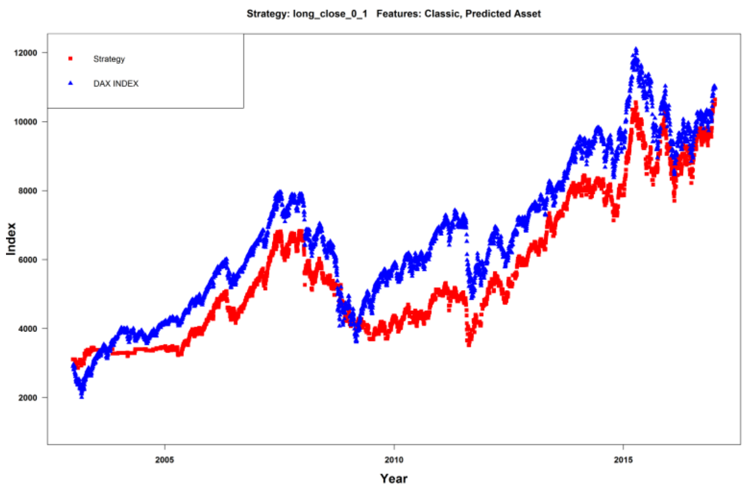


Figure 35: Investment Performance DAX - Single time series with ‘Classic Features’. Model (Red) versus benchmark (Blue).

(2) Single time series calculation with ‘Landmarks’: MCC = 0.06 and Chi-squared = 49.56 -> in line with other results, no real progress.

Table 38: Confusion Matrix DAX - Single time series with ‘Landmarks’.

		Predicted Class		
		-1	0	1
Correct Class	-1	425	454	194
	0	452	658	155
	1	471	438	210

(3) Single time series calculation with ‘EW-Features‘: MCC = 0.07 and Chi-squared = 47.27 -> the investment strategy is less volatile with similar returns, compared to the benchmark.

Table 39: Confusion Matrix DAX - Single time series with ‘EW-Features’.

		Predicted Class		
		-1	0	1
Correct Class	-1	202	325	239
	0	130	430	209
	1	204	305	237

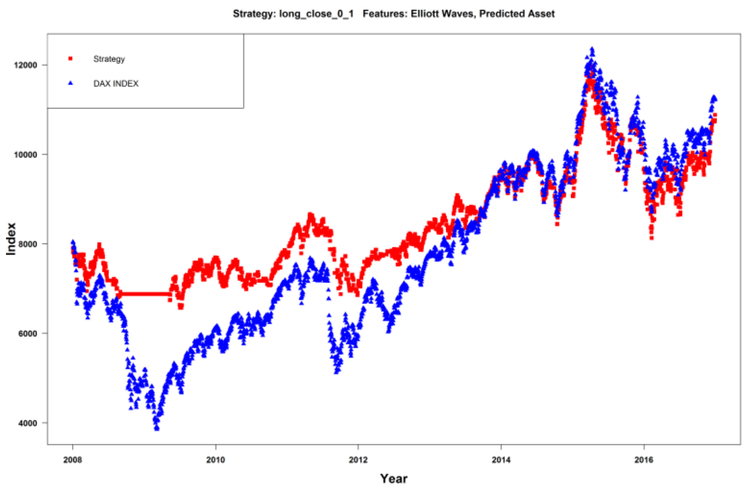


Figure 36: Investment Performance DAX - Single time series with 'EW-Features'. Model (Red) versus benchmark (Blue).

(4) Multi-asset calculation with Granger Causality and 'Classic Features': With the multi-asset approach we have the data 'time-overlap issue' again, less pronounced than in Asia though. We still calculate the investment strategy on an OOB basis. The results for the statistics are MCC = 0.11 and Chi-squared = 85.92 -> better than Single time series calculations, but the effect is less pronounced as with the Nikkei.

Table 40a, 40b: Confusion Matrix & Performance Statistics DAX – Multi-asset with Granger and 'Classic Features'.

		Predicted Class		
		-1	0	1
Correct Class	-1	398	440	273
	0	335	686	278
	1	298	469	374

	Maximum Drawdown	Sharpe Ratio
Index	4350	0.40
long 0_1	2863	0.39

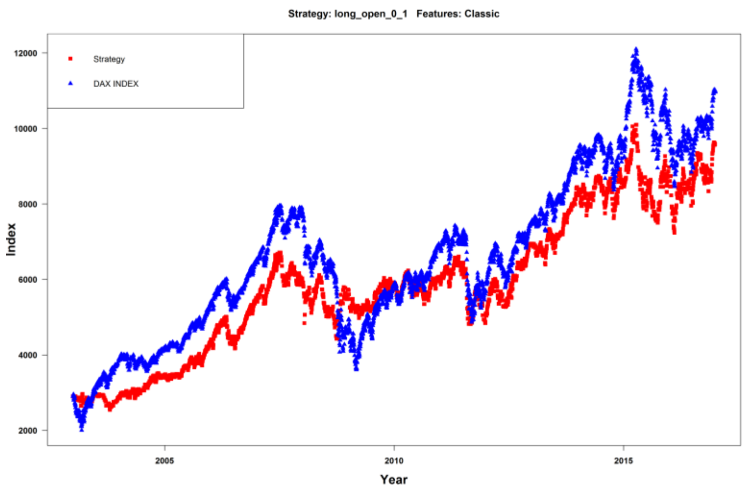


Figure 37: Investment Performance DAX - Multi-asset with Granger and 'Classic Features'. Model (Red) versus benchmark (Blue).

(5) Multi-asset calculation with Granger Causality and 'Landmarks': With MCC = 0.07 and Chi-squared = 38.59 -> one of the worst trend predictor, underperforms the Null, despite slightly positive MCC.

Table 41a, 41b: Confusion Matrix & Performance Statistics DAX – Multi-asset with Granger and 'Landmarks'.

		Predicted Class		
		-1	0	1
Correct Class	-1	384	371	240
	0	336	543	287
	1	364	378	299

	Maximum Drawdown	Sharpe Ratio
Index	3592	0.43
long 0_1	3823	0.13

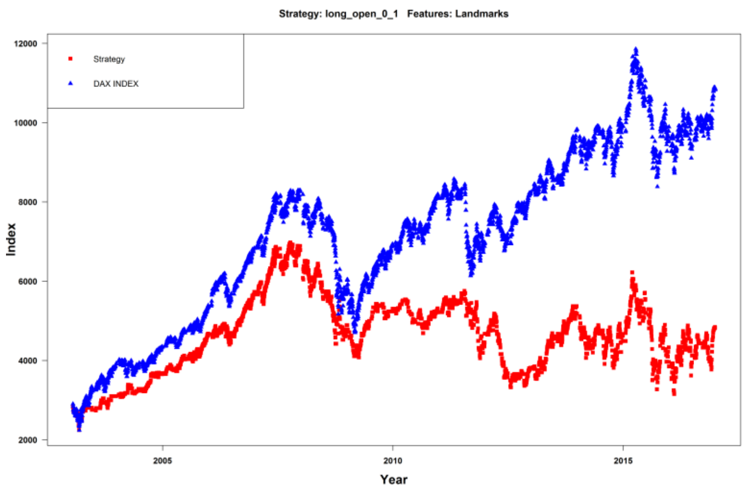


Figure 38: Investment Performance DAX - Multi-asset with Granger and 'EW-Features'. Model (Red) versus benchmark (Blue).

(6) Multi-asset calculation with Granger Causality and 'EW-Features': With MCC = 0.06 and Chi-squared = 32.73 -> again only a weak improvement over benchmarks.

Table 42a, 42b: Confusion Matrix & Performance Statistics DAX – Multi-asset with Granger and 'EW-Features'.

		Predicted Class		
		-1	0	1
Correct Class	-1	245	341	180
	0	228	416	125
	1	235	314	197

	Maximum Drawdown	Sharpe Ratio
Index	4190	0.21
long 0_1	3372	0.21

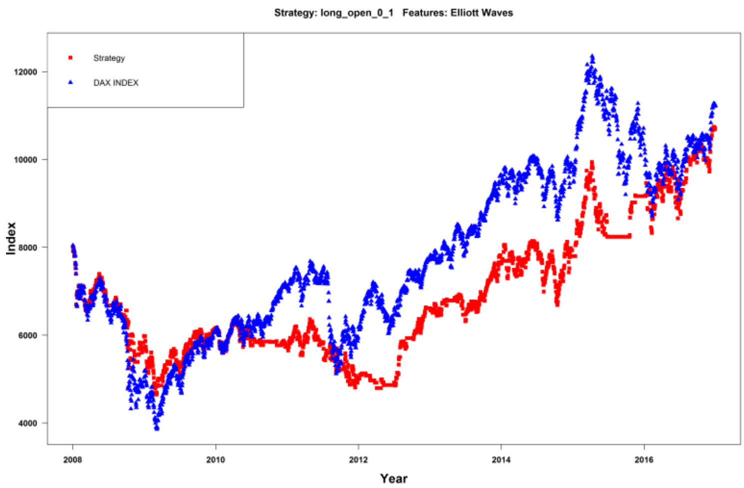


Figure 39: Investment Performance DAX - Multi-asset with Granger and 'EW-Features'. Model (Red) versus benchmark (Blue)

(7) Multi-asset calculation with Granger Causality and 'Residuals': see section 2.3.7.5.

(8) and (9) Multi-asset calculation with Granger Causality and 'Landmarks' or with 'Residuals & Landmarks light' (see also 2.3.7.3.) -> the results are in line, marginally worse than (7).

⇨ Final conclusion for the DAX Index: The best strategies are again multi-asset strategies, yet the effect is smaller than with the Nikkei. Overall performance of the models is surprisingly robust. Comparing Random Forest calculations with XGBoost we find that XGBoost has a slight edge over RF.

8.2.2.4. UKX Index, United Kingdom

The UKX index is a European index outside the EURO zone. When looking at the UKX results we find an outstanding model performance. This leads us to further investigate this anomaly. We find an interesting reason. This is a perfect example for the issues when operating learning algorithms in real life.

(1) Multi-asset calculation with Granger Causality and ‘Residuals’: MCC = 0.12 and Chi-squared = 118.49 are fairly good, but do not compare to the outstanding performance of the investment strategy - see the prominent performance graph and the Sharpe Ratio of 1.64. The inconsistency is similarly reflected in the Confusion Matrix.

Table 43, 43b: Confusion Matrix & Performance Statistics UKX – Multi-asset with Granger and ‘Residuals’.

		Predicted Class		
		-1	0	1
Correct Class	-1	508	335	281
	0	453	477	342
	1	333	333	476

	Maximum Drawdown	Sharpe Ratio
Index	3220	0.23
long 0_1	1019	1.64

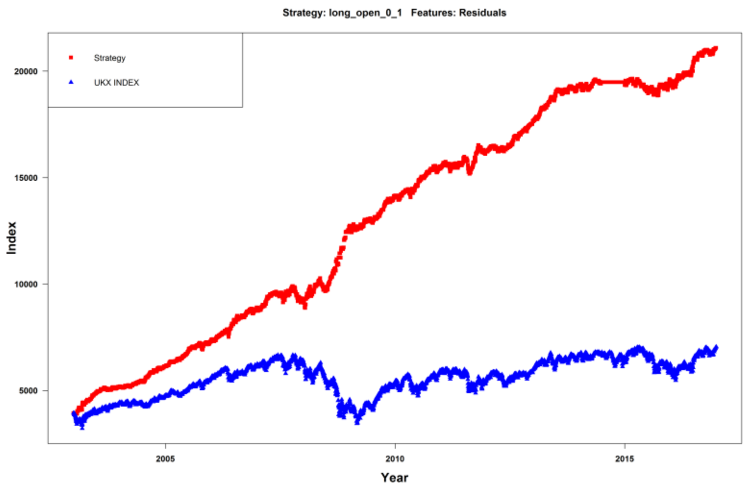


Figure 40: Investment Performance UKX - Multi-asset with Granger and ‘Residuals‘. Model (Red) versus benchmark (Blue).

(2) Multi-asset calculation with Granger Causality and ‘Residuals & Landmarks light’: see section 2.3.7.3.

⇨ Final conclusion for the UKX index:

(a) Based on the statistics the multi-asset models perform well and show a significant improvement over Null.

(b) We are surprised by the strong outperformance through the implicit time-overlap, see 2.3.7.3.[167]

8.2.2.5. SPX Index & Dow Jones, USA

The US indices reflect the world Leading markets. It is difficult to imagine that they take their lead from around the world. It is interesting to see, whether the difference in performance will be as pronounced between the Single time series calculations and the multi-asset ones.

(1) a) Single time series calculation with ‘Classic Features’: MCC = 0.12 and Chi-squared = 162.48 -> strongest performance of this specific model compared to other indices.

Table 44: Confusion Matrix SPX - Single time series with ‘Classic Features’.

		Predicted Class		
		-1	0	1
Correct Class	-1	322	400	365
	0	310	704	257
	1	348	369	428

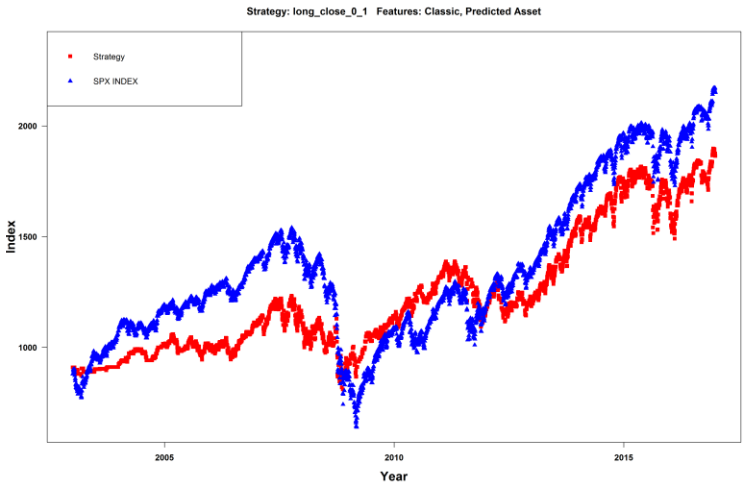


Figure 41: Investment Performance SPX - Single time series with ‘Classic Features’. Model (Red) versus benchmark (Blue).

[167] We did expect some effect, but not to that degree, as there is still a substantial part of the trading day left, for which there is no overlap.

(1) b) Single time series calculation with ‘Classic Features’ with strategy change. Long only strategy solely for ‘+1’ classifications: With identical statistics, significantly improved investment performance.

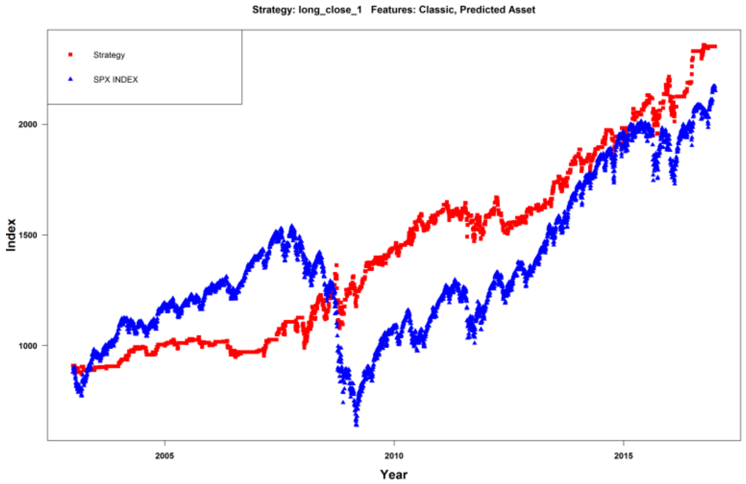


Figure 42: Investment Performance SPX - Single time series with ‘Classic Features’ with strategy change: ‘long-only +1’.
Model (Red) versus benchmark (Blue).

(2) Single time series calculation with ‘Landmarks’: see section 2.3.7.4.

(3) Single time series calculation with ‘EW-Features‘: MCC = 0.12 and Chi-squared = 97.77 -> in line with others, no benefit through the EW features.

Table 45: Confusion Matrix SPX - Single time series with ‘EW-Features’.

		Predicted Class		
		-1	0	1
Correct Class	-1	103	257	353
	0	100	433	251
	1	102	251	402

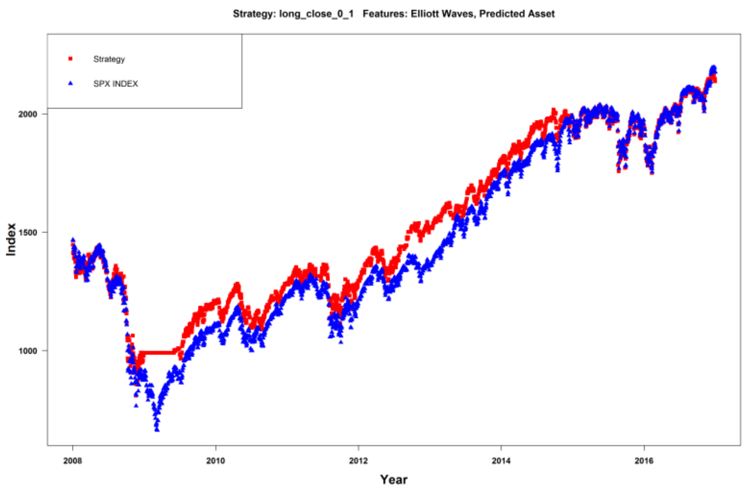


Figure 43: Investment Performance SPX - Single time series with ‘EW-Features’. Model (Red) versus benchmark (Blue).

(4) Multi-asset calculation with Granger Causality and ‘Classic Features‘: With the US indices there is no time-overlap issue, as they close last of the 300 underlying assets. Therefore we calculate on a COB basis. MCC = 0.10 and Chi-squared = 115.54 -> robust results, but slightly worse than Single time series calculations - interesting result.

Table 46a, 46b: Confusion Matrix & Performance Statistics SPX – Multi-asset with Granger and ‘Classic Features’.

		Predicted Class		
		-1	0	1
Correct Class	-1	305	410	372
	0	282	718	271
	1	298	447	400

	Maximum Drawdown	Sharpe Ratio
Index	901	0.39
long 0_1	422	0.51

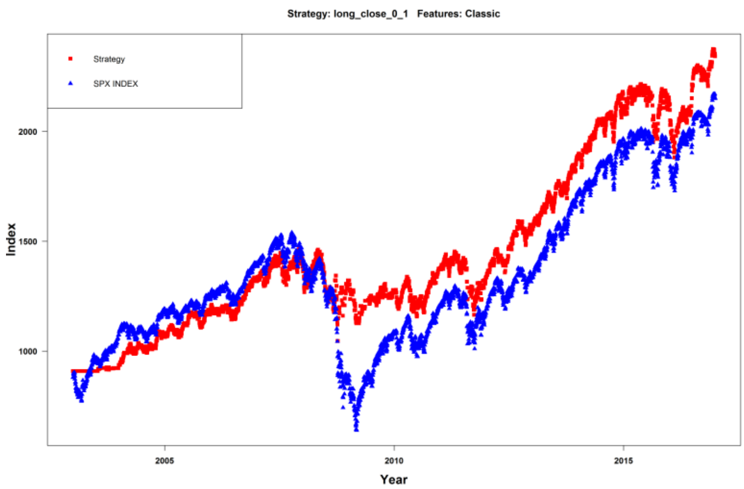


Figure 44: Investment Performance SPX - Multi-asset with Granger and ‘Classic Features‘. Model (Red) versus benchmark (Blue).

(5) + (6) Multi-asset calculation with Granger Causality and ‘Landmark‘ or ‘EW-Features‘:
-> similar results, slightly weaker than (4).

(7) Multi-asset calculation with Granger Causality and ‘Residuals‘ for SPX index and Dow Jones index:

a) SPX Index: MCC = 0.06, Chi-squared = 33.00 -> once again weaker than Single time series results

Table 47a, 47b: Confusion Matrix & Performance Statistics SPX – Multi-asset with Granger and ‘Residuals’.

		Predicted Class		
		-1	0	1
Correct Class	-1	351	431	312
	0	347	608	319
	1	362	429	365

	Maximum Drawdown	Sharpe Ratio
Index	889	0.41
long 0_1	359	0.45

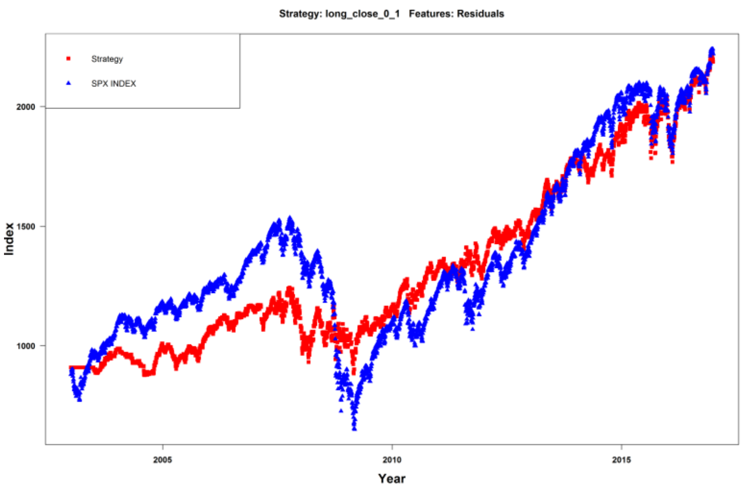


Figure 45: Investment Performance SPX - Multi-asset with Granger and ‘Residuals‘. Model (Red) versus benchmark (Blue).

b) INDU Index: MCC = 0.06, Chi-squared = 29.42 -> results in line with SPX index.

Table 48a, 48b: Confusion Matrix & Performance Statistics INDU – Multi-asset with Granger and ‘Residuals’.

		Predicted Class		
		-1	0	1
Correct Class	-1	351	442	290
	0	383	640	298
	1	335	445	340

	Maximum Drawdown	Sharpe Ratio
Index	7617	0.40
long 0_1	5969	0.37

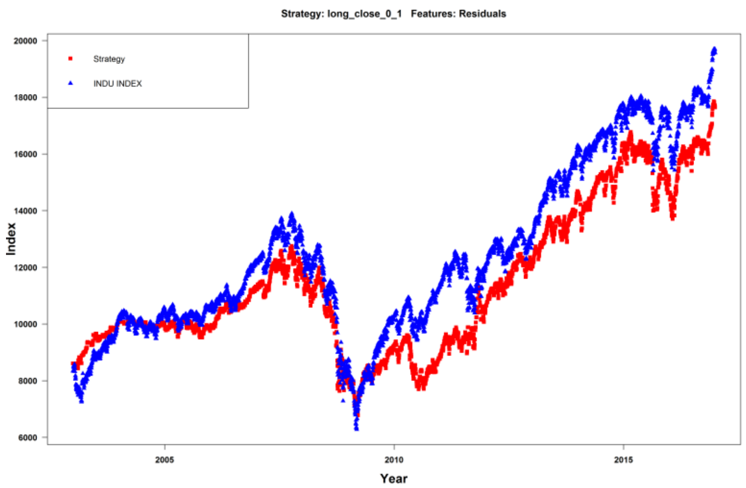


Figure 46: Investment Performance INDU - Multi-asset with Granger and ‘Residuals‘. Model (Red) versus benchmark (Blue).

(8) Multi-asset calculation with Granger Causality and ‘Residuals‘ & ‘Landmarks light‘ for SPX index and Dow Jones index:

a) SPX Index: MCC = 0.09, Chi-squared = 82.02 -> slightly weaker than Single time series results

Table 49a, 49b: Confusion Matrix & Performance Statistics SPX – Multi-asset with Granger and ‘Residuals’ & ‘Landmarks light’.

		Predicted Class		
		-1	0	1
Correct Class	-1	335	442	337
	0	292	680	302
	1	341	429	386

	Maximum Drawdown	Sharpe Ratio
Index	889	0.41
long 0_1	360	0.47

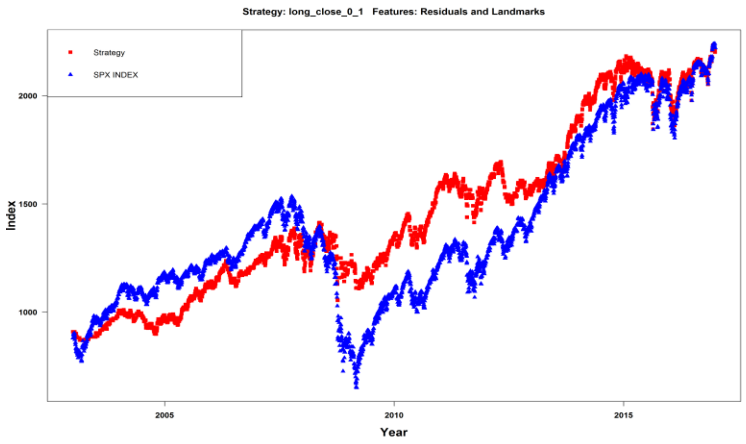


Figure 47: Investment Performance SPX - Multi-asset with Granger and 'Residuals & Landmarks light'. Model (Red) versus benchmark (Blue).

b) INDU Index: MCC = 0.08, Chi-squared = 54.08

Table 50a, 50b: Confusion Matrix & Performance Statistics INDU – Multi-asset with Granger and 'Residuals' & 'Landmarks light'.

		Predicted Class		
		-1	0	1
Correct Class	-1	322	459	302
	0	325	716	280
	1	311	467	342

	Maximum Drawdown	Sharpe Ratio
Index	7617	0.40
long 0_1	3930	0.46

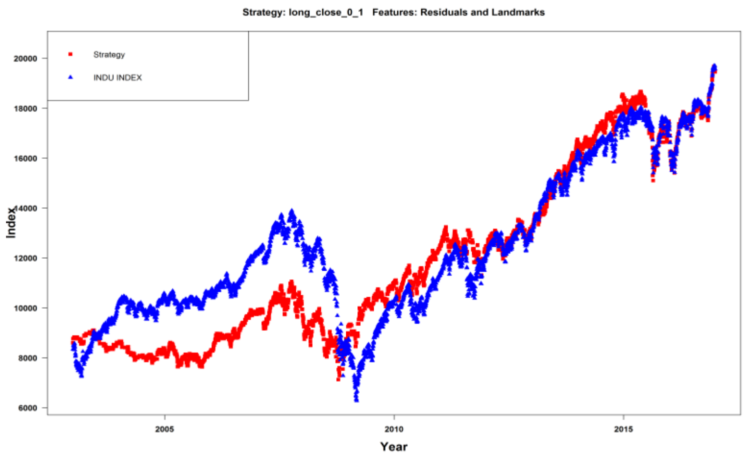


Figure 48: Investment Performance INDU - Multi-asset with Granger and 'Residuals & Landmarks light'. Model (Red) versus benchmark (Blue).

8.2.2.6. Volatility and Aggregated Class Zero Classifications

Figure 49 shows that class ‘0’ categorisations drop significantly in periods of high market volatility. The aggregated number of class ‘0’ predictions picks up again when markets calm down.

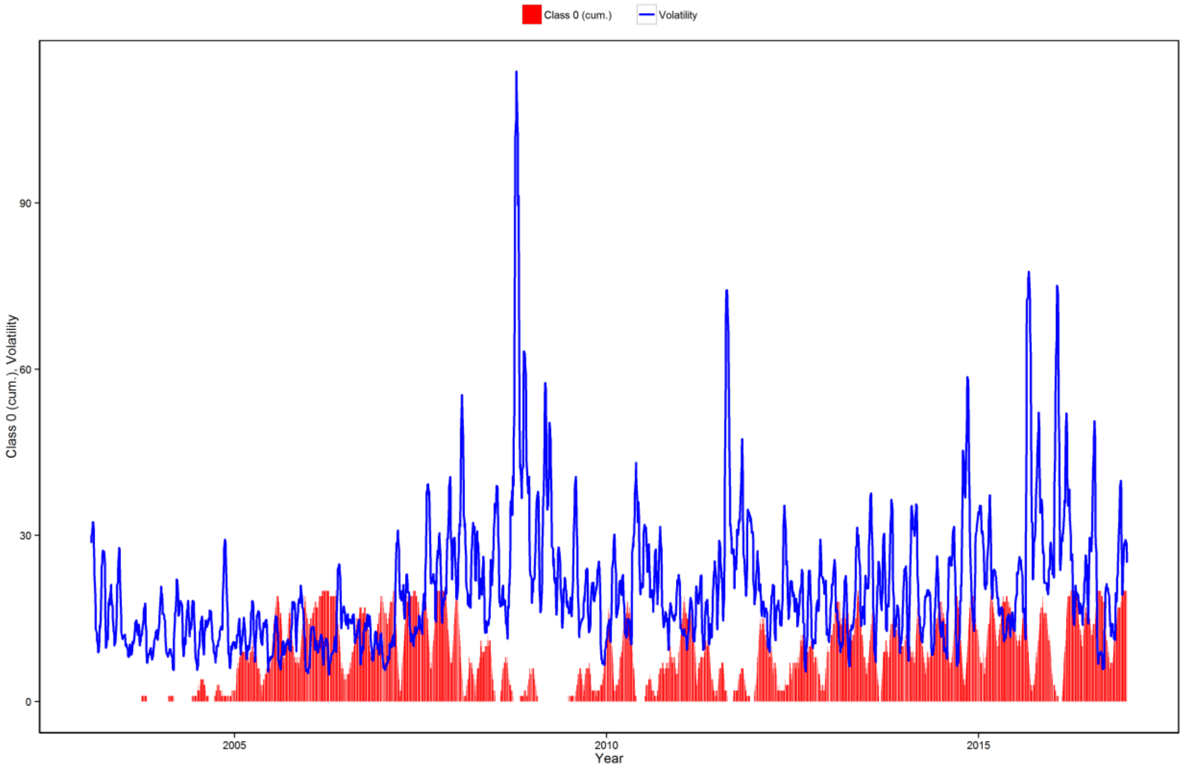


Figure 49: Comparing Volatility and aggregated class 0 results, calculated over a 20 day period. (SPX Index)

8.2.3. Appendix on Portfolio Optimisation

The Trend Classification was run at COB each day for each Index. The underlying portfolio consists of NKY, HIS, DAX, SPX and INDU indices. The Markowitz Model[168] is run the following morning due to time-overlap issues. The Markowitz model takes the opening prices as the input to the model. We ‘execute’ accordingly at the opening price. The Markowitz model uses the results from the Trend Classification models the night before as a pre-filter.

[168] The Markowitz model we use was coded / implemented / run by Vineet Gupta (MSc Econ).

The results with Risk Parameter = 0 are:

Enhanced Markowitz:		*Traditional Markowitz:*	
Annualised Return	0.0678	Annualised Return	0.0102
Annualised Std. Dev.	0.1721	Annualised Std. Dev.	0.1600
Annualised Sharpe[169]	0.3324	Annualised Sharpe	0.0009

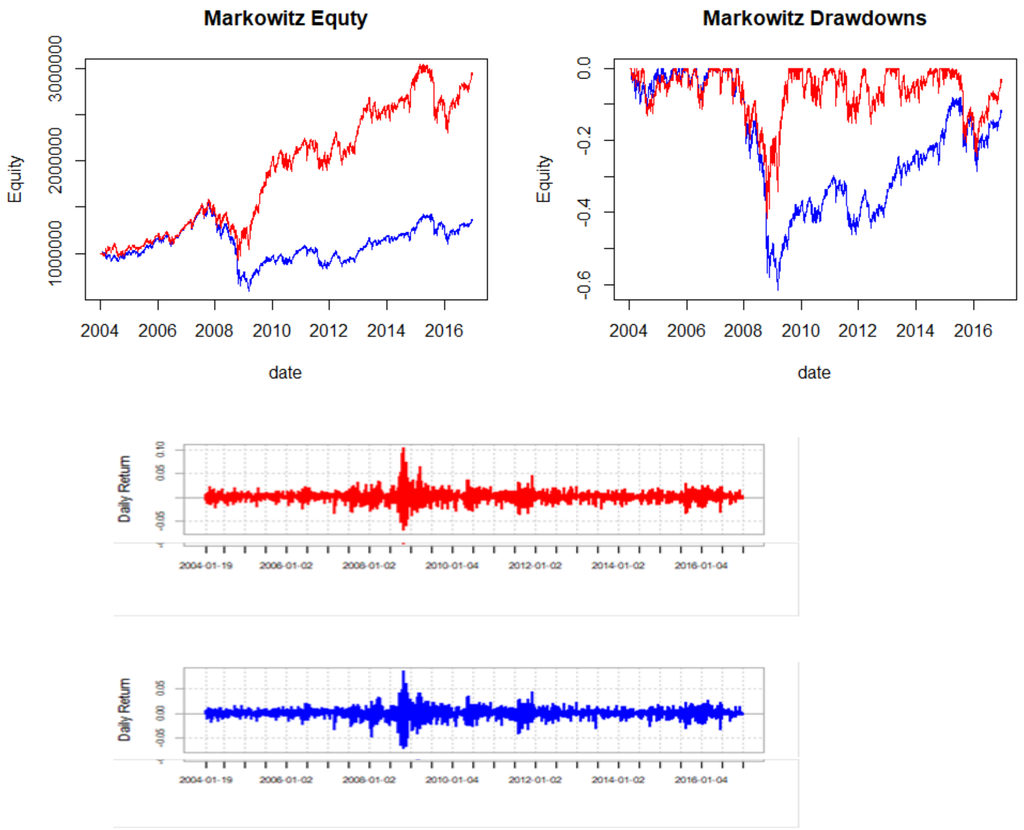


Figure 50a, 50b, 50c, 50d: Enhanced Markowitz (Red) vs. Traditional Markowitz Portfolio (Blue). Shown are the aggregated Investment Performance, the daily Drawdowns and the daily Returns for each Markowitz model.

169 Rf = 1%

8.3. Appendix on Institutional Customer Behaviour

8.3.1. Applied Algorithms - see Friedman *et al.* (2000)

(a) For binary classification, the following algorithm gives likelihood gradient boosting using regression trees:

1. $F_o(x) = \frac{1}{2} log \frac{1+\bar{y}}{1-\bar{y}}$
2. For m = 1 to M do:
3. $\tilde{y}_i = 2y_i/(1 + \exp(2y_i F_{m-1}(\mathrm{x}_i)))$
4. $\{R_{jm}\} = J - terminal\ node\ tree(\{\bar{y}_i, \mathrm{x}_i\}_1^N)$
5. $\gamma_{jm} = \sum_{\mathrm{x} \in R_{jm}} \tilde{y}_i / \sum_{\mathrm{x} \in R_{jm}} |\tilde{y}_i|\ (2 - |\tilde{y}_i|)$
6. $F_m(\mathrm{x}) = F_{m-1}(\mathrm{x}) + \sum_{j=1}^{J} \gamma_m\, 1(\mathrm{x} \in R_{jm})$

End For

End Algorithm

(b) For multiclass classification using K-class logistic gradient boosting the algorithm is:

1. $F_{ok}(x) = 0 \quad k = 1\ to\ K$
2. For m = 1 to M do:
3. $p_k(\mathbf{x}) = \exp\big(F_k(\mathbf{x})\big)/\sum_{l=1}^{K} \exp(F_l(\mathbf{x}))$
4. $For\ k = 1\ to\ K\ do$:
5. $\tilde{y}_i = y_{ik} - p_{k,}(\mathbf{x}_i), i = 1, N$
6. $\{R_{jm}\} = J - terminal\ node\ tree(\{\bar{y}_i, \mathrm{x}_i\}_1^N)$
7. $\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{\mathbf{x}_i \in R_{jkm}} \tilde{y}_{ik}}{\sum_{\mathbf{x}_i \in R_{jkm}} |\tilde{y}_{ik}|(1-|\tilde{y}_{ik}|)}, j = 1, J$
8. $F_m(\mathrm{x}) = F_{m-1}(\mathrm{x}) + \sum_{j=1}^{J} \gamma_m\, 1(\mathrm{x} \in R_{jm})$

End For

End Algorithm

8.3.2. Feature Engineering

Table 50 displays the most significant features used in the fitting process.

Table 51: Customer Coverage Feature set

Name	Description
Days Since Last Trade	Day count since last trade
Overdue-ness	(Days Since Last Trade – Mean Days Between Trades) / (Mean Days Between Trades) Better to use the version adjusted for variance
Month	Categorical and binary-ised
Trade Leader Activity	Categorical, based on statistical significance between client segments on trade leadership (does one segment induce another to trade). Take weekly total trades by all clients in the leading segment then view the current #trades in last 7 calendar days of that group and equate categories as follows: >75% = HIGH 25%-75% = MED <25% = LOW
Customer Segmentation	See Appendix 8.3.3
isPeriodic	Boolean, has statistically significant trade periodicity at 1 of 5 levels (Yearly, Semi-Annually, Quarterly, Bi-Monthly, Monthly)
isInPeriod	Boolean, second part to the above feature, denotes whether the client is close the expected periodicity

8.3.3. Customer Segmentation

We differentiate between five distinct customer groups as clusters:

1. Irregular & Extremely Infrequent: Members, with irregular purchases;
2. Financial Biased & Infrequent: Members, with less than 100 trades per year 20%, 60% and 20%;
3. Regular & Frequent: Members, with 300 trades per year, purchase market portfolio 50%, 45% and 5%;
4. Financial Biased & Frequent: Members, with 500 trades per year 20%, 60% and 20%; and
5. Regular & Extremely Frequent: Members, with thousands of trades per year 20%, 45% and 35%