

FORECASTING MACROECONOMICAL INDICES WITH MACHINE LEARNING: IMPARTIAL ANALYSIS OF THE RELATION BETWEEN ECONOMIC FREEDOM AND QUALITY OF LIFE

Jonathan Staufer and Patricia Brockmann

Computer Science Department
Technische Hochschule Nuernberg Georg Simon Ohm
Nuremberg, Germany

ABSTRACT

The importance of economic freedom has often been stressed by supporters of liberalism, but can its actual effect be observed in a data driven, objective way? To analyze this relation the Economic Freedom of the World (EFW) index and the Human Development Index (HDI) were examined with modern machine learning algorithms and a wide-ranging approach. Considering the EFW index's preference of a liberalistic oriented economic policy, an objective recommendation for creating an economic policy that improves people's everyday lives might be derived by the analysis results. It was found that these more advanced algorithms achieve a considerably stronger correlation between both indices than pure statistical means yet leave a small room for interpretation towards a counter-liberalistic implementation of demand-driven economic policy.

KEYWORDS

Data Mining, Machine Learning, Neural Networks, Economic Freedom of the World Index, Human Development Index

1. INTRODUCTION

Economic freedom is seen by liberalists as a valuable asset of an economy as it promotes economical growth which is followed by a better standard of living [1].

The question arises to what degree this statement is actually true. Can a relationship between the economic freedom of a population and its prosperity, development and quality of life be found — is economic freedom truly a way to improve peoples everyday life?

Objective insights may be provided by data analysis of economical indices using state of the art machine learning algorithms. An indicator of economic freedom is the Economic Freedom of the World Index (EFW Index) of the Canadian Fraser Institute. It quantifies the degree of economic freedom of various countries since 1996. Based on the EFW, this research aims to forecast the Human Development Index (HDI) released by the United Nations Human Development Report to analyse dependencies and relationships between both indices. The HDI aims to put prosperity, quality of life and the overall development of countries into figures.

Depending on the insights, it might even be possible that a general recommendation for the formulation of economic policy could be derived to develop a country in a positive manner.

Various publications have dealt with the EFW Index in the past [2]. For example, a correlation between data from the EFW index and population satisfaction was identified [3]. A noticeable feature of former publications was that were mostly investigated with statistical means. However, state of the art machine-learning algorithms could reveal and illustrate previously undiscovered knowledge and relationships within data.

2. OBJECTIVES

The general objective of this research was to use machine learning to forecast macroeconomic indices. The basic requirement to achieve this was to pick a suitable machine learning environment from the tools which have become available over the last few years. These will be looked at in Section 4.2.

To move beyond a simple forecast, two macroeconomic indices, which could be related, have been chosen to be predicted for the next time period: The Economic Freedom of the World Index as the basis. Its individual values were used to predict the Human Development Index for the following year. Previous work has shown that there is a statistical correlation between these two indices, yet modern machine learning algorithms hadn't yet been used to analyse this correlation more closely [4].

In German economic thought, there are different approaches to shape successful economic policy. The continuous discussion among scientists on how to execute this policy in detail still has not led to a clear result. One approach is similar to the *laissez-fair* liberalistic approach, where a country should not interfere directly with markets. Others see an active involvement of the state, where necessary, as reasonable. Section 3.2 will expand on these competing points of view. Depending on the results of the analysis, it may be possible to get closer to an answer regarding the effects of a direct interaction of the government and markets. The economic freedom of the world index favours a liberal standpoint of non-government involvement in its rankings which could lead to new insights regarding government involvement.

2. STATE OF THE ART AND BACKGROUND

2.1. Machine Learning

As a research field of artificial intelligence, machine learning has been influenced by a variety of different areas. At its core, it utilizes statistical analysis and computer science. Moreover, it has also been influenced by concepts from biology and psychology [5].

Machine learning algorithms analyse data for patterns either to highlight that structure, reduce its complexity or to generate predictions. As Arthur L. Samuel stated in 1959, ideally it is not necessary to create specific code to find a problem's solution, as the algorithm learns and approximates its variables to the problem itself [6]. This also works for complex systems with multiple influences — as long as there is enough data available, a lack of knowledge can be made up for by it [5].

To choose the right algorithm for an analysis, several aspects should be considered:

- **Type of available data:** The choice of algorithms is limited by the type of data available [7]. Labelled data can be used with different algorithms (supervised learning) than unlabeled data (unsupervised learning). Unlabeled data might get labelled during the process [5].
- **The goal of the analysis:** What is the target of an analysis? Usually the classes are distinguished between regression, classification, clustering and associative analysis.
- **Knowledge representation:** Depending on the choice of algorithms, the complexity of a model and the knowledge contained in it also differs.

Choosing an easy to understand knowledge representation is important if the results and the knowledge within the model are of interest. Accessibility of knowledge and its presentation often depends on the representation of knowledge [7].

As the objectives described in Section 2 emphasise, this project aims to create a precise numerical prediction. This classifies as a regression problem, where the accuracy of the prediction as well as the correlation between both indices provide information about their relationship.

During the evaluation phase, multi-layer Perceptron (MLP) networks showed good results in forecasting and thereby mapping the relation between both indices. The original type of neuron, called Perceptron, usually uses simple algorithms and delivers a single output value to its successors. Stacking Perceptrons generates a multi-layered network which computes only in one direction: A feed-forward multi-layer Perceptron neural network. Multi-layered Perceptrons are usually trained with backpropagation, an algorithm which reviews each neuron after an iteration, where it tries to reduce the encountered prediction error via a gradient descent optimization algorithm [8].

An alternative to neural networks are boosting algorithms, which use simpler prediction algorithms, such as decision trees [9]. The algorithm presented in this paper merges a multitude of weak models to train a regression tree, which has shown that it can predict with similar accuracy as a neural network.

2.2. Macroeconomics

Economic systems can be seen as complex environments, which are subject to a variety of internal and external influences [10]. This complexity makes it difficult to understand and model such systems. All of the influencing factors and components react to each other with varying degrees of intensity [11]. This complexity makes it difficult to create a model of an economic system. Machine learning however, could perhaps create an appropriate approximation of such a system through the abstraction of underlying patterns by utilizing macro-economic indices.

A market economy is a liberal economic system which implements private ownership of the means of production and a dynamic self-optimizing function based on supply and demand [12]. It is a system of markets, where supply and demand meet to form a price-determining intersection. Most common is the idea of a free market economy based on the concept of liberalism. Here, the government just establishes general conditions, which leaves individuals and companies as much space as possible to act economically. It has become clear in the past that a free market system based on liberalism has led to widespread social grievances and the creation of monopolies [13]. Based on these previous experiences, a new approach to creating a market economy was established in Germany after WW2: the social market economy. With the idea of “prosperity for all”

[14] social aspects, such as social security systems, health insurance and fair income distribution were combined with aspects of the free market economy. The goal is to secure a minimum standard of living for every citizen [15].

Within the framework of a market economy, the government has the option to exert influence on the market through laws and regulations. In a market economy based on liberalism, this influence is undesirable. In a social market economy, however, laws and regulations show that there is active interference. The extent of this interference is a subject of controversy among economists. This can be observed in the Republic of Germany's policy consultants and the different opinions present among that group [16]. It is generally agreed on that it is necessary to create order and a framework for markets to function — a regulatory policy should be established. To further control markets in detail, a process policy has to extend that regulatory policy. The extent of this policy is a controversial topic. An additional point of disagreement is the orientation of policy. Should the starting point be supply- or demand-oriented politics?

In view of these aspects, two opposing general schools of thought have emerged: Ordo liberalism and Keynesianism [16].

Ordo liberalism builds upon order and liberalism/liberty. Its key aspect is to establish order in markets. Guidance occurs solely through creation of that order, with as little further interference as possible [16]. Ordo liberalism usually favors supply-oriented policies over demand orientation. Thus, some consider it as in support of privileged and rich people [17]. In general, it is similar to liberalism and aims to create fair competition. A further goal is to dissolve power centers in markets. At the same time, it acknowledges the need for a minimum of social security [18, 19].

Keynesianism, named after the British economist John M. Keynes, describes a contradictory economic direction. In addition to regulatory policy, it also includes process policy. When required, demand promotion is recognized as a necessary tool to successfully shape markets and economic policy [16]. During weak economic phases, active economic policy should be applied to overcome the emerging recession by using fiscal policy funds. According to Keynesianism, the assumption that supply creates an appropriate demand itself is questioned. A socially acceptable self-regulation of the market is not necessarily expected. The government therefore actively intervenes in the market in order to steer the processes of an "imperfect" market with regulatory policy, with the aim of the common good.

As ordo liberalism shares many aspects of traditional liberalism, it might be possible to find additional insights about the effects of liberalism and of ordo liberalism on the quality of life and the development of countries by exploring the relationship of appropriate macroeconomic indices.

2.3. Economic Indices

Indicators are economic metrics that describe the development of a region or country in relation to an economic aspect. They serve to investigate, assess and predict the development of the object of observation [20]. Indices consisting of several components are often converted into a meaningful value in order to enable better communication outside of specialist areas.

To meet the objectives described in Section 2, two appropriate indices have been chosen: The "Economic Freedom of the World" index, which rates economic freedom under the aspects of liberalism, and the "Human Development Index", a known prosperity and quality of life index published by the United Nations. Both indicators consist of different measures and components that are combined into an annual country ranking.

2.3.1 Economic Freedom of the World Index

The Economic Freedom of the World Index (in short EFW) of the Canadian Fraser Institute, a Canadian “think tank”, rates the individual economic freedom of countries on a yearly basis. Economic freedom describes the degree of freedom of the individual with which they can pursue their own independent economic interests [21]. The key to such an environment is considered to be free trade, independent competition and the protection of persons and property [22]. Part of this is also that individuals can make free decisions, no obstacles to market entry exist and that the government does not influence markets [23]. Especially the last point has been made clear in the past by Gwartney, Lawson and Gratzke, some of the index’s authors [24].

The EFW has been published annually since 1996, with the aim of using objective and transparent sources as a database for its ratings [2]. Subjective sources, such as surveys, are preferably avoided. In general, the aim is to use public and reliable data sources in order to avoid any influence by the authors of the indicator.

The EFW consists of 24 individual components which are aggregated into five sub-indices as intermediate step:

- Size of government
- Legal system and property rights
- Sound money
- Freedom to trade internationally
- Regulation

These sub-indices are aggregated as well into the overall index. The correlation between the 24 indicators considered individually is not always pronounced, something that could be observed in the first approach to the analysis as well. The reason for this, Bologna and Hall [25] argue, is that different kinds of freedom are expressed differently.

2.3.2 Human Development Index

Published on a yearly basis in the Human Development Report of the United Nations, the Human Development Index (HDI) is an indicator for prosperity and quality of life [26] (United Nations Development Programme, 2013a). The HDI was originally developed by the Pakistani economist Mahbub ul Haq with the aim of creating a basis for evaluation that goes beyond pure financial prosperity [27]. The rating is rated with values between 0 and 1. It is a compound indicator like the EFW, with three sub-indices [28]:

- **GNI index:** This index aims to rate the standard of living based on the GNI per capita.
- **Education index:** Considering the mean years of schooling and the expected years of schooling this index tries to measure the quality of education and knowledge within a country.
- **Life expectancy index:** To measure lifespan and healthfulness this index compares life expectancy at birth compared to maximum life expectancy (which is a value adjusted anew every year).

The idea behind the selection of these three components is the realization that a pure analysis of purchasing power does not allow a meaningful assessment of the quality of life [26]. The richness

of an economy does not at the same time reflect the fulfilment of a human life, even though it may be part of it [27].

Prior to 2010, the HDI was calculated on the basis of other attributes, such as the adult literacy rate and different minimum and maximum values [29]. However, these formula changes are seen by the United Nations as having little impact on the results. In addition to this possible restriction, the HDI is often criticized for not including all relevant attributes in its calculation [31]. Values such as inequalities, poverty or security of the population are not explicitly considered, which is why an extension of the basic HDI to these aspects has been implemented.

3. METHODOLOGY

To achieve the desired goals two more general aspects had to be considered. First, a fitting tool or environment to plan and execute the analysis had to be found. Out of all available alternatives three interesting tools were selected. These were compared with a scoring model. The analysis itself was performed under the aspects of Fayyad's data mining or "knowledge discovery in databases" process [32].

3.1 Data Mining

Data mining describes the automated recognition of patterns and trends in data [7]. The extension of pattern recognition by data preparation and visualization or interpretation was originally called "Knowledge Discovery in Databases (KDD)" [32], but the term data mining is often used synonymously. The process defined by Fayyad et al. includes multiple steps to extract knowledge out of data, which were used to methodically approach the analysis discussed in Section 5.

The process begins with selecting the required data available, then pre-processing the selected data by adjusting datatypes and scanning for outliers as well as transforming the data into the required format for analysis. After these preparatory steps the actual data mining step applies machine learning algorithms to recognize and extract patterns. To conclude the process, the interpretation and evaluation of the findings result in the knowledge gained.

3.2 Selecting the Appropriate Tool

To select a suitable data mining tool, it is necessary to initially define requirements for such. This was achieved through a scoring matrix that illustrated the relevant aspects of a machine learning environment. Key characteristics that were most relevant were:

- tool documentation and support
- online integration
- functional scope
- user experience
- visualization

Characteristics like "extendibility" and "type of license" (e.g. open source) were deemed relevant to an extent. "Interfaces" to various applications were considered but not given too much weight in the final evaluation. With the requirements (and weights) set up, a preliminary selection was necessary to break down the abundance of tools available. A guide to achieve this was the annual survey executed by the well-recognized data science community website "KDNuggets". In the 2016 survey the statistical programming language R led the ranks [33]. The popularity of R and

the amount of information material and instructions available made this language interesting for further consideration. Rapid Miner [34], the survey's first actual data analysis software tool, promised ease of use and a comprehensive library of algorithms for data analysis and was thereby added for closer consideration. When looking at the alternatives for a third comparison solution, it was striking that the machine learning algorithms of the "Weka" tool from the University of Waikato, New Zealand, were often integrated into other environments. This tool was ranked fourth (yet first actual tool) in the 2013 KDNuggets polls [35], but has since fallen behind in rank. Weka was chosen as the third tool for comparison due to its being named frequently as an open source alternative with a complete graphical user interface.

To be able to compare R more accurately with the chosen alternatives, RStudio, a development environment for R was included in the comparison. R supports a wide selection of algorithms and can (or sometimes has to) be extended with additional code, making it highly adaptable to a wide range of use cases. It is accessible for free with no additional cost and has pretty adequate documentation and assistance [36]. The integrated support seems rather academic and the examples provided are quite abstract. The characteristics it lacks however, are user experience and visualization. The complexity for novice users to use R is the same as learning a new programming language. Error messages are sometimes inconclusive and the visualization is a simple image (.jpg format) with no scalability.

Weka is the Open Source Machine Learning and Data Mining application of the University of Waikato, New Zealand [37]. It is a collection of machine-learning algorithms with the possibility to apply them to data. Similar to RapidMiner, Weka makes it possible to select and filter data within the tool. It provides three different views as the starting point of the analysis, including a workflow editor. Pentaho offers a commercial version of Weka with an improved UI. Documentation and support is extensive for Weka. While integrated guides and information are rather limited there are whole lectures by Weka's creators and developers available on the internet. The tool itself is written in Java and can be extended with additional code at any time. Its functional scope is large and allows the combination of multiple algorithms to build more powerful models. However, some basic elements, such as correlation matrixes, are missing. Integration to Hadoop and of R is available if desired. The user experience is good, but the tool is overall more complex than RapidMiner. Also, the work flow editor is limited in available elements and connections which makes it harder to handle and reduces use cases. Error messages are somewhat hidden, and the tool only displays java error messages with no further insight. The user interface is in old-fashioned java look. Visualizations are similar to that art style but provide an interactive user interface.

RapidMiner is a fee-based licenced tool. Its system core is publicly available under an open source license [38]. The free version comes with some limitations in usage but is rather generous. It covers the entire data analysis process, from importing and cleansing the data to visualizing the results of the analysis. The core element of the software is a workflow designer, which allows users to link predefined process steps into a complete process via drag and drop elements. The tool itself has an integrated tutorial, documentation and sufficient support of the connected community for help. Anyone can add new Java code. RapidMiner also supports various scripting languages. As mentioned before, besides the already implemented extensive library of processing steps and algorithms, it also utilises Weka-algorithms by providing freely accessible easy to use extensions to the tool. The software provides interfaces to common data sources like Hadoop. From a user experience standpoint, it offers an easy to use workflow editor. The editor merely requires the user to learn what process steps do and how to configure them, which is made easy by the extensive integrated descriptions. Error messages are easy to understand and even include advice about how to resolve them. Visualisation is another strength of RapidMiner. The modern user interface allows for multiple configurations and interactive analysis of results.

After mapping these aspects these to the defined characteristics and translating them into values from 1 to 10, RapidMiner was the best tool for the planned analysis. Weka fell behind on visualisation and slightly increased complexity. R's complexity was something that could not be coped with during the duration of this project, but might be of interest for future work.

4. ANALYSIS

To approach the analysis, the first step was to create a combined dataset consisting of the aggregated value for the EFW index, its five sub-indices and the HDI as well as its value of the following year. The available attributes were adjusted for each experimental setup so that predictions couldn't be made on related value such as HDI and its next year follow-up value.

The analysis itself was setup similar to the process described in Section 4.2. Each step was iterative and repeated with different ideas and approaches. The evaluation started with a basic exploration by applying simple models and statistical methods. This aimed for learning about general correlations and insights. The information gained was utilized in the following experiments for optimization. This resulted in choosing the necessary complexity of algorithms for an in-depth analysis from a variety of perspectives. The comparison of results was based on RMSE and correlation of models.

4.1 Exploration

To learn about dependencies between attributes, a correlation matrix is a good starting point for exploration. Correlation means setting covariance and variance into a relation, to check whether there is a similar development within the data for attributes. It should be noted that even data which do not have any interrelationships could still show a statistical connection. Therefore, the correlation coefficients which are found should always be examined critically. In the case of this evaluation, there are various interrelationships the Fraser Institute recognized itself, such as the inhabitants of countries with a higher degree of economic freedom enjoy better education, longer lives and higher net income [23]. Assuming a correlation between the two values is therefore rational.

With a value of 0.701, the matrix showed that there is a clear relationship between the aggregated values of the EFW index and the HDI. Among the sub-indices and the HDI, the strongest relationship exists between "Legal system and property rights" with a value of 0.760. Other sub-indices to HDI correlations are between 0.5 to 0.67. These are still strong, but not as distinct. The characteristic of the "size of government" is also noteworthy - not because of its correlation with other values, but rather because of the lack of it. It does not seem to be of relevance for the EFW index as a standalone sub-index. It is even slightly inversely related to the HDI. The lack of correlation with other attributes suggests that the size of government has no greater relevance to the HDI. The size of government indicator might be of relevance in combination with other sub-indices of the EFW index. This also indicates that a larger government might be beneficial for quality of life and development, contrary to the beliefs of ordo liberalism.

Now, a first attempt to predict values can be made. A simple method of regression to start off with is linear regression. It averages the available data points and creates a forecast based on a straight line. This implies that the progression of a dataset is also linear. To create a prediction of future events, the HDI of the following year was to be forecasted. Using K-fold cross-validation, a simple forecast was made based on the aggregated EFW index for the HDI of the following year. Linear regression achieved a correlation of 0.702 and a RMSE of 1.177 for this forecast —

a meagre result that clearly had to be surpassed. Further experimentation showed that polynomial regression achieved a similar (linear) result with worse performance.

The next logical step is to increase complexity to allow for a better mapping. Multiple linear regression allows for linear regression analysis based on multiple input values. This extended approach made it possible to break down the EFW index into its sub-indices, in order to forecast the HDI. At this point, it is possible to either select or to omit individual sub-indices. Hall and Lawson, however, criticized this. They argued that the EFW index is more than just the sum of its parts [2]. As a consequence, it was decided that only the algorithms themselves are allowed to reduce the importance of attributes on a mathematical basis. Multiple linear regression achieved a correlation of 0.807 and a RMSE of 0.973. Both values improved and showed that an increase in complexity was the right path to take. During the analysis, performed with M5 regression trees, the attribute “regulations” was dropped. Polynomial regression once again showed that it wasn’t a suitable method to analyze the data and was therefore discarded.

To conclude the initial exploration, a visual and a cluster analysis was performed (see Figure 1). Tests showed that four clusters were the optimum number for reliable differentiation, since the cluster boundaries were in comparison most clearly defined.

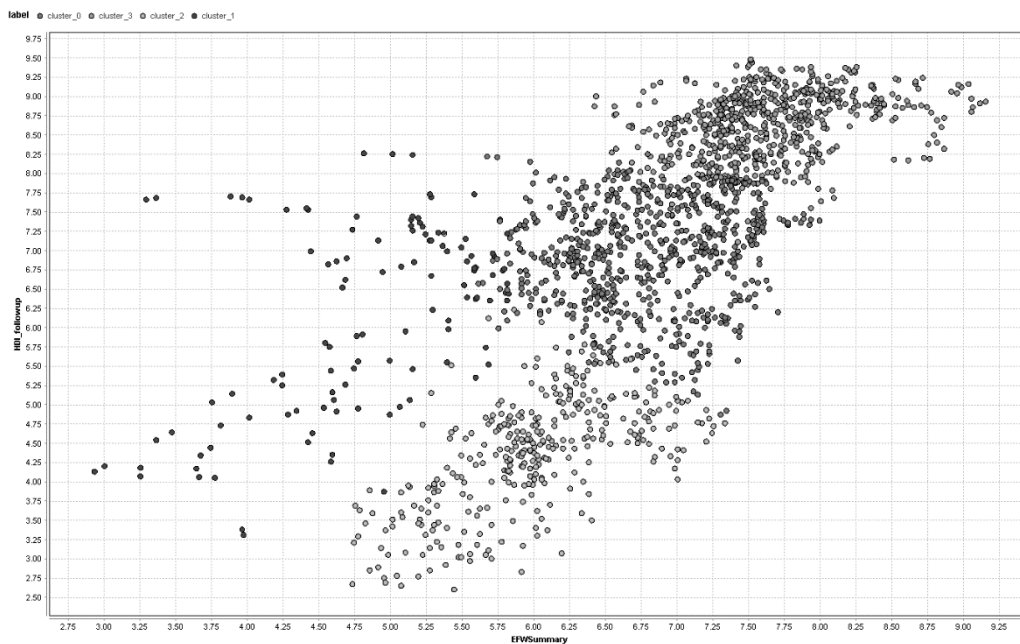


Figure 1: Comparison of HDI(left) and EFW index (Bottom) divided into four clusters

The comparison of HDI and EFW index showed that, to a certain degree, both values can develop independently from each other. To achieve the highest rating, however, they seem to be dependent on each other. These observations indicate that economic freedom and prosperity or development of a country complement each other when both indices are strongly developed, but are not solely responsible for the development of the compared index value.

All in all, the results confirm the assumptions already made during linear regression: The data set is subject to many interdependencies and rules, which cannot be mapped easily. To analyze this case, more advanced algorithms are needed that look at the individual variables independently of each other.

4.2 In-Depth Analysis

RapidMiner, the software used for the analysis, provides a variety of algorithms for complex evaluations. Based on the method of optimization, they all share one characteristic: Unsteady results. To counter this aspect, ten random numbers were chosen as seeds for the initialization and the average of all runs were calculated for comparison. All results shown were created by this method.

The alternatives available in RapidMiner are all MLPs (multi-layer Perceptrons) trained with backpropagation, which causes stochastic results.

The most simplistic implementation “Neural Net” cannot handle “polynomial” attributes, like text or dates. As this type of neural network provided only slightly better results compared to multiple linear regression, this approach was rejected. The neural net implementation is unable to evaluate polynomial values. Results showed that the “countries” attribute, which was non-evaluable, was the cause of its insufficient performance.

“Weka-MLP” is a more advanced algorithm that can be integrated into RapidMiner by using the Weka extension provided at the RapidMiner market place. It allows the analysis of any kind of data via the usage of a filter for transformation. Besides this, Weka-MLP also allows a more detailed configuration.

The third and last neural net alternative is called “Deep Learning”. It is based on a third-party implementation of H2O, which is an open source machine learning platform that allows the creation of models in the big data environment. The deep learning process is a powerful tool with numerous configuration options and additional features. The relevance of individual attributes of the data set can also be calculated and output, which provides further insights into the relationships between the attributes.

“Gradient Boosted Trees” are part of the H2O Framework as well. By gradually improving the prediction accuracy with optimization procedures, an "ensemble" of weak regression models created in this way becomes much more reliable than conventional regression tree algorithms [39].

To obtain a meaningful evaluation, the dataset created previously was split into two sets: training data and test data. Records before 2010 were used to train the model. Records from 2010 to 2014 served as validation data, in order to assess the model’s ability to reliably forecast future HDI values off the EFW index’s sub-indices. With this setup, two challenges emerged:

1. Since 2011, the security situation has deteriorated, especially in the Middle East, but also in Africa. Accordingly, some countries no longer followed the patterns learned by the model.
2. In 2010, the formula for calculating HDI was changed. The influence of the change was to be investigated in order to assess a possible influence on the model’s reliability.

To counter the discontinued development for the HDI and EFW for unstable countries, the strongest deviations in prediction accuracy were examined. Three countries, which had bad forecast accuracy that even got worse over subsequent years, were excluded from the dataset for modeling and testing. Syria suffered from civil war. Other exclusions were Iran and Central Africa.

The influence of the HDI’s calculation formula change was considered small by the United Nations [29]. To examine the influence of this change, a new model on a smaller dataset of values from before the conversion (2010) was created. After excluding Syria, Iran and Central Africa

from the analysis, the RMSE rose very slightly by up to 0.02 points (on a scale of 1 to 10). Thereby, either minimal or zero influence of the prediction by the formula change of the HDI can be assumed.

After optimizations of the dataset and the available algorithms were complete, the results of each could be compared directly (as seen in table 1). The RMSE illustrates the accuracy of the prediction, while the correlation describes the mutual development of the values.

Table 1: Results of forecasting accuracy on average for the used Algorithms

| | Weka-MLP | Deep Learning | Gradient Boosted Trees |
|-------------|-----------------|----------------------|-------------------------------|
| RMSE | 0.1216 | 0.1225 | 0.14 |
| Correlation | 0.998 | 0.997 | 0.997 |

Considering that this prediction is on a scale from 1 to 10 and created not through time series prediction but a completely different index, the forecast is quite accurate. This is also represented by the high correlation. Weka-MLP achieved the best results. Deep learning has shown a tendency to overfit during development. This result indicates that further confounding factors might still be hidden in the dataset. Even though they showed the worst result, gradient boosted trees were the only algorithm in RapidMiner’s library that could closely rival a neural net’s abilities in complexity. Nevertheless, gradient boosted trees prediction is still quite accurate.

Exploring the idea of creating an even more precise prediction, models for six individual countries were created with the result that they require too much effort to be further perused. The results will be discussed in Section 5.

4.3 Findings

A wide range of analytical approaches, algorithms and their implementations, results of modified data sets as well as the development of the data itself were examined. A direct dependence of the cumulated EFW index and the HDI could only be determined to a limited extent. Breaking the overall index into its components however, provided a useful basis for very accurate forecasts. Despite the advancing globalization, which is even a sub-component of the EFW index, the country reference in the overall model was of utmost relevance for a reliable prediction. Due to this, only algorithms that could evaluate non-numerical data were able to create a precise regression model.

The best forecasting results were provided by an overall model, from which countries without a reliable pattern were filtered out. Weka-MLP achieved the best results in this application, closely followed by the Deep Learning module.

Overall, based on the available data, a reliable prediction of the HDI value with the components of the EFW index as a basis is possible; the correlation of the models is very high. The initial RMSE of approximately 0.9 error points could be reduced by a factor of eight to approximately 0.12 through more complex algorithms, optimization of modeling and adaptation of the data set. These results indicate that, based on the available data, there is a strong dependency between the two indices. It should be noted that the filtering has shown an approximation of RMSE and the mean absolute error (MAE) which indicates a more stable prognosis.

As mentioned in Section 4.2 an in-depth evaluation for six specific nations was tested. It should be noted that forecasts for stable governments and countries were already reliable beforehand.

This approach resulted in slightly improved forecasts for predictions which were already precise, but a noticeable decline for forecasts of lesser accuracy. This indicates that international relationships have a greater influence on unstable nations and, if considered, thereby help to stabilize the prediction somewhat.

Cluster analysis has shown that HDI and EFW indexes complement each other at high levels of both indices and to a certain extent strike a balance between economic freedom and human development. However, dependencies between the two indices are not necessarily present in the mid to lower scale.

The 2010 HDI formula change in 2010 had only a negligible impact on the reliability of the forecast, especially after the countries Syria, Iran and the Central African Republic were excluded from the dataset.

The correlation and prediction accuracy between the liberal EFW index and HDI data is very high: Data and results show that economic freedom complements the development of a population. Due to the effects of economic freedom on a population as determined by the Fraser Institute, its connection with aspects of HDI and the results of the forecasts of both indices, a dependency between economic freedom and prosperity can be deemed probable, according to the available data. Based on the results of the analysis, the regulatory policy and supply orientation aimed for by ordo liberalists is of great importance for the prosperity of a population.

However, it is not possible to make a definitive statement on the relevance of the process policy (e. g. anti-cyclical economic policy), as it was not possible to draw sufficient differentiation from the data. The correlation of the model with the real data is almost perfect. However, if there are no further countries with uncertain patterns, this could be an indication of the additional relevance of process- or demand-oriented economic policy.

5. CONCLUSION

According to the available data, the importance of economic freedom for the prosperity and development of a country's population seems to be very high. Following this, regulatory policy clearly is important. Statements about the relevance of process policy, however, have proven to be rather difficult.

In general, it has been found that small sized governments do not bring any advantages in terms of quality of life and development, quite the contrary. This pattern is contrary to the views of ordo liberalism.

Machine learning algorithms are more complex than common statistical methods. The complexity of an economic system was approximated considerably more accurately, which led to a very high correlation of 0.998 points. Similar to statistical methods, only values for countries with sufficiently stable patterns could be predicted. This led to the exclusion of three countries from the dataset: Iran, Syria and the Central African Republic. Reliable patterns usually originate from stable states or governments, which makes it easier to predict values for these.

The patterns inherent in the data sets confirm the dependence of unstable nations on international partners. By looking at the complete dataset in an overall model, the forecasts for such countries became noticeably more accurate even if the model wasn't optimized and tuned just for their prediction.

Finally, a preference for ordo liberalism or Keynesianism could not be clearly defined as dependencies exist. Should the opportunity to obtain information on the primary character of economic policy decisions arise, a more precise analysis of regulatory or procedural policy could provide sufficient differentiation for further insights. Also, a study on inequality-adjusted HDI (IHDI) could lead to additional findings.

ACKNOWLEDGEMENTS

This research was conducted with support from a research grant from the Technische Hochschule Nuernberg Georg Simon Ohm, Germany.

REFERENCES

- [1] Derbal, H., Abdelkafi, R., & Chkir, A. (2011). The Effects of Economic Freedom Components on Economic Growth: An Analysis with A Threshold Model. *Journal of Politics and Law*, 4(2). <https://doi.org/10.5539/jpl.v4n2p49>.
- [2] Hall, J. C., & Lawson, R. A. (2014). Economic Freedom of the World: An Accounting of the Literature. *Contemporary Economic Policy*, 32(1), 1–19. <https://doi.org/10.1111/coep.12010>.
- [3] Gropper, D. M., Lawson, R. A., & Throne Jr., J. T. (2011). Economic Freedom and Happiness. *Cato Journal*. (31), 237–255. Retrieved from https://business.fau.edu/images/business/ourcollege/deans_office/dean_groppers_publications/Economic-Freedom-and-Happiness.pdf.
- [4] Nikolaev, B. (2014). Economic Freedom and Quality of Life: Evidence from the OECD's Your Better Life Index. (29). Retrieved from <http://borisnikolaev.com/wp-content/uploads/2014/08/Economic-Freedom-and-Quality-of-Life.pdf>
- [5] Alpaydin, E. (2010). Introduction to machine learning (2nd ed.). Adaptive computation and machine learning. Cambridge, Mass: MIT Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=307676>
- [6] Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*. (Volume:3, Issue: 3), 210.
- [7] Cleve, J. (2016). Data Mining (2nd ed.). De Gruyter Studium. Berlin: De Gruyter. Retrieved from <http://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4793920>
- [8] Rey, G. D., & Wender, K. F. (2011). Neuronale Netze: Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung (2., vollst. überarb. und erw. Aufl.). Bern: Huber. Retrieved from http://sub-hh.ciando.com/book/?bok_id=67571
- [9] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). An introduction to statistical learning: With applications in R (Corr. at 4. print). Springer texts in statistics. New York NY u.a.: Springer.
- [10] Luhmann, N. (1984). Die Wirtschaft der Gesellschaft als autopoietisches System. *Zeitschrift für Soziologie*, 13(4), 308–327. Retrieved from <http://www.zfs-online.org/index.php/zfs/article/view-File/2528/2065>
- [11] Dörner, D. (2000). Die Logik des Mißlingens: Strategisches Denken in komplexen Situationen (95. - 100. Tsd). Rororo: Vol. 19314. Reinbek bei Hamburg: Rowohlt.
- [12] Starbatty, J. (2017). Marktwirtschaft, Soziale Marktwirtschaft, Konrad-Adenauer-Stiftung. Retrieved from <http://www.kas.de/wf/de/71.11494/>

- [13]Farmer, K. (Ed.). (2006). Marktwirtschaft und Ethik: Vol. 10. Theorie der Wirtschaftspolitik, Entwicklungspolitik und Wirtschaftsethik: Festschrift für Werner Lachmann zum 65. Geburtstag. Wien: Lit-Verl.
- [14]Erhard, L., & Langer, W. (1964). Wohlstand für alle (8th ed.). Köln: Anaconda Verl.
- [15]Berger, J. (2012a). Soziale Marktwirtschaft. Retrieved from <http://www.bpb.de/politik/grundfragen/deutsche-verhaeltnisse-eine-sozialkunde/138633/soziale-marktwirtschaft>
- [16]Arnold, L. (2010). Unabhängige Wirtschaftspolitik. Wiesbaden: VS Verlag für Sozialwissenschaften.
- [17]Schüller, A. (2017). Ordoliberalismus Soziale Marktwirtschaft. Retrieved from <http://www.kas.de/wf/de/71.10255/>
- [18]Berger, J. (2012b). Soziale Marktwirtschaft. Retrieved from <http://www.bpb.de/politik/grundfragen/deutsche-verhaeltnisse-eine-sozialkunde/138633/soziale-marktwirtschaft>.
- [19]Scherer, J. F. (2016). Impulse - Freiheit, Gleichheit, Inklusivität: Der Ordoliberalismus als Ausgangspunkt einer Neuen Sozialen Ordnungspolitik. Retrieved from <http://library.fes.de/pdf-files/managerkreis/12642.pdf>
- [20]Döhrn, R. (2014). Konjunkturdiagnose und - prognose: Eine anwendungsorientierte Einführung. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [21]Fraser Institute. Economic Freedom Basics. Retrieved from <https://www.fraserinstitute.org/economic-freedom/economic-freedom-basics>
- [22]Gwartney, J., Lawson, R., & Fraser Institute. (2002). Economic Freedom of the World: 2002 Annual Report.
- [23]Fraser Institute. (2015). The Benefits of Economic Freedom. Retrieved from <https://www.fraserinstitute.org/economic-freedom/economic-freedom-basics>
- [24]Gwartney, J., Lawson, R., & Gartzke, E. (2005). Economic Freedom of the World: 2005 Annual Report.
- [25]Bologna, J., & Hall, J. (2014). Economic Freedom Research: Some Comments and Suggestions. The Annual, 2014, 123–136. Retrieved from https://www.beloit.edu/upton/assets/VOL_VI.Bologna.Hall.pdf
- [26]United Nations Development Programme. (2013a). Bericht über die menschliche Entwicklung 2013: Der Aufstieg des Südens: Menschlicher Fortschritt in einer ungleichen Welt. Bericht über die menschliche Entwicklung: Vol. 2013. Bonn: UNO-Verl.
- [27]United Nations Development Programme. (2017). About Human Development | Human Development Reports. Retrieved from <http://hdr.undp.org/en/humandev>
- [28]United Nations Development Programme. (2013b). Technical Notes. Retrieved from http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf
- [29]United Nations Development Programme. (2015a). The 2010 Human Development Index (HDI). Retrieved from http://hdr.undp.org/sites/default/files/hdi_training.pdf
- [30]United Nations Development Programme. (2015b). The 2010 Human Development Index (HDI). Retrieved from http://hdr.undp.org/sites/default/files/hdi_training.pdf

- [31]United Nations Development Programme. (2016). Human Development Index (HDI). Retrieved from <http://hdr.undp.org/en/content/human-development-index-hdi>
- [32]Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. (17). Retrieved from <https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>
- [33]Piatetsky, G. (2016). R, Python Duel As Top Analytics, Data Science software – KDnuggets 2016 Software Poll Results. Retrieved from <http://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html/2>
- [34]RapidMiner GmbH. (2017). Gradient Boosted Trees - RapidMiner Documentation. Retrieved from https://docs.rapidminer.com/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html
- [35]Piatetsky, G. (2013). KDnuggets Annual Software Poll:RapidMiner and R vie for first place. Retrieved from <http://www.kdnuggets.com/2013/06/kdnuggets-annual-software-poll-rapidminer-r-vie-for-first-place.html>
- [36]The R Foundation. (2016). R: What is R? Retrieved from <https://www.r-project.org/about.html>
- [37]University of Waikato. Data Mining with Open Source Machine Learning Software in Java. Retrieved from <http://www.cs.waikato.ac.nz/ml/weka/>
- [38]Mierswa, I. (2016). Introducing RapidMiner 7.2, Free Versions of Server & Radoop, and New Pricing | RapidMiner. Retrieved from <https://rapidminer.com/introducing-new-rapidminer-pricing-free-versions-server-radoop/>
- [39]Mason, L., Baxter, J. O., Bartlett, P. L., & Frean, M. R. (1999). Boosting Algorithms as Gradient Descent. Retrieved from <http://papers.nips.cc/paper/1766-boosting-algorithms-as-gradient-descent.pdf>

Authors

Jonathan Staufer completed his Bachelor of Science and Master of Science Degrees in Information Systems at the Technische Hochschule Nuernberg

Patricia Brockmann is a professor for Information Systems at the Technische Hochschule Nuernberg, Germany