

Executive Summary

Introduction

Over the past decade, the influence of social media has increased in many areas of our lives. Also in the financial world, social media has left its mark. More and more news is produced and consumed on such portals. Therefore, in recent years, the share of traditional financial information has been steadily declining, and the sources and volumes of alternative financial information are continuing to grow. Since news containing financial information moves the market, this alternative financial information must not be neglected. Therefore, news analytics plays a key role in financial markets. According to a growing body of research literature, news and media influence the investors sentiment and hence the stock prices (Tetlock (2007), Mitra and Mitra (2011), Leinweber and Sisk (2011), Barber and Odean (2011)). But since financial news from new sources is often in an unstructured and textual form, it is hard to quantify. Therefore, there is a need for effective models which are able to incorporate sentiment data.

The Swiss FinTech start-up Sentifi, which is short for “Sentiments in Finance”, helps to meet this need. Sentifi uses a combination of supervised and unsupervised machine learning to score the financial relevance of the content published by people on social media and blogs. The “Sentifi Engine” helps to extract and quantify sentiment data so that it can be easily applied to the investment decision-making process. The task of this thesis is to analyse the sentiment data provided by Sentifi and to check, whether it is possible to generate an excess return, also referred to as alpha by Jensen (1967). The research question, whether investing with collective financial intelligence does generate alpha, will be answered by developing and testing different strategies based on the sentiment data provided by Sentifi.

The question of whether it is possible to generate alpha with the help of sentiment data, both from traditional and new sources, has been a concern for scientists for a long time. While the followers of Fama (1970)’s efficient market hypothesis argue that capital markets are efficient (at least in the semi-strong form) and therefore, that stock prices fully reflect past stock prices and all publicly available information, various studies have shown that a sentiment-based trading strategy can achieve a certain excess return (e.g. Breitmayer, Pelster, and Massari (2016), Hafez (2011), Tetlock (2007), Lugmayr (2013), Leinweber and Sisk (2011)). The author states the hypothesis that it is possible to generate alpha with Sentifi’s unique data set. Thus, he also implies the hypothesis that the semi-strong form of the efficient market hypothesis is not given. By exploiting a very unique and new data set provided by Sentifi, this thesis contributes to the research on social media, crowd investing, sentiment analysis and stock markets, significantly differing from the existing literature.

Method

The data sample consists of all stocks which were listed in the S&P 500 index, the Stoxx Europe 600 index and the Swiss Performance index as of 01.12.2017. The sample period is from 01.12.2017 until 15.06.2019, which is the maximum period Sentifi can extract at the moment. This amounts to a total of 1'284 shares and a total of 738'468 daily Sentiment-Scores considered in the sample.

In order to test whether the data supplied by Sentifi can generate alpha, the author develops various strategies in a first step. These strategies are based on the Sentiment-Score - which indicates the crowd sentiment towards an asset - and the Attention-Buzz - a measure that indicates a company's above-average attention. These strategies are then tested in the form of a backtest over the observation period. Due to the short observation period, it is not possible to divide the data set by date. In order to still perform an out of sample test and thus minimize the problem of over-fitting, the data set is divided, and a sub-sample is created for each index. The strategies are only developed on the S&P 500 subset and afterwards applied to the other two subsets. In this approach - which could be called a "cross-section out of sample test" - the S&P 500 subset serves as the training set and the other two subsets serve as the test sets.

Since a backtest has only limited validity, the author examines in a further step the achieved portfolio returns with the concept of abnormal returns. The concept of abnormal returns is borrowed from the event study technique and makes it possible to isolate the effect of stock-specific reaction to the sentiment data (Hafez (2009)). To test whether a strategy generates excess return, time-series regressions are being run on a model consisting of the five Fama and French (2015) factors and the momentum factor from Carhart (1997). If the exposure to the respective factors captures all variation in expected returns, the intercept α_i would be zero. If there is excess return, the intercept is positive. To test the robustness of this model, further models were applied.

Results

The backtests carried out show that all the strategies developed beat the respective benchmark over the observation period. The executed time-series regressions prove that it is possible to generate an abnormal return with Sentifi's sentiment data, which doesn't just reflect more loadings on the economic risk factors. A long/short strategy based on the Sentifi Sentiment-Score and Attention-Buzz data generated a positive and significant alpha on the S&P 500, the Stoxx Europe 600 and the SPI Index between 0.1% (SPI & Stoxx Europe 600) and 0.3% (S&P 500) over the observed period. Furthermore, other strategies have also generated a positively significant alpha on various indices. These results confirm the hypothesis initially formulated and further imply that the semi-strong form of the efficient market hypothesis is not given.

Evaluation

This study presents interesting findings regarding the possibilities provided by the sentiment data of Sentifi. However, these results must take into account that the study deals with major and minor shortfalls. In particular, the transaction costs not taken into account, and the short observation period must be mentioned here. The transaction costs incurred would reduce the alpha generated to a large extent or eliminate it completely. Additionally, due to the short observation period, it is not possible to estimate how the strategies will perform in different economic cycles.

This and the great possibilities that the Sentifi Engine offers and will continue to offer in the future, however, provide the opportunity for further studies in this field. Especially the Sentifi event recognition tool and the combination of sentiment data and other trading relevant information, such as volatility analytics, allow the development and testing of additional exciting strategies that are left to further research.