Social Media Sentiment Analysis, Cryptocurrency Returns and Intraday Bitcoin Predictability

Inaugural-Dissertation

to obtain the degree of Doctor of Business Administration
(doctor rerum politicaum – Dr. rer. pol.)

submitted to the

Faculty of Business Administration and Economics
Heinrich-Heine University Düsseldorf

presented by

Lars Manfred Kürzinger, M.Sc.

Research Associate at the Chair of Business Administration,
esp. Financial Services
Heinrich-Heine University Düsseldorf
Universitätsstraße 1, 40225 Düsseldorf, Germany

Supervisor: Prof. Dr. Christoph J. Börner

Düsseldorf, 4th of June 2024

In loving memory of Wolfgang Ruthert & everything he has done for me

Dedication & Acknowledgement

I dedicate this dissertation to my family, whose unwavering support and encouragement have shaped me into who I am today. To my parents and grandparents, who have always been there for me with their guidance and assistance, this work is a testament to your love and belief in me. Especially to my girlfriend, Nadine Kranz, who has been my rock throughout the conception and execution of this work, standing by me through thick and thin.

I extend my deepest gratitude to my doctoral advisor, Christoph J. Börner, whose conceptual and expert guidance made this work possible. I am also indebted to my co-author and mentor, Dr. Ingo Hoffmann, whose insights and feedback have been invaluable throughout this journey. Special thanks to my colleagues at the Chair of Financial Services for their patience and support, particularly to my co-author and close friend, Philipp Stangor, whose collaboration significantly contributed to achieving my doctoral degree and continually challenged and enriched me intellectually.

I am grateful to all discussants of my various research works and participants of diverse conferences and workshops for their valuable insights and contributions. I also thank my second examiner, Prof. Dr. Eva Lutz, for her efforts and fruitful discussions.

Düsseldorf in the year 2024

Lars M. Kürzinger

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Anderson-Darling
ADF	Augmented Dickey-Fuller test
AG	Stock Company
ANOVA	Analysis of Variance
API	Application Programming Interface
ARCH	Autoregressive Condiional Heteroskedasticity
BERT	Bidirectional Encoder Representations from Transformers
BTC	Bitcoin
CC	Cryptocurrency
CEO	Chief Executive Officer
cf.	Confer
CvM	Cramer von Mises
D.C.	District of Columbia
Def.	Definition
Diff	Difference
DOGE	Dogecoin
e.g.	For example
EMH	Efficient Market Hypothesis
ELMo	Embeddings from language model
EmoLex	NRC-Emotion Association Lexicon
Eq.	Equation
et al.	Et alii
etc.	Et cetera
EWCI	Equal-Weighted Cryptocurrency Index
EWCI-	Equal-Weighted Cryptocurrency Index
Fig.	Figure
FLO	FLO
F-Stat	F-Statistic

FTC	Feathercoin
GED	Generalized extreme value distribution
GI	General Inquirer
GLD0	Generalized logistic distribution type 0
GLD3	Generalized logistic distribution type 3
GLoVe	Global Vectors
GPD	Generalized Pareto distribution
GPT	Generative Pre-trained Transformer
H	Hypothesis
HAC	Heteroskedasticity and Autocorrelation
HDS	Hartigans Dip Statistics
heiCAD	Heine Center for Artificial Intelligence and Data Science
HIT	Human intelligence task
HVB	Hypovereinsbank
i.a.	In accordance
i.e.	Id est
IFC	Infinitecoin
IID	Independent and Identically Distributed
IPO	Initial Public Offering
KS	Kolmogorov Smirnov Distance
LOCF	Last Observation Carried Forward
LM	Loughran and McDonald
LT	Lower Tail
N	Normal Distribution
NASDAQ	National Association of Securities Dealers automated quotations
NLP	Natural Language Processing
NXT	Nxt
p.m.	Post meridiem
QRK	Quark
Sec.	Section

SDI	Stable Distribution(s)
S&P	Standard and Poor's
Tab.	Table
TF-IDF	Term frequency - inverse document frequency
TRM	Text Represantation Method
USA	United States of America
VADER	Valence Aware Dictionary and sEntiment Reasoner
WDC	Worldcoin
Word2Vec	Word to Vector
XPM	Primecoin
XRP	Ripple

List of Symbols

Latin Symbols

a	Indirect effect
b	Indirect effect
Be	Bearish
Bu	Bullish
c	Direct effect (in Section 2)
c	Drift coefficient (in Section 4)
c'	Direct effect
$Closing$	Closing Price
$Corr$	Correlation
E	Expected value
EL	EmoLex Dictionary
EM	Emotion
F	Empirical distribution function
$\hat{F}$	Probablity function
$\hat{f}$	Function
F_Sen	Financial Sentiment
GI	Harvard GI Dictionary
HE	Henry Dictionary
i	Counting index
$Intraday$	Intraday return
j	Counting index
k	Counting index
LM	Loughran McDonald Dictionary
m	Time indicator
N	Negative Financials (in Section 2)
N	Number (in Section 4)
N	Normal distribution

n	Counting index
NN	Negative financials and negative tweets
NP	Negative financials and positive tweets
$Opening$	Opening Price
P	Positive Financials
P_t	Stock Price
p	Lag coefficient
PP	Positive financials and positive tweets
PN	Negative financials and positive tweets
PN_i	Positive/Negative Dictionary i
q	Time shift
R^2	Determination coefficient
$\hat{R}$	Weighted mean square error
r	Return(s)
S	Stable distribution function
SC	Stock Capital
$Sentiment$	Investor Sentiment
$sign$	Signum function
T	Return Observations (in Section 2)
t	Time intervall
T_Sen	Tweet Sentiment
u	Tail truncating threshold
v	Conditional Value at Risk
VaR	Value at Risk
w	Weight function
y	Return time series
X	Random variable
Z	Z-Score

Greek Symbols

$\boldsymbol{\alpha}$	Shape parameter
$\hat{\boldsymbol{\alpha}}$	Estimated shape parameter
β	Skewness Parameter
$\bar{\beta}$	Average coefficient
β_i	Regression coefficient i
$\hat{\beta}_i$	Estimated regression coefficient i
$\tilde{\beta}$	Standardized regression coefficients
$\hat{\beta}$	Estimated skewness Parameter
γ	Scale Parameter
$\boldsymbol{\delta}$	Location Parameter (in case of SDI)
$\boldsymbol{\delta}$	Deterministic trend coefficent (ADF test)
ϵ	Innovation process
$\boldsymbol{\mu}$	Mean
$\hat{\boldsymbol{\mu}}$	Estimated mean
$\boldsymbol{\xi}$	Tail parameter
$\hat{\boldsymbol{\xi}}$	Estimated tail parameter
π	Pi
$\boldsymbol{\sigma^2}$	Variance
$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$	Estimated variance
φ	Autoregressive coefficient

Mathematical Symbols

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of non-negative real numbers
$\triangle$	Change

1. Introduction

In the wake of rapid technological advancements and the digitization of society over recent decades, new inquiries have emerged in the field of economics, necessitating novel methodologies utilizing vast amounts of data. The introduction of Bitcoin byNakamoto (2008) marked the genesis of a crypto-economy, setting the stage for subsequent developments in cryptocurrency (CC) (Sixt (2017)). Initially conceived as a currency, CCs and especially Bitcoin evolved into an additional asset class due to their unique characteristics, such as the potential absence of intrinsic value (Baur et al. (2018b)), prompting inquiries into their suitability as investments and the analysis of their future behavior.

Concurrently, the popularity of social media platforms has surged, becoming significant hubs for communication, news consumption, and entertainment. The increased usage and participation in these online services generate substantial amounts of data, which can be analyzed for various purposes. One such method of analysis is sentiment analysis, which evaluates the underlying mood of social media messages. These sentiments can then be used to assess and predict the behavior of assets, such as CCs.

Against this backdrop, this dissertation explores the influence of sentiment on individuals' investment decisions, the potential improvement offered by multidimensional sentiment analysis compared to two-dimensional approaches, and the use of sentiment in predicting Bitcoin returns, particularly within intraday intervals. To address these inquiries systematically, this dissertation is structured into six sections.

The first section lays the groundwork by discussing the unique characteristics of CCs and their heavy tails, attempting to represent the tails and bodies of return distributions from a basket of CCs using a combination of two distribution functions. Additionally, it provides an overview of sentiment analysis, including its development, various methods, and its utilization for predicting asset price movements, while also addressing open research questions.

Building upon this foundation, section 2 demonstrates, through the analysis of a specially designed experiment, that sentiment influences individuals' investment decisions, thereby justifying its usage in predicting asset price developments. Section 3 explores the superiority of multidimensional sentiment analysis based on dictionaries over two-dimensional analysis in various aspects and provides recommendations regarding emotion-based sentiment analysis.

Subsequently, section 4 delves into the analysis of return distributions of 27 different CCs. Utilizing a methodology developed by Hoffmann and Börner (2021), it subdivides the individual return distributions into body and tail sections, employing two distributions to depict these segments. This approach accounts for the heavy tail characteristic of CCs, which poses a challenge in both practical and retail investor contexts.

Section 5 four examines the use of sentiment to predict intraday Bitcoin returns. It reveals

that a span from 18 to 108 minutes appears particularly promising for predicting BTC returns using sentiment.

Finally, Section 6 five concludes with a summary of the findings and outlines future research tasks.

1.1. Information Efficiency

The question of whether capital markets are informationally efficient has occupied the academic literature since the establishment of the Efficient Market Hypothesis (EMH) by Fama (1970). According to this hypothesis, the main function of capital markets is the efficient allocation of capital between the demand and supply sides. A market is considered efficient if the prices at all times fully reflect all available information (Fama (1970); Wurgler (2000)). This requires that all market participants always have the same information, interpret it in the same (rational) way, and immediately act based on this information. As a result, no systematic excess return can be achieved through any kind of information, as this information is reflected in the behavior of all market participants. Therefore, no market participant should have monopolistic access to information to allow such a market to exist, at least in theory. The existence of insider information however, which can be held by market makers or company employees, is noted by both Fama (1970) and Scholes (1969).

In addition to this strictly informationally efficient form of a capital market, two other forms with less stringent conditions can be distinguished. A semi-strong form is conceivable, in which all publicly available information is considered in asset pricing. Consequently, no systematic excess return can be achieved through the analysis of fundamental data or technical analyses.

Lastly, Fama (1970) discusses a third form of market informational efficiency. In this weak form, only historical data are fully reflected in asset prices. Therefore, while no systematic excess return can be achieved through technical analysis of past information, it can still be achieved through fundamental analysis and insider information (Fama (1970)).

Whether financial markets are informationally efficient, and if so, to what extent, has been a subject of academic inquiry at least since Fama's foundational typology. The findings of relevant literature addressing this question are, however, inconclusive. There are various studies that suggest efficient capital markets (see, i.a., Jensen (1978); Tirole (1982); Rubinstein (2001); Malkiel (2005))[1], while others reject this assumption (see, i.a., Lehmann (1990); Shiller (2000); Shleifer (2000); Lee et al. (2010)).

[1]For a more detailed overview of the history of the EMH, see Sewell (2011).

1.2. Behavioral Finance

However, it became apparent that the insights gained from these theories are not suitable for explaining all phenomena observed in the financial market. For instance, Shiller (2003) finds that the EMH cannot provide a satisfactory explanation for the development of dividends and volatility in the overall market, particularly in the 1980s. Such anomalies, which can be viewed as deviations from the expectations of the EMH, consequently question its validity and the assumptions and models of neoclassical economics as a whole.

In general, three types of capital market anomalies are particularly in the focus of research. Firstly, there are fundamental anomalies, where the expected value of an asset deviates from the observed value based on the available fundamental information (Chatterjee and Maniam (2011)). To form fundamental expectations, one can consider, for example, price-to-earnings (de Bondt and Thaler (1985)), price-to-book (Mitchell and Stafford (2000); Fama and French (1992)), price-to-sales (Desai et al. (2004)), or dividend yields (Keim (1985)).

Furthermore, technical anomalies are observed, which use past price information of an asset to predict its future development. For such analysis, moving average methods (Brock et al. (1992)), trading range breaks that generate buy or sell signals when assets surpass their previous highest or lowest valuation (Brock et al. (1992)), and momentum analyses that utilize short-term autocorrelation observations of asset prices (Hon and Tonks (2003)) can be employed. According to Fama (1970), such effects should not be possible in a weakly efficient market, thus contradicting the EMH.

Additionally, calendar anomalies can be observed in financial markets, leading to abnormal returns at certain periods or points in time. These can be seasonal or occur on specific days, weeks, or months of the year, thereby also contradicting the EMH (Boudreaux (1995)). For example, Gibbons and Hess (1981) and Smirlock and Starks (1986) find abnormally low returns on Mondays and abnormally high returns on Fridays. Agrawal and Tandon (1994) also observe this phenomenon but note that it occurs on different weekdays for different countries. A possible explanation for this observation could be related to market opening hours, so the effect observed on Mondays could be a form of compensation for weekend days when no trading takes place (Rogalski (1984)).

A similar pattern emerges when considering returns at the beginning and end of the month. Ariel (1987) finds higher average returns at the beginning of the month, while there is typically no change at the end of the month. However, this effect is not consistent across countries (Jaffe and Westerfield (1989)). A possible explanation for this observation is that investors build short positions primarily at the end of the month and reevaluate and stock up on new shares at the beginning of the new month. A similar argument can be observed in the so-called turn-of-the-year effect, which suggests that stocks show positive abnormal

returns in the first/last January/December months (Wachtel (1942); Rozeff and Kinney (1976); Keim (1983); Agrawal and Tandon (1994)). Such observed effects could arise from window-dressing, where investors try to present a positive overall trading balance at the end of the year (Agrawal and Tandon (1994)) or from tax optimization considerations. Potential liquidity surpluses from investors at the end of the year could also contribute to explaining these observed anomalies (Branch (1977); Ligon (1997)).

Besides the presented anomalies, other phenomena are also observed in the literature, such as the Super Bowl Indicator Stovall (1989) or the influence of weather on investor behavior (Stovall (1989); Cao and Wei (2005); Chang et al. (2008)).

In efficient markets, these deviations of asset prices from their fundamental value should be exploited risk-free and unlimitedly by rational investors, so no arbitrage opportunity would exist.[2] However, it becomes apparent that this is not generally the case in financial markets. The possibility of arbitrage, especially with hard-to-value assets or assets with high price volatility, is by no means risk-free and makes it difficult for professional arbitrageurs to exploit these suspected price differences. Consequently, anomalies and observed price differences may persist because arbitrageurs are unwilling to take on the risk associated with exploiting the differences (Shleifer and Vishny (1997)).

These limits of arbitrage and the anomalies observed in financial markets contradict the EMH and have led scholars to critically question the assumptions of the EMH regarding investors. In this context, investors' decisions have increasingly been considered in light of psychological aspects, moving away from the assumption that investors act rationally.

Thus, the assumption of risk-neutral investors has been critically questioned (Palomino (1996)), and more emphasis has been placed on risk-averse investors, along with varying utility functions in the sense of Pratt (1964).

Furthermore, there has been an increased focus on potential behavioral biases that consciously or unconsciously influence investors' investment decisions. Festinger et al. (1956) already discuss cognitive dissonance, the tendency of individuals to hold on to their beliefs despite evident facts (Festinger et al. (1956); Akerlof and Dickens (1982)). Additionally, Tversky and Kahneman (1973) and Tversky and Kahneman (1974) show that individuals tend to make decisions based on heuristics. For example, subjects may make incorrect assessments of the likelihood of events occurring based on the information available (availability heuristic) or the erroneous assumption that uncorrelated events are representative of each other (representativeness heuristic). Decision-making may also be influenced by anchoring effects, where

[2] It does not necessarily have to be assumed that all investors act rationally. A small number of professional and rational financial market participants would theoretically suffice to exploit the deviation of asset prices from their fundamental value and correct the price back to the fundamental value (Shleifer and Vishny (1997)).

economic agents set a mental anchor to evaluate future developments (anchoring effect).

Based on these insights, the authors subsequently developed the Prospect Theory, which thus represents an alternative to the expected utility theory (Kahneman and Tversky (1979)). The Prospect Theory replaces the probabilities previously used in the literature with decision weights and shows that individuals perceive losses and gains of the same magnitude differently, with a loss of the same magnitude as a gain causing a stronger utility loss than the gain causing a utility gain. Consequently, the utility function is not symmetric but takes on a convex shape in the loss area and a concave shape in the gain area (Kahneman and Tversky (1979)). In this context, Thaler (1980) first observed the endowment effect, where investors hold loss positions for too long to avoid realizing a potential loss while selling gain positions too early out of fear of losing a potential gain and thus avoiding emotional damage (Benartzi and Thaler (1995)).

Aside from the heuristics and the Prospect Theory presented so far, other effects that can be classified as behavioral biases have also been observed. These are based on the cognitive limitations of individuals that influence their investment decisions. In the relevant literature, herd behavior (Trueman (1994)), i.e., blindly following the prevailing market opinion without factual evidence, overconfidence, the overestimation of one's own abilities (Frank (1935); Langer and Roth (1975); Miller and Ross (1975); Barber and Odean (2000); Gervais and Odean (2001)), and mental accounting (Thaler (1980)), the separate tracking of goals, are observed and discussed.[3]

According to Tversky and Kahneman (1974), thinking in heuristics and the resulting behavioral biases affect not only laymen but also experts (in Tversky and Kahneman (1974), these are scientists and statisticians) when they are required to make intuitive decisions in complex situations.

In summary, it can thus be argued that financial market participants behave more risk-averse rather than risk-neutral, as assumed by, among others, Sharpe Fama (1970). Furthermore, their behavior is less in line with that of rational investors in the sense of a homo economicus but can be seen as irrational in the sense that, despite given information, this information is interpreted differently due to individual psychology and hence deviating actions are derived from the available information based on cognitive dissonance, heuristics, and behavioral biases (Hirshleifer (2001)).

Although there is still a lively debate about the presence of informational efficiency in markets, the degree to which they are informationally efficient, and whether arbitrage

[3] A detailed presentation and discussion of the development of Behavioral Finance and further behavioral biases can be found in Fu (2022).

reliably corrects price deviations from the fundamental value, it can be assumed based on the developments presented in the academic literature that due to the individual psychology of investors, deviations from the fundamental value can occur repeatedly.

1.3. Sentiment Analysis

1.3.1. Fundamentals

Additionally, within the realm of behavioral finance, deviations between stock prices and their intrinsic values can arise due to potentially irrational investor behaviors. Bullish or bearish sentiments among noise traders could consequently sway stock prices (De Long et al. (1990)). For instance, individuals may tend to overestimate the value of opinions from conversation partners (DeMarzo et al. (2003)) or exhibit a heightened willingness to invest in specific assets simply because they have captured their attention, consciously or unconsciously (Barber and Odean (2008)). These behavioral biases may contribute to the allure of social media sentiment analysis in financial contexts, offering insights into the erratic behaviors of individuals (Black (1986)) while also elucidating the reasons behind individuals' engagement on platforms like StockTwits or Twitter, where they openly express their beliefs.

In scenarios depicted by DeMarzo et al. (2003) and Giannini et al. (2018), it could even be deemed rational for institutional investors, often assumed to be less prone to biases, to follow opinion leaders, recognizing their potential to influence market movements or even become influential themselves. Moreover, interpersonal communication among market participants seems conducive to persuading hesitant investors to commit to particular assets, as they become aware of others sharing similar investment views (Cao et al. (2002); Antweiler and Frank (2004)). Understanding these behavioral patterns might even incentivize individuals to deliberately propagate rumors about assets, aiming to profit from anticipated reactions by their followers (van Bommel (2003)), thereby shedding light on the motivations behind informed investors' decisions to disseminate information (Xiong et al. (2019)). So ultimately, bullish or bearish sentiments among noise traders possess the capability to impact stock prices De Long et al. (1990); Black (1986).

Therefore, in the context of my dissertation, I am particularly drawn to the insights of Stiglitz and Grossman (1980) within behavioral finance. They propose that market participants can garner slight excess returns through continual information gathering, a notion that challenges the concept of market efficiency outlined by Jensen (1978). These surplus returns are construed as incentives for monitoring and analyzing market data, offsetting the costs involved in processing and interpreting market signals. Nevertheless, within a competitive market milieu, these marginal excess returns are expected to be short-lived. This expectation stems from the assumption that professional investors will swiftly exploit any pertinent

information to gain a competitive edge. Thus, sentiment analysis forms an interface between market efficiency and the incentive for information acquisition (Renault (2017)). Since asset prices can temporarily deviate from their fundamental value due to sentiment-driven noise traders (De Long et al. (1990)) and, as discussed in section 1.2, there can be limits to arbitrage (De Long et al. (1990); Pontiff (1996); Shleifer and Vishny (1997)), the question regarding the influence of sentiment on asset prices is no longer whether sentiment affects prices, but how strong this effect is and which measurement method should be used to determine sentiment (Baker and Wurgler (2006)).

In the context of this dissertation, sentiment is understood as the general psychological mood in markets. Several proxies for investor sentiment have been used in the literature to predict price movements. For example, surveys among investors, such as the Consumer Sentiment Index, can be used to gauge investor mood in the markets and subsequently use this for price prediction (Brown and Cliff (2005)). However, these surveys are often conducted infrequently, leading to low-frequency data that is inadequate for assessing short-term excess returns. Moreover, there's a lack of strong motivation for participants to provide truthful responses, introducing potential biases into survey outcomes (Singer (2010)).

Furthermore, attempts can be made to determine investor sentiment using market data. For instance, Lee et al. (1991) and Neal and Wheatley (1998) use the discount on closed-end funds as a sentiment indicator, while Baker and Wurgler (2006) adds other market indicators such as first-day IPO returns, the equity share in new issues, the dividend premium, and the share turnover of the New York Stock Exchange. The sentiment is approximated based on the behavior of market participants. However, this method of sentiment measurement suffers from the influence of many different factors affecting investor behavior and their interdependencies (Qiu and Welch (2004); Da et al. (2015)).

For these reasons, this dissertation focuses on using text analysis in the social media domain to generate a proxy for investor sentiment. The advantage of using social media posts for sentiment analysis lies partly in the volume of available data. Due to the widespread use of social media platforms such as Twitter or, in a financial context, StockTwits, and the sheer amount of user-generated content, social media data enables high-frequency sentiment calculations, which are needed to address the previously postulated short-term influence of sentiment. Another advantage is the 'living lab' characteristic of the data, meaning that, unlike in surveys, there is no incentive-driven bias expected in the results (Renault (2017)).

Various methods exist for using Natural Language Processing (NLP) to analyze sentiment, differing in complexity and capable of analyzing different aspects of texts concerning the present sentiment as development progresses. These methods and their characteristics will be presented in the following Section 1.3.2.

1.3.2. Methods of Social Media Sentiment Analysis

The analysis of sentiment in given texts was initially conducted through text representation methods[4] concerning the underlying sentiment, which involves identifying and categorizing expressed opinions and the corresponding attitude of a subject regarding a specific topic or product (Mishev et al. (2020)). In the field of psychology, this initially involved matching individual words using a pre-created lexicon that assigns a connotation to each existing word. Typically, all words within a text that do not contain valuable information (e.g., filler words) are first identified and removed. A common issue encountered when utilizing word lexicons to determine sentiment scores (especially in a social media context) is the omission of various word formats. For instance, a matching algorithm might overlook the word 'lovers' if only the root 'love' is included in the lexicon. To address this issue, linguistics offers two potential solutions: stemming and lemmatization. Stemming algorithms aim to identify a word's root by identifying and removing suffixes (e.g., transforming 'lovers' to 'lover' and 'loves' to 'love'), while lemmatization seeks to group inflected forms into a single category (e.g., transforming 'lovers' and 'loves' to 'love') (Mechura (2016); Renault (2017)).

Each remaining word can then be assigned a connotation (usually positive/negative). Thus, a positive (negative) sentiment for the considered text would result if the number of words classified as positive exceeds the number of words classified as negative (see, i.a., Antweiler and Frank (2004), Baker et al. (2007), Gao and Yang (2017), Kim and Kim (2014) and Sun et al. (2016)).[5]

In the financial sector, commonly used dictionaries include the Harvard General Inquirer (GI) as used by Tetlock (2007); Engelberg et al. (2012); Da et al. (2015); Mishev et al. (2020) and the dictionary by Loughran and McDonald (2011) (LM). The former has no finance reference, while the latter was specifically developed for evaluating financial reports and has been frequently used in research in this context (Dougal et al. (2012); Engelberg et al. (2012); Kearney and Liu (2014); Chen et al. (2014); Da et al. (2015)).

To automatically create word lists and count the number of individual words within a text, Count Vectorizer or Term Frequency-Inverse Document Frequency (TF-IDF) can also be used. While Count Vectorizer can list and count all unique words or features within a text, TF-IDF penalizes the frequent use of words or features and thus weights them differently. Both methods serve to analyze texts and can freely search for desired features. They often serve as a basis for complex machine learning algorithms (Mishev et al. (2020)).

However, both Count Vectorizer and TF-IDF are not capable of analyzing semantic rela-

[4] A representation of the individual methods and the most commonly used models can be found in Fig. 1. A more detailed overview of the application, strengths, and weaknesses can be found in Mishev et al. (2020).

[5] A more detailed presentation of sentiment analysis using dictionaries can be found in Section 3.5.3.

tionships and the context of a given text, which is essential for understanding the sentiment and the author's intent. For this reason, word encoders have been developed, which convert the words of the text into high-dimensional vectors (embeddings) and then place them in the context of the sentence. This method, also known as distributional semantics, can identify words used in similar contexts and assign them similar information vectors (Landauer and Dumais (1997); Sahlgren (2006); Mikolov et al. (2013a); Turney and Pantel (2010)). The most popular word encoders include Word2Vec, GloVe, FastText, and ELMo (Mishev et al. (2020)).

While GloVe offers improvements over Word2Vec by considering the co-occurrences of words, neither word encoder can handle unknown words that were not included in the training of the models (Mikolov et al. (2013b); Pennington et al. (2014)). This problem, however, can be addressed by FastText Bojanowski et al. (2017), an extension based on Word2Vec, as well as by ELMo (Sarzynska-Wawer et al. (2021)). ELMo stands out because it can determine a sentence-specific context for the analyzed words, allowing the same word to be assigned different contexts in different sentences (Sarzynska-Wawer et al. (2021)).

A similar application is found in so-called sentence encoders, which deal not only with the context of individual words but also with the meaning and context of entire sentences, representing an advancement of word encoders Mishev et al. (2020). Some of these models, such as the Universal Sentence Encoder (Cer et al. (2018)), can even extend this process to entire paragraphs.

While the previously presented text representation methods (TRMs) provide satisfactory results for analyzing text corpora, they lack context-based mutability, leading to weaknesses in interpreting the same words in different contexts. This issue can be resolved by NLP transformer models. One such model is the Bidirectional Encoder Representations from Transformers (BERT). This transformer model uses a bidirectional context understanding, determining the context in the text both forwards and backwards, thus improving text comprehension. Additionally, individual words can take on different meanings depending on the sentence and text. Ultimately, due to its structure, BERT can also be trained for various tasks. BERT can be used to predict masked words with probabilities or to predict following sentences. Moreover, the base BERT model can be fine-tuned depending on the task at hand. This involves fine-tuning the existing model with new data, creating a universal application possibility (Devlin et al. (2018); Mishev et al. (2020)) and training the model, for instance, on financial language (FinBERT) (Araci (2019)).

An advancement of the standard BERT model is DistilBERT, which enhances the performance of the BERT model (60% faster) based on knowledge distillation while retaining 97% of its language understanding capabilities.

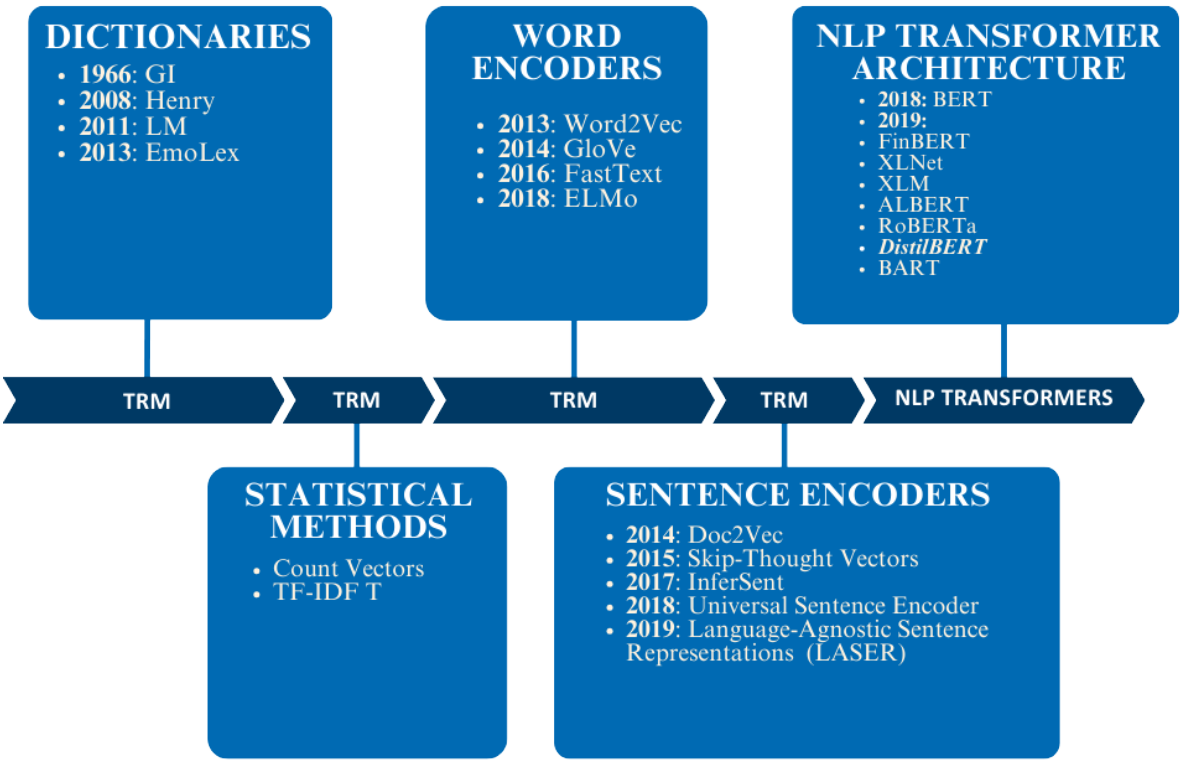


Fig. 1 Developement of Sentiment Analysis

Apart from the technical advancements in sentiment analysis, there have also been parallel developments in the field of cryptography, impacting financial markets. This topic will be addressed in the following section.

1.4. Cryptocurrencies

1.4.1. Fundamental Information

The global financial crisis of 2007 shook the trust in financial markets and their institutions, highlighting the inability of regulatory authorities and central banks to prevent such a crisis or effectively limit its effects. Against this backdrop, the year 2008 saw the introduction of the first CC, Bitcoin, which to this day holds approximately 53% market share according to Coinmarketcap.com and was presented in a white paper by an individual or group of individuals using the pseudonym 'Satoshi Nakamoto' , marking the beginning of the development of the CC market (Nakamoto (2008); Bariviera et al. (2017); Bouri et al. (2017a); Kaya Soylu et al. (2020)).

CCs can be viewed as a libertarian response to the perceived failures of financial market institutions. Originally conceived as a peer-to-peer electronic cash system, they leverage a shared computer/network architecture of hardware, such as processing or storage capacities, enabling users to operate across borders independently of governmental institutions

(Nakamoto (2008); Weber (2014); Bariviera et al. (2017); Baur et al. (2018b)). The use of distributed ledger technology, employing a blockchain, constitutes the core innovation of CCs in this context (Bariviera et al. (2017)).

Conventional payment methods rely on a central authority, acting as a clearinghouse, to verify and validate transactions. These central entities typically consist of financial intermediaries, governments, or other regulatory institutions, aiming to ensure smooth payment processes, market functionality, and fraud prevention (Bariviera et al. (2017); Kim (2020); Wingreen et al. (2020)). In contrast, CCs utilize a different guarantee mechanism based on automatic authorization and verification by other network members. Half of the network members verify transactions between two participants by storing transaction data on their computers as a guarantee for future payments. Decentralized verification across a multitude of users aims to make retroactive adjustments virtually impossible. Each transaction is timestamped and forms a record of transactions (Nakamoto (2008); Shi (2016); Kim (2020)). Bitcoin's design, inspired by gold's fixed supply, limits the total number of coins to 21 million. To circulate these coins, they must be 'mined', a process that rewards network nodes for facilitating transactions. Every 10 minutes, a certain amount of new bitcoins is released. Participants invest in computing power to compete for these new coins by solving cryptographic puzzles posed by the system whenever a transaction is made. The first node to solve the puzzle timestamps the transaction, proving it involves a unique coin that hasn't been spent before. This proof is added to a public ledger, which tracks all transactions to prevent double spending while keeping participants' identities anonymous. The updated ledger is verified by the network and then copied by all nodes, maintaining a decentralized transaction history. Successful miners receive bitcoins as a reward, but as more bitcoins are released, the puzzles become harder and the rewards decrease. Thus, Bitcoin's money supply is intrinsically linked to its payment system operations (Weber (2014)).

Due to their decentralized nature, CCs are not subject to geographical restrictions, enabling anonymous international transactions across borders and jurisdictions. In addition to worldwide applicability and anonymity, other advantages of this technology include low transaction costs due to the peer-to-peer design and independence from the existing financial system and its institutions (Bariviera et al. (2017), Baur et al. (2018b); Corbet et al. (2018b, 2019); Kakinaka and Umeno (2020)). However, alongside the perceived benefits of anonymity and independence from the financial system, there are also challenges. CCs are often associated with (cyber-)crime and money laundering and could potentially contribute to the destabilization of the financial system due to their unique characteristics, as discussed in the next chapter.

1.4.2. Characteristics

Although CCs, as the name suggests, were originally conceived as currencies, their current suitability as currency in the classical sense is doubted both by user behavior and within academic literature. Studies such as Baur et al. (2018b), Glaser et al. (2014), and Cheah and Fry (2015) show that a significant portion of Bitcoin users are considered inactive or dormant, suggesting that Bitcoin is not actively used as a means of payment but rather viewed as a long-term investment and ultimately as an asset.

Moreover, academic literature questions the suitability of Bitcoin, representing most other CCs[6], as a currency instrument. According to general consensus, a currency should fulfill three functions. A currency should therefore be a store of value, a unit of account, and a medium of exchange (Keynes (2011 [1930]); Ingham (2004); Weber (2014); Cheah and Fry (2015); Dwyer (2015); Bariviera et al. (2017); Baur et al. (2018b)). Whether Bitcoin fulfills all of these functions is questionable. Considerable volatility compared to traditional currencies and assets (Polasik et al. (2015); Balcilar et al. (2017)) and the associated heavy tails of the return distribution (Osterrieder et al. (2017); Gkillas et al. (2018); Gkillas and Katsiampa (2018)) pose a problem in this context. Due to Bitcoin's rapid value changes, prices for goods and services denominated in Bitcoin need to be continually updated and fluctuate significantly. Furthermore, the fulfillment of the store of value function is dubious, as volatility can lead to significant losses of value in this potential currency in a short time. Additionally, CCs are not backed by a governmental entity, and their value is not guaranteed (van Alstyne (2014); Cheah and Fry (2015); Yermack (2015)).

Thus, Bitcoin is primarily seen as a speculative asset (Cheah and Fry (2015); Weber (2014); Baur et al. (2018b); Dwyer (2015)), whose value is derived especially from buyers and sellers (Baek and Elbeck (2015)). The view of CC as its own asset class is confirmed by the fact that while CCs exhibit high correlation among themselves, they have very low correlation with other asset classes. Consequently, they are suitable for portfolio diversification and risk reduction within an investor's portfolio, as they react differently to external events and shocks, and thus can have a positive portfolio effect especially in times of crisis (Baur et al. (2018b); Corbet et al. (2018c)). However, based on these unique properties distinct from other asset classes, the asset class of CCs carries its own idiosyncratic risk that is difficult to hedge against (Corbet et al. (2018c)).

[6]As a multitude of CCs exist, Bitcoin is meant to represent the asset class of CCs, eventhough there exist technical differences between the Bitcoin and other CCs.

1.4.3. Valuation

However, in order to establish CCs, including Bitcoin, as a widely accepted asset class utilized by institutional investors, the characteristics of CCs and Bitcoin in particular, their value developments, and the associated risks must be analyzed to derive insights for investors and recommendations for risk management (Majoros and Zempléni (2018); Osterrieder et al. (2017); Gkillas et al. (2018); Gkillas and Katsiampa (2018); Corbet et al. (2018c)).

Thus, the question arises as to how CCs differ from traditional assets. The most obvious criterion for differentiation lies in assessing the true value of CCs and Bitcoin. Traditional asset pricing models and standard risk factors commonly utilize fundamental data such as cash flows, earnings, or dividends to evaluate stocks or other asset titles. However, in the case of CCs and Bitcoin in particular, this approach is challenging as CCs neither generate cash flows nor declare earnings or pay dividends. Consequently, the question arises regarding a valuation basis for this novel asset class (Liu and Tsyvinski (2021)).

This issue is unanimously discussed in academic literature but has not yet been conclusively answered. Various researchers arrive at different conclusions when addressing this question. For instance, among others, Cheah and Fry (2015) argue that the fundamental value of Bitcoin is 0 due to the lack of a fundamental basis, thus potentially explaining the formation of bubbles (Cheah and Fry (2015); Corbet et al. (2018a); Liu and Tsyvinski (2021)). On the other hand, Dyhrberg (2016a) and Hayes (2019) conclude that Bitcoin does indeed possess an intrinsic value. The value of Bitcoin is therefore attributed, in part, to the costs of mining in terms of the decentralized resources required, electricity used, and the underlying blockchain technology. In this context, the valuation of Bitcoin resembles that of gold in some respects. The costs for gold, for example, also arise from the decentralized resource extraction by multiple private providers, allowing comparisons with Bitcoin since both assets are scarce and difficult to mine.

Nevertheless, Dyhrberg (2016a) concludes that these factors are insufficient to explain the observed prices, thus indicating an overvaluation of Bitcoin.

One possible explanation for Bitcoin's price movements in light of the knowledge of its unresolved fundamental value, lack of regulation, and hence, a low number of institutional investors, along with, as stated by Glaser et al. (2014), a low level of professionalism among Bitcoin investors might lie in market sentiment towards this asset (Dwyer (2015); Weber (2014); Cheah and Fry (2015)).

1.4.4. Bitcoin & Sentiment

As explored in section 1.4.3, traditional asset pricing models and conventional risk factors fall short in accurately forecasting Bitcoin returns due to the absence of fundamental infor-

mation like dividends or earnings (Liu and Tsyvinski (2021)). Yet, the challenges in valuing Bitcoin extend beyond this fundamental gap.

Firstly, Bitcoin's investor base is predominantly comprised of uninformed retail investors or passionate enthusiasts, often lacking professional financial acumen and susceptible to irrational decision-making (Yelowitz and Wilson (2015); Almeida and Gonçalves (2023)).

Secondly, the profound influence of social dynamics and public sentiment on investor behavior adds another layer of complexity to Bitcoin's valuation. Opinions, trends, and emotions wield power over its price trajectory (Mai et al. (2018); Goczek and Skliarov (2019); Bianchi (2020); Naeem et al. (2021); Almeida and Gonçalves (2023)), demanding a contemporary understanding of psychological frameworks to decode the intricacies of investment decisions (Shrotryia and Kalra (2022)).

Thirdly, CC investors often engage in behavioral trading strategies, capitalizing on fleeting trends and high-sentiment, high-volume trades at short intervals. This behavior contributes to the noise prevalent in CC markets, challenging traditional evaluation methods (Mishev et al. (2020); Karaa et al. (2021)).

To confront these obstacles, harnessing social media sentiment emerges as a promising avenue (Naeem et al. (2021)). Social media sentiment analysis offers a means to navigate the dominance of retail investors, gauge emotions from online discourse, and conduct intraday analysis amidst the deluge of data, thereby accommodating the volatile trading patterns and short-term impacts.

1.5. Research Questions

Against the backdrop of the current state of literature, which has been outlined and summarized in the previous chapters, this dissertation aims to contribute to the knowledge base surrounding the topic of social media sentiment analysis and the field of CCs, with a particular focus on the effect channel of sentiment, an expansion of existing measurement methods to include additional dimensions, and their interaction with the CC Bitcoin.

1.5.1. Causality and Effect Channel of Social Media Sentiment

In this regard, Section 2 specifically addresses the question of how the precise effect channel of social media sentiment influences investor decisions. Considerations regarding the impact of sentiment on investors and ultimately asset prices have already been made in the relevant literature, assuming imperfect financial markets and limitations of arbitrage. Text analysis using modern methods, such as BERT, and leveraging the substantial data generated by large social media platforms, has become a widespread and popular analytical approach, aiming to approximate (irrational) retail investor behavior.

Given the constraints of investors' attention spans, their investment decisions often lean towards assets that capture their attention, whether consciously or subconsciously. This tendency, influenced by framing techniques and amplified by social media platforms, underscores the sway these platforms can hold over individual investment choices (Barber and Odean (2008); Liu (2020)). Sentiment, as noted by Johnson and Tversky (1983), possesses a remarkable ability to shape investors' risk perceptions and behavior, with personal factors like happiness also exerting an influence (Kaplanski et al. (2015)). While the debate no longer questions whether sentiment impacts market participants, the focus has shifted towards understanding the depth of its impact and devising effective measurement strategies (Baker et al. (2007)).

Despite empirical findings overwhelmingly suggesting a relationship between investor sentiment and asset prices, discussions regarding causality, particularly the direction and channels, have surfaced prominently, igniting scholarly interest. This area has been explored experimentally across various papers in economic literature, with each study adding a significant layer of understanding. Hales et al. (2011) make a substantial contribution by advancing linguistic analysis in financial accounting research. Their work showcases that investors are markedly more susceptible to the influence of vivid language compared to dull language of the same sentiment in financial reporting—a phenomenon notably accentuated when the underlying information is preference inconsistent. Echoing these findings, Tan et al. (2014) and Rennekamp and Witz (2021) emphasize the substantial impact text can have on investors' judgments, particularly when readability is low or the language used is informal. Furthermore, Miller (2010) underscores how lengthy and less readable filings directly lead to reduced trading, prompting small investors to halt trading activities—an insight with profound implications for market dynamics.

The chosen information channel also emerges as pivotal; Kelton and Pennington (2020) highlight investors' stronger identification with a CEO when communication occurs through Twitter compared to the company's website, signaling the critical role of platform selection in shaping investor perceptions. A recent and pivotal study by Boulu-Reshef et al. (2023) specifically delves into the influence of emojis in social media posts (tweets) on financial professionals, revealing a significant albeit marginal impact of these messages on investment decisions, underscoring the evolving landscape of communication in financial markets.

Despite these compelling experimental findings, the intricate mechanisms underlying these effects remain insufficiently understood, presenting a tantalizing avenue for further research. A robust examination of the influential channels is imperative to substantially advance comprehension of individuals' investment behavior, paving the way for more informed decision-making in financial markets. Thus, section 2 aims to add to the existing literature

by thoroughly investigating individuals' investment choices and their perceptions of financial and social media sentiment within an experimental framework that encompasses a diverse array of financial and social media information sources, thereby enriching the understanding of sentiment and its influence on inverstors.

1.5.2. Multidimensional Sentiment Analysis

As described in section 1.3.2 sentiment analysis for economic texts has made significant progress in recent years, introducing new methods using large amount of data.

In the realm of linguistic text analysis, understanding the connotations of words is pivotal for deciphering their intended meaning and sentiment. Traditionally, this has been accomplished through the use of linguistic dictionaries, which map words to predefined connotations based on psychological analysis. These dictionaries typically categorize words into 'positive' and 'negative' sentiments, providing a binary framework for sentiment analysis. This approach has been widely utilized in various fields, including economics, finance, and social media analysis (see, i.a., Antweiler and Frank (2004), Baker et al. (2007), Gao and Yang (2017), Kim and Kim (2014) and Sun et al. (2016)).

However, when applying linguistic text analysis in an economic context, one encounters the challenge of whether the connotations assigned to words in psychological dictionaries align with their meanings in economic discourse. For instance, a word like 'risk' may carry predominantly negative connotations in psychological contexts, but its interpretation in economic settings may differ. This incongruity necessitates the development of sentiment dictionaries specifically tailored for economic analysis. These dictionaries, pioneered by researchers like Henry (2008) and Loughran and McDonald (2011), are intentionally designed to capture the nuances of language in economic contexts, providing a more accurate framework for sentiment analysis in financial markets and related domains.

One notable example of such a sentiment analysis tool is the *Valence Aware Dictionary and sEntiment Reasoner* (VADER), introduced by Hutto and Gilbert (2014). VADER goes beyond traditional linguistic dictionaries by considering additional linguistic components such as punctuation, slang, irony, and emoticons, which are prevalent in social media data. By incorporating these elements, VADER offers a more nuanced understanding of sentiment in textual data, enabling the estimation of the degree of positive or negative sentiment expressed in microblogging texts with greater accuracy.

Despite the advancements made with dictionary-based approaches like VADER, newer techniques have emerged in recent years, particularly in the field of Natural Language Processing. Transformative models like BERT (Devlin et al. (2018)) , XLNet (Yang et al. (2019)), and XLM (Lample and Conneau (2019)) have revolutionized sentiment analysis by leveraging deep learning techniques to achieve higher accuracies in text classification tasks. These

NLP Transformers excel at capturing complex linguistic patterns and context dependencies, making them highly effective in analyzing sentiment in diverse textual datasets.

Section 3 embarks on a critical examination of sentiment analysis methodologies from the beginning, questioning whether a multidimensional approach might yield superior results compared to traditional two-dimensional frameworks. The utilization of emotion-based analysis is proposed, which moves beyond simplistic positive-negative classifications to consider a spectrum of emotions associated with each word. By employing the NRC-Emotion Association Lexicon (EmoLex) developed by Mohammad and Turney (2013), words can be categorize based on up to eight distinct emotions, offering a more nuanced understanding of sentiment in textual data.

EmoLex, unlike traditional economic sentiment dictionaries, lacks an explicit economic context. This divergence prompts to establish a suitable benchmark for comparison. Therefore, a positive-negative dictionary without an economic background is used, serving as a reference point to evaluate the efficacy of our multidimensional approach. This comparative analysis allows for assessing the performance of emotion-based sentiment analysis against traditional methods and determining its viability in enhancing the accuracy and efficiency of sentiment analysis tasks. This systematic exploration aims to contribute to the ongoing discourse on sentiment analysis methodologies and advance the understanding of linguistic sentiment representation in economic contexts and the influence of emotion on investor decision making.

1.5.3. Returns Distributions of Cryptocurrencies

While the technical intricacies of cryptocurrencies are well documented, their behavior remains enigmatic and requires deeper analysis. Much of the prevailing economic research has concentrated on prominent CCs like Bitcoin, Ethereum, and Ripple, given their dominance in terms of total market capitalization (Glas (2019)). Studies by Baur et al. (2018a) and Glas (2019) assert that Bitcoin and other CCs exhibit an uncorrelated nature with traditional assets during financial distress, contrasting with fiat currencies as demonstrated by Gkillas et al. (2018). Additionally, volatility (Polasik et al. (2015); Balcilar et al. (2017)), diversification concerns (Brière et al. (2015); Selgin (2015); Corbet et al. (2018b); Schmitz and Hoffmann (2021)), and safe haven attributes (Bouri et al. (2017b); Urquhart (2018)) have been explored in various studies.

To fully grasp the market risk posed by CCs, a comprehensive analysis of their return distributions is imperative. Numerous studies have observed non-Gaussian behavior and heavy-tailed distributions in CC returns (Osterrieder et al. (2017); Gkillas et al. (2018); Gkillas and Katsiampa (2018)), necessitating the adoption of distribution models tailored to these characteristics. Recent endeavors by Majoros and Zempléni (2018) and Kakinaka and

Umeno (2020) have utilized stable distributions (SDIs) to tackle these challenges. However, Kakinaka and Umeno (2020) found SDIs inadequate in capturing heavy tails compared to alternative distributions, prompting the utilization of the generalized Pareto distribution (GPD) in subsequent studies (Gkillas et al. (2018); Gkillas and Katsiampa (2018)).

Building upon this foundation, section 4 proposes a novel methodology that integrates both the SDI and the GPD to more accurately model CC return distributions. By identifying the onset of the tail in return distributions, the data is partitioned and distinct distributions to the tail and body segments are applied, respectively. The GPD is employed to estimate extreme losses in the tail, while the SDI is applied to the remaining data body owing to its superior fit.

Section 4 aims to significantly contribute to the existing literature by offering an advanced modeling approach for CC return distributions. Furthermore, while prior studies have predominantly focused on individual CCs like Bitcoin, Ethereum, and Ripple, this analysis endeavors to provide a more comprehensive understanding of the entire CC market. By bridging these gaps, section 4 extend the breadth of analysis and offer actionable insights for enhanced risk assessment, potentially benefiting both regulators and investors alike.

1.5.4. Bitcoin Returns and Intraday Twitter Sentiment

Executing effective sentiment analysis demands a sophisticated approach informed by prior research. Language nuances, varying across financial and non-economic contexts, demand attention (Henry (2008); Loughran and McDonald (2011); Renault (2017)). Moreover, the complexity of language structures, including syntax and tone, necessitates analytical methods capable of parsing these intricacies within the contextual framework (Devlin et al. (2018); Peters et al. (2018); Mishev et al. (2020)), like the ones discussed in section 1.3.2. Finally, a shift towards emotion-based sentiment analysis offers a deeper understanding of decision-driving sentiments, mitigating data loss and enhancing analytical precision (see section 3).

In addressing these challenges head-on, section 5 aims to leverage social media sentiment as a tool in deciphering market dynamics and refining risk assessment methodologies for Bitcoin valuation.

Section 5 contributes to the scientific discourse in two significant ways. Firstly, a state-of-the-art NLP Transformer model called EmTract, specifically trained for the financial context and social media based on DistilBERT, is used. This allows the model to account for language usage as well as more complex linguistic features. Secondly, the relevant literature indicates a short-term influence of (social) media sentiment on Bitcoin returns. While individual (intraday) intervals have been used for predictions (Behrendt and Schmidt (2018); Broadstock and Zhang (2019); Guégan and Renault (2021)), to the best of my knowledge a consistent analysis of all possible intervals on a minute-by-minute basis within a day has not yet been

conducted. This analysis is carried out in Section 5, aiming to provide deeper insights into the 'short-term' effects of sentiment on Bitcoin returns and to identify whether there are specific time horizons for which sentiment analysis appears particularly promising.

2. The relevance and influence of social media posts on investment decisions - An experimental approach based on tweets

2.1. Abstract

We conducted an experiment to examine the role of positive and negative tweets (generated by AI) on investment behavior, comparing them with provided historical and fundamental financials. Through mediator analysis, we discovered that positive tweets have a significantly positive mediating effect on investment amounts, while negative tweets have a negative impact. Importantly, we found that this effect is not primarily driven by the perception of the tweets; rather, positive tweets influence individuals' perception of a company's financials which is the most influencing factor in individuals' investment decision. In this manner our study contributes to the existing literature by (1) proving evidence for a causal effect of social media investor sentiment on investment behavior on capital markets and especially (2) focussing how the influence channels are built.

2.1.1. Introduction

Predominantly starting with Kyle (1985) and Black (1986) the influence of noise in financial markets has aroused the interest of many researchers in the field of Behavioral Finance. In financial research the role of noise traders has been widely discussed as noise trading is supposed to explain why stock prices could differ from their fundamental value. This idea contradicts the idea of information-efficient markets stated in the EMH by Fama (1970). Fama (1965) himself argues that irrational noise traders would meet rational traders on financial markets who trade against them. This should result in systematic losses for noise traders who will leave the market because of the behavior of rational arbitrageurs. De Long et al. (1990) oppose that there are limits to arbitrage due to risk aversion and short time horizons allowing noise traders to temporarily diverge prices from the fundamental value. Consequently, the development, identification (and prediction) of noise has become a main interest of research in financial research.

Market or investor sentiment defined as market's general, psychological environment is believed to wield considerable influence over noise trading, thereby anticipated to impact stock prices. Given the non-trivial nature of observing investor sentiment, the debate on its influence within financial markets pivots on identifying the most appropriate measure. Over time, three main distinct measurement approaches have emerged: market-based, survey-based, and text-based methodologies.[7]

[7] A comprehensive overview about the three measurements is given for example in Aggarwal (2022).

The approach last mentioned, which has gained and continues to enjoy widespread popularity, aligns with the ascent of social media platforms like Twitter, Facebook, and Instagram. Their expanding user bases, coupled with increasingly accessible textual analysis tools such as BERT with nearly 9,000 trained models on *Huggingface.co*, have propelled this approach. Consequently, researchers have probed the potential impact of a platform's content on stock market performance. Given investors' limited attention spans, their investment decisions often exhibit biases toward assets that consciously or subconsciously grab their attention — such as through framing techniques (Barber and Odean (2008)). As a result, social media platforms may indeed sway individual investment choices (Liu (2020)). Johnson and Tversky (1983) previously noted that sentiment has the power to influence investors' risk perceptions. Kaplanski et al. (2015) corroborate this observation, even detecting the effects of investors' personal happiness on their investment behavior. Additionally, Baker et al. (2007) conclude that the debate no longer revolves around whether sentiment influences market participants but rather focuses on the intensity of its impact and how best to measure it.

Despite empirical findings predominantly suggesting relationships, discussions surrounding causality, particularly the causal direction and channels, have surfaced. This area has been experimentally explored across various papers in economic literature. Hales et al. (2011) contribute to linguistic analysis in financial accounting research (e.g. Tetlock (2007), Tetlock et al. (2008), Feldman et al. (2010)) by demonstrating that investors are more susceptible to the influence of vivid language compared to dull language of the same sentiment in financial reporting. This effect is especially pronounced when the underlying information is preference inconsistent. Studies by Tan et al. (2014) and Rennekamp and Witz (2021) echo these findings, suggesting that text can significantly impact investors' judgments, particularly when the readability of the text is low or when the language used is informal. Moreover, Miller (2010) finds that lengthy and less readable filings lead to reduced trading, prompting small investors to halt trading activities. The chosen information channel also plays a role. Kelton and Pennington (2020) note that investors tend to identify more with a CEO when communication occurs through Twitter compared to the company's website. A recent and comparable study by Boulu-Reshef et al. (2023) specifically examines the influence of emojis in social media posts (tweets) on financial professionals. Their research indicates a significant yet marginal impact of these messages on investment decisions.

Despite the specific experimental findings, there remains a limited understanding of the intricate mechanisms underlying these effects. A deeper examination of the influential channels could significantly enhance our comprehension of individuals' investment behavior. Thus, we aim to contribute to the aforementioned literature by investigating individuals' investment choices and their perceptions of financial and social media sentiment within an experimental

setting encompassing various financial and social media information sources.

Through the application of mediation analysis, our study seeks to scrutinize whether and through which channels these distinct information sources exert an influence on perceived sentiment. Subsequently, we aim to explore how these perceptions, in turn, impact investment decisions. We go in line with prior findings, but also find using mediator analysis that the tweets do not have significant influence on investment decision directly as well as over the mediator perceived tweet Sentiment. Moreover, the tweets influence the perceived Financial Sentiment which has a large and significant influence on the investment decision.

The remainder of this paper is structured as follows: Section 2.2 provides a detailed description of the methods utilized to gather financial and social media data within the experimental framework, aiming for authenticity. It further delves into the implementation process, concluding with the formulation of hypotheses based on the established setting. Section 2.4 offers a concise overview of the collected data, leading into the presentation of our findings. This includes a mediation analysis elucidating the impact on investment decisions. Finally, Section 2.5 serves as the conclusion, where we summarize our observations in light of previous literature, and highlight potential avenues for future research.

2.2. Experimental Design

Our experimental design aims to assess the impact of social media posts, specifically tweets on the platform 'X' (formerly 'Twitter'), on the investment behavior of individuals. Taking into consideration aspects of loss aversion following prospect theory by Kahneman and Tversky (1979), we are also interested in observing this behavior with positive and negative versions of provided financials and tweets. To achieve this, we divided our test subjects into six different groups, as outlined in Tab. 1.

Group	**Tag**	**Financials**	**Twitter**
1	PP	Positive	Positive
2	PN	Positive	Negative
3	NP	Negative	Positive
4	NN	Negative	Negative
5	P	Positive	*none*
6	N	Negative	*none*

Tab. 1 Grouping

In the following subsections, we describe the specified investment setting along with the design of positive and negative financials and tweets. We conclude our introduction to the experimental design by detailing the incentive system. Subsequently, we derive our hypotheses based on our key findings in the introduction and our experimental design.

2.2.1. *Investment Setting*

Test participants were instructed to gather information about the fictional company 'Glubon AG'[8] of which they already owned 100 stocks, each valued at 10€ (resulting in a total stock capital of 1000€). Based on a brief company description (refer to Fig. 8 in 2.6.1), stock charts, financial metrics (see Section 2.2.2) and (for groups 1 to 4) posts on the platform Twitter[9] ('Tweets', see Section 2.2.3), participants had to decide whether to sell or buy stocks at a rate of 10€ each. Each participant also possessed 1000€ of free capital, and the decision was limited to holding between zero stocks and 2000€ of free capital or holding 200 stocks and 0€ of free capital at the end of the experiment. After all participants made their decisions, a new stock price per group would be calculated, as explained in Section 2.2.4. This calculation also affected the total capital (and consequently, the number of lottery tickets) of the participants. Therefore, the experimental setting is limited to one period and each participant makes only one decision.

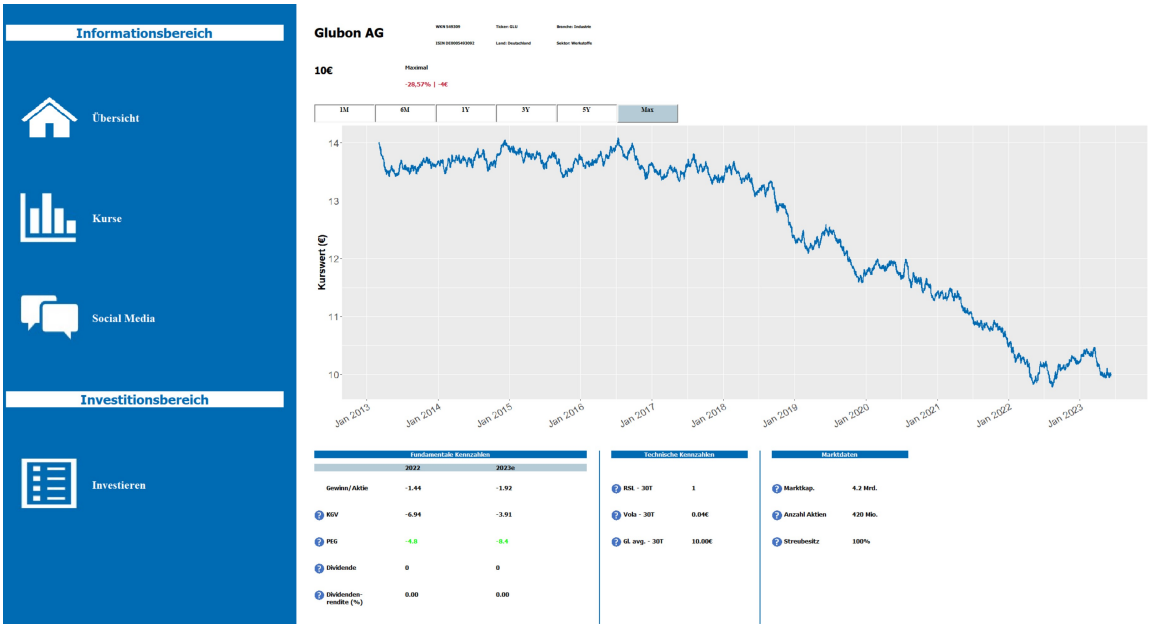


Fig. 2 Platforms Interface (*Financials* tab opened, negative version)

All information was presented on a self-designed, Visual Basic-based information and trading platform, exemplified by the opened (negative) *Financials* tab in Fig. 2. On this platform, our participants could freely navigate between three tabs: *company description*, *Financials*, and *social media*, to gather information for the final decision in the *investment*

[8] AG is the German abbreviation for 'Aktiengesellschaft', which translates to 'stock company'.

[9] Before the conclusion of our experiment, 'Twitter' had unexpectedly been rebranded to 'X'. We chose to keep using the name Twitter, as most participants might not be familiar with the new branding and the name 'Twitter' has been used to provide information to the participants.

decision tab. Thanks to the autonomous coding of the platform, we were also able to track all transitions between tabs and monitor the time spent within each tab.

2.2.2. Financials

The structure of the financials tab is modeled after financial websites such as Yahoo! Finance, presenting charts for different time horizons along with financial figures. The positive and negative cases can be found in Fig. 9 and 10 in the appendix.

The stock price development was simulated using a random walk with drift, as described in formulas 1 and 2. To enhance the authenticity of the development, a new drift α was drawn from a normal distribution with a positive mean for the positive case every 30 days, as detailed in formula 3.

$$P_t = P_{t-1} + \alpha_i + \epsilon_t \tag{1}$$

with

$$\epsilon_t \sim N(0, 1) \tag{2}$$

and an every 30 days t changing α_i

$$\alpha_i \sim N(1, 25) \tag{3}$$

For the negative case, daily returns were reversed, and both stock price developments were scaled to a price of 10€ on the last day.

Additionally, participants could find financial figures below the charts, designed to appeal to economically educated participants who assumed the market, following Fama (1970), to be semi information-efficient. Even less economically educated participants could benefit from this information, as each figure was explained by clicking the '?' buttons next to the figure. The provided positive (negative) financial figures included positive (negative) profits per share, positive (no) dividends/dividend returns, positive (negative) price-earning ratios for the previous year as well as expected for the current year. Furthermore, figures for low (high) volatilities, relative strength, 30 days moving average, as well as information about the market capitalization, free float, and number of shares, were presented.

Consequently, we are aware of possible biases in the perception of the financials of Glubon as 'positive' and 'negative', especially for the charts, due to prior findings in behavioral finance (in this case, especially the disposition effect empirically introduced by Shefrin and Statman (1985)). Therefore, we ask the participants about their perception as well as their judgment

regarding plausibility and trustworthiness of the given financials after the investment decision.

2.2.3. Tweets

tweets were presented as the result of a search for the cashtag '$GLU' of the imaginary Glubon AG on the platform Twitter. The content of the tweets was generated using OpenAI's ChatGPT queries mentioned in 2.6.1. Due to different queries, positive, negative, and neutral tweets were created by the AI using varying maximum lengths (20, 70, or 140 characters) as well as in colloquial and non-colloquial language. From the created database of 180 Tweets, we sampled 40 Tweets each for groups 1 & 3 and groups 2 & 4, as stated in Tab. 2.

Query specification			**Occurences per group**		
sentiment	colloquial	max character	1 & 3	2 & 4	5 & 6
Positive		20	5		0
Positive		70	5		0
Positive		140	5	randomly	0
Positive	X	20	5	picked 3	0
Positive	X	70	5		0
Positive	X	140	5		0
Neutral		20			0
Neutral		70			0
Neutral		140	randomly	randomly	0
Neutral	X	20	picked 7	picked 7	0
Neutral	X	70			0
Neutral	X	140			0
Negative		20		5	0
Negative		70		5	0
Negative		140	randomly	5	0
Negative	X	20	picked 3	5	0
Negative	X	70		5	0
Negative	X	140		5	0
		$\sum$	40	40	0

Tab. 2 Queries and Presence of Tweet Type per Group

The tweets provided on the platform for group 1 & 3 not only contain positive tweets but also a minor number of neutral and negative tweets for authenticity reasons. The same holds true vice versa for the tweets provided to group 2 & 4. To ensure that this does not affect the treatment, participants were asked for their perception of the tweets after the investment decision. To enhance authenticity further, we added ChatGPT-generated German usernames as well as randomly picked profile pictures from the academic dataset delivered by the company 'Generated photos'. The picture dataset, including estimators for gender, race, and the emotion shown in the picture, allowed us to pick a diverse spectrum of mostly happy profile pictures. While we randomly ordered the sampled tweets per group, the order of profile names and pictures is the same in every group. Ultimately, replies, retweets, likes and impressions were drawn from a normal distribution with a higher mean if the tweet sentiment

fits the group's social media treatment than for tweets of another sentiment as those factors can also influence investors' perception following Cade (2018) or Rennekamp and Witz (2021). All these operations lead to a social media tab as exemplified in Fig. 3.[10]

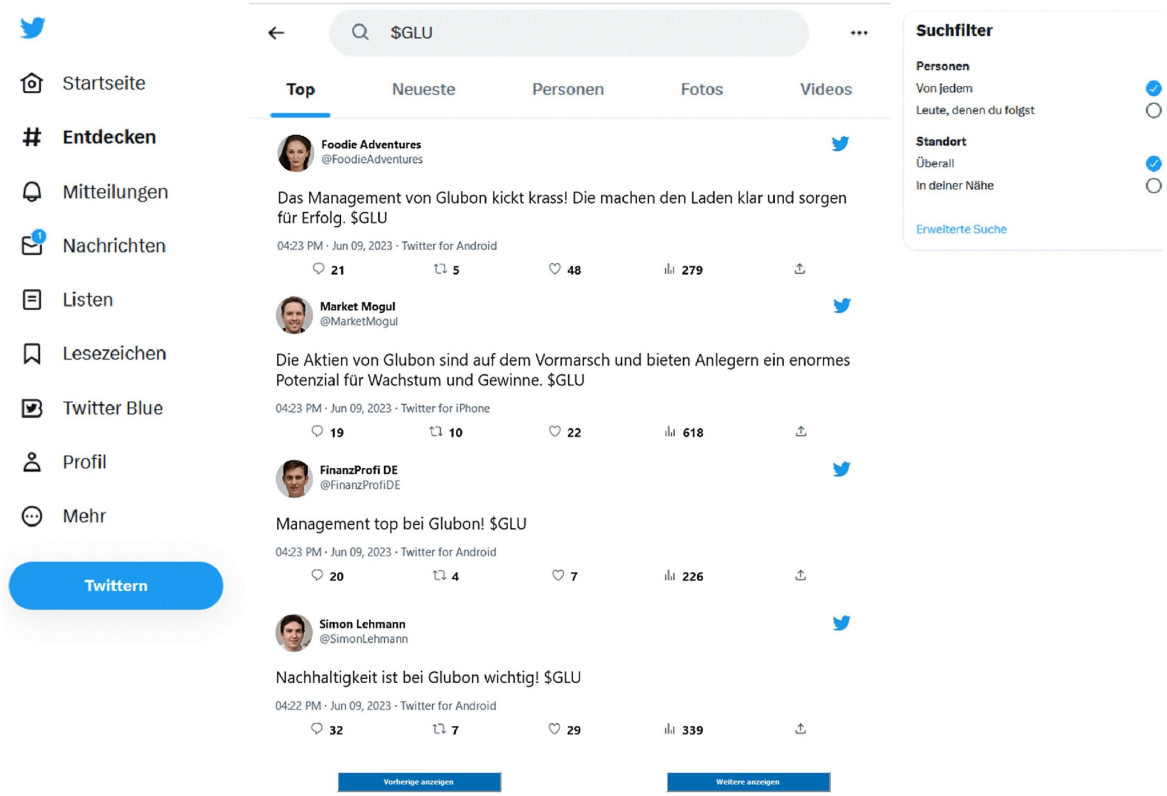


Fig. 3 Social Media Tab, site 1 of 10 opened, positive version

Consequently, this operationalization does not mimic a potential 'timeline' of the users and can be more accurately compared to a search for the company's cashtag ($) in the Twitter feed. We assume that potential effects reported in section 2.4 would be more pronounced if tweets had been posted by users our test participants would have decided to follow in real life, which would not have been possible to mimic reliably in an experiment. Additionally, the AI-generated content could possibly be recognized by the users. Therefore, we asked the participants for their assessment of the trustworthiness of the tweets.

2.2.4. Implementation

The experiment took place in a lab at the Heinrich-Heine-University Duesseldorf in July and August 2023 with an open registration for everyone speaking German fluently. Over time we collected data from 300 participants mainly containing economic students but also

[10] A translated example for a tweet of every query type mentioned in Tab. 2 can be found in Tab. 9 of 2.6.1.

professionals and students from other disciplines. From the 300 participants we use 259 responses for our dataset excluding 41 participants who failed at answering at least 3 of 4 control questions regarding the given setting and incentive system correctly.

In addition to fixed compensation, participants were incentivized by a lottery which ensures conscientious behavior by the participants (Holt and Laury (2002)). Each participant started the experiment with a total capital of 2000€ (1000€ stock capital, 1000€ free capital), which translated into 2000 tickets for the lottery (1€ equals 1 ticket). Depending on the decisions made within each reference group, a new stock price was calculated, affecting the stock capital and total capital of each participant based on their decision. Fig. 4 illustrates how the decision to buy or sell 50 stocks affects the total capital, and consequently, the number of lottery tickets, if the stock price increases to 15€ (blue situation) or decreases to 5€ (green situation).

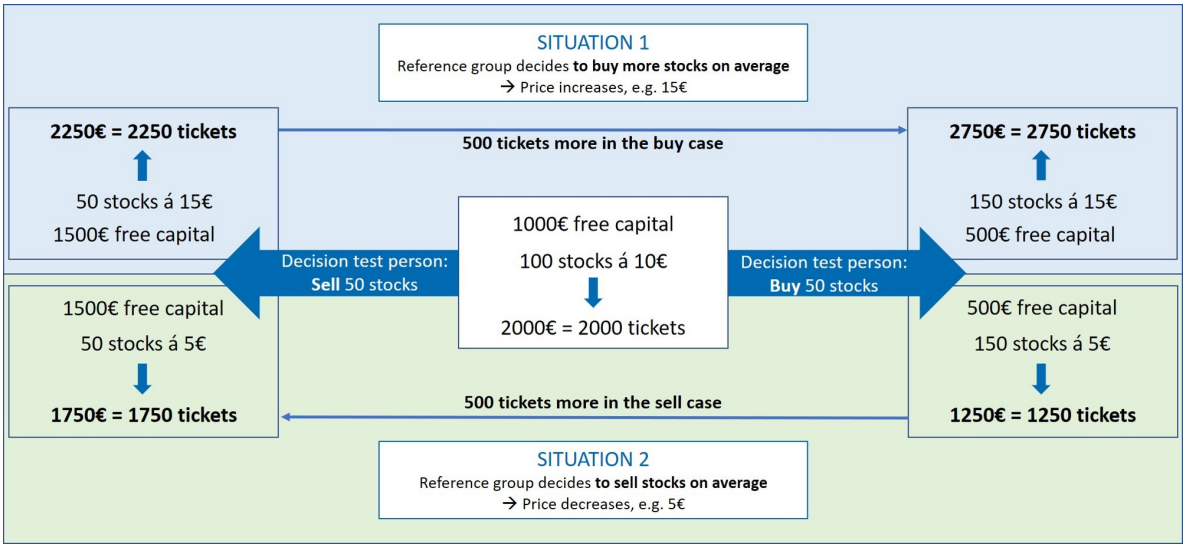


Fig. 4 Ticket Outcomes under Different Situations and Decisions

For the calculation of the new stock price, P_1, in each group i with N_i participants, we use a simplified stock pricing formula that interprets the return of the stock, r_i, as the ratio between the change in cumulated stock capital in t_1, $SC_{i,1}$, and the cumulated stock capital in t_0, $SC_{i,0}$:

$$r_i = \frac{SC_{i,1} - SC_{i,0}}{SC_{i,0}} \tag{4}$$

Consequently, the new price per group i ($P_{i,1}$) is calculated as

$$P_{i,1} = P_0 * (1 + r_i) \tag{5}$$

which is limited between

$$\lim_{SC_{i,1} \to 0} P_1 = 0 \tag{6}$$

and

$$\lim_{SC_{i,1} \to 2000N_i} P_1 = 20. \tag{7}$$

Further, we collected variables for controlling purposes regarding participants' demographics (as gender, age, income & risk tolerance following Holt and Laury (2002)), financial experience and social media usage.

2.3. Hypotheses

In the context of the EMH (Fama (1970)), it can be assumed that economic agents process information provided to them appropriately, thereby adjusting their actions to the existing information environment. As indicated by the relevant literature and various economic studies, both social media (see i.a. Antweiler and Frank (2004); Baker and Wurgler (2006); Da et al. (2015); Das and Chen (2007); Renault (2017); Sun et al. (2016); Tetlock (2007)) and financial indicators influence the investment calculus of individuals. However, Tversky and Kahneman (1974), in their highly regarded study considered the starting point of Behavioral Finance, demonstrated that due to behavioral biases, the available information is inadequately processed using experience and heuristics (Ritter (2003)). In this context, differences may arise in the consideration of various information sources and their interpretation leading to departures from rational decision-making calculations, as exemplified by phenomena such as noise trading. Thus, it can be assumed that different economic agents may consider various information sources differently based on their experiences and perceptions. In our specific case, economic agents have access to social media posts in the form of tweets and financials (historical and fundamental) for their investment decisions. The goal of this study is to examine whether the provided information has an impact on individuals' investment decisions. However, in the context of the presented behavioral biases, it is also necessary to investigate how the tweets and financial information were perceived by each participant (sentiment) and whether this sentiment also influences the investment decision. To address this question, a mediation analysis will be employed, aiming to answer the following main hypotheses:

H 1: *There is a mediating effect of Financial Sentiment on the investment decisions of individuals.*

H 2: *There is a mediating effect of Tweet Sentiment on the investment decisions of individuals.*

In our analysis, we draw insights from Baron and Kenny (1986) and Zhao et al. (2010) to elucidate the intricate mechanism by which provided information and the associated sentiment shape investment decisions. Our approach involves examining both the direct impact of tweets and financials on investment decisions and their indirect effects mediated by two factors: *Tweet Sentiment* and *Financial Sentiment*. Furthermore, we also examine the influence of tweets on Financial Sentiment and the influence of financials on Tweet Sentiment to account for a potential deviation from rational decision-making in the context of Behavioral Finance. Hence, the following sub-hypotheses arise:

H 1.1: *There is an indirect effect of tweets via the mediator Tweet Sentiment on the investment decisions of individuals.*

H 1.2: *There is an indirect effect of tweets via the mediator Financial Sentiment on the investment decisions of individuals.*

H 1.3: *There is a direct effect of tweets on the investment decisions of individuals.*

H 2.1: *There is an indirect effect of financials via the mediator Financial Sentiment on the investment decisions of individuals.*

H 2.2: *There is an indirect effect of financials via the mediator Tweet Sentiment on the investment decisions of individuals.*

H 2.3: *There is a direct effect of financials on the investment decisions of individuals.*

2.4. Results

2.4.1. Participants' Information

Before proceeding with the analysis of the data from the conducted experiment in the next section, we will first delve into the collected information of the participants. To do this, the data is divided into three categories, with the last category further subdivided into three more categories. All information discussed below can be found in Tab. 3.

The 'Participants' behavior' category encompasses the 'Stocks held' by participants at the end of the experiment, thus reflecting their investment decision. By definition, the values in this category can only be integers in the interval [0, 200], where 0 represents the sale of all initially (100) held stocks, and 200 represents the maximum purchase of 100 additional stocks within the available budget. This interval was utilized, as evident from the maximum and minimum values, with participants acquiring, on median, an additional 10 stocks, while, on average, only 1.6 additional stocks were acquired by a standard deviation of 61.73 stocks.

The second category, 'participants' sentiment', includes the sentiment of the participants regarding the given tweets and financials. After making their investment decisions, participants were tasked with using a Likert scale ranging from 1 to 5 to assess how they perceived the given tweets and financials.

	Min	Max	Med	Mean	Sd	Dummy
Participants' behavior						
Stocks held	0	200	110	101.60	61.73	
Participants' sentiment						
Tweets	1	5	2	2.63	1.59	
Financial	1	5	3	3.01	1.43	
Participants' characteristics						
Demographic						
Age	17	62	23	24.93	7.25	
Male	0	1	1	0.60	0.49	X
Risk	0	10	5	4.86	1.79	
Income	0	10	1	1.77	2.00	
Financial						
Economic	0	1	1	0.61	0.49	X
Cap market	0	1	1	0.60	0.49	X
Social Media						
Usage	0	14	2	2.48	1.74	
Twitter	0	1	0	0.26	0.44	X

Tab. 3 Participants' Information

In this context, a value of 1 corresponds to a very negative sentiment, 3 to a neutral one, and 5 to a very positive sentiment. These pieces of information serve in the further development of the work both to validate whether the given treatment was perceived by the participants according to its intention and to highlight whether perception, rather than the actual information, has an impact on investment decisions. The entire possible interval of $[1, 5]$ was also utilized by the participants for both Social Media and Financial Sentiment, with the Social Media Sentiment being more negative on both average and median compared to the Financial Sentiment.

The last category, 'Participants' characteristics', includes characteristics of the participants regarding their demographic information, financial experience, and social media usage. The category of 'Demographics' includes the age, gender, risk attitude and income of the participating individuals. The youngest participant was 17 years old, and the oldest person was 62 years old. Based on the median (23) and the average age (24.93), it can be observed that, as expected, it is a relatively young participant group since this study was conducted at an university.

The variable 'Male' is a dummy variable, which takes the value 1 for participants who identify as male. To account for the three different gender specifications of the participants and considering that only one observation is labeled as gender-diverse, a dummy variable is used. As indicated by the median and the mean, there is a slight majority of male participants

in the present dataset.

The 'Risk' variable measures the risk tolerance of each participant with values ranging from $[0, 10]$, which was determined using the Holt-Laury test (Holt and Laury (2002)).[11] A value of 0 indicates a high risk appetite, while a value of 10 reflects a pronounced risk aversion. In the present dataset, the majority of participants are therefore more risk-averse.

Furthermore, participants were asked about their monthly income, which could be indicated in increments of 500. Thus, the number 0 represents an income of 0-500€, and the number 10 (the maximum in this dataset) represents an income of more than 5000€. Hence, we observe a relatively low income level of 1.77 on average, which again, is to be expected since the experiment was conducted at an university.

Aside from demographic information, additional data was collected on participants' financial background and social media usage to consider their effects in the further analysis. In terms of economic characteristics, there is a dummy variable indicating whether a participant has an economic-related background in form of an university degree or an apprenticeship. The variable 'Cap market' indicates whether a participant has been active in a capital market. In terms of social media characteristics, the dummy variable 'Twitter' differentiates whether a participant uses or has used the social media platform Twitter, as this study focuses primarily on this platform for social media posts. Additionally, the variable 'Usage' indicates how many hours per day a participant uses social media channels.

Overall, the majority of participants have been active in the capital market and are currently or have previously pursued a study with an economic background. However, most participants do not use the social media platform Twitter. Furthermore, participants spend an average of 2.48 hours (2 hours in median) per day on social media channels. However, it is important to note that one participant with a daily usage of 14 hours is a clear outlier, which needs to be critically considered in the subsequent ANOVA analysis.

The collection of the data described above allows, on one hand, drawing conclusions about the characteristics of the participating individuals to assess the generalizability of the results of the present study. On the other hand, these variables serve as control variables in a later section to check the robustness of the results.

After examining participants' behavior, sentiment, and characteristics, the next step is to take a closer look at these factors for each group. Since this study aims to contribute to

[11] Holt and Laury measure individuals' risk aversion by presenting two lotteries. Participants are asked to choose between a less risky and a riskier but potentially more profitable lottery in 10 different scenarios, with the probability of the higher payoff increasing in each iteration. The degree of risk aversion is determined by the switching point from the less risky to the riskier lottery, with the rational switch based on expected value occurring after the fourth iteration. Therefore, values above 4 indicate increased risk aversion. For a more detailed overview, see Holt and Laury (2002).

the explanation of individuals' investment behavior, Fig. 5 and Fig. 6 are used to provide an overview of the differences in investment behavior between the individual groups.[12].

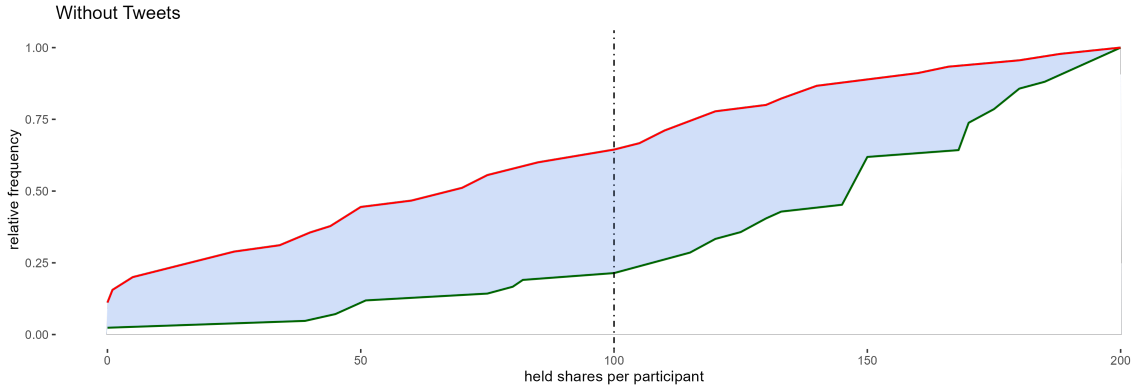


Fig. 5 Cumulative Relative Frequency of Stocks Held (without Tweets)

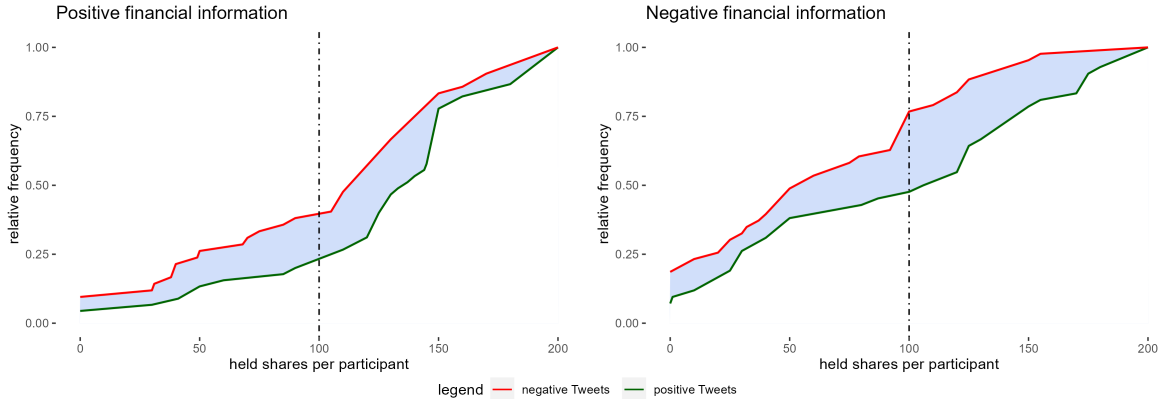


Fig. 6 Cumulative Relative Frequency of Stocks Held (with Tweets)

Firstly, the cumulative relative frequency of Stocks held for the groups without tweets is examined (Fig. 5). The two groups only differ in the provided financials. It can be seen that the group with positive financials (P), represented in green, holds more stocks throughout the entire distribution compared to the comparison group with negative financials (N). Looking at the density distribution of the other groups (Fig. 6), which were provided with tweets, a similar pattern emerges. The compared groups always differ in the provided tweets, while the financials do not differ in the individual comparisons. It becomes clear that both in the case of positive and negative financials, there is a difference in the held stocks. In both cases, participants who were provided with positive tweets (PP, NP) hold more stocks throughout the entire distribution compared to the groups with negatively connotated tweets (PN, NN).

[12]For an overview of the different groups see Tab. 1.

2.4.2. Analysis of Variance & Post-hoc Test

Based on these observations, an Analysis of Variance (ANOVA) is conducted subsequently to examine whether the held stocks differ significantly among the individual groups. In addition to differences in participant behavior, an examination will also be conducted to determine whether there are differences in participants' sentiment and characteristics among the individual groups. The ANOVA results and the means for every aspect analyzed are depicted in Tab. 4.

The F-statistic of the ANOVA clearly indicates that there are significant differences between individual groups regarding the average number of Stocks held at the end of the experiment. On average, groups with positive financials hold more stocks than those with given negative financials. In particular, the control group with positive indicators without social media posts (P) holds the most stocks on average. Furthermore, a difference can be observed between the groups with positive financials and positive or negative social media treatment (PN & PP). Participants in the group with positive social media posts (PP) hold, on average, about 23 more stocks compared to participants with negative posts (PN), which might hint towards an influence of the given social media treatment. A similar pattern emerges when examining the groups with negative financials and different social media treatments (NN & NP). Participants in the group with positive social media posts (NP) hold, on average, about 30 more stocks than the comparison group with negative posts (NN). The NN group also holds the lowest number of stocks on average, even when compared to the control group with negative financials and no social media posts (N).

Moreover, the results of the ANOVA regarding participants' perceptions reveal that the given treatments (social media posts) were perceived by the participants in accordance with their intended sentiment. There are significant differences in the perception of the sentiment of social media posts among the individual groups, as measured on a Likert scale. The groups with positive tweets (PP & NP) perceive these posts significantly more positively on average (deviation of approximately 2.5 units) compared to the groups with negatively formulated tweets. A different perception also exists regarding the financials. The groups with positive financials (PP & PN & P) perceive them on average significantly more positively than the groups with given negative financials (NP & NN & N). These results suggest that the treatments were perceived according to their intended purpose.

Finally, ANOVA was used to compare participant characteristics across the individual groups (for characteristics where such a method is meaningful). The results indicate that there are no significant differences in terms of the participants' characteristics, suggesting a

	PP	PN	NP	NN	P	N	F-Stat
				Participants' behavior			
Stocks held	128.77	105.97	97.35	66.86	136.52	75.15	10.53***
				Participants' sentiment			
Tweets	3.91	1.38	3.80	1.37	NA	NA	127.99***
Financial	4.15	4.04	2.23	1.72	4.23	1.68	93.54***
				Participants' characteristics			
Demographic							
Age	26.68	25.33	23.28	25.79	23.71	24.62	1.37
Male	0.511	0.38	0.42	0.34	0.35	0.4	0.63
Risk	5.17	4.61	4.73	4.88	4.61	5.06	0.75
Income	2.07	1.93	1.64	1.91	1.36	1.69	1.72
Financial							
Economic	0.60	0.62	0.62	0.49	0.79	0.53	0.00
Cap market	0.53	0.57	0.69	0.58	0.67	0.53	0.07
Social Media							
Usage	2.29	2.21	3.28	2.24	2.57	2.32	0.00
Twitter	0.20	0.21	0.33	0.33	0.26	0.20	0.06

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Tab. 4 ANOVA between the Different Groups

balanced distribution of participants.[13]

Although the results of the ANOVA indicate significant differences between the means of the six groups in terms of participant behavior and perception, such an analysis does not provide insight into the specific nature of these differences. Therefore, a post-hoc test, specifically the Tukey post-hoc test, is employed (Tukey (1992)). This test allows for detailed comparisons between each group and the others, enabling a pairwise comparison across all groups. The results of the post-hoc test can be found in Tab. 5.

Group comparison	**Stocks held**		**Tweet Sentiment**		**Financial Sentiment**	
	diff	p adj.	diff	p adj.	diff	p adj.
PN - PP	-22.802	0.420	-2.530	0.000	-0.108	0.992
NP - PP	-31.420	0.105	-0.102	0.991	-1.917	0.000
NN - PP	-61.917	0.000	-2.539	0.000	-2.435	0.000
P - PP	7.746	0.988			0.083	0.998
N - PP	-53.622	0.000			-2.467	0.000
NP - PN	-8.619	0.982	2.429	0.000	-1.810	0.000
NN - PN	-39.116	0.020	-0.009	1.000	-2.327	0.000
P - PN	30.548	0.137			0.190	0.911
N - PN	-30.821	0.118			-2.359	0.000
NN - NP	-30.497	0.134	-2.437	0.000	-0.517	0.063
P - NP	39.167	0.021			2.000	0.000
N - NP	-22.202	0.451			-0.549	0.036
P - NN	69.663	0.000			2.5171	0.000
N - NN	8.295	0.983			-0.032	1.000
N - P	-61.368	0.000			-2.549	0.000

Tab. 5 Post-hoc Test between each Group

The group-wise comparison of participant behavior (Stocks held) reveals that groups with opposing financials significantly differ in their purchasing behavior (NN-PP, N-PP, NN-PN, P-NP, P-NN, N-P), with groups having negative financials, as expected, holding fewer stocks. Furthermore, the results from the preceding ANOVA analysis is confirmed in the sense that the treatments of sentiment and financials were perceived by the participants according to their intended purpose. Thus, the groups with divergent sentiment in social media posts consistently differ statistically highly significantly in their perception of tweets.

The same applies to the treatment of financials. The metrics are perceived as intended by the authors. However, two group comparisons stand out. Although groups NP, NN, N were

[13] As previously noted, there is an outlier with 14 hours of social media usage. The effect of this outlier is evident in the elevated mean of social media usage for the NP group. However, in this context, this outlier should not pose a problem, as even when considering this outlier, there is no significant difference between the individual groups. Moreover, if the outlier were to be excluded, the average of this group should align even more closely with the lower average of the other groups.

each provided with the same financial information, these pieces of information were perceived statistically significantly differently. In the N-NP comparison, this difference is significant at a 5% level, and in the NN-NP group comparison, it is still significant at a 10% level.

Since the respective groups all received the same financial information, they differ only in the sentiment of the provided social media posts. In both group comparisons (NN-NP and N-NP), participants received positively connotated tweets. Thus, it can be presumed that the sentiment, especially if the tweets contain positiv sentiment, of the given tweets has an influence on individuals' perception of financial information, which in turn might influence an individuals investment decision. To test this hypothesis, a statistical analysis using a mediation analysis will be conducted subsequently.

2.4.3. Mediation Analysis

Mediation analysis (Baron and Kenny (1986)) is used to measure the effect of (an) independent variable(s) on a dependent variable. For this purpose, both the direct influence of the independent variable(s) on the dependent variable and the indirect effect of the independent variable through a mediator are estimated.

In the present analysis, due to the identified group differences, there is reason to believe that the provided tweets and financials have a direct impact on the investment decisions of the participants (H1.3 & H2.3). Thus, these variables are chosen as independent variables to assess their direct influence on the investment decision made. Furthermore, the results of the preceding section provide grounds to assume that the actual manifestations of tweets and financials influence how these variations are perceived by the participants, and in turn, this sentiment has an impact on the investment decision (H1.1 & H2.1). First evidence that tweets (financials) can also frame perceived Financial (Tweet) sentiment (H1.2 & H2.2) can be seen in Tab. 5 as the perceived Financial Sentiment was significantly more positive when the tweets were of a positive nature. Hence, through the mediation analysis, the model illustrated in Fig. 7 is estimated.

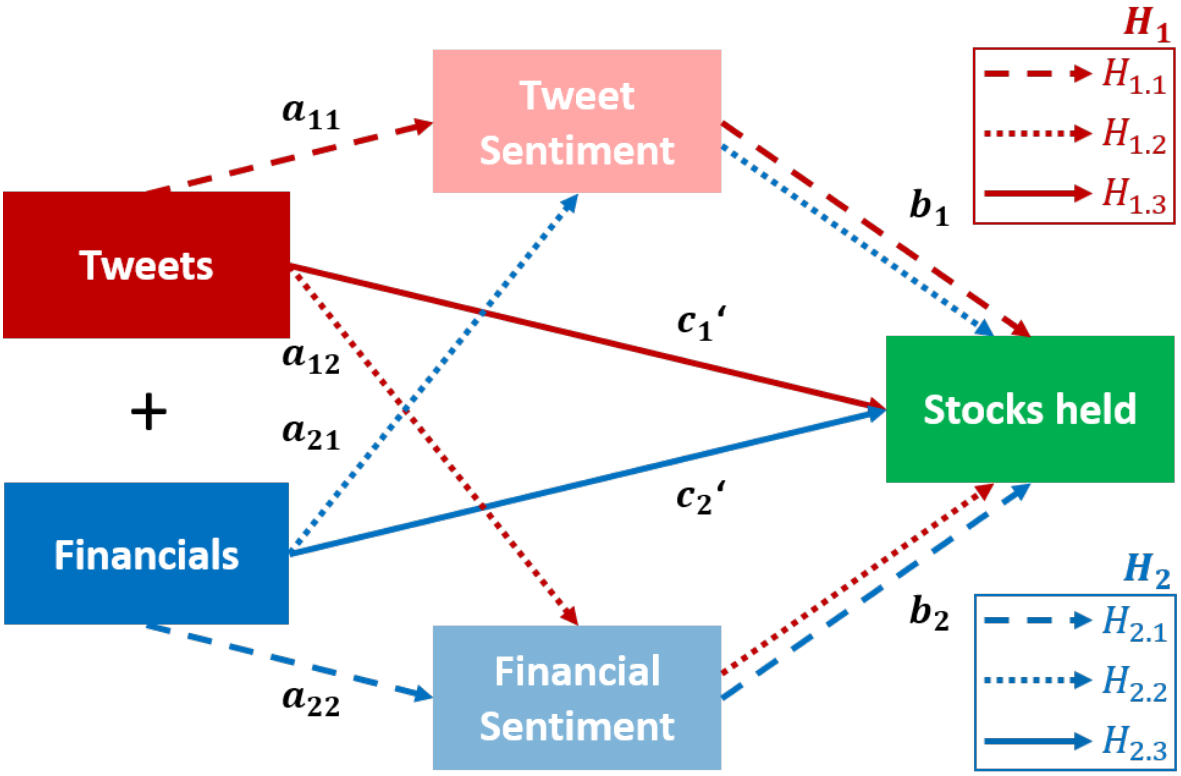


Fig. 7 Mediation Analysis

This model uses the provided tweets and financials as dependent variables and the perception of their sentiment as mediators to explain the Stocks held by the participants and test our hypotheses. In the presented base model (A) of a two-mediator model, a total of 4 different regressions need to be estimated to determine the direct and indirect effects of each regressor and takes the following form:

$$Stocks\_held = i_1 + c_1 * Tweets + c_2 * Financials + \epsilon_1 \tag{8}$$

$$\begin{aligned} Stocks\_held = {} & i_2 + c_1' * Tweets + c_2' * Financials \\ & + b_1 * Tweet\_Sentiment \\ & + b_2 * Financial\_Sentiment + \epsilon_2 \end{aligned} \tag{9}$$

$$Tweet\_Sentiment = i_3 + a_{11} * Tweets + a_{21} * Financials + \epsilon_3 \tag{10}$$

$$Financial\_Sentiment = i_4 + a_{12} * Tweets + a_{22} * Financials + \epsilon_4 \quad (11)$$

To check the robustness of the results of this base model, additional control variables are subsequently added to the estimation. Model (B) includes the demographic information about the participants already presented earlier. In contrast, model (C) has been expanded to include financial and social media characteristics, while model (D) contains both demographic information and financial and social media characteristics.

Please be aware that for assessing the influence of Tweet Sentiment, it is imperative to exclusively consider the groups provided with tweets, given that participants in groups P and N were not exposed to any tweets, thus rendering them incapable of developing any Tweet Sentiment. Consequently, the models are estimated with $N = 172$ observations. The results of these estimation models can be found in Tab. 6.

The results of model (A) show that the given tweets do not have a direct impact on Stocks held. However, as expected, the given tweets have a strong and highly significant influence (a_{11}) on the first mediator, the Tweet Sentiment (T_Sen). However, this mediator does not have a statistically significant impact (b_1) on Stocks held either, so in this case, we can neither assume a mediating or direct effect, contradicting H1.1 & H1.3. This is also confirmed by the statistically insignificant indirect effect $a_{11} * b_1$. The given financials do not have a statistically significant direct influence on Stocks held, which rejects H2.3. Although the Financial's direct effect does not exert a statistically significant influence (c'_2), there is an indirect impact of the financials on Stocks held through the mediator Financial Sentiment. Stocks held are primarily influenced by the Financial Sentiment and therefore by the perception of the nature of the financial information provided. This indirect effect ($a_{22} * b_2$) is statistically highly significant and substantial, thereby confirming H2.1. In this case, we can speak of full mediation (Baron and Kenny (1986), Zhao et al. (2010)). As suspected from the results of the previous section, the mediator Financial Sentiment is also influenced by the tweets at a 5% significance level (a_{12}). Thus, Financial Sentiment serves as a mediator for both the financials and tweets to explain Stocks held. The indirect effect of tweets on Stocks held through Financial Sentiment ($a_{12} * b_2$) is relatively smaller than the indirect effect $a_{22} * b_2$; however, it is significant and thus provides a first explanation for the group differences with the same financials (NN-NP, N-NP) from Tab. 5 and confirms H1.2. However, H2.2 must be rejected, as the financials do not exert a significant influence on the perception of tweets.

These effects remain significant even with the gradual inclusion of control variables concerning the participants' demographics, their financial background and social media usage

	Effect type		(A)	(B)	(C)	(D)
Stocks held	**Direct**					
	Tweets	(c'_1)	0.108	0.108	0.087	0.087
			(0.115)	(0.113)	(0.116)	(0.114)
	Financials	(c'_2)	-0.142	-0.153	-0.153	-0.157
			(0.106)	(0.107)	(0.099)	(0.100)
	T_Sen	(b_1)	0.059	0.076	0.086	0.096
			(0.115)	(0.111)	(0.113)	(0.110)
	F_Sen	(b_2)	0.560***	0.583***	0.575***	0.590***
			(0.115)	(0.114)	(0.111)	(0.111)
	Age			0.066		0.053
				(0.081)		(0.080)
	Male			0.098		0.046
				(0.064)		(0.067)
	Income			-0.105		-0.120
				(0.089)		(0.084)
	Risk			-0.057		-0.054
				(0.059)		(0.060)
	Economic				-0.045	-0.047
					(0.063)	(0.063)
	Cap Market				0.146*	0.150*
					(0.064)	(0.067)
	Usage				-0.026	-0.035
					(0.075)	(0.080)
	Twitter				-0.070	-0.051
					(0.065)	(0.066)
	Indirect					
	$a_{11} \rightarrow b_1$	($a_{11} * b_1$)	0.046	0.059	0.068	0.075
			(0.090)	(0.087)	(0.089)	(0.086)
	$a_{21} \rightarrow b_1$	($a_{21} * b_1$)	0.001	0.001	0.002	0.002
			(0.004)	(0.004)	(0.005)	(0.005)
	$a_{12} \rightarrow b_2$	($a_{12} * b_2$)	0.063*	0.065*	0.065*	0.066*
			(0.027)	(0.028)	(0.028)	(0.028)
	$a_{22} \rightarrow b_2$	($a_{22} * b_2$)	0.429***	0.447***	0.441***	0.452***
			(0.096)	(0.095)	(0.093)	(0.093)
T_Sen	**Direct**					
	Tweets	(a_{11})	0.784***	0.784***	0.784***	0.784***
			(0.047)	(0.047)	(0.047)	(0.047)
	Financials	(a_{21})	0.018	0.018	0.018	0.018
			(0.048)	(0.048)	(0.048)	(0.048)
F_Sen	**Direct**					
	Tweets	(a_{12})	0.112*	0.112*	0.112*	0.112*
			(0.048)	(0.048)	(0.048)	(0.048)
	Financials	(a_{22})	0.767***	0.767***	0.767***	0.767***
			(0.048)	(0.048)	(0.048)	(0.048)
R^2	Stocks held		0.261	0.285	0.293	0.311
	T_Sen		0.615	0.615	0.615	0.615
	F_Sen		0.605	0.605	0.605	0.605

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Regressions estimate equations 8 to 11 for the models (A) to (D) with gradual inclusion of controls and the respective sample size $N = 172$ for each model.

Tab. 6 Results Mediation Analysis Models (A)-(D)

(models (B) to (D)). The direct effect of tweets does not excert a significant influence on Stocks held in any model leading to the continued rejection of hypothesis H1.3. The strength of the significant direct and indirect effects on Stocks held ($a_{12} * b_2$ and $a_{22} * b_2$) in model (A) is slightly increased in models (B) to (D), while most control variables do not exert a significant influence on Stocks held. When all control variables are included in model (D), only the previous experience in capital markets at a 5% significance level has an impact on the Stocks held. In case of existing experience in capital markets more stocks are held by participants.

According to the respective R^2 values for the two mediators, the presented models explain above 60% of the total variance of the perceived Financial Sentiment and the perceived Tweet Sentiment. Also the investment decision of Stocks held can be explained with an R^2 of over 30%.

The measured effect $a_{12} * b_2$ provides an explanation for the group differences in Stocks held, as depicted in Fig. 7, when the financials are the same. However, especially Tab. 5 provides grounds to assume that tweets primarily affect Financial Sentiment when the financials are of a negative nature (NN-NP, N-NP), as in these cases, there are significant differences in perception at a 10% level for NN-NP and a 5% level for N-NP respectively, which is why a more in-depth analysis of this observation is needed.

Therefore, in the next step, we divide our overall dataset into participants who received positive financials and participants who were given negative financials for their investment decision. Subsequently, we estimate further separate mediator models for both groups. The base models for positive and negative financials (E) and (F) without control variables take the following form:

$$Stocks\_held = i_1 + c_1 * Tweets + \epsilon_1 \tag{12}$$

$$\begin{aligned} Stocks\_held = i_2 &+ c_1' * Tweets \\ &+ b_1 * Tweet\_Sentiment \\ &+ b_2 * Financial\_Sentiment + \epsilon_2 \end{aligned} \tag{13}$$

$$Tweet\_Sentiment = i_3 + a_{11} * Tweets + \epsilon_3 \tag{14}$$

$$Financial\_Sentiment = i_4 + a_{12} * Tweets + \epsilon_4 \tag{15}$$

Both basic models are consequently expanded with the demographic, financial, and social media characteristics to check the robustness of the estimations. The results of the estimation of these models (G) and (H) are depicted in Tab. 7.

The results of the estimations (E) and (F) confirm, on the one hand, the highly significant direct effect of Financial Sentiment on Stocks held (b_2) and, as expected, the highly significant influence of tweets on Tweet Sentiment. However, on the other hand by dividing the overall dataset, differences in the impact of positive and negative tweets become evident. In the case of positive financials (E), unlike the estimation with negative financials (F) and the previously estimated models (A) and (B), tweets do not exert a significant influence on the Financial Sentiment (a_{12}) and, consequently, exert no indirect effect ($a_{12} * b_2$) on the Stocks held, either. Therefore, the observable variance of Financial Sentiment, which has the dominant influence on Stocks held, can be explained to a significantly lesser extent in the model with positive financials (E) in comparison to the model with negative financials (F) since the nature of the given financials does exert an influence on the investment decision of individuals. As a result, the Financial Sentiment can be explained to a slightly but higher extent in model (F) than in model (E).

All results remain robust for both models even when control variables are included, where model (G) represents the model with control variables and positive financials, and model (H) includes control variables and negative financials. Overall, our observations align with the initial assumptions and indicate that the Financial Sentiment is particularly influenced when the available financials are negative, and the tweets contradict them in their statements. In addition, it can be seen that individuals tend to have a loss aversion as b_2 is considerable higher for negative (models (F) and (H)) than positive (models (E) and (G)) financials.

Transfering this idea of loss aversion to the given tweets we also divide the dataset by the nature of tweets in Tab. 8 estimating the following equations:

$$Stocks\_held = i_1 + c_2 * Financials + \epsilon_1 \tag{16}$$

	Effect type		(E)	(F)	(G)	(H)
Stocks held	**Direct**					
	Tweets	(c_1')	0.021	0.165	-0.013	0.119
			(0.214)	(0.133)	(0.205)	(0.138)
	T_Sen	(b_1)	0.200	-0.039	0.272	-0.007
			(0.209)	(0.136)	(0.194)	(0.135)
	F_Sen	(b_2)	0.290***	0.442***	0.305***	0.475***
			(0.097)	(0.116)	(0.101)	(0.111)
	Age				0.001	-0.060
					(0.117)	(0.082)
	Male				0.095	-0.009
					(0.105)	(0.097)
	Income				-0.075	-0.178
					(0.114)	(0.150)
	Risk				-0.036	-0.070
					(0.098)	(0.086)
	Economic				-0.111	0.020
					(0.093)	(0.093)
	Cap Market				0.166	0.150
					(0.097)	(0.099)
	Usage				-0.099	-0.028
					(0.116)	(0.114)
	Twitter				-0.121	-0.024
					(0.093)	(0.092)
	Indirect					
	$a_{11} \rightarrow b_1$	($a_{11} * b_1$)	0.163	-0.029	0.221	-0.005
			(0.170)	(0.103)	(0.157)	(0.102)
	$a_{12} \rightarrow b_2$	($a_{12} * b_2$)	0.021	0.114*	0.022	0.123*
			(0.030)	(0.046)	(0.032)	(0.049)
T_Sen	**Direct**					
	Tweets	(a_{11})	0.813***	0.756***	0.813***	0.756***
			(0.062)	(0.071)	(0.062)	(0.071)
F_Sen	**Direct**					
	Tweets	(a_{12})	0.073	0.257*	0.073	0.257*
			(0.106)	(0.105)	(0.106)	(0.105)
R^2	Stocks held		0.142	0.244	0.224	0.301
	T_Sen		0.660	0.572	0.660	0.572
	F_Sen		0.005	0.066	0.005	0.066
Group	N		87	85	87	85
	Financials		Positive	Negative	Positive	Negative

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Regressions estimate equations 12 to 15 for the models (E) to (H) with gradual inclusion of controls and the respective sample size N for each model.

Tab. 7 Results Mediation Analysis Models (E)-(H)

	Effect type		(I)	(J)	(K)	(L)
Stocks held	**Direct**					
	Financials	(c_2')	-0.063	-0.413*	-0.100	-0.425**
			(0.132)	(0.162)	(0.125)	(0.155)
	T_Sen	(b_1)	0.006	0.129	0.066	0.159
			(0.101)	(0.094)	(0.098)	(0.091)
	F_Sen	(b_2)	0.479***	0.847***	0.551***	0.849***
			(0.145)	(0.181)	(0.141)	(0.175)
	Age				0.003	0.062
					(0.111)	(0.119)
	Male				0.079	0.032
					(0.101)	(0.109)
	Income				-0.168	-0.079
					(0.101)	(0.138)
	Risk				-0.001	-0.117
					(0.017)	(0.089)
	Economic				-0.110	-0.023
					(0.083)	(0.099)
	Cap Market				0.195	0.091
					(0.101)	(0.090)
	Usage				-0.100	0.064
					(0.116)	(0.132)
	Twitter				-0.085	-0.066
					(0.096)	(0.101)
	Indirect					
	$a_{21} \rightarrow b_1$	($a_{21} * b_1$)	0.000	0.001	0.003	0.001
			(0.014)	(0.014)	(0.008)	(0.018)
	$a_{22} \rightarrow b_2$	($a_{22} * b_2$)	0.325***	0.743***	0.374***	0.744****
			(0.110)	(0.161)	(0.109)	(0.154)
T_Sen	**Direct**					
	Financials	(a_{21})	0.042	0.007	0.042	0.007
			(0.699)	(0.109)	(0.699)	(0.109)
F_Sen	**Direct**					
	Financials	(a_{22})	0.679***	0.877***	0.679***	0.877***
			(0.080)	(0.052)	(0.080)	(0.052)
R^2	Stocks held		0.193	0.296	0.293	0.340
	T_Sen		0.002	0.000	0.002	0.000
	F_Sen		0.461	0.769	0.461	0.769
Group	Observations		87	85	87	85
	Tweets		Positive	Negative	Positive	Negative

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Regressions estimate equations 16 to 19 for the models (I) to (L) with gradual inclusion of controls and the respective sample size N for each model.

Tab. 8 Results Mediation Analysis Models (I)-(L)

$$Stocks\_held = i_2 + c_2' * Financials + b_1 * Tweet\_Sentiment + b_2 * Financial\_Sentiment + \epsilon_2 \quad (17)$$

$$Tweet\_Sentiment = i_3 + a_{21} * Financials + \epsilon_3 \quad (18)$$

$$Financial\_Sentiment = i_4 + a_{22} * Financials + \epsilon_4 \quad (19)$$

In the case of negative tweets the effect of perceived Tweet Sentiment (b_1) remains insignificant. Nevertheless, it might be noteworthy that the b_1 coefficients in the negative models (J) and (L) of 0.129 and 0.167 are higher than in the positive models (I) and (K) and also show smaller standard erorrs leading to p-values decreasing from 95% to 17%, respectively from 57% to 7%. This observation could give justification for not rejecting H 1.1 but should not be overvalued as the effect is negligible aligning with the observations of Boulu-Reshef et al. (2023). In contrast, no significant differences for the given financials between positive and negative tweets can be found.[14]

2.5. Conclusion

The objective of this study is to illuminate the causal pathway of available information on the investment decisions of economic agents. Specifically, the focus is on a detailed examination of the impact of social media posts and their perception. To achieve this goal, a laboratory experiment was conducted, providing participants with various pieces of information in the form of financial data and tweets to inform an investment decision. The aim is to draw conclusions about the causal channels of the provided information based on the investment decisions made by the participants at the end of the experiment. Following their investment decisions, participants were surveyed regarding their perception of the financials and tweets using a Likert scale. This allows for an examination of whether participants perceived the information in line with the author's intentions. As significant differences in participants' perceptions between the individual groups were expected, it can be inferred that

[14]Following Zhao et al. (2010) we can observe a competitive mediation with $a_{22} * b_2 * c_2' < 0$ in the models (J) and (L) leading to a summed whole effect of the financials which is nearly the same as of the positive pendants (I) and (K).

the information was perceived as intended. Furthermore, the Financial and Tweet Sentiment provide an opportunity for a more in-depth analysis of the causal pathway of these two pieces of information.

To address this, the method of mediation analysis was employed to separate the influence of the given information into direct and indirect effects. It was revealed that particularly the perception of information has a significant effect on the investment decisions of economic agents. While the Tweet Sentiment does not directly influence investment decisions (or just with a neglible impact when tweets are negative), the tweets do impact the perception of financials, which in turn significantly influences investment decisions. This result is in line with existing literature in two different ways. On the one hand we show that social media sentiment does influence the investment decisions of individuals, which has previously also been shown by i.a. Antweiler and Frank (2004); Baker and Wurgler (2006); Da et al. (2015); Das and Chen (2007); Renault (2017); Sun et al. (2016); Tetlock (2007). On the other hand, our results align with the findings of Behavioral Finance. Contrary to the participants' self-reported statements, their investment decisions are subconsciously influenced by the provided tweets, indicating the existence of biases in the information processing process.

In this specific case, the behavior of the participants suggests the presence of the anchoring effect, as presented by Tversky and Kahneman (1974). According to this effect, the tweets, with their content, act as a mental anchor that distorts the interpretation of the financial information. Additionally, we observe a differential impact of tweets on Financial Sentiment when the financials are positive or negative. Our results suggest that an influence exists when negative financial information is present, and the tweets contradict it, i.e., they are positively framed. This could be rooted in the prospect theory, wherein, in the case of losses expressed through negative financials, participants, due to their risk aversion, behave differently than in the case of positive financials. In this scenario they may be more susceptible to information from tweets that deviate from the financials. The results of our study provide three starting points for further research and the practical application of sentiment analysis regarding the precise direction of the impact of social media sentiment we presented. Firstly, the models discussed could be expanded to include moderators that could serve as catalysts for the strength of the effect of social media sentiment. This could provide insights into relevant factors influencing the susceptibility of economic agents to social media sentiment. However, such an analysis would require a broader participant base and, consequently, a higher number of observations per study group than was the case in this study.

Secondly, the influence of bot-generated tweets on our participants suggests that despite the automated generation of these tweets, an impact on economic agents occurs. It seems possible to influence the assessment of a company's financial situation using computer-generated social

media content. In light of the advancing development of AI, an accurate measurement of this approach compared to the use of human-generated tweets appears necessary.

Finally, our results indicate that the influence of social media sentiment on investor decisions is of an indirect nature. Therefore, it seems advisable to take this into greater consideration in future analyses. To the best of our knowledge, this is the first experimental study that dissects the causal pathway of social media sentiment through a mediation analysis into direct and indirect effects, aiming to gain a deeper understanding of its impact on the investment behavior of economic agents.

2.6. Appendix

2.6.1. Platform's Interface and Content

Company description

Fig. 8 Company Description Interface (German Language, Translation Below)

Translation: **We are Glubon** - Glubon improves the everyday life with intelligent solutions for multiple generations. Since 125 years we are driven by our vision every day improving our all and future generation's life with our innovative and sustainable products and technologies. At our company everything is dedicated to our guiding principle: 'grow responsible'.

With over 120,000 employees in over 50 different countries we belong the worldwide leading suppliers of industry and consuming goods. To our innovation and product range count multiple intelligent solutions in the sections plastics, carbon, metal and glas.

Financials

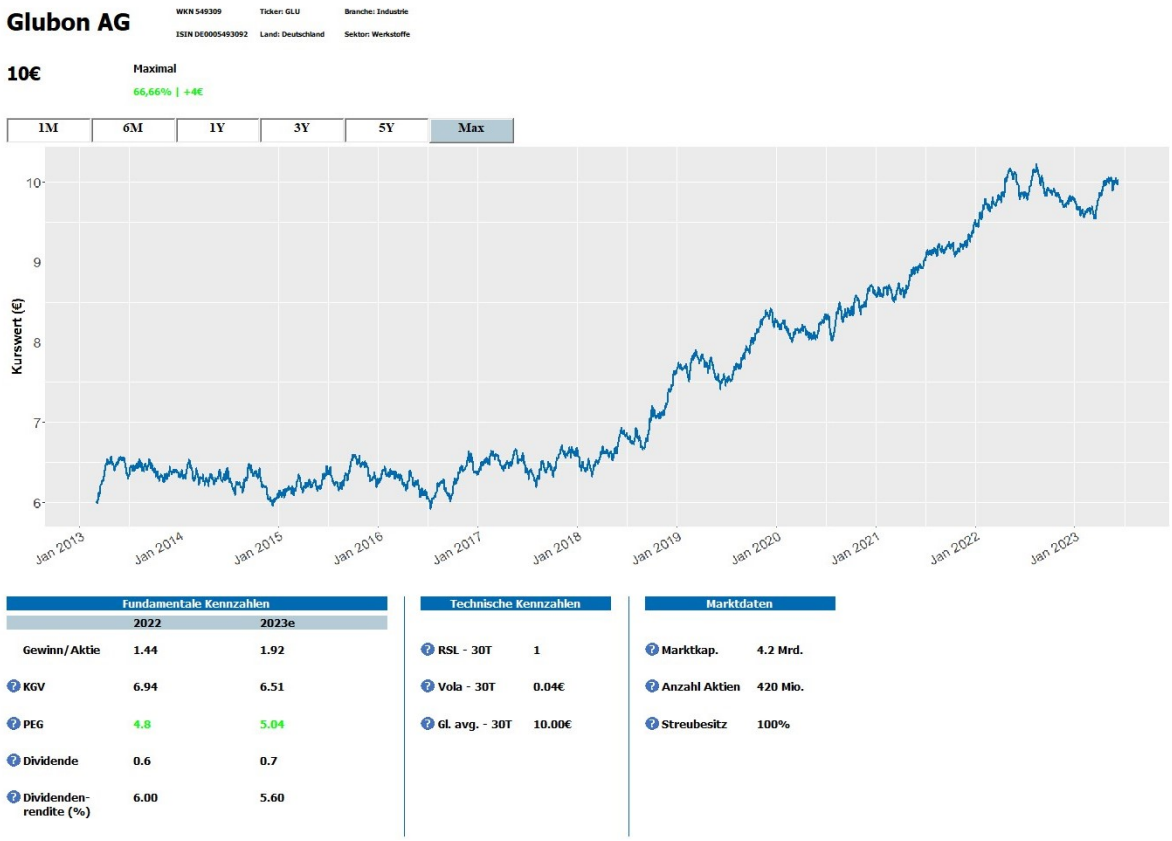


Fig. 9 *Financials* Tab, Max Chart Opened (Positive Version)

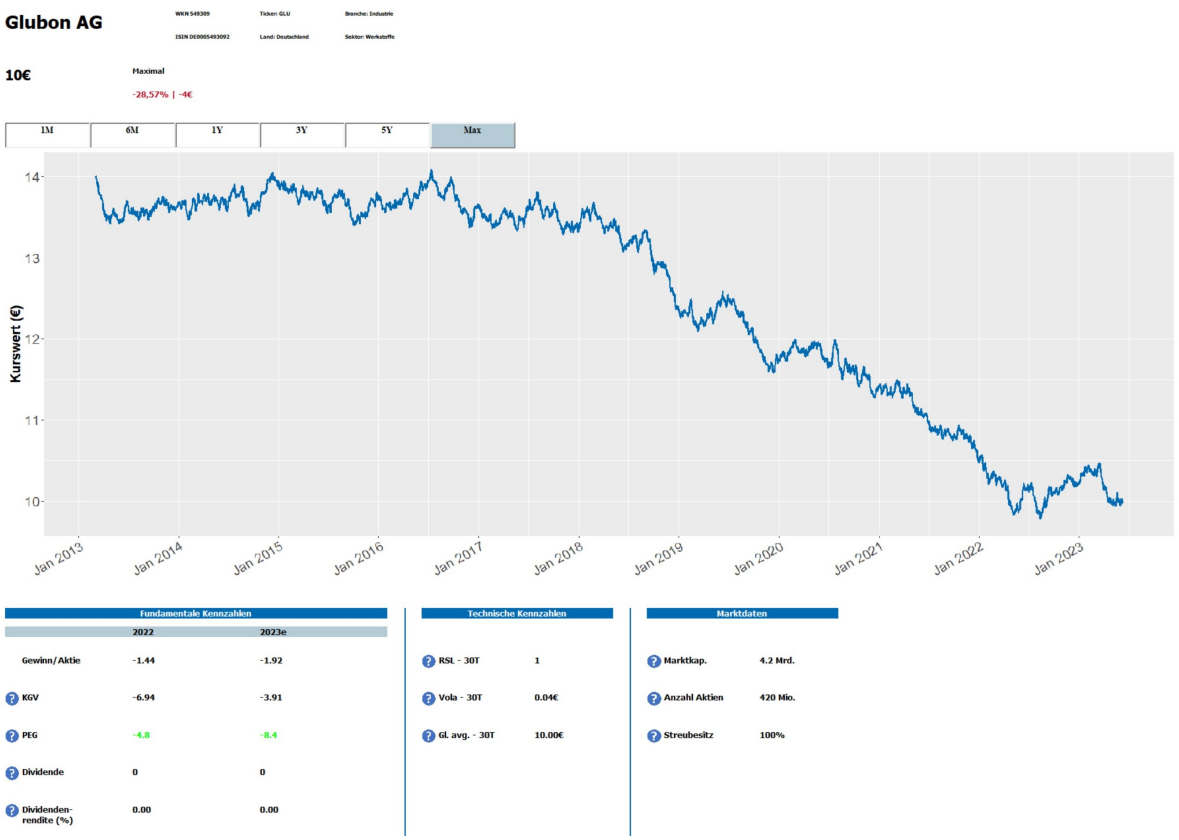


Fig. 10 *Financials* Tab, Max Chart Opened (Negative Version)

Tweets

German ChatGPT query: Generiere mir 10 <*colloquial*> deutsche <*sentiment*> Tweets über die imaginäre Firma Glubon bezüglich Ihrer Aktien, Finanzen, Strategie, Nachhaltigkeit oder Ihres Managements mit maximal <*max length*> Zeichen und dem Cashtag $GLU sowie keinen Emojis.

Translated ChatGPT query: Generate 10 <*colloquial*> German <*sentiment*> Tweets about the imaginary company Glubon regarding their stocks, Financials, strategy, sustainability or management with maximal <*max length*> characters and the cashtag $GLU as well as no emojis for me.

<*colloquial*> = {'colloquial', ' '}
<*sentiment*> = {'positive', 'negative', 'neutral'}
<*max length*> = {20, 70, 140}

Original Tweet	Translation	Max length	sentiment	colloquial
Glubon zeigt beeindruckende Finanzergebnisse und beweist erneut, warum sie ein solider Wert für langfristige Investitionen sind. $GLU	Glubon shows impressive financial results and proves again why they are a solid value for long-term investments. $GLU	140	positive	no
Glubon-Aktien performen hervorragend und bieten Anlegern eine solide Rendite. $GLU	Glubon-stocks perform excellently and deliver investors a solid return. $GLU	70	positive	no
Top-Finanzen bei Glubon! $GLU	Top-Financials at Glubon! $GLU	20	positive	no
Die Aktien von Glubon sind der Hammer, Leute! Die machen richtig Knete und lassen uns alle mitverdienen. $GLU	The Glubon stocks are amazing, folks! They're making serious dough and letting all of us earn a share. $GLU	140	positive	yes
Glubon-Aktien ballern richtig! Hier gibt's fette Gewinne, Brudi! $GLU	The Glubon stocks are really booming! There are fat profits here, bro! $GLU	70	positive	yes

Glubon-Aktien abgefahren! $GLU	Glubon stocks are off the charts! $GLU	20	positive	yes
Glubon berücksichtigt Nachhaltigkeitsaspekte in ihrem Geschäft und strebt einen verantwortungsbewussten Umgang mit Ressourcen an. $GLU	Glubon considers sustainability aspects in their business and aims for responsible resource management. $GLU	140	neutral	no
Glubon legt Wert auf Nachhaltigkeit und Ressourcenschonung. $GLU	Glubon emphasizes sustainability and resource conservation. $GLU	70	neutral	no
Strategie solide. $GLU	Strategy is solid. $GLU	20	neutral	no
Die Aktien von Glubon sind ganz okay, nichts Weltbewegendes, aber auch keine Totalausfälle. Mal sehen, wie's weitergeht. $GLU	The Glubon stocks are just okay, nothing groundbreaking, but not total disappointments either. Let's see how it goes. $GLU	140	neutral	yes

Finanzen bei Glubon okay, nix Besonderes, aber auch nicht im Keller. So mittel halt. $GLU	Finances at Glubon are okay, nothing special, but not at rock bottom either. Just average. $GLU	70	neutral	yes
Management ganz okay. $GLU	Management is quite okay. $GLU	20	neutral	yes
Die Strategie von Glubon ist zum Scheitern verurteilt, kein Wunder, dass sie den Markt nicht dominieren können. $GLU	Glubon's strategy is doomed to fail; no wonder they can't dominate the market. $GLU	140	negative	no
Finanzen bei Glubon katastrophal, rote Zahlen ohne Ende. Keine gute Wahl für Anleger. $GLU	Finances at Glubon are catastrophic, endless red figures. Not a good choice for investors. $GLU	70	negative	no
Strategie bei Glubon schwach. $GLU	Strategy at Glubon is weak. $GLU	20	negative	no

Ey, die Aktien von Glubon sind voll der Reinfall, voll im Keller! Wer da investiert, hat echt 'nen Schaden. Finger weg! $GLU	Hey, Glubon stocks are a complete flop, way down in the dumps! Investing there is a real mistake. Stay away! $GLU	140	negative	yes
Finanziell geht's bei Glubon den Bach runter, die sind pleite! $GLU	Financially, Glubon is going downhill, they're bankrupt! $GLU	70	negative	yes
Nachhaltigkeit Fehlanzeige. $GLU	No sustainability in sight. $GLU	20	negative	yes

Tab. 9 Tweet Examples per ChatGPT Query

2.6.2. Robustness Checks

Tab. 10: Removal of slowest and fastest participants

	Effect type		(M)	(N)	(O)
Stocks held	**Direct**				
	Tweets	(c_1')	0.096	0.101	0.111
			(0.118)	(0.118)	(0.122)
	Financials	(c_2')	-0.138	-0.180	-0.161
			(0.117)	(0.099)	(0.117)
	T_Sen	(b_1)	0.088	0.073	0.063
			(0.115)	(0.113)	(0.117)
	F_Sen	(b_2)	0.571***	0.617***	0.599***
			(0.115)	(0.111)	(0.130)
	Age		0.057	0.083	0.086
			(0.082)	(0.081)	(0.085)
	Male		0.061	0.058	0.073
			(0.069)	(0.067)	(0.069)
	Income		-0.140	-0.138	-0.160
			(0.088)	(0.087)	(0.092)
	Risk		-0.043	-0.066	-0.058
			(0.064)	(0.061)	(0.066)
	Economic		-0.043	-0.040	-0.034
			(0.064)	(0.064)	(0.065)
	Cap Market		0.148*	0.165*	0.164*
			(0.068)	(0.068)	(0.069)
	Usage		-0.037	-0.021	-0.023
			(0.081)	(0.080)	(0.080)
	Twitter		-0.055	-0.031	-0.035
			0.068	(0.064)	(0.066)
	Indirect				
	$a_{11} \rightarrow b_1$	($a_{11} * b_1$)	0.069	0.057	0.049
			(0.090)	(0.088)	(0.092)
	$a_{21} \rightarrow b_1$	($a_{21} * b_1$)	0.002	0.001	0.002
			(0.005)	(0.004)	(0.005)
	$a_{12} \rightarrow b_2$	($a_{12} * b_2$)	0.065*	0.067*	0.066*
			(0.028)	(0.030)	(0.030)
	$a_{22} \rightarrow b_2$	($a_{22} * b_2$)	0.457***	0.447***	0.484***
			(0.110)	(0.093)	(0.112)
T_Sen	**Direct**				
	Tweets	(a_{11})	0.786***	0.784***	0.787***
			(0.049)	(0.049)	(0.050)
	Financials	(a_{21})	0.024	0.018	0.024
			(0.049)	(0.049)	(0.050)
F_Sen	**Direct**				
	Tweets	(a_{12})	0.113*	0.109*	0.110*
			(0.046)	(0.049)	(0.047)
	Financials	(a_{22})	0.801***	0.773***	0.809***
			(0.046)	(0.050)	(0.047)
R^2	Stocks held		0.302	0.319	0.309
	T_Sen		0.620	0.616	0.671
	F_Sen		0.659	0.612	0.621
N	Observations		163	163	154

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$
Regressions estimate equations 8 to 11 for the models (M) to (O) with gradual exclusion of the 5% fastest (M)/slowest (N)/fastest and slowest participants(O))

Tab. 10 Results Mediation Analysis Models (M)-(O)

Tab. 11: Results per check questions correctly answered

	Effect type		(P)	(Q)	(R)
Stocks held	**Direct**				
	Tweets	(c_1')	0.049	0.061	0.087
			(0.114)	(0.113)	(0.114)
	Financials	(c_2')	-0.160	-0.146	-0.157
			(0.096)	(0.097)	(0.100)
	T_Sen	(b_1)	0.137	0.136	0.096
			(0.109)	(0.109)	(0.110)
	F_Sen	(b_2)	0.564***	0.562***	0.590***
			(0.106)	(0.107)	(0.111)
	Age		0.023	0.019	0.053
			(0.070)	(0.070)	(0.080)
	Male		0.029	0.039	0.046
			(0.067)	(0.067)	(0.067)
	Income		-0.125	-0.122	-0.120
			(0.076)	(0.076)	(0.084)
	Risk		-0.076	-0.084	-0.054
			(0.059)	(0.059)	(0.060)
	Economic		-0.040	-0.039	-0.047
			(0.062)	(0.062)	(0.063)
	Cap Market		0.141*	0.123	0.150*
			(0.066)	(0.065)	(0.067)
	Usage		-0.064	-0.041	-0.035
			(0.081)	(0.079)	(0.080)
	Twitter		-0.000	-0.012	-0.051
			(0.065)	(0.065)	(0.066)
	Indirect				
	$a_{11} \rightarrow b_1$	($a_{11} * b_1$)	0.108	0.107	0.075
			(0.086)	(0.086)	(0.086)
	$a_{21} \rightarrow b_1$	($a_{21} * b_1$)	0.003	0.003	0.002
			(0.007)	(0.007)	(0.005)
	$a_{12} \rightarrow b_2$	($a_{12} * b_2$)	0.071**	0.070**	0.066*
			(0.027)	(0.026)	(0.028)
	$a_{22} \rightarrow b_2$	($a_{22} * b_2$)	0.430***	0.428***	0.452***
			(0.088)	(0.089)	(0.093)
T_Sen	**Direct**				
	Tweets	(a_{11})	0.787***	0.785***	0.784***
			(0.045)	(0.046)	(0.047)
	Financials	(a_{21})	0.022	0.022	0.018
			(0.047)	(0.047)	(0.048)
F_Sen	**Direct**				
	Tweets	(a_{12})	0.126**	0.125**	0.112*
			(0.046)	(0.047)	(0.048)
	Financials	(a_{22})	0.763***	0.762***	0.767***
			(0.048)	(0.048)	(0.048)
R^2	Stocks held		0.287	0.297	0.311
	T_Sen		0.621	0.618	0.615
	F_Sen		0.608	0.604	0.605
N	Observations		182	180	120

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$
Regressions estimate equations 8 to 11 for the models (P) to (R) including all particpants (P) and participants who correctly answered at least 2(Q) or 4(R) control questions.)

Tab. 11 Results Mediation Analysis Models (P)-(R)

2.7. Declaration of (Co-)Authors and Record of Accomplishments

Title:	The relevance and influence of social media posts on investment decisions − An experimental approach based on tweets
Author(s):	**Lars M. Kürzinger** (Heinrich-Heine University Düsseldorf) Philipp Stangor (Heinrich-Heine University Düsseldorf)
Conference(s):	Participation and presentation at 'Forschungskolloquium Finanzmärkte', 25^{th} January 2023, Düsseldorf, Germany Participation and presentation at 'HVB Doctoral Colloquium', 2^{nd} – 3^{rd} February 2024, Münster, Germany
Publication:	SSRN published. Submitted to the 'Journal of Behavioral and Experimental Finance', double-blind peer-reviewed journal. Current status: Under review.

Share of contributions:

Contributions	**Lars M. Kürzinger**	**Philipp Stangor**
Research Design	**65%**	**35%**
Development of research question	50%	50%
Method developement	80%	20%
Research performance & analysis	**50%**	**50%**
Literature review and framework development	50%	50%
Data collection, preparation and analysis	20%	80%
Analysis and discussion of results	80%	20%
Derivation of implications and conclusions	50%	50%
Manuscript preparation	**45%**	**55%**
Final draft	45%	55%
Finalization	40%	60%
Overall contribution	**60%**	**40%**

31.05.2024, 

Date, Lars M. Kürzinger

31.05.2024, 

Date, Philipp Stangor

3. Measuring investor sentiment from Social Media Data - An emotional approach

3.1. Abstract

We employ a multidimensional approach extracting investor sentiment from social media data using the NRC-Emotion Association Lexicon. Considering a vast number of short text messages from the financial microblogging platform StockTwits, we analyze different emotions contained in each message. Subsequently, we classify these posts as bullish or bearish signals on basis of their emotional profile using machine learning techniques to develop aggregated investor sentiment. This classification outperforms comparable classifications based on non-economic or two-dimensional dictionaries in terms of accuracy and data efficiency. Consequently, we are able to predict intraday returns for the S&P 500 and NASDAQ 100.

This paper contributes to the existing literature by outlining the advantages of a multidimensional analysis and pointing out three key factors for designing accurate field-specific dictionaries, which need to be context specific, emotion-based *and* economic-related.

3.2. Introduction

With the rise of social media platforms such as Twitter, Facebook and Instagram and their growing popularity, many researchers have investigated the potential influence of a platform's content on the performance of stock markets. As investor's attention is found to be limited, their investment behavior tends to be biased towards investments that consciously or unconsciously attract their attention (Barber and Odean (2008)). In this case, social media platforms might affect an individual's investment decision (Liu (2020)). Johnson and Tversky (1983) already observed that sentiment is able to affect investors' perception of risk. Kaplanski et al. (2015) confirm this finding, even going so far as to detect the effects of investors' personal happiness on their investment behavior. Furthermore, Baker et al. (2007) conclude that what is in question is no longer whether sentiment influences market participants but rather how strong its effect may be and how its measured.

In this regard, we extract aggregated investor sentiment by analyzing a vast number of social media posts and examine the sentiment's influence on market movement.

Fig. 11 outlines the progress made in economic text analysis. Linguistic text analysis initially classified single words by matching them with their predefined connotation in a linguistic dictionary that was originally derived from psychological analysis. Hence, a word's connotation is usually distinguished between 'positive' and 'negative' (see, i.a., Antweiler and Frank (2004), Baker et al. (2007), Gao and Yang (2017), Kim and Kim (2014) and Sun et al. (2016)). However, when conducting this analysis in an economic setting, one faces

the question of whether a word's meaning in a psychological context might differ from its meaning in an economic context. Thus, the word 'risk' might be connotated (very) negatively in the first setting, while this might not be the case in an economic analysis. Therefore, starting with Henry (2008) and Loughran and McDonald (2011), sentiment dictionaries intentionally designed for economic uses of language have been created and used for a more economically specific analysis. Nevertheless, further linguistic challenges such as punctuation, slang, irony or emoticons were not considered, which led to the rise of rule-based models to further evaluate social media data. One of most prominent dictionaries and sentiment analysis tools in this context is the so-called *Valence Aware Dictionary and sEntiment Reasoner* (VADER), which is able to consider the abovementioned linguistic components, making it possible to estimate the degree to which a microblogging text contains positive or negative sentiment (Hutto and Gilbert (2014)). Apart from those dictionary based approaches newer techniques such as NLP Transformers like i.a. BERT (Devlin et al. (2018)), XLNet (Yang et al. (2019)) and XLM (Lample and Conneau (2019)) recently emerged and have been optimized continuously. These NLP Transformers make use of different kinds of machine learning techniques to achieve high accuracies in text classification tasks.[15]

However, in this work we take a step back to answer the question of whether a multidimensional analysis might present a better starting point for sentiment analysis than a two-dimensional approach. We find that a multidimensional approach using emotions outperforms comparable classifications based on non-economic or two-dimensional dictionaries in terms of accuracy and data efficiency. When using the NRC-Emotion Association Lexicon created by Mohammad and Turney (2013) (also known as 'EmoLex'), we do not match positive or negative connotations with a given microblogging text but rather with up to eight different emotions associated with each word. As EmoLex is a dictionary without an economic background (Fig. 11: A1) a suitable benchmark is a positive-negative dictionary without an economic context (Fig. 11: B1).

To further validate our results, we compare the accuracy of our approach with other benchmark dictionaries that are already widely used in (economics) literature and practice and possess an economic background, as well (Fig. 11: B2).

Our results emphasize the need for (more specific) emotion-based *and* economic-related dictionaries. To the best of our knowledge, we are unaware of other studies using this explicit technique in the same way. With our results, we encourage economic research in textual sentiment analysis to focus more on multidimensional emotional approaches than on

[15]For a more extensive overview concerning diffenrent approaches of sentiment analysis (including dictionary based approaches and NLP Transformers) see Mishev et al. (2020).

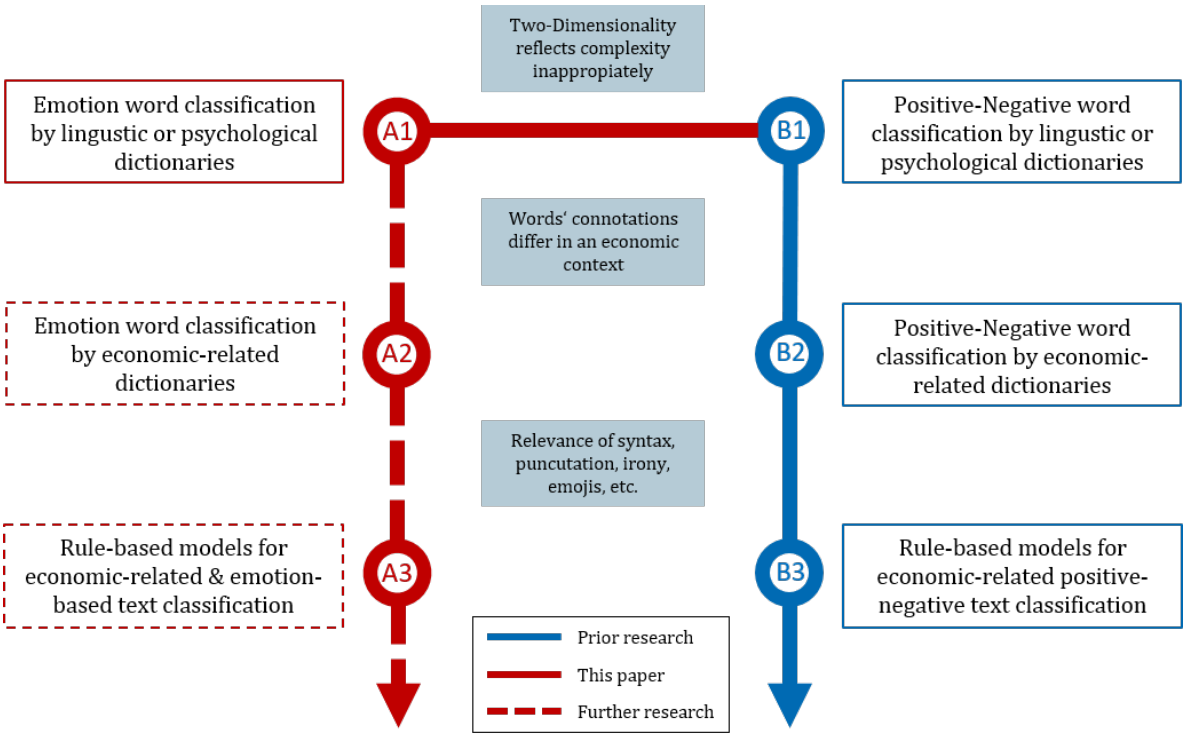


Fig. 11 Progress of Economic-related Text-analysis Research

two-dimensional approaches as the most prominent positive-negative approaches used in the majority of related research. We expand the existing literature by outlining three main factors determining the success of a field-specific sentiment analysis dictionary: multidimensional scoring (for example emotions), economic word connotation and type of text. Our dictionary based results from the beginnings of sentiment analysis (Fig. 11, A1) also give implications for more sophisticated approaches of sentiment analysis. When considering our findings, one could expect the results of other approaches like more advanced dictionaries (Fig. 11, A3) or NLP Transformers to profit from a more dimensional analysis further improving classification results. Future research should take this hypothesis into consideration and validate our basic findings.

The remainder of this paper is structured as follows: Section 3.3 gives a short overview of the related literature regarding sentiment analysis. Section 3.4 presents our data, namely, the ideas from the social media platform StockTwits and the chosen stock market data for proving the economic relevance of our results. In Section 3.5, we describe our method, which leads to our results presented in Section 5.4. Section 3.7 concludes the paper, relates our observations with prior results found in the literature and provides an outlook on possible future research topics.

3.3. Literature Review

Beginning with the work of Antweiler and Frank (2004), internet stock messages have been investigated for to their suitability to measure market sentiment and thus to predict the movement of markets. Antweiler and Frank (2004), among other studies, (see, i.a., Das and Chen (2007), Kim and Kim (2014)) do not find a significant relationship between sentiment and market returns but reveal a correlation among social media activity, trading volume and return volatility. Although Kim and Kim (2014) do not find any relationship of the abovementioned kind, other studies do find significant relationships between intraday sentiment and intraday returns (Sun et al. (2016), Gao and Yang (2017)) or overnight returns (Renault (2017)). One possible explanation for the differing results might lie in the changing composition of social media users and their behavior over time, as Renault (2017) argues.

Following Stiglitz and Grossman (1980), however, we assume that market participants are able to obtain small excess returns as compensation for continuous information gathering, contradicting market information efficiency (Jensen (1978)). These additional returns can be viewed as a reward for monitoring and analyzing market information that compensate market participants for the costs associated with monitoring and maintaining the market's signals. In a competitive market setting, however, small excess returns are assumed to be short-lived since professional investors will exploit any value-relevant information to gain an information advantage over their competitors (Renault (2017)).

Therefore, individuals will make use of any institution that reduces information costs by centralizing, selecting and verifying information, which explains the emergence of information service providers such as Reuters or Bloomberg. Usually, fees for using these services exceed small investors' capabilities. Social media platforms may represent one means of filling this gap, making it easier to obtain potential value-relevant information. This finding is in line with Baker et al. (2007), for example, who argue that sentiment effects hold especially for 'small-capitalization, younger, unprofitable, high-volatility, non-dividend-paying, growth companies or stocks of firms in financial distress' since they might be more difficult to value due to increasing information costs.

Furthermore, in a behavioral finance context, stock prices may differ from their fundamental value due to possibly irrational investor behavior. Bullish or bearish expectations among noise traders might therefore be able to move stock prices (De Long et al. (1990)). For example, individuals tend to overvalue a conversation partner's opinion (DeMarzo et al. (2003) or may be more willing to invest in certain assets because they have aroused their attention consciously or unconsciously (Barber and Odean (2008)). Behavioral biases such as these might be one of the reasons that social media sentiment analysis appears tempting in a financial setting, as it may provide an explanation for individuals' noisy behavior in the

sense of Black (1986) and simultaneously provide an explanation for why people participate in social media platforms such as StockTwits and publish their beliefs. In the setting described by DeMarzo et al. (2003) and Giannini et al. (2018), it might even be rational for institutional investors, who are often assumed to be less susceptible to biases, to follow opinion leaders since they are able to move markets or even become influential themselves. Furthermore, communication between market participants appears to be suitable to convince hesitant market participants to invest in certain assets, as they learn of other individuals who share a similar opinion about an investment possibility (Cao et al. (2002), Antweiler and Frank (2004)). Knowledge of these ways of behavior might even provide incentives to individuals to deliberately spread rumors about assets in an attempt to profit from the expected reactions their followers might take (van Bommel (2003)), thereby explaining questions concerning the motivation of informed investors to publish their information (see Xiong et al. (2019)). Bullish or bearish expectations among noise traders are therefore able to move stock prices (De Long et al. (1990), Black (1986)).

For this purpose, we define sentiment as a market's general, psychological environment. Currently, three different methods to obtain a market's sentiment can be found in the literature. The first alternative resembles the analysis of market-based data such as trading volumes, IPO returns or IPO volumes using high-frequency data (e.g., Lee et al. (1991) or Baker and Wurgler (2006)). However, Qiu and Welch (2004) and Da et al. (2015) argue that these types of studies suffer from the vast number of potential variables at hand and their interdependencies. Second, surveys such as the Consumer Sentiment Index represent another method of measuring investor sentiment (i.a. Brown and Cliff (2005)) but are only frequently conducted and therefore suffer from low frequency, making them unsuitable for analyzing short-lived excess returns. Additionally, little incentive exists to truthfully answer survey questions, resulting in potentially biased survey results (Singer (2010)). As a consequence, we employ the third alternative in the form of a textual-based analysis with reference to Tetlock (2007) and Renault (2017), using a linguistic approach to evaluate text data from the microblogging platform StockTwits. This approach enables us to make use of high-frequency text data created by the platform's users and the data's living lab properties, negating the abovementioned issues.

3.4. Data

3.4.1. StockTwits

In this study, we use data from the microblogging platform StockTwits as formerly done in, for example, Renault (2017), Giannini et al. (2018) and Cookson and Niessner (2020). Ranked by the website analytics tool Alexa as the 768th most popular website in the USA

as of April 2022, the platform addresses individuals, professionals and institutions who want to share their opinions, thoughts and ideas about financial topics. Sprenger et al. (2014) correctly note how many (early) results in the field of financial textual sentiment research lack statistical significance because of using un- or inadequately filtered data. In this manner, the platform's concept addresses those emerging problems in a way not done by other microblogging platforms (e.g., Twitter) without losing the advantage of generating a considerable amount of real-time data. Another noteworthy benefit of StockTwits data is the user's ability to flag their ideas as 'bullish' or 'bearish', thereby eliminating the need for researchers to manually classify ideas into each category. In prior research, this issue has often led to the problem of misclassification due to subjective classification. An additional noteworthy feature of our data is the possibility for users to reveal information about themselves within the shown categories in Tab. 12:

Category	**Possible Expressions**
Trading Experience	Novice, Intermediate, Professional
Holding Period	Day Trader, Position Trader, Swing Trader, Long Term Investor
Trading Approach	Technical, Momentum, Growth, Fundamental, Global Macro, Value
Trading Asset	Equities, Options, Forex, Futures, Bonds, Private Companies

Tab. 12 User Categories and Possible Expressions

As in every self-classification task, there is obvious potential for misclassification by the users, especially due to the possible benefits of over- or underestimating themselves. In our case, we do not expect systematic problems to occur for the last three categories of Tab. 12, since the categories are well distinguished from one another and understandable for the users interested in participating on such a platform and there is no incentive for misclassification. However, there might exist an incentive for users to overestimate their trading experience before the community such that self-classification could suffer from bias. Nevertheless, differences between the expressions ('novice', 'intermediate' and 'professional') can be interpreted in the following.

At the end of 2020, StockTwits had traffic of over 40,000 active users[16] sharing nearly 300,000 ideas on average per day. Fig. 12 illustrates that (1) both numbers have strongly increased in recent years and (2), for this reason, the latter results are constantly in need of updates and improvements (e.g., Renault (2017) also states himself).

To update the latter research results, we use all ideas published on the platform from January 2012 until the end of December 2020 by accessing the StockTwits Developer API.

[16] As active users per day, we define the number of users who published at least one idea on the platform on that day.

After clearing the data of ideas that are not suitable for the measurement of textual sentiment, for example, ideas that only contain 'cashtags' as identifiers for several stocks ($), pictures or hyperlinks, 250,321,511 ideas remain in the chosen time horizon; 75,414,994 (30.13%) have been classified by the StockTwits community – 62,826,233 (25.10%) as 'bullish' (N_{Bu}) and 12,588,761 (5.03%) as 'bearish' (N_{Be}). To the best of our knowledge, this is one of the largest StockTwits data samples used in published relevant research to date.

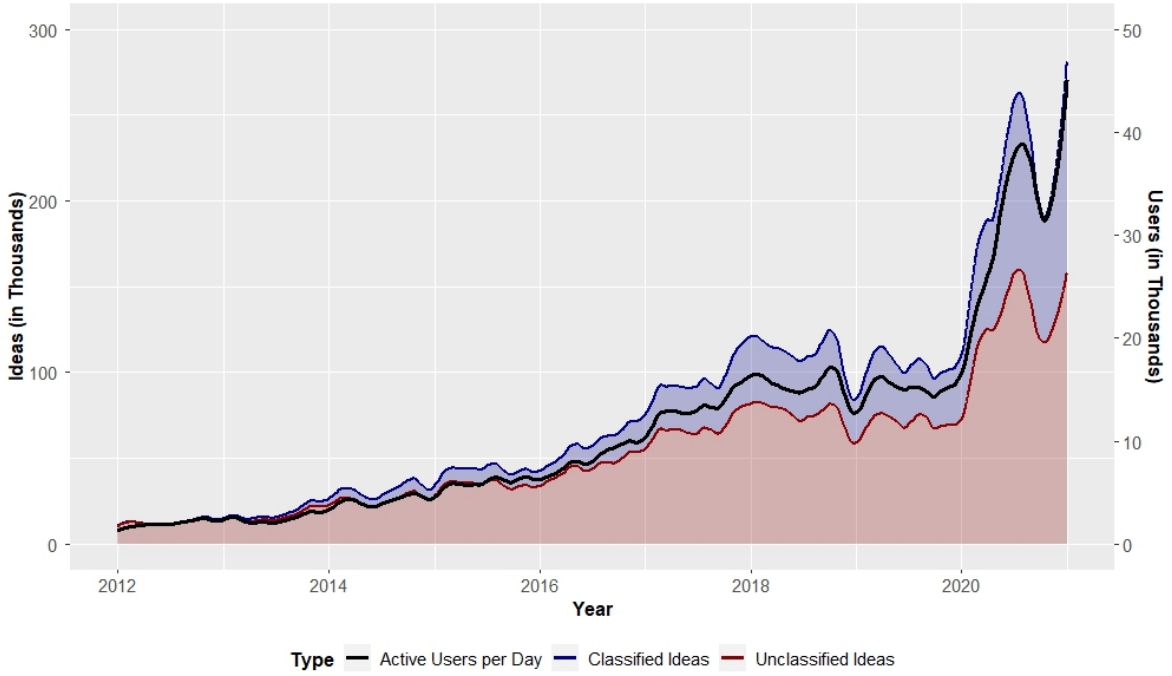


Fig. 12 Number of Shared Ideas and Active Users per Day (Loess-smoothed)

The higher rate of bullish ideas can be explained by the predominantly bullish market conditions in the chosen time horizon and the fact that individuals generally tend to share positive rather than negative news. The major share of unclassified ideas illustrates how important a suitable classification with the help of emotion scores is. On average, 34,833 ideas were classified on StockTwits per day in 2019. Assuming an equal distribution over time, approximately 1,451 ideas per hour or only 24 ideas per minute were published on average in 2019. Considering the discussed economic theory, enlarging the dataset by classifying all published ideas improves prediction quality and allows for a more detailed analysis (e.g., for single stocks).

3.4.2. Stock Data

In addition to our main research topic of measuring investor sentiment, we want to emphasize the economic relevance of this generated sentiment by attempting to forecast intraday returns. We do so by observing the development of derived investor sentiment

shortly before stock market closing on the previous day ($t-1$) and after the opening on the next day (t). The stock data we use to analyze the predictive power of investor sentiment are retrieved from *Thomson Reuters Eikon*. As depicted in Fig. 13, the timeframe with the highest average activity on StockTwits coincides with the opening hours of the American stock markets in contrast to the European and Asian markets. This observation most likely derives from the fact that according to Alexa around 54% of all visitors to the platform StockTwits originate from the United States (49.2%) and Canada (5.2%) as of April 2022.

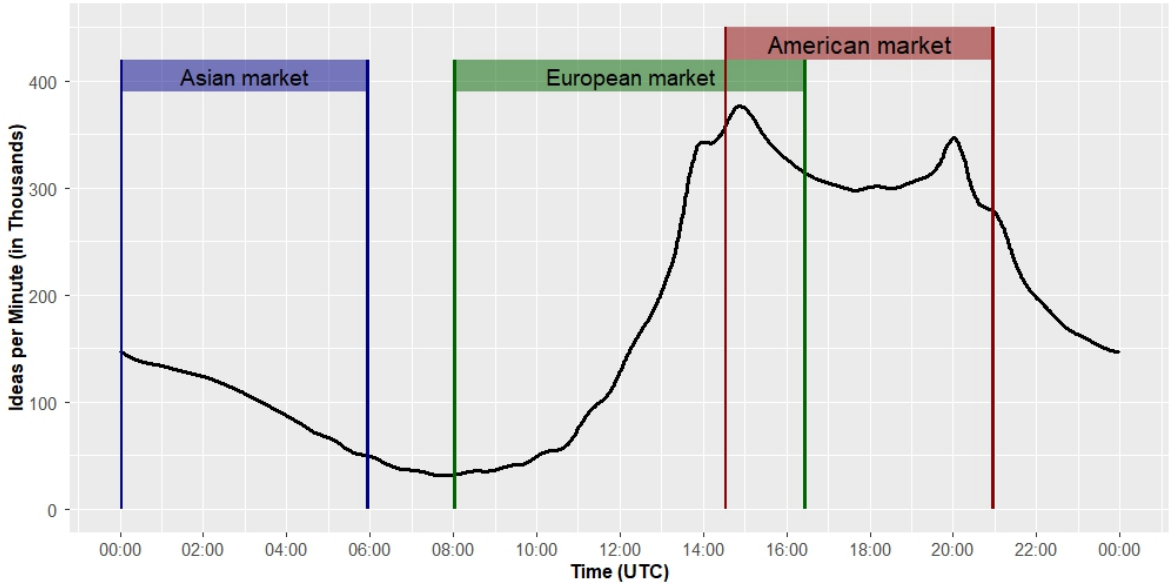


Fig. 13 Creation Time of Shared Ideas on StockTwits

As we expect most conversation to be held about topics concerning the US stock market and we find, in accordance to Cookson and Niessner (2020), the platform's users to have an affinity for technology companies, we obtain besides the S&P 500 data for the NASDAQ 100. The NASDAQ 100 appears to be suitable to appropriately represent the North American financial market in general but is also more focused on technology stocks, thereby addressing user affinity. The examined time period corresponds to the time interval selected for our StockTwits data spanning from 01/2012 to 12/2020 ($T = 2263$). Due to their statistically desirable characteristics, we use logarithmic returns as a steady measure of performance. We compute the intraday return of the S&P 500 and NASDAQ 100 on a given trading day t ($Intraday_t$) as formula (20)

$$Intraday_t = ln\left(\frac{Closing_t}{Opening_t}\right) \tag{20}$$

depicts, where $Opening_t$ denotes the opening price on the given day t and $Closing_t$ the

closing price on the same day.

3.5. Methodology

3.5.1. Converting Text to Emotion Scores

As previously defined, our aim is to improve previous research results, which we seek to accomplish by using the NRC Word-Emotion Association Lexicon (also known as 'EmoLex') introduced by Mohammad and Turney (2013) as a multidimensional text classification approach. In contrast to the predominantly used sentiments 'positive' and 'negative' in related literature, Mohammad and Turney (2013) created EmoLex containing the basic emotions 'anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise' and 'trust' proposed by Plutchik (1984). Using the R package 'syuzhet' developed by Jockers (2015), we are able to access the collected word list from Mohammad and Turney (2013) containing 14,182 unigrams and 25,000 word senses. The word list consists of the most frequently used unigrams and bigrams measured by the Google n-gram corpus, which are part of the *Macquarie Thesaurus* dictionary of words from the WordNet Affection Lexicon, and at most word-sense pairs from the General Inquire, which have at least two or three senses. The authors split the classification task into independently solvable 'human intelligence tasks' (HITs), which are solved by users (so-called 'turkers') on the Amazon platform 'Mechanical Turk'. Thus, emotion scores can be extracted from the individual classification by turkers (Mohammad and Turney (2013)).

A further problem that often emerges while working with word lexicons to identify sentiment scores – regardless of whether positive-negative polarity or emotions are examined – is that word lexicons do not include all possible formats a word might take. For example, a matching algorithm would miss the word 'lovers' if only the root 'love' is part of the lexicon. With *stemming* and *lemmatization*, linguistics proposes two possible solutions for this issue. While stemming algorithms attempt to determine a word's root by detecting and removing suffixes ('lovers' to 'lover' and 'loves' to 'love'), lemmatization attempts to group inflected forms into a single group ('lovers' and 'loves' to 'love'). For our analysis, we use the lemmatization list (41,531 words) created by Mechura (2016), which we access via the R package 'textstem' from Rinker (2018).

Tab. 13 illustrates how three representative ideas had been edited before we matched them with EmoLex to extract their emotion scores. In addition to the lemmatization of the strings, we remove whitespaces, stopwords, hyperlinks, hash- (#) or cashtags ($) and punctuation.

Furthermore, Tab. 14 depicts the mean emotion scores within the dataset of classified ideas grouped by the classification of the users. Bearish ideas tend to be loaded with words associated with anger, disgust, fear and sadness, while bullish ideas tend to be loaded with

Idea		Emotion scores							
Origin	Edited	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Costco should report stronger than expected December comps, says Stifel Nicolaus - $COST - http://stks.co/1k8b	Costco report strong expect December comps say Stifel Nicolaus	0	1	0	0	0	0	1	1
What word would you use to describe your feelings about $USDCAD since 1.0655? hatred..anger.. boredom..frustration.. #forex #cad	What word use describe feeling since hatred anger boredom frustration forex cad	5	1	3	2	1	3	1	2
Finally $TZA I am in green. I am off to enjoy my weekend. Signing off early. Lot of stress and anxiety. Need a break. Good luck to all.	Finally I green I enjoy weekend sign early Lot stress anxiety Need break Good luck	1	5	1	1	5	1	4	4

Tab. 13 Conversion from Original Ideas to Edited Ideas and Resulting Emotion Scores

words associated with anticipation, joy, surprise and trust. Except for the emotions anticipation and surprise, which do not appear to be clearly assignable to one of the classifications, all results match intuition. Using a Welch two-sample t-test, we check whether the difference in means is different from zero. For all types of emotions, we can reject the null hypothesis that groups' mean scores do not differ with a significance level below 0.01%.

Emotion	Classification		Welch Two Sample t-test
	Bullish	Bearish	
Anger	0.2388	**0.3161**	-407.47 ***
Anticipation	**0.6055**	0.5236	310.52 ***
Disgust	0.1121	**0.2015**	-620.27 ***
Fear	0.2449	**0.3282**	-428.09 ***
Joy	**0.3745**	0.2975	400.79 ***
Sadness	0.1682	**0.2635**	-562.90 ***
Surprise	**0.2891**	0.2880	5.86 ***
Trust	**0.5599**	0.5049	205.64 ***

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$
Sample: Classified ideas between 01/2012 and 12/2020 ($N_{Bu} + N_{Be} = 75,414,994$)

Tab. 14 Mean Emotion Scores of Classified Ideas per Group ('Bullish'/'Bearish')

With the generated emotion scores, we train a machine learning algorithm with the aim of classifying the 69.87% unclassified ideas as 'bullish' or 'bearish' using their emotion scores. As our dataset is strongly unbalanced with many 'bullish' ideas and less 'bearish' ones we first balance it by randomly picking ideas from each group with the following sample size:

$$N_{Sample} = \frac{min(N_{Bu}, N_{Be})}{2} = 6,294,380 \tag{21}$$

Furthermore, we divide the sample ($N_{Bu,Sample} + N_{Be,Sample} = 2 * N_{Sample} = 12,588,760$) into a training and a test dataset with a proportion of 80 to 20. Subsequently, we divide the training dataset with the same proportion into two further datasets that the algorithm uses for training and validation. The model used contains three dense layers, the first two layers deliver 64 units using a ReLU activation function while the last layer delivers one unit using an sigmoid activation which is the probability that an ideas is classified as 'bullish'.[17] In this manner, we classify ideas with a probability above or equal to 0.5 as 'bullish' and below 0.5 as 'bearish' using the test dataset in a first step. Thus, we define the accuracy of our model as the percentage of its correct classification within the test dataset.

3.5.2. Benchmarks

Consequently, we need to choose qualified benchmarks to compare the accuracy results of ideas classified by EmoLex with the classification results of other dictionaries to evaluate the performance of our approach. Therefore, we conduct the same classification task as described in Section 3.5.1 with other dictionaries commonly used in the economic literature. As mentioned at the beginning of this work, our aim is to underline the need to create an emotion-based *and* economic-related dictionary. For this purpose, we separately analyze the benefits of both dictionary types, defining the two following hypotheses:

H 2 (H1): *The classification accuracy and economic relevance of emotion-based dictionaries are higher than the accuracy of positive-negative-based dictionaries in text with an economic background.*

As noted in the introduction, we expect that an emotion-based dictionary such as the EmoLex dictionary with its eight dimensions (emotions) is more suitable to capture the complexity of (everyday) language. The compared dictionaries need to be created for the same type of language because words differ in connotation across contexts. This aspect brings us to our second hypothesis (H3), in which we expect that field-specific dictionaries are more capable of classifying words correctly from a text originating in this specific field.

[17] Beforehand, we've also tried out linear regression and logit/probit models, which have already been outperformed by a simple neural network with three dense layers in terms of accuracy of classification. As the classification problem is not that complex, more complex networks only delivered small increases in accuracy and the use of them has been rejected by the authors with respect to proportionality.

H 3 (H2): *The classification accuracy and economic relevance of economic-related dictionaries are higher than the accuracy of non-economic-related dictionaries in text with an economic background.*

Tab. 15 presents an overview of the properties of the chosen benchmark dictionaries. To examine the first hypothesis (H2), we use the accuracy rates of a positive-negative dictionary that is not economically related. The simplest approach is the use of the positive and negative scores of EmoLex (PN_{EL}), which we will also check for their accuracy contribution but which we are also questioning with respect to their independence from EmoLex emotion scores. Hence, we implement the positive and negative scores from the Harvard General Inquirer (PN_{GI}), since they have been used in many economic studies since the beginning of textual analysis research. Eventhough, this dictionary is not economic related (i.a. Da et al. (2015), Engelberg et al. (2012) or Tetlock (2007)).

Dictionary		**Score type**		**Economic-related**	**No. of words**
Name	Symbol	Emotions	Pos.-Neg.		
EmoLex	EM_{EL}	X			14,154
	PN_{EL}		X		
H GI	PN_{GI}		X		3642
LM	PN_{LM}		X	X	2709
Henry	PN_{HE}		X	X	190

Tab. 15 Properties of Commonly Used Dictionaries in Economic Literature

For the examination of the second hypothesis (H3), we need an economic-related positive-negative dictionary. Most prominent in this context is the work of Loughran and McDonald (2011), who created such a dictionary for evaluating the text tone of financial reports (PN_{LM}). Despite the broad use of this dictionary in economic research (i.a. Da et al. (2015), Chen et al. (2014), Kearney and Liu (2014), Engelberg et al. (2012), Dougal et al. (2012)) we also consider the dictionary by Henry (2008), as it is one of the first economic-related dictionaries that focuses on the influence of earnings press releases' tone on investor decision-making (PN_{HE}).[18]

3.5.3. Deriving Investor Sentiment

As the next and last step, we use the received classification to measure investor sentiment. This task is only needed for the investigation of the relevance of our main results – the accuracy of the different dictionaries. Intuitively, we expect times of high investor sentiment on the platform to be characterized by a high number of bullish ideas ($N_{Bullish}$) relative to

[18]Please note, that none of the dictionaries considered in this work have been developed to catch the tone of language used in social media as discussed in Section 3.2.

the number of bearish ideas ($N_{Bearish}$) and vice versa. In Fig. 12, we show that the number of ideas is increasing over time, and thus we need to correct the bullish-bearish spread with the number of classified ideas in total. Following Antweiler and Frank (2004), we define investor sentiment on a given day t derived from a specific dictionary $i = \{EM_{EL}, PN_{EM}, PN_{GI}, PN_{LM}, PN_{HE}\}$ as

$$Sentiment_{i,t} = \frac{N_{Bu,i,t} - N_{Be,i,t}}{N_{Bu,i,t} + N_{Be,i,t}} \tag{22}$$

where $N_{Bu,t}$ and $N_{Be,t}$ are the numbers of bullish and bearish ideas, respectively, on a given day t. The resulting measure is bounded in the interval $[-1, 1]$, where a value of 1 denotes the best possible investor sentiment and one of -1 the worst.

3.6. Results

3.6.1. Classification Accuracy

Before we use the derived measure for intraday return prediction, we compare the accuracy of all scored ideas within the analyzed dictionaries. Tab. 16 shows various descriptive statistics of the summed generated scores for the four dictionaries.

The mean number of scores per idea of EM_{EL}, which is 2.54, is nearly 40% higher as the next highest score of PN_{GI}, which takes a value of 1.83. As the emotion scores contain eight different emotions, it was predictable that EmoLex emotions (EM_{EL}) would exhibit the highest statistics per idea on average. The emotions 'anticipation' and 'trust' from EM_{EL} exhibit the highest scores on average. As the majority of ideas in our dataset have been classified as bullish, this result was to be expected as well. As already highlighted in Tab. 14, both of these emotions have a strong bullish connotation and occur especially in bullish ideas.

Nevertheless, the difference in scores between the economic-related and the non-economic-related positive-negative dictionaries attracts our attention. Both non-economic-related dictionaries PN_{EL} and PN_{GI} possess considerably higher scores than PN_{LM} and PN_{HE}, which are economically related. This finding hints at two possible conclusions. On the one hand, the mean score of PN_{HE} might suffer from its short amount of words (see Tab. 15) relative to PN_{LM}. On the other hand, as PN_{LM} and PN_{HE} possess a mainly economic background, it seems that the language used on social media platforms, in our case StockTwits, is distinct from language in economic texts as for example financial reports. This finding emphasizes that in addition to the claims 'emotional-based' and 'economic-related', a perfectly designed field-specific dictionary for social media text analysis should focus on the text type used on such platforms.

This conclusion is further strengthened when considering the median score of 0 for PN_{LM} and PN_{HE}, indicating that less than 50% of all ideas contain at least one word that can be

Dictionary		Mean	Median	Sd	Min	Max	Unique expressions
EM_{EL}	**All**	**2.54**	**1**		**0**	**112**	**638,712**
	anger	0.23	0	0.54	0	24	19
	anticipation	0.56	0	0.88	0	19	20
	disgust	0.13	0	0.39	0	21	19
	fear	0.25	0	0.56	0	25	22
	joy	0.33	0	0.64	0	16	17
	sadness	0.19	0	0.48	0	24	20
	surprise	0.27	0	0.56	0	12	13
	trust	0.57	0	0.90	0	21	21
PN_{EL}	**All**	**1.23**	**1**		**0**	**59**	**484**
	positive	0.77	0	1.15	0	27	28
	negative	0.46	0	0.82	0	48	31
PN_{GI}	**All**	**1.83**	**1**		**0**	**245**	**1,062**
	positive	1.01	1	1.53	0	245	75
	negative	0.82	0	1.25	0	198	74
PN_{LM}	**All**	**0.36**	**0**		**0**	**107**	**265**
	positive	0.17	0	0.49	0	107	40
	negative	0.19	0	0.50	0	62	33
PN_{HE}	**All**	**0.22**	**0**		**0**	**77**	**208**
	positive	0.15	0	0.45	0	47	33
	negative	0.07	0	0.28	0	76	27

Sample: 01/2012 - 12/2020 (N = 250, 321, 511)

Tab. 16 Descriptive Statistics of Generated Scores from Textual Analysis

Dictionary	All Ideas (in %)			Scored Ideas (in %)			
	All	Bullish	Bearish	All	Bullish	Bearish	Loss
EM_{EL}	55.73	53.88	61.12	58.17	56.63	60.64	36.51
PN_{EL}	55.37	54.05	57.94	57.69	57.77	57.62	40.26
PN_{GI}	54.66	53.89	55.79	56.14	57.08	55.41	29.53
PN_{LM}	53.32	51.97	60.30	60.36	60.61	60.11	73.13
PN_{HE}	52.38	60.97	51.33	57.74	62.26	55.66	82.12

Sample: 01/2012 - 12/2020 (N_{Bu} + N_{Be} = 75, 414, 994)
Results differ max. ±0.05% using only for the test datset for validation.

Tab. 17 Accuracy of Scoring by Different Dictionaries

classified as positive or negative. Another noteworthy feature in Tab. 16 is the number of unique expressions of word scores found in the data when classifying ideas with different dictionaries. It becomes apparent that due to its higher dimensionality and mean score, the most unique combinations of scores by far occur when using EM_{EL} (638,712), confirming the proposed ability to capture the underlying complexity of (social media) text in greater detail than two-dimensional approaches do.

To further compare the different dictionaries and their performance, we analyze their respective classification accuracies for the full classified dataset. Tab. 17 illustrates that when including all 75,414,994 ideas previously classified as 'bullish' or 'bearish' by the users, EM_{EL} scores highest with an accuracy of 55.73%, followed by both non-economic-related dictionaries PN_{EL} and PN_{GI} with accuracies of 55.37% and 54.66%, respectively. Again, PN_{LM} and PN_{HE} surprisingly obtain the lowest accuracies, with 53.32% and 52.38% at first glance. As explained above, the relatively low accuracy rate of PN_{LM} and PN_{HE} derives from the low classification rate of the words contained in the analyzed ideas.

This is why we further compute the accuracy rate for all ideas that contain at least one scored word in each dictionary. When only considering ideas containing at least one score, all dictionaries experience a growth in accuracy. In particular, the accuracy rates of PN_{LM} and PN_{HE} disproportionately increase to 60.36% and 57.74%, exceeding all other growth rates. Nevertheless, this increase comes with the loss of approximately 73.13% and 82.12%, respectively, of all potentially available ideas. Apparently, a tradeoff between accuracy and data loss exists and needs to be considered when using either dictionary. Therefore, exclusively observing the accuracy rate might not be adequate. Apart from the two economic-related dictionaries PN_{LM} and PN_{HE}, EM_{EL} provides the highest accuracy rate (58.17%) with the second lowest percentage of ideas lost (36.51%).

On this basis, Fig. 14 shows the relationship between the share of excluded data and the accuracy of our classification. Gradually, we exclude data points whose prediction values from the trained algorithm are most uncertain by moving simultaneously from 0.5 to 0 (bearish predicts) and 0.5 to 1 (bullish predicts). In general, all dictionaries profit in accuracy from this operation. Nevertheless, some dictionaries profit more than others. At a data loss level of 95%, the accuracy of the economic-related positive-negative dictionaries reaches around 67% for PN_{LM} or nearly 70% for PN_{HE}, while the non-economic-related positive-negative dictionaries only reach an accuracy of approximately 62% (PN_{GI}) and 66% (PN_{EM}). Furthermore, the emotion-based and non-economic-related dictionary EmoLex (EM_{EL}) performs even stronger than the dictionary by Henry (2008) with around 73% accuracy at a degree of data excluded slightly above 95%. The last mentioned EmoLex dominates all other dictionaries without exception at every degree of excluded data. It is clear that this dominance grows with the

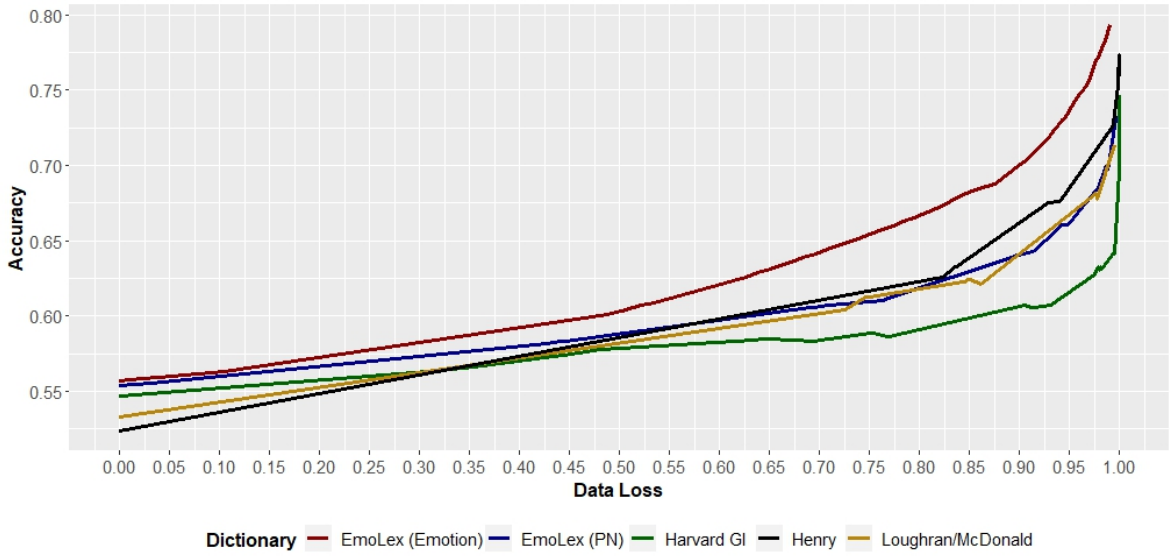


Fig. 14 Relationship Between the Data Loss and Classification Accuracy of Different Dictionaries

number of excluded, most uncertain predicted ideas.

Plotting the histograms of the resulting prediction values for each dictionary suggests the abovementioned observations. As Fig. 15 illustrates, all histograms show a high density around a value of 0.5, which is mainly caused by ideas without any score. Consequently, both economic-related dictionaries PN_{LM} and PN_{HE} with the highest rate of unscored ideas show the highest density at approximately 0.5, downgrading their accuracy when using the full dataset. Nevertheless, the prediction values of these dictionaries and EmoLex possess a considerably higher kurtosis than the positive-negative dictionaries (PN_{EM} and PN_{GI}), illustrating their power to classify economic text as 'bullish' or 'bearish' in a more certain way. Observing the tails of the prediction distributions by calculating the 2.5% and 97.5% quantiles shows that EM_{EL} and PN_{HE} make the most safe predictions, leading to the highest accuracy rates when more than 95% of the data are excluded.

Overall, the data show that the emotion-based dictionary performs slightly better than positive-negative dictionaries using full data. Nevertheless, the feature of multidimensionality leads to safer predictions in the tails of the prediction distribution. The same is true for the economic-related dictionary because of the use of accurate connotations. Subsetting the data into the three in Tab. 12 mentioned self-classified trading experience groups (Novice, Intermediate, Professional) and observing the kurtosis of the extracted prediction values within each group gives a hint that the type of the observed text is important for accurate predictions, as well. This finding strengthens prior assumptions found in related literature by e.g. Giannini et al. (2018).

Tab. 18 shows that despite the kurtosis of all dictionaries differ, most dictionaries also

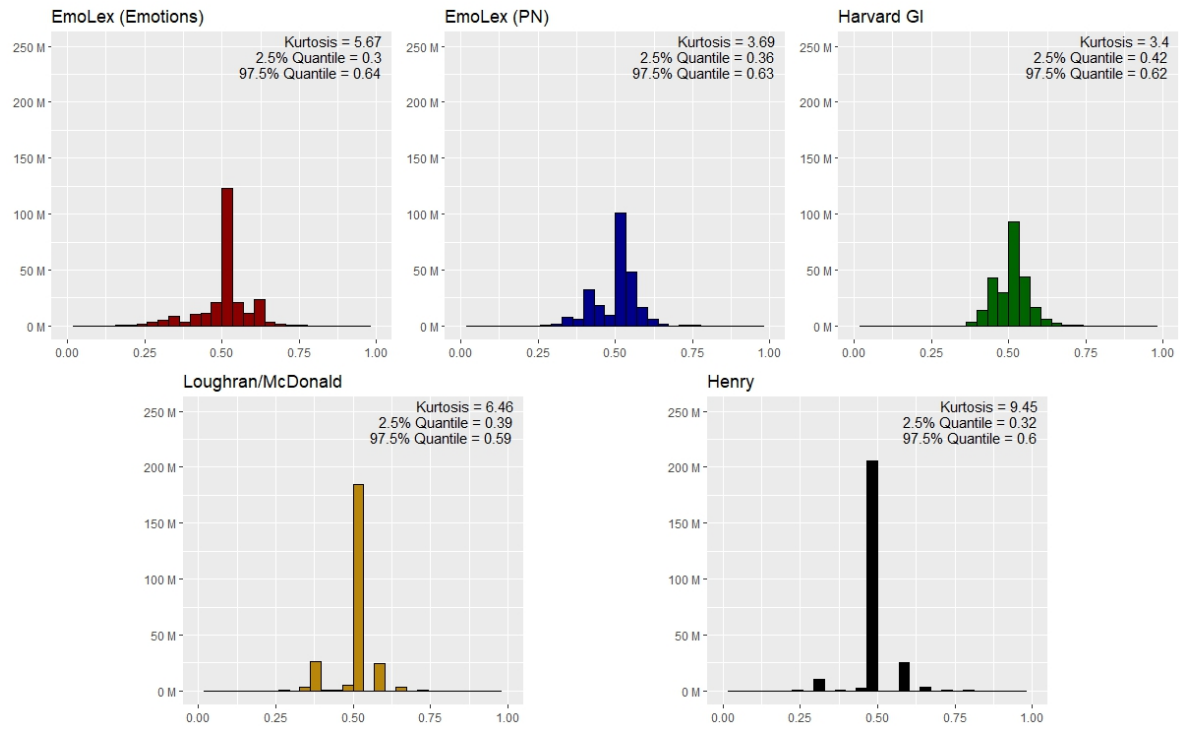


Fig. 15 Histograms of Prediction Values of Different Dictionaries ($N = 250,321,511$)

show the same positive tendency in kurtosis moving to professional text. Assuming that the language used between all groups differs, the need for wordlists adressing the language of other author groups is illustrated as the economic relevance of their posts can not be ignored.

	EM_{EL}	PN_{EL}	PN_{GI}	PN_{LM}	PN_{HE}
All	**5.67**	**3.69**	**3.40**	**6.46**	**9.45**
Novice	5.55	3.48	3.07	5.56	9.34
Intermediate	5.70	3.53	3.13	5.54	9.06
Professional	6.18	3.56	3.09	5.97	9.23

Tab. 18 Kurtosis of Prediction Values of Different Dictionaries

Consequently, it remains to illustrate the hypothesized economic relevance of our findings. For this purpose we use the generated scores from all dictionaries to measure investor sentiment and compare their explanatory power for intraday stock returns in the following.

3.6.2. Economic Relevance

As according to Fig. 13 most ideas are published around the opening of the US stock market, and we attempt to use this amount of information to predict the intraday return of a specific trading day t by using the shift in sentiment between one hour before market opening

of that trading day and the last market hour of the previous trading day $t-1$.[19] In detail, we therefore calculate investor sentiment ($Sentiment_{i,t,m}$) for each dictionary i on trading day t and with a time indicator m subsetting the classified ideas used.

$$Sentiment_{i,t,m} = \begin{cases} \text{m=1 for 01.30 p.m. to 02.30 a.m. (UTC)} \\ \text{m=2 for 08.00 p.m. to 09.00 p.m. (UTC)} \end{cases} \tag{23}$$

Hence, we define the shift in investor sentiment ($\triangle Sentiment_{i,t}$) derived by dictionary i on trading day t as:

$$\triangle Sentiment_{i,t} = Sentiment_{i,t,1} - Sentiment_{i,t-1,2} \tag{24}$$

We use this as an explanatory variable explaining the intraday return of the S&P 500 and the NASDAQ 100 on trading day t ($Intraday_t$), as defined in 20. As many studies offer the critique that identified relationships between investor sentiment and stock returns might simply be driven by autocorrelation, we introduce the intraday return on the previous trading day $t-1$ as a second explanatory variable to control for this effect in our regression model (i.e., Xiong et al. (2019)). Thus, we estimate the following linear model.

$$Intraday_t = \beta_0 + \beta_1 * Intraday_{t-1} + \beta_i * \triangle Sentiment_{i,t} + \epsilon_t \tag{25}$$

Tab. 19 illustrates the results of the regression for both indices for each dictionary. We calculate standardized coefficients ($\tilde{\beta}_1$ and $\tilde{\beta}_i$) because all derived investor sentiment measures have different statistical properties and the size of coefficients between the estimated models would not be comparable.

Starting with the full dataset, differences in sentiment for all dictionaries except Harvard GI (PN_{GI}) have a significantly positive influence on the intraday return of the S&P 500 while the influence on NASDAQ 100 returns is for all dictionaries lower and no longer significant for the Henry dictionary (PN_{HE}). Surprisingly, observing the standardized coefficients and the goodness of fit measured by adjusted R^2 shows that (1) the sentiments derived from the EmoLex dictionary by Mohammad and Turney (2013), EM_{EL}, possess the highest influence on intraday returns with standardized coefficients of 0.0641 for the NASDAQ 100, but (2) the influence for the economic-related dictionaries predicting S&P 500 return is even higher with

[19] By doing so, we assume ideas to be published shortly before stock market closing on day $t-1$ mostly contain a summary of that day, while users/investors focus primarily on upcoming events in the hour before market opening of the next trading day t.

		$\tilde{\beta_1}$	$\tilde{\beta_i}$	R^2	$Adj.R^2$
S&P 500	EM_{EL}	-0.1031*	0.0646**	0.0151	0.0142
		(-2.3398)	(3.1252)		
	PN_{EL}	-0.1058*	0.0573**	0.0142	0.0134
		(-2.4094)	(2.6168)		
	PN_{GI}	-0.1046**	0.0375	0.0123	0.0115
		(-2.3812)	(1.8901)		
	PN_{LM}	-0.1042*	0.0656**	0.0152	0.0144
		(-2.3839)	(3.1109)		
	PN_{HE}	-0.1069*	0.0833***	0.0179	0.0170
		(-2.4458)	(3.4206)		
NASDAQ 100	EM_{EL}	-0.1523***	0.0641**	0.0273	0.0264
		(-4.5966)	(3.2067)		
	PN_{EL}	-0.1531***	0.0533*	0.0260	0.0252
		(-4.6213)	(2.3474)		
	PN_{GI}	-0.1522***	0.0214	0.0236	0.0228
		(-4.594)	(1.1112)		
	PN_{LM}	-0.1518***	0.0462*	0.0253	0.0245
		(-4.5953)	(2.4250)		
	PN_{HE}	-0.1528***	0.0188	0.0235	0.0227
		(-4.6135)	(0.8673)		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (We use robust White standard errors.)
Regressions estimate formula 25 with the sample: 01/2012 - 12/2020 ($T = 2263$)

Tab. 19 Intraday Return Predictability Using Different Sentiment Measures

(highly) significant coefficients of 0.0833 and 0.0656.

As these results are predominantly in line with the accuracy results in Tab. 17, we also calculate the same regressions while gradually excluding ideas with the most uncertain prediction values. Economically, such indifferent ideas could be interpreted as a 'hold' signal by the publisher. The results of this operation on the standardized coefficients ($\tilde{\beta_i}$) can be observed in Fig. 16.

First, all dictionaries show a significant positive influence on intraday returns at data loss degrees above 55% despite of the Harvard GI dictionary. Nevertheless, investor sentiment derived from EmoLex or Henry (2008) (and at a high degree of uncertain predictions excluded also from Loughran and McDonald (2011)) dominates the non-economic-related positive-negative dictionaries in predictive power, with standardized coefficients reaching maximum values of approximately 0.16 and an explained variance of up to 3% measured by adjusted R^2.[20]

Regarding our stated hypotheses (H2 and H3) these findings can only be validated reliably by testing for differences between the coefficients. Therefore, following Clogg et al. (1995) and Paternoster et al. (1998) we calculate the Z-scores using the formula:

[20]More detailed regression results can be found in the appendix in Tab. 22.

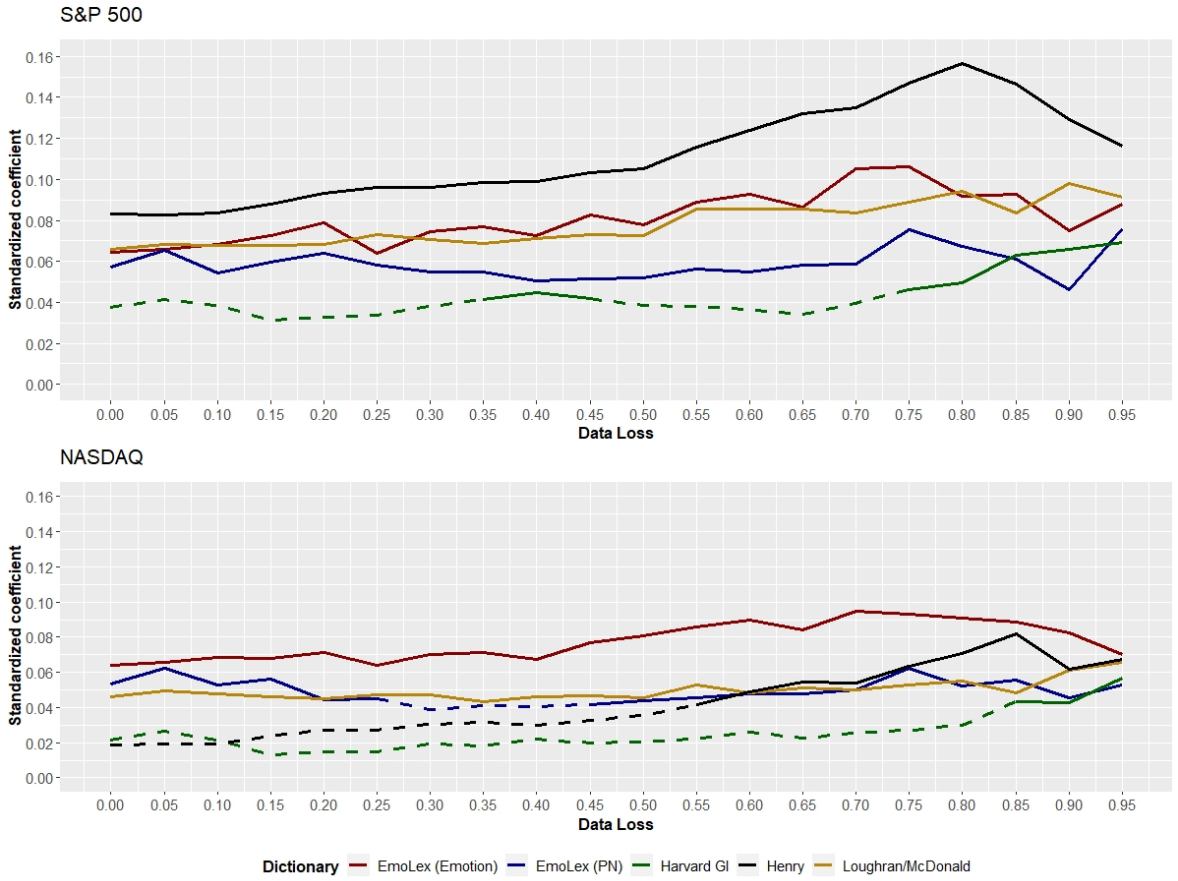


Fig. 16 Development of the Standardized Coefficients (Dashed if $p > 0.05$)

$$Z = \frac{\tilde{\beta}_i - \tilde{\beta}_j}{\sqrt{\sigma_{\tilde{\beta}_i} + \sigma_{\tilde{\beta}_j}}} \tag{26}$$

where $i, j = \{EM_{EL}, PN_{EM}, PN_{GI}, PN_{LM}, PN_{HE}\}$ and $i \neq j$. Fig. 17 and Fig. 18 show the development of the Z-scores testing the relevant differences in coefficients for both hypotheses.

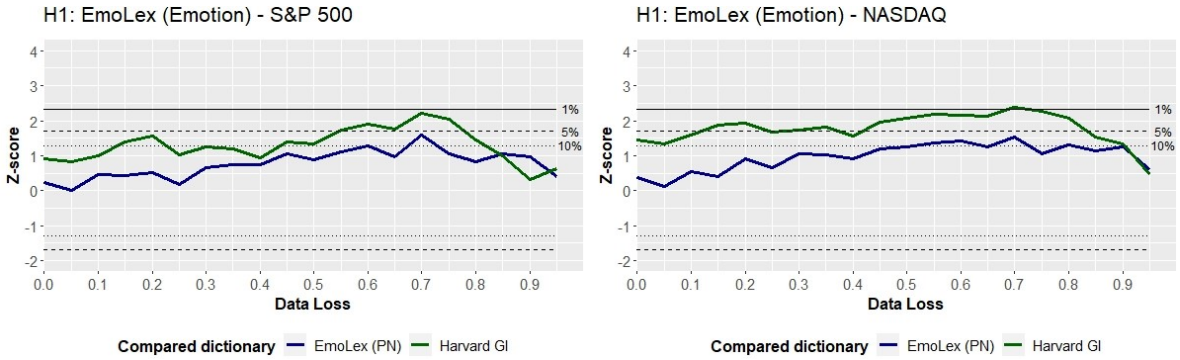


Fig. 17 Development of the Z-Scores Proving for H2

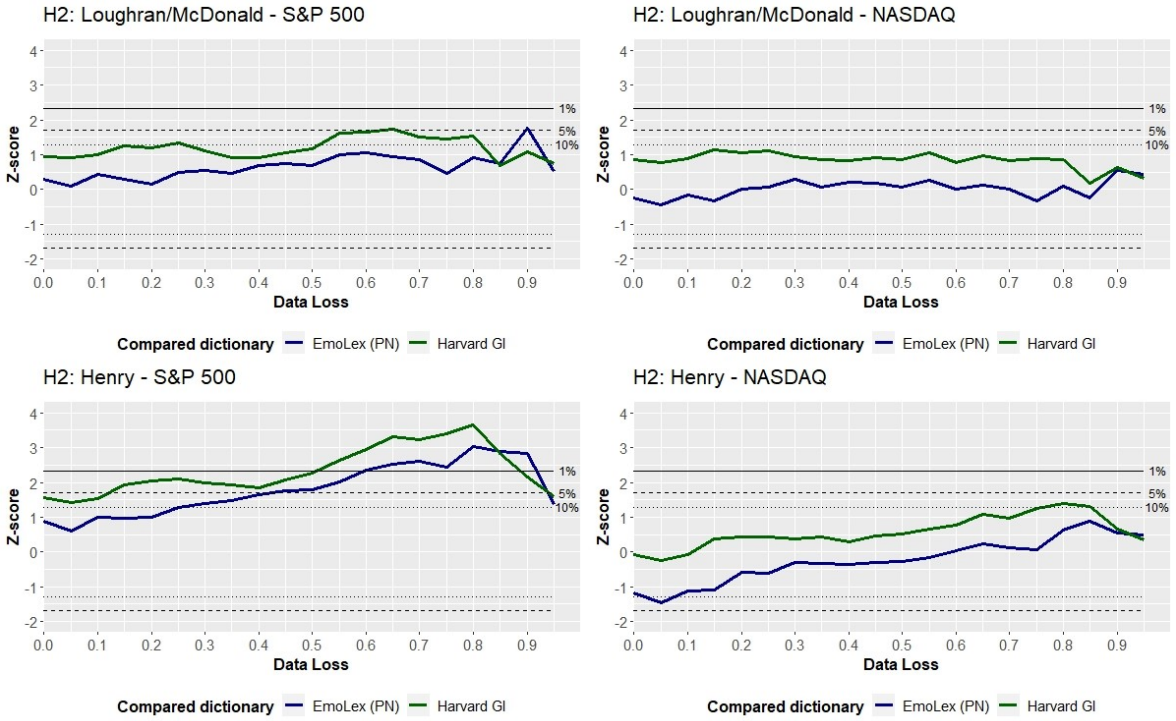


Fig. 18 Development of the Z-Scores Proving for H3

Proving H2 we observe the difference between EM_{EL} and PN_{EL}/PN_{GI} in Fig. 17. For the S&P 500 as well as the NASDAQ 100 differences for the Harvard GI are positive and

highly significant, especially for higher degrees of excluded data. Differences with the related positive-negative version of EmoLex are positive but only become significant for higher levels of data excluded.

Proving H3 we observe more mixed results while testing for differences between PN_{LM}/PN_{HE} and PN_{EL}/PN_{GI}. While for the S&P 500 both economic-related dictionaries overperform the Harvard GI (and Henry also the positive-negative EmoLex version), the Z-scores for the NASDAQ 100 are mainly insignificant around zero.

Overall, also bearing the accuracy results from Section 3.6.1 in mind, our first hypothesis (H2) cannot be rejected, as EmoLex emotion scores reach a higher accuracy in classifying 'bullish' or 'bearish' signals. Furthermore, the shift in investor sentiment has greater predictive power than the shift in investor sentiment derived from the non-economic-related positive-negative dictionaries, PN_{EM} and PN_{GI}, at all levels of excluded data.

In line with prior research (for example by Renault (2017)) the same holds true for the second hypothesis (H3) as the accuracy and economic relevance of investor sentiment derived with the help of economic-related dictionaries are higher than those of dictionaries without economic relations - even though Z-scores only indicated a significant difference between the coefficients for the S&P 500. If we consider this hypothesis in a more precise way, we also find that within economic-related dictionaries, accuracy and economic relevance can differ. Hence, the results of the dictionary created by Henry (2008), which is based on (financial) earnings press releases, are stronger than those of the dictionary created by Loughran and McDonald (2011), which originated in an accounting background, for our field-specific application.

Further, our prior analysis indicates that in addition to the economic word connotation and the emotional scoring as a third factor, the type of text plays an important role in deriving investor sentiment from text. By dividing the published ideas into the three self-classified trading groups – 'novice', 'intermediate' and 'professional' – and repeating the regression defined in formula (25) for each of those groups, as is clear from Tab. 20, the predictive power of most dictionaries reported in the full regression (Tab. 19) highly differs between user types, this is especially true for the two economic-related dictionaries.

We assume the language used by user who classify themselves as 'professional' to be that specific that it only fits well for the financial earnings related dictionary PN_{HE}. For other dictionaries the ideas differ too much from their origin texts which vice versa perform better for text written by users who classify themselves as 'novice' or 'intermediate' [21]. Despite the disadvantage of not being economic-related, shifts in investor sentiment from EmoLex

[21] For different degrees of data excluded no further findings could be made. Nevertheless, all relevant figures are reported in the appendix in Fig. 19.

		SP500		NASDAQ	
		$\tilde{\beta}_i$	$Adj.R^2$	$\tilde{\beta}_i$	$Adj.R^2$
Novice	EM_{EL}	0.0681***	0.0147	0.0428*	0.0250
		(3.6878)		(2.3803)	
	PN_{EL}	0.0646***	0.1420	0.0528**	0.0260
		(3.3921)		(2.8138)	
	PN_{GI}	0.0763***	0.0159	0.0479**	0.0246
		(4.0782)		(2.6742)	
	PN_{LM}	0.0764***	0.0159	0.0518**	0.0250
		(3.8073)		(2.6368)	
	PN_{HE}	-0.0302	0.0110	0.0074	0.0224
		(1.5879)		(0.4191)	
Intermediate	EM_{EL}	0.0488*	0.0125	0.0346	0.0235
		(2.5574)		(1.8814)	
	PN_{EL}	0.0510**	0.0127	0.0337	0.0235
		(2.6280)		(1.6675)	
	PN_{GI}	-0.0018	0.0101	-0.0128	0.0225
		(-0.0949)		(-0.7026)	
	PN_{LM}	-0.0437*	0.0120	0.0290	0.0232
		(2.2568)		(1.5514)	
	PN_{HE}	0.0709***	0.0151	0.0600**	0.0259
		(3.7072)		(3.1817)	
Professional	EM_{EL}	0.0476*	0.0123	0.0599**	0.0259
		(2.423)		(3.0924)	
	PN_{EL}	0.0463*	0.0122	0.0662**	0.0267
		(2.1606)		(3.2627)	
	PN_{GI}	0.0431*	0.0119	0.0407*	0.0240
		(2.2132)		(2.1496)	
	PN_{LM}	0.0364	0.0114	0.0185	0.0227
		(1.4446)		(0.8339)	
	PN_{HE}	0.0961***	0.0193	0.0255	0.0230
		(4.3384)		(1.2016)	
No Group	EM_{EL}	0.0409*	0.1170	0.0410*	0.0240
		(2.0418)		(2.1417)	
	PN_{EL}	0.0157	0.0103	0.0068	0.0224
		(0.7327)		(0.3277)	
	PN_{GI}	0.0186	0.0104	-0.0039	0.0223
		(0.8642)		(-0.2038)	
	PN_{LM}	0.0359	0.0114	0.0271	0.0230
		(1.6339)		(1.4423)	
	PN_{HE}	0.0349	0.0113	-0.0073	0.0224
		(1.7575)		(-0.4032)	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (We use robust White standard errors.)
Regressions estimate formula 25 with the sample: 01/2012 - 12/2020 (T = 2263)
Subgroup ideas: Novice 16,172,895 (6.46%) / Intermediate 42,480,800 (16.97%) /
Professional 31,464,736 (12.57%) / No Group 160,203,326 (64.00%)

Tab. 20 Intraday Return Predictability Using Different Sentiment Measures by Trader Group

emotion scores exhibit their slightly weaker predictive power for intraday returns by each of the three (or rather four) groups, which makes it more applicable for analyzing text from which the trader group of the publisher is unknown, as is common in most social media text data.

3.7. Conclusion

The field of social media sentiment analysis is fast moving due to the rapid growth of data published on platforms such as Twitter, Facebook or, in our case, StockTwits. This makes it necessary to regularly reevaluate preexisting research, adjust former methodologies and propose new methodologies.

The first part of this paper addressed the economic background of sentiment analysis and why it might be desirable to consider sentiment analysis when attempting to predict stock market movements. We do so by reviewing the preceding related literature. In the following, we present our obtained dataset, which we generated from the microblogging platform StockTwits. Furthermore, we provide detailed insight into the way we extract eight emotions from ideas published on StockTwits by using the NRC Word-Emotion Association Lexicon (EmoLex), enabling us to correlate these underlying emotions with an individual's self-revealed bullish or bearish sentiment by using machine learning algorithms.

Consequently, this allows us to classify further ideas that have not been classified into bullish or bearish sentiment categories, thereby enriching our database. We make use of this extended database, which comprises approximately 250 million classified ideas, to correlate our sentiment findings with US stock market performance. We find that investor sentiment classified by emotion scores can be used to predict stock market movements weakly but more accurately than positive-negative scores derived from non-economic-related dictionaries. Although we are able to classify more ideas than the analyzed two-dimensional dictionaries, many ideas still cannot be scored by our approach due to missing occurrences of ideas' words in dictionaries' wordlists. This weakness is in line with prior research results and illustrates the need for further improvement.

In detail, our results define three main factors that determine the success of deriving investor sentiment with the help of textual sentiment in an economic context: multidimensional scoring (for example emotions), economic word connotation and type of text. By using supervised machine learning algorithms without taking common benchmark dictionaries into account, many researchers address those three factors with mostly noteworthy results. Nevertheless, for example, Renault (2017) comes to the same conclusion as Kearney and Liu (2014) and shows that field-specific dictionaries are more applicable than rough benchmark dictionaries as well as machine learning algorithms. Since the classification of a text by its publisher for training purposes, as in StockTwits data, is a rare feature of text data and

self-classification of text by researchers often leads to misclassification, there remains a need to create multiple dictionaries addressing those three factors.

In accordance to our results it can be expected that more advanced dictionaries (Fig. 11, A3) or NLP Transformers might also profit from the factors outlined above - especially multidimensional scoring. These considerations therefore give a reasoning for the recent emergence of field-specific and emotion-based NLP transformers (as for example specifications of RoBERTa/DistilRoBERTa)[22].

[22] An overview of various specification can be found on the model hub for NLP: Hugging Face.

3.8. Appendix

		Data Loss				
		10%	30%	50%	70%	90%
EM_{EL}	$\tilde{\beta}_1$	-0.1033*	-0.1025*	-0.1029*	-0.1032*	-0.1045*
		(-2.3451)	(-2.3353)	(-2.3465)	(-2.3627)	(-2.3862)
	$\tilde{\beta}_i$	0.0680**	0.0745***	0.0798***	0.1053***	0.0751***
		(3.2626)	(3.7124)	(3.7523)	(5.1316)	(3.8235)
	R^2	0.0156	0.0165	0.0170	0.0220	0.0166
	$Adj.R^2$	0.0147	0.0156	0.0161	0.0212	0.0157
PN_{EL}	$\tilde{\beta}_1$	-0.1057*	-0.1059*	-0.1059*	-0.1050*	-0.1045*
		(-2.4058)	(-2.4228)	(-2.4170)	(-2.4018)	(-2.3838)
	$\tilde{\beta}_i$	0.0544*	0.0546*	0.0521**	0.0585**	0.0462*
		(2.5038)	(2.5574)	(2.6970)	(3.1564)	(2.4923)
	R^2	0.0139	0.0139	0.0137	0.0144	0.0131
	$Adj.R^2$	0.0130	0.0131	0.0128	0.0135	0.0122
PN_{GI}	$\tilde{\beta}_1$	-0.1047*	-0.1046*	-0.1044*	-0.1049*	-0.1052*
		(-2.3848)	(-2.3836)	(-2.3782)	(-2.3918)	(-2.3998)
	$\tilde{\beta}_i$	0.0384	0.0378*	0.0385*	0.0396*	0.0658**
		(1.9375)	(1.9798)	(2.0029)	(2.1778)	(3.5362)
	R^2	0.0124	0.0124	0.0124	0.0125	0.0152
	$Adj.R^2$	0.0115	0.0115	0.0116	0.0116	0.0144
PN_{LM}	$\tilde{\beta}_1$	-0.1043*	-0.1042*	-0.1043*	-0.1042*	-0.1042*
		(-2.3870)	(-2.3878)	(-2.3883)	(-2.3910)	(-2.3812)
	$\tilde{\beta}_i$	0.0676**	0.0707**	0.0726**	0.0838***	0.0979***
		(3.1862)	(3.2686)	(3.4762)	(3.9877)	(4.9409)
	R^2	0.0155	0.0159	0.0162	0.0180	0.0205
	$Adj.R^2$	0.0146	0.0151	0.0153	0.0171	0.0197
PN_{HE}	$\tilde{\beta}_1$	-0.1070*	-0.1079*	-0.1079*	-0.1091*	-0.1073*
		(-2.4458)	(-2.4696)	(-2.4721)	(-2.5154)	(-2.4513)
	$\tilde{\beta}_i$	0.0836***	0.0961***	0.1052***	0.1349***	0.1294***
		(3.4497)	(4.2358)	(4.9281)	(6.3827)	(6.3361)
	R^2	0.0179	0.0202	0.0220	0.0291	0.0277
	$Adj.R^2$	0.0171	0.0193	0.0211	0.0283	0.02687

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (We use robust White standard errors.)
Regressions estimate formula 25 with the sample: 01/2012 - 12/2020 (T = 2263)

Tab. 21 S&P 500 Intraday Return Predictability Using Different Sentiment Measures at Different Degrees of Excluded Uncertain Predictions

		Data Loss				
		10%	30%	50%	70%	90%
EM_{EL}	$\tilde{\beta}_1$	-0.1525***	-0.1522***	-0.1524***	-0.1529***	-0.1527***
		(-4.6052)	(-4.5947)	(-4.6054)	(-4.6259)	(-4.6232)
	$\tilde{\beta}_i$	0.0683**	0.0700***	0.0807***	0.0948***	0.0823***
		(3.3783)	(3.6775)	(3.9760)	(4.7052)	(4.3509)
	R^2	0.0279	0.0281	0.0297	0.0322	0.0300
	$Adj.R^2$	0.0270	0.0272	0.0288	0.0313	0.0291
PN_{EL}	$\tilde{\beta}_1$	-0.1531***	-0.1530***	-0.1532***	-0.1527***	-0.1526***
		(-4.6213)	(-4.6176)	(-4.6236)	(-4.6109)	(-4.6113)
	$\tilde{\beta}_i$	0.0526*	0.0387	0.0438*	0.0499**	0.0454*
		(2.3474)	(1.8911)	(2.3668)	(2.7391)	(2.5473)
	R^2	0.0259	0.0247	0.0251	0.0257	0.0252
	$Adj.R^2$	0.0251	0.0238	0.0242	0.0248	0.0244
PN_{GI}	$\tilde{\beta}_1$	-0.1522***	-0.1523***	-0.1522***	-0.1541***	-0.1526***
		(-4.5947)	(-4.5999)	(-4.5966)	(-4.6035)	(-4.6075)
	$\tilde{\beta}_i$	0.0216	0.0191	0.0203	0.0253	0.0428*
		(1.1067)	(1.0505)	(1.221)	(1.4427)	(2.3948)
	R^2	0.0236	0.0235	0.0236	0.0238	0.0250
	$Adj.R^2$	0.0228	0.0227	0.0227	0.0230	0.0241
PN_{LM}	$\tilde{\beta}_1$	-0.1519***	-0.1520***	-0.1520***	-0.1517***	-0.1514***
		(-4.5971)	(-4.6048)	(-4.6062)	(-4.5944)	(-4.5763)
	$\tilde{\beta}_i$	0.0477*	0.0471*	0.0455*	0.0498*	0.0614**
		(2.4708)	(2.4640)	(2.4888)	(2.5340)	(3.1260)
	R^2	0.0255	0.0254	0.0252	0.0257	0.0267
	$Adj.R^2$	0.0246	0.0245	0.0244	0.0248	0.0261
PN_{HE}	$\tilde{\beta}_1$	-0.1529***	-0.1532***	-0.1533***	-0.1541***	-0.1531***
		(-4.6145)	(-4.6260)	(-4.6282)	(-4.6487)	(-4.6120)
	$\tilde{\beta}_i$	0.0193	0.0302	0.0356	0.0538**	0.0618**
		(0.8788)	(1.3968)	(1.7201)	(2.6888)	(3.042)
	R^2	0.0236	0.0241	0.0244	0.0261	0.0270
	$Adj.R^2$	0.0227	0.0232	0.0236	0.0252	0.0261

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (We use robust White standard errors.)
Regressions estimate formula 25 with the sample: 01/2012 - 12/2020 (T = 2263)

Tab. 22 NASDAQ 100 Intraday Return Predictability Using Different Sentiment Measures at Different Degrees of Excluded Uncertain Predictions

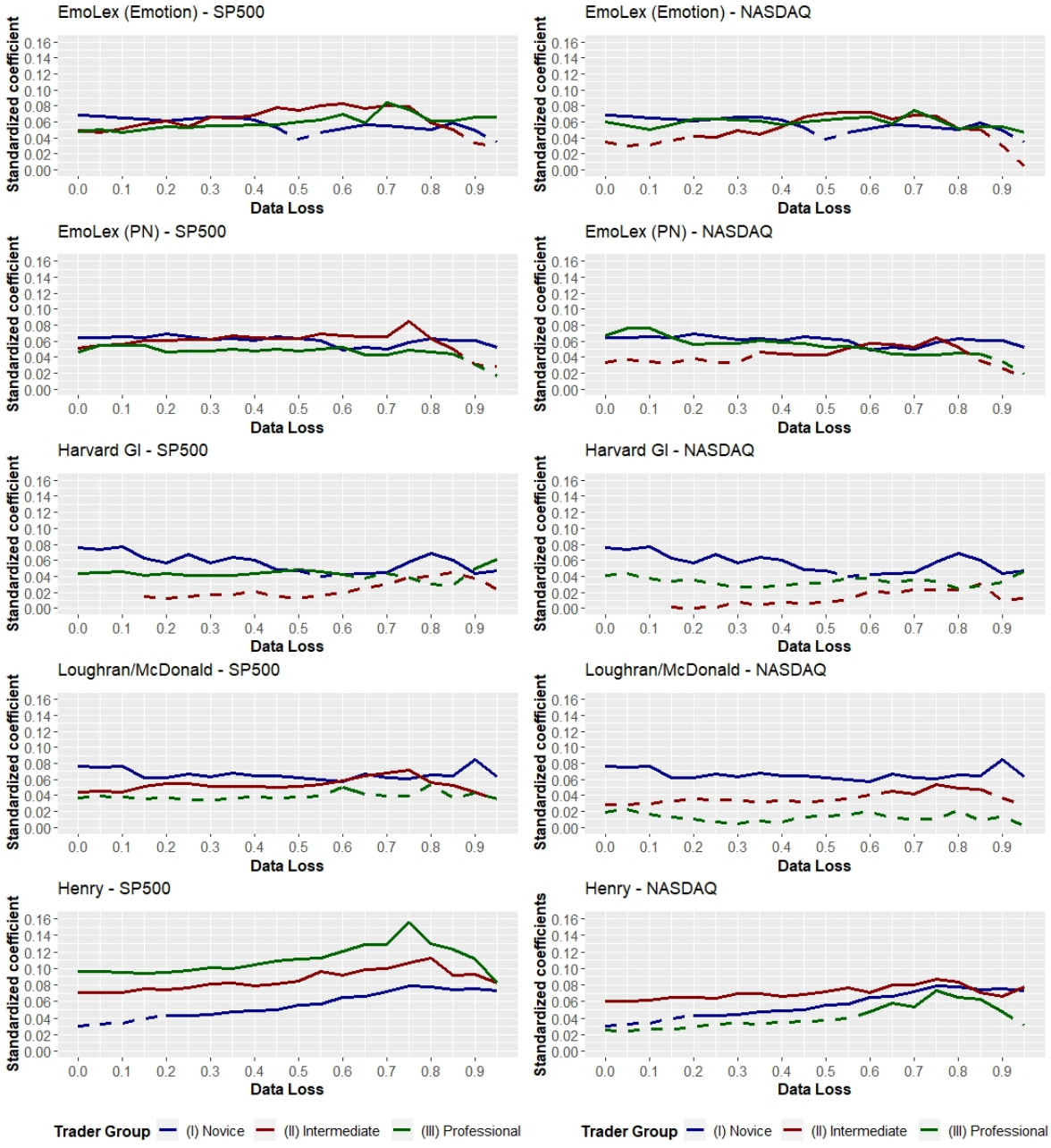


Fig. 19 Development of the Standardized Coefficients (Dashed if $p > 0.05$)

3.9. Declaration of (Co-)Authors and Record of Accomplishments

Title:	Measuring investor sentiment from Social Media Data – An emotional approach
Author(s):	Philipp Stangor (Heinrich-Heine University Düsseldorf) **Lars M. Kürzinger** (Heinrich-Heine University Düsseldorf)
Conference(s):	Participation and presentation at 'Forschungskolloquium Finanzmärkte', 4^{th} November 2020, Düsseldorf, Germany Participation and presentation at 'HVB Doctoral Colloquium', 15^{th} – 16^{th} January 2021, online hosted by University of Paderborn, Germany Participation and presentation at '58^{th} Annual Meeting of the Eastern Finance Association (EFA)', 6th-9th April 2022, Washington D.C., United States of America Participation and presentation at 'HeiCAD Lightning Talks', 24^{th} June 2022, Düsseldorf, Germany
Publication:	SSRN published. Submitted to 'The Quarterly Review of Economics and Finance', double-blind peer-reviewed journal. Current status: Under review.

Share of contributions:

Contributions	Philipp Stangor	Lars M. Kürzinger
Research Design	**85%**	**15%**
Development of research question	90%	10%
Method developement	80%	20%
Research performance & analysis	**70%**	**30%**
Literature review and framework development	20%	80%
Data collection, preparation and analysis	90%	10%
Analysis and discussion of results	90%	10%
Derivation of implications and conclusions	80%	20%
Manuscript preparation	**100%**	**0%**
Final draft	100%	0%
Finalization	100%	0%
Overall contribution	**85%**	**15%**

31.05.2024, 

Date, Philipp Stangor

31.05.2024, 

Date, Lars M. Kürzinger

4. On the Return Distributions of a Basket of Cryptocurrencies and Subsequent Implications

4.1. Abstract

This study evaluates the risk associated with capital allocation in CCs using a basket of 27 CCs and the CC index EWCI-. We apply basic statistical tests to model the body distribution of CC returns. Consistent with prior research, the stable distribution (SDI) is the most suitable model for the body distribution. However, due to less favorable properties in the tail area for high quantiles, the generalized Pareto distribution is employed. A combination of both distributions is utilized to calculate Value at Risk and Conditional Value at Risk, revealing distinct risk characteristics in two subgroups of CCs.

4.2. Introduction

Following the financial crisis of 2007 and the following period of extreme uncertainty and volatility, trust in the financial system and its institutions, such as central banks and their monetary policies, were shattered(Bouri et al. (2017a); Kaya Soylu et al. (2020)).

Against this background, the market for CCs started to emerge in 2009 with the development of Bitcoin by Nakamoto (2008). This innovative peer-to-peer electronic cash system is not accountable to any subordinating institution, but is managed and controlled by its own community using the blockchain technology. Furthermore, the anonymity and security of transactions represent another noteworthy feature Bitcoin promises its users (Kakinaka and Umeno (2020)), resulting in increasing trading volumes and prices (Corbet et al. (2019)). This development raises questions for both investors and regulators alike regarding the CC market's characteristics and risk profile, which need to be answered for CCs to become an investable asset class for a wide range of investors (Osterrieder et al. (2017); Gkillas and Katsiampa (2018); Majoros and Zempléni (2018)) and to provide guidance for risk management in general (Kakinaka and Umeno (2020)). In this context, this study investigates the question of which family of distribution functions suitably and most accurately models the returns of CCs. We answer this question using a novel approach to separate a distribution's body from its tail proposed by Hoffmann and Börner (2021). By doing so, we are able determine the risk and statistical properties associated with CCs and provide valuable implications for portfolio management and regulators alike.

Although the technical properties of CCs are well understood, CCs' behavior remains to be fully comprehended and analyzed. Thus far, many economic studies have focused on Bitcoin and other prominent CCs such as Ethereum and Ripple since they dominate the CC market

due to their high proportion of total market capitalization (Glas (2019)). In this regard, Baur et al. (2018a) and Glas (2019) find Bitcoin and other CCs to be uncorrelated with traditional assets in times of financial distress. Additionally, Gkillas et al. (2018), demonstrate that CC's behavior also differs from that of fiat currencies. Furthermore, various studies analyze certain characteristics of CCs, e.g., volatility (Polasik et al. (2015); Balcilar et al. (2017)), diversification issues (see, i.a., Brière et al. (2015); Selgin (2015); Corbet et al. (2018b); Schmitz and Hoffmann (2021)) and safe haven properties(Bouri et al. (2017b); Urquhart (2018)).[23] However, to evaluate the corresponding market risk and to completely understand the whole CC market (with its typical features), we follow a literature strand of studies analyzing return distributions of different selected CCs. As numerous studies observe non-gaussian behavior and heavy tails in return distributions (Osterrieder et al. (2017); Gkillas et al. (2018); Gkillas and Katsiampa (2018)), a distribution model accounting for these observed characteristics ought to be implemented. To account for such characteristics Majoros and Zempléni (2018) and Kakinaka and Umeno (2020) use stable distributions (SDIs) in their recent studies. This paper is based on these results. The findings there are reproduced here in an expanded database and possibilities are shown to mitigate the weaknesses of the concept in the risk assessment, especially in the tail area. However, Kakinaka and Umeno (2020) observe the SDI to be unable to efficiently grasp the heavy tails of the analyzed return distributions in all scenarios in comparison to other possible distributions. Given this background, the generalized Pareto distribution (GPD), a statistical distribution that appears to more accurately model heavy tail properties, is used in further studies (Gkillas et al. (2018); Gkillas and Katsiampa (2018)). Following the approach presented in Hoffmann and Börner (2021), we therefore attempt to use a combination of both described distributions. Analytically discovering the beginning of the tail of the analyzed return distributions enables us to divide the given data into a body and a tail, and we implement a different distribution for each. For the first sample, containing the tail of potential losses, we implement the GPD, since the literature and our tests show its goodness of fit for estimating tail values. For the remaining body of our data, we apply the SDI because its fit outperforms that of other possible distributions.

The aim of our study is to add to the existing literature by implementing a novel approach intended to achieve higher quality modeling of the return distributions observed in the CC market. Furthermore, thus far, most studies have merely been concerned with analyzing characteristics of Bitcoin or the most prominent CCs, e.g., Ethereum, Ripple and Litecoin (Baur et al. (2018a); Bouri et al. (2017b); Osterrieder et al. (2017); Gkillas et al. (2018);

[23] For a more extensive literature review, see Corbet et al. (2019).

Gkillas and Katsiampa (2018); Majoros and Zempléni (2018); Kakinaka and Umeno (2020)). Therefore, most studies do not consider the entirety of the CC market, which, as Glas (2019) notes, might lead to potential bias. Hence, we join Glas (2019), ElBahrawy et al. (2017) and Schmitz and Hoffmann (2021) in an attempt to provide a broader overview of the CC market. By doing so, we address two existing gaps in literature identified by Corbet et al. (2019) in their extensive literature review. Namely, we extend the number and size of the analyzed data in an attempt to analyze CCs as an asset class. Furthermore, our research provides practical relevance in the form of an improved risk assessment. By analytically separating a distribution's body from its tail and implementing different distributions, we are able to more precisely estimate risk measures in form of the value at risk and the conditional value at risk, both of which are important for regulators and investors alike.

The remainder of this paper is structured as follows: In Sec. 4.3, we present and describe the data used in our following analysis. In Sec. 4.4, we perform a series of statistical analyses and tests that lead us to the family of SDIs as the best model choice for the body of the CC return distribution. Based on these findings, Sec. 4.5 is concerned with the assessment of the tail risk inherent in an investment in a single CC or the basket aggregated in the EWCI^{-}. We compare both methods of risk assessment at high quantiles. We first use the body model and then adapt the GPD as a tail model for risk assessment in terms of the value at risk and the conditional value at risk. The last section summarizes our most import results and provides an overview of future research topics.

4.3. Data

As a starting point of our data collection, we follow different studies in the literature by extracting daily prices of selected CCs. Central studies were, e.g., Fry and Cheah (2016); Hayes (2017); Brauneis and Mestel (2018); Caporale et al. (2018); Gandal et al. (2018); Glas (2019). The relevant data originates from the website Coinmarketcap.com. and is downloaded for each of the $n = 66$ CCs from the Coinmarketcap Market Cap Ranking (reference date: 2014-01-01), see Tab. 23, assuming an observation period from 2014-01-01 to 2019-06-01, as it was originally done by Schmitz and Hoffmann (2021) and – based on this study – also in later derivative works (e.g. by Börner et al. (2022)). We follow both studies and replicate their way of sample selection (as described above) and data editing (as described below) and thus end up with a replicated version of their original dataset:

Cryptocurrency Sample at 2014-01-01: n = 66 CCs *(as in: Schmitz and Hoffmann (2021))*					
→ **Thereof:** n = 27 CCs considered in the final dataset (printed bold) *(as in: Schmitz and Hoffmann (2021); Börner et al. (2022))*					
CC	ID	CC	ID	CC	ID
Anoncoin	ANC	**FLO**	FLO	**Omni**	OMNI
BitBar	BTB	**Freicoin**	FRC	**Peercoin**	PPC
Bitcoin	BTC	**GoldCoin**	GLC	**Primecoin**	XPM
CasinoCoin	CSC	**Infinitecoin**	IFC	**Quark**	QRK
Deutsche e-Mark	DEM	**Litecoin**	LTC	**Ripple**	XRP
Diamond	DMD	**Megacoin**	MEC	**TagCoin**	TAG
Digitalcoin	DGC	**Namecoin**	NMC	**Terracoin**	TRC
Dogecoin	DOGE	**Novacoin**	NVC	**WorldCoin**	WDC
Feathercoin	FTC	**Nxt**	NXT	**Zetacoin**	ZET
→ **Thereof:** n = 39 CCs excluded from the final dataset due to the data gap critereon *(as in: Schmitz and Hoffmann (2021); Börner et al. (2022))*					
CC	ID	CC	ID	CC	ID
Argentum	ARG	Elacoin	ELC	Luckycoin	LKY
AsicCoin	ASC	EZCoin	EZC	MemoryCoin	MMC
BBQCoin	BQC	FastCoin	FST	MinCoin	MNC
BetaCoin	BET	Fedoracoin	TIPS	NetCoin	NET
BitShares PTS	PTS	Franko	FRK	Noirbits	NRB
Bullion	CBX	Globalcoin	GLC	Orbitcoin	ORB
ByteCoin	BTE	GrandCoin	GDC	Philosopher Stones	PHS
CatCoin	CAT	HoboNickels	HBN	Phoenixcoin	PXC
Copperlark	CLR	I0Coin	IOC	SexCoin	SXC
CraftCoin	CRC	Ixcoin	IXC	Spots	SPT
Datacoin	DTC	Joulecoin	XJO	StableCoin	SBC
Devcoin	DVC	Junkcoin	JKC	Tickets	TIX
Earthcoin	EAC	LottoCoin	LOT	TigerCoin	TGC

(Source: The table is based on an original table by Schmitz and Hoffmann (2021) for an identical market snapshot (full sample and reduced subsamples), which was also created with CC market data by CoinMarketCap.com. The subsampled version of this original dataset by Schmitz and Hoffmann (2021) was also used in the derivative work of Börner et al. (2022)).

Tab. 23 Derivation of the Dataset Under Study

We aim to consider as many CCs as possible from this original sample for our final analyses in order to illustrate a preferably high share of the CC market. Nonetheless, we need to exclude all those CCs in the dataset with longer data gaps (here: with five or more consecutive missing observations). By using a Last Observation Carried Forward (LOCF) procedure, as in Trimborn et al. (2020) and Schmitz and Hoffmann (2021), we are then able to consider all the remaining CCs (means: those CCs with smaller data gaps) in our final dataset.

After conducting these steps, we also end up with 27 remaining CCs, e.g. as in Börner et al. (2022) and Schmitz and Hoffmann (2021). These CCs are depicted in Tab. 23. In the

next steps, we are again guided by the procedure described in Schmitz and Hoffmann (2021): Using the daily USD-EUR exchange rates collected from Thomson Reuters Eikon, the CC price data (originally denoted in USD) is converted to EUR. Furthermore, the resulting daily prices are converted to weekly prices in a subsequent step to avoid any possible weekday biases. Moreover, we do not only use our weekly CC prices on an individual CC level, but also calculate an equally weighted CC index (EWCI), following Schmitz and Hoffmann (2021) to get an aggregated perspective. As we exclude more CCs than the beforementioned study, we will call this index $EWCI^{-}$ for a more precise distinction.

Using their respective price data, we follow Börner et al. (2022) and finally calculate logarithmic returns (simply abbreviated as 'returns' in the remainder of this study) for all the considered individual CCs and the aggregated $EWCI^{-}$ index.

4.4. The Return Distribution of Cryptocurrencies

For an initial classification of CCs, key simple statistics from the standard repertoire of empirical statistics are used below. The description and evaluation of additional statistical properties of CCs, for example value at risk or lower partial moments, are carried out using a suitable distribution function.

Our results show that the family of SDIs is a suitable model for the examined returns of CCs. Hence, this family of distributions is used in Sec. 4.5 for a more in-depth analysis of the statistical properties of CCs.

4.4.1. Determination of Basic Key Statistical Figures of the Cryptocurrencies

Standard procedures lead us to estimates of the set of basic key statistical figures: mean $\hat{\mu}$, variance $\hat{\sigma}^2$ and bandwidth (Tab. 24.).

Crypto ID	Mean $\hat{\mu}$	Variance $\hat{\sigma}^2$	Bandwidth min	Bandwidth max	HDS test H0	HDS test p-value	SIG test H0	SIG test p-value
EWCI$^-$	0.3	1.6	-40.8	37.7	0	79.2	0	10.8
ANC	-1.0	14.3	-241.4	162.9	0	99.0	0	17.1
BTB	-0.7	11.7	-201.2	150.4	0	97.4	0	37.2
BTC	0.9	1.0	-30.4	42.2	0	99.8	0	59.2
CSC	-1.2	30.5	-697.9	169.0	0	92.2	0	51.2
DEM	-1.4	9.3	-100.8	145.8	0	99.6	0	85.8
DMD	0.0	4.1	-84.7	102.1	0	99.0	0	43.9
DGC	-1.7	9.0	-217.4	116.2	0	96.6	0	31.1
DOGE	0.9	3.7	-60.8	144.9	0	97.6	1	0.1
FTC	-0.9	7.6	-144.7	171.7	0	83.2	0	10.8
FLO	0.7	6.8	-67.9	162.2	0	84.8	0	43.9
FRC	-0.5	21.3	-332.1	338.9	0	100.0	0	95.3
GLC	0.2	5.6	-74.1	91.0	0	100.0	0	51.2
IFC	-0.6	10.6	-126.4	296.1	0	50.0	0	51.2
LTC	0.6	2.1	-34.2	87.5	0	99.0	0	21.1
MEC	-1.6	5.6	-112.9	133.1	0	96.6	0	85.8
NMC	-0.9	3.0	-110.2	82.4	0	100.0	0	25.8
NVC	-1.0	5.5	-235.7	128.2	0	99.6	0	8.4
NXT	-0.1	4.2	-83.9	106.5	0	93.8	0	8.4
OMNI	-1.4	5.4	-73.1	116.9	0	82.6	0	43.9
PPC	-0.9	2.7	-60.2	73.8	0	86.8	0	21.1
XPM	-1.0	4.1	-67.7	117.7	0	91.2	1	4.9
ORK	-1.1	7.2	-94.1	137.9	0	96.8	0	37.2
XRP	1.1	3.8	-72.9	109.7	0	100.0	1	0.0
TAG	-1.1	5.3	-63.6	136.3	0	100.0	0	59.2
TRC	-1.0	6.7	-72.7	162.6	0	94.6	0	17.1
WDC	-1.6	6.8	-121.7	110.3	0	99.8	0	51.2
ZET	-1.0	6.6	-97.7	131.4	0	94.8	0	95.3

Units in percent and boolean; see text. Note: Although Schmitz and Hoffmann (2021) calculate similar descriptive statistics for the same CC sample, different results occur due to the usage logarithmic returns in this work.

Tab. 24 Basic Key Statistical Figures and Tests on Unimodality and Symmetry.

The returns scatter strongly around a center close to zero. While the variance and thus the standard deviation indicate leptokurtic behavior and therefore a concentration of returns, the sometimes considerable bandwidth indicates a strong blur of returns over a wide measurement range. This leads to the preliminary conclusion that the returns of CCs in the middle value range follow a concentrated distribution that has pronounced fat tails in the outer areas. In particular, the large variance ($\sim 7\%$) and the bandwidth ($\sim 300\%$) of CCs clearly show the completely different character of CCs compared to traditional asset classes. Comparable values on a weekly basis for the traditional asset classes (stocks, bonds, real estate, etc.) fall in the range of $\sim 0.02\%$ (variance) or $\sim 0.1\%$ (bandwidth) on average. Hence, in comparison there is considerable risk associated with CCs.

Additionally, in this step, we use the Hartigan dip test (Hartigan and Hartigan (1985)) to test the null hypothesis H0 that the empirical distribution is unimodal and symmetric. Thus, for each dataset, the Hartigans dip statistics (HDS) are calculated and evaluated. To test for symmetry, the simple sign test, cf., e.g., Gibbons and Chakraborti (2011), is carried out for each dataset. The last four columns in Tab. 24 show the results of both tests.[24]

The results and especially the high p-values of the HDS test strongly suggest that all datasets obey a unimodal distribution. The CC Infinitecoin (IFC) shows the lowest p-value. This indicates that the empirical distribution could be multimodal. In fact, the histogram of returns for the IFC suggests a multipeaked nature. There are no indications of a fundamental structural break, and hence, this tends to be more of a random nature and is due to insufficient statistics (cf. the theorem of Glivenko (1933); Cantelli (1933)), which might occur with short samples in particular. When adjusting a unimodal distribution function later in Sec. 4.4.3, we expect a lower quality of the distribution model for the returns of this currency.

Apart from three exceptions, a clear result can also be seen in the SIG test for symmetry of the empirical distribution of returns. For the vast majority of CCs, the assumption of a symmetrical distribution of returns at a moderately high level of significance cannot be rejected. While the assumption of a symmetrical distribution of the returns is narrowly rejected for the CC Primecoin (XPM), the rejection for the CCs Dogecoin (DOGE) and Ripple (XRP) is almost clear. We therefore assume that the distributions are slightly skewed. Going forward, we assume the returns of the CCs to be concentrated around zero and the empirical distribution to have a fat tail due to the large bandwidth. Furthermore, we expect the empirical distribution to have a unimodal and essentially symmetrical shape. We cannot rule out that the datasets in question may be (slightly) skewed. We will take this into account when selecting and adapting a suitable distribution function in Sec. 4.4.3.

4.4.2. Statistical Tests to Further Reduce the Variety of Possible Distributions

A number of mathematical models are available for the statistical description of CC returns. On the basis of some characteristic features of the dataset, the family of models can be narrowed down, and a suitable family of functions for representing the distribution can be deduced. In the following, a series of statistical tests are conducted to infer possible function families for the description of our datasets. The same tests are also carried out for the EWCI^{-} defined in Sec. 4.3. In total, $N = 28$ time series are considered in the tests described below.

Overall, the statistical tests in Sec. 4.4.1 and the following are used to examine whether the combined hypothesis that the datasets have a unimodal, symmetrical and stationary distri-

[24]Note that for all tests performed in this paper, the following applies: The boolean '0' indicates that the null hypotheses cannot be rejected at the 5% level, and alternatively the boolean '1' indicates a rejection.

bution must be rejected. Furthermore, we assess whether the hypothesis of an independent, identical distribution of the individual returns must be rejected, which is an important property required to specify a distribution function.

The augmented Dickey-Fuller (ADF) test (Dickey and Fuller (1979); Wooldridge (2020)) is used to test a possible rejection of the stationarity hypothesis. Finally, the autoregressive conditional heteroskedasticity (ARCH) test according to Engle (Engle (1982, 2002)) is used to check whether the hypothesis of homoskedasticity of the innovation process ϵ_t for the individual CCs and the EWCI$^-$ must be rejected.

Augmented Dickey-Fuller Test

The ADF test is performed considering the autoregressive model for the CC return time series, y_t, of each CC[25]:

$$y_t = c + \delta t + \varphi y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \epsilon_t, \tag{27}$$

with a drift coefficient c, a deterministic trend coefficient δ, an AR(1) coefficient φ and the coefficients β_i for the lag terms $i = 1, \ldots, p$ up to the order $p = 10$. In Eq. (27) ϵ_t denotes the innovation process. The aim of the test is to examine the hypothesis of trend stationarity, i.e., $\delta = 0$ is the null hypothesis, in the tables denoted by H0; see, e.g., Wooldridge (2020).

The heatmap in Fig. 20 visualizes the structure of the fitted autoregressive model Eq. (27).

We find the deterministic trend coefficient δ to be comparable to zero in all CCs considered. The results in Tab. 25 in the first four columns provide deeper insight. For the vast majority of CCs, the p-values are comfortably high that a rejection of the null hypothesis of trend stationarity is not indicated here. However, as can be seen in the table, the ADF test rejects the null hypothesis for the CCs FLO (FLO), Quark (QRK) and Worldcoin (WDC) at the 5% confidence level. Next, we more closely examine the corresponding trend coefficients: $\delta_{\text{FLO}} = 1.6\text{e-}03$, $\delta_{\text{QRK}} = 1.5\text{e-}04$ and $\delta_{\text{WDC}} = 1.0\text{e-}04$. Since all the coefficients δ are close to zero, the influence of a possible trend is likely to be of significantly less importance. Therefore, in the following, we assume trend stationarity ($\delta = 0$) for the time series of CCs.

The third column in the heatmap in Fig. 20 illustrates the value of φ. We find the parameter φ to be greater than 0.9 for all CCs and, generally, clearly close to 1. The latter is an important condition to be fulfilled for the assumption of a random walk ($\varphi = 1$).

The ADF test was performed for lags p up to order ten. More complicated dynamics with serial correlation are made apparent in the analysis by the fact that the coefficients of the terms corresponding to the lags are clearly different from zero.

[25]Note that the returns r_t are calculated in this notation according to $r_t = y_t - y_{t-1}$.

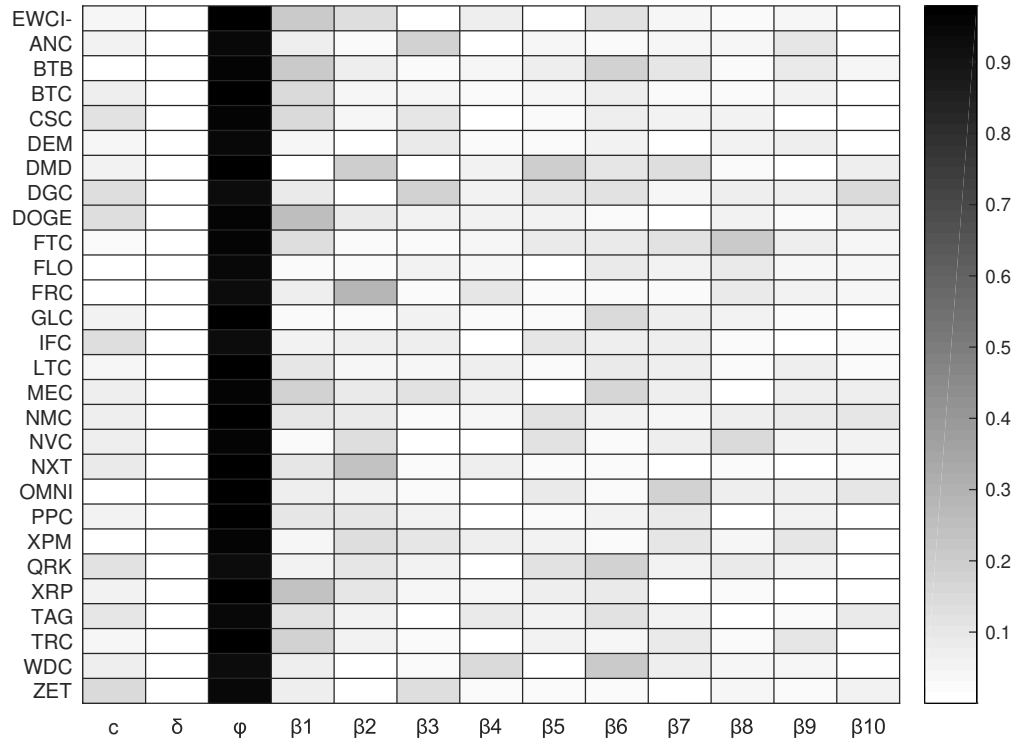


Fig. 20 The heatmap reflects the structure of the autoregressive model with drift coefficient c, deterministic trend coefficient δ, AR(1) coefficient φ and coefficients β_i for the lag terms up to ten time shifts for each CC.

We find the absolute values of the coefficients (that means $|\beta_i|$) to be close to zero. In fact, the majority of the absolute coefficients $|\beta_i|$ are clearly smaller than 0.2 as an upper limit[26] and a basic model $y_t = c + y_{t-1} + \epsilon_t$ can be assumed (in Tab. 25, third column denoted by 'G'). Only a few single coefficients exceed the value 0.2 do so by a small margin, and in these cases, a model with a slightly influential lag structure $y_t = c + y_{t-1} + \bar{\beta} \times (\text{Lag-Structure}) + \epsilon_t$ with an average coefficient $\bar{\beta} = 0.03$ could be considered. By marking the corresponding CCs with 'L', Tab. 25 shows for which CCs this is the case. Due to the observed insignificance of the lag structure, we assume a basic model 'G' in these cases and expect statistical inaccuracies to distort the result due to the limited length of the time series. We therefore postulate possible serial correlation in the datasets to be of minor importance. Thus, the detailed analysis of the results of the ADF test suggest that the model $r_t = c + \epsilon_t$ cannot not be rejected for the returns r_t. Here, c denotes the individual time-constant drift term for each CC, and as above, ϵ_t denotes the innovation process.

[26] By similar argumentation as in (Wooldridge, 2020, Chapter 11 therein), a repeated substitution causes the effectiveness of the corresponding terms to fall below the 5% mark in the next time step and thus become largely insignificant.

Test of Autoregressive Conditional Heteroskedasticity

The ARCH test is performed for time shifts q up to the order of ten. Up to this lag, the null hypothesis that the innovation process of returns is homoskedastic could not be rejected for most CCs (see Tab. 25); i.e., the basic model $\epsilon_t = \sigma z_t$ with constant volatility σ and z_t being an independent and identically (IID) distributed process with mean 0 and variance 1 could not be rejected for the majority of CCs. Note that the results found in literature show a heterogeneous picture, and we do not find ARCH effects in contrast to other related studies (Peng et al. (2018); Dyhrberg (2016b); Avital et al. (2014)). Ultimately, only 13 of 28 time series examined show ARCH effects. The differences across studies may derive from different sampling frequencies or the differently chosen time period or its length and show that the design of the data collection may influence the results.

CC	ADF test			ARCH test		Distribution
ID	H0	Model	p-value	H0	p-value	
EWCI$^-$	0	L	29.4	1	0.0	-
ANC	0	G	11.7	0	6.6	IID
BTB	0	L	22.9	1	0.0	-
BTC	0	G	48.7	1	3.1	≈ IID
CSC	0	G	49.2	0	50.8	IID
DEM	0	G	24.0	1	0.0	-
DMD	0	G	75.8	0	18.0	IID
DGC	0	G	11.0	1	0.0	-
DOGE	0	L	14.7	0	69.8	IID
FTC	0	L	14.9	1	0.0	-
FLO	1	L	4.0	0	63.2	≈ IID
FRC	0	L	14.3	0	71.6	IID
GLC	0	G	65.2	0	16.6	IID
IFC	0	G	6.4	0	38.3	IID
LTC	0	G	34.2	0	25.2	IID
MEC	0	G	17.7	1	1.5	≈ IID
NMC	0	G	42.0	0	38.1	IID
NVC	0	G	22.8	0	92.6	IID
NXT	0	L	50.4	1	0.0	-
OMNI	0	G	35.7	0	70.2	IID
PPC	0	G	41.5	1	0.4	≈ IID
XPM	0	G	12.7	0	11.1	IID
QRK	1	L	2.0	1	0.3	≈ IID
XRP	0	L	35.1	1	0.0	-
TAG	0	G	6.0	0	60.4	IID
TRC	0	G	46.0	1	1.9	≈ IID
WDC	1	L	2.5	1	0.0	-
ZET	0	G	22.6	0	6.0	IID

Units in percent and boolean; see text.

Tab. 25 Results of the Statistical Tests.

The aforementioned results would justify the following calculation: $\mathrm{E}[r_t] = \mu = c$ and $\mathrm{Var}[r_t] = \sigma^2$ for the corresponding CCs. Estimates for the mean and the variance of the individual returns are noted in Tab. 24. Their calculation is also justified with the combined consideration of the test results described above.

When combining the two tests, we noted a characterization of the return distribution in the last column of Tab. 25. For the majority of CCs, IID returns can be assumed. Another part is approximately independent and identical distributed ($\approx$ IID), because either the lag structure is less important in the ADF model or the rejection of homoskedasticity based on the p-value is only weakly justified. For eight CCs, the assumption of IID returns is clearly rejected.

Note that the test for IID returns could have been performed with a turning point test (Bienaymé (1874); Kendall and Stuart (1977)). However, as we are interested in a deeper analysis of the possible serial correlation in our datasets, we use a combination of the ADF test and the ARCH test instead.

All tests conducted thus far do not reject the assumption that the returns obey an essentially symmetrical, unimodal distribution. Furthermore, for the majority of CCs, the assumption of IID or nearly IID returns holds. In the first case (IID), the modeling of the empirical distribution with a distribution function is justified. In the second case ($\approx$ IID), the model represents a coarser approximation. In the latter case, if the assumption of IID returns were to be rejected, the distribution function could only be used as a rough approximation and must be – as in case ($\approx$ IID) – examined more precisely and critically in individual cases, as we show in the following section.

4.4.3. Determination of the Appropriate Return Distribution Function

When modeling the empirical distribution of returns, we focus on families of unimodal distribution functions that are defined over the entire axis (infinite support). Hence, a more detailed investigation of the following distribution functions suggests itself: normal distribution (N), the generalized extreme value distribution (GED), the generalized logistic distribution type 0 and type 3 (GLD0, GLD3) and the SDI.

The analysis below proves that the family of SDIs is the most suitable alternative for modeling the distributions of the CC returns under study. The analyses performed thus confirm the results found in the literature (Majoros and Zempléni (2018); Kakinaka and Umeno (2020)). Consequently, we present this family of functions afterwards in more detail.
Following (Nolan, 2020, Def. 1.4 therein) the SDIs represent a family of distributions appropriate for modeling heavy-tailed and skewed data. In this context, it is noteworthy, that the linear combination of two IID and stably distributed random variables follows the same distributional characteristics as both individual variables. A random variable X follows the SDI $S(\alpha, \beta, \gamma, \delta)$ if its characteristic function can be defined as follows:

$$\mathrm{E}\left[\exp\left(\mathrm{i}tX\right)\right] =$$

$$\begin{cases} \exp\left(\mathrm{i}\delta t - |\gamma t|^{\alpha}\left[1 + \mathrm{i}\beta\mathrm{sign}(t)\ \tan\left(\frac{\pi\alpha}{2}\right)\left(|\gamma t|^{1-\alpha} - 1\right)\right]\right) & \alpha \neq 1 \\ \exp\left(\mathrm{i}\delta t - |\gamma t|\ \left[1 + \mathrm{i}\beta\mathrm{sign}(t)\ \frac{2}{\pi}\ln\left(|\gamma t|\right)\right]\right) & \alpha = 1 \end{cases} \tag{28}$$

The first parameter of the distribution, $0 < \alpha \leq 2$ (named: *shape parameter*), is used to model the tail of the distribution. The second parameter of the distribution, $-1 \leq \beta \leq +1$, is used as a *skewness parameter*: For $\beta < 0$ ($\beta > 0$) the distribution is left-skewed (right-skewed). The distribution is symmetric, if $\beta = 0$. When α is small, the skewness of β is significant. As α increases, the effect of β decreases. Furthermore, $\gamma \in \mathbb{R}^{+}$ is used as a *scale parameter*, and $\delta \in \mathbb{R}$ is a *location parameter*.

For the special case of $\alpha = 2$, the SDI's characteristic function, see Eq. (28), reduces to $\mathrm{E}\left[\exp\left(\mathrm{i}tX\right)\right] = \exp\left(\mathrm{i}\delta t - (\gamma t)^2\right)$ and therefore becomes independent of β, so that the SDI becomes equal to N with mean δ and standard deviation $\sigma = \sqrt{2}\gamma$. For a more detailed description, compare Nolan (2020). For other applications of SDIs in the context of CCs, see, e.g., Börner et al. (2022).

Evaluation of Distance Measurements to Compare Model Quality

In the following, we use standard distance measures to determine and compare the model qualities of N, GED, GLD0, GLD3 and SDI for the individual CCs. There are several distance measures available that are suitable to measure the potential differences between an empirical distribution function and a modeled distribution function. The distance measures from Cramér (1928) von Mises (1931) (W^2-Distance), Anderson and Darling (1952, 1954) (A^2-Distance) and Kolmogorov (1933) Smirnov (1936, 1948) (KS-Distance) are widely used in the literature. A brief summary of the distance measures employed here is given in 4.7.1.

In Tab. 26, we summarize the results for the Anderson-Darling (AD) distance (A^2). The results for the Kolmogorov Smirnov (KS) distance and the Cramér von Mises (CvM) distance (W^2) are compiled in Tab. 33 and Tab. 32 in 4.7.1.

CC ID	Anderson-Darling Distance A^2 N	GED	GLD0	GLD3	SDI	Best Choice
EWCI$^-$	3.41	147.4	1.76	0.81	0.79	SDI
ANC	8.65	22.4	2.53	1.61	0.39	SDI
BTB	3.36	12.3	0.76	0.48	0.22	SDI
BTC	2.07	224.8	0.88	0.60	0.97	GLD3
CSC	21.80	42.6	3.90	n.d.	0.43	SDI
DEM	2.64	6.7	0.54	0.17	0.20	GLD3
DMD	3.80	23.7	1.25	0.29	0.54	GLD3
DGC	6.84	26.6	2.05	0.65	0.78	GLD3
DOGE	9.56	9.1	4.10	1.77	0.53	SDI
FTC	9.41	14.2	1.77	1.05	0.17	SDI
FLO	1.74	1.5	0.54	0.54	0.32	SDI
FRC	17.36	n.d.	5.21	1.98	0.39	SDI
GLC	1.53	16.8	0.40	0.19	0.42	GLD3
IFC	n.d.	n.d.	4.79	3.43	2.57	SDI
LTC	7.10	13.7	2.72	0.72	0.71	SDI
MEC	9.19	18.8	3.71	1.25	0.45	SDI
NMC	6.02	60.5	1.73	0.48	0.47	SDI
NVC	17.24	51.3	4.08	1.47	0.47	SDI
NXT	6.69	17.6	2.39	0.96	0.41	SDI
OMNI	1.25	4.4	0.25	0.24	0.24	SDI
PPC	4.93	31.9	1.63	0.61	0.48	SDI
XPM	5.66	5.0	1.53	0.67	0.27	SDI
QRK	6.32	9.5	2.12	0.53	0.52	SDI
XRP	13.90	13.0	5.71	2.79	0.62	SDI
TAG	4.85	5.4	1.23	0.30	0.64	GLD3
TRC	4.36	5.5	0.80	0.45	0.13	SDI
WDC	9.68	24.1	3.95	1.35	0.57	SDI
ZET	5.23	12.3	1.58	0.41	0.48	GLD3

Tab. 26 Anderson-Darling Distance for Different Body Model Distributions.

The values of the various distance measures show that the GED is least suitable to model the empirical distribution function. This may derive from the fact that the GED contains a fundamental skewness, which can only be slightly influenced via parameter selection. Furthermore, once the shape parameter becomes different from zero, a fundamental change in the distribution model occurs, and the definition interval on the x-axis becomes restricted. Additionally, associated with a change in the sign of the shape parameter is a fundamental change in the distribution model and an abrupt change in the sign of the upper (or lower) bound on the x-axis; see, e.g., Embrechts et al. (1997). In the present case, these properties of the GED make it difficult to precisely adapt the distribution to the dataset.

On closer inspection of the calculated distances, the N also does not appear to be suitable as a model, since it is neither suitable for the modeling of empirical distribution functions with fat tails nor for those with a slight skew. Overall, the GLD0 shows significantly smaller distances across all CCs but is also not ideally suited, as it is completely symmetrical and is therefore

not able to model any slight skewness. The best results in terms of the smallest distance can be achieved with the GLD3 and the SDI. When comparing all CCs, the corresponding distances are very close to one another. If only the Cramér von Mises distance and KS distance are considered, see 4.7.1, we find that approximately half of the empirical distributions of CC returns can be modeled with one or the other distribution. However, once more attention is paid to the tail, i.e., if deviations in the tail area should receive comparably higher weightings to account for tail risks, and the AD distance A^2 is considered, the share of CCs for which the SDI is the most suitable model predominates.

This result ties in with those of Majoros and Zempléni (2018); Kakinaka and Umeno (2020). Using intraday and daily time series for a small sample of CCs, they show the SDI family to be the best choice for modeling the empirical distribution function of intraday and daily returns of CCs.

For a broader sample of CCs, we found that the SDI is, on average, much better suited to model both slight skewness and pronounced tails in the empirical distribution function. Therefore, we use the SDI for all CCs to model the distribution function of returns.

The Stable Distribution Family as a Model for Cryptocurrencies

Tab. 27 shows the results of the parameter estimation for the SDI $S(\alpha, \beta, \delta, \gamma)$. For the individual parameters of the SDI, the 95% scatter intervals are also provided. The latter can be determined from the covariance matrix of the estimated parameters. The covariance matrix of the parameter estimates is a matrix in which the off-diagonal element (i, j) resembles the covariance between the estimates of the i-th parameter and the j-th parameter. For the CC FLO, these scatter intervals cannot be determined numerically using the dataset at hand. This is because the corresponding empirical distribution on the far right shows a very long, pronounced tail and is strongly skewed to the right. The estimated parameter $\hat{\beta}$ of the SDI accordingly takes a value of 1; see Tab. 27. It may be the case that the single right tail data point, i.e., the return in period 62, represents an outlier, which is difficult to determine and correct afterwards without any further knowledge.

CC	Parameter of the SDI $S(\alpha, \beta, \delta, \gamma)$								AD test	
ID	$\hat{\alpha}$	$\pm\Delta\alpha$	$\hat{\beta}$	$\pm\Delta\beta$	$\hat{\gamma}$	$\pm\Delta\gamma$	$\hat{\delta}$	$\pm\Delta\delta$	H0	p-value
EWCI$^-$	1.62	0.18	0.46	0.39	0.07	0.01	-0.01	0.02	0	48.9
ANC	1.41	0.17	0.26	0.31	0.16	0.02	-0.03	0.03	0	85.8
BTB	1.67	0.18	0.38	0.45	0.18	0.02	-0.03	0.04	0	98.3
BTC	1.78	0.16	-0.21	0.61	0.06	0.01	0.01	0.01	0	37.5
CSC	1.45	0.18	0.27	0.32	0.16	0.02	-0.04	0.03	0	81.4
DEM	1.65	0.18	-0.04	0.45	0.17	0.02	-0.01	0.03	0	99.0
DMD	1.56	0.18	0.10	0.39	0.11	0.01	-0.01	0.02	0	70.7
DGC	1.57	0.18	0.31	0.38	0.14	0.02	-0.04	0.03	0	49.7
DOGE	1.33	0.17	0.35	0.27	0.08	0.01	-0.02	0.02	0	72.0
FTC	1.63	0.18	0.54	0.38	0.12	0.01	-0.05	0.03	0	99.7
FLO	1.88	n.d.	1.00	n.d.	0.16	n.d.	-0.02	n.d.	0	92.3
FRC	1.24	0.16	0.06	0.26	0.13	0.02	-0.02	0.03	0	86.0
GLC	1.80	0.16	0.43	0.63	0.15	0.02	-0.02	0.03	0	82.5
IFC	1.47	0.18	0.25	0.33	0.12	0.01	-0.04	0.02	1	4.6
LTC	1.37	0.17	0.06	0.30	0.06	0.01	-0.01	0.01	0	55.2
MEC	1.25	0.16	-0.01	0.26	0.09	0.01	-0.02	0.02	0	79.9
NMC	1.48	0.18	0.07	0.34	0.08	0.01	-0.01	0.02	0	78.0
NVC	1.42	0.17	0.28	0.31	0.08	0.01	-0.03	0.02	0	77.5
NXT	1.48	0.18	0.37	0.32	0.10	0.01	-0.03	0.02	0	83.8
OMNI	1.82	0.16	0.26	0.71	0.14	0.01	-0.03	0.03	0	97.7
PPC	1.50	0.18	0.10	0.35	0.08	0.01	-0.02	0.02	0	76.8
XPM	1.58	0.18	0.42	0.37	0.10	0.01	-0.04	0.02	0	95.6
QRK	1.45	0.18	0.17	0.33	0.12	0.02	-0.03	0.02	0	72.6
XRP	1.33	0.17	0.35	0.27	0.07	0.01	-0.03	0.01	0	63.2
TAG	1.63	0.18	0.10	0.44	0.12	0.01	-0.02	0.02	0	61.1
TRC	1.64	0.18	0.29	0.43	0.13	0.02	-0.04	0.03	0	99.9
WDC	1.26	0.16	0.08	0.26	0.10	0.01	-0.03	0.02	0	67.2
ZET	1.53	0.18	0.15	0.36	0.13	0.02	-0.02	0.03	0	76.5

Tab. 27 Parameters of the SDI and Goodness of Fit Test.

On average, the estimated parameter $\hat{\alpha}$ exceeds 1.5. With parameter α increasing to its limit value of 2.0, it can be seen that the distribution function becomes similar to the N and the skewness parameter ($\beta \neq 0$) becomes increasingly insignificant. The parameters β and α of the SDI are mutually dependent, and as described above, the meaning of β decreases when α increases. Thus, it is generally difficult to infer the skewness from the value β alone. A relative comparison of the distributions with respect to skewness is only possible if α has the same value. Hence, for a more precise analysis of a distribution's skewness, other methods are necessary. In this manner, we used the SIG test as an example and noted the results in Tab. 24.

For some CCs and the EWCI$^-$ index, Fig. 21 shows the empirical densities and the density of the corresponding SDI in comparison.

In addition, the complete AD goodness-of-fit test was performed. The last two columns of Tab. 27 show that for almost all CCs with very high significance (high p-values), the null

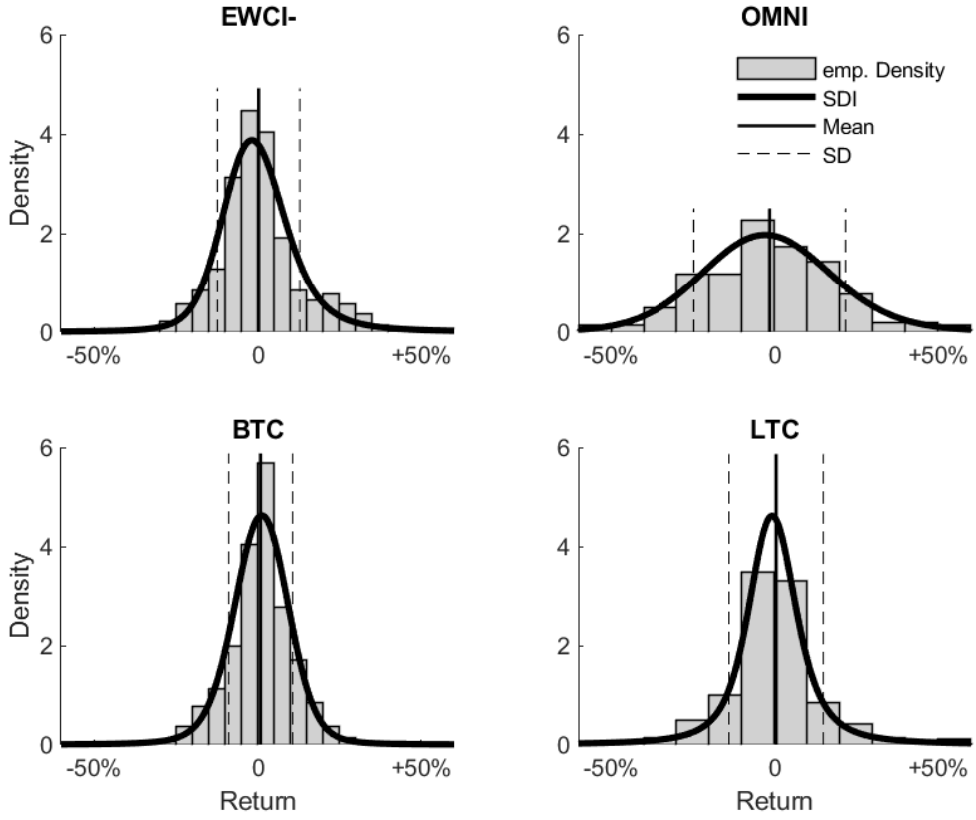


Fig. 21 The Empirical Densities and the Results of the Modeling of the Distributions with the Family of SDIs can be seen for the EWCI$^-$ and Selected CCs.

hypothesis that the adjusted SDI models the dataset cannot be rejected. Only for the CC IFC this assumption is rejected at the 5% level. This is probably because the empirical distribution suggests a slight bimodal distribution. This peculiarity of the CC IFC is also indicated in the results of the HDS test in Tab. 24.

Overall, the SDI family represents a suitable framework for modeling the distribution function of CC returns. We will exploit this finding for the assessment of tail risks and the comparison of our results with other modeling approaches.

4.5. Assessment of Tail Risks

4.5.1. Modeling of Cryptocurrencies' Tail Risks

Especially when considering high quantiles in the risk assessment process, we follow e.g. Hoffmann and Börner (2020a, 2021) and make use of a separated modeling of the parent distribution's tail. In practice, the GPD is used predominantly as a tail model for such a modeling (Basel Commitee on Banking Supervision (2009)). Henceforth, we briefly discuss the main steps of the tail modeling using the GPD in this section.

It is well known and studied that for a wide class of distribution functions, GPD is suitable as a model for the limiting distribution in the tail region if the tail truncating threshold u is large enough (Gnedenko (1943); Balkema and de Haan (1974); Pickands III (1975)).

The GPD is a distribution with (typically) two input parameters and follows the distribution function (Embrechts et al. (1997); McNeil et al. (2015)):

$$F(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}}, \tag{29}$$

where $\sigma > 0$ is the scale parameter and ξ is the shape parameter (a.k.a. tail parameter). The density function is described as

$$f(x) = \frac{1}{\sigma}\left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1+\xi}{\xi}}, \tag{30}$$

with $0 \leq x < \infty$ for $\xi \geq 0$ and $0 \leq x \leq -\frac{\sigma}{\xi}$ when $\xi < 0$. The mean and variance are depicted as $\mathrm{E}[x] = \frac{\sigma}{1-\xi}$ and $\mathrm{Var}[x] = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}$, respectively.

After the foundations of the GPD were introduced by Pickands III (1975), not only theoretical advancements, but also practical applications were built on this work (Davison (1984); Smith (1984, 1985); van Montfort and Witter (1985); Hosking and Wallis (1987); Davison and Smith (1990); Embrechts et al. (1997); Choulakian and Stephens (2001); McNeil et al. (2015); Hoffmann and Börner (2020a,b, 2021)). Besides applications in engineering, the GPD is also the most widely used and recommended distribution function in finance for risk assessment at high quantiles(Embrechts et al. (1997); McNeil et al. (2015); Basel Commitee on Banking Supervision (2009)).

For the means of parameter estimation, the standard maximum likelihood method is the standard approach in the related literature (Davison (1984); Smith (1984, 1985); Hosking and Wallis (1987); Embrechts et al. (1997)). This method is also used here to separately estimate the parameters of the tail distributions of all CC return series under study (belonging to the 27 single CCs and the EWCI^{-} index). Nevertheless, a plausibility check of the results is highly recommended. As we show in Sec. 4.5.2, when assessing tail risks, it is advisable to evaluate the results to avoid possible misinterpretations in individual cases.

In the course of a separated tail modeling, we now only need to consider data points, which belong to the tail area of the underlying empirical distribution function, i.e., the data that belongs to the area below a threshold u in case of the loss tail, for the parameter estimation of the GPD. The correct determination of the threshold u is of crucial importance. Following Hoffmann and Börner (2020a, 2021), the recently developed fully automated process that does not require any user intervention or additional parameters is used, to determine the threshold

u. A brief description of the procedure is given in 4.7.2.

Tab. 28 depicts the estimated parameters of the GPD for the different CCs. The second column reports the proportion of the whole dataset belonging to the loss tail. The proportion of the return data below the threshold $\hat{u}$ is used to fit the parameters ξ, σ of the GPD. The threshold value lies within the bandwidth shown in Tab. 24 and when considering the loss tail closer to the lower interval limit of the bandwidth. Due to the results of different standard goodness of fit tests (here: CvM and AD), the null hypothesis that the GPD is a suitable model for the tail of the CC return distribution cannot be rejected at any significance level. In addition, the p-values for the lower tail (LT) statistics according to Ahmad et al. (1988) are given in Tab. 28. The corresponding statistic AL^2 is defined in 4.7.2 and used here to determine the threshold value u. As can be seen in column six of Tab. 28, the LT statistics also present high confidence levels.

CC ID	Prop.	GPD Parameter			Goodness of Fit		
		$\hat{u}$	$\hat{\xi}$	$\hat{\sigma}$	p-values		
					LT	CvM	AD
EWCI$^-$	45.7	-1.7	-0.09	0.09	88.0	84.9	92.6
ANC	4.3	-52.0	-0.08	0.61	98.5	98.1	93.9
BTB	4.3	-51.0	0.63	0.13	99.0	96.1	98.1
BTC	24.1	-3.9	-0.30	0.10	93.0	97.5	98.1
CSC	9.9	-35.5	0.82	0.11	96.7	99.7	94.3
DEM	51.4	-1.3	-0.05	0.22	54.2	58.2	60.8
DMD	4.3	-32.5	0.04	0.13	94.5	89.0	93.2
DGC	10.3	-31.7	0.61	0.08	99.8	99.6	99.5
DOGE	21.6	-9.2	0.13	0.09	99.1	99.4	98.9
FTC	3.5	-37.0	0.99	0.05	99.7	99.4	93.7
FLO	16.3	-20.3	-0.25	0.17	93.9	91.5	89.2
FRC	11.0	-28.8	0.28	0.31	98.8	98.3	99.6
GLC	51.8	0.2	-0.19	0.20	99.9	99.9	100.0
IFC	20.9	-18.2	0.55	0.07	55.0	70.4	52.8
LTC	44.3	-0.7	-0.18	0.11	57.5	67.7	54.1
MEC	56.4	0.0	0.16	0.12	99.8	99.2	93.3
NMC	46.1	-1.9	0.15	0.09	78.0	95.7	94.0
NVC	9.2	-19.3	0.72	0.05	77.2	93.6	86.0
NXT	2.8	-34.6	1.27	0.03	60.1	76.7	67.7
OMNI	15.2	-23.0	-0.11	0.14	95.9	96.2	97.1
PPC	8.5	21.3	0.00	0.09	96.7	97.1	98.8
XPM	4.6	29.5	-0.08	0.12	91.4	93.1	96.7
QRK	63.5	-1.6	0.05	0.16	47.7	60.8	61.1
XRP	2.8	-25.6	0.76	0.04	99.2	98.8	99.1
TAG	53.2	-0.5	-0.13	0.17	91.2	92.7	96.8
TRC	35.5	-9.9	-0.06	0.15	64.2	70.9	81.0
WDC	62.8	0.8	0.18	0.13	94.8	94.4	88.9
ZET	20.9	-17.8	0.27	0.11	59.8	83.6	82.5

Units in percent.

Tab. 28 Parameters of the GPD and Goodness of Fit Test for the Loss Tail.

4.5.2. Risk Assessment at High Quantiles

In this section, we use the SDI as the body model and the GPD as the tail model to determine the risk parameters of value at risk (as a quantile) and the conditional value at risk (as a weighted loss when the loss threshold is exceeded); see, i.a., Embrechts et al. (1997); Hull (2018). The dataset includes $T = 282$ return observations for each CC, so that a comparison of the results with the empirically determined values is possible for moderately high confidence levels ($\approx 99\%$). The corresponding values for the confidence level of 99.9%, which is important for regulatory purposes (Basel Commitee on Banking Supervision (2004); European Parliament (2009, 2013a,b)), can only be estimated for data records of this length using a previously fitted body or tail model. The calculation of the quantiles is also subject to a statistical spread, and the estimation error increases the fatter the tail is and the higher the

confidence level selected; see, e.g., Hoffmann and Börner (2020b).

Value at risk

Tab. 29 illustrates the results of the risk assessment for the most common confidence levels found in literature and regulatory requirements. The value at risk for the observed CCs is shown for the various models.

CC	Value at risk empirical				Tail Model (GPD)				Body Model (SDI)			
ID	95%	97%	99%	99.9%	95%	97%	99%	99.9%	95%	97%	99%	99.9%
EWCI-	26	29	38	41	19	23	30	43	18	22	33	116
ANC	101	133	212	241	42	73	136	252	46	60	119	582
BTB	75	91	154	201	49	56	82	251	47	55	82	276
BTC	21	24	28	30	16	19	24	31	15	18	28	91
CSC	117	166	424	698	46	58	110	595	47	61	116	535
DEM	72	84	100	101	51	61	82	122	48	60	100	367
DMD	44	52	72	85	30	37	52	85	30	38	69	280
DGC	65	82	153	217	39	46	71	229	40	49	81	314
DOGE	36	43	56	61	23	29	42	77	23	31	64	343
FTC	50	60	109	145	36	38	50	205	32	37	53	177
FLO	49	54	66	68	38	44	55	70	37	42	52	68
FRC	110	142	256	332	57	78	136	333	54	79	185	1159
GLC	49	55	68	74	38	44	56	74	36	42	56	140
IFC	61	77	114	126	34	43	75	251	35	45	82	365
LTC	27	31	34	34	20	24	31	41	21	30	62	321
MEC	59	71	99	113	37	46	70	137	38	56	128	787
NMC	42	50	92	110	27	34	51	97	25	33	63	284
NVC	47	63	141	236	23	28	46	188	24	31	59	278
NXT	42	48	78	84	33	34	42	210	27	33	59	258
OMNI	48	55	67	73	37	43	55	76	37	42	56	142
PPC	35	40	55	60	26	31	41	61	25	33	60	258
XPM	40	46	62	68	29	35	47	70	27	33	51	191
QRK	62	71	91	94	41	50	70	117	38	50	96	440
XRP	30	36	57	73	24	25	32	85	22	30	61	330
TAG	47	53	62	64	35	41	53	73	34	42	69	252
TRC	51	57	69	73	37	44	57	83	36	44	68	238
WDC	64	79	109	122	39	50	76	150	40	57	128	769
ZET	59	73	91	98	37	46	70	151	37	47	85	357

Losses with a positive sign and units in percent.

Tab. 29 Value at Risk of the CCs for Different Confidence Levels and Different Calculation Methods.

Overall, in the overwhelming number of individual cases, the assessment of risk with the tail model (GPD) is closer to the empirical value at risk values. This applies to the lower confidence levels in particular, but even a high confidence level of 99.9%, better estimates are possible in individual cases than with the body model. This becomes apparent from the statistical parameters of the deviation analysis shown in Tab. 30. The mean value and standard

deviation over the set of CCs are shown. For this purpose, the deviation between the modeled variable and the corresponding empirical value at risk was determined. On average, adopting the GPD as the tail model leads to better risk estimates.

ΔVaR				Confidence levels			
				95%	97%	99%	99.9%
Mean	GPD	./.	Emp.	0.5	-0.2	-4.6	15.5
	SDI	./.	Emp.	-0.3	0.4	10.5	214.1
SD	GPD	./.	Emp.	2.4	2.1	7.3	43.1
	SDI	./.	Emp.	1.8	4.1	19.3	211.7

Tab. 30 Average Deviation from the Empirical Value at Risk and Scattering.

When comparing CCs with one another, a heterogeneous picture emerges; see Tab. 29. If the empirical value at risk for the 99.9% confidence interval is taken as a measure, two subgroups can be defined for the cutoff value $\text{VaR}_{99.9\%} \approx 100\%$. One group possesses a significant LT risk ($\text{VaR}_{99.9\%} < 100\%$) as the corresponding $\hat{\xi}$ values in Tab. 28 indicate. The other group ($\text{VaR}_{99.9\%} > 100\ \%$) partly embodies a significantly higher tail risk. Correspondingly, large $\hat{\xi}$ values can be determined for these CCs.

Fig. 22 shows the empirical distribution function for the EWCI$^-$ and Bitcoin (BTC), the SDI as a body model and the GPD as a tail model in comparison. The graphics on the right portray an enlargement of the loss area. Particularly in this region, the GPD models the empirical distribution function very well. Considering the analysis above, we find that the GPD is ideally suited to conduct risk assessment at high quantiles. Therefore, we exclusively consider the GPD to estimate the conditional value at risk as a further risk indicator in the following.

Conditional Value at Risk

The following Tab. 31 shows the conditional value at risk calculated with the tail model (see Tab. 28) for the individual CCs. The calculation of the conditional value at risk can be conducted using the mean excess function of the GPD:

$$e(v) = \frac{\sigma + \xi v}{1 - \xi}, \tag{31}$$

with v being greater than the lower bound of the definition interval of the GPD to consider the loss tail and the parameters σ, ξ of the GPD. The mean excess function is bound to the restrictions $\xi < 1$ and $\sigma + \xi v > 0$; see, e.g., (Embrechts et al., 1997, Theorem 3.4.13). In Tab. 31, we used Eq. (31) to estimate the conditional value at risk for each CC, setting $v = \text{VaR}_{p\%}$.

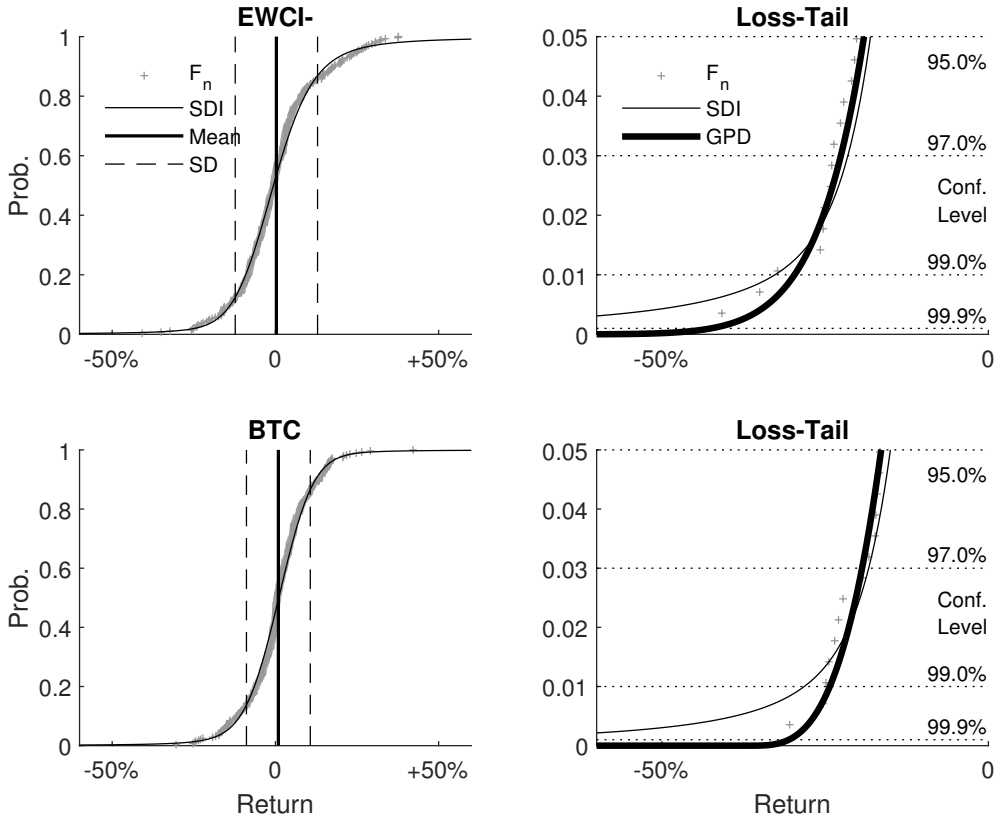


Fig. 22 The empirical distribution function and the distribution function modeled with the SDI can be seen for the EWCI⁻ and an example CC (left panels). The right graphics focus on the loss tail. The GPD adopted as a tail model and the confidence levels that are important for the regulator are also shown.

Again, the grouping of CCs described above can be seen. A group of CCs with a fat tail and therefore higher tail risk can be distinguished from a group with moderate risk; see Tab. 31.

CC ID	Conditional value at risk 95%	97%	99%	99.9%
EWCI-	25	29	35	47
ANC	96	125	184	292
BTB	169	188	258	717
BTC	20	23	26	31
CSC	308	374	655	3294
DEM	70	79	99	138
DMD	45	52	68	102
DGC	118	136	200	600
DOGE	37	43	59	98
FTC	3219	3408	4343	16648
FLO	44	49	58	69
FRC	122	151	231	504
GLC	49	54	64	79
IFC	91	112	181	571
LTC	26	29	35	44
MEC	59	70	99	178
NMC	43	51	71	125
NVC	99	115	180	679
NXT	-134	-138	-165	-784
OMNI	46	51	62	80
PPC	35	40	49	70
XPM	38	43	55	76
QRK	59	69	90	140
XRP	114	121	147	367
TAG	46	52	62	80
TRC	49	55	68	92
WDC	63	76	107	197
ZET	66	79	112	223

Losses with a positive sign and units in percent.

Tab. 31 Conditional Value at Risk of CCs for Different Confidence Levels Calculated with the Corresponding Tail Model (GPD).

Furthermore, two peculiarities are noticeable concerning the CCs Feathercoin (FTC) and Nxt (NXT). For both CCs, the tail is modeled on a small number of data points that have been assigned to the tail. This individual property of the dataset deriving from the random distribution of the data in the tail area is assumed to be given and, as noted above, is not corrected. In particular, when estimating the parameter ξ, small samples lead to large statistical errors. Regrading the conspicuous CCs, the parameter is very close to 1 in one case (FTC) and even higher in the other case (NXT), see Tab. 28. As a result, the calculation of the conditional value at risk for the CC NXT is not possible and must be discarded; cf. Eq. (31) and the restriction $\xi < 1$. On the other hand, the calculation of the FTC with $\xi \approx 1$ has to be questioned critically. Hence, the conditional value at risk may only be a rough estimate in this case.

4.6. Conclusion

The aim of this study is to find a distribution that most accurately models CC returns and does not suffer from restrictions in specific parts of the distribution. In former research, the SDI and GPD have been found to adequately model the body and the tail of the CC return distributions, respectively. Nevertheless both distributions prove to be unsuitable to appropriately model the entirety of the distribution. Therefore, using a novel approach to separate the distribution's tail from its body, we model the entire distribution by combining the model abilities of the SDI for the body and the GPD for the tail.

We select 27 CCs from the broad market of CCs according to predefined criteria and construct the representative index EWCI^{-}. Overall, we find independent, identical distributions such as the GPD and the SDI to be well suited for the most part and the family of SDIs in particular to be able to model the slightly skewed empirical distributions, especially in the body region. A comparison between different distribution functions shows that the SDI has outstanding modeling properties across the entire dataset. However, we show that the assessment of risks associated with fat tails can be performed more precisely with the GPD. The analysis of tail risks in the CC market using the GPD further hints at a certain internal structure of the CC market. The CC market can roughly be divided into two sets: CCs with moderate risk and CCs with high risk. This finding provides valuable information for both investors and regulators alike. Hence, our results are not only relevant for scientific applications and extensions but also for conceivable future regulation if the CC asset class is to become a permanent, noteworthy component of institutional investors' portfolios in the financial sector in the future. In this regard, numerous future extensions and research topics are conceivable. On the one hand, the SDI's suitability to model CC returns in portfolio optimization remains to be investigated. On the other hand, further research considering the segmentation of the CC market could enrich the understanding of CCs and improve forecasts concerned with the fundamental behavior of different CCs.

4.7. Appendix

4.7.1. Appendix A: Distance Measures – Tables of Results

In what follows, a brief summary of the used distance measures is given.

Cramér von Mises and Anderson-Darling distance measures

Following Hoffmann and Börner (2020a, 2021), our first choice to measure the distance between the empirical distribution functions $F_n(x)$ (Kolmogorov (1933)) and a model $F(x)$,

is a weighted mean square error calculated as

$$\hat{R}_n = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \; w(F(x)) \; \mathrm{d}F(x) \tag{32}$$

and originally introduced by Cramér (1928); von Mises (1931); Smirnov (1936) in the context of statistical (hypothesis) testing, cf. also Shorack and Wellner (2009). From a more decision-theoretical point of view (Ferguson (1967)), numerous studies also used the weighted mean square error as an application to determine distribution parameters by using minimum distance approaches (Wolfowitz (1957); Blyth (1970); Parr and Schucany (1980); Boos (1982)). This measure of error is also used when adapting tail models (Hoffmann and Börner (2020a, 2021)). Therefore, in Sec. 4.5.1, we applied this distance measure in connection with the adaption of a suitable tail model for CC returns. A brief overview of the procedure to fit a suitable tail model is given in 4.7.2.

Using a (non-negative) weight function $w(t)$, the formula in Eq. (32) is able to consider the differences between the different distribution functions more accentuated in those areas, where the respective distance measure should be particularly sensitive (Hoffmann and Börner (2020a, 2021)). Usually the weight function

$$w(t) = \frac{1}{t^a (1-t)^b} \tag{33}$$

with parameters $a, b \geq 0$ and $t \in [0, 1]$ is considered. Here, a affects the weight at the lower tail and b at the upper tail. For $a = b = 0$, Eq. (32) provides the CvM distance W^2 used in the corresponding statistic (Cramér (1928); von Mises (1931)). For the case of $a = b = 1$ (means: a heavy weighting of the tail area), the resulting expression becomes equal to the AD distance A^2, which is used in the corresponding statistic by Anderson and Darling (1952, 1954). Thus, potential differences between the two distributions in the upper and lower tails of the distribution $F(x)$ have a higher weighting in the calculation of the AD distance.

CC ID	Cramér von Mises Distance W^2 N	GED	GLD0	GLD3	SDI	Best Choice
EWCI$^-$	0.62	29.89	0.23	0.07	0.12	GLD3
ANC	1.51	4.25	0.41	0.24	0.06	SDI
BTB	0.55	2.22	0.11	0.06	0.04	SDI
BTC	0.40	41.81	0.16	0.11	0.19	GLD3
CSC	3.94	8.22	0.60	0.15	0.09	SDI
DEM	0.43	1.20	0.08	0.02	0.04	GLD3
DMD	0.70	5.19	0.22	0.05	0.10	GLD3
DGC	1.21	5.33	0.33	0.06	0.14	GLD3
DOGE	1.86	1.70	0.64	0.20	0.07	SDI
FTC	1.54	2.37	0.21	0.09	0.03	SDI
FLO	0.29	0.23	0.07	0.07	0.05	SDI
FRC	3.26	6.35	0.94	0.33	0.07	SDI
GLC	0.27	3.66	0.06	0.02	0.08	GLD3
IFC	2.92	4.43	0.91	0.72	0.58	SDI
LTC	1.35	2.98	0.46	0.11	0.12	GLD3
MEC	1.75	3.78	0.67	0.22	0.06	SDI
NMC	1.11	13.40	0.32	0.08	0.09	GLD3
NVC	3.09	10.46	0.65	0.17	0.09	SDI
NXT	1.20	3.57	0.33	0.09	0.07	SDI
OMNI	0.18	0.84	0.03	0.04	0.04	GLD0
PPC	0.88	7.07	0.26	0.09	0.08	SDI
XPM	0.99	0.86	0.20	0.05	0.05	SDI
QRK	1.15	1.74	0.36	0.08	0.10	GLD3
XRP	2.57	2.44	0.79	0.28	0.09	SDI
TAG	0.87	0.94	0.22	0.04	0.13	GLD3
TRC	0.71	0.96	0.09	0.04	0.02	SDI
WDC	1.80	5.03	0.70	0.23	0.10	SDI
ZET	0.92	2.33	0.25	0.05	0.09	GLD3

Tab. 32 Cramér von Mises Distance for Different Body Model Distributions.

The results for the AD distance are shown in Tab. 26 in the main text in Sec. 4.4.3.

Kolmogorov-Smirnov Distance Measure

Furthermore, we also determine the well-known distance between the empirical distribution function and the distribution model of Kolmogorov (1933) and Smirnov (1936, 1948). The KS distance calculates the supremum of the absolute difference between the empirical and the estimated distribution functions. Hence, the KS distance quantifies possible differences between both the theoretically assumed and the empirically observed distribution functions of the CC returns under study. A more theoretical overview and comparisons to other distance measures can be found in, e.g. Stephens (1974); Shorack and Wellner (2009).

CC ID	Kolmogorov-Smirnov Distances KS N	GED	GLD0	GLD3	SDI	Best Choice
EWCI⁻	0.100	0.509	0.063	0.043	0.051	GLD3
ANC	0.132	0.233	0.069	0.070	0.047	SDI
BTB	0.087	0.156	0.056	0.037	0.042	GLD3
BTC	0.077	0.591	0.058	0.047	0.060	GLD3
CSC	0.179	0.296	0.085	0.053	0.047	SDI
DEM	0.071	0.127	0.048	0.027	0.039	GLD3
DMD	0.094	0.233	0.056	0.039	0.050	GLD3
DGC	0.117	0.241	0.066	0.034	0.045	GLD3
DOGE	0.159	0.132	0.099	0.053	0.042	SDI
FTC	0.134	0.141	0.061	0.045	0.031	SDI
FLO	0.069	0.060	0.036	0.035	0.038	GLD3
FRC	0.166	0.250	0.097	0.061	0.047	SDI
GLC	0.072	0.190	0.038	0.025	0.040	GLD3
IFC	0.194	0.251	0.142	0.144	0.106	SDI
LTC	0.133	0.184	0.077	0.043	0.058	GLD3
MEC	0.144	0.206	0.096	0.070	0.041	SDI
NMC	0.095	0.368	0.056	0.040	0.037	SDI
NVC	0.167	0.338	0.092	0.052	0.042	SDI
NXT	0.134	0.196	0.077	0.051	0.039	SDI
OMNI	0.052	0.113	0.028	0.031	0.036	GLD0
PPC	0.109	0.262	0.063	0.048	0.053	GLD3
XPM	0.116	0.095	0.055	0.048	0.035	SDI
QRK	0.108	0.139	0.073	0.048	0.048	SDI
XRP	0.171	0.168	0.101	0.061	0.039	SDI
TAG	0.101	0.103	0.056	0.037	0.048	GLD3
TRC	0.099	0.120	0.039	0.033	0.023	SDI
WDC	0.144	0.236	0.103	0.075	0.052	SDI
ZET	0.123	0.154	0.077	0.049	0.047	SDI

Tab. 33 Kolmogorov-Smirnov Distance for Fifferent Body Model Distributions.

4.7.2. Appendix B: F indT heT ail – Determining Threshold u

As a foundation of our CC tail modeling application, we start with the common assumption, that there is a threshold u, which divides the underlying (parent) distribution into a body and a tail as separately modeled areas (Embrechts et al. (1997); McNeil et al. (2015); Hoffmann and Börner (2021)). This separation is a common approach to capture high quantiles of distributions more accurately (European Parliament (2009)).

Various authors have proposed methods for determining the appropriate threshold u and subsequently the GPD as a model for the tail from empirical data. Most methods require the setting of parameters, which often requires experience and hinders full automation of the modeling process. We follow Hoffmann and Börner (2020a, 2021) and use their full automated process for the determination of the threshold u and the parametrization of the tail model.

Starting with a suitable distance measure $\hat{R}_n = \hat{R}_n(F_n, \hat{F})$ as a function of the estimated

GPD $\hat{F}(x)$ and the empirical distribution function F_n Kolmogorov (1933), an automated modeling process can be constructed using the following pseudo algorithm (Hoffmann and Börner (2020a, 2021)):

1. We arrange the (random) sample data, which is assumed to be drawn from an unknown (parent) distribution, in a descending order: $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$.

2. Assuming $k = 2, \ldots, n$, we now estimate the parameters of the GPD for each k. (Note: For numerical reasons, the process starts at $k = 2$).

3. We then calculate the probabilities $\hat{F}(x_{(i)})$ for $i = 1, \ldots, k$ with the estimated GPD, and determine the distance $\hat{R}_k$ for $k = 2, \ldots, n$.

4. At last, we identify the index k^*, which is relevant for the minimum distance $\hat{R}_k$.

Building on this beforementioned algorithm, we can now estimate the optimal threshold ($\hat{u} = x_{(k^*)}$), and finalize the tail modeling of our unknown (parent) distribution, which is here proxied by the estimated GPD $\hat{F}(x)$ derived from the abovementioned subset $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(k^*)}$.

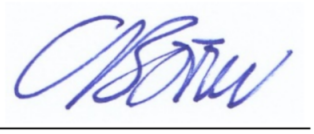
As proposed by Hoffmann and Börner (2020a, 2021) the distance measure defined by Ahmad et al. (1988) is used in the algorithm above. This distance measure is also based on the weighted mean square error, Eq. (32), and can be noted in two variants. The two variants of the distance measure of Ahmad et al. (1988) are derived from Eq. (32) when the integral is calculated with the weight functions Eq. (33) and $(a, b) = (1, 0)$ for the lower tail ($= AL^2$) and $(a, b) = (0, 1)$ for the upper tail ($= AU^2$). With the asymmetrical weight function defined so far the distance measure $\hat{R}_n$ take more account of the difference between the measured and the modeled data, especially in the tail region.

4.8. Declaration of (Co-)Authors and Record of Accomplishments

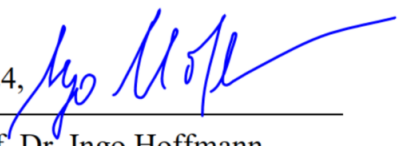
Title:	On the Return Distributions of a Basket of Cryptocurrencies and Subsequent Implications
Author(s):	Prof. Christoph J. Börner (Heinrich-Heine University Düsseldorf) Dr. Ingo Hoffmann (Heinrich-Heine University Düsseldorf) **Lars M. Kürzinger** (Heinrich-Heine University Düsseldorf) Dr. Tim Schmitz (Heinrich-Heine University Düsseldorf)
Conference(s):	Participation and presentation at 'Forschungskolloquium Finanzmärkte,' 7^{th} July 2021, Düsseldorf, Germany
Publication:	Submitted to 'Research in Economics', single-blind peer-reviewed journal. Current status: Under review.

Share of contributions:

Contributions	Prof. Dr. Christoph J. Börner	Prof. Dr. Ingo J. Hoffmann	Lars M. Kürzinger	Dr. Tim Schmitz
Research Design	**30%**	**40%**	**0%**	**30%**
Development of research question	30%	40%	0%	30%
Method developement	30%	40%	0%	30%
Research performance & analysis	**0%**	**20%**	**50%**	**30%**
Literature review and framework development	0%	40%	60%	0%
Data collection, preparation and analysis	0%	10%	10%	80%
Analysis and discussion of results	0%	30%	60%	10%
Derivation of implications and conclusions	0%	20%	70%	10%
Manuscript preparation	**0%**	**0%**	**100%**	**0%**
Final draft	0%	0%	100%	0%
Finalization	0%	0%	100%	0%
Overall contribution	**10%**	**20%**	**50%**	**20%**

03.06.2024, 

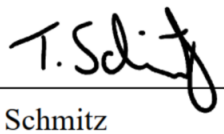
Date, Prof. Dr. Christoph J. Börner

03.06.2024, 

Date, Prof. Dr. Ingo Hoffmann

03.06.2024, 

Date, Lars Kürzinger

03.06.2024, 

Date, Dr. Tim Schmitz

5. The Influence of Intraday Sentiment on Bitcoin Returns

5.1. Abstract

The assessment of Bitcoin, the oldest and most prominent CC, presents a challenge due to the scarcity of fundamental data, the prevalence of uninformed retail investors, and their erratic trading behaviors. To tackle these hurdles, I delve into Twitter sentiment analysis, examining seven emotional dimensions found in tweets related to Bitcoin, aiming to forecast its price movements. Employing the Natural Language Processing Transformer model 'EmTract' by Vamossy and Skog (2023), I scrutinize a substantial volume of Bitcoin-related tweets from Twitter. My analysis focuses on evaluating the impact of intraday valence and its varaince on Bitcoin's future price trajectory across both daily and intraday levels.

I discover that valence significantly influences Bitcoin's price dynamics, particularly during the 18 to 108−minute intervals, while subsequent intervals exhibit more erratic patterns. This study corroborates previous research by confirming the short-term influence of social media sentiment on Bitcoin prices. Moreover, it underscores that while a stable effect exists in the short term, the significance of results in later intervals hinges on the choice of the underlying timeframe, often subject to randomness.

5.2. Introduction

After the financial crisis of 2007 and the disrupted trust of investors in traditional assets and the financial market (Bouri et al. (2017b); Kaya Soylu et al. (2020)), Bitcoin was developed by Nakamoto (2008). The CC Bitcoin, an electronic peer-to-peer cash system, has been gaining popularity since its inception resulting in increasing trading volumes, prices and public attention (Corbet et al. (2019)) making it the CC with the highest market capitalization today. As of 2023 Bitcoin thus comprises around 53% of the total market capitalization of all CCs according to Coinmarketcap.com.

Therefore, investors, academia and regulators alike urge to understand and predict the price movement of Bitcoin and its corresponding returns. When considering classical financial theory, markets operate efficiently, so that all available information is fully reflected in the observed asset prices and only the fundamental value influences stock prices (Naeem et al. (2021)). However, traditional asset pricing models and standard risk factors fail to predict Bitcoin returns properly, as missing fundamental information like dividends, earnings or other cashflows complicate predictions (Liu and Tsyvinski (2021)). Additionally, there are further challenges in the valuation of Bitcoin.

Firstly, Bitcoin investors largely consist of uninformed retail investors or enthusiasts who often lack professional financial knowledge and act irrationally (Yelowitz and Wilson (2015); Almeida and Gonçalves (2023)).

Secondly, this results in a significant impact of social influence and public sentiment on investors' decision-making processes, rendering the valuation and return prediction of Bitcoin challenging. Opinions, popularity and emotions exert a substantial influence on its price development (Mai et al. (2018); Goczek and Skliarov (2019); Bianchi (2020); Naeem et al. (2021); Almeida and Gonçalves (2023)) emphasizing the necessity for modern psychological doctrines to comprehend the dynamic nature of investment decisions (Shrotryia and Kalra (2022)).

Thirdly, CC investors often adopt behavioral trading strategies, focusing on fleeting trends and engaging in high-sentiment and high-volume trades at hourly and daily intervals. This observation reaffirms the presence of noisy trading activity within the CC market. According to the EMH, stock prices incorporate all past information and swiftly assimilate new information to shape future stock prices. Therefore, analyzing Bitcoin over short time periods should lead to advantages in its evaluation (Mishev et al. (2020); Karaa et al. (2021)).

A solution to these challenges could involve utilizing social media sentiment as a non-fundamental factor (Naeem et al. (2021)) that addresses the three highlighted challenges. In this study, social media sentiment is consequently used to account for the dominance of retail investors in the Bitcoin market, analyze their emotions based on the available social media posts, and facilitate intraday analysis due to the abundance of data available, thus accounting for the depicted noisy trading patterns and short-term effects.

To conduct a tailored sentiment analysis effectively, I draw upon insights from previous research, which have already identified several relevant factors to consider when calculating (social media) sentiment through language analysis. Firstly, language usage plays a central role, differing between financial and non-economic settings (Henry (2008); Loughran and McDonald (2011); Renault (2017)). Furthermore, within the economic context, language usage varies across different application areas, resulting in differences in usage between announcements (Rosa and Verga (2007); Amaya and Filbien (2015); Picault and Renault (2017)), articles (Renault (2017)), or social media due to the short and informal nature of the language used (Loughran and McDonald (2016); Renault (2017)). Additionally, differences in language usage can also be observed across various social media platforms (Hutchison et al. (2013)). Secondly, since language can exhibit complex structures, including syntax, irony, and negations, a method of language analysis is required that can capture these intricacies and ideally also takes into account the context in which a word/sentence is situated (Devlin et al. (2018); Peters et al. (2018); Mishev et al. (2020)). Third and lastly, analyzing sentiment based

on emotions, as opposed to a simple positive-negative assessment, allows for a more precise categorization of the posts made. This enables a more in-depth analysis of decision-relevant emotions and reduces data loss due to ambiguous classifications (Stangor and Kuerzinger (2021)).

To meet these criteria, the NLP Transformer model 'EmTract', introduced in Vamossy and Skog (2023), is utilized to ascertain Twitter[27] sentiment. EmTract assigns probabilities to each individual social media post for the presence of seven emotional states.[28] It has been fine-tuned specifically for financial context and social media analysis.

Through sentiment analysis of 54,461,335 tweets related to Bitcoin, using EmTract, I can demonstrate that Twitter sentiment is suitable for predicting Bitcoin returns. Particularly for the 18 to 108 minute intervall, a stable and highly significant influence can be assumed. Specifically, the change in overall emotions ($Change\ of\ Valence$) proves to be effective, while the measure of agreement ($Variance\ of\ Valence$) is not reliable in these intervals but shows increasing influence in longer time intervals. Furthermore, the use of various bootstraps shows that the observed results are robust against the selection of underlying tweets and that cleansing the dataset of all tweets lacking informational content has positive effects on the estimation results, especially in the case of the $VarianceofValence$.

Additionally, my results indicate that the estimation results in later intervals do indeed exhibit (highly) significant results. However, this significance appears to be strongly dependent on the chosen time interval, suggesting a random effect. Consequently, this study contributes to the scientific discourse by demonstrating that sentiment analysis of Bitcoin is promising in short intervals as expected, and that the achieved results in individual intervals should always be checked for stability in the surrounding intervals.

5.3. Data & Methodology

5.3.1. Bitcoin

The Bitcoin price data necessary for analysis is obtained through the Bloomberg API, encompassing the timeframe from May 2022[29] to the conclusion of March 2023[30] on a minute-by-minute basis. This results in 481,886 minute-by-minute price observations. In addition to tracking Bitcoin price movements, minute-by-minute data on the number of

[27] Since the name change of Twitter to X occurred after the time interval considered in this work, the term Twitter will be used instead of X in the following.

[28] The seven emotions considered by EmTract align with the emotional states outlined in Breaban and Noussair (2018).

[29] Minute-level data is only available in Bloomberg for the past six months, hence an earlier date for analysis could not be selected.

[30] Since the Twitter research API used in this paper is no longer available, it was not possible to analyze a longer time period.

transactions is also accessible. The Bitcoin price movement and the number of transactions are depicted in Fig. 1.

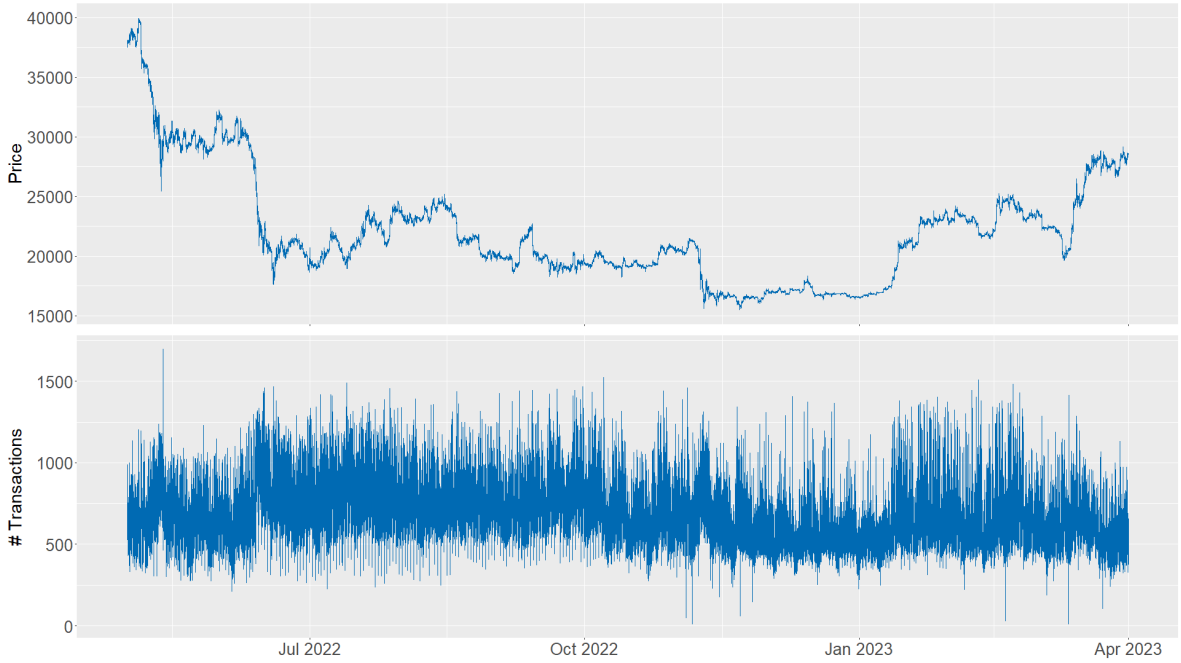


Fig. 23 BTC Price and Number of Transactions over Time

5.3.2. Twitter

In this study, I utilize Twitter data related to the CC Bitcoin. Twitter stands as one of the most popular and widely-used social media platforms to date, generating substantial volumes of data, particularly in the form of tweets. As of 2021 with some topics, such as Bitcoin, generating over a hundred tweets per minute. Given that this study focuses on intraday analysis of social media sentiment, there's a requirement for significant amounts of data within short intervals. For this reason, leveraging the capability provided by Twitter's API to access this data, this social media platform is employed for the current analysis. Using the Twitter API for researchers, I was able to obtain 54,461,335 tweets concerning Bitcoin from 01 Mai 2022 until 31 March 2023. The API prompt considered every tweet which was written in Englisch and included one or more of the following: '#BTC','#Bitcoin', '#Bitcoins', '$BTC'. The obducted data contain the tweets' texts and further information like the time of publication, the Tweet ID etc..

5.3.3. Sentiment Analysis

EmTract

To evaluate the predictive influence of tweets on Bitcoin's performance, it is crucial to discern the sentiment within the tweet texts. This analysis employs EmTract, a recently

developed NLP Transformer model introduced by Vamossy and Skog (2023), which represents a fine-tuned iteration based on DistilBERT. In general the advantage of using such algorithms, which are based on machine learning techniques, lies in their ability to consider linguistic features such as negations, irony and word order. Features which frequently occur in social media texts and are highly relevant for understanding a message's content correctly.

Furthermore, EmTract in particular provides additional advantages to address the challenges discussed in section 5.2. Language is always to be understood within its expressed context and is dependent on the expressions used (Mishev et al. (2020)). Therefore, an algorithm like EmTract is well-suited for its analysis, as this tool has been fine-tuned specifically for language analysis in a financial and social media context. The language and expressions used on social media platforms differs from the language found in reports, newspaper articles, or other forms of media. Therefore, fine-tuning a model for this specific use case should yield improvements in analyzing the particular language used in these contexts. Furthermore, EmTract has the capability to interpret emojis and emoticons, commonly found in social media posts. This inclusion has been shown to enhance predictive accuracy (Vamossy and Skog (2023)), particularly when emojis and emoticons are analyzed in their original formats rather than being grouped into broader categories (Felbo et al. (2017)), as is the approach taken by EmTract.

Lastly, EmTract's focus on analyzing emotions, as opposed to a purely positive-negative perspective, allows for a nuanced analysis of the content of social media texts. It can capture different facets of a tweet more effectively, enabling a multidimensional analysis (Stangor and Kuerzinger (2021); Vamossy and Skog (2023)). Furthermore, this type of analysis takes into account emotionally driven investor behaviors which are especially pronounced in CC markets (Mai et al. (2018); Naeem et al. (2021); Almeida and Gonçalves (2023)). The emotions extracted from the tweets — $neutral$, $happy$, sad, $anger$, $disgust$, $surprise$, and $fear$ — align with the seven emotional states of physiological expressions, as identified in Breaban and Noussair (2018) through facereading software. These values represent probabilities, ensuring that the sum of the seven emotions per tweet always equals 1.

Sentiment

Before extracting emotions from individual tweets using EmTract, preprocessing steps are necessary to ensure that the NLP Transformer model can effectively process the content. In line with this, I adhere to the methodology presented by Vamossy and Skog (2023). This involves removing images, hyperlinks, and tags from the original texts. Subsequently, the text is transformed to lowercase, and contractions (e.g. 'you're' to 'you are') are expanded, while common misspellings are corrected using the Python package SymSpell (e.g. 'ilike' to 'i like'). Numbers, stock tickers, company names, usernames, and unknown tokens are replaced with

<number>, <ticker>, <company>, <user> or <unknown> where appropriate. Additionally, emojis and emoticons are converted using the emoji package in Python.[31] Then the messages undergo tokenization, a process wherein words are transformed into numerical representations. Following tokenization, the messages are segmented into individual sentences and the emotion scores are computed for each individual tweet.[32]

The described approach is now implemented for all tweets within the observation period, resulting in each tweet being assigned a probability score for each of the seven emotions. Fig. 24 illustrates the proportions of the highest respective emotion per tweet.

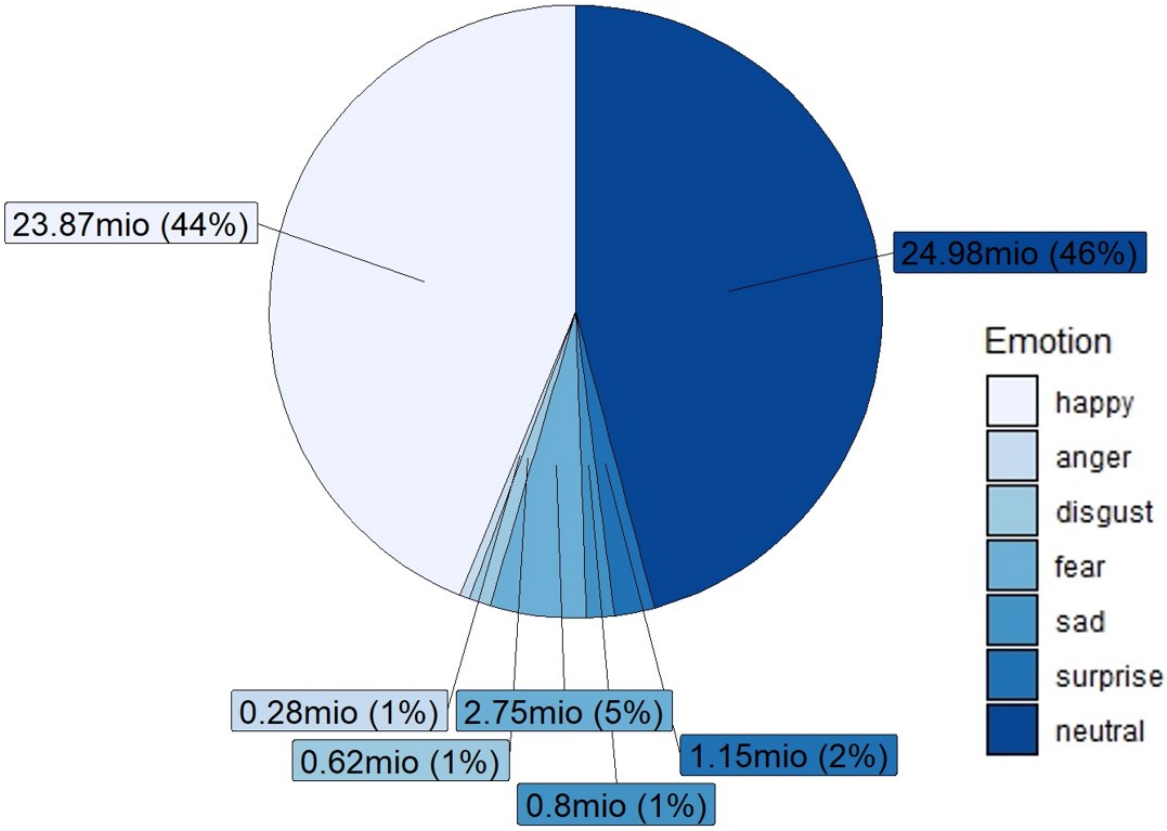


Fig. 24 (Percentage) Share of Strongest Emotion per Tweet

From this figure, it is evident that the largest proportion of all tweets ($\approx$ 46%) is attributed to the *neutral* category. This result suggests a high number of bots, spam, or contentless tweets related to Bitcoin on Twitter, which is not particularly surprising. However, it is noteworthy that there is also a significantly high proportion of tweets classified as *happy* ($\approx$ 44%), compared to the combined proportion of all other emotions ($\approx$ 10%). This is

[31] Vamossy and Skog (2023) provide the described cleaning function under: 'https://github.com/dvamossy/EmTract/tree/main/emtract/processors/cleaning.py'.

[32] The described procedure is illustrated using a sample tweet in Fig. 32 in the Appendix.

surprising given the previously perceived negative trend in the Bitcoin price. According to Vamossy and Skog (2023) who find similar results in their study concerning stock prices, this result implies that investors might be more inclined to express their enthusiasm on social media than pessimism. Other explanations may stem from the more nuanced nature of the other emotions, which do not differ as distinctly from each other as they do from *happy*. This consideration is supported by the correlations depicted in Fig. 25. It is evident that particularly the negative emotions (*sad*, *anger*, *disgust*, *fear*) are relatively positively correlated, while *happy* exhibits a negative correlation with all other emotions. Furthermore, there is a negative correlation of 0.78 with *neutral*, suggesting a clear differentiation between positive and *neutral* tweets, whereas the distinction between negative emotions and *neutral* does not appear to be as clear-cut.

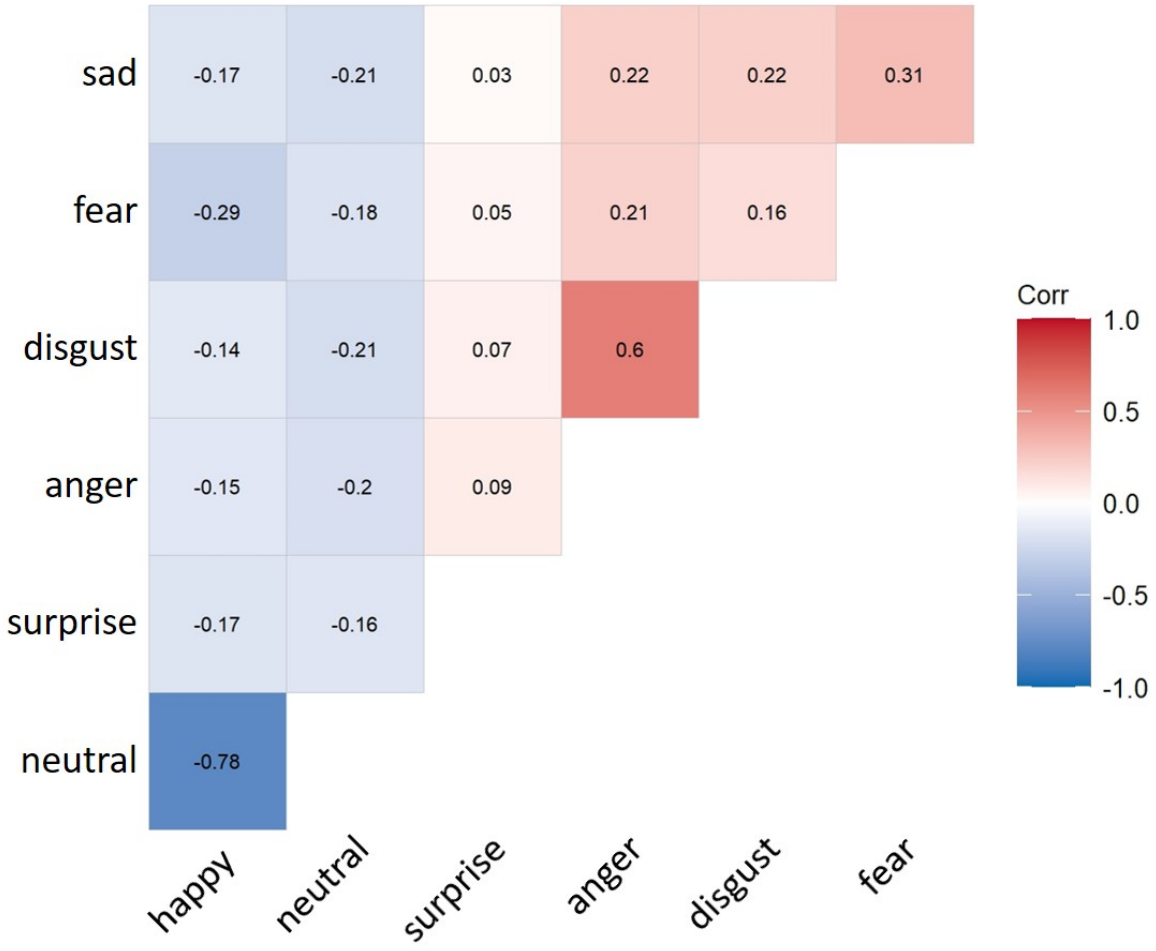


Fig. 25 Correlation of Emotions

To conduct a sentiment analysis using the provided emotions, I follow relevant literature (Breaban and Noussair (2018); Vamossy and Skog (2023)) and translate the positive emotion

$happy$ and the negative emotions (sad, $anger$, $disgust$, $fear$) into a valence measure for each tweet i of the following form:

$$Valence_i = happy_i - sad_i - anger_i - fear_i - disgust_i \tag{34}$$

The $Valence$ can range from $[-1, 1]$, which is derived from the intervals of the individual emotions in the range $[0, 1]$. A value greater than 0 indicates a predominance of positive emotions, while a value less than 0 indicates a predominance of negative emotions. A higher absolute value indicates the strength of the respective predominance, with a maximum at $|1|$. Against this backdrop, Fig. 26 illustrates both the distribution of the $Valence$ within the interval [-1, 1] for all tweets i and the distribution of the individual emotions within the interval [0, 1]. As anticipated from the predominance of tweets classified primarily as $neutral$ or $happy$, as shown in Fig. 24, a significant number of tweets demonstrate a $Valence$ around 0 and in the higher positive ranges. Conversely, only a relatively small proportion of tweets display a negative $Valence$ measure. To accommodate this observation, the analysis in the following section will consistently focus on the $Change\ of\ Valence$ between different time periods rather than its absolute level.

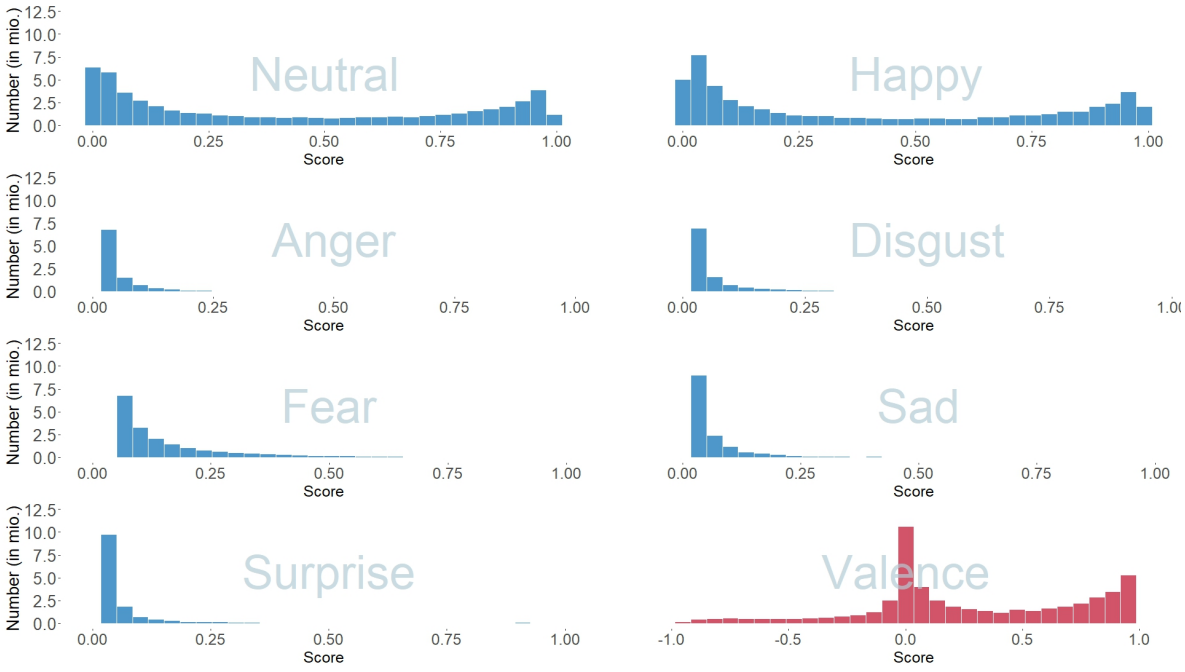


Fig. 26 Distribution of Emotions

It should be noted that so far no cleaning of the selected data for potential spam, bots, duplicates, or empty tweets has occurred to replicate the reality on the social media platform, where also no cleaning of content according to these categories takes place. Instead, the

user decides which tweets contain important information for them. Moreover, this implies refraining from subjectively interfering with the database. In light of this, cleansing the database could potentially be more effectively based on the informational content, as it could better reflect the user's own information selection behavior. Therefore, in addition to a dataset containing all tweets, a subset of this dataset is formed excluding all tweets for which the category *neutral* has the highest probability among all emotions, to account for irrelevant information. To assess the effectiveness of these assumptions, the results of the subsequent analysis will be compared in the next section. As a result, approximately 24,977,890 tweets from the original 54,461,335 are excluded from analysis, leaving 29,483,445 observations. Accordingly, the distributions of emotions in this subset are depicted in Fig. 27.

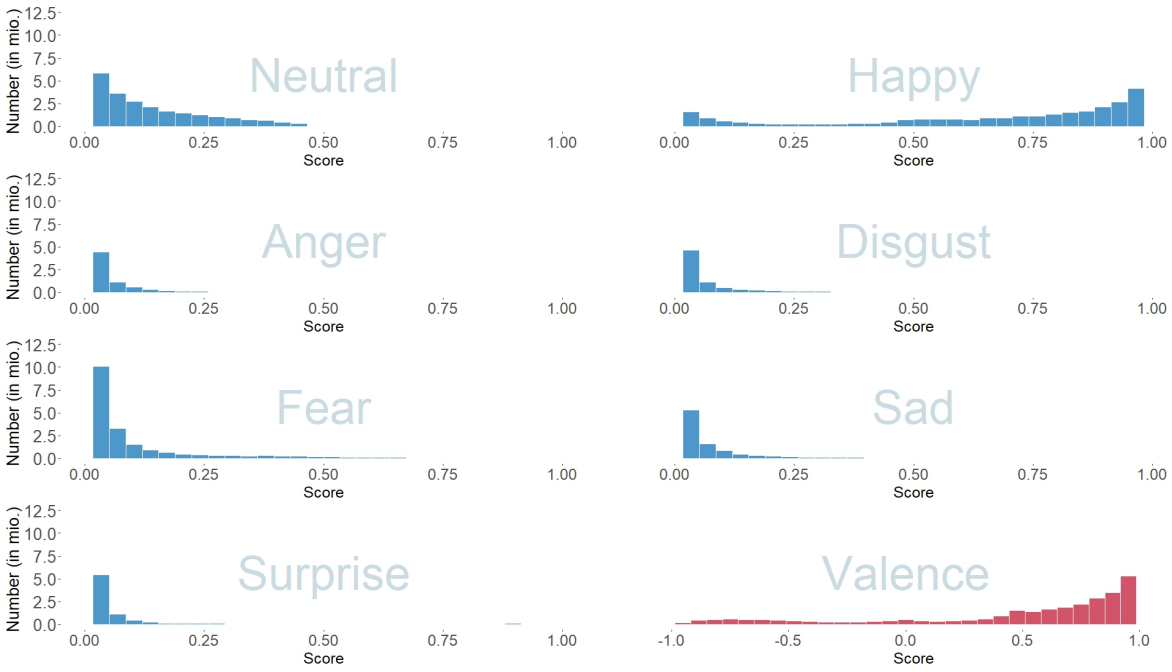


Fig. 27 Distribution of Emotions Excluding *neutral* Tweets

It becomes evident that by excluding tweets primarily classified as *neutral*, the majority of tweets with a $Valence$ of around 0 are significantly reduced, potentially facilitating a sharper distinction in the subsequent analysis.

To forecast Bitcoin's return development, it's crucial to align the sentiment data of the 54,461,335 (29,483,445) million tweets with the Bitcoin price data. Notably, there are frequently multiple tweets per second, whereas Bitcoin price data is on a minute-by-minute basis. Hence, aggregating the $Valence$s of individual tweets becomes essential. Defining the $Valence$ within the time period from t_1 to t_2, where n represents the number of observations, involves computing the average of the $Valences$ of all tweets during this interval as:

$$Valence_t = \frac{1}{n} \sum_{i=1}^{n} Valence_i \tag{35}$$

Since a positive $Valence$ indicates an abundance of the $happy$ emotion, an increasing $Valence$ should correspond to a positive influence on Bitcoin's price development, while a decreasing $Valence$ should lead to the opposite effect.

A consideration of the average $Valence$ per time interval, however, does not provide insights into how the average value is achieved. The $Valences$ of tweets within the time interval may all be close to the average value or exhibit both more extreme positive and negative expressions, which cancel out in the average analysis. For this reason, another variable, the $Variance\ of\ Valence$ within the time interval t, is considered, as follows:

$$\sigma_{Valence_t} = \sum_{i=1}^{n} ((\frac{1}{n} \sum_{i=1}^{n} Valence_i) - Valence_i)^2 \tag{36}$$

By examining the variance of individual $Valences$, an assessment can be made regarding the consistency of social media contributions in terms of $Valence$. A higher variance indicates increased inconsistency in $Valence$ within the observed interval. Therefore, it is reasonable to assume that an increasing $Variance\ of\ Valence$ equates to growing uncertainty in Bitcoin evaluation, potentially exerting a negative influence on Bitcoin price development.

5.4. Results

5.4.1. Daily Analysis

The analysis begins with a daily estimation to verify whether the results align with the relevant literature and to establish a benchmark for the subsequent estimations of intraday influence.

To estimate the influence of the measured Twitter sentiment in the form of $Valence$ and its variance on the future return development of Bitcoin, I first estimate an OLS model (37) of the following form:

$$Return_{BTC,t} = \beta_0 + \beta_1 * \triangle Valence_{t-1} + \beta_2 * \sigma_{Valence_{t-1}} + \epsilon_t \tag{37}$$

To address the effect of the additional information available and to counter potential endogeneity issues, another model with additional control variables is estimated, as well. For

this purpose, the number of tweets ($Number_{tweets,t-1}$) and the number of Bitcoin transactions ($Number_{trans,t-1}$) within the observed time interval, as well as the Bitcoin return in ($Return_{BTC,t-1}$), are initially considered. Subsequently, the percentage difference from the previous period is calculated for both the number of tweets and Bitcoin transactions to obtain a measure of change comparable to that of the returns used. In contrast to stocks, Bitcoin can be traded at any time of day, so changes in the number of tweets and Bitcoin transactions should take this into account. Therefore the extended estimation (38) of the following form is carried out:

$$\begin{aligned} Return_{BTC,t} = \beta_0 + \beta_1 * \triangle Valence_{t-1} + \beta_2 * \sigma_{Valence_{t-1}} + Return_{BTC,t-1} \\ + \triangle Number_{trans,t-1} + Number_{tweets,t-1} + \epsilon_t \end{aligned} \tag{38}$$

with:

$$\triangle Number_{trans,t-1} = log(Number_{trans,t-1}) - log(Number_{trans,t-2}) \tag{39}$$

$$\triangle Number_{tweets,t-1} = log(Number_{tweets,t-1}) - log(Number_{tweets,t-2}) \tag{40}$$

The results of the baseline model based on formula 37 for both the entire dataset (I) and the dataset excluding *neutral* tweets (II), as well as the results of the extended model based on formula 38 for all (III) and excluding *neutral* tweets (IV), can be found in Tab. 34.

According to the regression results considering all tweets, the $Change\ of\ Valence$ has a significant impact on predicting Bitcoin returns on a daily basis. For the estimated models I and IV, the significance level is 5%, while in the case of the entire dataset for the extended model (III), it even reaches the 1% significance level. In all estimated models, the $Change\ of\ Valence$ has a positive effect on Bitcoin returns. A higher $Valence$ signifies an excess of the emotion '$happy$' and an increase in its change thus indicates an increase in this emotion. Therefore, this result aligns with the expected relationships for all models. Additionally, this finding is consistent with relevant literature, which, on one hand, identifies a positive correlation between positive sentiment and Bitcoin returns (Guégan and Renault (2021)) as well as stock returns (see i.a. Bollen et al. (2011); Renault (2017); Breaban and Noussair (2018); Vamossy (2024)). On the other hand, the direction of effect of $Valence$ also corresponds to the results of Vamossy and Skog (2023).

EmTract	(*I*)	(*II*)	(*III*)	(*IV*)
$Intercept$	0.000	0.000	0.000	0.000
	(0.054)	(0.054)	(0.054)	(0.054)
$\triangle Valence_{t-1}$	0.141*	0.153*	0.243**	0.157*
	(0.061)	(0.061)	(0.089)	(0.067)
$\sigma_{Valence_{t-1}}$	-0.092	-0.179*	-0.090	-0.188*
	(0.064)	(0.078)	(0.061)	(0.076)
$Return_{BTC_{t-1}}$			0.002	-0.015
			(0.070)	(0.067)
$N_{trans_{t-1}}$			0.033	0.049
			(0.059)	(0.057)
$N_{tweets_{t-1}}$			-0.196	-0.026
			(0.125)	(0.093)
N_{days}	335	335	355	355
$\sum tweets$	54,461,335	29,483,445	54,461,335	29,483,445
$F-Stat$	4.899	7.198	3,764	3.062
R^2	0.029	0.042	0.054	0.045
$adj.R^2$	0.023	0.035	0.040	0.030

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$
This table depicts the regressions estimates using formula 37 and formula 38 respectively. Columns I and III show the estimation results for all tweets ($n = 54,461,335$) while columns II and IV show the results when $neutral$ tweets are excluded.
All models have been estimated using robust White standard errors and standardized coefficients.

Tab. 34 Estimation Results on Daily Basis

The $Variance\ of\ Valence$ does not exert a significant effect when the entire dataset is used for estimation as models (I) and (III) show. When $neutral$ tweets are excluded in model (II) and (IV) however, it becomes significant on a 5% level. By excluding tweets perceived as $neutral$, which, as previously explained, could be seen as irrelevant information, there appears to be an apparent improvement in the differentiation of individual opinions and emotions within the remaining tweets, leading to the illustrated significant influence. An increase in the $Variance\ of\ Valence$ has a negative effect on Bitcoin returns. This result again corresponds to the expected effect, as an increase in the $Variance\ of\ Valence$ reflects inconsistency among social media users regarding their emotions expressed towards Bitcoin. This uncertainty consequently has a negative impact on Bitcoin returns.

The adjusted proportion of explained variance ($adj.R^2$) of Bitcoin returns in model (III) and model (IV) may be relatively low, but it aligns with the findings of relevant literature in this context (see i.a. Renault (2017); Guégan and Renault (2021)). The regression results of the individual models thus show that, the $Change\ of\ Valence$ and, if $neutral$ tweets are excluded, the $Variance\ of\ Valence$ do exert a significant effect on Bitcoin returns. The estimated effects remain stable even when additional control variables are included in the regressions.

To further examine the robustness of the results, in the next step, I conduct three different

bootstraps for each of the presented models, randomly selecting 50%, 10%, and 1% of the tweets and estimating the model with the new sample again. Thus, the influence of the selection of tweets and the data quantity on the estimated effects is intended to be examined. Subsequently, the models are re-estimated 1,000 times using these tweets. Since the extended models previously demonstrated significantly better performance than the baseline models, only the results of these bootstraps are presented in Fig. 28. Additionally, due to the previously highlighted relevance of the *Change of Valence* and the *Variance of Valence*, Fig. 28 also focuses on these regressors.

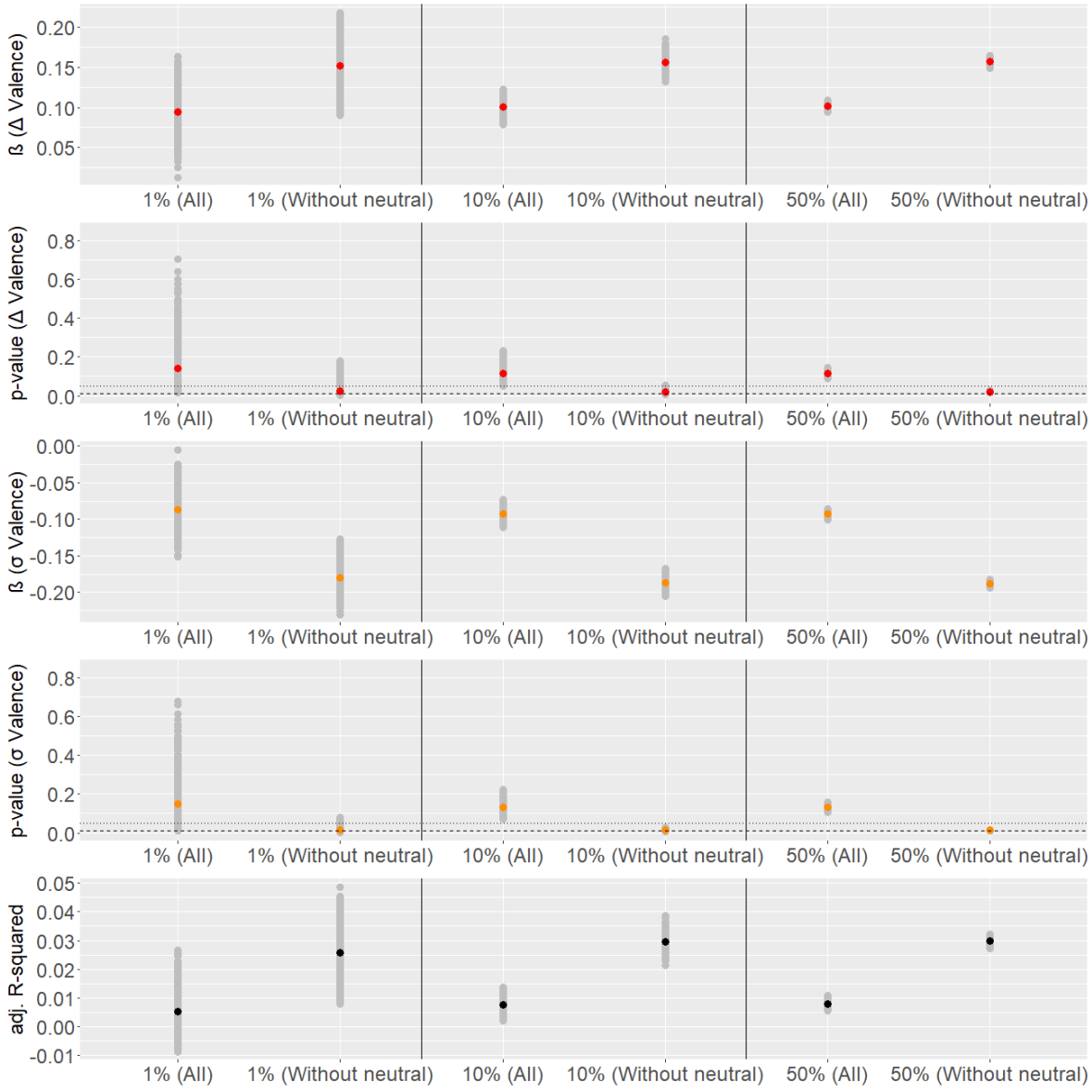


Fig. 28 Results Daily Bootstrap

In Fig. 28, the gray dots represent the estimates for the *Change of Valence* and *Variance*

$of\ Valence$ for each of the 1,000 estimations with random draws. The beta of the $Change\ of\ Valence$, its variance, the corresponding $p-values$, and the $adj.R^2$ are depicted for both models. The red and orange dots represent the median value of the 1,000 estimates.

It is evident that the significance levels decrease as the amount of data decreases, as expected, and the variance between individual estimates with the same amount of data increases, as indicated by the range of estimate results. This demonstrates the differing information content of the selected tweets. For individual estimates, in the case of the availability of 1% of the data, it may occur that the significance level is below 1%, while other estimates show levels, for example, of 40%. In this context, the advantage of cleaning the basic dataset of all $neutral$ tweets becomes apparent. Across all randomly drawn datasets, the significance levels of the estimates without $neutral$ tweets are always lower than those with the total dataset. Therefore, the estimates of the model without considering $neutral$ tweets show significance levels below 5% for the $Change\ of\ Valence$ ($Variance\ of\ Valence$) even when using only 10% of the available tweets in 99.9% (100%) of estimations, whereas in the case of the total dataset, it is only 0% (0%).

Thus, cleaning the dataset of $neutral$ tweets allows for the use of smaller amounts of data, ensuring that even when excluding a large portion of the dataset, the variance of the estimate results remains significantly more stable. However, more meaningful observations naturally provide more conclusive results in terms of significance levels. The reduction of the data volume is not unlimited, as excluding 99% of the data significantly increases the variance of the estimate results. The previously posited hypothesis that $neutral$ tweets can be excluded due to their lack of informational content can thus be confirmed, at least in daily analysis. Furthermore, the realization of the improvement in estimate results with less available data could be helpful in the case of intraday analysis, as smaller time intervals result in fewer tweets per return observation.

These analysis results are subsequently verified in the following section through the intraday analysis.

5.4.2. Intraday Analysis

Due to the noisy trading patterns and short-term behavior of small investors highlighted in section 5.2, in addition to the already conducted daily analysis, an intraday analysis is performed. Although intraday analyses have already been conducted for individual intervals in the relevant literature (see i.a. the works of Behrendt and Schmidt (2018); Broadstock and Zhang (2019); Guégan and Renault (2021)), to the best of my knowledge, no study has comprehensively analyzed the entirety of possible intraday intervals for Bitcoin sentiment thus far. The aim of such an analysis is to determine whether sentiment analysis appears particularly promising for certain periods of several intervals and to ascertain the time frame

that the often postulated 'short-term' effect truly encompasses. Additionally, such an analysis offers the opportunity to gain a more precise insight into the robustness of the estimated results. This is verified both through the development of $p-values$ of the individual estimators across all intervals and, similar to the daily analysis conducted, also using the bootstrap method.

To perform the intraday analysis, the $Valences$ of individual tweets must first be aggregated per interval, similar to the previous daily-based analysis. This analysis considers 1439 different intervals, corresponding to the number of minutes per day (the next day begins at minute 1440), necessitating the aggregation of data for each interval. Subsequently, for each of these intervals, another OLS estimation is conducted with the model configuration using formula 38 and robust White standard errors, both for all tweets and again excluding $neutral$ tweets. The results of these total 1439 estimations for each dataset can be found in Fig. 29 and Fig. 30, respectively.

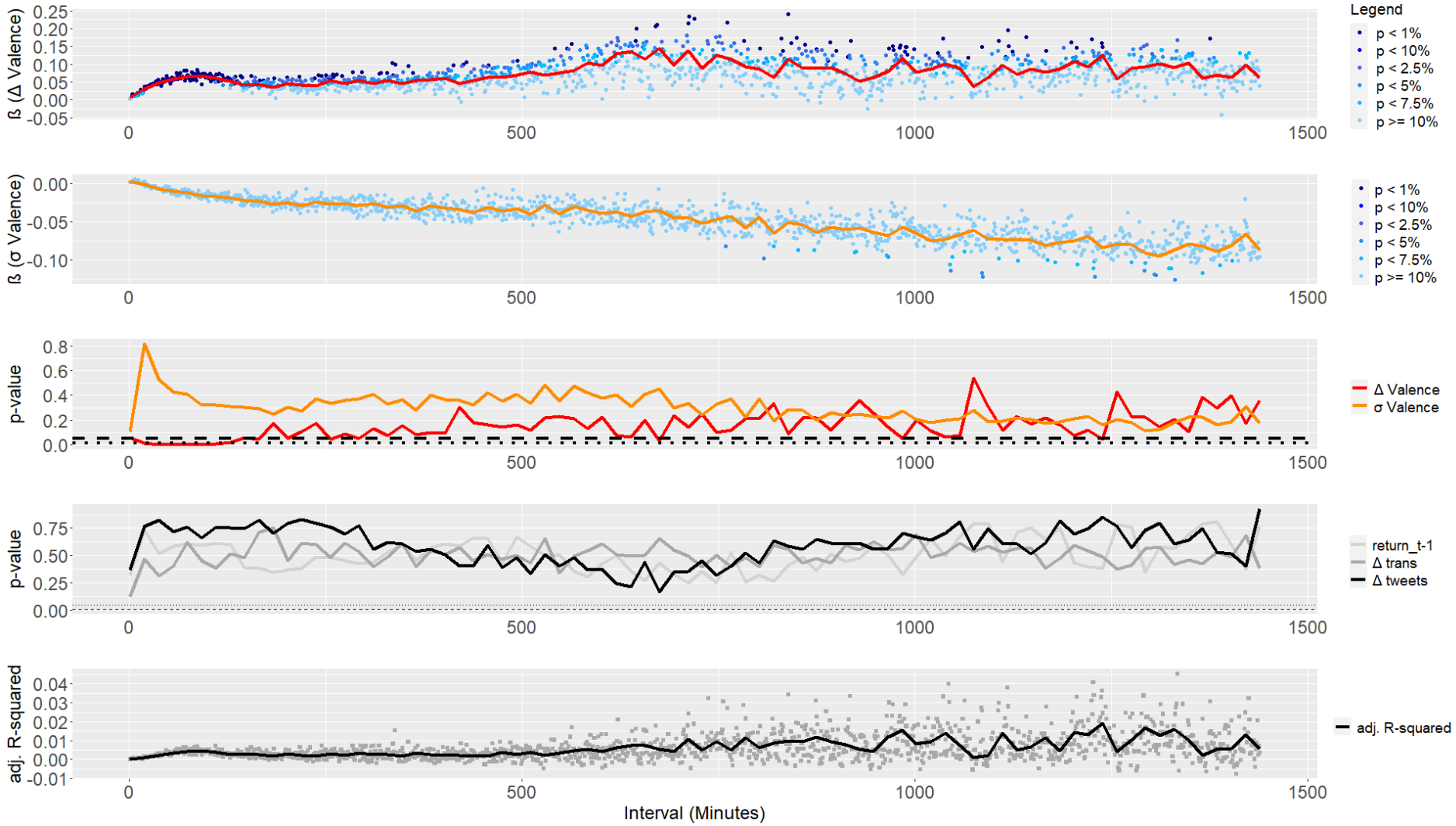


Fig. 29 Results Intraday Analysis

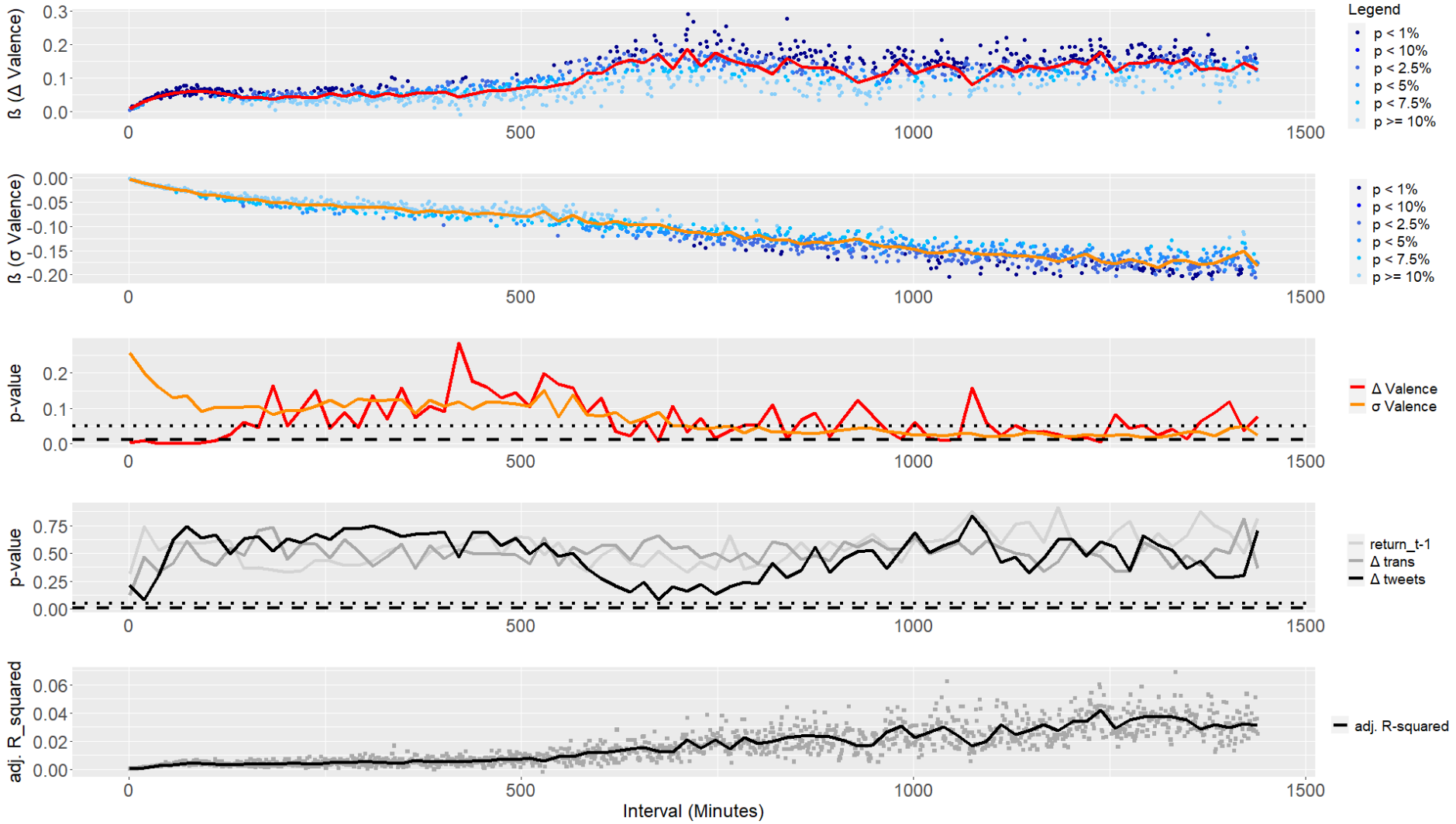


Fig. 30 Results Intraday Analysis Excluding *neutral* Tweets

All data

Fig. 29 depicts, in the upper part, the betas for the estimators $Change\ of\ Valence$ and $Variance\ of\ Valence$ for each time interval (x-axis), whereby each of the points shown corresponds to the magnitude of the beta for a specific interval. Furthermore, the individual points have been colored according to their significance, with darker shading indicating a higher level of significance. The darkest shading represents a significance level < 1%, while a light shading corresponds to insignificant ($p - value > 0.1$) observations. Fig. 29 indicates that the impact of the $Change\ of\ Valence$ on future Bitcoin returns appears especially significant in shorter time intervals, as shown by the red line consistently crossing the 5% and 1% significance levels. Assuming a minimum significance level of 5%, exclusively significant results are obtained in the range from minute 17 to 127. Assuming a target level of 1%, a range from minute 18 to 108 can be identified. The significance of these two intervals is further emphasized by examining the variance of the p-values. Within the interval for the 1% (5%) significance level, the variance is only approximately 0.000025 (0.000032), while outside the interval it is approximately 0.038996 (0.039180).

The depiction of the beta values of the $Change\ of\ Valence$ reveals that the estimated relationship is particularly prevalent in the aforementioned intervals and increases in strength with minimal fluctuation. This is further supported by the lower variance of the beta values within the interval for the 1% (5%) significance level, which is approximately 0.000158 (0.000154), while outside this interval it is approximately 0.0015145 (0.001529).

In longer time intervals, although there are occasional significant results, they appear increasingly random and exhibit a significantly higher variance in both their estimated magnitude and significance. A similar pattern is observed for the $adj.R^2$, depicted in the lower part of the figure. While the explained portion of the variance of Bitcoin returns can occasionally be higher in longer intervals compared to the beginning, however this cannot reliably be extrapolated to surrounding intervals, suggesting a random effect in this aspect as well.

For the $Variance\ of\ Valence$, however, a completely different picture emerges. This estimator is insignificant for the majority of all regressions and only sporadically achieves significant results in later intervals. For a better overview of the development of significance levels, a smoothed representation of the evolution of the $p - values$ of the two regressors can be found in the middle part of Fig. 29. To enhance clarity, the 5% significance level is represented by a dotted line, and the 1% significance level by a dashed line. It is evident that only the $Change\ of\ Valence$ exhibits a significant influence on the future returns of Bitcoin, aside from occasional outliers for the $Variance\ of\ Valence$.

Data excluding neutral tweets

Fig. 30 depicts the results of the intraday analysis when all *neutral* tweets are excluded. For the beta coefficient of the $Change\ of\ Valence$, a similar pattern emerges as before. Particularly in smaller intervals, this estimator exhibits a significance level consistently below 1%. Although, as also evident in the middle part of the figure when examining the $p-values$, there are increased significant estimates compared to the analysis using the entire dataset, these do not consistently reach the 1% significance level and are still affected by higher variance in both the beta coefficients and their significance levels. Assuming a minimum significance level of 5%, exclusively significant results are obtained in the range from minute 9 to 127. Assuming a target level of 1%, a range from minute 18 to 108 can be identified, which correspondents exactly with the results from the previous analysis using the entire dataset. On the other hand, the $Variance\ of\ Valence$ shows significant results on a 5% level, especially in longer intervals, in the case of excluding *neutral* tweets. However, the significant results do not occur with the same regularity as with the $Change\ of\ Valence$. While there is an increase in the relative frequency of significant results, there are alternating occurrences of insignificant estimator results, as indicated by the examination of the color-coded beta values. Against this backdrop, it cannot be assumed that there is a consistently reliable relationship.

Bootstrap results

As already done for daily analysis, I apply the bootstrap method for each individual intraday interval to verify the robustness of the results regarding the selection of tweets used and the number of tweets available. Thus, 1,000 times, 50%, 10%, and 1% of tweets are randomly drawn from the population of all tweets. For each of the 1439 intervalls 1,000 random draws are conducted, for each of which the underlying model is estimated again afterward. This operation is carried out for both the entire dataset and the dataset without *neutral* tweets. The results for the bootstrap with 50% of the data with and without *neutral* tweets can be found in Fig. 31. [33]

[33]The results for bootstraps using 10% and 1% of the data can be found in Fig. 33 and Fig. 34.

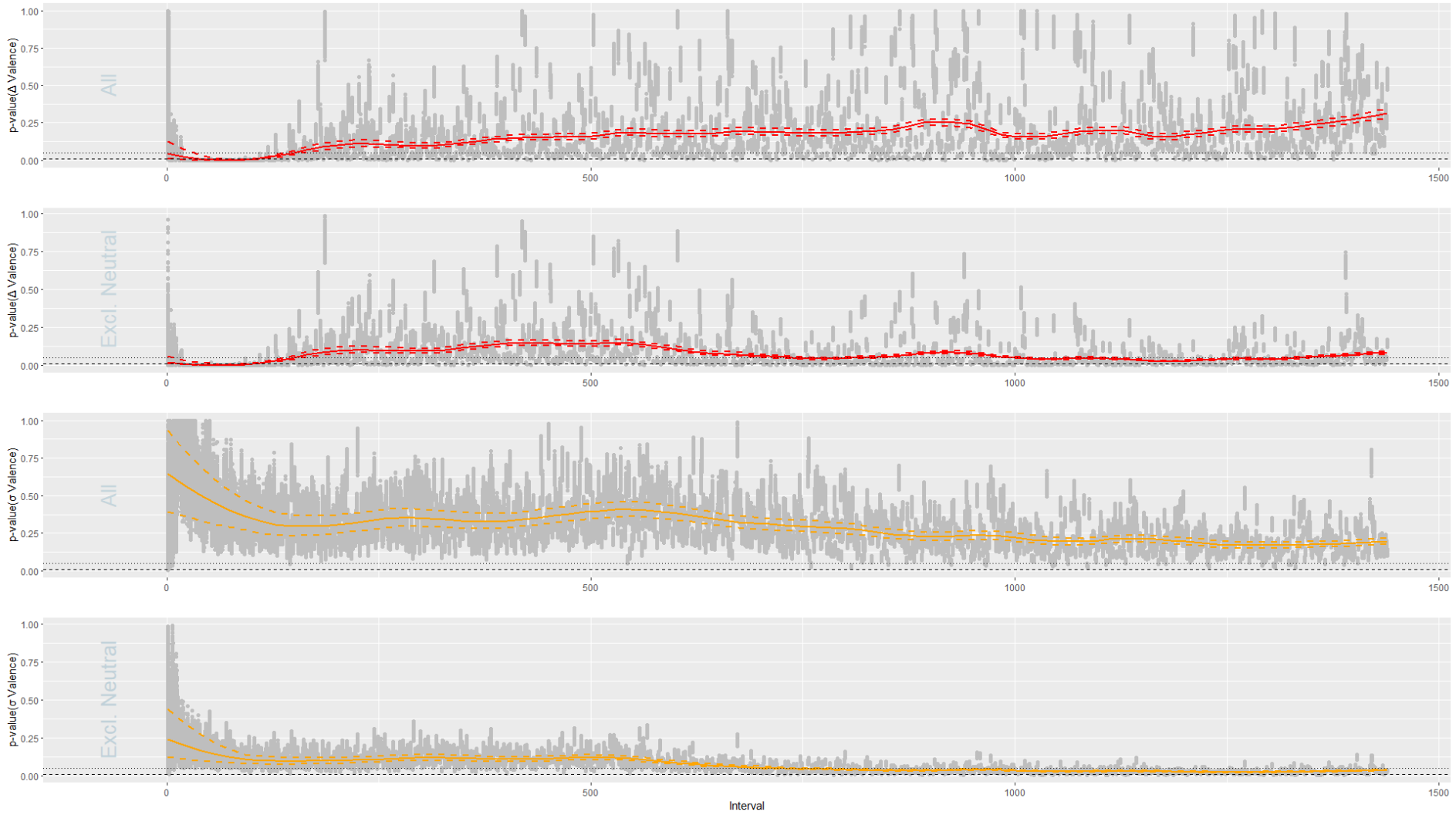


Fig. 31 Results Intraday Bootstrap (50%)

In Fig. 31, the p-values of the $Change\ of\ Valence$ (upper part) as well as the $Variance\ of\ Valence$ (lower part) are depicted. Here, the gray dots represent the p-values for each of the 1,000 estimations per interval. The red (orange) lines indicate the median of the p-values of the $Change\ of\ Valence$ ($Variance\ of\ Valence$), and the dashed lines represent the range within which 95% of all estimated values lie. Additionally, the 1% and 5% significance levels are indicated again by the dashed and dotted black lines, respectively.

The results of the bootstraps indicate that the previously identified interval from minute 18 to 108 remains stable in the bootstrap results, with significantly lower fluctuation compared to other intervals.

A different pattern emerges when considering the $Variance\ of\ Valence$. Again, there is no clear interval indicating a stable relationship when using all tweets. However, there is a tendency for significance of results to increase for longer intervals, accompanied by a decrease in variance of the p-values. Excluding $neutral$ tweets reveals that the median of the p-values is statistically significant at a 5% significance level for longer intervals, yet there are still outliers both above and below.

Thus, the results of the conducted bootstraps confirm the effects observed in previous sections, which remain stable even when excluding 50% of the underlying tweets. Furthermore, excluding $neutral$ tweets enhances predictive power, particularly for the $Variance\ of\ Valence$, allowing predictions even with smaller tweet samples (10%, 1%), as shown in Fig. 33 and Fig. 34.

5.5. Conclusion

The results of this analysis thus allow for the following interpretations. Firstly, there is an influence of sentiment measured by $Changes\ of\ Valence$ and its variance on future Bitcoin returns both on a daily and intraday basis. The influence of the $Change\ of\ Valence$ is particularly statistically significant in the minutes 18 to 108 both when $neutral$ tweets are considered and excluded. However, the variance of the estimated results increases significantly in longer intervals, suggesting the presence of random effects based on the underlying tweets. These results confirm, on one hand, the assumptions made regarding the trading behavior of small and uninformed Bitcoin investors and, on the other hand, align with the findings of comparable literature, which also highlights an increased influence of sentiment in short intervals (Behrendt and Schmidt (2018); Broadstock and Zhang (2019); Guégan and Renault (2021)).

However, the analysis across all intervals also indicates that caution should be exercised in deriving insights from longer intervals due to the varying results, as while individual intervals may show statistical significance, a possible relationship may be questioned due to potential random occurrences. An analysis across multiple intervals and robustness checks

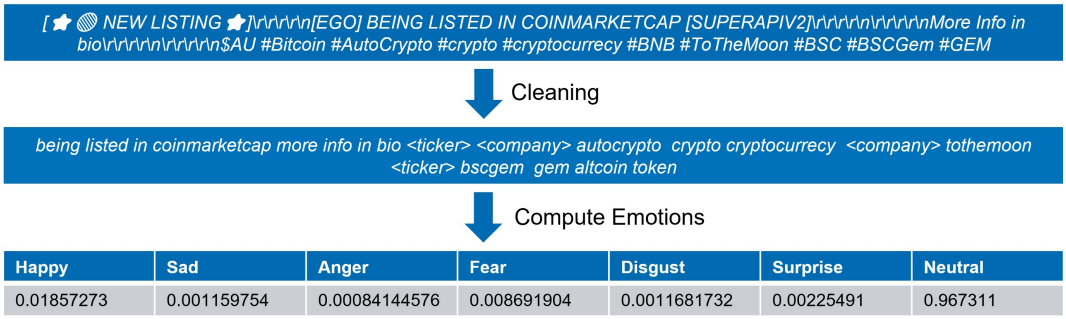
(e.g., using bootstrapping) thus appear necessary. In the case of the *Variance of Valence*, however, while a statistically significant relationship can be observed in the analysis of future Bitcoin returns, especially in longer intervals, this significance does not persist across multiple intervals.

Furthermore, the results of this study show that data cleansing using 'EmTracts' *neutral* category to remove non-informative social media contributions in the form of tweets increases the reliability of the results. Especially through the implementation of multiple bootstraps, it becomes apparent that even with smaller amounts of data, data cleaning enables a reliable analysis and reduces the variance of individual results, while still achieving similar significant estimations. hence, the results of this study appear robust against using smaller datasamples and cleaning the dataset of all tweets considered non-informative.

It should be noted, however, that the results of this study could have benefited from a longer observation period. Unfortunately, due to the data availability restrictions introduced by Twitter, extending the time frame of analysis was not feasible. As outlined in section 5.3.2, the Bitcoin price trend during the observation period was predominantly negative. Given this context, a broader dataset encompassing longer periods of positive developments would have been desirable. Another limitation lies in the use of 'EmTract' as the NLP Transformer model. As detailed in section 5.3.3, 'EmTract' was chosen for its properties necessary for social media sentiment analysis, but it is not specifically fine-tuned for Twitter; rather, it was optimized based on StockTwits data. Although both social media platforms share many similarities, there can be differences in language usage between different platforms. Thus, it can be presumed that the use of an NLP Transformer model specifically optimized for Twitter would be desirable. However, the validation of this consideration could not be conducted at the time of this study due to the absence of such a model.

Thus, in future research contributions, it is important to verify the obtained results over longer observation periods and to compare the results achieved by EmTract with those obtained through multiple NLP Transformer models. Additionally, an analysis using other CCs and asset classes would yield further insight into the influence of intraday social media sentiment. Moreover, a relatively simple model for estimation was used. A more comprehensive examination, for example, of the interactions between individual variables (Kürzinger and Stangor (2024)), the use of individual emotions instead of the *Valence* (Vamossy (2024)), or the inclusion of additional control variables would offer the opportunity to further examine the robustness of the results and to delve deeper into the effect of sentiment on explanatory power.

5.6. *Appendix*



Happy	Sad	Anger	Fear	Disgust	Surprise	Neutral
0.01857273	0.001159754	0.00084144576	0.008691904	0.0011681732	0.00225491	0.967311

Fig. 32 Example of Emotion Computing

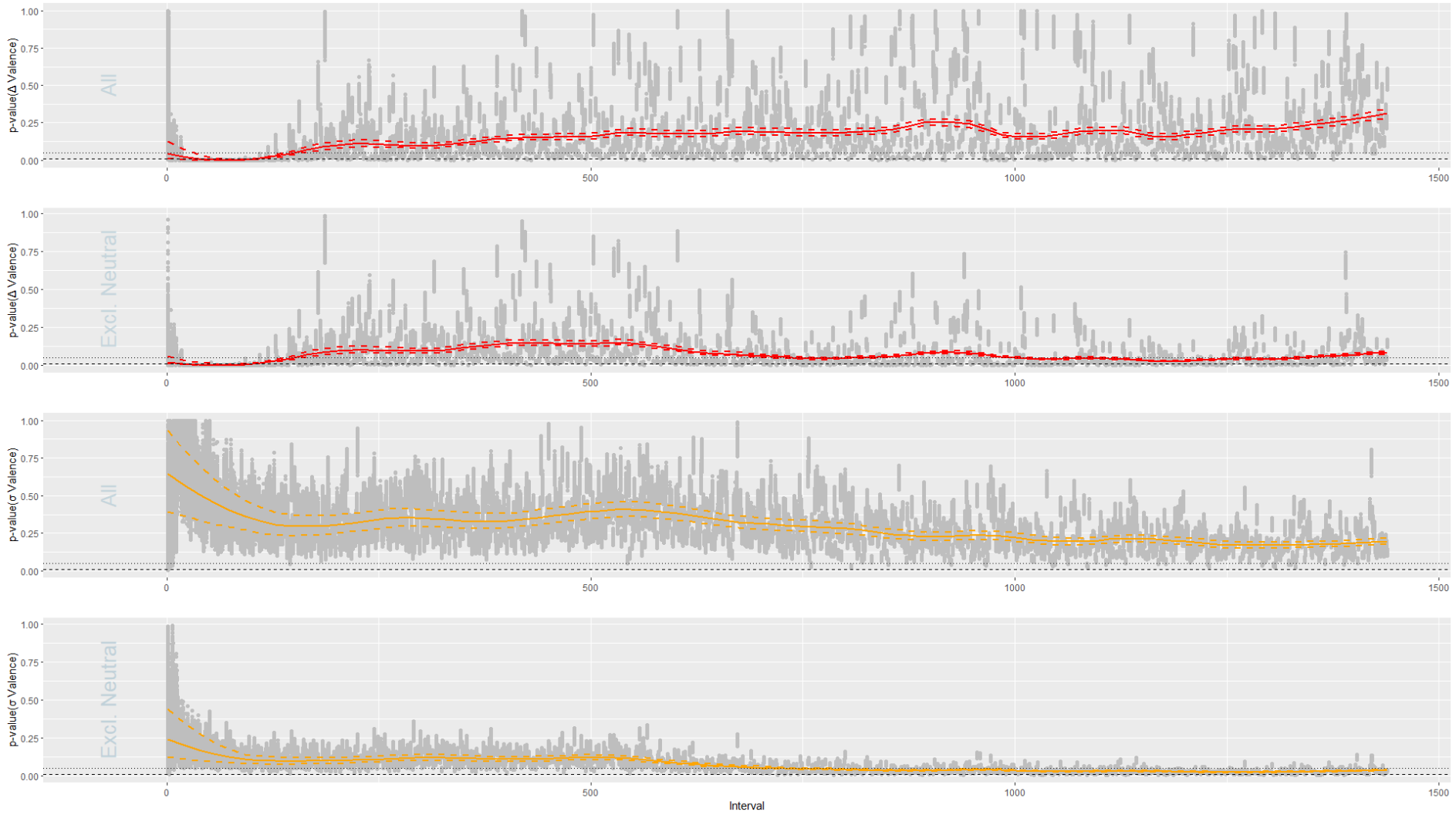


Fig. 33 Results Intraday Bootstrap (10%)

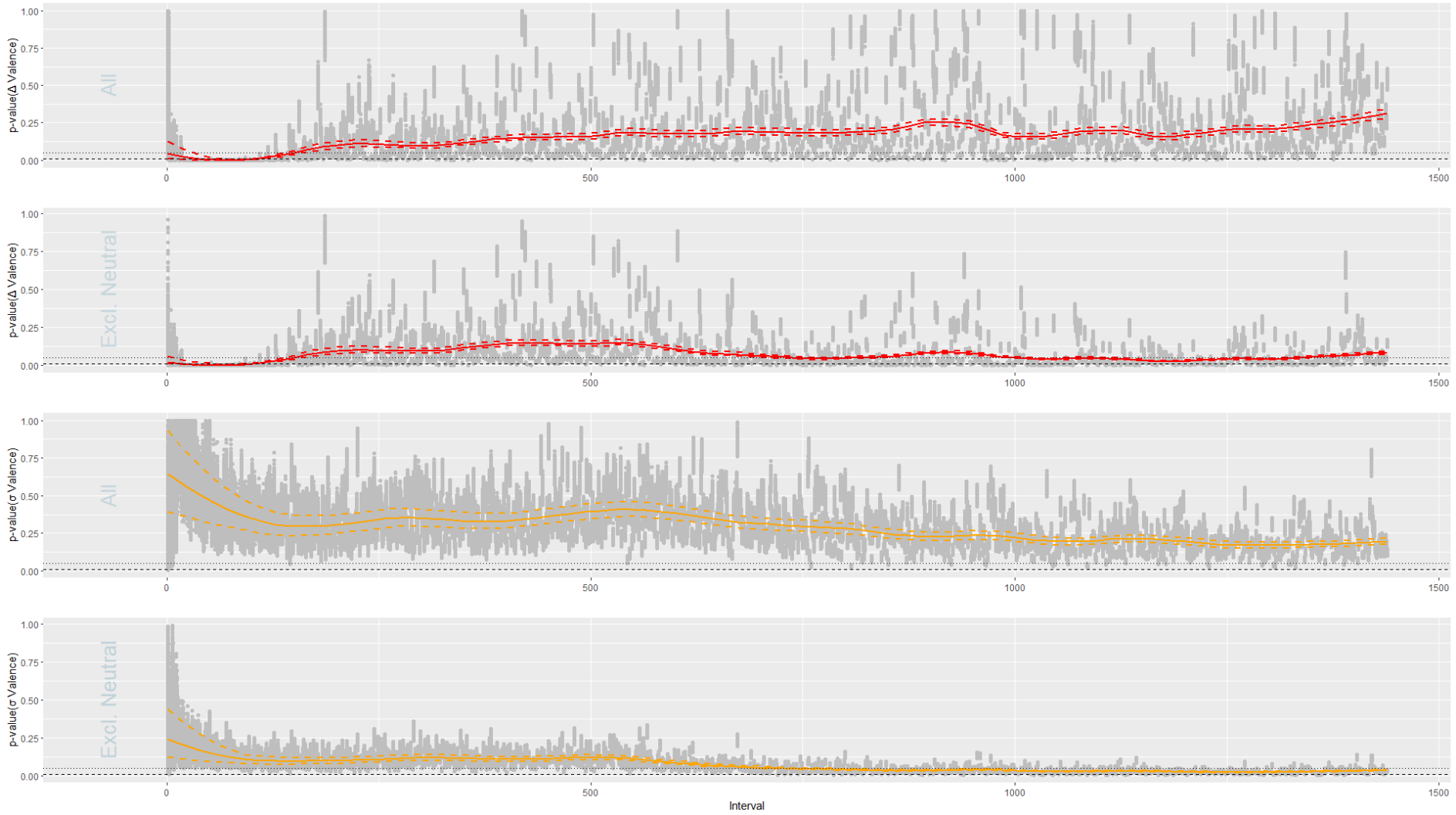


Fig. 34 Results Intraday Bootstrap (1%)

Title: The Influence of Intraday Sentiment on Bitcoin Returns

Author(s): **Lars M. Kürzinger** (Heinrich-Heine University Düsseldorf)

Conference(s): Participation and presentation at 'Forschungskolloquium Finanzmärkte', 7^{th} February 2024, Düsseldorf, Germany

Publication: SSRN published. Submitted to the 'International Review of Financial Analysis', single-blind peer-reviewed journal.
Current status: Under review.

Share of contributions:

Contributions	**Lars M. Kürzinger**
Research Design	**100%**
Development of research question	100%
Method developement	100%
Research performance & analysis	**100%**
Literature review and framework development	100%
Data collection, preparation and analysis	100%
Analysis and discussion of results	100%
Derivation of implications and conclusions	80%
Manuscript preparation	**100%**
Final draft	100%
Finalization	100%
Overall contribution	**100%**

31.05.2024,

Date, Lars M. Kürzinger

6. Conclusion and Outlook

After presenting and discussing the foundations of the relevant literature, the research questions, and the individual research projects within this dissertation, it can be summarized that this work has contributed to scientific research in the following aspects. This work commenced with an examination of the causal relationship between social media posts and investor decisions, along with its mediating mechanisms, in section 2. Subsequently, in section 3, it was demonstrated that a multidimensional analysis of social media sentiment supersedes a two-dimensional approach in terms of data efficiency. Furthermore, in section 4, the modeling of CC returns revealed the distinctive internal structure of the CC asset class. Additionally, section 5 conducted an intraday analysis using sentiment to predict Bitcoin returns, indicating promising predictive potential, particularly in short intraday intervals, while results in longer intervals appear to be driven by randomness. The specific results of each research project and assignment can be found in the following sections.

6.1. Results Social Media Sentiment & Investor Decision Making

Firstly, the question of the causality of sentiment on investor decisions and its specific impact channel on investment decisions was examined using an experimental design in Section 2. The mediation analysis used showed that the nature of the given tweets did not have a direct influence on the investment decisions of the participants. However, there was an indirect influence in terms of the anchoring effect (Tversky and Kahneman (1974)) through the perception of the given financial metrics.

Furthermore, the investigation unveils a nuanced impact of tweets on Financial Sentiment contingent upon the positivity or negativity of financial indicators. The findings suggest a discernible effect when adverse financial data is juxtaposed with positively framed tweets. This phenomenon may find its roots in prospect theory, where individuals, confronted with negative financial outcomes, exhibit risk-averse tendencies distinct from those seen with positive financial outcomes. Consequently, they may display heightened susceptibility to tweets that diverge from the prevailing financial narrative.

The outcomes of section 2 furnish three avenues for future research and practical applications in sentiment analysis regarding the specific directionality of social media sentiment's influence as elucidated in section 2.

Firstly, extending the discussed models to integrate moderators capable of augmenting the impact of social media sentiment could unveil essential factors that shape the responsiveness of economic agents to such sentiment. However, such an analysis necessitates a broader participant base and a larger volume of observations per study group compared to our current study.

Secondly, the discernible influence of bot-generated tweets on the participants suggests that despite their automated nature, they still wield influence over economic agents. This implies the potential to manipulate perceptions of a company's financial health through computer-generated social media content. A rigorous comparison between this approach and human-generated tweets becomes imperative given the rapid advancements in AI technology.

Lastly, the findings underscore the indirect nature of social media sentiment's influence on investor decisions. This underscores the importance of considering this indirect impact in future analyses.

6.2. Results Emotion Analysis

The results of this section highlight three key factors influencing the success of deriving investor sentiment from textual sentiment in an economic context: multidimensional scoring (such as emotions), economic word connotations, and text type. Many researchers have addressed these factors using supervised machine learning algorithms, yielding promising results. However, recent studies, like Renault (2017), echo earlier findings, suggesting that field-specific dictionaries outperform generic benchmark dictionaries and machine learning algorithms. Given the rarity of text classification by publishers, as seen in StockTwits data, and the propensity for misclassification when self-classifying text, there remains a need for multidimensional dictionaries addressing these three factors.

Based on our findings, it's reasonable to expect that more advanced dictionaries or NLP transformers, particularly those incorporating multidimensional scoring, will benefit from the factors outlined above. These insights provide a rationale for the emergence of field-specific and emotion-based NLP transformers, such as RoBERTa/DistilRoBERTa specifications, as evidenced by platforms like Hugging Face's model hub for NLP.

6.3. Results Body & Tail Anaylsis für Cryptocurrencies

In summary, this analysis indicates that both the GPD and the SDI are effective in modeling the empirical distributions of CC returns observed, with the SDI showing particular promise in capturing slightly skewed distributions, especially within the central region. While the SDI demonstrates strong modeling capabilities across the dataset, the GPD also posseses superior precision in assessing risks associated with fat tails.

Furthermore, the examination of tail risks in the CC market using the GPD reveals a discernible internal structure, delineating CCs into categories of moderate and high risk. This segmentation provides valuable insights for investors and regulators, underscoring the importance of understanding risk dynamics within the CC market.

These findings hold implications not only for academic research and future extensions but also for potential regulatory considerations, especially as CCs gain prominence within

institutional investors' portfolios. Moving forward, exploring the SDI's applicability in CC return modeling for portfolio optimization and investigating the segmentation of the CC market offer promising avenues for further research, enriching our understanding of CC behavior and facilitating more informed decision-making in the financial sector.

6.4. Results Intraday Sentiment & Bitcoin Analysis

The analysis reveals several key findings. Firstly, there's a notable impact of sentiment, as measured by changes in valence and its variance, on future Bitcoin returns, both daily and intraday. This influence is particularly significant within shorter time frames, indicating heightened sensitivity to sentiment among investors. However, caution is advised when interpreting results over longer intervals due to increased variance and potential random effects.

Moreover, the study demonstrates the effectiveness of data cleansing using the 'EmTracts' $neutral$ category, enhancing result reliability. Despite using smaller data samples, cleaning non-informative tweets leads to more consistent outcomes. However, limitations arise from the restricted observation period and the choice of 'EmTract,' not specifically fine-tuned for Twitter.

Future research should validate these findings over longer time periods, for other asset classes and CCs. Additionally these presented results should be compared with alternative NLP Transformer models, and more sophisticated modeling approaches should be explored. This includes investigating individual emotions' effects, their interaction with other varaibles as outlined in section 2 and incorporating additional control variables to enhance the analysis's robustness and explanatory power.

6.5. Final Remarks

In summary, the presentation of the individual chapters and research contributions of this dissertation highlights that the topic of social media sentiment analysis has been comprehensively examined. Based on the identified research gaps, the addressed research questions were developed, which contribute both to basic research in the area of sentiment and its impact on investment decisions, and have revealed new insights into the use of advanced methods in intraday analysis. Significant results were also achieved in the area of cryptocurrencies, particularly Bitcoin. Consequently, this dissertation has provided a holistic view, especially on the subject of social media sentiment analysis, which in the future could be further enhanced by the already identified open research questions, as well as a combination of social media sentiment analysis with the distribution of cryptocurrency return distributions.

References

Aggarwal, D., 2022. Defining and measuring market sentiments: a review of the literature. Qualitative Research in Financial Markets 14, 270–288. doi:10.1108/QRFM-03-2018-0033.

Agrawal, A., Tandon, K., 1994. Anomalies or illusions? evidence from stock markets in eighteen countries. Journal of International Money and Finance 13, 83–106. doi:10.1016/0261-5606(94)90026-4.

Ahmad, M.I., Sinclair, C.D., Spurr, B.D., 1988. Assessment of flood frequency models using empirical distribution function statistics. Water Resources Research 24, 1323–1328. doi:10.1029/WR024i008p01323.

Akerlof, G.A., Dickens, W.T., 1982. The economic consequences of cognitive dissonance. The American economic review 72, 307–319.

Almeida, J., Gonçalves, T.C., 2023. A systematic literature review of investor behavior in the cryptocurrency markets. Journal of Behavioral and Experimental Finance 37, 100785. doi:10.1016/j.jbef.2022.100785.

Amaya, D., Filbien, J.Y., 2015. The similarity of ecb's communication. Finance Research Letters 13, 234–242. doi:10.1016/j.frl.2014.12.006.

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. The Annals of Mathematical Statistics 23, 193–212. doi:10.1214/aoms/1177729437.

Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. Journal of the American Statistical Association 49, 765–769. doi:10.2307/2281537.

Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance 59, 1259–1294. doi:10.1111/j.1540-6261.2004.00662.x.

Araci, D., 2019. Finbert: Financial sentiment analysis with pre-trained language models URL: http://arxiv.org/pdf/1908.10063v1.

Ariel, R.A., 1987. A monthly effect in stock returns. Journal of Financial Economics 18, 161–174. doi:10.1016/0304-405X(87)90066-3.

Avital, M., Leimeister, J.M., Schultze, U., 2014. ECIS 2014 proceedings: 22th European Conference on Information Systems ; Tel Aviv, Israel, June 9-11, 2014. AIS Electronic Library. URL: http://aisel.aisnet.org/ecis2014/.

Baek, C., Elbeck, M., 2015. Bitcoins as an investment or speculative vehicle? a first look. Applied Economics Letters 22, 30–34.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of economic perspectives 21, 129–152.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680. doi:10.1111/j.1540-6261.2006.00885.x.

Balcilar, M., Bouri, E., Gupta, R., Roubaud, D., 2017. Can volume predict bitcoin returns and volatility? a quantiles-based approach. Economic Modelling 64, 74–81. doi:10.1016/J.ECONMOD.2017.03.019.

Balkema, A.A., de Haan, L., 1974. Residual life time at great age. The Annals of Probability 2, 792–804. doi:10.1214/aop/1176996548.

Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. The Journal of Finance 55, 773–806. doi:10.1111/0022-1082.00226.

Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785–818. doi:10.1093/rfs/hhm079.

Bariviera, A.F., Basgall, M.J., Hasperué, W., Naiouf, M., 2017. Some stylized facts of the bitcoin market. Physica A: Statistical Mechanics and its Applications 484, 82–90. doi:10.1016/j.physa.2017.04.159.

Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51, 1173–1182. doi:10.1037/0022-3514.51.6.1173.

Basel Commitee on Banking Supervision, 2004. International convergence of capital measurement and capital standards - a revised framework.

Basel Commitee on Banking Supervision, 2009. Observed range of practice in key elements of advanced measurement approaches (ama).

Baur, D.G., Dimpfl, T., Kuck, K., 2018a. Bitcoin, gold and the us dollar–a replication and extension. Finance Research Letters 25, 103–110.

Baur, D.G., Hong, K., Lee, A.D., 2018b. Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money 54, 177–189. doi:10.1016/j.intfin.2017.12.004.

Behrendt, S., Schmidt, A., 2018. The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. Journal of Banking & Finance 96, 355–367. doi:10.1016/j.jbankfin.2018.09.016.

Benartzi, S., Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. The Quarterly Journal of Economics 110, 73–92. doi:10.2307/2118511.

Bianchi, D., 2020. Cryptocurrencies as an asset class? an empirical assessment. The Journal of Alternative Investments 23, 162–179. doi:10.3905/jai.2020.1.105.

Bienaymé, I.J., 1874. Sur une question de probabilités. Bulletin de la Société Mathématique de France 2, 153–154.

Black, F., 1986. Noise. The Journal of Finance 41, 528–543. doi:10.1111/j.1540-6261.1986.tb04513.x.

Blyth, C.R., 1970. On the inference and decision models of statistics. The Annals of Mathematical Statistics 41, 1034–1058.

Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146. doi:10.1162/tacl{\textunderscore}a{\textunderscore}00051.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1–8. doi:10.1016/j.jocs.2010.12.007.

de Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? The Journal of Finance 40, 793–805. doi:10.1111/j.1540-6261.1985.tb05004.x.

Boos, D.D., 1982. Minimum anderson-darling estimation. Communication in Statistics-Theory and Methods 11, 2747–2774.

Börner, C.J., Hoffmann, I., Krettek, J., Schmitz, T., 2022. Bitcoin: like a satellite or always hardcore? a core–satellite identification in the cryptocurrency market. Journal of Asset Management 23, 310–321. doi:10.1057/s41260-022-00267-z.

Boudreaux, D.O., 1995. The monthly effect in international stock markets: evidence and implications. 1 ed., Journal of Financial and Strategic Decisions.

Boulu-Reshef, B., Bruneau, C., Nicolas, M., Renault, T., 2023. An experimental analysis of investor sentiment, in: Bourghelle, D., Grandin, P., Jawadi, F., Rozin, P. (Eds.), Behavioral Finance and Asset Prices. Springer International Publishing, Cham. Contributions to Finance and Accounting, pp. 131–154. doi:10.1007/978-3-031-24486-5$\backslash$textunderscore.

Bouri, E., Gupta, R., Tiwari, A.K., Roubaud, D., 2017a. Does bitcoin hedge global uncertainty? evidence from wavelet-based quantile-in-quantile regressions. Finance Research Letters 23, 87–95. doi:10.1016/j.frl.2017.02.009.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I., 2017b. On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? Finance Research Letters 20, 192–198. doi:10.1016/j.frl.2016.09.025.

Branch, B., 1977. A tax loss trading rule. The Journal of Business 50, 198. doi:10.1086/295930.

Brauneis, A., Mestel, R., 2018. Price discovery of cryptocurrencies: Bitcoin and beyond. Economics Letters 165, 58–61.

Breaban, A., Noussair, C.N., 2018. Emotional state and market behavior. Review of Finance 22, 279–309. doi:10.1093/rof/rfx022.

Brière, M., Oosterlinck, K., Szafarz, A., 2015. Virtual currency, tangible return: Porfolio diversification with bitcoin. Journal of Asset Management 16, 365–373.

Broadstock, D.C., Zhang, D., 2019. Social-media and intraday stock returns: The pricing power of sentiment. Finance Research Letters 30, 116–123. doi:10.1016/j.frl.2019.03.030.

Brock, W., LAKONISHOK, J., LeBARON, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance 47, 1731–1764. doi:10.1111/j.1540-6261.1992.tb04681.x.

Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. The Journal of Business 78, 405–440. doi:10.1086/427633.

Cade, N.L., 2018. Corporate social media: How two-way disclosure channels influence investors. Accounting, Organizations and Society 68-69, 63–79. doi:10.1016/j.aos.2018.03.004.

Cantelli, F.P., 1933. Sulla determinazione empirica delle leggi di probabilità. Giornale dell'Istituto Italiano degli Attuari 1933, 421–424.

Cao, H.H., Coval, J.D., Hirshleifer, D., 2002. Sidelined investors, trading-generated news, and security returns. Review of Financial Studies 15, 615–648. doi:10.1093/rfs/15.2.615.

Cao, M., Wei, J., 2005. Stock market returns: A note on temperature anomaly. Journal of Banking & Finance 29, 1559–1573. doi:10.1016/j.jbankfin.2004.06.028.

Caporale, G.M., Gil-Alana, L., Plastun, A., 2018. Persistence in the cryptocurrency market. Research in International Business and Finance 46, 141–148. doi:10.1016/j.ribaf.2018.01.002.

Cer, D., Yang, Y., Kong, S.y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.H., Strope, B., Kurzweil, R., 2018. Universal sentence encoder doi:10.48550/arXiv.1803.11175.

Chang, S.C., Chen, S.S., Chou, R.K., Lin, Y.H., 2008. Weather and intraday patterns in stock returns and trading activity. Journal of Banking & Finance 32, 1754–1766. doi:10.1016/j.jbankfin.2007.12.007.

Chatterjee, A., Maniam, B., 2011. Market anomalies revisited. Journal of Applied Business Research (JABR) 13, 47. doi:10.19030/jabr.v13i4.5740.

Cheah, E.T., Fry, J., 2015. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. Economics Letters 130, 32–36. doi:10.1016/j.econlet.2015.02.029.

Chen, H., De, P., Hu, Y., Hwang, B.H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. Review of Financial Studies 27, 1367–1403. doi:10.1093/rfs/hhu001.

Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit tests for the generalized pareto distribution. Technometrics 43, 478–484.

Clogg, C.C., Petkova, E., Haritou, A., 1995. Statistical methods for comparing regression coefficients between models. American Journal of Sociology 100, 1261–1293. URL: https://www.jstor.org/stable/2782277.

Cookson, J.A., Niessner, M., 2020. Why don't we agree? evidence from a social network of investors. The Journal of Finance 75, 173–228. doi:10.1111/jofi.12852.

Corbet, S., Lucey, B., Urquhart, A., Yarovaya, L., 2019. Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis 62, 182–199. doi:10.1016/j.irfa.2018.09.003.

Corbet, S., Lucey, B., Yarovaya, L., 2018a. Datestamping the bitcoin and ethereum bubbles. Finance Research Letters 26, 81–88.

Corbet, S., Lucey, B.M., Peat, M., Vigne, S., 2018b. Bitcoin futures - what use are they? Economics Letters 172, 23–27.

Corbet, S., Meegan, A., Larkin, C., Lucey, B., Yarovaya, L., 2018c. Exploring the dynamic relationships between cryptocurrencies and other financial assets. Economics Letters 165, 28–34. doi:10.1016/j.econlet.2018.01.004.

Cramér, H., 1928. On the composition of elementary errors: Second paper: Statistical applications. Scandinavian Actuarial Journal 1, 141–180.

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all fears investor sentiment and asset prices. Review of Financial Studies 28, 1–32. doi:10.1093/rfs/hhu072.

Das, S.R., Chen, M.Y., 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. Management Science 53, 1375–1388. doi:10.1287/mnsc.1070.0704.

Davison, A.C., 1984. Modelling excess over high threshold, with an application: Statistical Extremes and Applications. Reidel Publishing Company, Dordrecht, Netherlands.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with comments). Journal of the Royal Statistical Society, Series B (Methodological) 52, 393–442.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. Journal of political economy 98, 703–738. doi:10.1086/261703.

DeMarzo, P.M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. The Quarterly Journal of Economics 118, 909–968. doi:10.1162/00335530360698469.

Desai, H., Rajgopal, S., Venkatachalam, M., 2004. Value-glamour and accruals mispricing: One anomaly or two? The Accounting Review 79, 355–385. doi:10.2308/accr.2004.79.2.355.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding URL: http://arxiv.org/pdf/1810.04805v2.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dougal, C., Engelberg, J., García, D., Parsons, C.A., 2012. Journalists and the stock market. Review of Financial Studies 25, 639–679. doi:10.1093/rfs/hhr133.

Dwyer, G.P., 2015. The economics of bitcoin and similar private digital currencies. Journal of Financial Stability 17, 81–91.

Dyhrberg, A.H., 2016a. Bitcoin, gold and the dollar–a garch volatility analysis. Finance Research Letters 16, 85–92. doi:10.1016/j.frl.2015.10.008.

Dyhrberg, A.H., 2016b. Hedging capabilities of bitcoin. is it the virtual gold? Finance Research Letters 16, 139–144. URL: https://www.sciencedirect.com/science/article/pii/S1544612315001208, doi:10.1016/j.frl.2015.10.025.

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., Baronchelli, A., 2017. Evolutionary dynamics of the cryptocurrency market. The Royal Society Open Science 4.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg. doi:10.1007/978-3-642-33483-2.

Engelberg, J.E., Reed, A.V., Ringgenberg, M.C., 2012. How are shorts informed? Journal of Financial Economics 105, 260–278. doi:10.1016/j.jfineco.2012.03.001.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society 50, 987–1007. doi:10.2307/1912773.

Engle, R.F., 2002. A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20, 339–350.

European Parliament, 2009. Directive 2009/138/ec of the european parliament and of the council of 25 november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii). Official Journal of the European Union 52, 1–155.

European Parliament, 2013a. Directive 2013/36/eu of the european parliament and of the council of 26 june 2013 on access to the activity of credit institutions and the prudential

supervision of credit institutions and investment firms, amending directive 2002/87/ec and repealing directives 2006/48/ec and 2006/49/ecand investment firms, amending directive 2002/87/ec and repealing directives 2006/48/ec and 2006/49/ec. Official Journal of the European Union 56, 338–436.

European Parliament, 2013b. Regulation (eu) no 575/2013 of the european parliament and of the council of 26 june 2013 on prudential requirements for credit institutions and investment firms and amending regulation (eu) no 648/2012. Official Journal of the European Union 56, 1–337.

Fama, E.F., 1965. The behavior of stock-market prices. The Journal of Business 38, 34–105.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25, 383. doi:10.2307/2325486.

Fama, E.F., French, K.R., 1992. The cross–section of expected stock returns. The Journal of Finance 47, 427–465. doi:10.1111/j.1540-6261.1992.tb04398.x.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S., 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm, in: Palmer, M., Hwa, R., Riedel, S. (Eds.), Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 1615–1625. doi:10.18653/v1/D17-1169.

Feldman, R., Govindaraj, S., Livnat, J., Segal, B., 2010. Management's tone change, post earnings announcement drift and accruals. Review of Accounting Studies 15, 915–953. doi:10.1007/s11142-009-9111-x.

Ferguson, T.S., 1967. Mathematical statistics: A decision theoretic approach. Probability and mathematical statistics a series of monographs and textbooks, 1, Academic Press, New York.

Festinger, L., Riecken, H.W., Schachter, S., 1956. When prophecy fails. University of Minnesota Press, Minneapolis. doi:10.1037/10030-000.

Frank, J.D., 1935. Some psychological determinants of the level of aspiration. The American Journal of Psychology 47, 285. doi:10.2307/1415832.

Fry, J., Cheah, E.T., 2016. Negative bubbles and shocks in cryptocurrency markets. International Review of Financial Analysis 47, 343–352. doi:10.1016/j.irfa.2016.02.008.

Fu, C., 2022. Behavioural finance: A synthetic review of literature and future development. SSRN Electronic Journal doi:10.2139/ssrn.4107830.

Gandal, N., Hamrick, J.T., Moore, T., Oberman, T., 2018. Price manipulation in the bitcoin ecosystem. Journal of Monetary Economics 95, 86–96.

Gao, B., Yang, C., 2017. Forecasting stock index futures returns with mixed-frequency sentiment. International Review of Economics & Finance 49, 69–83. doi:10.1016/j.iref.2017.01.020.

Gervais, S., Odean, T., 2001. Learning to be overconfident. Review of Financial Studies 14, 1–27. doi:10.1093/rfs/14.1.1.

Giannini, R., Irvine, P., Shu, T., 2018. Nonlocal disadvantage: An examination of social media sentiment. The Review of Asset Pricing Studies 8, 293–336. doi:10.1093/rapstu/rax020.

Gibbons, J.D., Chakraborti, S., 2011. Nonparametric statistical inference. volume 198 of *Statistics*. 5. ed. ed., Chapman & Hall/CRC Press, Boca Raton, Fla.

Gibbons, M.R., Hess, P., 1981. Day of the week effects and assets returns. Journal of Business 54, 579–596.

Gkillas, K., Bekiros, S., Siriopoulos, C., 2018. Extreme correlation in cryptocurrency markets. SSRN Electronic Journal doi:10.2139/ssrn.3180934.

Gkillas, K., Katsiampa, P., 2018. An application of extreme value theory to cryptocurrencies. Economics Letters 164, 109–111. doi:10.1016/j.econlet.2018.01.020.

Glas, T.N., 2019. Investments in cryptocurrencies: Handle with care! The Journal of Alternative Investments 22, 96–113.

Glaser, F., Zimmarmann, K., Haferhorn, M., Weber, M.C., Siering, M., 2014. Bitcoin - asset or currency? revealing users' hidden intentions .

Glivenko, V., 1933. Sulla determinazione empirica delle leggi di probabilità. Giornale dell'Istituto Italiano degli Attuari 1933, 92–99.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics 44, 423–453.

Goczek, Ł., Skliarov, I., 2019. What drives the bitcoin price? a factor augmented error correction mechanism investigation. Applied Economics 51, 6393–6410. doi:10.1080/00036846.2019.1619021.

Guégan, D., Renault, T., 2021. Does investor sentiment on social media provide robust information for bitcoin returns predictability? Finance Research Letters 38, 101494. doi:10.1016/j.frl.2020.101494.

Hales, J., Kuang, X.I., Venkataraman, S., 2011. Who believes the hype? an experimental examination of how language affects investor judgments. Journal of Accounting Research 49, 223–255. doi:10.1111/j.1475-679X.2010.00394.x.

Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. The Annals of Statistics 13, 70–84.

Hayes, A.S., 2017. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. Telematics & Informatics 34, 1308–1321. doi:10.1016/j.tele.2016.05.005.

Hayes, A.S., 2019. Bitcoin price and its marginal cost of production: support for a fundamental value. Applied Economics Letters 26, 554–560. doi:10.1080/13504851.2018.1488040.

Henry, E., 2008. Are investors influenced by how earnings press releases are written? Journal of Business Communication 45, 363–407. doi:10.1177/0021943608319388.

Hirshleifer, D., 2001. Investor psychology and asset pricing. The Journal of Finance 56, 1533–1597. doi:10.1111/0022-1082.00379.

Hoffmann, I., Börner, C., 2021. Body and tail: an automated tail-detecting procedure. The Journal of Risk 23, 43–69. doi:10.21314/JOR.2020.447.

Hoffmann, I., Börner, C.J., 2020a. The risk function of the goodness-of-fit tests for tail models. Statistical papers , 1–17.

Hoffmann, I., Börner, C.J., 2020b. Tail models and the statistical limit of accuracy in risk assessment. The Journal of Risk Finance 21, 201–216. doi:10.1108/JRF-11-2019-0217.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655. doi:10.1257/000282802762024700.

Hon, M.T., Tonks, I., 2003. Momentum in the uk stock market. Journal of Multinational Financial Management 13, 43–70. doi:10.1016/S1042-444X(02)00022-1.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29, 339–349. doi:10.2307/1260343.

Hull, J.C., 2018. Options, futures, and other derivatives. Ninth edition, global edition ed., Pearson Education, Harlow.

Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Rau, P.L.P. (Eds.), 2013. Cross-Cultural Design. Cultural Differences in Everyday Life. Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-39137-8.

Hutto, C., Gilbert, E., 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media 8. URL: https://ojs.aaai.org/index.php/icwsm/article/view/14550.

Ingham, G., 2004. The nature of money. economic sociology. perspectives and conversations 5, 18–28. URL: https://EconPapers.repec.org/RePEc:zbw:econso:155831.

Jaffe, J., Westerfield, R., 1989. Is there a monthly effect in stock market returns? Journal of Banking & Finance 13, 237–244. doi:10.1016/0378-4266(89)90062-9.

Jensen, M.C., 1978. Some anomalous evidence regarding market efficiency. Journal of Financial Economics 6, 95–101. doi:10.1016/0304-405x(78)90025-9.

Jockers, M.L., 2015. Syuzhet: Extract sentiment and plot arcs from text URL: https://github.com/mjockers/syuzhet.

Johnson, E.J., Tversky, A., 1983. Affect, generalization, and the perception of risk. Journal of Personality and Social Psychology 45, 20–31. doi:10.1037/0022-3514.45.1.20.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263. doi:10.2307/1914185.

Kakinaka, S., Umeno, K., 2020. Characterizing cryptocurrency market with lévy's stable distributions. Journal of the Physical Society of Japan 89, 024802. doi:10.7566/JPSJ.89.024802.

Kaplanski, G., Levy, H., Veld, C., Veld-Merkoulova, Y., 2015. Do happy people make optimistic investors? The Journal of Financial and Quantitative Analysis 50, 145–168. doi:10.1017/S0022109014000416.

Karaa, R., Slim, S., Goodell, J.W., Goyal, A., Kallinterakis, V., 2021. Do investors feedback trade in the bitcoin—and why? The European Journal of Finance , 1–21doi:10.1080/1351847X.2021.1973054.

Kaya Soylu, P., Okur, M., Çatıkkaş, Ö., Altintig, Z.A., 2020. Long memory in the volatility of selected cryptocurrencies: Bitcoin, ethereum and ripple. Journal of Risk and Financial Management 13, 107. doi:10.3390/jrfm13060107.

Kearney, C., Liu, S., 2014. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis 33, 171–185. doi:10.1016/j.irfa.2014.02.006.

Keim, D., 1985. Dividend yields and stock returns: Implications of abnormal january returns. Journal of Financial Economics 14, 473–489. doi:10.1016/0304-405X(85)90009-1.

Keim, D.B., 1983. Size-related anomalies and stock return seasonality. Journal of Financial Economics 12, 13–32. doi:10.1016/0304-405X(83)90025-9.

Kelton, A.S., Pennington, R.R., 2020. If you tweet, they will follow: Ceo tweets, social capital, and investor say-on-pay judgments. Journal of Information Systems 34, 105–122. doi:10.2308/isys-52449.

Kendall, M.G., Stuart, A., 1977. The advanced theory of statistics: In three volumes. 4 ed., Charles Griffin & Co Ltd, London.

Keynes, J.M., 2011 [1930]. A Treatise on Money. Martino, Eastford.

Kim, S.H., Kim, D., 2014. Investor sentiment from internet message postings and the predictability of stock returns. Journal of Economic Behavior & Organization 107, 708–729. doi:10.1016/j.jebo.2014.04.015.

Kim, S.T., 2020. Bitcoin dilemma: Is popularity destroying value? Finance Research Letters 33, 101228. doi:10.1016/j.frl.2019.07.001.

Kolmogorov, A.N., 1933. Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari , 83–91.

Kürzinger, L.M., Stangor, P., 2024. The relevance and influence of social media posts on investment decisions - an experimental approach based on tweets. SSRN Electronic Journal doi:10.2139/ssrn.4739088.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315. doi:10.2307/1913210.

Lample, G., Conneau, A., 2019. Cross-lingual language model pretraining doi:10.48550/arXiv.1901.07291.

Landauer, T.K., Dumais, S.T., 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104, 211–240. doi:10.1037/0033-295X.104.2.211.

Langer, E.J., Roth, J., 1975. Heads i win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. Journal of Personality and Social Psychology 32, 951–955. doi:10.1037/0022-3514.32.6.951.

Lee, C.C., Lee, J.D., Lee, C.C., 2010. Stock prices and the efficient market hypothesis: Evidence from a panel stationary test with structural breaks. Japan and the World Economy 22, 49–58. doi:10.1016/j.japwor.2009.04.002.

Lee, C.M.C., Shleifer, A., Thaler, R.H., 1991. Investor sentiment and the closed-end fund puzzle. The Journal of Finance 46, 75–109. doi:10.1111/j.1540-6261.1991.tb03746.x.

Lehmann, B.N., 1990. Fads, martingales, and market efficiency. The Quarterly Journal of Economics 105, 1. doi:10.2307/2937816.

Ligon, J.A., 1997. A simultaneous test of competing theories regarding the january effect. Journal of Financial Research 20, 13–32. doi:10.1111/j.1475-6803.1997.tb00234.x.

Liu, B., 2020. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

Liu, Y., Tsyvinski, A., 2021. Risks and returns of cryptocurrency. Review of Financial Studies 34, 2689–2727. doi:10.1093/rfs/hhaa113.

Loughran, T.I., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance 66, 35–65. doi:10.1111/j.1540-6261.2010.01625.x.

Loughran, T.I., McDonald, B., 2016. Textual analysis in accounting and finance: A survey. Journal of Accounting Research 54, 1187–1230. doi:10.1111/1475-679X.12123.

Mai, F., Shan, Z., Bai, Q., Wang, X., Chiang, R.H., 2018. How does social media impact bitcoin value? a test of the silent majority hypothesis. Journal of Management Information Systems 35, 19–52. doi:10.1080/07421222.2018.1440774.

Majoros, S., Zempléni, A., 2018. Multivariate stable distributions and their applications for modelling cryptocurrency-returns. Working Paper 2018. URL: http://arxiv.org/pdf/1810.09521v1.

Malkiel, B.G., 2005. Reflections on the efficient market hypothesis: 30 years later. The Financial Review 40, 1–9. doi:10.1111/j.0732-8516.2005.00090.x.

McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative risk management: Concepts, techniques and tools. Princeton series in finance. revised edition ed., Princeton University Press, Princeton and Oxford.

Mechura, M.B., 2016. Lemmatization list - english (en) URL: http://www.lexiconista.com.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013a. Efficient estimation of word representations in vector space doi:10.48550/arXiv.1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013b. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems 26.

Miller, B.P., 2010. The effects of reporting complexity on small and large investor trading. The Accounting Review 85, 2107–2143. doi:10.2308/accr.00000001.

Miller, D.T., Ross, M., 1975. Self-serving biases in the attribution of causality: Fact or fiction? Psychological Bulletin 82, 213–225. doi:10.1037/h0076486.

von Mises, R.E., 1931. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Franz Deuticke, Leipzig und Wien.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., Trajanov, D., 2020. Evaluation of sentiment analysis in finance: From lexicons to transformers. Ieee Access 8, 131662–131682. doi:10.1109/Access.2020.3009626.

Mitchell, M.L., Stafford, E., 2000. Managerial decisions and long–term stock price performance. The Journal of Business 73, 287–329. doi:10.1086/209645.

Mohammad, S.M., Turney, P.D., 2013. Crowdsourcing a word-emotion association lexicon. Computational Intelligence 29, 436–465. doi:10.1111/j.1467-8640.2012.00460.x.

Naeem, M.A., Mbarki, I., Shahzad, S.J.H., 2021. Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears. International Review of Economics & Finance 73, 496–514. doi:10.1016/j.iref.2021.01.008.

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system. Consulted 1, 2012.

Neal, R., Wheatley, S.M., 1998. Do measures of investor sentiment predict returns? The Journal of Financial and Quantitative Analysis 33, 523. doi:10.2307/2331130.

Nolan, J.P., 2020. Univariate Stable Distributions: Models for Heavy Tailed Data. Springer Series in Operations Research and Financial Engineering. 1st ed. 2020 ed., Springer International Publishing and Imprint: Springer, Cham. doi:10.1007/978-3-030-52915-4.

Osterrieder, J., Lorenz, J., Strika, M., 2017. Bitcoin and cryptocurrencies - not for the faint-hearted. International Finance and Banking 13, 145–193.

Palomino, F., 1996. Noise trading in small markets. The Journal of Finance 51, 1537–1550. doi:10.1111/j.1540-6261.1996.tb04079.x.

Parr, W.C., Schucany, W.R., 1980. Minimum distance and robust estimation. Journal of the American Statistical Association 75, 616–624.

Paternoster, R., Brame, R., Mazerolle, P., Piquero, A., 1998. Using the correct statistical test for equality of regression coefficients. Criminology 36, 859–866. doi:10.1111/j.1745-9125.1998.tb01268.x.

Peng, Y., Albuquerque, P.H.M., Camboim de Sá, J.M., Padula, A.J.A., Montenegro, M.R., 2018. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. Expert Systems with Applications 97, 177–192. doi:10.1016/j.eswa.2017.12.004.

Pennington, J., Socher, R., Manning, C., 2014. Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Stroudsburg, PA, USA. doi:10.3115/v1/d14-1162.

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. doi:10.48550/arXiv.1802.05365.

Picault, M., Renault, T., 2017. Words are not all created equal: A new measure of ecb communication. Journal of International Money and Finance 79, 136–156. doi:10.1016/j.jimonfin.2017.09.005.

Pickands III, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131. doi:10.1214/aos/1176343003.

Plutchik, R., 1984. A general psychoevolutionary theory of emotion, in: Plutchik, R. (Ed.), Emotion: Theory, research, and experience. Acad. Press, Orlando, pp. 3–33. doi:10.1016/b978-0-12-558701-3.50007-7.

Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R., Lightfoot, G., 2015. Price fluctuations and the use of bitcoin: An empirical inquiry. International Journal of Electronic Commerce 20, 9–49.

Pontiff, J., 1996. Costly arbitrage: Evidence from closed-end funds. The Quarterly Journal of Economics 111, 1135–1151. doi:10.2307/2946710.

Pratt, J.W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122. doi:10.2307/1913738.

Qiu, L., Welch, I., 2004. Investor sentiment measures. National Bureau of Economic Research URL: https://www.nber.org/papers/w10794, doi:10.3386/w10794.

Renault, T., 2017. Intraday online investor sentiment and return patterns in the u.s. stock market. Journal of Banking & Finance 84, 25–40. doi:10.1016/j.jbankfin.2017.07.002.

Rennekamp, K.M., Witz, P.D., 2021. Linguistic formality and audience engagement: Investors' reactions to characteristics of social media disclosures*. Contemporary Accounting Research 38, 1748–1781. doi:10.1111/1911-3846.12661.

Rinker, T.W., 2018. Textstem: Tools for stemming and lemmatizing text URL: http://github.com/trinker/textstem.

Ritter, J.R., 2003. Behavioral finance. Pacific-Basin Finance Journal 11, 429–437. doi:10.1016/S0927-538X(03)00048-9.

Rogalski, R.J., 1984. New findings regarding day–of–the–week returns over trading and non–trading periods: A note. The Journal of Finance 39, 1603–1614. doi:10.1111/j.1540-6261.1984.tb04927.x.

Rosa, C., Verga, G., 2007. On the consistency and effectiveness of central bank communication: Evidence from the ecb. European Journal of Political Economy 23, 146–175. doi:10.1016/j.ejpoleco.2006.09.016.

Rozeff, M.S., Kinney, W.R., 1976. Capital market seasonality: The case of stock returns. Journal of Financial Economics 3, 379–402. doi:10.1016/0304-405X(76)90028-3.

Rubinstein, M., 2001. Rational markets: Yes or no? the affirmative case. Financial Analysts Journal 57, 15–29. doi:10.2469/faj.v57.n3.2447.

Sahlgren, M., 2006. The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. volume 44 of *SICS dissertation series*. Dep. of Linguistics, Stockholm Univ, Stockholm.

Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., Okruszek, L., 2021. Detecting formal thought disorder by deep contextualized word representations. Psychiatry research 304, 114135. doi:10.1016/j.psychres.2021.114135.

Schmitz, T., Hoffmann, I., 2021. Re-evaluating cryptocurrencies' contribution to portfolio diversification – a portfolio analysis with special focus on german investors. Working Paper URL: http://dx.doi.org/10.2139/ssrn.3625458.

Scholes, M., 1969. A Test of the Competitive Hypothesis: : The Market for New Issues and Secondary Offerings. Unpublished PH.D. URL: https://scholar.google.com/citations?user=sfkqfdgaaaaj&hl=de&oi=sra.

Selgin, G., 2015. Synthetic commodity money. Journal of Financial Stability , 92–99.

Sewell, M., 2011. History of the efficient market hypothesis. RN/11/04. URL: http://www.cs.ucl.ac.uk/fileadmin/ucl-cs/images/research_student_information/rn_11_04.pdf.

Shefrin, H., Statman, M., 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. The Journal of Finance 40, 777. doi:10.2307/2327802.

Shi, N., 2016. A new proof-of-work mechanism for bitcoin. Financial Innovation 2. doi:10.1186/s40854-016-0045-6.

Shiller, R.C., 2000. Irrational exuberance. Philosophy and Public Policy Quarterly 20, 18–23.

Shiller, R.J., 2003. From efficient markets theory to behavioral finance. The Journal of Economic Perspectives 17, 83–104. doi:10.1257/089533003321164967.

Shleifer, A., 2000. Inefficient Markets: An Introduction to Behavioural Finance. Clarendon Lectures in Economics Ser, Oxford University Press, Incorporated, Oxford.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. The Journal of Finance 52, 35–55. doi:10.1111/j.1540-6261.1997.tb03807.x.

Shorack, G.R., Wellner, J.A., 2009. Empirical processes with applications to statistics. volume 59 of *Classics in applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, Pa.

Shrotryia, V.K., Kalra, H., 2022. Herding in the crypto market: a diagnosis of heavy distribution tails. Review of Behavioral Finance 14, 566–587. doi:10.1108/RBF-02-2021-0021.

Singer, E., 2010. The use of incentives to reduce nonresponse in houshold surveys, in: Groves, R.M. (Ed.), Survey nonresponse. Wiley, New York, NY. A Wiley-Interscience publication.

Sixt, E., 2017. Bitcoins und andere dezentrale Transaktionssysteme. Springer Fachmedien Wiesbaden, Wiesbaden. doi:10.1007/978-3-658-02844-2.

Smirlock, M., Starks, L., 1986. Day-of-the-week and intraday effects in stock returns. Journal of Financial Economics 17, 197–210. doi:10.1016/0304-405X(86)90011-5.

Smirnov, N.V., 1936. Sur la distribution de w^2-criterion (critérion de r. von mises). Comptes Rendus de l'Académie des Sciences Paris 202, 449–452.

Smirnov, N.V., 1948. Table for estimating the goodness of fit of empirical distributions. The Annals of Mathematical Statistics 19, 279–281. doi:10.1214/aoms/1177730256.

Smith, R.L., 1984. Threshold methods for sample extremes: Statistical extremes and applications. D. Reidel Publishing Company 1984, 621–638.

Smith, R.L., 1985. Maximum likelihood estimations in a class of nonregular cases. Biometrika 72, 67–90.

Sprenger, T.O., Tumasjan, A., Sandner, P.G., Welpe, I.M., 2014. Tweets and trades: the information content of stock microblogs. European Financial Management 20, 926–957. doi:10.1111/j.1468-036X.2013.12007.x.

Stangor, P., Kuerzinger, L., 2021. Measuring investor sentiment from social media data - an emotional approach. SSRN Electronic Journal doi:10.2139/ssrn.3976224.

Stephens, M.A., 1974. Edf statistics for goodness of fit and some comparisons. Journal of the American Statistical Association 69, 730–737. doi:10.2307/2286009.

Stiglitz, J.E., Grossman, S.J., 1980. On the impossibility of informationally efficient markets. The American economic review 70, 393–408. doi:10.7916/D8765R99.

Stovall, R.H., 1989. The super bowl predictor. Financial World 158.

Sun, L., Najand, M., Shen, J., 2016. Stock return predictability and investor sentiment: A high-frequency perspective. Journal of Banking & Finance 73, 147–164. doi:10.1016/j.jbankfin.2016.09.010.

Tan, H.T., Ying Wang, E., Zhou, B.O., 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. Journal of Accounting Research 52, 273–302. doi:10.1111/1475-679X.12039.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance 62, 1139–1168. doi:10.1111/j.1540-6261.2007.01232.x.

Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance 63, 1437–1467. doi:10.1111/j.1540-6261.2008.01362.x.

Thaler, R., 1980. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization 1, 39–60. doi:10.1016/0167-2681(80)90051-7.

Tirole, J., 1982. On the possibility of speculation under rational expectations. Econometrica 50, 1163. doi:10.2307/1911868.

Trimborn, S., Li, M., Härdle, W.K., 2020. Investing with cryptocurrencies - a liquidity constrained investments approach. Journal of Financial Econometrics 18, 280–306.

Trueman, B., 1994. Analyst forecasts and herding behavior. Review of Financial Studies 7, 97–124. doi:10.1093/rfs/7.1.97.

Tukey, J.W., 1992. Exploratory data analysis. Addison-Wesley series in behavioral science: quantitative methods. 16. print ed., Addison-Wesley, Reading, Mass. u.a.

Turney, P.D., Pantel, P., 2010. From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research 37, 141–188. doi:10.1613/jair.2934.

Tversky, A., Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5, 207–232. doi:10.1016/0010-0285(73)90033-9.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science (New York, N.Y.) 185, 1124–1131. doi:10.1126/science.185.4157.1124.

Urquhart, A., 2018. What causes the attention of bitcoin? Economics Letters 166, 40–44.

Vamossy, D.F., 2024. Social media emotions and market behavior doi:10.48550/arXiv.2404.03792.

Vamossy, D.F., Skog, R., 2023. Emtract: Investor emotions and market behavior. SSRN Electronic Journal doi:10.2139/ssrn.3975884.

van Alstyne, M., 2014. Why bitcoin has value. Communications of the ACM 57, 30–33.

van Bommel, J., 2003. Rumors. The Journal of Finance 58, 1499–1520. doi:10.1111/1540-6261.00575.

van Montfort, M.A.J., Witter, J.V., 1985. Testing exponentiality against generalized pareto distribution. Journal of Hydrology 78, 305–315.

Wachtel, S.B., 1942. Certain observations on seasonal movements in stock prices. Journal of Business of the University of Chicago 15, 184. doi:10.1086/232617.

Weber, B., 2014. Bitcoin and the legitimacy crisis of money. Cambridge Journal of Economics 40, 17–41.

Wingreen, S.C., Kavanagh, D., Dylan-Ennis, P., Miscione, G., 2020. Sources of cryptocurrency value systems: The case of bitcoin. International Journal of Electronic Commerce 24, 474–496. doi:10.1080/10864415.2020.1806469.

Wolfowitz, J., 1957. The minimum distance method. The Annals of Applied Statistics 28, 75–88.

Wooldridge, J.M., 2020. Introductory econometrics: A modern approach. Seventh edition ed., CENGAGE, Boston, MA.

Wurgler, J., 2000. Financial markets and the allocation of capital. Journal of Financial Economics 58, 187–214. doi:10.1016/S0304-405X(00)00070-2.

Xiong, X., Luo, C., Zhang, Y., Lin, S., 2019. Do stock bulletin board systems (bbs) contain useful information? a viewpoint of interaction between bbs quality and predicting ability. Accounting & Finance 58, 1385–1411. doi:10.1111/acfi.12448.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V., 2019. Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems 32. doi:10.48550/arXiv.1906.08237.

Yelowitz, A., Wilson, M., 2015. Characteristics of bitcoin users: an analysis of google search data. Applied Economics Letters 22, 1030–1036.

Yermack, D., 2015. Is bitcoin a real currency? an economic appraisal, in: Lee Kuo Chuen, D. (Ed.), Handbook of Digital Currency. Academic Press, Amsterdam, pp. 31–43.

Zhao, X., Lynch, J.G., Chen, Q., 2010. Reconsidering baron and kenny: Myths and truths about mediation analysis. Journal of Consumer Research 37, 197–206. doi:10.1086/651257.

Statutory Declaration

I, Lars Manfred Kürzinger, swear that I am writing this dissertation independently and without inadmissible outside help, taking into account the 'principles for ensuring good scientific practice at the Heinrich-Heine University Düsseldorf'.

Düsseldorf, June 4, 2024

_____

Signature