Journal of Banking and Financial Economics

No 2(16)2021

University of Warsaw Faculty of Management

ISSN 2353-6845

Editorial Office: Faculty of Management, University of Warsaw, Szturmowa Str. 1/3, Postal Code 02-678 Warsaw, Email: jbfe@wz.uw.edu.pl

Publisher: Uniwersytet Warszawski, Krakowskie Przedmieście 26/28, 00-927 Warsaw

Webpage of Journal of Banking and Financial Economics:

https://jbfe.wz.uw.edu.pl/resources/html/cms/MAINPAGE

Overview

The ***Journal of Banking and Financial Economics*** **(JBFE)** is an open access journal. The submission of manuscripts in free of fee payment. This journal follows a double-blind reviewing procedure.

Aims and Scope

JBFE publishes high quality empirical and theoretical papers spanning all the major research fields in banking and financial economics. The aim of the journal is to provide an outlet for the increasing flow of scholarly research concerning banking, financial institutions and the money and capital markets within which they function. The journal also focuses on interrelations of financial variables, such as prices, interest rates and shares and concentrates on influences of real economic variables on financial ones and vice versa. Macro-financial policy issues, including comparative financial systems, the globalization of financial services, and the impact of these phenomena on economic growth and financial stability, are also within the *JBFE*'s scope of interest. The Journal seeks to promote research that enriches the profession's understanding of the above mentioned as well as to promote the formulation of sound public policies.

Main subjects covered include, e.g.: [1] ***Valuation of assets***: Accounting and financial reporting; Asset pricing; Stochastic models for asset and instrument prices; [2] ***Financial markets and instruments:*** Alternative investments; Commodity and energy markets; Derivatives, stocks and bonds markets; Money markets and instruments; Currency markets; [3] ***Financial institutions, services and regulation***: Banking efficiency; Banking regulation; Bank solvency and capital structure; Credit rating and scoring; Regulation of financial markets and institutions; Systemic risk; [4] ***Corporate finance and governance****:* Behavioral finance; Empirical finance; Financial applications of decision theory or game theory; Financial applications of simulation or numerical methods; Financial forecasting; Financial risk management and analysis; Portfolio optimization and trading.

Special Issues

JBFE welcomes publication of Special Issues, whose aim is to bring together and integrate work on a specific theme; open up a previously under-researched area; or bridge the gap between formerly rather separate research communities, who have been focusing on similar or related topics. Thematic issues are strongly preferred to a group of loosely connected papers.

Proposals of Special Issues should be submitted to at jbfe@wz.uw.edu.pl. All proposals are being reviewed by the Editorial Team on the basis of certain criteria that include e.g.: the novelty, importance and topicality of the theme; whether the papers will form an integrated whole; and the overall 'added value' of a Special Issue.

CONTENTS

Sentiment Analysis of German Texts in Finance: Improving and Testing the BPW Dictionary

Matthias Pöferlein[1]
University of Bayreuth, Germany
matthias.poeferlein@uni-bayreuth.de
https://orcid.org/0000-0002-8967-6087

Received: 4 November 2021 / Revised: 13 December 2021 / Accepted: 15 December 2021 / Published online: 20 December 2021

ABSTRACT

Using the dictionary-based approach to measure the sentiment of finance-related texts is primarily focused on English-speaking content. This is due to the need for domain-specific dictionaries and the primary availability of those in English. Through the contribution of Bannier et al. (2019b), the first finance-related dictionary is available for the German language. Because of the novelty of this dictionary, this paper proposes several reforms and extensions of the original word lists. Additionally, I tested multiple measurements of sentiment. I show that using the edited and extended dictionary to calculate a relative measurement of sentiment, central assumptions regarding textual analysis can be fulfilled and more significant relations between the sentiment of a speech by a CEO at the Annual General Meeting and subsequent abnormal stock returns can be calculated.

JEL Classification: G12; G14

Keywords: textual analysis, textual sentiment, sentiment analysis, content analysis, annual general meeting, CEO speeches.

1. INTRODUCTION

In recent years, textual analysis has become an important part of accounting and finance research. This is due to the fact that the availability and quantity of digitally available texts are constantly increasing. Additionally, the information encoded in those texts in the form of sentiment can be obtained in an easier and more targeted way through recent developments in the field of textual analysis (Bannier et al., 2019b, pp. 82f.; Gentzkow et al., 2019, p. 535; Loughran & McDonald, 2015, p. 1).

Algaba et al. (2020, p. 2) define sentiment "[…] as the disposition of an entity toward an entity, expressed via a certain medium. […] This disposition can be conveyed numerically but is primarily expressed qualitatively through text, audio, and visual media." The two most common methods for transforming qualitative sentiment data into quantitative sentiment variables are the dictionary-based approach (also referred to as bag-of-words) and machine learning (Kearney

[1] Matthias Pöferlein – University of Bayreuth, Germany, matthias.poeferlein@uni-bayreuth.de, University of Bayreuth, Chair of Finance and Banking, Universitätsstr. 30, 95447 Bayreuth, Germany, Tel.: +49 (0)921/55-6258, Fax: +49 (0)921/55-6272.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

& Liu, 2014, pp. 174f.). The dictionary-based approach is a rule-based approach that uses an algorithm to classify a text's words or phrases into different categories based on predefined rules or categories like dictionaries[2] (Li, 2010, p. 146). More specifically, the dictionary assigns words into different categories like positive or negative. Using the total count of positive, negative, and all words, several measurements of sentiment can be calculated (Loughran & McDonald, 2015, p. 1). The machine learning or statistical approach relies on statistical techniques to classify the content of documents (Kearney & Liu, 2014, p. 175; Li, 2010, p. 146).

When using the dictionary-based approach, the chosen dictionary has a specific importance (Bannier et al., 2019b, p. 80; Loughran & McDonald, 2015, p. 1). As described in the following section, the newly developed word list provided by Bannier et al. (2019b) (BPW Dictionary) gives researchers the possibility to analyze German-speaking texts in finance in a more targeted way.

Due to the novelty of this BPW Dictionary, I propose several reforms and extensions with the objective of improving its performance. Therefore, the main hypothesis of this paper is that the edited version of the BPW (BPW_N) can improve results compared to its original (BPW_O). So far, the BPW Dictionary has been used primarily to analyze the market reaction to the sentiment of CEO speeches held at the Annual General Meeting (AGM) of German stock companies (Bannier et al., 2017, 2019a). Therefore, this paper also uses comparable speeches for testing the possible improvements.

As stated in the following course of this paper, there are several different possibilities to measure the sentiment of textual documents in a dictionary-based approach. Given the fact that this is the first German domain-specific dictionary for the field of finance, the additional research question is which sentiment measure is the most appropriate for measuring the tone of textual documents in the field of finance using a German domain-specific dictionary. This topic is especially relevant, given the previous use of exclusively four different measurements of sentiment using the BPW Dictionary (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

The contribution of this paper to the literature on textual analysis of German texts is the extension and reform of the only existing German finance-related dictionary and testing the performance of the original against the new dictionary. Additionally, the suitability of the primarily used measures of sentiment in a business context is analyzed. This should make it possible for researchers to measure the sentiment of German texts in finance more accurately and more thoroughly.

The paper proceeds as follows. In the following section, I will give a short review of the relevant literature regarding textual analysis with a particular focus on analyzing financial texts. The data and the parsing procedure applied to it, as well as the used dictionaries form the third section. The used measurements of sentiment and the empirical approach to obtain the results given in section five are presented in the fourth section. Section six concludes.

2. LITERATURE REVIEW

The extensive field of textual analysis in finance is ideally pictured in the surveys of Kearney and Liu (2014) and the online appendix of Bannier et al. (2019b). Other important surveys giving additional information and areas of caution regarding textual analysis in finance are Algaba et al. (2020) and Loughran and McDonald (2016).

One of the first steps in measuring the tone of a text is selecting a dictionary or word list (Loughran & McDonald, 2015, p. 1). According to Loughran and McDonald (2016, p. 1200), four different word lists have been primarily used by researchers classifying English finance-related

[2] As stated in Loughran and McDonald (2015, p. 10), the terms dictionary and word list are used interchangeably.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

texts. These are the two general dictionaries – General Inquirer (Stone et al., 1966) and DICTION (Hart, 2000) – and the two word lists generated for finance-related texts: Henry (Henry, 2006, 2008) and Loughran and McDonald (Loughran & McDonald, 2011).

In the contributions of Henry (2006, 2008) and Loughran and McDonald (2011), the usage of general word lists for different forms of textual content like news, earnings press releases or annual reports was widely criticized in favor of domain-specific word lists, because of the high possibility of misclassification (Algaba et al., 2020, pp. 13–15; Lewis & Young, 2019, pp. 598f.; Mengelkamp et al., 2016, p. 7; Price et al., 2012, p. 1006). Loughran and McDonald (2011, p. 49) analyzed that 73.8% of negative words in the general dictionary General Inquirer do not have a negative meaning in a business context.

Despite the fact that the Henry word lists have been used for different purposes like conference calls (Davis et al., 2015, pp. 641, 647; Price et al., 2012, pp. 996f.) or news (Jandl et al., 2014, pp. 4, 7), the lists provided by Loughran and McDonald have become predominant (Kearney & Liu, 2014, p. 175) in the field of finance. They have been used in the classification of many different kinds of written financial content like news (Garcia, 2013, pp. 1272, 1274; Gurun & Butler, 2012, pp. 562, 566), conference calls (Mayew & Venkatachalam, 2012, pp. 2, 20) and annual reports (Ahmed & Elshandidy, 2016, p. 179; Jegadeesh & Wu, 2013, pp. 713, 715).

Due to the absence of a German domain-specific dictionary for the field of finance, research was limited to different versions of general dictionaries like LIWC (Meier et al., 2018; Wolf et al., 2008) or SentiWS (Remus et al., 2010), resulting in little research (Ammann & Schaub, 2016; Dorfleitner et al., 2016; Fritz & Tows, 2018). The first public available business-related dictionary for the German language was introduced by Bannier et al. (2019b). The introduced word lists are based on the predominant lists by Loughran and McDonald (Bannier et al., 2019b, p. 79) and have already been successfully used (Bannier et al., 2017, 2019a; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019).

As stated in Bannier et al. (2019a, p. 2), the contributions of Bannier et al. (2017, 2019a) are the primary studies analyzing the information content of CEO speeches delivered at the Annual General Meeting. Thus, this paper is also an essential complementary contribution to the information content of CEO speeches.

3. DATA

3.1. Data Source

I collected the transcripts of the CEO speeches from the companies' homepages, since there is no database for German CEO speeches delivered at the AGM. I screened the web pages of all companies listed in the DAX, MDAX, SDAX or TECDAX between 2008 and 2019 for transcripts of CEO speeches delivered at the AGM. Since not all companies publish transcripts on their homepage, I could find 976 speeches of 139 companies for the initial sample. I had to remove 53 speeches that were not delivered by the CEO. All available additional information, such as annotations, audio and video material provided by the company or other providers, was evaluated to confirm that the speeches were initially delivered in German. Therefore I had to exclude another 50 speeches. Additionally, 49 transcripts contained speeches of several speakers and required filtering of the relevant parts. Due to a delisting, I had to delete one additional speech. The final sample consists of 872 speeches from 125 companies. Comparing the contributions of Bannier et al. (2017, p. 10) (338 speeches) and Bannier et al. (2019a, p. 7) (457 speeches), this is the most comprehensive collection of German CEO speeches so far. An overview of the sample creation is given in Table 1. I obtained all other variables from Thomson Reuters Datastream.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 1
Sample creation

Source/Filter	Sample Size	Removed Observations
CEO speeches found on the companies' homepages	976	
Speeches not held by the CEO	923	53
Speeches held initially in English	873	50
Speeches where no CAR or CAV could be calculated	872	1
Final Sample	872	

Source: Author's calculation.

3.2. Used Dictionaries

The mutated vowels "ä", "ö" and "ü" in the German language can alternatively be written as "ae", "oe" and "ue". To get the updated form of the BPW_O (BPW_N), the first step is to add the alternative spelling of words with mutated vowels because the BPW_O does not include those. As a part of the parsing procedure, I deleted hyphens. Therefore, stop words written with hyphens had to be included without hyphens. Overall, I deleted 21 words that also appear on the positive and negative list of the BPW_O from the stop word list. In total, 144 stop words occurred twice and had to be deleted, because 110 surnames match company or given names (e.g. "kummer"). After extending for mutated vowels and hyphens, another 34 words occurred twice. Finally, I added 244 additional stop words through a translation of the generic list provided by Loughran and McDonald (2020) (LMD stop words). A summary of the conducted steps and the resulting alteration of the number of words on the different lists is given in Table 2.

Table 2
Updating of the BPW

	Positive	Negative	Stop words
BPW_O total words	2,223	10,147	3,682
Adding mutated vowels	+ 626	+ 2,514	+ 218
Including words without hyphens			+ 153
Delete doubles (positive/negative)			- 21
Delete doubles			- 144
Adding additional LMD stop words			+ 244
BPW_N total words	2,849	12,661	4,132

Source: Author's calculation.

Due to the update of the BPW_O, this paper examines the suitability of two different dictionaries.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

3.3. Parsing

Given expressed criticism regarding unspecified parsing rules and the related difficulty to replicate existing studies (Loughran & McDonald, 2015, p. 2), I give a detailed overview of performed text manipulation.

In the first step, the collected PDF files were transferred into TXT files using UTF-8 encoding (Bannier et al., 2017, p. 10, 2019a, p. 9; Meier et al., 2018, p. 29). In order to automatically process the speeches, they need to be parsed. Due to the unique and unsystematic character of the collected texts, manual corrections need to be conducted before using an automated parser. Those include the removal of headlines, disclaimers, legal notices, and additional information (e.g. the positioning of slides).

The subsequent automated parser was programmed using python. First of all, I replaced typographic ligatures (Bannier et al., 2017, p. 10, 2019a, p. 9) and hyphens (Loughran & McDonald, 2011, internet appendix) and converted all words to lowercase (Fritz & Tows, 2018, p. 61; Picault & Renault, 2017, p. 139). Additionally, I removed special characters (Allee & Deangelis, 2015, p. 247; Mengelkamp et al., 2016, p. 4), numbers (Boudt & Thewissen, 2019, p. 84; Schmeling & Wagner, 2016, p. 8), punctuation (Gentzkow et al., 2019, p. 538; Loughran et al., 2009, p. 41), and multiple whitespaces (González et al., 2019, p. 7; Schmeling & Wagner, 2016, p. 8). Finally, I removed words with fewer than three characters (Bannier et al., 2017, p. 10, 2019a, pp. 9f.; Loughran et al., 2009, p. 42). Depending on the used dictionary (BPW_O or BPW_N), I deleted the predefined individual stop words. Stop words are very common words but have relatively little meaning or rarely contribute information on their own, despite being essential to the grammatical structure of a sentence (Bannier et al., 2017, p. 10; Gentzkow et al., 2019, p. 538).

Furthermore, I included an important automated alteration[3] of the words "betrug" and "sorgen" prior to the automated parser. When written in lowercase, the words were changed to "betrugnoneg" and "sorgennoneg." This is because of the very frequent occurrence of those words in the analyzed texts (betrug: 812, sorgen: 344) and the characteristics of the German language. When written with a first capital letter, both words are nouns, where the word "Betrug" means "fraud" and the word "Sorgen" means "sorrow," which are both negative words in a business context and due to that are justifiably on the list of negative words. But when written entirely in lowercase, both words are verbs. In this case, the word "betrug" means "amounted" and "sorgen" means "care," which does not have a negative connotation. Without this automated alteration, the exclusive use of lowercase words would lead to a wrong and exaggerated number of negative words.

4. METHDOLOGY

4.1. Measurement of Sentiment

Using python, I counted the occurrence of positive (p) and negative (n) words from each of the two dictionaries as well as the total number of words (w) for each document. By using those three numbers, a variety of measurements of sentiment can be calculated. Even though the notations differ in several contributions, this paper focuses on the most widely used measurements to evaluate which sentiment measure is the most appropriate for the tone of textual documents in finance.

[3] Note that this automated alteration was only implemented when using the updated form of the dictionary provided by Bannier et al. (2019b) (BPW_N).

DOI: 10.7172/2353-6845.jbfe.2021.2.1

First of all, I calculated a simple share of negative and positive words as in Loughran and McDonald (2011, p. 46), Ferguson et al. (2015, p. 7) and Ammann and Schaub (2016, p. 2):

$$N = \frac{n}{w} \tag{1}$$

$$P = \frac{p}{w} \tag{2}$$

Other studies, as stated below, use the relation of positive and negative words rather than their individual fractions. However, there are different approaches to measure this relation. In this paper, I used the three most prominent relative measurements of sentiment.

Following the approach of Davis et al. (2015, p. 646), Loughran and McDonald (2015, p. 4), and Picault and Renault (2017, p. 141), I measured the sentiment of a text as the number of positive words minus the number of negative words divided by the total number of words:

$$Tone = \frac{p - n}{w} \tag{3}$$

Other contributions switch the numerator while retaining the notation "*Tone*" (Franke, 2018, p. 9; Kim & Meschke, 2014, p. 33). To prevent misinterpretations, this paper uses the term *ITone* for inverted tone.

$$ITone = \frac{n - p}{w} \tag{4}$$

In contrast to *Tone* and *ITone*, the variable *NTone* used by Henry (2008, p. 386), Price et al. (2012, p. 998), and Henry and Leone (2016, p. 159) only focuses on the number of positive and negative words and is not altered by the length of the analyzed text. It therefore gives the NetTone:

$$NTone = \frac{p - n}{p + n} \tag{5}$$

Also, a fourth relative variable *NToneSQ* as in Henry (2008, p. 393) is estimated, by squaring the variable *NTone*.

Given this variety of six different measurements of sentiment, this paper adds the two measurements *InvTone* and *NToneSQ* to the four already tested calculations, when using the BPW_O (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

In this paper, following Apel and Blix Grimaldi (2012, p. 9), Davis et al. (2015, p. 653), and Bannier et al. (2017, p. 15), all words found are weighted equally. This approach makes it possible for other researchers to replicate and further develop the results of this contribution, due to the independence of the weighting scheme from the used texts. This approach and the superiority of equal weighting is also supported by Henry and Leone (2016, p. 166).

4.2. Empirical Approach

By using linear regressions, I conduct one of the most common approaches for analyzing the impact of sentiment on stock prices (Kearney & Liu, 2014, p. 177). Therefore, I performed several linear regressions for ten different dependent variables in the following form:

$$Dep_j = \alpha_0 + \alpha_1 Sentiment_j + \sum_{k=1}^{K} \propto_k Control_{kj} + \varepsilon_j \tag{6}$$

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Dep represents two different forms of variables to measure the effect of speech sentiment on stock prices and trading.

To obtain the effect on stock prices, I calculated cumulative abnormal returns (*CAR*). The abnormal returns are calculated by the market adjusted model using the value weighted market index CDAX. Following Henry (2006, p. 5, 2008, p. 385), Loughran and McDonald (2011, p. 41), Henry and Leone (2016, p. 159), and Bannier et al. (2017, p. 12, 2019a, p. 8), the CARs are calculated through cumulating the abnormal returns (*AR*) over a predefined event period (event window) with length *T*. I obtained the individual *AR*s by subtracting the returns (*R*) of the analyzed stock (*j*) from the return of the CDAX for a given day (*t*):

$$AR_{j,t} = R_{j,t} - R_{CDAX.t} \tag{7}$$

$$CAR_{j,T} = \sum_{t=0}^{T} AR_{j,t} \tag{8}$$

Based on Loughran and McDonald (2011, p. 41), Boudt and Thewissen (2019, p. 95) and Bannier et al. (2019a, p. 9), this paper solely uses event windows beginning on the day of the AGM ($t = 0$), to only measure the effect of the CEO speeches. Therefore, the five different trading day event windows [0,1], [0,3], [0,5], [0,15], and [0,30] were used following contributions examining similar texts like CEO letters or CEO conference calls (Bannier et al., 2019a, p. 9; Boudt & Thewissen, 2019, p. 95; Doran et al., 2012, p. 412; Loughran & McDonald, 2011, p. 41; Mayew & Venkatachalam, 2012, p. 20).

Additionally, I performed all regressions with cumulative abnormal trading volumes (*CAV*) for the five different event windows. I calculated the different CAVs according to Bannier et al. (2017, p. 47, 2019a, p. 38) and Price et al. (2012, p. 1000) as:

$$AV_{j,t} = \frac{VOL_{j,t}}{\overline{VOL_{j,t}}} - 1 \tag{9}$$

$$CAV_{j,T} = \sum_{t=0}^{T} AV_{j,t} \tag{10}$$

Here $VOL_{j,t}$ is the trading volume for firm *j* at day *t*, and $\overline{VOL_{j,t}}$ is the mean volume for firm *j* from trading day $t = -252$ to $t = -2$. Due to different estimation windows in the primary studies of Bannier et al. (2017, p. 47, 2019a, p. 38), I selected a combined period of time in accordance with Price et al. (2012, p. 1000).

I used the six above mentioned measurements of sentiment separately for each of the ten different dependent variables *Dep*.

The comprehensive set of control variables *Control* consist of eleven different variables (*K*), which include the firm size (*SIZE*), the market to book value (*M2B*), leverage (*LEV*), volatility (*VOLA*), volume (*VOL*), number of words (*COUNT*), individual words (*IND*), return on assets (*ROA*), the earnings surprise (*EPS_SP*), and the dividend surprise (*DIV_SPP* and *DIV_SPN*) (Bannier et al., 2017, p. 47, 2019a, pp. 38f.; Doran et al., 2012, p. 426; Loughran & McDonald, 2011, p. 63). The calculation of the individual control variables can be found in the appendix.

I used the variables *SIZE*, *VOL*, and *COUNT* in a logarithmic form. When using *CAV*, the variable *VOL* is excluded from the regression. Additionally, I used year fixed effects.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

5. RESULTS

5.1. Summary Statistics

I report summary statistics for the analyzed sample of 872 CEO speeches in the following three tables.

Table 3 provides descriptive statistics for all calculated CARs and CAVs. While I could calculate CARs for all different event windows, the calculation of CAVs is only partially possible based on the availability of data. As stated in Bannier et al. (2017, p. 16), the means of all CARs are economically small, indicating no market reaction due to the AGM. In comparison, CAVs are in the mean higher than 1, indicating an abnormal trading volume caused by the AGM.

Table 3
Descriptive statistics for CARs and CAVs

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
CAR01	872	0.001	0.027	-0.184	0.104	-0.013	0.015
CAR03	872	-0.0002	0.031	-0.285	0.116	-0.017	0.018
CAR05	872	-0.002	0.037	-0.171	0.138	-0.021	0.018
CAR015	872	-0.004	0.059	-0.271	0.229	-0.035	0.033
CAR030	872	-0.005	0.087	-0.459	0.321	-0.057	0.046
CAV01	849	2.790	2.192	0.041	32.141	1.654	3.195
CAV03	841	4.825	3.076	0.054	37.987	3.130	5.645
CAV05	839	6.787	3.705	0.087	41.084	4.604	7.927
CAV015	827	16.498	7.859	0.595	82.829	12.060	19.007
CAV030	817	30.614	12.434	0.931	124.574	23.843	35.132

Source: Author's calculation based on data from Thomson Reuters Datastream.

Because of the extension of the stop word list, the mean words counted are 22.7% lower for BPW_N, as given in Table 4. In addition to the change of sentiment measures, the reduction of words also improves calculation times of algorithms for measuring textual sentiment. The deletion of positive words from the stop words list leads to an increase in the number of positive words. In contrast, the mean number of negative words decreases due to the treatment of the words "betrug" and "sorgen." The combination of those changes leads to an increase in all six sentiment measures on average. The mean number of positive and negative words combined with positive means for the measurements *Tone*, *NTone*, and *NToneSQ* show that the speeches delivered by the CEOs are on average positive. This positivity of speeches is slightly higher for the BPW_N dictionary. As stated in Doran et al. (2012, p. 414) for earnings conference calls using the Henry word list, it is not surprising that the general sentiment is positive, reflecting the effort of CEOs to present their information as positive as possible. This positive wording is also reflected in the characteristics of values of *NTone*, which by construction is bounded between -1 and 1. While the minimum value is -0.455 and thus relatively far from the highest possible minimum, the maximum value of 0.941 for BPW_O and 0.943 for BPW_N shows that in the most positive speeches hardly any negative words were used. This finding is additionally confirmed by the positivity of the 25% quartile and by the minimum number of one negative and eleven positive words.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 4
Descriptive statistics for sentiment variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
COUNT_BPW_O	872	2,411.709	834.021	759	5,625	1,817.5	2,909
IND_NUM_BPW_O	872	1,153.603	334.053	433	2,402	920.8	1,331.5
IND_BPW_O	872	0.490	0.046	0.368	0.642	0.457	0.519
P_NUM_BPW_O	872	90.142	32.124	11	206	65	112
N_NUM_BPW_O	872	38.556	25.082	1	152	21	49
N_BPW_O	872	0.015	0.007	0.001	0.046	0.010	0.019
P_BPW_O	872	0.038	0.009	0.010	0.068	0.032	0.044
Tone_BPW_O	872	0.023	0.013	-0.029	0.062	0.014	0.032
NTone_BPW_O	872	0.428	0.241	-0.455	0.941	0.283	0.606
ITone_BPW_O	872	-0.023	0.013	-0.062	0.029	-0.032	-0.014
NToneSQ_BPW_O	872	0.241	0.188	0.000	0.886	0.083	0.367
COUNT_BPW_N	872	1,864.443	646.324	589	4,431	1,405	2,247.2
IND_NUM_BPW_N	872	1,098.989	326.592	399	2,323	873	1,277
IND_BPW_N	872	0.602	0.052	0.456	0.777	0.566	0.634
P_NUM_BPW_N	872	92.905	32.992	11	212	68	116
N_NUM_BPW_N	872	37.361	24.830	1	149	20	48
N_BPW_N	872	0.019	0.010	0.001	0.062	0.012	0.024
P_BPW_N	872	0.051	0.011	0.015	0.095	0.043	0.058
Tone_BPW_N	872	0.031	0.017	-0.039	0.090	0.020	0.043
NTone_BPW_N	872	0.454	0.238	-0.455	0.943	0.304	0.630
ITone_BPW_N	872	-0.031	0.017	-0.090	0.039	-0.043	-0.020
NToneSQ_BPW_N	872	0.263	0.195	0.000	0.889	0.095	0.396

Source: Author's calculation.

I conducted a dependent-samples t-test to compare the alteration of positive and negative words found. There was a significant difference in the number of positive words found concerning the use of the BPW_O (M = 90.142, SD = 32.124) and BPW_N (M = 92.905, SD = 32.992), $t(871) = -22.939$, $p < .001$. This also applies to the number of negative words found when using the BPW_O (M = 38.556, SD = 25.082) and the BPW_N (M = 37.361, SD = 24.830), $t(871) = 18.471$, $p < .001$.

Table 5 gives the descriptive statistics for the additional control variables used in the regression. In accordance with Bannier et al. (2017, p. 17), the number of observations in which the dividend per share is unchanged compared to the previous year is 31.1%. In 51.4% the dividend per share increased, and in 17.5% decreased.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 5
Descriptive statistics for control variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
SIZE	870	9,883.827	16,996.830	30.200	104,226.900	845.245	10,287.470
M2B	869	2.208	2.267	-17.640	19.070	1.160	2.930
LEV	865	0.637	0.209	0.094	1.811	0.519	0.753
VOLA	872	0.020	0.010	0.002	0.130	0.014	0.024
VOL	852	2,108.435	4,949.786	0.100	47,270.600	67.925	1,518.850
ROA	865	0.037	0.065	-0.483	0.679	0.007	0.063
EPS_SP	848	1.685	16.275	-140.625	196.193	-1.607	2.625
DIV_SPP	872	0.514	0.500	0.000	1.000	0.000	1.000
DIV_SPN	872	0.175	0.381	0.000	1.000	0.000	0.000

Note: The definitions of all variables are given in the appendix.

Source: Author's calculation based on data from Thomson Reuters Datastream.

Overall, editing stop words leads to a word reduction of 22.7% (477,216 words), as stated in Table 6. Deleting the 21 words from the stop word list that are also on the positive and negative list leads to 3.1% (2,409) more positive words found, with only eight more individual words. Although there are three more individual negative words, the number of negative words found decreases by 3.1% (1,042). This is because of the correction for "betrug" and "sorgen" described in the parsing process.

Table 6
Total number of words

	BPW_O	BPW_N
	All words	
Number of words	2,103,010	1,625,794
Individual words	100,151	99,970
	Positive words	
Number of words	78,604	81,013
Individual words	1,123	1,131
	Negative words	
Number of words	33,621	32,579
Individual words	2,180	2,183

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 7 displays the number and cumulative fraction of the ten most frequent positive words in all speeches after correcting for stop words. The only difference is the deletion of the word "große" from the stop word list of the dictionary BPW_N.

Table 7
Ten most frequent positive words

BPW_O			BPW_N		
Word	**Number**	**cumulative %**	**Word**	**Number**	**cumulative %**
erfolgreich	2,143	2.73%	erfolgreich	2,143	2.65%
erfolg	2,015	5.29%	erfolg	2,015	5.13%
erreicht	1,624	7.36%	erreicht	1,624	7.14%
erreichen	1,566	9.35%	erreichen	1,566	9.07%
großen	1,546	11.31%	großen	1,546	10.98%
besser	1,515	13.24%	besser	1,515	12.85%
positiv	1,157	14.71%	große	1,209	14.34%
stärker	1,089	16.10%	positiv	1,157	15.77%
positive	1,040	17.42%	stärker	1,089	17.11%
stärken	1,035	18.74%	positive	1,040	18.40%

Source: Author's calculation.

As Table 8 illustrates, the adjustment in the parsing process for the words "betrug" and "sorgen" leads to an extensive decrease of those words, to the extent to which they do not appear in the ten most frequent negative words.

Table 8
Ten most frequent negative words

BPW_O			BPW_N		
Word	**Number**	**cumulative %**	**Word**	**Number**	**cumulative %**
herausforderungen	1,019	3.03%	herausforderungen	1,019	3.13%
betrug	876	5.64%	krise	845	5.72%
krise	845	8.15%	schwierigen	792	8.15%
schwierigen	792	10.51%	rückgang	728	10.39%
rückgang	728	12.67%	gegen	650	12.38%
gegen	650	14.60%	minus	483	13.86%
minus	483	16.04%	verfügung	476	15.33%
verfügung	476	17.46%	wider	415	16.60%
wider	415	18.69%	leider	356	17.69%
sorgen	398	19.87%	finanzkrise	330	18.71%

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

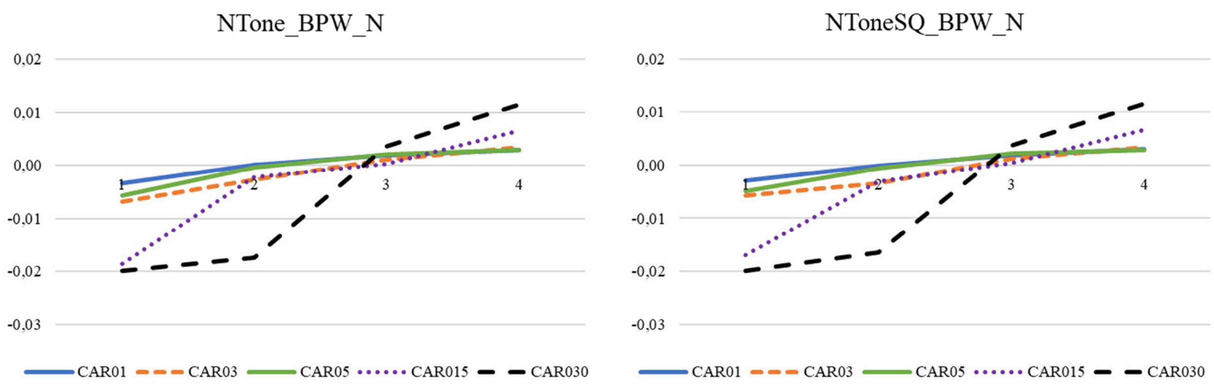
An English translation of all words listed in Table 7 and Table 8 is given in the appendix. Note that an important distinction of German words through small and capital letters is not possible due to the nature of the parsing procedure and structure of the dictionaries. Because of their impact, I only considered this distinction for the words "betrug" and "sorgen."

Of the 2,223 (BPW_N: 2,849) positive words available, I only found 1,123 (BPW_N: 1,131) words. A comparably small fraction of those words found is able to account for 18.74% (BPW_N: 18.40%). The same applies to the more extensive list of 10,147 (BPW_N: 12,661) negative words. Of this list, I only found 2,180 (BPW_N: 2,183) words in the speeches, with ten words accounting for 19.87% (BPW_N: 18.71%) of all negative words found. These results clearly indicate that the correct words are more important than the mere extent of the used list.

5.2. Sentiment Measurement

Following Loughran and McDonald (2011, pp. 50f.), the assumption that the sentiment of certain texts is relevant leads in the case of CEO speeches to the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns and higher abnormal trading volumes. By dividing all texts into quartiles based on the different sentiment measures[4] and analyzing the median CARs and CAVs, a visual examination can be conducted. Figure 1 gives the only two measurements that meet the stated assumptions. Using the sentiment measures *NTone* and *NToneSQ*, it is possible to have ascending quartile medians for all five event windows.

Figure 1
CARs by quartiles (sufficient)



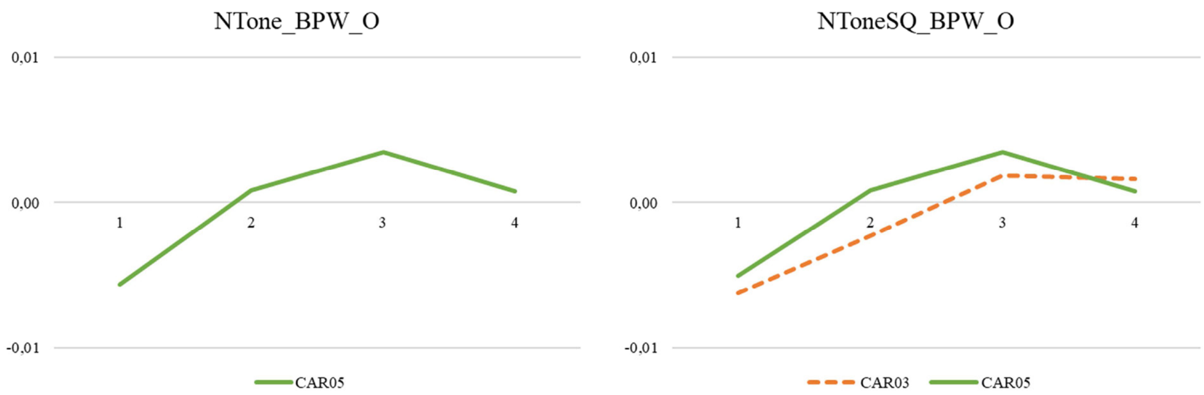
Source: Author's calculation.

The equivalent measures for the BPW_O cannot provide comparable sufficient results for all analyzed event windows. The affected windows and the not sufficient results for the associated quartiles are given in Figure 2. Here the window CAR [0,5] does not meet the assumptions for the sentiment measurement *NTone*. The same applies to the two windows CAR [0,3] and CAR [0,5] for *NToneSQ*. Other measurements of sentiment using the BPW_O or BPW_N do not meet this assumption either and therefore are not discussed further.

4 Note that only the share of negative words (*N*) was sorted in the descending order. All other sentiment measures are sorted in the ascending order.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Figure 2
CARs by quartiles (not sufficient)

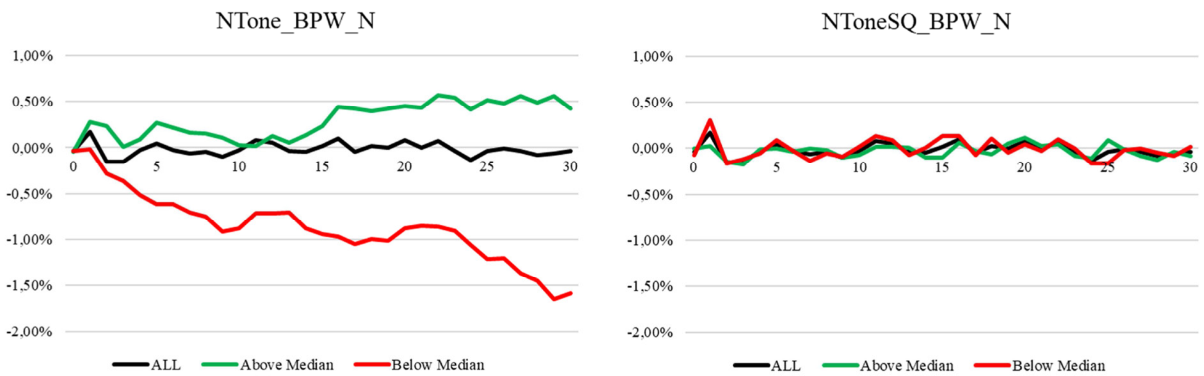


Source: Author's calculation.

With regard to the visual examination of the CAVs for different sentiment measures, no measure meets the above stated assumptions. Therefore, I excluded those figures.

Another essential assumption independent of certain event windows is the separation of above and below average abnormal returns through the use of sentiment measures as precisely as possible. Therefore, following Bannier et al. (2019a, pp. 17f., 37) and Price et al. (2012, pp. 1001f.), Figure 3 gives the average cumulative abnormal returns for up to 30 days following the AGM, divided by the above and below median sentiment measures *NTone* and *NToneSQ*. Additionally, the average CARs for all days are given[5].

Figure 3
CARs over time



Source: Author's calculation.

The accumulation of abnormal returns in Figure 3 for up to 30 days following the AGM shows that the average CARs are close to zero. By dividing the different observations into above and below median *NTone*, it is possible to separate positive and negative CARs. This is in accordance with the results of Bannier et al. (2019a, pp. 17f., 37). This separation can only be conducted using *NTone*. The same analysis using *NToneSQ* allows no distinction of positive and negative CARs using above and below median *NToneSQ*.

It therefore can be stated as an interim result that only the usage of the reformed and extended BPW_N dictionary with *NTone* as a sentiment measure is able to meet one of the central assumptions stated in the pioneer paper by Loughran and McDonald (2011, pp. 50f.) and the additional assumption of distinction.

[5] Due to the results stated in Figure 1 and Figure 2, only the results for *NTone* and *NToneSQ* calculated using the BPW_N are given.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

5.3. Significance of Results

Based on the preceding results, this section examines the relation between *NTone* and CARs for different event windows in a multivariate context using the control variables that I described above. Table 9 reports the regression results for *NTone* using the BPW_N and the five different event windows for CARs.

Table 9
Regression of NTone_BPW_N and CARs

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
NTone_BPW_N	0.014***	0.018***	0.018***	0.035***	0.064***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_N	0.009**	0.004	0.011**	0.011	0.014
	(0.004)	(0.005)	(0.005)	(0.008)	(0.012)
IND_BPW_N	0.071***	0.045	0.041	0.042	0.056
	(0.027)	(0.032)	(0.036)	(0.055)	(0.082)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0002	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.028	-0.091	-0.144	-0.679**	-1.162**
	(0.201)	(0.247)	(0.210)	(0.326)	(0.499)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.084***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00001	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
DIV_SPP	-0.0002	0.003	0.007**	0.007	0.018***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.119**	-0.064	-0.115*	-0.128	-0.152
	(0.046)	(0.052)	(0.062)	(0.097)	(0.141)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.032	0.050	0.053	0.073	0.121
Adjusted R2	0.004	0.022	0.026	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.149	1.826**	1.977***	2.747***	4.800***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

The results show a high statistical significance of the coefficient of the sentiment measurement *NTone* that I calculated using the BPW_N and the five different CARs as dependent variables. Thus, more positive speeches of CEOs can be associated with higher abnormal returns. An increase in *NTone* by the interquartile change of 0.326 leads to a minor increase of 0.42% in CAR [0,1], but a major increase of 1.53% in CAR [0,30]. This role as a key factor in the market reaction to AGMs becomes more interesting, when other variables, based on the performance or the dividend policy are considered. The ROA negatively relates to all five event windows and is only significant for the first three windows. The dividend surprise can only partially account for the significance of the longer event windows. I could verify only a significant association with individual event windows for the analyzed control variables. None of the variables are able to explain all windows.

Regarding the significant relation of *NTone* as a relative measurement of sentiment and short- and long-term event windows, the results are consistent with Price et al. (2012, pp. 1004f.) and Bannier et al. (2017, p. 37, 2019a, p. 34).

Despite the insufficient fulfillment of the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns for *NTone* using the BPW_O, Table 10 shows that the positive relation between this measurement and the different CARs is almost as significant as the usage of BPW_N. Only for the event windows CAR [0,1] and CAR [0,5], the coefficient is significant at a 5% level. Due to the smaller interquartile change of 0.323, a change in *NTone* by this change leads to a 0.39% higher CAR [0,1] and a 1.45% higher CAR [0,30]. Interestingly, these results show higher significance than Bannier et al. (2019a, p. 34), where maximum significance at the 5% level was achieved (CAR [0,30]: 10%).

Table 10
Regression of NTone_BPW_O and CARs

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
NTone_BPW_O	0.012**	0.016***	0.017**	0.034***	0.062***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_O	0.008**	0.003	0.010*	0.009	0.011
	(0.004)	(0.005)	(0.005)	(0.009)	(0.012)
IND_BPW_O	0.075**	0.036	0.037	0.022	0.032
	(0.031)	(0.037)	(0.042)	(0.064)	(0.097)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0001	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.022	-0.098	-0.149	-0.688**	-1.172**
	(0.202)	(0.248)	(0.210)	(0.324)	(0.496)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.083***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00002	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 10 (cont.)

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
DIV_SPP	-0.00001	0.003	0.007^{**}	0.008	0.018^{***}
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022^{**}
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.110^{**}	-0.043	-0.104	-0.096	-0.111
	(0.048)	(0.054)	(0.065)	(0.098)	(0.145)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.029	0.047	0.052	0.072	0.120
Adjusted R2	0.001	0.020	0.025	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.042	1.736^{**}	1.937^{***}	2.721^{***}	4.790^{***}

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: Author's calculation.

Based on the already stated results for the necessary assumptions of the cumulative abnormal trading volumes under 5.2, I will not discuss those regressions further.

6. CONCLUSION

This paper focuses on textual analysis as an important part of accounting and finance research using the dictionary-based approach with the first available finance-related dictionary for the German language (BPW_O). Due to the novelty of this dictionary, the aim of this paper is to propose several reforms and extensions (BPW_N) to improve its performance and to find the most appropriate measurement of sentiment.

Based on the visual examination of the two central assumptions that speeches with a more positive measurement of sentiment lead to higher abnormal returns and that it is possible to separate above and below average abnormal returns through the use of sentiment measures, the use of the measurement *NTone* calculated using the BPW_N should be preferred. Additionally, I was able to supplement the significance of these results by several regressions. Here the use of *NTone*, calculated by using the BPW_N, could provide highly statistically significant results for all five analyzed event windows. Thus, more positive speeches of CEOs can be associated with higher abnormal returns following the Annual General Meeting. Based on the event window, an increase in *NTone* by the interquartile change of 0.326 leads to an increase in cumulative abnormal returns ranging from 0.42% (CAR [0,1]) to 1.53% (CAR [0,30]).

Using the most comprehensive collection of German CEO speeches so far, this paper is able to give two contributions to the literature on textual analysis of German texts. Through implementing reforms and extensions, I improved the results of the original BPW_O and confirmed the stated hypothesis. Additionally, the combination of the BPW_N and the relative measurement of sentiment *NTone* has proven to be the most suitable one for measuring business texts and therefore answers the additional research question.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Due to the results of the proposed adjustments on the newly developed BPW_O, additional improvements should be considered and tested. Moreover, this new version of the BPW (BPW_N) should be compared to old and new versions of general German dictionaries. As there is a wide range of publicly available textual data, the BPW_N should be used to analyze other types of corporate disclosures.

References

Ahmed, Y., & Elshandidy, T. (2016). The effect of bidder conservatism on M&A decisions: Text-based evidence from US 10-K filings. International Review of Financial Analysis, 46, 176–190. https://doi.org/10.1016/j.irfa.2016.05.006

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. Retrieved on 02.04.2020 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2652876. https://doi.org/10.1111/joes.12370

Allee, K. D., & Deangelis, M. D. (2015). The structure of voluntary disclosure narratives: Evidence from tone dispersion. Journal of Accounting Research, 53(2), 241–274. https://doi.org/10.1111/1475-679X.12072

Ammann, M., & Schaub, N. (2016). Social interaction and investing: Evidence from an online social trading network (Working Paper). Retrieved on 11.07.2018 from https://www.rsm.nl/fileadmin/home/Department_of_Finance__VG5_/PAM2016/Final_Papers/Nic_Schaub.pdf

Apel, M., & Blix Grimaldi, M. (2012). The information content of central bank minutes. Sveriges Riksbank Working Paper Series, (261). Stockholm: Sveriges Riksbank. Retrieved on 13.02.2020 from http://archive.riksbank.se/Documents/Rapporter/Working_papers/2012/rap_wp261_120426.pdf. https://doi.org/10.2139/ssrn.2092575

Bannier, C. E., Pauls, T., & Walter, A. (2017). CEO-speeches and stock returns (Working Paper). Retrieved on 15.08.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3051151_code1882913.pdf?abstractid=3051151&mirid=1

Bannier, C. E., Pauls, T., & Walter, A. (2019a). The Annual General Meeting revisited: The role of the CEO speech (Working Paper). Retrieved on 11.12.2021 from https://ssrn.com/abstract=2869785

Bannier, C. E., Pauls, T., & Walter, A. (2019b). Content analysis of business specific text documents: Introducing a German dictionary. Journal of Business Economics, 89(1), 79–123. https://doi.org/10.1007/s11573-018-0914-8

Boudt, K., & Thewissen, J. (2019). Jockeying for position in CEO letters: Impression management and sentiment analytics. Financial Management, 48(1), 77115. https://doi.org/10.1111/fima.12219

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. Review of Accounting Studies, 20(2), 639–673. https://doi.org/10.1007/s11142-014-9309-4

Doran, J. S., Peterson, D. R., & Price, M. S. (2012). Earnings conference call content and stock price: The case of REITs. Journal of Real Estate Finance and Economics, 45(2), 402–434. https://doi.org/10.1007/s11146-010-9266-z

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending: Evidence from two leading European platforms. Journal of Banking & Finance, 64, 169–187. https://doi.org/10.1016/j.jbankfin.2015.11.009

Ferguson, N. J., Philip, D., Lam, H. Y. T., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. Multinational Finance Journal, 19(1), 1–31. https://doi.org/10.17578/19-1-1

Franke, B. (2018). Qualitative information and loan terms: A textual analysis (Working Paper). Retrieved on 15.09.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3209201_code1660824pdf?abstractid=3152458&mirid=1

Fritz, D., & Tows, E. (2018). Text mining and reporting quality in German banks: A cooccurrence and sentiment analysis. Universal Journal of Accounting and Finance, 6(2), 54–81. https://doi.org/10.13189/ujaf.2018.060204

Garcia, D. (2013). Sentiment during recessions. The Journal of Finance, 68(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535–574. https://doi.org/10.1257/jel.20181020

González, M., Guzmán, A., Téllez, D. F., & Trujillo, M. A. (2019). What you say and how you say it: Information disclosure in Latin American firms. Journal of Business Research, 127(3), 427 443. https://doi.org/10.1016/j.jbusres.2019.05.014

Gurun, U. G., & Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. The Journal of Finance, 67(2), 561–598. https://doi.org/10.1111/j.1540-6261.2012.01725.x

Hart, R. P. (2000). DICTION 5.0. Retrieved on 09.06.2020 from https://rhetorica.net/diction.htm

Henry, E. (2006). Market reaction to verbal components of earnings press releases: Event study using a predictive algorithm. Journal of Emerging Technologies in Accounting, 3, 1–19. https://doi.org/10.2308/jeta.2006.3.1.1

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Henry, E. (2008). Are investors influenced by how earnings press releases are written?. Journal of Business Communication, 45(4), 363–407. https://doi.org/10.1177/0021943608319388

Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. The Accounting Review, 91(1), 153–178. https://doi.org/10.2308/accr-51161

Jandl, J.-O., Feuerriegel, S., & Neumann, D. (2014). Long- and short-term impact of news messages on house prices: A comparative study of Spain and the United States. Paper presented at Thirty Fifth International Conference on Information Systems, Auckland. Retrieved on 15.09.2019 from https://aisel.aisnet.org/icis2014/proceedings/DecisionAnalytics/17/

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. Journal of Financial Economics, 110(3), 712–729. https://doi.org/10.1016/j.jfineco.2013.08.018

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

Kim, Y. H., & Meschke, F. (2014). CEO interviews on CNBC (Working Paper). Retrieved on 12.02.2020 from http://dx.doi.org/10.2139/ssrn.1745085. https://doi.org/10.2139/ssrn.1745085

Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. Accounting and Business Research, 49(5), 587–615. https://doi.org/10.1080/00014788.2019.1611730

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. Journal of Accounting Literature, 29, 143–165.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2015). The use of word lists in textual analysis. Journal of Behavioral Finance, 16(1), 1–11. https://doi.org/10.1080/15427560.2015.1000335

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 54, 1187–1230. https://doi.org/10.1111/1475-679X.12123

Loughran, T., & McDonald, B. (2020). Stop words. Retrieved on 21.01.2021 from https://drive.google.com/file/d/0B4niqV00F3mseWZrUk1YMGxpVzQ/view?usp=sharing

Loughran, T., McDonald, B., & Yun, H. (2009). A wolf in sheep's clothing: The use of ethics-related terms in 10-K reports. Journal of Business Ethics, 89(1), 39–49. https://doi.org/10.1007/s10551-008-9910-1

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. The Journal of Finance, 67(1), 1–43. https://doi.org/10.1111/j.1540-6261.2011.01705.x

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). »LIWC auf Deutsch«: The development, psychometrics, and introduction of DE-LIWC2015. Retrieved on 08.03.2019 from https://osf.io/tfqzc/. https://doi.org/10.31234/osf.io/uq8zt

Mengelkamp, A., Wolf, S., & Schumann, M. (2016). Data driven creation of sentiment dictionaries for corporate credit risk analysis. Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). Retrieved on 10.07.2018 from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1058&context=amcis2016

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. https://doi.org/10.1016/j.jimonfin.2017.09.005

Price, M. S., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking & Finance, 36(4), 992–1011. https://doi.org/10.1016/j.jbankfin.2011.10.013

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - A publicly available German-language resource for sentiment analysis. Proceedings of the 7th International Language Ressources and Evaluation (LREC'10) (pp. 1168–1171). Retrieved on 19.12.2018 from http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf

Röder, F., & Walter, A. (2019). What drives investment flows into social trading portfolios?. The Journal of Financial Research, 42(2), 383–411. https://doi.org/10.1111/jfir.12174

Schmeling, M., & Wagner, C. (2016). Does central bank tone move asset prices?. Paper presented at the 77th Annual Meeting of American Finance Association (AFA 2017). Retrieved on 29.06.2018 from https://research.cbs.dk/en/publications/does-central-bank-tone-move-asset-prices(c6401864-a921-401c-90db-57d42d6b5022).html

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). The General Inquirer: A computer approach to content analysis, Cambridge, Mass.: The M.I.T. Press.

Tillmann, P., & Walter, A. (2018). ECB vs Bundesbank: Diverging tones and policy effectiveness. Joint Discussion Paper Series in Economics, (20). Retrieved on 13.02.2020 from https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/paper_2018/20-2018_tillmann.pdf

Tillmann, P., & Walter, A. (2019). The effect of diverging communication: The case of the ECB and the Bundesbank. Economics Letters, (176), 68–74. https://doi.org/10.1016/j.econlet.2018.12.035

Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. Diagnostica, 54(2), 85–98. https://doi.org/10.1026/0012-1924.54.2.85

DOI: 10.7172/2353-6845.jbfe.2021.2.1

APPENDIX

Table 11
Description of variables

Variable	Description
SIZE	Firm Size: Daily market value
M2B	Market to Book Value: Ratio of the market value of the ordinary (common) equity to the balance sheet value of the ordinary (common) equity
LEV	Leverage: Ratio of the total liabilities to the total assets
VOLA	Volatility: Standard deviation of the daily returns for the ninety trading-day window ending ten days prior to the AGM
VOL	Volume: Number of shares traded on the day of the AGM
COUNT	Total number of Words. Due to different stop word lists calculated individually for BPW_O and BPW_N
IND_NUM	Number of individual words. Due to different stop word lists calculated individually for BPW_O and BPW_N.
IND	Individual Words: *IND_NUM* divided by *COUNT*
ROA	Return on Assets: Net income divided by total assets
EPS_SP	Earnings Surprise: Calculated according to Bannier et al., 2017: The difference between the last reported earnings per share at time t minus the latest reported earnings per share in the year prior to date t, divided by the stock price one year before t times 100 $EPS_{SP} = \frac{EPS_t - EPS_{t-1}}{Price_{t-1}} \cdot 100$
DIV_SPP	Dividend Surprise Positive: Calculated according to Bannier et al., 2017: *DIV_SPP* equals one if the dividend per share is increased compared to the previous year, zero otherwise
DIV_SPN	Dividend Surprise Negative: Calculated according to Bannier et al., 2017: *DIV_SPN* equals one if the dividend per share is decreased compared to the previous year, zero otherwise
P_NUM	Number of positive words
N_NUM	Number of negative words

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 12
Translation of ten most frequent words

positive words		negative words	
German	**English**	**German**	**English**
besser	better	betrug	fraud, amounted
erfolg	success	finanzkrise	financial crisis
erfolgreich	successful	gegen	against
erreichen	achieve	herausforderungen	challenges
erreicht	achieved	krise	crisis
große	large	leider	unfortunately
großen	large	minus	minus
positiv	positive	rückgang	decline
positive	positive	schwierigen	difficult
stärken	strenghten	sorgen	sorrow, care
stärker	stronger	verfügung	decree
		wider	against

Note that the listed translations represent only one of several possibilities. Due to the nature of the parsing procedure and structure of the dictionaries, an important distinction of German words through small and capital letters is not possible.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Concept of Peer-to-Peer Lending and Application of Machine Learning in Credit Scoring

Aleksy Klimowicz
University of Warsaw, Faculty of Economic Sciences
aleksy.klimowicz@gmail.com

Krzysztof Spirzewski[1]
University of Warsaw, Faculty of Economic Sciences
kspirzewski@wne.uw.edu.pl
https://orcid.org/0000-0002-9646-5667

Received: 29 September 2021 / Revised: 23 November 2021 / Accepted: 26 November 2021 / Published online: 20 December 2021

ABSTRACT

Numerous applications of AI are found in the banking sector. Starting from the front-office, enhancing customer recognition and personalized services, continuing in the middle-office with automated fraud-detection systems, ending with the back-office and internal processes automatization. In this paper we provide comprehensive information on the phenomenon of peer-to-peer lending in the modern view of alternative finance and crowdfunding from several perspectives. The aim of this research is to explore the phenomenon of peer-to-peer lending market model. We apply and check the suitability and effectiveness of credit scorecards in the marketplace lending along with determining the appropriate cut-off point.
We conducted this research by exploring recent studies and open-source data on marketplace lending. The scorecard development is based on the P2P loans open dataset that contains repayments record along with both hard and soft features of each loan. The quantitative part consists in applying a machine learning algorithm in building a credit scorecard, namely logistic regression.

JEL Classification: G21; C25

Keywords: artificial intelligence, peer-to-peer lending, credit risk assessment, credit scorecards, logistic regression, machine learning.

1. INTRODUCTION

The recent explosive growth of brand-new alternative financial possibilities has brought about a lot of discussions and studies. One of such possibilities is the peer-to-peer alternative finance sector. The primary focus has been put on the analysis of a possible expansion of the peer-to-peer (P2P) finance industry with sequential inversion of the existing structural and institutional organization of banking. There are numerous instances of how peer-to-peer technology may

[1] Corresponding author.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

affect a particular industry. Considerable changes have already occurred in lodging, file sharing, multimedia, etc. A decentralized network of credit relations increasingly captures the credit market and challenges traditional banking pillars. P2P lending is characterized by the improvement of service and higher economic efficiency. On the other hand, P2P technology brings about various risks that have to be addressed.

In our paper we aimed to understand the structure and key features of a peer-to-peer lending market model, its role in financial intermediation, and investigate the main advantages and drawbacks of marketplace lending. Once we develop a clear understanding, the objective is to apply and check the suitability and effectiveness of credit scorecards in the marketplace lending along with determining the appropriate cut-off point.

The research is conducted by exploring recent studies and open-source data on marketplace lending. The scorecard development is based on the P2P loans open data set that contains repayments record along with both hard and soft features of each loan. The quantitative part consists in applying a machine learning algorithm in building a credit scorecard, namely logistic regression. The objective is to select, through descriptive and quantitative analysis, the best features that allow differentiating the loan performance in the marketplace lending environment and process the data, followed by scorecard construction and quality assessment.

The research paper is divided into three parts, each part having its particular objectives. Section 2 of the research is dedicated to developing a broad picture of the traditional financial system, as well as exploring the origins, explaining the structure and features of marketplace lending. The emphasis is put on the general mechanism of the platform's intermediation. Section 3 is intended to study the P2P lending system from the perspective of an end-user, along with the determination of risks involved in marketplace lending and an overview of current regulatory frameworks and practices. As an empirical part of the chapter, breakdowns of the alternative finance market in the European Union and in the United Kingdom are prepared. Section 4 contains an analysis of credit risk in marketplace lending. A credit scorecard is created based on the Logistic Regression, utilizing the best practices of variable processing and modelling. The last section number 5 provides conclusion of the paper.

2. BANKING SYSTEM AND MODERN LENDING

2.1. Traditional Banking and Modern Lending

Banking has its roots deep in the past. The evolution of the banking system intensely changed and created an intricate structure of services offered by the banking sector and banking structure itself in the process of time. Historically, the first and the only objective of a bank was to securely store consumer savings. The primary function of a contemporary bank is still accepting deposits from legal entities as well as individuals, acting as a borrower; and providing loans on a time-interest basis, acting as a lender, which enables a bank to perform transformations of savings to investments, in other words, asset transformation.

These days the financial system performs this fundamental function. It serves as a platform for funds channeling: those who have a surplus of their funds (savers) may lend them to spenders, i.e., those who are willing to borrow money. This fundamental mechanism may be of either direct or indirect nature. In the first case, funds are transferred from lenders directly to the financial market and channeled via financial securities to borrowers as a claim for their future income. Thus, securities are assets for creditors and liabilities for debtors. In the latter case, financial intermediaries step in, savers lend their funds to financial institutions, and they, in turn, may lend these funds via financial market or directly to borrowers. Above-mentioned relations foster the

DOI: 10.7172/2353-6845.jbfe.2021.2.2

productivity of the economic system, solving the problems of inefficient capital allocation and lack of liquidity.

Initially, the term ‘peer-to-peer’ (P2P) was created to indicate the process of direct interaction between two parties without the need for the central intermediary being involved. The name originally described a computer network system in which any computer may act both as a server or as a client relative to other machines operating in this network; therefore, a centralized server was no longer required for the network functioning. A sequence of information technology innovations that took place in the first decade of the 21st century led to an enormous expansion of broadband internet usage and peer-to-peer (also interpreted as people-to-people) technology implementation in diverse ways. The P2P technology made a colossal impact of P2P on file sharing. For instance, the appearance of BitTorrent is one of the most popular communication protocols used in the distribution of data and electronic files over the internet. Digitalization created a framework for numerous platform-based markets and aggregators that perform as instruments for buyers and sellers of various goods and services, where main determinants of prices are genuinely demand and supply in the long run and the auction processes or fixed-price offers in a short run. This changed numerous market sectors, including accommodation services (Airbnb, launched in 2008), transport (Uber, launched in 2009), etc. Similarly, technological progress opened new horizons and opportunities for the financial sector by smoothing out the distance and reducing obstacles to access, allowing the market to expand and new services to arise. The FinTech expansion brought in a disturbance to the financial intermediation market in the form of brand-new crowdfunding projects and ventures.

Initially, the P2P lending market consisted of individual investors and small businesses. Over time, large firms and investors have entered the market, and the term “P2P lending” has become less descriptive. The new name – marketplace lending – has come into use. There are some misunderstandings related to the usage of these two terms. However, they are mostly interchangeable and stand for fundamentally the same mechanism that allows matching lenders and borrowers directly through online services. The only difference is in parties involved. In P2P lending, primarily individuals and small businesses are engaged in the lending cycle, whereas in marketplace lending institutional investors enter the market. Nowadays, the marketplace lending may be broken up into consumer lending, business lending, and property lending. Consumer lending constitutes a significant part of marketplace lending and is granted for various purposes, including debt consolidation, credit card refinancing, home improvements, and major purchases. Business lending is actively utilized by manufacturing and engineering companies, as well as businesses operating in transport and utilities. Property lending firms provide services and products and flexible financing models starting from bridging finance to commercial and residential mortgages, and construction and development investment opportunities. The very first P2P lending platforms were Zopa, established in the UK in 2005[2] and Prosper, launched in 2006 in the US These companies laid the foundation for the development of the decentralized marketplace, which enables borrowers and lenders to deal directly with each other without the involvement of a mediator, broker, or intermediary. Zopa is now one of the largest European P2P lending platforms, having the market share on the UK market of around 28.79%.[3]

It is, however, of fundamental importance to take into account that different government regulations apply to P2P platforms and to banks. Generally, fewer regulatory requirements allow broader operational scope at the lower costs. This, however, generates additional risk.

2 BBC UK. (2005). Q&A: Online lending exchange.

3 P2PMarketData. (2019). Accessed October 31, 2019. https://www.p2pmarketdata.com.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

3. LITERATURE REVIEW

3.1. Credit Grade Assigned by a Platform Reduces Information Asymmetry

Recent studies have covered the topic of risk of credit default in marketplace lending. Studies included analysis of loan/borrower characteristics that affect the loan performance. The analysis of 143,654 matured P2P loans funded in 2012–2013 did not reject the hypothesis stating that the credit grade assigned by a platform reduces information asymmetry (Möllenkamp, 2017). That study, entitled *Determinants of Loan Performance in P2P Lending*, found that credit grade is a prevalent determining factor of bad debt, hence a lower credit grade increases the probability of bad debt. Factors that were positively correlated with high loan performance included annual income, debt-to-income ratio, and inquiries in the last six months. The inverse relationship was found between the loan amount and debt performance. The paper *Determinants of Default in P2P Lending* (Serrano-Cinca et al., 2015) studied the determining factors within each credit grade. As in the previous research, annual income, debt-to-income ratio, and inquiries in the past two years along with "Credit Card" and "Small Business" loan purposes were once again found as efficient predictors for each grade class. Also, revolving credit utilization and delinquency in the past two years are useful in the low-risk category (grade A), whereas the length of credit history has shown high efficiency in high-risk (grade C) loan class.

The problem of information asymmetry is addressed in *Disrupting Finance: FinTech and Strategy in the 21st Century* (Lynn et al., 2018). A borrower has nearly complete information, while the information provided by the platform guides the investor most of the time. The book highlights the importance of credit grade assigned by the platforms' preliminary screening based on hard information[4] (i.e., debt-income ratio, number of opened credit lines, etc.). It is argued that for better information disclosure and improvement in decision-making credit scores should be used instead of credit grades, since the latter may not accurately serve as estimates of debtors' creditworthiness.

An empirical investigation of a large sample of PRC's P2P platform containing data on repayment records included in the working paper entitled *Adverse Selection and Credit Certificates: Evidence From a P2P Platform* (Hu et al., 2019) has shown that borrowers tend to attract lenders with high-grade certificates. Certificates are a technique of signaling the presence of information asymmetry. In theory, such licenses have been designed to distinguish borrowers with lower delinquency. Consequently, more funds are loaned to borrowers holding certificates. Despite this, the study has shown that borrowers holding certificates with higher grades have a propensity to higher ex-post delinquency and default rates. The research on **investors** is mainly focused on investment decisions and learning behavior. *A Trust Model for Online Peer-to-Peer Lending: A Lender's Perspective* study (Chen et al., 2014) examined the trust of lenders in borrowers and their willingness to lend via P2P lending intermediaries. The first finding was that the platform's level of service quality and protection significantly affects the lender's trust in the intermediary. The second conclusion was that *"The information quality of borrowers' loan requests is the most important factor influencing lenders' trust in borrowers…"* (Chen et al., 2014). Investors who have suffered financial loss are more liable to herd, thereby lend higher amounts to loan requests that are highly trusted by other creditors (Gonzalez, 2018). The research on the investor side carried out by (Vallée & Zeng, 2019) has confirmed that advanced investors tend to assess loans in a different way than those who are less sophisticated. Moreover, it was proven on the empirical data, that there is a tendency of outperforming by more sophisticated creditors when analyzing loans. However, this outperformance decreases when the platform reduces the applicant's characteristics available to the investor.

[4] Hard information is such information that could be accurately quantified and efficiently transmitted.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

The article *Research on Risk Factors Identification of P2P Lending Platforms* (Lu and Zhang, 2018) complements subject-related literature with the analysis of P2P platform attributes (profitability, risk control, transparency, operation time, etc.) that can determine the probability of a platform being problematic. Data from 2259 P2P lending platforms were taken as a sample from binary logistic regression. It turned out that platforms with higher active operating time and average loan periods tend to be less problematic. The presence of fund custody (support of a third-party managed funds) secures the capital. Furthermore, companies that allow the transfer of creditors' rights and support automatic bidding tend to operate better. Meanwhile, the average interest rate negatively correlates with the platform's riskiness.

3.2. The Overview of Crowdfunding and Other P2P Financial Services

The term "crowdfunding" arose in the early 2006 as a part of a broader concept – crowdsourcing, a name coined by Jeff Howe earlier the same year.[5] Crowdsourcing may be defined as a practice of mobilizing the resources of a substantial number of people to solve specific problems in different areas voluntarily.

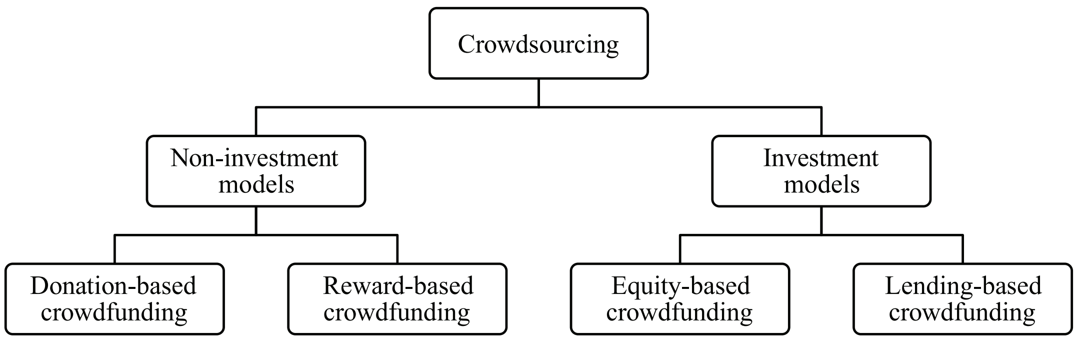
Crowdfunding represents a specific mechanism of fundraising, in which borrowers (capital seekers) may access a pool of capital through interacting with investors (capital givers) by means of a web-based crowdfunding intermediary (peer-to-peer platform). After the rapid technological development accompanied by rapid social media networks growth, capital seekers could easily approach a wide range of individuals interested in supporting innovative business initiatives and ideas. Crowdfunding serves as a general term to describe any type of web-based collective gathering of small contributions from a relatively large number of platform participants for further financing of a recipient (e.g., venture, project). A crowdfunding platform, which is often operated by a third party, manages arising transactions, provides payment facilities, and in some cases, carries out a fundamental analysis of a project before its introduction.

Different forms of crowdfunding may be distinguished by the type of remuneration the capital-givers receive (Lynn et al., 2018). Those types are as follows:

A. **Non-investment models**

B. **Investment models**

Figure 1
Breakdown of crowdsourcing by type of remuneration



Source: Lynn, Theo, John G. Mooney, Pierangelo Rosati, and Mark Cummins (2018). Disrupting Finance: FinTech and Strategy in the 21st Century. London: Palgrave Studies in Digital Business & Enabling Technologies.

[5] WIRED. 2006. The Rise of Crowdsourcing 2006. CNMN Collection.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

A. Non-investment models:

a. Donation-based crowdfunding implies that ventures are funded on a charitable or sponsorship basis and donor[6] has no anticipation of monetary or material return. In general, this type of crowdfunding is used to raise funds for projects not related to entrepreneurship. Experiment is an example of a donation-based platform. The platform serves for "All-Or-Nothing"[7] crowdfunding for scientific research projects.
b. Reward-based crowdfunding is similar to the donation-based one because the backer does not receive any financial remuneration, yet may expect a non-financial reward as a return for a contribution to a project. In this crowdfunding model, backers are driven not only by inherent or societal incentives and opportunity to be credited as funders but also by the possibility to receive merchandise ranging from small symbolic gifts to final products depending on the size of the pledge. Reward-based crowdfunding platforms may operate in either "All-Or-Nothing" or "Keep-It-All." Examples of such platforms are Kickstarter ("All-Or-Nothing) and GoFundMe ("Keep-It-All"). The indicator of total transaction value in the reward-based crowdfunding segment amounted to $6.9 billion in 2019 and is predicted to reach $12.0 billion by 2023 with the compound annual growth rate of 14.7% in 2019–2023 (Statista 2019).

B. Investment models:

Capital providers, involved in the mechanism of investment crowdfunding, may expect to receive some sort of remuneration in the form of financial return.

a. Equity-based crowdfunding (also: crowd investing): investors receive shares in a business, shares in profit generated by this business, and/or the voting power. This form of crowdfunding serves as an instrument for early-stage funding for young and innovative companies and may also help them to bridge the funding gap. The entire procedure may be broken into four steps. In the first step, the company submits its application, including the detailed plan, description, and other required information to the platform. The firm then undergoes a preliminary screening of its appropriateness to crowdfunding, the possibility of being deceitful, reputation, etc. Based on that, a subsequent decision is made on whether to place the business on a platform or to reject the application. The second step is uploading the presentative and investment-encouraging materials for potential shareholders. The third step is gathering the funds, and it continues within the timeframe specified by the platform (case of "All-Or-Nothing" model), funds are held at the escrow account within the funding window. After the deadline, money is transferred to the entrepreneurs provided that the funding target has been achieved; otherwise, funds are returned to the investors. The transaction value of the segment amounted to $4,794.9 million in 2019, with average values of funding $104,115 per application (Statista,2019).
b. Lending-based crowdfunding, the main target of this paper, is, similarly to equity-based crowdfunding, a commercial subtype of crowdfunding. The object of crowdlending is a debt agreement that contains the lender's credit claim to receive interest and redemption payments in the future. This type of crowdfunding is well-developed, holding a significant share of market volume in the industry of crowdfunding. The next section will examine lending-based crowdfunding in detail.

There are more peer-to-peer phenomena aside crowdfunding that are worth studying; however, they are less common. Foreign currency exchange platforms and invoice discounting (a.k.a. invoice trading) platforms that are based on the P2P concept are also interesting examples; however, they will not be studied in this research.

---

[6] According to the CROWD-FUND-PORT terminology, contributors in donation-based crowdfunding are referred as donors, in reward-based crowdfunding as backers, in equity-based crowdfunding as investors and in lending-based crowdfunding as lenders.

[7] Under "All-Or-Nothing" model, the project receives foundation only if the stated funding target is reached withing the prescribed timeframe (Bellefamme, Lambert & Schwienbacher, 2010).

DOI: 10.7172/2353-6845.jbfe.2021.2.2

3.3. Model of Marketplace Lending and Critical Distinctions From Traditional Banking

The primary function of all platforms is generally the same – to serve as a two-sided intermediary and connect the borrower with the lender. Nonetheless, there might be differences in operating mechanisms. Apart from the traditional lending platforms (e.g., Zopa, LendingClub), other ones are launched with the aim to specialize and operate in particular industries, such as AgFunder, focused on the agri-food tech industry. A significant decrease in the number of intermediaries in the process of loan origination and the appliance of new practices to ease financial "frictions" such as information asymmetry and transactional costs considerably decreased the platform's charge on loan transactions. Moreover, several platforms do not charge anything for loan transactions.

There are several methods of categorizing marketplace lending platforms. Firstly, by application domain; companies may be divided into two groups: general platforms and professional platforms. (Wang et al., 2017) General platforms operate in a broad scope of individuals and small and medium-sized enterprises irrespective of loan purposes and intentions. The very first P2P lending platforms (i.e., Prosper and Zopa) originated as general types. Recently, various professional platforms focused on particular application areas have emerged. For example, previously mentioned AgFunder performs as an online venture platform for certified investors to finance agriculture and agricultural technology companies. Another example of a professional platform is LandlordInvest that specializes in supporting borrowers who are having difficulties with borrowing from traditional lenders due to an adverse credit event. The platform enables them to receive financing through buy-to-let mortgages and bridging loans. Although a great deal of marketplace lending platforms rely on unsecured borrowing, LandlordInvest is a representative platform of property-backed marketplace lending. Another form of differentiating between marketplace lending platforms is based on the type of trading rule. (Wang et al., 2017) There are two groups in this category: auction-based and fundraising (nonauction-based) platforms. On platforms operating under auction basis, the price (i.e., interest rate) is determined by the Dutch Auction Rule. A borrower is obliged to construct a loan requirement specification list, which, apart from the information on creditworthiness and other necessary data depending on the platform's regulations, includes the highest interest rate accepted, soliciting duration (i.e., the time interval during which the listing will be open for bids from investors) and the required amount funded.

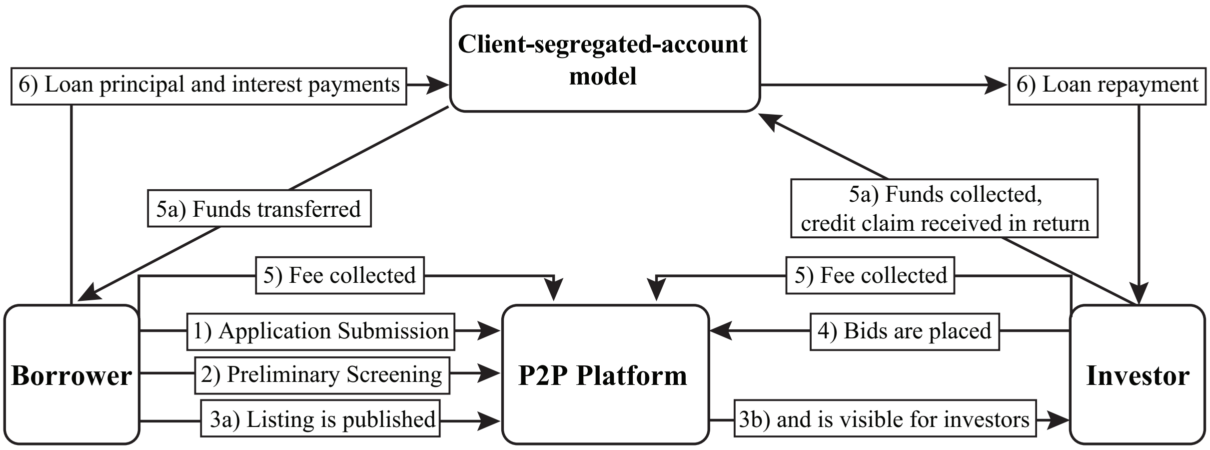
Provided that the platform accepts the loan request, it is posted and is observable for lenders. If a lender is willing to fund this listing during its soliciting interval, a bid is created that reflects the amount of money to be financed, and the minimum interest rate accepted. If cumulative bid amount of a particular listing exceeds the required amount in its soliciting duration, competition among bids will occur based on the interest rate, i.e., bids with higher rates will be outbid, and the bids with lower rates will be accepted. After the soliciting duration, the final treading rate is the same for all investors whose bids succeeded in an auction and is defined as the maximum rate of all successful submissions. As in the "All-Or-Nothing" principle, if the listing fails to gather the stated amount funded withing the soliciting period, it is expired, and all bids made are canceled. Based on the foregoing process, investors may also analyze the probabilities of their bid winning the auction on the particular listing and the likelihood of this listing being fully funded withing the soliciting period when making an investment decision.

Due to the complexity of an auction, most platforms ended their auction process and changed the trading rule. For the sake of high quality customer service and trading efficiency, they decided to carry out a less sophisticated procedure – fundraising. Thereby, company Prosper ended its auction after five years of operating in 2010.[8] Fundraising may employ either a fixed ("All-Or-Nothing") or flexible ("Keep-It-All") principles of setting the funding target, which were discussed in the previous section.

[8] Renton, Peter. (2019). Prosper.com Ending Their Auction Process. December 16. Accessed December 27, 2019. https://www.lendacademy.com/prosper-com-ending-their-auction-process-dec-19th.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Figure 2
A process map of the client-segregated-account model



Source: Lenz, Rainer. (2016). Peer-to-Peer Lending: Opportunities and Risks. European Journal of Risk Regulation, pp. 688–700.

A general model may be described in chronologically ordered steps:

1. An individual or institutional borrower sends an application via the internet platform. The application consists of the amount requested and the maturity of the loan. Also, depending on the platform, the borrower is inquired to hand over additional information, such as borrowing history, credit certificates, debt to income ratio, employment length in years, amount of opened credit lines, etc.
2. After the application submission platform conducts a preliminary assessment of underlying credit risk based on the information provided and decides on whether the applicant matches the platform's risk categories. Some platforms assign a credit grade or score to reflect the riskiness. Finally, the platform offers the risk-appropriate interest rate to the borrower.
3. At this stage borrower may reject the proposal and exit the market. Otherwise, the application is listed on the platform for a defined soliciting duration. Individual loan listings are usually anonymized, while institutional ones are published with the borrower's title.
4. To become an investor, one needs to sign an agreement with the platform and complete the due diligence proceeding as a part of the Anti-Money Laundering Rules. Investors remain anonymous on the platform and assigned a coded username. During the soliciting period, investors may place their bids and observe the remaining amount required to match the funding target.
5. If a listing collects the funding target within the soliciting period, the loan money is obtained from the investors and is transferred to the borrower. Investors, in return, receive a document that writes down credit claims with the corresponding portion of the total loan principal and interest to be repaid by the borrower. Before that, the platform collects a fee from both parties: investors and borrowers. The critical point is the fact that the platform does not store the funds collected from investors. Transfers of funds are conducted simultaneously as counterclaims. (Lenz, 2016) There are three main loan origination models (Havrylchyk & Verdier, 2018):
   a. In the "client-segregated account" model, mostly exercised by the UK platforms, the platform itself originates the loan, but all the money flows through legally segregated client accounts. It is kept strictly separated from the platform's balance sheet. In the case of platform insolvency, creditors have no claim on the platform's client funds, and the contractual agreements of peer-to-peer loans remain valid.
   b. Opposite to the UK, the US and most European countries have different national banking regulations: origination of loans is allotted to licensed banks only. A "notary" model with credit institution (for the most part – commercial banks) involvement turns out to

DOI: 10.7172/2353-6845.jbfe.2021.2.2

be obligatory for loan origination and payment service. After the borrower's application collects the funding target from investors on the platform, the loan package is hand to the partner bank, which originates the loan in the required amount. In 2–3 days, once the partner bank transfers funds to the borrower, the loan is sold to the marketplace company. At this point, the borrower's repayment obligation is transferred to the bank-affiliated marketplace company. The latter eventually issues notes to lenders, which reflect the corresponding share of funds that have been invested. The remaining steps mirror the "client-segregated account" model. The charge for White-Label-Banking intermediation depends on the volume of credit and ranges typically from 0.5% to 1%. As a rule, the identity of the partner bank is not revealed to the end-user.

c. In the "guaranteed-return" model, the platform acts similarly to the «client-segregated account" model and manages the investments of borrowers and repayments of lenders directly. However, a guaranteed return rate for borrowers is set by the platform. (CreditEase in China).

6. The last step is servicing the loan, collecting and dealing out interest and possible recovery payments up until the loan maturity date. Generally, marketplace loans are arranged in a form of monthly annuity loans. In the event of debtor's default, the platform is to arrange the collection of payments for account of crowd investors. Nevertheless, the platform is not legally responsible for possible losses carried by lenders. Some platforms practice sale of defaulted loans for the account of lenders to a debt-collecting agency for an agreed price in order to partially recover the credit claim. Others have developed automated litigation and recovery processes for defaulted credit lines. In the latter case, the recovery rates are higher.

As in traditional lending, the problem of information asymmetry may arise when the platform attempts to assess the borrower's creditworthiness. In the case of conventional banking, the assessment is mainly based on the analysis of systemized, implicit, hard information (i.e., financial statements, tax reports, etc.). Apart from this type of data, banks often possess non-codified information that was collected through an interview or obtained from previous credit history while dealing with a long-time customer. In P2P lending, the company is unable to acquire such information due to the lack of personal contact with a customer and the time scarcity devoted to deciding on the approval and level of the interest rate. A concept of big data comes into play instead. The structure of contemporary social media services inevitably leads to an individual's digital social footprint in the form of social media activities, preferences, age, education, social circle, etc. These data may effectively substitute the personal interviews and other conventional methods of forming the level of interpersonal trust and assigning a credit score. Companies use special software that is often based on machine learning to conduct credit scoring, pricing and to decide whether to accept or reject the borrower's loan request, autonomously and without the involvement of the platform's management. As already mentioned, if the proper software architecture is used, there is the negligible cost of assessing a marginal loan request. However, the target percentage of failures to predict the outcome has to be met.

Another substantial difference from traditional banking is the lack of credit risk presence on platforms' balance sheets. This fact relaxes the requirement for an equity loss-absorption buffer and the need for partial coverage of the originated loan with their equity capital. Thus, there is a lack of dependence between the value of queries and the equity requirement. Platform clients benefit from the lower cost of funds for borrowers and/or higher returns for the investor. The aggregate benefit equals the banks' interest margin, which is not charged in this case, less platform fees. In traditional banking, an institution obtains profit relying on interest margin between deposits held and loans provided. This does not apply to marketplace lending companies

DOI: 10.7172/2353-6845.jbfe.2021.2.2

since they derive revenues from the transaction, servicing, loan origination, and other fees. Their profits, therefore, are directly unaffected by interest rate market fluctuations. Loan origination fees are deducted from the loan before transferring funds to a borrower. Origination fees vary across platforms and depend on the value of credit and type of borrowers, starting from 1% for large businesses and reaching 6% for SMEs. The servicing fees are calculated per annum based on the amount outstanding on any loan and are deducted from the loan repayments made by borrowers. Servicing fees vary less and are, on average, around 1%.[9] Companies are indeed interested in processing as many queries as possible since their revenue is partially subject to it. At the same time, an intermediary is motivated to act prudently and conduct adequate credit risk assessments since the platform's reputation and revenues are subject to the rate of return yielded for investors.

4. HYPOTHESIS AND PEER-TO-PEER LENDING MODEL

We verify the following research hypothesis: The method of credit scoring is applicable in alternative lending environment. Additionally, the quality of the final version of the logistic regression model and, thus, the scorecard, may be enhanced by more advanced variable pre-processing. In our case, variables binning based on selected indices (Weight of Evidence and Information Value) allowed to pre-select the most meaningful explanatory features. Investors select the preferred cut-off point subject to their risk acceptance level. To do so, they apply an expected profit/loss method, and based on the specificity and sensitivity values, choose the cut-off point subject to the highest expected profit.

In order to confirm or deny the above-mentioned hypothesis, the research which explores recent studies and open-source data on marketplace lending is done. The scorecard development is based on the P2P loans open data set that contains repayments record along with both hard and soft features of each loan. The quantitative part consists in applying a machine learning algorithm in building a credit scorecard, namely logistic regression. The objective is, through descriptive and quantitative analysis, to select the best features that allow for differentiating the loan performance in the marketplace lending environment and process the data, followed by scorecard construction and quality assessment.

4.1. Marketplace Lending From the Lender's and Borrower's Perspective

Investors may estimate the annual risk-adjusted returns received by subtracting the annual servicing fee and annualized bad debt loss from the gross profit (gross interest rate). Table 1 represents the annualized return less fees and bad debt losses by platform and year of loan origination. The values of net ROI varied significantly in 2015; however, the variance has decreased, accompanied by an increase in average return approaching 2020. These values, however, are applicable only in the case of a well-diversified portfolio containing a high number of loans. At this point, investors may benefit from diversification software instruments that may process automatic order placement depending on the preset amount invested per loan, risk grade, maturity, etc.

[9] Oxera Consulting LLP. (2016). The economics of peer-to-peer lending. Independent economic assessment, Oxford: Peer-to-Peer Finance Association.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Table 1
Annualized return less fees and bad debt losses by platform and year of loan origination

Platform / Year	2015	2016	2017	2018	2019	2020
Lending Club (US, SME, PL*)	4.69%	4.31%	4.75%	4.81%	6.66%	N/A
Funding Circle US (US, SME)	2.6–2.8%	4.1–4.9%	5.3–6.2%	5–6.3%	5.7–7.8%	N/A
Rate Setter (US, PL)	4.8%	4.3%	4.0%	4.4%	4.4%	N/A
LendingCrowd (UK, SME)	6.92%	5.24%	5.53%	8.05%	7.94%	9.16%
MarketFinance (UK, SME)	2.88%	4.46%	4.83%	5.96%	6.39%	7.25%

* Personal loans.

Source: Funding Circle (2019), LendingClub (2019), RateSetter (2020), LendingCrowd (2020), MarketFinance (2020).

A comparison of these values with interest rates that are offered on deposit bank accounts shall also be avoided. The investments on the P2P lending market are, most of the time, unsecured, and the capital invested is fixed until the maturity date. In contrast, funds on the bank account (except time deposit account and other non-transaction accounts) may be withdrawn on demand and without a fee. Despite the existence of secondary marketplace lending market, there is no guarantee of exit without high expense as a result of a discount. Moreover, according to the EU Directive on Deposit Guarantee Schemes, deposits on bank accounts at EU banks are guaranteed by EU member states up to a level of €100,000 per person per bank.

The investment risk in a particular loan request may vary. A classical concept of risk-return tradeoff is applicable, similarly to the one present in the case of portfolio provided by a corporate bond investment fund that consists of corporate loans. The risk also depends on the type of loan, since some platforms host not only unsecured loans but also asset-backed ones (e.g., property-backed). The existing and properly managed buffer fund may considerably reduce the lender's risk burden and smoothen the investment result in case of a bad debt or recession.

A large number of platforms make their up-to-date statistics (including annualized returns, projected and historical bad debt rates, lifetime default rates, the volume of buffer fund, etc.) publicly available on their webpages. Investors may collect their portfolio performance for a given period. However, neither these indices nor techniques of their calculation are standardized. The industry lacks a framework of rules and regulations for clear, well-defined standards for performance evaluation. Likewise, disclosure standards for information about borrowers or platforms' credit assessment methods are yet to be defined. As a result, this may create an obstacle for an investor to compare platforms adequately and to decide which platform to select. The regulatory issue will be studied more broadly in the following chapter.

Borrowers benefit in terms of additional choice of loan options offered by marketplace lending, which are now broadly comparable to traditional banking solutions when it comes to the cost of borrowed funds. The emergence of marketplace lending brought an additional portion of the competition to the lending industry. As a result, SMEs may access funds from an additional source. That is, the share of funds borrowed by SMEs from traditional channels has fallen by more than a fifth in recent years. The Funding Circle in their survey of SME clients has noticed that the rise of popularity of alternative sources of finance is caused by shorter period from submitting application and loan pay-out (31% of customers) and simplicity of obtaining a loan (28% of customers).

P2P lending is accessible online at any time of the day; the number of documents and forms is rather low, which reduces bureaucracy. Other borrowers also notice the lack of collateral required for the majority of loan requests and the possibility of premature loan cancellation without a fee imposed. Borrowers with bad credit history and those unable to access banks benefit from an

DOI: 10.7172/2353-6845.jbfe.2021.2.2

additional source of funding. 21% of Funding Circle customers report that they wouldn't be able to access the funds through a bank.[10] One may presume the presence of adverse selection: borrowers with low default risk will borrow from banks, and those with higher default risk will enter the marketplace lending. There is, however, no empirical evidence to prove that statement.

The major drawback of the model of P2P lending is that a potential borrower cannot be sure if they will get the required funds even if a platform accepts the application. Given the specific loan volume, interest rate, maturity, and credit grade, lenders may refuse to supply the needed amount of funds. To address this problem, platforms often raise the interest rate until the offer becomes sufficiently attractive. The next shortcoming of the marketplace model is that credit risk assessment lacks disclosure; borrowers are not aware of the data that the platform uses to analyze one's creditworthiness. This may bring a possible problem of discrimination into the industry based on gender, race, migration status, etc. The problem may be solved by introducing an appropriate legal framework.

5. RESEARCH OF METHOD OF SCORECARD CREDIT RISK ASSESSMENT

5.1. Concepts of Credit Scorecards and Linear Regression Machine Learning Algorithm

One of the most critical factors in investors' profitability and prosperity of their lending decisions is their ability to adequately measure credit risk involved in loan requests and borrower's creditworthiness in particular. One option is to refer to the subjective technique to estimate the probability of default (PD); alternatively, one may apply the objective approach to credit risk assessment – method of credit scoring. Credit scorecards are widely utilized by banks to distinguish "bad" clients from the "good" ones, since they may benefit from extensive client data collected from their experience or access databases of credit information bureaus.[11] Although a typical non-institutional marketplace lending investor has no access to such comprehensive data, this technique still may be of particular interest, since a platform discloses certain loan and borrower's features to investors. Among others, an investor may observe borrower's Debt-to-Income Ratio (DTI), the number of derogatory public records, total credit revolving balance, latest FICO Score[12] range, and many more. Also, listing-specific grade, interest rate assigned by platform itself as well as loan amount, and Equated Monthly Installment (EMI) are displayed. Thus, credit scorecards appear as quite an attractive objective technique for an investor to assess the creditworthiness of a particular loan request, since data are already provided.

There are some significant benefits of scorecards for credit assessment; for instance, it removes the possible bias which may arise when analyzing only good non-defaulted applications, thus minimizing the survivorship bias risk.[13] Given that credit scorecards are founded on fairly large data samples, they may include a wide range of features to extract the correlation between variables and bad loan performance. Despite the vast number of characteristics and observations, the algorithm's processing time is efficient, which minimizes process time and cost and produces fewer errors.

In the classical credit scoring approach, there are two types of scoring techniques: application and behavioral. The principal difference is that the application scorecard (AS) is created for

[10] Funding Circle. (2016). Small Business, Big Impact: The changing face of business finance. Evidence from Funding Circle, London: Centre for Economics and Business Research.

[11] An example of such credit bureau is Biuro Informacji Kredytowej S.A. (BIK) – an organization established by the Polish Bank Association and private banks, which gathers, processes and shares data on the credit history of banks', credit unions' customers and also some non-bank lending companies.

[12] FICO® score is one of the most well-known credit scores designed by the Fair Isaac Corporation.

[13] Survivorship Bias Risk is the risk that an investor's decision may be misguided when considering only "good" loan requests based on published return data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

a specific lending company and a particular product (e.g., revolving loans, mortgage loans) and utilizes its historical data to evaluate at the application stage. They may include characteristics such as personal data, application data[14], and information provided by credit bureaus. On the contrary, behavioral scoring (B.S.) is predicated on based on time-dependent attributes of debtors and on how these attributes change once the loan contract is originated. They may take into consideration the borrower's credit behavior (credit limits, number of current credit lines, open bank accounts, deposit balance, granted credits, etc.). The general problem of credit scorecards is the lack of an explicit theory behind the chosen independent variables in classifying the loan outcome. There are, however, some papers that provide advice on variable selection. The general recommendation is to select interpretable variables based on discriminatory power, future availability, legal issues, etc.[15] The number of variables in scorecard should lie in between 8 and 15 to provide stability and keep relatively high predictive power even if the profile of one or two variables changes. Scorecards with an insufficient number of characteristics are more vulnerable to minor changes from the applicant's profile, making the scorecard unable to remain stable over time. Recent research confirms that there is no universal number of variables that should be included in scorecard development.

The idea of a credit scorecard is to choose such a cut-off score in which the final sum of scores for each attribute is present in the scorecard for a particular application. There are various techniques to determine specific scores and cut-off points. Generally, these methods are divided into parametric ones, where the number of parameters is finite and fixed with respect to data (e.g., linear regression), and non-parametric ones, where the potential number of parameters is independent of data and may potentially be infinite (e.g., decision trees, neural networks). This paper is going to focus on parametric statistical techniques, or more precisely – on logistic regression. The logistic regression algorithm is a regression analysis technique that belongs to generalized linear models (GLMs), designed to analyze the relationship between a dependent (explained) variable and one or more independent (explanatory) variables, in other words – regressors. This model has close ties with the classical linear regression model (CLRM); however, the latter is intended for continuous dependent variables only, meanwhile the logistic regression functions with binary and categorical variables with more than two levels. Depending on the form of dependent variable models are classified as: binary logistic regression – model with binary dependent variable; multinomial logistic regression – model with unordered categorical dependent variable with more than two levels; and ordinal logistic regression – model with ordered categorical dependent variable with more than two levels.

Binary logistic regression is a suitable instrument for credit scorecards development since the dependent variable is a good/bad flag that represents the loan outcome – bad meaning failure to pay and good – successful repayment. In contrast to CLRM, it calculates the conditional probability of dependent variable taking a specific value (0 or 1 if the dependent variable is coded as a binary variable) subject to the values of independent variables, for instance, in the case of one independent variable $p(X) = Pr(Y = 1/X)$, where Y is dependent, and X is independent variables. Parameters reflect the relationship between explained and explanatory variables, such that: $p(X) = \beta_0 + \beta_1 X$. Fitting a straight line would be inappropriate in case of a binary outcome; therefore the sigmoid-shaped function is used: $p(X) = \dfrac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \dfrac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$; alternative form is $\dfrac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 X}$; where the left-hand side (LHS) of the equation is defined as odds ratio that ranges from 0 to $+\infty$, indicating low and high probabilities of event $p(X) = Pr(Y = 1/X)$ correspondingly. Taking the natural logarithm of both sides gives the logistic

[14] E.g. term, requested amount, EMI, purpose, joint or individual application, collateral, etc.

[15] Siddiqi, Naeem (2017). Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards. Hoboken: John Wiley & Sons.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

regression function (logit): $\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$. Similary to CLRM, a one-unit increase in X increases the value of LHS (logit) by β_0. The change in conditional probability, $p(X)$, depends on the value of an independent variable. For multiple independent variables: $p(X) = \frac{e^{X\beta}}{1+e^{X\beta}}$; where X is the matrix of independent variables, and β is the matrix of parameters.

The Maximum Likelihood Estimation (MLE) method is used to find the matrix of estimates for parameters β. The Likelihood Function takes the following form $L(\beta) = \prod_{i=1}^{n} p(x_i)^{y_i}(1-p(x_i))^{(1-y_i)}$. Once parameters are estimated, the probability that the dependent variable takes value 1 may be found for a specific combination of independent variables. For one unit increase in an independent variable x_k, the change in odds ratio is e^{β_k}.

5.2. Database and Features Description. Initial Data Cleaning and Processing

The main instrument of the quantitative part of research and modeling is an integrated development environment for R language – RStudio 1.2 combined with the smbinning package. The initial data set contains full Lending Club information on accepted loan applications for the period from 2007 up to the 3rd quarter of 2019 with 150 variables and 2 650 550 observations. Some variables require significant cleaning. Several characteristics are available only ex-post from the database; thus, they are not visible for an investor on the platform's website. Given that the aim is to construct a scorecard that will be useful in practical terms, one shall choose among variables that are available for an investor when deciding to lend money or to forgo a particular listing. Variables of interest were picked, provided that they are available on the platform website. The dependent variable is Loan Status, it is a categorical (factor) variable with eight levels, according to the LendingClub data dictionary:

– Charged Off – Loan for which there is no longer a reasonable expectation of further payments. Generally, Charge Off occurs no later than 30 days after the Default status is reached.
– Default – loan has not been current for 121 days or more.
– Fully Paid – loan has been fully repaid, either at the expiration of the 3- or 5-year term or as a result of a prepayment.
– Issued – a new loan that has been approved by LendingClub reviewers, received full funding, and has been issued.
– Current – loan is up to date on all outstanding payments.
– In Grace Period – loan is past due but within the 15-day grace period
– Late (16–30 days) – loan has not been current for 16 to 30 days.
– Late (31–120 days) – loan has not been current for 31 to 120 days.

The defaulted credit line is assigned default status once the payment is delayed for 121 days (i.e., for an extended time). The charged-off state is consecutively assigned to defaulted loan, and the remaining principal balance of the note is deducted from the investor's account balance. Thus, these statuses indicate the same practical loan outcome – default and differ in a formal principal deduction from an account. In this research, a bad loan outcome is recognized as either Charged Off or Default status of the credit line. The Fully Paid state is perceived as good loan outcome. Listings with other states are disregarded and removed.

Table 2 presents the description of the dependent variables that have been selected from the initial pool of features. After the variable selection and data cleaning, the approximate number of observations is more than 1.2 million. The handling of such large amount of data is resource-consuming. Therefore, after removing listings with missing information, 400 000 observations were randomly selected from this data set. An additional binary variable (good/bad flag) "DEF" was introduced with values 1 for bad loan outcome and 0 for good loan outcome.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Table 2
Description of independent variables

Variable title in R	Description
total_acc	The total number of credit lines currently in the borrower's credit file. **Numerical variable.**
term	Loan duration. Values are in months and can be either 36 or 60. **Factor variable** with two levels: "36", "60".
revol_util	Revolving line utilization rate. **Numerical variable**.
revol_bal	Total credit revolving balance. **Numerical variable**.
pub_rec	Number of derogatory public records. **Numerical variable**.
home_ownership	Home ownership status provided by the borrower or obtained from the credit report. **Factor variable**. Levels*: "Rent", "Own", "Mortgage".
inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage). **Factor variable** with nine levels from "0" to "8".
open_acc	The number of open credit lines in the borrower's credit file. **Numerical variable**.
mort_acc	Number of mortgage accounts. **Numerical variable**.
loan_amnt	The listed amount of the loan applied for by the borrower in $ US **Numerical variable**.
avg_fico**	The average of upper and lower boundary range values the borrower's last FICO belongs to. **Numerical variable**.
int_rate	Interest Rate on the loan. **Numerical variable**.
installment	Equated Monthly Installment (EMI) in $ US **Numerical variable**.
grade	Loan grade assigned by LendingClub. **Factor variable**. Levels: "A", "B", "C", "D", "E", "F", "G".
emp_length	Employment length in years. **Factor variable**. Levels: 12 level from "< 1 year" to "10+ years".
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LendingClub loan, divided by the borrower's self-reported monthly income. **Numerical variable**.
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years. **Numerical variable**.
annual_inc***	The self-reported annual income in $ US provided by the borrower during registration. **Numerical variable**.

* Initially, the variable contained the level "Other", which has been omitted.
** Generated as an arithmetic average of "last_fico_range_low" and "last_fico_range_high" variables.
*** Observations only with verified annual income are included in the final data set.

Source: LendingClub. (2018). "Data Dictionaries." LendingClub. Accessed March 28, 2020. www.help.lendingclub.com/hc/en-us/articles/216127307-Data-Dictionaries.

Table 3 presents the summary descriptive statistics: mean values, standard deviation, as well as minima and maxima values of each explanatory numeric variable. The first three variables in the table (i.e., annual income, revolving balance, and loan amount) have very high standard deviation values, standing out from the rest of features and generating quite a diverse data set with diverse applicants.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

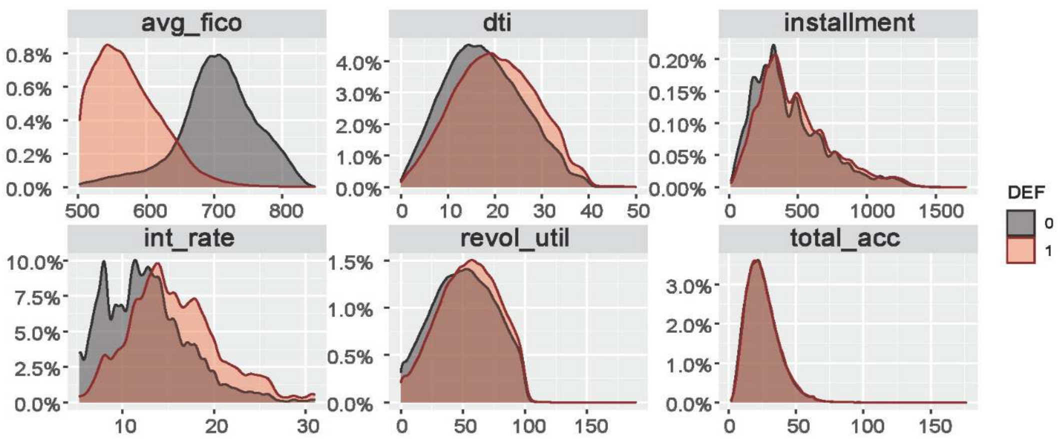
Table 3
Descriptive statistics for independent numeric variables

Variable	Mean	Std.Dev.	Min	Max
annual_inc	77568.86	71367.47	2500	9550000
revol_bal	16473.48	22523.11	0	2560703
loan_amnt	14454.94	8706.56	1000	40000
avg_fico	680.26	76.21	502	848
installment	439.58	261.33	14	1720
revol_util	51.50	24.46	0	189
total_acc	25.30	12.05	2	176
dti	18.18	8.38	0	50
int_rate	13.18	4.75	5	31
open_acc	11.74	5.52	1	84
mort_acc	1.68	2.01	0	37
delinq_2yrs	0.33	0.89	0	30
pub_rec	0.22	0.60	0	47
Obs.			400,000	

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

Figure 3 consists of kernel density approximations for several continuous variables by loan outcome. Despite the generally positive (right) skewness tendency, most variables are approximately bell-shaped. At this step, conclusions about the data may already be drawn. Some variables have quite high (e.g., Average FICO and Interest Rate) and moderate (e.g., Debt-to-Income) discriminatory power. Whereas some variables (e.g., Total Number of Credit Lines) have negligible differences in distributions depending on loan outcome.

Figure 3
Kernel density estimations for selected numeric variables

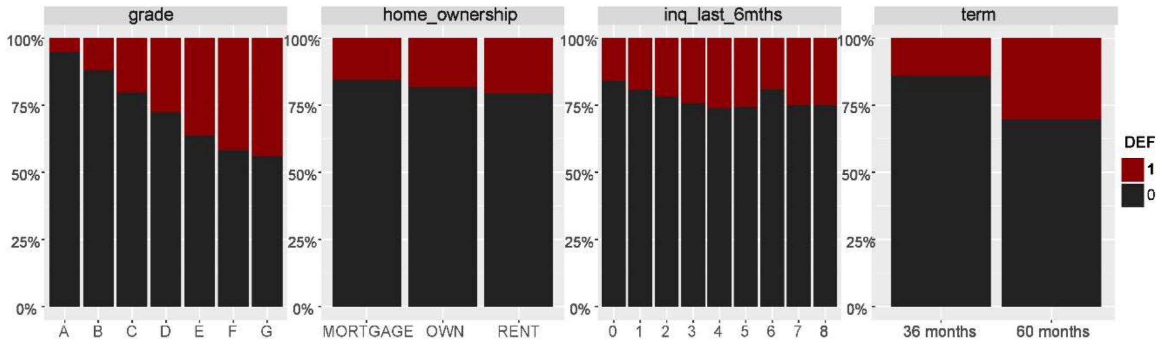


Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Figure 4 allows for graphical analysis of selected factor variables subject to the loan outcome. The situation is similar, the percentage of defaulted loans differs noticeably by grade and term. However, the relation is not that distinctive in case of home ownership and inquiries during the last 6 months.

Figure 4
Levels of factor variables by loan outcome



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data

Moreover, the inquiries in the last 6 months are an ordinal factor variable, and the general relation in positive, the percentage of defaulted loans grows as the number of inquiries increases; however, there is an apparent nonlinearity in from of bad rate drop created by level "6".

5.3. Variables Pre-Processing. Fine and Coarse Classing

Since the scorecard development is based on logistic regression, explanatory variable transformations and addressing data issues are required. Rather than proceeding with an analysis of variables predictive power, solving problems of nonlinearities and outliers manually for each feature, this research suggests implementing an algorithmic method of variable transformation as the first step of variables pre-processing.

As a screening benchmark for pre-processing, this research employs the Fine Classing concept. It helps to reveal the structure of every single variable and its relationship with the dependent variable. Fine classing suggests that the variable is binned based on Weight of Evidence (WoE) and Information Value (IV) indices. This research uses the quantile approach, meaning that the number of bins is subject to the type of quantiles. More precisely, the decile method is applied through smbinning.custom function. As a result, the number of bins is always fixed and is equal to 10.

When it comes to the factor variables, at this point of initial pre-processing, factors are not changed. Weight of Evidence (WoE), a measure of the predictive power of the independent variable, discloses the relationship between dependent and explanatory variable and may be calculated for i-th bin as $\mathrm{WoE}_i = \ln\left(\frac{\%\text{ of non-defaults}_i}{\%\text{ of defaults}_i}\right)$.[16] As follows, the higher the relative share of non-defaults in a particular bin, the higher the WoE for that bin and, therefore, observations related to that bin are less prone to default.

[16] $\%\text{ of defaults}_i = \frac{\text{no. of defaults subject to bin}}{\text{total number of defaults}}$; $\%\text{ of non-defaults}_i = \frac{\text{no. of non-defaults subject to bin}_i}{\text{total number of non-defaults}}$.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Table 4
Indices for univariate analysis

Variable	IV	GINI	Correlation
avg_fico	4.1023	0.8680	
grade	0.4555	0.3586	int_rate
int_rate	0.4398	0.3576	grade
term	0.1930	0.1984	
dti	0.0832	0.1639	
loan_amnt	0.0501	0.1248	installment
installment	0.0419	0.1059	loan_amnt
mort_acc	0.0277	0.0911	home_ownership
revol_util	0.0287	0.0900	
inq_last_6mths	0.0279	0.0847	
annual_inc	0.0206	0.0798	
home_ownership	0.0231	0.0792	mort_acc
open_acc	0.0119	0.0621	
emp_length	0.0112	0.0397	
revol_bal	0.0034	0.0326	
pub_rec	0.0056	0.0286	
delinq_2yrs	0.0041	0.0249	
total_acc	0.0016	0.0183	

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

The next step is discriminatory power assessment of variables and univariate analysis by dint of: GINI index (G) – measure of discriminatory power, higher values indicate higher discriminatory power; Information Value (IV) – another distinguishing power index, higher values indicate higher predictive ability. To recognize collinearity, Kendall's Tau[17] is calculated. The summary of indices and correlation analysis for each transformed feature are represented in table 4. Variables for which the Kendall's Tau exceeds **0.5** are displayed in the last column pairwise. Variables with a GINI index lower than **0.9** or IV lower than **0.25** are considered as weak predictors and are omitted in further analysis. If two variables are highly correlated and both satisfy GINI and IV thresholds, then the one with lower GINI is omitted. Variables that meet the above conditions are: "avg_fico", "grade", "term", "dti", "loan_amount", "mort_acc" and "revol_util". Appendix 1 contains complete sets of fine classing algorithm output graphs with descriptions for the abovementioned variables.

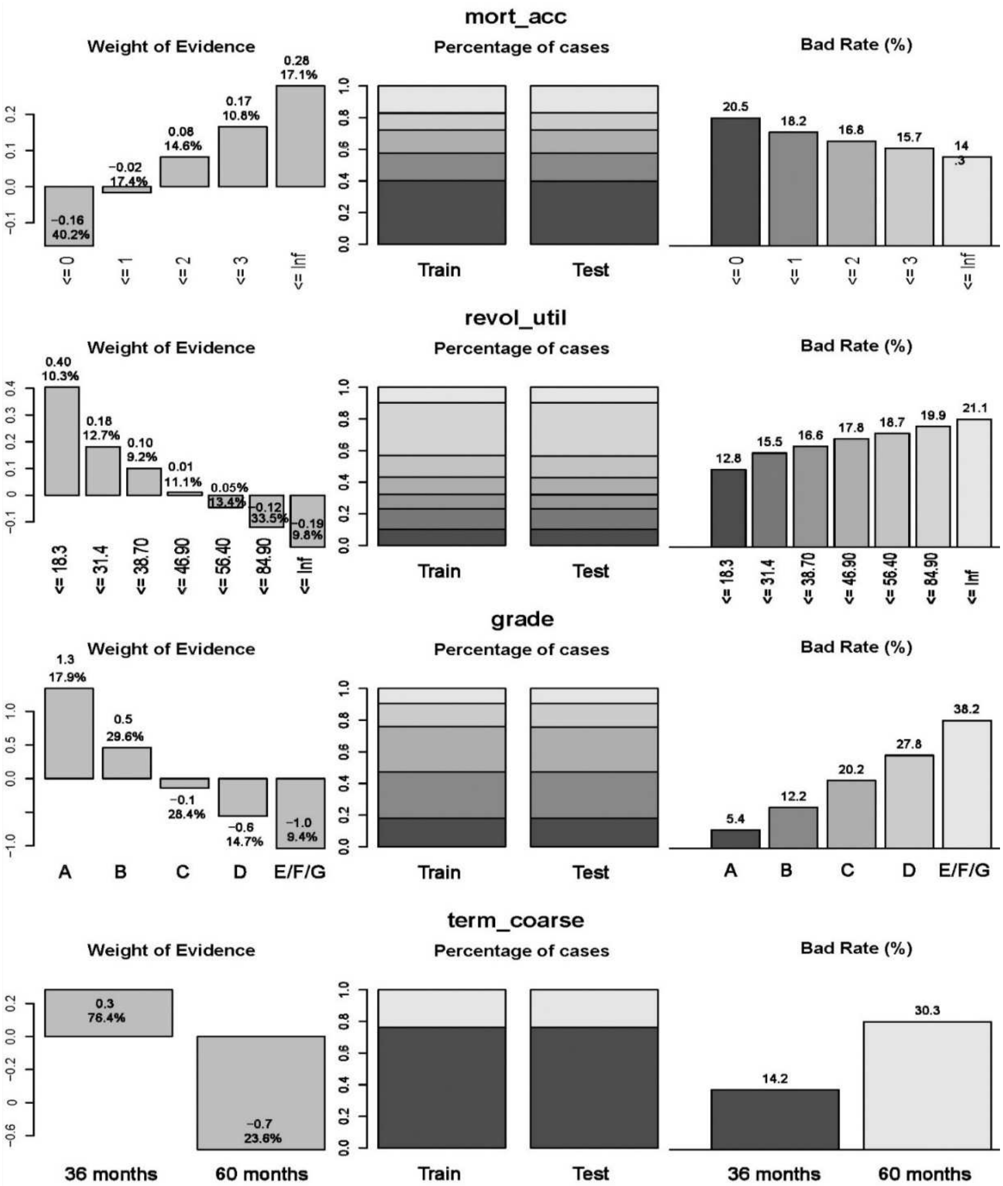
The last phase of sample pre-processing is generating a test subsample used to build a scorecard and train subsample used for validation. Best practices suggest that in case of sufficiently large samples, the train subsample constitutes from 70% to 80% of initial data. (Siddiqi, 2017) To ensure the preservation of initial bad and good outcomes' proportions, sampling with stratification (proportional sampling) is used. After the data splitting, train sample contains 70% of observations, and the percentage of defaulted loans is equal to 18%.

[17] This coefficient is appropriate for the calculation of correlation between ranked (binned) data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Bins generated by Fine Classing are not used in regression analysis. Coarse Classing is the following step to create more representative classes that will be used in modeling. Although Coarse Classing uses the same statistical measures, it is a more advanced technique. The smbinning package works in a tree-like method. Using the Conditional Inference Trees algorithm, it iteratively splits and then merges bins with similar WoE with respect to the dependent variable and maximizes the difference between classes, at the same time keeping the Information Value above the target level. The lower bound of IV is set at 0.1. Results of Coarse Classing of train sample for numeric variables are presented in Figures 5 and 6. Weight of Evidence diagrams give a picture of WoE values for each bin of specific variables (values on the top/bottom of each bar). Under these values, the share of observations contained for that specific bin in percentage is displayed.

Figure 5
Summary graphs of Coarse Classing, part I

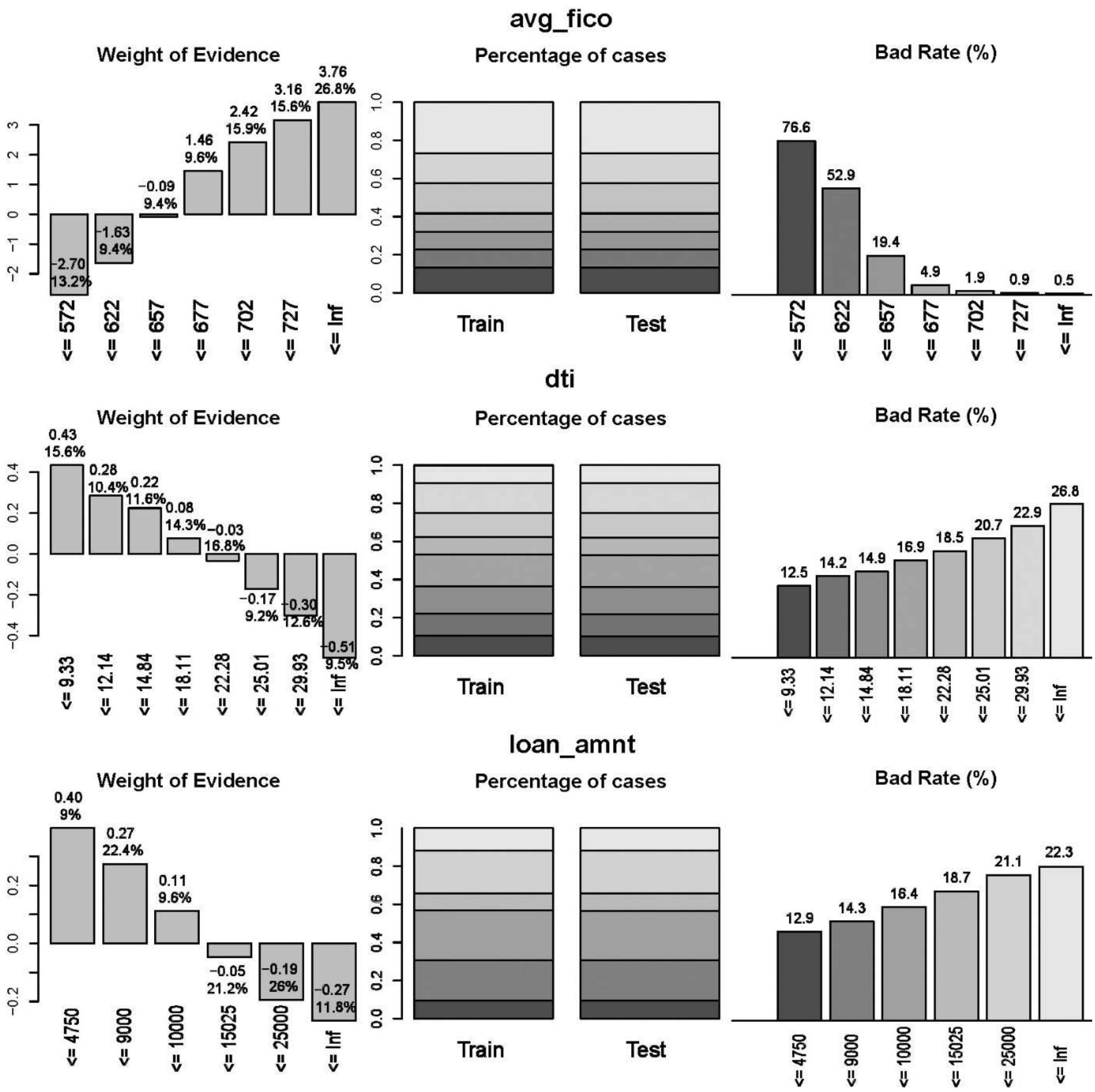


Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

Percentage of cases bar plots can be used to compare the share of observation contained in each bin in train and test subsamples for each variable. Generally, it is preferred, that these values are approximately the same.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Figure 6
Summary graphs of Coarse Classing, part II



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

The third graph in each set – Bad Rate (%), simply illustrates the percentage of defaulted loans in each bin of a specific variable. These sets of charts may be used to analyze the quality and adequateness of Coarse Classing. There are several details to be checked:

– each category (bin) should have at least 5% of the observations. Fine Classing indicated that variable *"grade"* has two underrepresented classes, namely "F" and "G". Since the WoE values of these classes were comparable and to prevent the overfitting, classes "E", "F" and "G" were merged into one level "E/F/G" with the cumulative percentage of 9.4%
– each category (bin) should be non-zero for both non-events and events. Neither Fine Classing nor Coarse Classing has shown that issue. Bad Rate is non-zero for all bins of each variable
– the WoE should be distinct for each category. Similar groups should be aggregated. Although, after Fine Classing, there were some bins with similar/same WoE, after Coarse Classing, this issue was eliminated
– the WoE should be monotonic, i.e., either growing or decreasing with the groupings. Fine Classing revealed the lack of monotonicity for variable *"loan_amnt"*. The problem was resolved by increasing the lower bound for each bin up to 9% in smbinning function.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Since each point of the checklist is satisfied, the obtained discretization is appropriate. Initial independent variable values that are contained in the same bin are replaced with the WoE value of that particular bin for further logistic regression modeling. Thus, the amount of unique values for a variable is equal to the number of bins after Coarse Classing. Classifying with respected bounds and WoE values obtained from analyzing train sample are also substituted into the test sample. Nevertheless, these variables are treated as continuous in further modeling.

5.4. Modeling. Scorecard Development

Table 5 contains summary table of the final logistic regression model. Since initial values of variables are substituted with WoE, all estimates have to be negative, as a property of WoE transformation. Variable *"revol_util_woe"* has been excluded, since it has non-meaningful positive value of estimate. All variables are individually statistically significant according to the *Z*-value of Wald Test even at significance level as low as 0.01.

At the next step, based on the estimated model, fitted values (i.e., probabilities of default (P.D.)) and values of logit function are assigned to each observation for both train and test samples. Then, P.D.s are scaled to obtain scores. The following formula is used:

$$Score_i = PS - \frac{PTD}{\ln\left(\frac{1}{2}\right)} * \ln\left(ODDS\right) + \frac{PTD}{\ln\left(\frac{1}{2}\right)} * \ln\left(\widehat{ODDS}_\iota\right)$$

where:
PS – base number of points which corresponds to having ODDS value.
ODDS – value of odds, which is related to having PS score.
PTD – points to double, number of points that causes a double decrease in odds.

Table 5
Logistic Regression summary

Deviance Residuals				
Min	**1Q**	**Median**	**3Q**	**Max**
-2.3444	-0.2368	-0.1238	-0.0738	3.6317
Coefficients				
	Estimate	**Std. Error**	**Z-Value**	**P-Value**
(Intercept)	-1.5139	0.0087	-174.118	< 2E-16
avg_fico_woe	-1.0183	0.0044	-231.896	< 2E-17
dti_woe	-0.7582	0.0254	-29.804	< 2E-18
loan_amnt_woe	-1.2907	0.0374	-34.483	< 2E-19
mort_acc_woe	-0.3002	0.0464	-6.475	9.50E-11
grade_woe	-0.0539	0.0128	-4.220	2.44E-05
term_woe	-0.8915	0.0206	-43.208	< 2E-16
Null deviance: 264060 on 279994 degrees of freedom				
Residual deviance: 123187 on 279988 degrees of freedom				
AIC: 123201				

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

The final form of the transformation formula:

$$Score_i = 660 - \frac{40}{\ln\left(\frac{1}{2}\right)} * \ln\left(\frac{1}{72}\right) + \frac{40}{\ln\left(\frac{1}{2}\right)} * \ln\left(\widehat{ODDS}_i\right)$$

Table 6 summarizes results of model quality assessment. The p-value of L.R. test is 0, thus, the null hypothesis about joint insignificance of variables is rejected. P-value of Osuis-Rojek goodness-of-fit test does not allow to accept the null which states that the model is well fitted to data. Hosmer-Lemeshow show p-value equal to 0, which a well does not allow to accept the null about wellness of fit. However, p-value of Pearson's goodness-of-fit test is 1, thus the hypothesis that the model fits the data well is not rejected. ROC curves from model with intercept only and final model are compared by DeLong's test. P-value of the test is 0, thus, the null hypothesis stating that ROC curves from both models are equally good is rejected. Values of Kolmogorov – Smirnov test statistics from both test and train samples are quite high (> 0.77), indicating that distributions of scores for defaulted and non-defaulted clients in both test and train samples differ significantly, which is a good indicator. Population Stability Index (PSI) takes value lower than 0.1 (common rule of thumb), indicating that the model is stable. P-value of Komogorov-Smirnov stability test also does not allow to reject the null, which states that data from two periods (test and train) come from the same distribution, i.e., the model is stable. GINI values for test and trains samples are presented along with 95% confidence intervals. Indicators takes quite high values, 0.8881 and 0.8891 for train and test samples respectively, meanwhile 95% confidence intervals for these values are rather narrow.

Table 6
Logistic Regression quality assessment summary

LR	Osius-Rojek	Hosmer-Lemeshow	Pearson's Test	ROC Comparison
0	0	0	1	0
K-S Statistic Train	**K-S Statistic Test**	**Population Stability Index**		**K-S Stability**
0.7793	0.7779	0.0003		0.9171
GINI Train = 0.8881		**GINI Train 95% CI: [0.8860; 0.8902]**		
GINI Test = 0.8891		**GINI Test 95% CI: [0.8859; 0.8922]**		

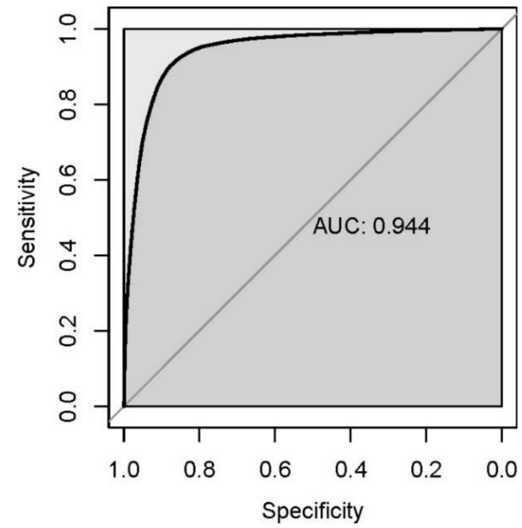
Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

Although two of three goodness-of-fit tests are rejected, one shall not rely on p-values only when operating with large samples, since p-values of test in such sample quickly go to zero. Moreover, goodness-of-fit tests are not assessing the predictive ability of the model, but rather check for deviations of functional S-shaped curve.

The area under the ROC curve (AUROC) presented in Figure 7 indicates quite a high distinguishing capability of binary classifier. The percentage of AUROC is around 94.4%. Histogram of assigned scores by loan outcome based on the train sample is pictured in Figure 8. Green and red-colored shares of histogram bins represent non-defaulted and defaulted cases, respectively. The distribution is left-skewed: the mean value of the score is shifted leftwards. This is explained by the prevalence of non-defaulted cases in the train sample, which tends to have higher scores.

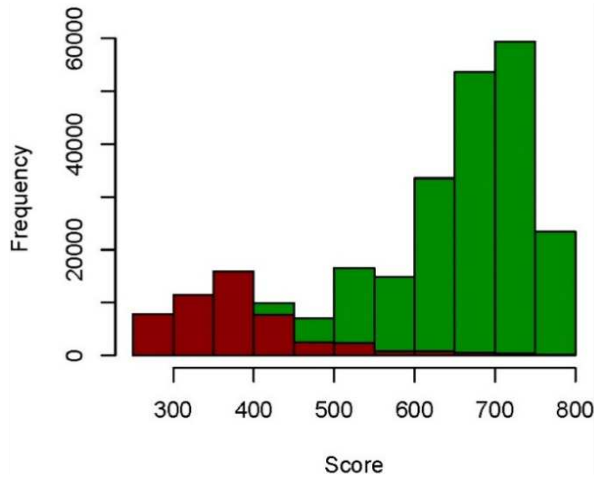
DOI: 10.7172/2353-6845.jbfe.2021.2.2

Figure 7
Receiver operating characteristic (ROC) curve in specificity and sensitivity space



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

Figure 8
Histogram of scores by loan outcome



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

The number of scores subject to each variable level is assigned by the following method:

$$points_{i,j} = \frac{\widehat{B}_J * \mathrm{WoE}_{i,j} * PTD}{\ln(0.5)} \; ; \; points_{intercept} = \frac{\ln\left(\frac{1}{e^{B_{intercept}}}\right) + \ln(ODDS) + \frac{\ln(2) * PS}{PTD}}{\ln(2)/PTD}$$

where:
$points_{i,j}$ – points subject to i-th level of j-th variable,
$\widehat{B}_J$ – an estimate of j-th feature.
$\mathrm{WoE}_{i,j}$ – WoE of i-th level of j-th variable.
$points_i$ – points subject to constant (initial score).
$B_{intercept}$ – value of intercept.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

The final scorecard is presented in Table 7. An amount of points that correspond to the specific level/interval of a variable is displayed in columns "Points". The base number of points is 500.57.

Table 7
Scorecard summary

Variable	Level	Points	Variable	Level	Points
Intercept	N/A	500.57	grade	E/F/G	-3.22
avg_fico	[0; 572]	-158.87	grade	D	-1.74
avg_fico	(572; 622]	-95.84	grade	C	-0.44
avg_fico	(622; 657]	-5.54	grade	B	1.43
avg_fico	(657; 677]	85.66	grade	A	4.16
avg_fico	(677; 702]	142.24	loan_amnt	(+ ∞; 25000)	-19.78
avg_fico	(702; 727]	185.98	loan_amnt	[25000; 15025)	-14.45
avg_fico	(727; + ∞)	221.01	loan_amnt	[15025; 10000)	-3.49
dti	[0; 9.33]	19.00	loan_amnt	[10000; 9000)	8.29
dti	(9.33; 12.14]	12.46	loan_amnt	[9000; 4750)	20.39
dti	(12.14; 14.84]	9.81	loan_amnt	(0; 4750]	29.59
dti	(14.84; 18.11]	3.32	mort_acc	0	-2.83
dti	(18.11; 22.28]	-1.48	mort_acc	1	-0.28
dti	(22.28; 25.01]	-7.54	mort_acc	2	1.42
dti	(25.01; 29.93]	-13.22	mort_acc	3	2.87
dti	(29.93; + ∞)	-22.24	mort_acc	(3; + ∞]	4.82
			term	60 months	-35.20
			term	36 months	14.55

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

The last step in scorecard development is finding an optimal cut-off score, which will be referred to when making an investment decision. There are several approaches. One of them is to maximize the portfolio performance based on the expected profit and expected loss from a good and bad client, respectively. Another approach is to set the target acceptance or default rate of the portfolio. However, the above practices are subject to expected profits and losses specific to good and bad loan outcomes. This paper, thus, focuses on the comparison of cutoff point calculations based on Diagnostic Accuracy Indices (DAI) that are constructed from True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) value (e.g., specificity, sensitivity). Where Negative outcome (0) stands for non-default and Positive (1) outcome is defaulted loan. Analyzed approaches are:

- minimization of the *Sum of misclassification costs* = *FN* + *FP*; i.e., the sum of False Bad (type I error) and False Good (type II error) clients
- minimization of the p-value (maximization of a statistic) of a chi-squared test on the confusion matrix, achieving maximum discrimination power
- *Youden index* = (*Sensitivity* + *Specificity* – 1) maximization
- cut off score subject to the point, such that the distance to (0,1) point on ROC in False Positive and True Positive space is minimized

DOI: 10.7172/2353-6845.jbfe.2021.2.2

– maximizing $F1\,Score = \dfrac{2 * TP}{2 * TP + FP + FN}$

– maximizing $Cohen's\ Kappa = \dfrac{Accuracy - P_e}{1 - P_e}$[18]

– maximizing $Matthews\ Correlation\ Coefficient\ (MCC) =$

$$= \frac{TP * TN - FP * FN}{Sqrt\left((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)\right)}.$$

For each method, values of cut-off points are calculated based on a train sample. Afterwards, each cut-off point is applied to the test sample and measures for classifier evaluation are calculated.

Table 8
Cut-off points metrics

Metric	Cut-off Point	Accuracy	Sensitivity	Specificity
Misspecification Cost	415.4483	0.9063	0.7463	0.9414
Cohen's Kappa	432.8462	0.9036	0.8028	0.9257
ROC (0,1)	445.0417	0.8999	0.8412	0.9128
MCC	447.3703	0.8991	0.8470	0.9106
Youden Index	499.6972	0.8819	0.9017	0.8775
F1 Score	500.6436	0.8814	0.9026	0.8767
P-value	586.0022	0.8039	0.9577	0.7701

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

Summary of cut-off points obtained from each approach are presented in Table 8. Methods are sorted by accuracy. ROC point and MCC approaches also are of similar accuracy; however, in this case, their Specificity and Sensitivity metrics are also comparable. They both offer higher Sensitivity, thus, accepting more loan applications, but at the cost of the greater share of False Negative rate. Cut-off points calculated based on the Youden Index and F1 Score metrics are of a virtually equal cut-off score. The P-value approach has the lowest accuracy. Misspecification cost minimization and Cohen's Kappa metric maximization are two methodologies that give the highest value of accuracy (i.e., the sum of correctly predicted outcomes as a share of the total number of applications). The difference in accuracy is negligible. There is, however, a noticeable tradeoff between sensitivity and specificity. The misspecification cost has higher specificity – an advantage in detecting True Negative outcomes; meanwhile, the share of correctly predicted Positive outcomes is higher in Cohen's Kappa approach.

[18] $P_e = \dfrac{(TP + FP) * (TP + FN) + (TN + FP) * (TN + FN)}{(TP + TN + FP + FN)^2}$

DOI: 10.7172/2353-6845.jbfe.2021.2.2

6. CONCLUSIONS

The aim of our research was to explore the phenomenon of peer-to-peer lending market model. In our paper a comprehensive view on the historical development of peer-to-peer lending in the financial environment, as well as the overview of the current situation on the alternative finance markets was presented.

Marketplace lending shows itself as one of the most promising and rapidly emerging forms of crowdfunding. It has developed enormously in recent years, providing more and more funding and investment opportunities for individuals and institutions. Among others, this form of crowdfunding is regarded as a potential competitor to traditional banking lending. The regulation of marketplace lending experienced a time lag; however, some countries with developed P2P lending industry have recently responded to the growing demand for adequate and industry-specific regulations with brand-new legal solutions.

The research hypothesis that the method of credit scoring is applicable in alternative lending environment is confirmed. The research has shown that scorecard derived from the logistic regression is a robust risk assessment instrument that can be used not only in the traditional financial environment but also in alternative lending, where both historical data and application-specific data are available.

Moreover, the research has shown that logistic regression approach to scorecard development provides high AUROC values, as well as sensitivity and specificity statistics that are comparable to more advanced machine learning models, provided that cut-off point is defined properly. Additionally, it was shown that quality of the final version of the logistic regression model and, thus, the scorecard, may be enhanced by more advanced variable pre-processing. In our case, variables binning based on Weight of Evidence (WoE) and Information Value (IV) indices allowed to pre-select the most meaningful explanatory features. The issue of choosing the appropriate cut-off point metrics was also addressed. Despite the fact that there might not be a huge absolute difference in accuracy, evidently, there is a clear trade off tendency between sensitivity and specificity for a given level of precision. Thus, investors should select the preferred cut-off point according to their risk acceptance level. Therefore, the latter two methods are the only ones that are similar in terms of accuracy, nonetheless with the apparent disparity in cut-off scores. One may try to apply an expected profit/loss method, and based on the specificity and sensitivity values, choose the cut-off point according to the highest expected profit.

The recent COVID-19 pandemic caused by SARS-CoV-2 virus has brought a noticeable disturbance to the financial market, particularly its lending division. The operational side of online platforms remained virtually unaffected, and employees continued their work remotely. Nevertheless, P2P lending platforms have faced a kind of "bank run". A particular group of investors who were alarmed by the previous crises want to retract their funds from platforms, regardless of the potential decrease in returns. Others actively use secondary markets to sell their investments with discounts. Some platforms, in turn, introduce withdrawal restrictions and increase the withdrawal processing time, since they are unable to service these outflows simultaneously. Although platforms are not directly affected by the increased number of defaults (investors bear this risk), they still finance their costs from the loan origination fees. The amount of originated loans has decreased, triggering platforms' liquidity issues.

On the contrary, many SMEs were in search of new funding solutions to resolve their liquidity issues. Thus, there may be a disparity of demand and supply of loanable funds on platforms. Government support aimed at SMEs may bring some relief to the market, as may deferred repayment solutions introduced by platforms for SMEs that are experiencing liquidity issues.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

References

BBC UK. (2005). Q&A: Online lending exchange. Retrieved on 28 November 2019 from news.bbc.co.uk/2/hi/business/4325761.stm.

Bellefamme, P., Lambert, T., & Schwienbacher, A. (2010). Crowdfunding: An industrial organization perspective. Prepared for the Workshop Digital Business Models: Understanding Strategies held in Paris on June 25–26.

Chen, D., Lai, F., & Lin, Z. (2014). A trust model for online peer-to-peer lending: A lender's perspective. Information Technology and Management, 239-254. https://doi.org/10.1007/s10799-014-0187-z

Crockett, Z. (2019). Kickstarter: A data analysis. The Hustle. Retrieved on 23 December 2019 from https://thehustle.co/archive/02102019d.

Financial Conduct Authority. (2016). Interim feedback to the call for input to the post-implementation review of the FCA's crowdfunding rules. Feedback Statement. London: Financial Conduct Authority.

Funding Circle. (2016). Small business, big impact: The changing face of business finance. Evidence from Funding Circle. London: Centre for Economics and Business Research. Retrieved on 7 February 2020 from www.fundingcircle.com/uk/statistics/.

Gerber, E., Hui, J., & Kuo, P. (2012). Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. Proceedings of the International Workshop on Design, Infuence, and Social Technologies: Techniques, Impacts and Ethics, II(11).

Gonzalez, L. (2018). Blockchain, herding and trust in peer-to-peer lending. Managerial Finance, 46(6), 815–831. https://doi.org/10.1108/MF-09-2018-0423

Havrylchyk, O., & Verdier, M. (2018). The financial intermediation role of the P2P lending platforms. Comparative Economic Studies, 60, 115–130. doi:10.1057/s41294-017-0045-1. https://doi.org/10.1057/s41294-017-0045-1

Havrylchyk, O., Mariotto, C., Rahim, T., & Verdier., M. (2019). The expansion of the peer-to-peer lending. SRNN. Retrieved on 12 December 2019 from https://ssrn.com/abstract=2841316.

Hu, M. R., Li, X., & Shi, Y. (2019). Adverse selection and credit certificates: Evidence from a P2P platform. ADBI Working Paper Series. Working Paper 942. Retrieved from https://www.adb.org/publications/adverse-selection-credit-certificates-evidence-p2p-platform. https://doi.org/10.2139/ssrn.3470048

Kritzinger, N., & Van Vuuren, G.W. (2018). An optimised credit scorecard to enhance cut-off score determination. South African Journal of Economic and Management Sciences, 21(1). https://doi.org/10.4102/sajems.v21i1.1571

LendingClub. (2018). Data Dictionaries. Retrieved on 28 March 2020 from www.help.lendingclub.com/hc/en-us/articles/216127307-Data-Dictionaries.

LendingClub. (2019). LendingClub Statistics. Retrieved on 7 February 2020 from www.lendingclub.com/info/demand-and-credit-profile.action.

LendingClub. (2020). LendingClub Loan Data. Database. Retrieved from www.kaggle.com/denychaen/lending-club-loans-rejects-data.

LendingCrowd. (2020). Marketplace Statistics. Retrieved on 7 February 2020 from app.lendingcrowd.com/statistics.

Lenz, R. (2016). Peer-to-peer lending: Opportunities and risks. European Journal of Risk Regulation, 7(4), 688–700. https://doi.org/10.1017/S1867299X00010126

Lu, C., & Zhang, L. (2018). Research on risk factors identification of P2P lending platforms. American Journal of Industrial and Business Management, 8(5), 1344–1357. https://doi.org/10.4236/ajibm.2018.85091

Lynn, T., Mooney, J.G., Rosati, P., & Cummins, M. (Eds.). (2018). Disrupting finance: FinTech and strategy in the 21st century. London: Palgrave Studies in Digital Business & Enabling Technologies. https://doi.org/10.1007/978-3-030-02330-0

MarketFinance. (2020). Investor Statistics. Retrieved on 7 February 2020 from marketfinance.com/investor-statistics.

Möllenkamp, N. (2017). Determinants of Loan Performance in P2P. Paper presented at the 9th IBA Bachelor Thesis Conference, Enschede: University of Twente.

Oxera Consulting LLP. (2016). The economics of peer-to-peer lending. Independent economic assessment. Oxford: Peer-to-Peer Finance Association.

P2PMarketData. (2019) Retrieved on 7 December 2019 from https://p2pmarketdata.com.

Prosper Funding LLC. (2019). Press Releases. Retrieved on 7 December 2019 from https://www.prosper.com.

RateSetter. (2020). RateSetter statistics. Retrieved on 7 February 2020 from https://www.ratesetter.com/invest/statistics.

Renton, P. (2019). Prosper.com ending their auction process. LendIt Fintech News. Retrieved on 27 December 2019 from https://www.lendacademy.com/prosper-com-ending-their-auction-process-dec-19th.

Sauermann, H., Franzoni, C., & Shafi, K. (2019). Crowdfunding scientific research: Descriptive insights and correlates of funding success. PLoS ONE, 14(1). https://doi.org/10.1371/%20journal.pone.0208384. https://doi.org/10.1371/journal.pone.0208384

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. PLoS ONE, 10(10). https://doi.org/10.1371/journal.pone.0139427

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Siddiqi, N. (2017). Intelligent credit scoring: building and implementing better credit risk scorecards. Hoboken, New Jersey: John Wiley & Sons. https://doi.org/10.1002/9781119282396

Vallée, B., & Zeng, Y. (2019). Marketplace lending: A new banking paradigm?. The Review of Financial Studies, 32(5), 1939-1982. https://doi.org/10.1093/rfs/hhy100

Wang, G., Chen, E., & Zhang, H. (2017). P2P lending survey: Platforms, recent advances and prospects. ACM Transactions on Intelligent Systems and Technology, 8(6), 1-28. https://doi.org/10.1145/3078848

Zhang, B., Ziegler, T., Mammadova, L., Johanson, D., Gray, M., & Yerolemou, N. (2018). The 5th UK alternative finance industry report. Cambridge: The Cambridge Centre for Alternative Finance (CCAF). https://doi.org/10.2139/ssrn.3084570

Ziegler, T., Johanson, D., King, M., Zhang, B., Mammadova, L., Ferri, F., … Yerolemou, N. (2018). Reaching new heights: The 3rd Americas alternative finance industry report. Cambridge: The Cambridge Centre for Alternative Finance (CCAF). https://doi.org/10.2139/ssrn.3106911

Ziegler, T., Johanson, D., Zhang. B, Shenglin, B., Wang, W., Mammadova, L., … Hao, X. (2018). The 3rd Asia Pacific region alternative finance industry report. Cambridge: The Cambridge Centre for Alternative Finance (CCAF).

Ziegler, T., Shneor, R., Wenzlaff, K., Odorović, A., Johanson, D., Hao, R., & Ryll, L. (2019). Shifting paradigms. The 4-th European alternative finance benchmarking report. Cambridge: The Cambridge Centre for Alternative Finance (CCAF).

Zopa Bank Limited. (2019). Zopa.com. Retrieved on 7 December 2019 from https://www.zopa.com.

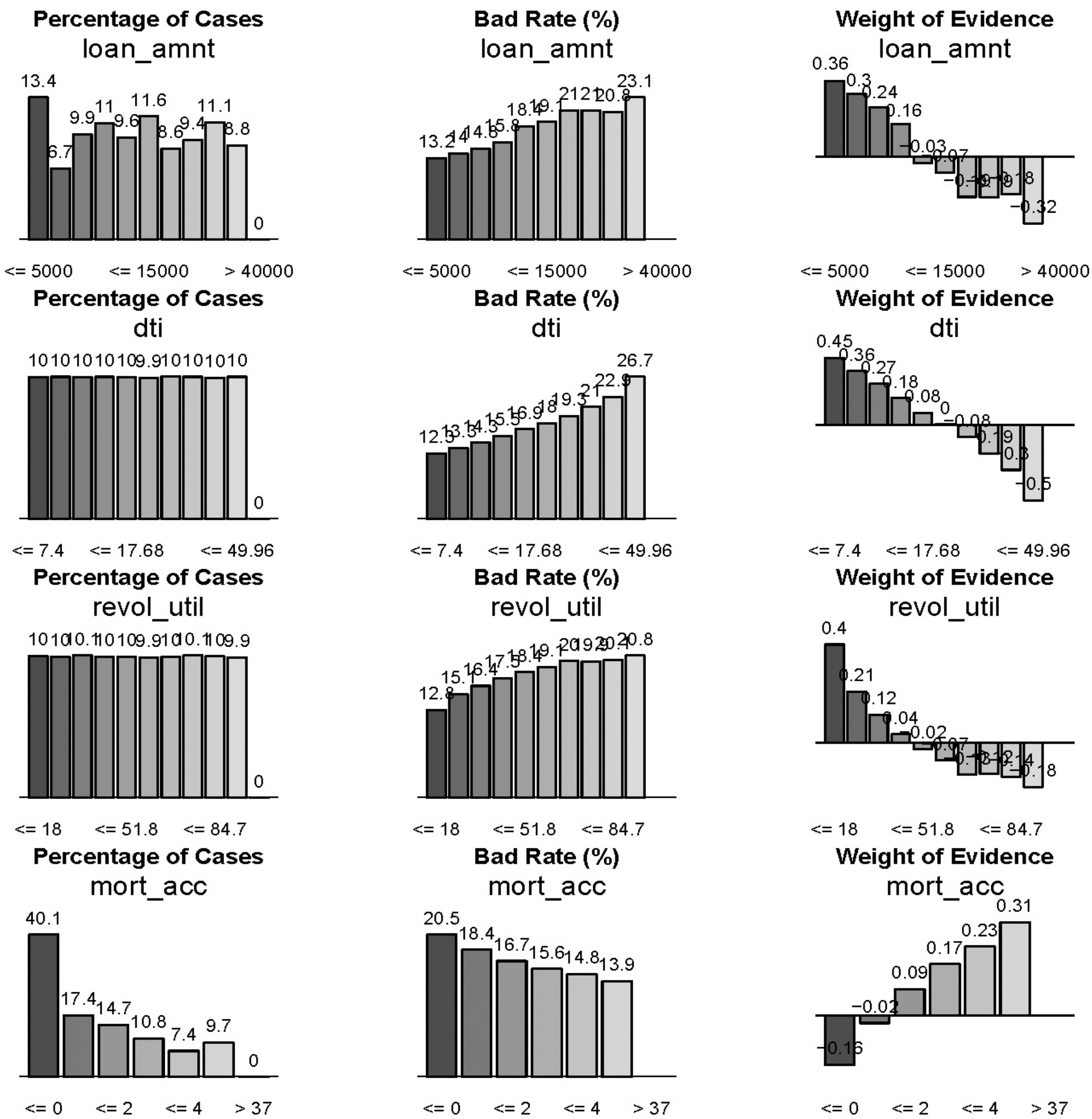
DOI: 10.7172/2353-6845.jbfe.2021.2.2

APPENDIX

Appendix 1. Summary of Fine Classing and Kendall's Tau analyses

Figure 9 contains sets of 3 graphs for each variable that were picked out as a result of variable quality assessment. The percentage of cases indicates the proportion of observations that falls into the specific bin. Bad Rate illustrates the percentage of defaults (G/B flag = 1) for a particular bin. Weight of Evidence displays calculated WoE for each bin.

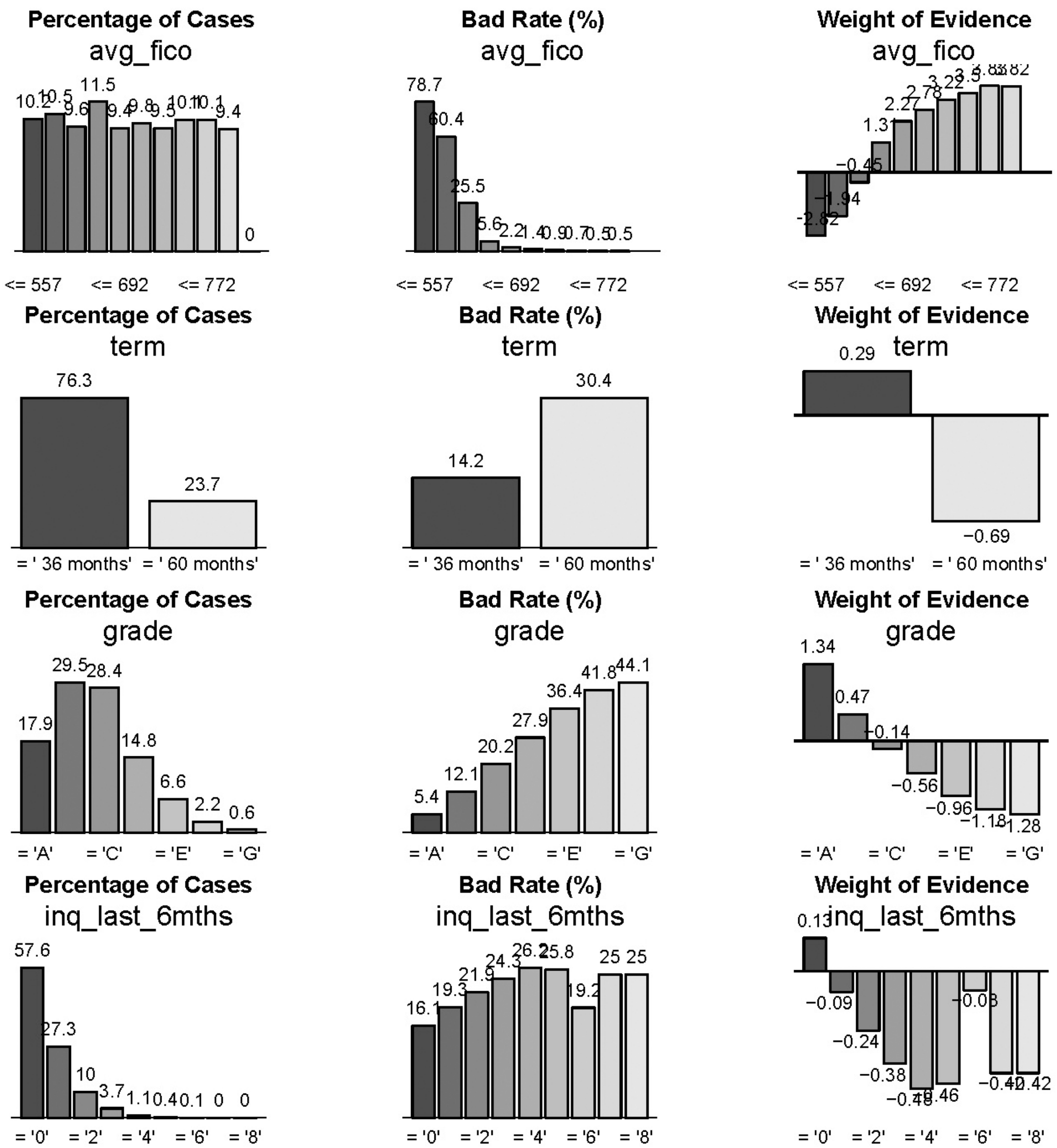
Figure 9
Summary graphs of fine classing, part I



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Figure 10
Summary graphs of fine classing, part II



Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data.

DOI: 10.7172/2353-6845.jbfe.2021.2.2

Table 9
Kendall's Tau rank correlation coefficients

	loan_amnt	int_rate	installment	dti	revol_util
loan_amnt	1				
int_rate	0.0809	1			
installment	**0.7788**	0.0925	1		
dti	0.0277	0.1317	0.0313	1	
revol_util	0.0911	0.1908	0.1037	0.1256	1
mort_acc	-0.1623	0.0732	-0.1372	0.0246	-0.0228
avg_fico	-0.0332	0.2649	-0.0217	0.0738	0.1306
term_	0.3439	0.3376	0.1955	0.0592	0.0558
grade	0.0855	**0.8836**	0.0934	0.1438	0.1928
home_ownership	-0.1343	0.0611	-0.112	-0.0041	-0.0189

	mort_acc	avg_fico	term	grade	home_ownership
mort_acc	1				
avg_fico	0.0822	1			
term_	-0.1031	0.0694	1		
grade	0.0761	0.2817	0.3610	1	
home_ownership	**0.5287**	0.0728	-0.0967	0.0649	1

Source: Own study based on: LendingClub Loan Data. San Francisco, February 2020. Database. www.kaggle.com/denychaen/lending-club-loans-rejects-data

DOI: 10.7172/2353-6845.jbfe.2021.2.2

The Impact of Equity Capital on the Bank's Profitability: Evidence From Vietnam's Banking System

Ngo, Thanh Xuan
School of Banking and Finance, National Economics University, Hanoi, Vietnam
xuantn@neu.edu.vn
https://orcid.org/0000-0002-1464-3942

Bui, Phương Linh
School of Banking and Finance, National Economics University, Hanoi, Vietnam
linh.bpl1401@gmail.com

Le, Minh Huong
School of Banking and Finance, National Economics University, Hanoi, Vietnam
huonglm08.ssc@gmail.com

Received: 10 November 2021 / Revised: 22 November 2021 / Accepted: 25 November 2021 / Published online: 20 December 2021

ABSTRACT

From 2008 to 2019, this research examines the effect of equity capital on the profitability of 24 Vietnamese commercial banks. The research findings indicate that, when ROAA and ROAE are used to measure the bank's profit, the equity capital ratio (CAP) has a statistically significant positive effect on the ROAA while having a negative effect on the ROAE. Between 2013 and 2019, the CAP variable has a positive effect on the ROAA and ROAE, indicating that banks with a larger equity capital ratio achieved higher profitability. Furthermore, the deposits-to-assets ratio (DTA) and loan-loss reserves ratio (LLR) both have a negative effect on both proxies for bank profitability, although bank size (SIZE) has a negligible effect on bank profits in the majority of circumstances. Additionally, the rate of GDP growth and inflation (INF) have a beneficial effect on the bank's profitability. The study's objective is to present some critical policy implications for bank executives about the importance of adequate equity capital for the bank's sustainability development.

JEL Classification: G20; G21

Keywords: bank equity capital, bank profitability, commercial banks.

1. INTRODUCTION

The COVID-19 epidemic triggered the world economy's worst crisis since the 2007–2009 global financial crisis, impairing the functioning of financial sectors, especially banks. It resulted in severe tightening of lending policy and a decrease in borrower creditworthiness, placing undue strain on the bank's buffer against risk – equity capital and profitability. In Vietnam, a wave of equity raising has lately swept across numerous commercial banks as a result of the Basel accords'

DOI: 10.7172/2353-6845.jbfe.2021.2.3

proposal for a stricter bank capital regime. Since January 1, 2020, when Circular 41/2016/TT-NHNN on the capital adequacy ratios (CAR) regulation was implemented, all Vietnamese banks and foreign bank branches have been required to maintain a minimum CAR of 8%. If they do not find ways to increase their CAR, businesses face a significant risk of being restricted in their credit expansion. According to bank financial documents, this research calculated an exceptional 18.61 per cent growth rate in equity capital from 2014 to 2019, compared to 3.12 per cent over the 2007–2014 period (Nguyen & Le, 2016). Due to the current rising trend in bank equity in Vietnam following the COVID crisis, our team conducted this research on the impact of equity capital on commercial bank profitability, utilizing 264 observations from 24 Vietnamese commercial banks from 2008 to 2019. Does a higher level of bank capital have an effect on the profitability of commercial banks in Vietnam during the 2008–2019 period and the post-crisis period (2013–2019)? Is the relationship comparable in terms of the State ownership structure: banks with State capital greater than 50% vs banks with State capital equal to or less than 50%? Which of the following variables has a greater impact on the profitability of commercial banks in Vietnam's economy during and after the crisis? Finally, what policy implications do these two types of banks have for the Vietnamese banking industry in the aftermath of the COVID crisis?

This paper also enriches the literature on banking and finance in this topic. Most existing literature on the relationship between the bank's equity capital and profit have been conducted in developed countries (Pettway, 1976; Berger, 1995; Goddard et al., 2004; Iannotta et al., 2007). Disagreements among various countries require further research to reach a suitable consensus on this issue. Many researchers approve of the positive correlation between the bank's equity capital and profits (Jacques & Nigro, 1997; Rime, 2001; Iannotta et al., 2007; Bitar et al., 2018; Bourke, 1989; Pasiouras & Kosmidou, 2007; Tan, 2016). In contrast, some studies explored a negative correlation between the bank's equity capital and profitability (Cavallo & Rossi, 2002; Goddard et al., 2010; Hermes & Vu, 2010; Nguyen, 2018; Dang, 2019). Although researchers are attempting to answer a similar research question in Vietnam, most Vietnamese papers ignore the heteroskedasticity and autocorrelation tests despite knowing that they are critical to confirm whether estimation results are reliable or not (Phan, 2016) or experiment on inadequate observants or in a short period of time (Do & Vu, 2019). As a result, this paper will fulfil the research gaps by providing more in-depth study based on two research time frames (12-year period comprising the crisis and 5 years later after crisis) and two types of research objectives (banks with State-owned capital greater than 50% and banks with State-owned equal to or less than 50%) and strengthen research technics to achieve more reliable results.

The rest of this paper is sequenced as follows: the second section contains relevant literature review about the bank's profitability and equity capital and their relationship. The third section contains the methodology, data collection sources, variables measurement and the mathematical model of the study. The fourth part contains discussions and an analysis of the results. Finally, suggestions and recommendations for further research are presented in the conclusion.

2. LITERATURE REVIEW

2.1. Bank's Profitability

Bank profits are recognized by how the bank uses its resources to generate income, reflecting its overall revenue and expenses, thus becoming an important financial indicator determining its effectiveness. The ratio of profit before (or after) tax/total assets (ROA) and the ratio of profit before (or after) tax/total equity (ROE) are the two profit indicators that managers, investors often use to assess the profitability and performance of banks (Berger, 1995; Naceur & Omran, 2011; Lee & Hsieh, 2013; Dang, 2019; Mishkin, 2013).

DOI: 10.7172/2353-6845.jbfe.2021.2.3

2.2. Bank Equity Capital

According to Mishkin (2013), equity capital is the bank's net worth, which is raised through fresh stock sales or retained earnings. Even though banks have the highest debt-to-equity ratios and typically have less than 10% equity in their capital structure, equity plays a key role in the bank (Rose & Hudgins, 2008). A new bank is required by law to raise a particular amount of legal equity capital in order to form, organize, and commence business. On the other hand, a bank's equity capital protects it against a decline in the value of its assets, which could force the bank into bankruptcy (Mishkin, 2013). This function of equity capital is to ensure that the bank is capable of mitigating risk. A high equity capital ratio fosters public trust and reassures creditors and borrowers that the bank will always be financially sound enough to meet their credit demands regardless of the economy's state (Rose & Hudgins, 2008). Additionally, capital adequacy has become a mandatory criterion for central bank oversight and regulation. The central bank strictly monitors bank activities based on the capital adequacy ratio in order to maintain the safety of banking operations and the financial system in general.

2.3. The Relationship Between Bank's Equity Capital and Profitability

There is a mix in results when researching the impact of the bank's equity capital on its profitability. There have been some research stating that the bank's equity capital positively relates to profitability (Berger, 1995; Jacque & Nigro, 1997; Demirgüç-Kunt & Huizinga, 1999; Rime, 2001; Goddard et al., 2004; Iannotta et al., 2007; Lee & Hsieh, 2013; Bitar et al., 2018). Berger (1995) used almost 80,000 observations to examine the link between a bank's equity capital and earnings for US commercial banks from 1983 to 1989. Granger causality tests revealed that a rise in equity capital results in an increase in profits and vice versa. Demirgüç-Kunt & Huizinga (1999) used bank-level data from 80 countries between 1988 and 1995 and found a positive correlation between bank equity capital, net interest margin (NIM), and profits before taxes (EBT) to total assets. Similarly, Goddard et al. (2004) discovered a strong and favourable association between the capital-to-assets ratio and return on equity (ROE) in six key European banking sectors throughout the 1990s. Iannotta et al. (2007) established a favourable correlation between the book value of equity to total assets and the operational profit to total assets ratio in a large number of banks. Private banks, in particular, are more profitable on average than mutual and public banks.

Lee & Hsieh (2013) recently adopted four profitability proxies: return on assets (ROA), return on equity (ROE), net interest margin (NIM), and net interest revenue as a percentage of average assets (N.R.). The authors acknowledged the ambiguity of their findings. Investment banks have the smallest positive capital effect on NIM and N.R.; banks in middle-to-high-income nations have the largest positive capital effect on ROE but the smallest on N.R. As a result, in lower-income countries, the equity capital of the bank has a greater impact on profitability. Bank capital and profit (excluding ROE) are positively associated across all samples. Similarly, Bitar et al. (2018) conducted an empirical study from 1999 to 2013 on 1,992 banks in 39 OECD countries and discovered that increased equity capital ratios significantly improve banking institutions' efficiency and profitability. Specifically, the author claimed that equity capital has a greater impact on larger and "too big to fail" banks, whereas high liquidity institutions utilize equity capital less effectively. During times of stress, highly capitalized banks have larger loan loss reserves, bigger net interest margins, and lower costs. This result is in line with Coccorese & Girardone's (2017) research, in which 4,414 banks from 77 countries over 2000–2013 were observed. This study found that the capital-profitability relationship is significantly stronger in crisis periods, in lower- and middle-income countries with higher corruption levels and larger banks. Several empirical studies further report a positive relationship between the bank's equity capital and profitability

DOI: 10.7172/2353-6845.jbfe.2021.2.3

(Munyambonera, 2013; Pervan et al., 2015; Ozili, 2017; Islam & Nishiyama, 2016; Abbas et al., 2019; Arshad, 2019).

Nguyen & Le (2016) are among researchers who support positive results when analyzing bank capital's effect on 30 Vietnamese commercial banks' risk and profit from 2007 to 2014. Nevertheless, the study contains limitations since its data do not include joint-venture banks and foreign bank branches in Vietnam; hence, the generalization is not high. Supporters of this result are Do & Vu (2019), whose research makes a difference using NIM besides ROA as proxies for bank profit. In addition, some notable independent variables are "growth deposit", "funding cost", "ownership" and "lend". The paper also reached different conclusions based on different bank sizes and types of ownership. Accordingly, the effect of capital on profit is larger for small banks than for large ones. Huynh (2019) obtained the same result but brought remarkable points in his research when measuring profitability by ROAA (return on average asset). Besides, independent variables such as net interest margin (NIM), cost-to-income ratio (CIR), loan loss provision (LLP), non-performing loan (NPL), and Herfindahl-Hirschman Index (HHI) are also incorporated in the model.

On the other hand, several investigations discovered the opposite. Pettway (1976) examined the negative association between bank equity capital and profitability in the United States of America for banks and bank holding corporations between 1971 and 1974. The author discovered that by combining the beta and P/E models, the equity capital requirement reduces operational efficiency, predicting a drop in bank profitability. Additionally, Altunbas et al. (2007) showed that inefficient European banks appear to have a higher level of equity capital and lower risk tolerance. According to Modigliani and Miller's (1963) "risk-reward trade-off concept," low risk levels result in low potential profits. Indeed, according to Fotios et al. (2009), capitalization has a statistically significant detrimental effect on both cost and profit efficiency. Additionally, Goddard et al. (2010) discovered that the average profitability of efficient and diverse banks is higher than that of heavily capitalized institutions. Between 1992 and 2007, a negative relationship between equity capital and profitability, implying an opportunity cost associated with high capital levels, tended to decrease European banks' shareholder returns.

Dang (2019) claimed that the higher the equity ratio banks have, the fewer risks banks take; hence, the profit would lower. Interestingly, the study found a nonlinear relationship that explains that credit risk lessens the impact of equity on returns. However, one disadvantage of the study is that the paper only applies traditional accounting methods and does not approach a more complete data set.

Mixed results can also be found in recent studies. Tran et al. (2016), who took British banks data from 1996–2013 into the VAR and the generalized method of moments model (GMM), pointed out a negative correlation with large-capitalized banks, yet a positive correlation with small-capitalized banks. Specifically, the researchers used three ways to measure bank capital: (i) the ratio of tier 1 capital to total risk-weighted assets (RWA); (ii) the ratio of total equity to total assets; (iii) tangible ordinary equity ratio to RWA, denoted as CARA, CARB, and CARC respectively. Besides, Hasnaoui & Fatnassi (2019) also applied the GMM method with the secondary data of 85 banks in the Gulf Cooperation Council (GCC) countries in the period 2003–2011 and described the following: (i) Islamic banks with high capitalization produce low profits, while conventional banks with high capitalization produce high profits; (ii) GCC banks (including Islamic and conventional banks) have greater risk compared to others; (iii) all the risk and return variables are statistically significant. Saona (2016) and Le & Nguyen (2020) concluded an inverse U-shaped relationship between the bank's capital ratios and profitability. In contrast, Barth et al. (2008) concluded that the equity capital and performance do not have a linear relationship.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

3. METHODOLOGY AND DATA

3.1. Research Hypotheses

Based on literature reviews, this paper will measure profitability by return on average equity (ROAE) and return on average assets (ROAA) for the following reasons: (i) the indicator depicts the evolution throughout time, not a single point in time; (ii) if the asset or equity value fluctuates significantly over time, the simple ROA and ROE ratios will not be as accurate as the average ratios (Abbas et al., 2019). These proxies were confirmed by Dietrich and Wanzenried (2011), Batten and Vo (2016), Chiaramonte and Casu (2017), Abbas et al. (2019).

In addition to the primary explanatory variable (capital-to-assets ratio), other variables used in this study have been verified by much prior research. Tan and Floros (2013), Lee and Hsieh (2013), and Hasnaoui and Fatnassi (2019) adopted the loan-loss-reserve ratio, loans-to-assets ratio, GDP growth rate, and inflation rate to estimate their impact on bank profitability. Besides, the bank's primary source of funds is derived from deposits. Hence the deposits-to-assets (DTA) ratio plays a vital role in the regression model. This variable is supported by Lee and Hsieh (2013), Ramlan and Adnan (2016), Ali et al. (2017), Dang (2019), Arshad and Iskandar (2020). On the other hand, we use the bank size variable to consider whether big banks or small banks generate more profit over time. Many researchers expressed concern about this issue, such as Berger and Bouwman (2013), Cohen and Scatigna (2016), Abbas et al. (2019), Kanga et al. (2020), Arshad and Iskandar (2020). Finally, Iannotta et al. (2007), Lee and Hsieh (2013), and Do & Vu (2019) considered State ownership factors affecting the bank's profitability.

Based on the reviewed literature, the study regresses the following variables to measure their impact on bank profitability and proposes 8 hypotheses as follows (Table 1):

Table 1
Description of the variables and expected correlation coefficient

	Indicator	Measured by	Notation	Related studies	Hypothesis
Dependent variable					
Profitability	Return on average equity	$\frac{\textit{Net income}}{\textit{Average total equity}}$	ROAE	Dietrich & Wanzenried (2011), Batten & Vo (2016), Abbas et al. (2019)	
	Return on average assets	$\frac{\textit{Net income}}{\textit{Average total assets}}$	ROAA	Dietrich & Wanzenried (2011), Batten & Vo (2016), Chiaramonte & Casu (2017), Abbas et al. (2019), Huynh (2019)	

DOI: 10.7172/2353-6845.jbfe.2021.2.3

Table 1 (cont.)

	Indicator	Measured by	Notation	Related studies	Hypothesis
Independent variable					
Internal control variables	Equity capital-to-total assets ratio	$\frac{\textit{Equity}}{\textit{Total assets}}$	CAP	Altunbas et al. (2007), Goddard et al. (2010), Dietrich & Wanzenried (2011); Lee & Hsieh (2013), Tan & Floros (2013), Nguyen & Le (2016), Dang (2019), Kanga et al. (2020)	+
	Loans-to-assets ratio	$\frac{\textit{Total loans}}{\textit{Total assets}}$	LTA	Iannotta et al. (2007), Tan & Floros (2013), Lee & Hsieh (2013), Nguyen & Le (2016), Coccorese & Girardone (2017), Hasnaoui & Fatnassi (2019), Le & Nguyen (2020), Kanga et al. (2020)	+
	Loan-loss-reserves ratio	$\frac{\textit{Loav loss reserves}}{\textit{Total loans}}$	LLR	Dietrich & Wanzenried (2011), Tan & Floros (2013), Ozili (2015), Ranajee (2018), Dang (2019), Abbas et al. (2019), Hasnaoui & Fatnassi (2019), Kanga et al. (2020)	-
	Deposits-to-assets ratio	$\frac{\textit{Total deposits}}{\textit{Total assets}}$	DTA	Acharya & Naqvi (2012), Lee & Hsieh (2013), Ramlan and Adnan (2016), Ali et al. (2017), Dang (2019), Arshad & Iskandar (2020)	-
	Bank size	Natural logarithm of total assets	SIZE	Altunbas et al. (2007); Lee & Hsieh (2013), Berger & Bouwman (2013), Cohen & Scatigna (2016), Abbas et al. (2019), Kanga et al. (2020), Arshad & Iskandar (2020)	+
	State ownership	= 1 if the States owns > 50% shares = 0 if the States owns ≤ 50% shares	OWN	Iannotta et al. (2007), Dietrich & Wanzenried (2011), Vu & Nahm (2013), Ongore & Kusa (2013), Lee & Hsieh (2013), Do & Vu (2019)	-
Macroeconomic control variables	GDP growth rate	World Bank data	GDP	Tan & Floros (2013), Lee & Hsieh (2013), Dietrich & Wanzenried (2014), Coccorese & Girardone (2017), Hasnaoui & Fatnassi (2019)	+
	Inflation rate	World Bank data	INF	Tan & Floros (2013), Lee & Hsieh (2013), Tan & Floros (2013), Dang (2019), Hasnaoui & Fatnassi (2019)	+

Notes:
(+) Independent variable has positive effect on profitability
(–) Independent variable has negative effect on profitability

Source: Authors' compilation, 2020.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

3.2. Model, Data and Analytical Methods

The data were compiled from the audited financial statements of 24 Vietnamese commercial banks over a 12-year period, from 2008 to 2019. The shortlisted banks must demonstrate that they are viable businesses with adequate financial disclosures during this time period. We omit banks that have been merged or acquired by other banks, joint venture banks, foreign bank branches, and banks that have ceased to exist. Additionally, macroeconomic data are derived from the World Bank's annual report. After gathering and compiling data indicators in Microsoft Excel, the authors run the models using the Stata 14 software.

First, the authors used the following two tests to determine which method is the most suitable for the research model.

Breusch & Pagan Lagrangian multiplier test

To determine whether the Ordinary Least Square (OLS) or Random Effects Model (REM) is more suitable, we use the Breusch & Pagan Lagrangian multiplier test.

H0: The OLS model is suitable and efficient

H1: The REM model is suitable and efficient

Hausman test

To select a more suitable approach between Fixed Effects Model (FEM) and Random Effects Model (REM), we use the Hausman test.

H0: REM is consistent and efficient

H1: REM is inconsistent

After choosing a suitable regression method, the authors examined the model for the following defects: multicollinearity, heteroscedasticity and autocorrelation. To correct these two problems, models with the robustness option should be performed.

Variation Magnification Factor (VIF)

The authors used the defect model to test multicollinearity based on the Variance Magnification Coefficient (VIF) to check if the eight explanatory variables of the model have high collinearity phenomenon or not. When the VIF coefficient is greater than 5, the model has high collinearity, if the VIF is greater than 10, the research model will definitely have multicollinear defects.

LM – Breusch & Pagan Lagrangian multiplier test (for REM model) or Wald (for FEM model)

H0: Model has homoscedasticity

H1: Model has heteroscedasticity

Wooldridge test

H0: There is no autocorrelation

H1: Model has autocorrelation

Table 2
Tests for selecting the most appropriate model and tests for defects

	Tests for selecting the most appropriate model		Tests for detecting problems	
	Breusch & Pagan Lagrangian multiplier test	**Hausman test**	**Heteroskedasticity test**	**Wooldridge test for autocorrelation**
H0	OLS is consistent and effective	REM is consistent and effective	Homoscedasticity	No first-order autocorrelation
Ha	REM is consistent and effective	FEM is consistent and effective	Heteroskedasticity problem	Autocorrelation problem

Source: Authors' compilation, 2020.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

The proposed research model is as follows:

$$ROAE_{it} = \beta 0 + \beta 1\ CAP_{it} + \beta 2\ LTA_{it} + \beta 3\ LLR_{it} + \beta 4\ DTA_{it} + \beta 5\ SIZE_{it} + \\ + \beta 6\ OWN_{it} + \beta 7\ GDP_{it} + \beta 8\ INF_{it} + \varepsilon_{it}\ \text{(Model 1)}$$

$$ROAA_{it} = \beta 0 + \beta 1\ CAP_{it} + \beta 2\ LTA_{it} + \beta 3\ LLR_{it} + \beta 4\ DTA_{it} + \beta 5\ SIZE_{it} + \\ + \beta 6\ OWN_{it} + \beta 7\ GDP_{it} + \beta 8\ INF_{it} + \varepsilon_{it}\ \text{(Model 2)}$$

4. RESULTS AND DISCUSSIONS

Table 3 summarizes the factors used statistically. The average ROAE and ROAA for dependent variables are 0.106 and 0.009, respectively. The lowest ROAE and ROAA were 0.0004 and 0.00004, respectively, achieved by Viet Capital bank in 2016; the highest ROAE is 0.291, earned by SGB in 2010, while the lowest ROAA is 0.059, acquired by LPB in 2008. In terms of the bank's internal factors, the CAP variable averages 0.098 and varies somewhat widely (0.028–0.462). The bank's equity capital has a rather high standard deviation of 0.058. Additionally, the loans-to-assets ratio (LTA) is frequently high, averaging 0.574, indicating that Vietnamese commercial banks continue to rely substantially on credit. Additionally, the results indicate that LLR and DTA have mean values of 0.012 and 0.759, respectively. On the other hand, the mean bank size (SIZE) is 11.516, with the largest and smallest banks measuring 14.188 and 7.790, respectively, as reported by BID in 2019 and TPB in 2008. Additionally, the State ownership variable (OWN) only includes four banks that have State equity greater than 50% in their capital structure: AGRI, VCB, CTG, and BID. In terms of macroeconomic control factors, the sample averages 0.061 GDP growth and 0.076 inflation.

Table 3
Summary statistics for variables

Variables	Mean	Std. Dev.	Min	Max
ROAE	0.1068556	0.0734671	0.0004372	0.2911836
ROAA	0.0098269	0.0079141	0.0000459	0.0595733
CAP	0.0983710	0.0582819	0.0289337	0.4624983
LTA	0.5740402	0.1356559	0.1139038	0.8604010
LLR	0.0127397	0.0062340	0.0005517	0.0646743
DTA	0.7590368	0.0846606	0.5098618	0.9138934
SIZE	11.5168300	1.3175450	7.7909620	14.1881800
OWN	0.1666667	0.3733267	0	1
GDP	0.0618104	0.0061956	0.052500	0.070800
INF	0.0767642	0.0644769	0.008800	0.231200

Source: Authors' calculations using Stata 14, 2020.

Table 4 demonstrates that the correlation among variables is acceptable because the correlation coefficient between the two independent variables is less than 0.8 (Kennedy, 2008). We perform a multicollinearity test based on the Variance Inflation Factor (VIF) (see Appendix 1) to reinforce this conclusion. According to the results, all the models' explanatory variables have VIF coefficients of less than five and an average VIF of 2.04. Therefore, we assert that there is no high multicollinearity between independent variables.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

Table 4
Correlation matrix

	ROAE	ROAA	CAP	LTA	LLP	DTA	SIZE	OWN	GDP	INF
ROAE	1.000									
ROAA	0.6895	1.000								
CAP	-0.1892	0.4359	1.000							
LTA	-0.0015	-0.1745	-0.1774	1.000						
LLR	0.0388	-0.0952	-0.1913	0.0021	1.000					
DTA	-0.1772	-0.3675	-0.3959	0.2775	-0.0508	1.000				
SIZE	0.3127	-0.1940	-0.7156	0.3566	0.2485	0.1579	1.000			
OWN	0.2223	-0.0831	-0.3057	0.4317	0.3351	-0.0365	0.6184	1.000		
GDP	0.0842	-0.0912	-0.2525	0.1937	-0.1152	0.2426	0.3339	-0.0002	1.000	
INF	0.1574	0.2937	0.3188	-0.3302	0.0880	-0.2867	-0.3577	-0.0002	-0.4283	1.000

Source: Authors' calculations using Stata 14, 2020.

After regressing Breusch & Pagan Lagrangian multiplier test and Hausman test (see Appendix 2), we conclude that FEM is best suited for model 1, while REM shows reliable results for model 2 2008–2019. Both models have heteroskedasticity and autocorrelation problems (see Appendix 3). Hence, we regress models with the robustness option to fix these defects and draw some conclusions as the following states (Table 5):

Table 5
Full sample: estimation results in period 2008–2019

Variables	ROAE		ROAA	
	FEM robust		REM robust	
	Coef	Robust Std. Dev	Coef	Robust Std. Dev
CAP	-0.3967729***	0.1262998	0.0513069**	0.0224545
LTA	0.0162335	0.0416814	-0.0045832	0.0052219
LLR	-2.5069490***	0.8052045	-0.1604198**	0.0814649
DTA	-0.2796916***	0.0489450	-0.0221347***	0.0058308
SIZE	0.0026836	0.0116522	0.001129	0.0008066
OWN	0	(omitted)	-0.0003426	0.0028326
GDP	1.8415200**	0.6841281	0.1048647*	0.0557855
INF	0.3162656***	0.0761189	0.0236709***	0.0078575
R-squared	0.1636		0.2874	

Note: ***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.

Source: Authors' calculations using Stata 14, 2020.

To begin, when a bank's profitability is measured using the dependent variables ROAE and ROAA, inconsistent results about the relationship between equity capital and profitability are discovered. The equity capital ratio has a negative correlation with ROAE but a positive correlation with ROAA. Specifically, the CAP variable in model 1 has a coefficient of -0.396, implying that

DOI: 10.7172/2353-6845.jbfe.2021.2.3

a 1% increase in capital reduces ROAE by 39.67%. This conclusion is logical given that enterprises have a high capital ratio, which increases risk aversion (Berger, 1995), and a high capital ratio also diminishes the beneficial effect of the tax shield (Modigliani & Miller 1958; Berger, 1995; Goddard et al., 2010). These measures may result in a decrease in profit and ROAE. On the other hand, the coefficient of CAP is 0.051 at a 5% significance level in model 2, indicating that a 1% rise in equity capital ratio results in a 5.13 per cent increase in ROAA, all other variables remaining constant. Better capitalized banks may demonstrate increased creditworthiness, management quality, and competitiveness, allowing them to earn a high profit while maintaining a low cost (Iannotta et al., 2007). Additionally, lower predicted bankruptcy costs associated with a greater equity capital ratio may result in increased profitability and ROAA (Berger, 1995).

Second, the loan-loss-reserves ratio (LLR) has a statistically significant and negative effect on both profitability variables with estimated coefficients of -2.506 and -0.160. Most banks increase the provisions for credit losses due to the increased non-performing loans ratio, leading to an increase in risk provision expenses and debt recovery costs, which reduce profits.

Third, the deposit-to-assets ratio (DTA) is inversely connected to both ROAE and ROAA. The regression coefficients on DTA are -0.279 and -0.022 for model 1 and model 2, respectively, at a 1% significance level. Individual deposits account for the majority of commercial banks' deposit ratio. The increase in the deposit ratio will attract additional rivals in the supplement market, such as insurance, pension funds, and people's credit funds. Simultaneously with inadequate loan quality management and control, banks take risks by raising leverage at a high cost, resulting in a lower profit.

Fourth, regarding the macroeconomic conditions, the regression results show that the GDP growth rate (GDP) and inflation (INF) have a positive effect on the bank's ROAE and ROAA. These figures indicate that a significant increase in GDP with a moderate increase in inflation will enhance the profitability of the banking system (Iannotta et al., 2007; Lee & Hsieh, 2013; Dang, 2019; Hasnaoui & Fatnassi, 2019).

To further test the models' validity, the authors decided to evaluate the impact of the bank's equity capital on profitability for five years after recovering from the 2008 global financial crisis. Then, the results are as follows (Table 6):

Table 6
Full sample: estimation results in period 2013–2019

Variables	**ROAE**		**ROAA**	
	FEM robust		**REM robust**	
	Coef	**Robust Std. Dev**	**Coef**	**Robust Std. Dev**
CAP	0.7377549***	0.2521130	0.1346286***	0.0195851
LTA	0.1433555	0.0862052	0.0103041	0.0077162
LLR	-4.6452200**	1.9431560	-0.2204951	0.1518059
DTA	-0.1680508**	0.0813887	-0.0196563***	0.0066072
SIZE	0.0718427***	0.0248148	0.0054861***	0.0012364
OWN	0	(omitted)	-0.0078536**	0.0030812
GDP	2.2263480*	1.1945660	0.2181179***	0.0754550
INF	1.0401030***	0.2464699	0.0822023***	0.0160999
R-squared	0.3171		0.4052	

Note: ***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.

Source: Authors' calculations using Stata 14, 2020.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

To begin, between 2013 and 2019, CAP has a significant beneficial effect on ROAE and ROAA at the 1% level. As previously stated, highly capitalized banks generate considerable profits as a result of their high creditworthiness and limited reliance on external financing. In accordance with National Assembly Resolution 24/2016/QH14 dated November 8, 2016, the State Bank of Vietnam implemented Basel II regulations in the domestic banking sector, requiring a minimum capital adequacy ratio based on risky assets. That is, a larger capitalization ratio suggests that banks own more hazardous assets (Iannotta et al., 2007), which also implies a greater projected return.

Second, the link between LLR and ROAE is negative, consistent with the finding of full sample estimation, but in model 2, this association is inconsequential. Bank size is positively related to profit at a 1% significance level. Hughes et al. (2001) pointed out that as banks' scale gets more extensive, they will gain better advantages from potential diversification, leading to a positive relationship between the operational efficiency and size, thereby increasing the bank's profit.

Third, the OWN variable is negatively related to ROAA and barely affects ROAE. The OWN coefficient of -0.007 implies that a 1% increase in equity capital decreases ROAA by 0.7%. Similarly, Iannotta et al. (2007) argue that private banks appear to be more profitable than both mutual and public banks. The remaining variables, including LTA, DTA, SIZE, GDP, and INF, have similar results with the full sample regression over the period 2008–2019.

To examine the effect of equity capital on bank profits according to the State ownership structure, the entire sample is separated into two subsamples: banks with more than 50% of State ownership and banks with equal to or less than 50% of State ownership. The computed coefficients for both categories are consistent with the full sample regression results for the period 2008–2019, which indicates that equity capital has a negative effect on ROAE and a positive effect on ROAA. The DTA is inversely connected to the dependent variables. Additionally, LTA is statistically significant and has a negative influence on the profitability of > 50% of State-owned banks at a 1% significance level. However, this effect is negligible for banks with a 50 per cent State control. For a developing country like Vietnam, the government controls a sizable portion of the banking sector (Qian et al., 2015). Government engagement in banks owned by the State at a level greater than 50% is greater than in other banks. These banks place a premium on large-scale projects and wholesale products, resulting in low loan profitability (Dang & Huynh, 2019). On the other hand, LLR has a negative effect on bank profitability for banks with less than 50% of State control but is positively associated with bank profitability for banks with more than 50% of State ownership. This conclusion could be explained by the fact that banks held by the State at a level greater than 50% receive benefits from government guarantees, which help them minimize default risk (Brown & Dinç, 2011). As a result, an increase in loan loss reserves suggests an increase in high-risk loans, which results in increased profitability (risk-reward trade-off) (Kanga et al., 2020). Meanwhile, State-owned banks are under immense pressure to manage credit risk; as a result, they must bear increased credit risk management expenses if LLR increases. Finally, macroeconomic factors (GDP and INF) have a favourable effect on the profitability of banks with less than 50% of State control but have no effect on banks with more than 50% of State ownership. Bolt et al. (2012) once stated that the relationship between macro variables and profitability is ambiguous.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

Table 7
Different state-ownership levels: estimation results for period 2008–2019

Variables	> 50% State-ownership banks		≤ 50% State-ownership banks	
	REM robust	REM robust	REM robust	REM robust
	ROAE	ROAA	ROAE	ROAA
CAP	-0.855384* (0.4574056)	0.0701839*** (0.0250811)	-0.3394482** (0.1333827)	0.0495747** (0.0243919)
LTA	-0.4030582*** (0.0855576)	-0.0234738*** (0.0045473)	0.0257587 (0.0390479)	-0.0031481 (0.0058385)
LLR	0.8252344* (0.4342548)	0.0620873*** (0.0178626)	-2.647073*** (0.9616007)	-0.197238* (0.1025384)
DTA	-0.2473062*** (0.0168161)	-0.013401*** (0.0016899)	-0.2761599*** (0.0492013)	-0.0241525*** (0.0062252)
SIZE	0.0074127 (0.0351398)	-0.0000602 (0.0019466)	0.0122636 (0.0106802)	0.0012912 (0.0009043)
GDP	2.808265 (2.311432)	0.1714896 (0.1305927)	1.381056* (0.7147939)	0.1011338 (0.0637513)
INF	0.0629943 (0.2772859)	-0.000029 (0.0144965)	0.3715141*** (0.0812546)	0.026919*** (0.0094111)
R-squared	0.3552	0.5545	0.1972	0.2950

Note: ***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.

Source: Authors' calculations using Stata 14, 2020.

5. CONCLUSION

The primary objective of this study is to examine in depth the influence of equity capital on bank profitability in a rising economy such as Vietnam, using secondary data compiled from 24 Vietnamese commercial banks during a 12-year period from 2008 to 2019. Our findings indicate that when a bank's profitability is measured across the full research period, the equity capital ratio has a negative effect on ROAE and a favourable effect on ROAA. A detailed examination of the period from 2013 to 2019, five years following the financial crisis, reveals that the CAP variable has a positive effect on both ROAA and ROAE, indicating that banks with a higher capital-on-assets ratio achieved greater profitability. The inconsistent outcomes are partially a result of the 2008 financial crisis detrimental influence on the commercial banking industry. Specifically, interest rates climbed significantly between 2009 and 2011 as a result of the government's tight monetary policy (Nguyen et al., 2020). This constrains the credit area, which was the primary activity of commercial banks, and results in decreased bank efficiency. To provide a more detailed explanation, we will conduct a follow-up study on "the impact of monetary policy on bank profitability." Additionally, this has been a point of contention in recent years as a result of COVID-19.

Along with contributing to the understanding of the relationship between bank equity capital and profitability, our research has some policy implications for banking operations management. To begin, banks should seek short-, medium-, and long-term capital. Second, commercial banks must strengthen their capital management capabilities, as this enables them to prepare capital budgets more accurately and efficiently.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

References

Abbas, F., Iqbal, S., & Aziz, B. (2019). The impact of bank capital, bank liquidity and credit risk on profitability in post crisis period: A comparative study of US and Asia. Cogent Economics & Finance, 7, 1–18. https://doi.org/10.1080/23322039.2019.1605683

Ali, T. Y., Asif, A., & Mosab, I. T. (2017). The impact of political instability, macroeconomic and bank-specific factors on the profitability of Islamic banks: An empirical evidence. Investment Management and Financial Innovations, 14(4), 30–39. https://doi.org/10.21511/imfi.14(4).2017.04

Altunbas, Y., Carbo, S., Gardener, E.P., & Molyneux, P. (2007). Examining the relationships between capital, risk and efficiency in European banking. European Financial Management, 13(1), 49–70. https://doi.org/10.1111/j.1468-036X.2006.00285.x

Barth, J. R., Caprio, G., & Levine, R. (2008). Rethinking bank regulation: Till angels govern. Economica, 74(239), 177–179. https://doi.org/10.1111/j.1468-0335.2006.00561.x

Berger, A. N. (1995). The relationship between capital and earnings in banking. Journal of Money, Credit and Banking, 27(2), 432–456. https://doi.org/10.2307/2077877

Berger, A. N., & Bouwman, C. H. S. (2013). How does capital affect bank performance during financial crises?. Journal of Financial Economics, 109(1), 146–176. https://doi.org/10.1016/j.jfineco.2013.02.008

Bitar, M., Pukthuanthong, K., & Walker, T. (2018). The effect of capital ratios on the risk, efficiency and profitability of banks: Evidence from OECD countries. Journal of International Financial Markets, Institutions and Money, 53, 227–262. https://doi.org/10.1016/j.intfin.2017.12.002

Bolt, W., De Haan, L., Hoeberichts, M., Van Oordt, M. R., & Swank, J. (2012). Bank profitability during recessions. Journal of Banking & Finance, 36(9), 2552–2564. https://doi.org/10.1016/j.jbankfin.2012.05.011

Bourke, P. (1989). Concentration and other determinants of bank profitability in Europe, North America and Australia. Journal of Banking & Finance, 13(1), 65–79. https://doi.org/10.1016/0378-4266(89)90020-4

Brown, C., & Dinç, I. S. (2011). Too many to fail? Evidence of regulatory forbearance when the banking sector is weak. Review of Financial Studies, 24(4), 1378–1405. https://doi.org/10.1093/rfs/hhp039

Cavallo, L., & Rossi, S. (2002). Do environmental variables affect the performance of european banking systems? A parametric approach using the stochastic frontier approach. European Journal of Finance, 8(1), 123–146. https://doi.org/10.1080/13518470110076277

Chiaramonte, L., & Casu, B. (2017). Capital and liquidity ratios and financial distress. Evidence from the European banking industry. The British Accounting Review, 49(2), 138–161. https://doi.org/10.1016/j.bar.2016.04.001

Coccorese, P., & Girardone, C. (2017). Bank capital and profitability: Evidence from a global sample. Working Paper Series, 17(2), 1–45.

Cohen, B. H., & Scatigna, M. (2016). Banks and capital requirements: Channels of adjustment. Journal of Banking & Finance, 69(S1), S56–S69. https://doi.org/10.1016/j.jbankfin.2015.09.022

Dang, V. D. (2019). Should Vietnamese banks need more equity? Evidence on risk-return trade-off in dynamic model of banking. Journal of Risk and Financial Management, 12(84), 1–13. https://doi.org/10.3390/jrfm12020084

Dang, V. D. & Huynh, J. (2019). The effects of loan portfolio diversification on Vietnamese banks' return. Studies in Computational Intelligence, 809, 928–939. https://doi.org/10.1007/978-3-030-04200-4_68

Demirgüç-Kunt, A., & Huizinga, H. (1999). Determinants of commercial bank interest margins and profitability: Some international evidence. World Bank Economic Review, 13(2), 379–408. https://doi.org/10.1093/wber/13.2.379

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. Journal of International Financial Markets, Institutions and Money, 21(3), 307–327. https://doi.org/10.1016/j.intfin.2010.11.002

Dietrich, A., & Wanzenried, G. (2014). The determinants of commercial banking profitability in low-, middle-, and high-income countries. The Quarterly Review of Economics and Finance, 54(3), 337–354. https://doi.org/10.1016/j.qref.2014.03.001

Do, H. L. & Vu, K. T. (2019). Impact of capital on profitability of banks: Evidence from Vietnamese commercial banks. Journal of Economics and Business, 2(2), 379–395. https://doi.org/10.31014/aior.1992.02.02.94

Fotios, P., Sailesh, T., & Constantin, Z. (2009). The impact of banking regulations on banks' cost and profit efficiency: Cross-country evidence. International Review of Financial Analysis, 18(5), 294–302 https://doi.org/10.1016/j.irfa.2009.07.003

Goddard, J., Liu, H., Molyneux, P., & Wilson, J. O. S. (2010). The persistence of bank profit. Journal of Banking & Finance, 35(11), 2881–2890. https://doi.org/10.1016/j.jbankfin.2011.03.015

Goddard, J., Molyneux, P., & Wilson, J. (2004). The profitability of European banks: A cross-sectional and dynamic panel analysis. Manchester School, 72(3), 363–381. https://doi.org/10.1111/j.1467-9957.2004.00397.x

Hasnaoui, H., & Fatnassi, I. (2019). The impact of bank capital on profitability and risk in GCC countries: Islamic vs. conventional banks. Afro-Asian Journal of Finance and Accounting, 9(3), 243–268. https://doi.org/10.1504/AAJFA.2019.100976

DOI: 10.7172/2353-6845.jbfe.2021.2.3

Hermes, N., & Vu, T. H. N. (2010). The impact of financial liberalization on bank efficiency: Evidence from Latin America and Asia. Applied Economics, 42(26), 3351–3365. https://doi.org/10.1080/00036840802112448
Huynh, M. N. (2019). Tác động của sự thay đổi vốn chủ sở hữu đến hiệu quả hoạt động tại một số ngân hàng thương mại cổ phần ở Việt Nam trong giai đoạn 2008–2017 (Master's thesis). University of Economics, Ho Chi Minh City.
Iannotta, G., Nocera, G., & Sironi, A. (2007). Ownership structure, risk and performance in the European banking industry. Journal of Banking and Finance, 31(7), 2127–2149. https://doi.org/10.1016/j.jbankfin.2006.07.013
Islam, M. S., & Nishiyama, S. I. (2016). The determinants of bank net interest margins: A panel evidence from South Asian countries. Research in International Business and Finance, 37(C), 501–514. https://doi.org/10.1016/j.ribaf.2016.01.024
Jacques, K., & Nigro, P. (1997). Risk-based capital, portfolio risk and bank capital: a simultaneous equations approach. Journal of Economics and Business, 49(6), 533–547. https://doi.org/10.1016/S0148-6195(97)00038-6
Kanga, D., Victor, M., & Soumare, I. (2020). Capital, risk and profitability of WAEMU banks: Does bank ownership matter?. Journal of Banking and Finance, 114(C), 1–22. https://doi.org/10.1016/j.jbankfin.2020.105814
Langton, J. (2020). Banks' capital built to weather Covid-19 crisis: EBA. Retrieved on 15 September 2020 from https://bitly.com.vn/9nCeW
Le, D. Q. & Nguyen, D. T. (2020). Capital structure and bank profitability in Vietnam: A quantile regression approach. Journal of Risk and Financial Management, 13(8), 1–17. https://doi.org/10.3390/jrfm13080168
Lee, C., & Hsieh, M. F. (2013). The impact of bank capital on profitability and risk in Asian banking. Journal of International Money and Finance, 32(C), 251–281. https://doi.org/10.1016/j.jimonfin.2012.04.013
Luong Xuan Minh (2020). Thực trạng năng lực tài chính của các ngân hàng thương mại cổ phần trên địa bàn TP. Hồ Chí Minh. Retrieved on 15 September 2020 from https://bitly.com.vn/7xF9J
Micco, A., Panizza, U., & Yanez, M. (2007). Bank ownership and performance. Does politics matter?. Journal of Banking & Finance, 31(1), 219–241. https://doi.org/10.1016/j.jbankfin.2006.02.007
Mishkin, F. S. (2013). Mishkin: The economics of money, banking and financial markets. Colombia: Pearson.
Modigliani, F., & Merton, H. M. (1963). Corporate income taxes and the cost of capital: A correction. American Economic Association, 53(3), 433–443.
Modigliani, F., & Miller, M., (1958). The cost of capital, corporation finance and the theory of investment. The American Economic Review, 48(3), 261–281.
Munyambonera, E. F. (2013). Determinants of commercial bank profitability in sub Saharan Africa. Int. J. Econ. Finance, 5(9), 134–147. https://doi.org/10.5539/ijef.v5n9p134
Naceur, S. B., & Omran, M. (2011). The effects of bank regulations, competition, and financial reforms on banks' performance. Emerging markets review, 12, 1–20. https://doi.org/10.1016/j.ememar.2010.08.002
National Assembly Resolution 24/2016/QH14 dated November 8, 2016 on Economic Restructuring Plan 2016–2020.
Nguyen, D. S, Luu, T. Q., Pho, K. H. & McAleer, M. (2020). Net interest margin of commercial banks in Vietnam. Advances in Decision Sciences, 24, 1–27. https://doi.org/10.47654/v24y2020i1p1-27
Nguyen, T. H. V. & Le, P. T. D. T. (2016). Effects of bank capital on profitability and credit risk: The case of Vietnam's commercial banks. Journal of Economics Development, 23(4), 117–137. https://doi.org/10.24311/jed/2016.23.4.06
Nguyen, T. K. A. (2018). Tác động của vốn ngân hàng đến khả năng sinh lời và rủi ro tín dụng của các ngân hàng thương mại cổ phần ở Việt Nam. An Giang University Journal of Science, 19(1), 59–66.
Noraziah, C. A. (2019). Bank specific characteristics and profitability of Islamic and conventional banks in Malaysia. International Journal of Islamic Business, 4(1), 39–53.
Ongore, V. O., & Kusa, G. B. (2013). Determinants of financial performance of commercial banks in Kenya. International Journal of Economics and Financial Issues, 3(1), 237–252.
Ozili, P. K. (2015). Determinants of bank profitability and Basel capital regulation: Empirical evidence from Nigeria. Research Journal of Finance and Accounting, 6(2), 124–131. https://doi.org/10.2139/ssrn.2544647
Ozili, P. K. (2017). Bank profitability and capital regulation: Evidence from listed and non-listed banks in Africa. Journal of African Business, 18(2), 143–168. https://doi.org/10.1080/15228916.2017.1247329
Pasiouras, F., & Kosmidou, K. (2007). Factors influencing the profitability of domestic and foreign commercial banks in the European Union. Research in International Business and Finance, 21(2), 222–237. https://doi.org/10.1016/j.ribaf.2006.03.007
Pervan, M., Pelivan, I., & Josip A. (2015). Profit persistence and determinants of bank profitability in Croatia. Economic Research-Ekonomska Istraživanja, 28(1), 284–298. https://doi.org/10.1080/1331677X.2015.1041778
Pettway, R. H. (1976). Market tests of capital adequacy of large commercial banks. Journal of Finance, 31(3), 865–875. https://doi.org/10.1111/j.1540-6261.1976.tb01929.x
Phan, T. H. (2016). Các nhân tố ảnh hưởng đến cấu trúc vốn của doanh nghiệp công nghiệp: nghiên cứu từ mô hình GMM. Tạp chí Tài chính, 6, 47–51.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

Qian, J., Strahan, P. E. and Yang, Z. (2015). The impact of incentives and communication costs on information production and use: Evidence from bank lending. Journal of Finance, 70(4), 1457–1493. https://doi.org/10.1111/jofi.12251

Ramlan, H., & Adnan, M. S. (2016). The profitability of Islamic and conventional bank: Case study in Malaysia. Procedia Economics and Finance, 35(1), 359–367. https://doi.org/10.1016/S2212-5671(16)00044-7

Ranajee, B., (2018). Factors influencing profitability of banks in India. Theoretical Economics Letters, 8, 3046–3061. https://doi.org/10.4236/tel.2018.814189

Rime, B. (2001). Capital requirements and bank behavior: Empirical evidence for Switzerland. Journal of Banking and Finance, 25(4), 789–805. https://doi.org/10.1016/S0378-4266(00)00105-9

Rose, P. S., & Hudgins, S. C. (2008). Bank management and financial services. New York: McGraw-Hill.

Saona, P. (2016). Intra-and extra-bank determinants of Latin American banks' profitability. International Review of Economics & Finance, 45(C), 197–214. https://doi.org/10.1016/j.iref.2016.06.004

SBV Circular No. 41/2016/TT-NHNN dated December 30, 2016 on prescribing the capital adequacy ratio for operations of banks and/or foreign bank branches.

Suppia, N. M. I., & Arshad, N. C. (2019). Bank specific characteristics and profitability of Islamic and conventional banks in Malaysia. International Journal of Islamic Business, 4(1), 39–53.

Tan, Y. (2016). The impacts of risk and competition on bank profitability in China. Journal of International Financial Markets, Institutions and Money, 40(C), 85–110. https://doi.org/10.1016/j.intfin.2015.09.003

Tan, Y., & Floros, C. (2013). Risk, capital and efficiency in Chinese banking. Journal of International Financial Markets, Institutions and Money, 26, 378–393. https://doi.org/10.1016/j.intfin.2013.07.009

The National Assembly. (2010). Law no. 47/2010/QH12, Law on credit institution.

Tran, V. T., Lin, C. T., & Nguyen, H. (2016). Liquidity creation, regulatory capital, and bank profitability. International Review of Financial Analysis, 48(C), 98–109. https://doi.org/10.1016/j.irfa.2016.09.010

Vu, H., & Nahm, D. (2013). The determinants of profit efficiency of banks in Vietnam. Journal of the Asia Pacific Economy, 18(4), 615–631. https://doi.org/10.1080/13547860.2013.803847

World Bank. (2008–2018). Annual report.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

APPENDICES

Appendix 1
Multicollinearity test results

Variables	VIF	1/VIF
CAP	2.71	0.369200
LTA	1.60	0.625876
LLR	1.20	0.835388
DTA	1.46	0.685076
SIZE	4.05	0.247098
OWN	2.41	0.414668
GDP	1.39	0.720573
INF	1.55	0.646200
Mean VIF	2.04	

Source: Authors' calculations using Stata 14, 2020.

Appendix 2
Full sample: tests to select the appropriate model 2008–2019

	Model 1		Model 2	
Breusch and Pagan Lagrangian multiplier test	Chibar2(1)	Prob > chibar2	Chibar2(1)	Prob > chibar2
	149.32	0.0000	62.83	0.0000
Hausman test	Chi2(7)	Prob > chi2	Chi2(7)	Prob > chi2
	14.62	0.0411	10.67	0.1539

Source: Authors' calculations using Stata 14, 2020.

Appendix 3
Problem testing

Test	Model 1		Model 2	
Heteroskedasticity tests	Modified Wald test	Prob > chi2	Breusch and Pagan Lagrangian multiplier test	Prob > chibar2
	485.63	0.0000	62.83	0.0000
Wooldridge test for autocorrelation	F (1, 23)	Prob > F	F (1, 23)	Prob > F
	56.165	0.0000	6.908	0.0150

Source: Authors' calculations using Stata 14, 2020.

DOI: 10.7172/2353-6845.jbfe.2021.2.3

The Impact of Investor Sentiment on Direction of Stock Price Changes: Evidence from the Polish Stock Market

Kamil Polak
Kozminski University, Warsaw, Poland
kamil.polak@kozminski.edu.pl
https://orcid.org/0000-0002-9897-8431

Received: 13 September 2021 / Revised: 8 December 2021 / Accepted: 18 December 2021 / Published online: 30 December 2021

ABSTRACT

The purpose of this research is to examine the impact of sentiment derived from news headlines on the direction of stock price changes. The study examines stocks listed on the WIG-banking sub-sector index on the Warsaw Stock Exchange. Two types of data were used: textual and market data. The research period covers the years 2015–2018. Through the research, 7,074 observations were investigated, of which 3,390 with positive sentiment, 2,665 neutral, and 1,019 negative. In order to examine the predictive power of sentiment, six machine learning models were used: Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, KNN Classifier, SVC and Gaussian Naive Bayes Classifier. Empirical results show that the sentiment of news headlines has no significant explanatory power for the direction of stock price changes in one-day time frame.

JEL Classification: G14; G17; G41

Keywords: sentiment analysis, natural language processing, machine learning, financial forecasting, behavioral finance.

1. INTRODUCTION

The dynamic technological development has significantly increased the information role of the Internet in recent times. The arrival of the era known as Web 2.0 gave rise to a completely new way of communication, namely social media. The exchange of information via online channels takes place almost immediately, which gives a significant information advantage over the traditional methods of data acquisition, such as board reports or articles. In hindsight, over the last few years social media platforms have become not only means of communication expressing their own opinions, but also publicly sharing emotions.

The development of technology has significantly contributed to the evolution of research tools and methods enabling the exploration of new areas, and thus an increase in interdisciplinary research. Behavioral economics, natural language processing and machine learning have been proved useful in analyzing and modeling the impact of emotional impulses on the behavior of financial markets.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

In addition to the development of technology, which resulted in changes in the functioning of financial markets, an equally important issue is the development of the market theory which is both a response to changes and an attempt to explain the mechanisms that form current financial markets. The result is the adaptive market hypothesis (Lo, 2005). The adaptive market hypothesis partly based on the assumptions of behavioral economics points out that price formation is influenced not only by market data, but also by how they are perceived by market participants, which can be examined using sentiment analysis.

Sentiment analysis deals with detecting the general mood prevailing in online resources and social media to understand how people think about a given topic (Nassirtoussi et al., 2015). It is mainly based on identifying positive and negative words and processing text to classify its emotional attitude as positive or negative. Sentiment analysis is based on two assumptions. First, some words express emotions. Second, there are words that can cause emotions when they are spoken (Pang et al., 2002). Thus, on the one hand, sentiment analysis indicates the emotional states of the author of the statement, on the other – it also serves to determine the emotional effect that a given statement can cause (Tomanek, 2014).

The results of research on the impact of information from social media cast new light on the problem of prediction of price changes on capital markets (Johnman et al., 2018; Pagolu et al., 2016; Pasupulety et al., 2019).

The main purpose of this research is to examine the impact of sentiment on the direction of stock price changes on the WIG-banking sub-sector index on the Warsaw Stock Exchange. Specifically, this paper uses several classification machine learning techniques to predict the direction of one-day-ahead stock price changes based on sentiment derived from news headlines.

A detailed analysis of the research results presented in the literature review section allowed for formulating the following hypothesis: The sentiment data extracted from news headlines allow for stock price predictions on the WIG-banking sub-sector index in a one-day time horizon.

It is worth noting that the studies conducted so far have been based, in most cases, on the American market, which has different characteristics compared to the Polish market, i.e. market capitalization, trading volume, number of the investors, specificity of the language, as well as some cultural differences that can cause different perception of the information published. My study attempts to close this gap by implementation of the existing research methods on the Warsaw Stock Exchange.

The paper is organized as follows. Section 2 presents the theory of adaptive markets hypothesis emphasizing the impact of emotional overtones accompanying emerging information about a given entity on the prices of financial instruments. The second part of the chapter presents the issues of sentiment in the context of the possibility of explaining changes on the financial market. Section 3 describes the data used in the study and the methodology of the research. Section 4 presents the results of the study. Finally, the last section concludes.

2. LITERATURE REVIEW

2.1. Adaptive Market Hypothesis

A major part of modern investment theory and practice is based on the Efficient Market Hypothesis (Fama, 1965). This concept assumes that markets fully, accurately and immediately incorporate all available information into market prices. At the root of this far-reaching idea is the assumption that market participants are rational economic entities, always acting in their own interests and making decisions in an optimal way (Lo, 2005). This means that stock prices cannot be predicted because they depend on new information rather than current/past prices. As a result,

DOI: 10.7172/2353-6845.jbfe.2021.2.4

stock prices are subject to random walks. In the updated study (Malkiel & Fama, 1970), the authors stated that efficiency can take three forms: weak, semi-strong and strong.

The implication of the Efficient Market Hypothesis is that the market cannot be beaten because all information that could predict performance is already incorporated into the stock price. However, several studies provide evidences contrary to the suggestion of the Efficient Market Hypothesis and Random Walk Theory (Bollen et al., 2011; Schumaker et al., 2012). These studies show that the stock market can be predicted to some extent, and thus question the basic assumptions of the above hypothesis. This phenomenon was explained by behavioral economics, which argues that markets are not efficient, and the element of random walk can be explained by human behavior, because ultimately people are responsible for making decisions, and as people they make irrational and systematic mistakes. These errors affect prices and returns, resulting in inefficiency of the market.

Lo (2005), based on the analysis of the Adaptive Markets Hypothesis (Lo, 2004) attempted to reconcile the Efficient Markets Hypothesis and behavioral economics theory. This hypothesis is based on some well-known principles of evolutionary biology – competition, mutation, reproduction and natural selection. Translating this into the realities of financial markets, this means that the degree of market efficiency is related to environmental factors, such as the number of competitors on the market, the scale of available profit opportunities and the ability to adapt participants to the changing market situation. In other words, it is unrealistic to expect perfectly efficient/inefficient markets – due to behavioral bias. The importance of Adaptive Markets Hypothesis is well documented in the literature (Charles et al., 2012; Hiremath & Narayan, 2016).

Literature provides many examples proving the assumption of Adaptive Markets Hypothesis. Zhou and Lee (2013) analyzed REITs listed on NYSE, AMEX and NASDAQ. Based on the sample of 7,570 daily observations from the period of January 1980 – December 2009, they proved that market efficiency depends on the behavior on given market and is variable over time. Therefore, it cannot be considered as a binary variable, which confirms the assumptions of the Adaptive Market Hypothesis. Kim et al. (2011) evaluated return predictability of the daily and weekly Dow-Jones Industrial Average indices from 1900 to 2009. Based on the analysis, they found strong evidence that stock returns, e.g. during fundamental economic or political crises, have been highly predictable with a moderate degree of uncertainty, which confirms that predictability is driven by changing market conditions.

As a result, building a predictive model based solely on the analysis of historical time series or micro/macroeconomic data puts a big question mark on its effectiveness. The reason for that should be seen in the fact that not only the above-mentioned data have an impact on financial markets, but also on the way they are perceived by market participants. As numerous studies show, sentiment analysis is a factor that significantly improves the effectiveness of prediction.

2.2. Sentiment Analysis

Sentiment analysis deals with detecting the general mood prevailing in online resources and social media to understand how people think about a given topic (Nassirtoussi et al., 2015). It is mainly based on identifying positive and negative words and processing text to classify its emotional attitude as positive or negative. Sentiment analysis is based on two assumptions. First, some words express emotions. Second, there are words that can trigger termination of emotions (Pang et al., 2002).

Thus, on the one hand, sentiment analysis indicates the emotional states of the author, on the other it also serves to determine the emotional effect that a given statement can cause (Tomanek, 2014).

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Sentiment analysis is performed at three levels. At the first, the document level, the entire content of the document is classified to determine whether it contains a positive or negative attitude. At the second level, the sentence level, it is determined whether the sentence contains a positive, negative or neutral opinion. Neutral overtones can also mean no opinion. The last level is the entity and aspect level (Liu, 2012).

In sentiment analysis, two methods are used to classify the text:

- dictionary,
- statistical.

The dictionary method can be based on a set of opinion words or the entire corpus of texts. It assumes that there are certain words that are often used to express emotions (Pang et al., 2002). The dictionary method takes into account the meaning of the analyzed words and lexical rules of a given language. Therefore, it is necessary to know the grammatical rules of the analyzed language and the specificity of utterances related to the vocabulary used (Tomanek, 2014).

Second, the statistical method, treats the text as an object, which is represented using quantitative data in the form of, e.g., the number of words or phrases. The statistical method represents the object in the form of a vector in a multi-dimensional space defined by a set of features (Tomanek, 2014).

Sentiment analysis of news can be an effective source for market forecasts, because it expresses the point of view and the mood of opinion leaders who, to some extent, form public opinion and cause public reactions. It is not surprising, then, that the impact of the sentiment of emerging information on price formation has become the focus of extensive research (Hagenau et al., 2013; P. Mehta et al., 2021; Schumaker et al., 2012; Urlam, 2021; Valle-Cruz et al., 2021).

Textual input data used for the sentiment analysis model can have multiple sources. Most of the research has used information gathered from platforms such as Bloomberg (Chatrath et al., 2014; Gumus & Sakar, 2021; Jin et al., 2013) and Yahoo Finance (Rechenthin et al., 2013; Xie et al., 2013).

Periodic financial reports published by the companies are another category of input data sources. It is worth noting that this type of data has a special feature which is the periodicity of publication. According to Huang et al. (2010), a strictly fixed data release schedule may affect predictive capabilities resulting from existing investor expectations for achieving specific financial results. Recently, there has been an increase in interest in the third, less formal source of information such as blogs, microblogs and forums (Yu et al., 2013).

The second category of input data necessary to quantify the impact of sentiment in the context of financial markets are market data in the form of historical quotations for a given financial instrument. Depending on the type of instrument and availability, it may be OHLC data (open-high-low-close) (Y. Mehta et al., 2021; Pagolu et al., 2016; Xie et al., 2013) or only closing price information (Chatrath et al., 2014; Jin et al., 2013; Kumar et al., n.d.). The frequency of the data depends on the frequency of the data containing sentiment.

3. METHODOLOGY AND DATA

3.1. Sample

The objective of the research was to evaluate sentiment derived from news headlines for stock price predictions on the WIG-banking sub-sector index. The composition of the WIG-banking index in the analyzed period, i.e. in the years 2015–2018 is presented in Table 1.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Table1
Composition of the WIB-banking index

Issuer	Ticker	ISIN
Alior Bank SA	ALR	PLALIOR00045
Banco Santander SA	SAN	ES0113900J37
Bank Handlowy w Warszawie SA	BHW	PLBH00000012
Bank Millennium SA	MIL	PLBIG0000016
Bank Ochrony Środowiska SA	BOS	PLBOS0000019
Bank Polska Kasa Opieki SA	PEO	PLPEKAO00016
Getin Holding SA	GTN	PLGSPR000014
Getin Noble Bank SA	GNB	PLGETBK00012
Idea Bank SA	IDA	PLIDEAB00013
ING Bank Śląski SA	ING	PLBSK0000017
mBank SA	MBK	PLBRE0000012
Powszechna Kasa Oszczędności Bank Polski SA	PKO	PLPKO0000016
Bank Zachodni WBK	SPL	PLBZ00000044
UniCredit S.p.A.	UCG	IT0005239360

Source: Self-preparation based on the Stock Exchange Annals published by GPW S.A.

The decision to focus on the WIB-banking index only was taken based on the analysis which showed a dominant share in WIG index across the research period (27.95%, 27.26%, 28.85%, 28.53% respectively). This fact implies the highest number of textual data available for the study among all the indexes. Furthermore, in the analyzed period, the composition of the index did not change, which eliminates potential disturbances resulting from the change in the characteristics of the index.

The study used two data sources: textual data from the InfoStrefa[1] website owned by the Polish Press Agency and the Warsaw Stock Exchange and market data in the form of historical prices of companies in the research sample. The research period is 01.01.2015–31.12.2018 representing a total of 1000 session days. Textual data were extracted based on web scrapping techniques. In order to do that, the web crawler was created.

Historical daily time series was derived from the InfoStrefa. For each session, the following data were collected:

- Open price,
- Minimum and maximum price in each trading session,
- Close price,
- Trading volume,
- Turnover value,
- Number of transactions.

[1] http://infostrefa.com/infostrefa/pl. The decision was taken for analysis headers based on the work by Huang et al. (2010), which demonstrated that the headlines are more direct than the entire text – a consequence of a lower level of information noise.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Based on the close price, the daily rate of return was computed in the following way:

$$r_i = \ln\left(\frac{p_i}{p_{i-1}}\right)$$

where:
r_i – daily rate of return,
p_i – stock price in day i,
p_{i-1} – stock price in day i–1.

In the next step, each daily rate of return was labeled in the following way:

$$DIRECTION: \begin{cases} r_i > 0 : 1 \\ r_i \leq 0 : 0 \end{cases}$$

which allows to construct dependent variable "DIRECTION".

3.2. Data Pre-Processing

Text data were the subject of a few pre-processing steps which contain:
- Tokenization,
- Stop words removal,
- Non-alphanumeric characters removal,
- Conversion text to lower case.

The above steps were performed using the RE and NLTK libraries available for Python.

In the next step the stemming procedure was applied. Stemming was performed based on PoliMorf[2]. Before PoliMorf was applied, the Bug-of-word method (Hájek, 2018) was used to represent the corpus in the form of a sparse matrix.

Extracting the stems from each token made it possible to reduce the number of inflectional forms and thus facilitated the further process of assigning sentiment to each of the headers.

Sentiment assignment was done based on the Polish sentiment dictionary[3]. The dictionary is a list of words with negative, neutral and positive sentiment. For each of the headline, the number of occurring words with given sentiment was counted. Then, based on the comparison, the overall sentiment of each headline was determined, i.e. if the number of words with negative sentiment prevailed in a given message, then a negative sentiment was assigned. Sentiment was labeled in the following way:

$$sentiment = \begin{cases} -1: negative \\ \ \ 0: neutral \\ \ \ 1: positive \end{cases}$$

The final stage of data pre-processing was the merging of datasets containing text data with the corresponding market data. The date was used as a key which enabled the merging process. The final dataset included 7,074 observations, of which 3,390 with positive sentiment, 2,665 neutral, and 1,019 negative.

[2] PoliMorf is the morphological dictionary for Polish resulting from the standardization and merger of Morfeusz SGJP and Morfologik financed by CESAR project.

[3] Polish sentiment dictionary was created by The Linguistic Engineering (LE) Group, which is part of the Department of Artificial Intelligence at the Institute of Computer Science, Polish Academy of Sciences.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

3.3. Machine Learning

To examine sentiment derived from news headlines for stock price predictions, several machine learning models were applied. Based on the literature review (Jabreel & Moreno, 2018; John & Vechtomova, 2017; Lango et al., 2016), the following algorithms were chosen:

- Gaussian Naive Bayes Classifier,
- Support Vector Classifier,
- KNN Classifier,
- Decision Tree Classifier,
- Random Forest Classifier,
- XGBoost Classifier.

The model training phase was preceded by a procedure of eliminating distortions in both the learning process and the result itself. The procedure consisted of the following:

- verification of data completeness,
- verification of correlations between variables,
- selection of independent variables and the dependent variable,
- splitting the data into training and test set,
- standardization of independent variables.

Data pre-processing as well as machine learning were performed using the Python programming language with dedicated libraries such as Pandas, Numpy, Matplotlib, Seaborn, and Scikit-learn.

Verification of completeness of data was aimed at checking whether there were any variables with missing values in the dataset.

To verify the correlation between variables[4], the Pearson correlation coefficient was calculated for each pair. As the Pearson correlation coefficient assumes the assumption of linear dependence of variables and normal distribution, for additional verification the Spearman correlation coefficient was also calculated.

In addition, the significance of correlations between strongly correlated variables was examined. A two-sided 95% confidence interval was used for the analysis.

Based on the above analysis, the following variables were selected. Independent variables (x): CLOSE, CHANGE, VOLUME, NUMBER OF TRANSACTIONS, SENTIMENT, COUNT_WORD, MEAN_WORD_LEN; dependent variable (y): DIRECTION.

In accordance with accepted practice found in the literature, the data were split into a training set and a test set based on which the learning performance was assessed. The ratio between the two sets is 80/2020.

The last stage preceding model training was the standardization of independent variables.

3.4. Model Evaluation

The predictive power of each model was assessed by comparing the model result with the set "y_test" containing the set of expected values of the DIRECTION variable. For each model, a classification report was prepared. To evaluate the model prediction, two metrics were taken into consideration: accuracy and AUC (Area Under Curve) (Huang & Ling, 2003; Nassirtoussi et al., 2015; Rokach & Maimon, 2014)

In addition, for each model the learning curve was constructed based on 10-fold cross validation to assess if the model is not underfitted/overfitted (Cawley & Talbot, 2010; Guyon, 2009; Guyon et al., 2010).

---

4 The list of all variables with their description is presented in Appendix 1.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Finally, the feature importance was computed, which allowed to increase the interpretability of the model results. The analysis was performed using the permutation importance method available in Eli5 library. The procedure is as follows:
(1) Get a trained model.
(2) Shuffle the values in a single column, make predictions using the resulting dataset. Use these predictions and the true target values to calculate how much the loss function suffered from shuffling. That performance deterioration measures the importance of just shuffled variable.
(3) Return the data to the original order (undoing the shuffle from step 2). Now repeat step 2 with the next column in the dataset until obtaining calculations of the importance of each column.

4. RESULTS

Table 2 presents the aggregated results of all machine learning models.

Table 2
Summary results

	SVC	KNN	Random Forest	Decision Tree	XGBoost	Gaussian Naive Bayes Classifier
Accuracy	0.9611	0.9307	0.9986	0.9957	0.9978	0.9102
AUC	0.9611	0.9307	0.9986	0.9957	0.9978	0.9103

Source: Self-preparation based on empirical results.

Empirical results showed that each model achieved the AUC and accuracy score above 90%. The best classification performance was achieved by CART algorithms, i.e. Decision Tree, Random Forest and XGBoost, where Random Forest has the highest accuracy as well as AUC. The difference between the best CART algorithm, i.e. Random Forest and model with the lowest score – Gaussian Naive Bayes Classifier was 8.84 p.p. for accuracy and 8.83 p.p. for AUC.

Neither of the algorithms showed any problem with underfitting or overfitting. In the case of the KNN Classifier and SVC algorithms, in the initial phase, there was a slight mismatch to the data, which, however, decreased with the increase in the number of training samples.

However, based on the permutation importance technique, it was identified that the impact of the variable SENTIMENT turned out to have an insignificant impact on the prediction result of the algorithms. In other words, the study showed that the sentiment of the data extracted from news headlines does not allow for stock price predictions on the WIG-banking sub-sector index in a one-day time horizon. Therefore, the research hypothesis cannot be accepted. Detailed results of each model including performance metrics, learning curve and future important analysis are presented in the following sections.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

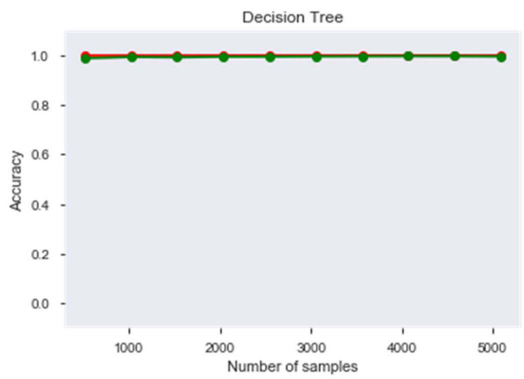
4.1. Estimating Results: Decision Tree Classifier

As shown in Figure 1, accuracy and AUC of Decision Tree Classifier is respectively: 0.9958 and 0.9957.

Figure 1
Classification report for Decision Tree Classifier

```
Model Performance metrics:
------------------------------
Accuracy: 0.9958
Precision: 0.9958
Recall: 0.9958
F1 Score: 0.9958

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.99      1.00      1.00       706
           0       1.00      0.99      1.00       709

    accuracy                           1.00      1415
   macro avg       1.00      1.00      1.00      1415
weighted avg       1.00      1.00      1.00      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                   1    0
Actual: 1        704    2
        0          4  705


AUC:  0.995762694934013
```

Figure 2
Learning curve for Decision Tree Classifier



The learning curve does not show any problem with underfitting or overfitting.

Figure 3
Feature importance report for Decision Tree Classifier

Weight	Feature
0.4779 ± 0.0325	Change
0.0174 ± 0.0038	Number of transactions
0.0113 ± 0.0025	Close
0.0010 ± 0.0014	count_word
0 ± 0.0000	Volume
-0.0001 ± 0.0016	mean_word_len
-0.0006 ± 0.0011	Sentiment

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Figure 3 shows the results of future importance assessment. The values towards the top are the most important features, and those towards the bottom matter least.

The first number in each row shows how much model performance decreased with a random shuffling (in this case, using “accuracy” as the performance metric).

Since there is some randomness in the exact performance change resulting from the shuffling of a specific column, the amount of randomness in permutation importance calculation is computed by repeating the process with multiple shuffles. The number after the ± measures how performance varied from one reshuffling to the next. The results indicate that variable CHANGE has the most significant impact on model prediction.

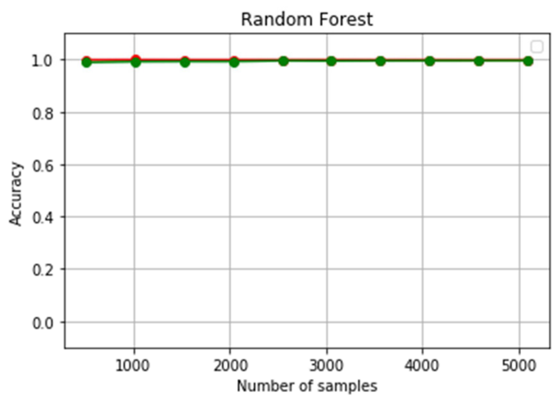
4.2. Estimating Results: Random Forest Classifier

Figure 4
Classification report for Random Forest Classifier

```
Model Performance metrics:
------------------------------
Accuracy: 0.9986
Precision: 0.9986
Recall: 0.9986
F1 Score: 0.9986

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       1.00      1.00      1.00       706
           0       1.00      1.00      1.00       709

    accuracy                           1.00      1415
   macro avg       1.00      1.00      1.00      1415
weighted avg       1.00      1.00      1.00      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                   1    0
Actual: 1        705    1
        0          1  708

AUC:  0.998586566084778
```

Random Forest Classifier results are slightly better than Decision Tree Classifier (28 bp for both accuracy and AUC).

Figure 5
Learning curve for Random Forest Classifier



DOI: 10.7172/2353-6845.jbfe.2021.2.4

Since the shape of learning curve is very close to the Decision Tree Classifier curve, there is no problem with model stability.

Figure 6
Feature importance report for Random Forest Classifier

Weight	Feature
0.4817 ± 0.0311	Change
0.0126 ± 0.0030	Close
0.0105 ± 0.0011	Number of transactions
0.0031 ± 0.0007	Volume
0.0001 ± 0.0006	mean_word_len
0.0001 ± 0.0011	Sentiment
-0.0006 ± 0.0006	count_word

The feature importance report presented in Figure 6 shows that variable CHANGE has the greatest impact on model performance.

4.3. Estimating Results: XGBoost Classifier

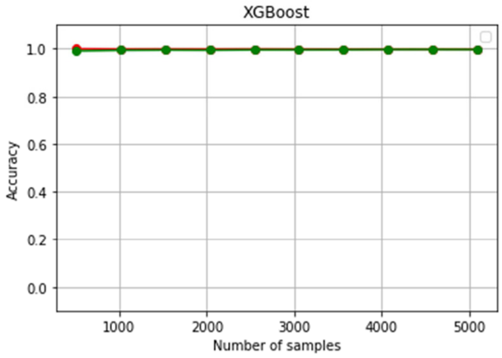
Figure 7
Classification report for XGBoost Classifier

```
Model Performance metrics:
------------------------------
Accuracy: 0.9979
Precision: 0.9979
Recall: 0.9979
F1 Score: 0.9979

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       1.00      1.00      1.00       706
           0       1.00      1.00      1.00       709

    accuracy                           1.00      1415
   macro avg       1.00      1.00      1.00      1415
weighted avg       1.00      1.00      1.00      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                   1    0
Actual: 1        704    2
        0          1  708


AUC:   0.9978783507873276
```

Performance metrics presented in Figure 7 show that compared to the previous two algorithms, XGBoost ranks between the Decision Tree Classifier and Random Forest Classifier algorithms.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Figure 8
Learning curve for XGBoost Classifier



The learning curve is very similar to the previous two algorithms. Thus, model was not affected by underfitting and overfitting.

Figure 9
Feature importance report for XGBoost Classifier

Weight	Feature
0.4841 ± 0.0293	Change
0.0131 ± 0.0034	Close
0.0045 ± 0.0014	Number of transactions
0.0023 ± 0.0006	Volume
0.0008 ± 0.0011	mean_word_len
0 ± 0.0000	Sentiment
-0.0003 ± 0.0011	count_word

Variable SENTIMENT has no significant impact on model performance, while variable CHANGE was the one with the highest significance.

4.4. Estimating Results: SVC

Figure 10
Classification report for SVC

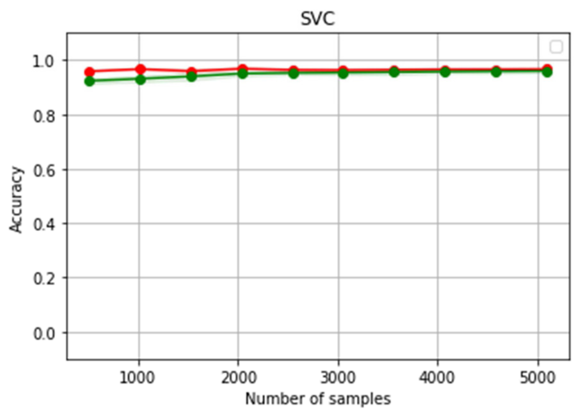
```
Model Performance metrics:
------------------------------
Accuracy: 0.9611
Precision: 0.9612
Recall: 0.9611
F1 Score: 0.9611

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.96      0.96      0.96       706
           0       0.96      0.96      0.96       709

    accuracy                           0.96      1415
   macro avg       0.96      0.96      0.96      1415
weighted avg       0.96      0.96      0.96      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                   1    0
Actual: 1        681   25
        0         30  679

AUC:  0.961138059030594
```

The classification report shows that SVC achieved slightly worse results that CART-based models, i.e. Decision Tree Classifier, Random Forest Classifier and XGBoost Classifier.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

The analysis of the learning curve presented in Figure 11 does not indicate model instability.

Figure 11
Learning curve for SVC



The feature importance results presented in Figure 12 are aligned with CART-based models, i.e. showing that variable SENTIMENT has no significant impact on model performance.

Figure 12
Feature importance report for SVC

Weight	Feature
0.4500 ± 0.0262	Change
0.0090 ± 0.0044	count_word
0.0083 ± 0.0035	Volume
0.0073 ± 0.0070	Number of transactions
0.0047 ± 0.0007	mean_word_len
0.0023 ± 0.0024	Close
0.0020 ± 0.0026	Sentiment

4.5. Estimating Results: KNN Classifier

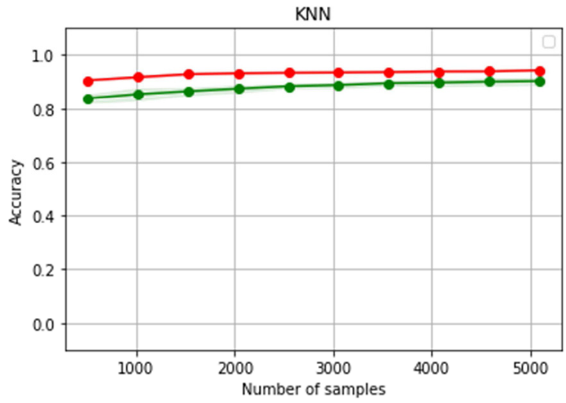
Figure 13
Classification report for KNN Classifier

```
Model Performance metrics:
------------------------------
Accuracy: 0.9208
Precision: 0.9209
Recall: 0.9208
F1 Score: 0.9208

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.93      0.91      0.92       706
           0       0.92      0.93      0.92       709

    accuracy                           0.92      1415
   macro avg       0.92      0.92      0.92      1415
weighted avg       0.92      0.92      0.92      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                   1    0
Actual: 1        645   61
        0         51  658

AUC:  0.9307117713573361
```

KNN Classifier performance is definitely lower than CART-based models, e.g. Decision Tree Classifier accuracy was 0.9958 which is better by 7.5 bp.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Figure 14
Learning curve for KNN Classifier



The learning curve presented in Figure 14 indicates a slight variance that decreases as the number of learning samples increases. The increase in the number of learning samples is also accompanied by an increase in classification accuracy.

Figure 15
Feature importance report for KNN Classifier

Weight	Feature
0.3963 ± 0.0267	Change
0.0228 ± 0.0108	mean_word_len
0.0225 ± 0.0067	count_word
0.0212 ± 0.0114	Number of transactions
0.0189 ± 0.0129	Close
0.0133 ± 0.0079	Sentiment
0.0107 ± 0.0076	Volume

The analysis of feature importance is aligned with the results of previous models.

4.6. Estimating Results: Naive Bayes Classifier

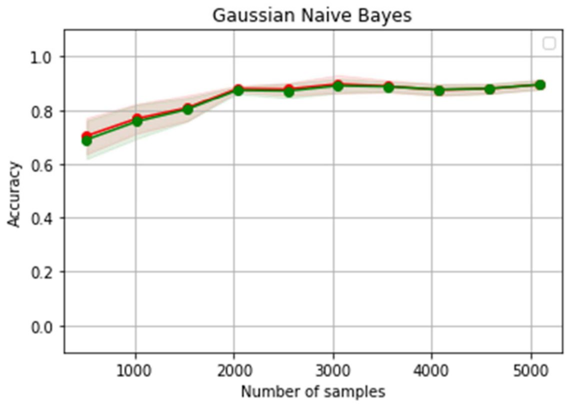
The classification report presented in Figure 16 shows that Naive Bayes Classifier performance is the worst among all applied models.

Figure 16
Classification report for Naive Bayes Classifier

```
Model Performance metrics:
------------------------------
Accuracy: 0.9102
Precision: 0.9154
Recall: 0.9102
F1 Score: 0.91

Model Classification report:
------------------------------
              precision    recall  f1-score   support

           1       0.87      0.97      0.91       706
           0       0.96      0.85      0.91       709

    accuracy                           0.91      1415
   macro avg       0.92      0.91      0.91      1415
weighted avg       0.92      0.91      0.91      1415


Prediction Confusion Matrix:
------------------------------
          Predicted:
                  1    0
Actual: 1       682   24
        0       103  606

AUC:   0.9103653152307243
```


DOI: 10.7172/2353-6845.jbfe.2021.2.4

Figure 17
Learning curve for Naive Bayes Classifier



Naive Bayes Classifier was not affected by underfitting and overfitting. It is worth noting the increase in model accuracy with an increase in the number of learning samples up to the level of 2000.

Figure 18
Feature importance report for Naive Bayes Classifier

Weight	Feature
0.3984 ± 0.0247	Change
0.0133 ± 0.0081	Number of transactions
0.0083 ± 0.0027	Close
0.0052 ± 0.0065	mean_word_len
0.0031 ± 0.0029	Volume
0.0007 ± 0.0039	count_word
0.0003 ± 0.0039	Sentiment

Similarly to the other models, analysis of feature importance shows that variable SENTIMENT has no significant impact on model performance.

5. CONCLUSION

The objective of the research was to examine the impact of sentiment derived from news headlines for stock price predictions on the WIG-banking sub-sector index. The text data as well as market data were derived from the InfoStrefa website owned by the Polish Press Agency and the Warsaw Stock Exchange. The research period is 01.01.2015–31.12.2018 representing a total of 1000 session days.

To examine the impact of sentiment, six machine learning models were used: Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, KNN Classifier, SVC and Gaussian Naive Bayes Classifier.

Empirical results show that each model achieved both accuracy and AUC above 90%, i.e. a good ability to predict the direction of price change in one-day time horizon. Furthermore, based on the analysis of learning curves, it was assessed that none of the models was affected by underfitting or overfitting.

However, the results of feature importance analysis show that for each of the model variable SENTIMENT, which contains information about emotional attitude, had no significant impact on its classification performance. Thus, it cannot be concluded that sentiment of news headlines has a significant impact on stock price changes.

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Therefore, the conclusion from my study is not similar to the results of most research conducted so far. It is worth noting, however, that the research carried out so far has been based on foreign markets, in most cases on the American market, which has different characteristics compared to the Polish market, i.e. market capitalization, number of investors as well as cultural differences which may cause different perception of published information. In terms of the Polish market, Wojarnik (2021) concluded that sentiment analysis of texts posted on discussion forums may be useful in analyzing the behavior of stock price. However, it should be noted that this research was devoted to three companies from the WIG-GAMES index and a different type of textual data was used.

In addition, it should be taken into account that the financial language, like any other industry language, has a number of specific phrases and terms not used in colloquial speech. According to Loughran and McDonald (2011), who created their own dictionary classifying sentences related to the field of economics and finance, nearly 75% of sentences classified based on the Harvard dictionary as negative after using their dictionary turned out to be positive. The difference was the result of a different sense context.

The PoliMorf and Polish sentiment dictionary were build on the basis of the Polish Language Grammatical Dictionary, which does not take into account the financial context. As a result, there is a risk of imprecisely identified sentiment of messages. Unfortunately, at the time of the research, there was no Polish dictionary available for the financial industry terminology.

On the basis of the conducted research, there are several future directions for this area of research that could be suggested. The first one is to attempt to create a dictionary of financial terms for the Polish language. Perhaps this will lead to better predictivity. The second future direction would be to investigate other machine learning techniques. While classical machine learning models have proven themselves in the textual financial domain, perhaps other more advanced techniques, e.g. deep learning models, could achieve better results. The third future direction would be to explore another source of text data, e.g. social media platforms such as Twitter or Facebook.

References

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

Cawley, G.C., & Talbot, N.L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11, 2079–2107.

Charles, A., Darné, O., & Kim, J.H. (2012). Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates. Journal of International Money and Finance, 31, 1607–1626. https://doi.org/10.1016/j.jimonfin.2012.03.003

Chatrath, A., Miao, H., Ramchander, S., & Villupuram, S. (2014). Currency jumps, cojumps and the role of macro news. Journal of International Money and Finance, 40, 42–62. https://doi.org/10.1016/j.jimonfin.2013.08.018

Fama, E.F. (1965). The behavior of stock-market prices. The Journal of Business, 38, 34–105. https://doi.org/10.1086/294743

Gumus, A., & Sakar, C.O. (2021). Stock market prediction by combining stock price information and sentiment analysis. International Journal of Advances in Engineering and Pure Sciences, 33, 18–27.

Guyon, I. (2009). A practical guide to model selection. In Proc. Mach. Learn. Summer School Springer Text Stat. (pp. 1–37).

Guyon, I., Saffari, A., Dror, G., & Cawley, G. (2010). Model selection: Beyond the bayesian/frequentist divide. Journal of Machine Learning Research, 11, 61–87.

Hagenau, M., Liebmann, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. Decision Support Systems, 55, 685–697. https://doi.org/10.1016/j.dss.2013.02.006

Hájek, P. (2018). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. Neural Computing and Applications, 29, 343–358. https://doi.org/10.1007/s00521-017-3194-2

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Hiremath, G.S., & Narayan, S. (2016). Testing the adaptive market hypothesis and its determinants for the Indian stock markets. Finance Research Letters, 19, 173–180. https://doi.org/10.1016/j.frl.2016.07.009

Huang, C.-J., Liao, J.-J., Yang, D.-X., Chang, T.-Y., & Luo, Y.-C. (2010). Realization of a news dissemination agent based on weighted association rules and text mining techniques. Expert Systems with Applications, 37, 6409-6413. https://doi.org/10.1016/j.eswa.2010.02.078

Huang, J., & Ling, C.X. (2003). Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 17, 299–310. https://doi.org/10.1109/TKDE.2005.50

Jabreel, M., & Moreno, A. (2018). EiTAKA at SemEval-2018 Task 1: An ensemble of n-channels ConvNet and XGboost regressors for emotion analysis of tweets. arXiv preprint arXiv:1802.09233. https://doi.org/10.18653/v1/S18-1029

Jin, F., Self, N., Saraf, P., Butler, P., Wang, W., & Ramakrishnan, N. (2013). Forex-foreteller: Currency trend modeling using news articles. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1470–1473). ACM. https://doi.org/10.1145/2487575.2487710

John, V., & Vechtomova, O. (2017). Sentiment analysis on financial news headlines using training dataset augmentation. arXiv preprint arXiv:1707.09448. https://doi.org/10.18653/v1/S17-2149

Johnman, M., Vanstone, B.J., & Gepp, A. (2018). Predicting FTSE 100 returns and volatility using sentiment analysis. Accounting & Finance, 58, 253–274. https://doi.org/10.1111/acfi.12373

Kim, J.H., Shamsuddin, A., & Lim, K.-P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. Journal of Empirical Finance, 18, 868–879. https://doi.org/10.1016/j.jempfin.2011.08.002

Kumar, K.S.M.V., Kumar, G.R., & Rao, J.N. (2020) Use sentiment analysis to predict future price movement in the stock market. International Journal of Advanced Research in Engineering and Technology, 11, 1123-1130.

Lango, M., Brzezinski, D., & Stefanowski, J. (2016). PUT at SemEval-2016 Task 4: The ABC of Twitter sentiment analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) (pp. 126–132). https://doi.org/10.18653/v1/S16-1018

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5, 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Lo, A.W. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. Journal of Investment Consulting, 7, 21–44.

Lo, A.W. (2004). The adaptive markets hypothesis. The Journal of Portfolio Management, 30, 15–29. https://doi.org/10.3905/jpm.2004.442611

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66, 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Malkiel, B.G., & Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25, 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476. https://doi.org/10.7717/peerj-cs.476

Mehta, Y., Malhar, A., & Shankarmani, R. (2021). Stock price prediction using machine learning and sentiment analysis. Paper presented at the 2nd International Conference for Emerging Technology (INCET) IEEE. https://doi.org/10.1109/INCET51464.2021.9456376

Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., & Ngo, D.C.L. (2015). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. Expert Systems with Applications, 42, 306–324. https://doi.org/10.1016/j.eswa.2014.08.004

Pagolu, V.S., Reddy, K.N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. Paper presented at the International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) IEEE. https://doi.org/10.1109/SCOPES.2016.7955659

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (Vol. 10, pp. 79–86). Association for Computational Linguistics. https://doi.org/10.3115/1118693.1118704

Pasupulety, U., Anees, A.A., Anmol, S., & Mohan, B.R. (2019). Predicting stock prices using ensemble learning and sentiment analysis. Paper presented at the IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). https://doi.org/10.1109/AIKE.2019.00045

Rechenthin, M., Street, W.N., & Srinivasan, P. (2013). Stock chatter: Using stock sentiment to predict price direction. Algorithmic Finance, 2, 169–196. https://doi.org/10.3233/AF-13025

Rokach, L., & Maimon, O. (2014). Data mining with decision trees: Theory and applications. World Scientific Publishing Co. https://doi.org/10.1142/9097

Schumaker, R.P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. Decision Support Systems, 53, 458–464. https://doi.org/10.1016/j.dss.2012.03.001

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Tomanek, K. (2014). Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych. Przegląd Socjologii Jakościowej, 10, 118–136.

Urlam, S. (2021). Stock market prediction using LSTM and sentiment analysis. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12, 4653–4658.

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2021). Does Twitter affect stock market decisions? Financial sentiment analysis during pandemics: A comparative study of the H1N1 and the COVID-19 periods. Cognitive Computation,1–16. https://doi.org/10.21203/rs.3.rs-39991/v1

Wojarnik, G. (2021). Sentiment analysis as a factor included in the forecasts of price changes in the stock exchange. Procedia Computer Science, 192, 3176–3183. https://doi.org/10.1016/j.procs.2021.09.090

Xie, B., Passonneau, R., Wu, L., & Creamer, G.G. (2013). Semantic frames to predict stock price movement, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. pp. 873–883.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. Decision Support Systems, 55, 919–926. https://doi.org/10.1016/j.dss.2012.12.028

Zhou, J., & Lee, J.M. (2013). Adaptive market hypothesis: Evidence from the REIT market. Applied Financial Economics, 23, 1649–1662. https://doi.org/10.1080/09603107.2013.844326

DOI: 10.7172/2353-6845.jbfe.2021.2.4

APPENDIX 1

Description of variables

Variable	Description
COMPANY	Company name
DAY	Date of trading session
TIME	Time of publication of the news
HEADLINE	Headline content
OPEN	Opening price
MAX_PRICE	Highest price
MIN_PRICE	Lowest price
CLOSE	Closed price
CHANGE	The percentage change in the price
DIRECTION	Binary variable: 0 – price decrease/no change, 1 – price increase
VOLUME	Volume
NUMBER OF TRANSACTIONS	Number of executed transactions
TURNOVER	Turnover (in PLN).
SENTIMENT	Variable which contains information about emotional attitude: –1 negative, 0 neutral, 1 positive
COUNT_WORD	Number of words in each headline
COUNT_LETTERS	Number of letters in each headline
MEAN_WORD_LEN	Average length of word in each headline

DOI: 10.7172/2353-6845.jbfe.2021.2.4

Re-Evaluating Sharpe Ratio in Hedge Fund Performance in Light of Liquidity Risk

Richard Van Horne
Poznan University of Economics and Business
Institute of Finance, Department of Investment and Financial Markets
Al. Niepodległości 10, 61-875 Poznań
https://orcid.org/0000-0002-4648-9601

Katarzyna Perez[1]
Poznan University of Economics and Business
Institute of Finance, Department of Investment and Financial Markets
katarzyna.perez@ue.poznan.pl
https://orcid.org/0000-0003-3331-8456

Received: 8 November 2021 / Revised: 17 December 2021 / Accepted: 18 December 2021 / Published online: 30 December 2021

ABSTRACT

This paper demonstrates how the Sharpe Ratio can be modified by altering the measure of "total risk" in the denominator of the Sharpe Ratio (i.e., the standard deviation) to include liquidity risk, a major risk for investors in hedge funds that is missing from the standard Sharpe Ratio formulation. We refer to our liquidity-risk-adjusted performance ratio as the LRAPR. The results of our analysis of 1186 hedge funds alive in 2012–2020 show that funds with higher liquidity risk exhibit higher Sharpe Ratios and higher Alphas (as estimated in a 7-factor model that does not incorporate liquidity risk). We posit that analysts and investors should not necessarily take these higher Sharpe Ratios and higher Alphas as indications of fund superiority; what appears to be superior manager skill may rather be a compensation for bearing liquidity risk. Our LRAPR is a tool that analysts or investors could use to compare funds on a more equal footing, adjusting for differential liquidity risk across funds.

JEL Classification: G12; G23; C18

Keywords: liquidity risk, liquidity risk factor, serial correlation, Sharpe ratio, hedge fund performance.

1. INTRODUCTION

In finance, standard deviation is referred to as a measure of "total risk" in that it incorporates both the systematic risk and unique risk of an investment. To us, the moniker "total risk measure" is a bit of a misnomer, in that standard deviation does not capture liquidity risk. Previous researchers have recognized this deficiency, especially in case of measuring investment fund performance

[1] Corresponding author: Katarzyna Perez – Institute of Finance, Department of Investment and Financial Markets, Al. Niepodległości 10, 61-875 Poznań; e-mail: katarzyna.perez@ue.poznan.pl; tel. +48 603 776 587

DOI: 10.7172/2353-6845.jbfe.2021.2.5

by the Sharpe ratio (Sharpe, 1966, 1994). They have pointed out that standard deviation may be understated and the Sharpe ratio therefore overstated in the presence of liquidity risk in a fund's portfolio. Lo (2002) has suggested methods for adjusting the Sharpe ratio to take liquidity risk into account, and Getmansky et al. (2004) have followed Lo (2002) in this by investigating hedge funds, for which liquidity is particularly important risk factor (C. Li et al., 2020; Siegmann & Stefanova, 2017).

In this study we propose an intuitive and simple modification to the Sharpe ratio that introduces into the calculation a proxy for liquidity risk, which is not directly captured in the standard deviation measure. This proxy is a serial correlation of fund returns. Our modification method results in a measure that we call a "liquidity-risk-adjusted performance ratio" (LRAPR). We calculate LRAPR for 1186 hedge funds alive in 2012–2020. Intuitively this measure should allow a better apples-to-apples comparison of funds exhibiting higher assumed liquidity risks with other funds where liquidity risk is absent or more muted. We find that in a certain group of higher-liquidity-risk funds in our fund universe, there is a strongly positive association between the level of liquidity risk and the fund Sharpe ratio. We also find that our LRAPR may be a more useful reward-for-risk measure than the Sharpe ratio in that the LRAPR seems to be independent of the fund's liquidity risk, so that differences in LRAPR across the funds may depend on more relevant forces, such as differential manager skill or exposure to more unusual risks that are harder to identify, measure, and intentionally incorporate into a diversified portfolio. To confirm our intuition, we calculate Alphas for all the hedge funds in the database using a 7-factor equilibrium model similar to that of Fung & Hsieh (2004). We find that higher-liquidity-risk funds have higher estimated Alphas in the model. This suggests that the extra Alpha for this group of funds may be a compensation for an eighth risk factor that is missing from the 7-factor model, namely, a liquidity risk factor.

We have organized our study as follows: In Section 2, we review the literature; in Section 3, we present the data and describe our methods, including our model and our liquidity risk factor; in Section 4, we present our findings; and in Section 5, we conclude.

2. LITERATURE REVIEW

There is a robust and historically important literature related to performance evaluation and measurement of hedge funds. Among many types of fund performance measures, two are the most commonly utilized. The first one is the Alpha coefficient proposed by Jensen (1969), developed by Fama & French (1993) and adjusted to the hedge fund industry by Fung and Hsieh (2004). Alpha measures the relationship between a fund return and a set of undiversifiable risk factors that influence this return. Because of its properties, this measure is a standard in academic research on hedge funds (e.g. Barras et al., 2010; Chen et al., 2016; Fung & Hsieh, 2001; Kosowski et al., 2007; Stulz, 2007). The second measure of particular popularity in practice is the reward-to-variability ratio of Sharpe (1966). The Sharpe ratio measures the relationship between the mean and the standard deviation of excess returns. It is one of the best-known and widely used metrics to measure and compare investment performance (Auer & Schuhmacher, 2013). For a long time it was treated as not appropriate for hedge funds because theoretically it was justified to be utilized in the case of normally distributed excess returns or quadratic investor preferences (Brooks & Kat, 2002; Mahdavi, 2004; Zakamouline & Koekebakker, 2009). More recent findings of Schuhmacher & Eling (2011, 2012) prove that the Sharpe ratio has a decision theoretic foundation even in the case of asymmetric or fat-tailed excess returns and thus it is applicable for the performance analysis of the hedge funds. Auer and Schuhmacher (2013) expand the analysis of the statistical properties of the Sharpe ratio and propose adequate testing that strengthen the theoretical plausibility of the Sharpe ratio as a hedge fund performance measure.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Sharpe ratio is a simple fund performance measure that can be easily interpreted by an ordinary investor. This is one of the reasons why it is widely and commonly used in practice to publish rankings of funds according to their performance. Such rankings serve then as a tool of investment advisors who help hedge fund clients to make investment decisions (Liang, 1999). However, the original version of the Sharpe ratio does not capture individual types of risks, especially liquidity risk that is of particular importance to hedge funds. Hedge funds are meaningful investors in markets of illiquid assets where they bear illiquidity risk as a major source of return, that becomes an "illiquidity premium." This premium is investigated by researchers in two levels. In the present research we are focused on what we might call "micro level" liquidity risk at the level of the fund and the fund portfolio, that is, asset liquidity. Other sources of liquidity risk are "macro level" risks, and we have indeed seen such risks impact the liquidity of hedge funds at times of the market stress. The Global Financial Crisis comes to mind, of course, as does the "double whammy" in August 1998 of the Russian ruble crisis and the Long Term Capital Management disaster. These macro-level liquidity shocks tend to affect all hedge funds more or less at the same time.

The literature on liquidity risk at the "macro level" focuses on the effect of systemic liquidity shocks (that is, aggregate market-wide liquidity risk as an undiversifiable risk factor) on market microstructure aspects such as bid/offer spreads, trading volume, and price impact, as well as changes in funding. Several studies have found that systemic liquidity risk accounts for or explains a significant portion of fund Alpha (e.g., Gibson et al. (2013)) or that the large losses experienced during global liquidity shocks counteract the generally good performance of illiquid funds in calmer markets (Sadka (2010) and Sadka (2012)). Previously, Pastor and Stambaugh (2003) employ proxies for system liquidity risk, such as bid/offer spreads and trading volume for stocks, to rank and sort stocks to create a no-investment, long-short liquidity risk factor. They find that the aggregate liquidity risk measure helped to explain the cross-section of stock returns. Billio et al. (2011) find that some hedge fund strategies perform particularly poorly during bouts of financial distress due to funding problems and illiquid assets.

Another focus of the literature connected to the liquidity risk at the "macro level" is on well-organized markets for the underlying securities, such as stock exchanges or other exchanges where bids and offers are posted and where the price data is transparent and trading volumes are available. In such markets the researchers look for liquidity timing ability of hedge fund managers and they find it: Cao et al. (2013) on the equity market, B. Li et al. (2017) on the fixed-income market and Luo et al. (2017) and Ch. Li et al. (2020) on the foreign exchange market. Yet many of the hedge fund strategies deal in securities that do not trade on those markets. And virtually no such hedge fund discloses its portfolio holdings in enough detail to enable an analyst to assess the liquidity of the fund through examination of its holdings. One of the possibilities to assess liquidity risk at more "micro level", at the level of the individual fund, is the model proposed by Lo (2002) who focuses on the Sharpe ratio. He finds the volatility of holding period returns, which serves as the risk measure, to be understated in the case of illiquid portfolio holdings. Lo (2002) establishes a fund's correlation to its own one-month lagged returns as a proxy for liquidity risk and uses this approach as the basis for restating the Sharpe ratio. Lo (2002) adjusts the Sharpe ratio for liquidity risk through the process of annualizing Sharpe ratios typically calculated on the basis of monthly fund data. He recognizes that the usual method of multiplying by $\sqrt{12}$ to annualize data based upon monthly returns is not appropriate in situations where the returns are non-IID, serially correlated returns being one example of returns that violate the assumption of IID. The Lo (2002) factors essentially reduce traditionally calculated Sharpe ratios for funds with positive serial correlation and increase Sharpe ratios for funds with negative serial correlation.

Getmansky et al. (2004) extend Lo (2002), relying on the regression coefficient in an AR(1) serial correlation model to serve as a proxy for a fund's liquidity risk. They point out that returns should be serially uncorrelated in an informationally efficient market. The presence of serially auto-correlated returns certain funds, then, they take as an implication of market inefficiency

DOI: 10.7172/2353-6845.jbfe.2021.2.5

and illiquidity. Getmansky et al. (2004) find that the presence of serial correlation in hedge fund returns can be caused by three factors: 1. the fund investment strategy and the nature of assets in the fund; 2. the method of month-end pricing; and 3. deliberate "smoothing" of returns by a fund manager. These factors bear directly on the liquidity of the fund's underlying assets and therefore of the fund itself. Generally large cap equity funds should have low levels of serial correlation, because they are liquid and easy to price so the temptation of a fund manager to "smooth" his returns in that type of hedge funds is small. However, the strategies of small cap equity, distressed debt, PIPES or fixed income arbitrage tend to have high serial correlation. Therefore they have a greater risk of dislocation and a large negative performance surprise. For these types of funds, the standard deviation may understate the actual risk, and the Sharpe ratio may overstate the reward-for-risk tradeoff. To the extent that analysts rely heavily on the Sharpe ratio measure in these circumstances in their investment decision making, they may overestimate the diversification benefit of including such a fund in their portfolios.

Khandani and Lo (2011) apply the analysis of Getmansky et al. (2004) to both hedge funds and mutual funds, as well as artificially created portfolios of stocks. They rank and sort these various portfolios by the autocorrelation of monthly returns to create a no-investment, long-short liquidity risk factor. They find a positive relationship between elevated levels of serial correlation and funds with longer redemption periods (i.e, longer lock ups) as well as funds with investment strategies known to be more illiquid (such as small-cap stocks, emerging market stocks, and mortgage-backed bonds).

Our interest in this research is the fund liquidity risk at the "micro level", where different funds are exposed to different levels of such risk at different times. We follow Lo (2002) and Getmansky et al. (2004) and use serial correlation of returns as a proxy of the liquidity risk and propose the liquidity risk adjusted performance ratio (LRAPR). By doing that we add a brick to the discussion on adjusting Sharpe ratio with the liquidity risk. Because of simplicity our measure could be easily used to rank hedge funds by the financial information systems and investment advisors who – then – could present them to the hedge fund investors.

3. DATA AND METHODOLOGY

We use the BarclayHedge database of "living" funds, with monthly returns through December 2020. Our first step is to eliminate duplicate funds[2]. To do this, we took the monthly returns across all the funds in the database for the period July 2015 to June 2020, and we calculated the pairwise correlation coefficients. For any $R>0.95$, we deemed this to be a duplicate fund. Per this procedure, we reduced the fund database from 5,000+ to 2,133 funds. Next, we decided on an analysis period of January 2012 through December 2020 as the period of the dynamic world capital market growth where the number on long "living" hedge funds was high and the reliable data on them available. We found there to be 1,186 funds for the January 2012 through December 2020 time period. For each fund for the nine-year period, we calculated the following metrics, using monthly holding period return data: CAGR; Standard Deviation; Risk-free rate; Sharpe Ratio[3]; AR(1) Beta – a one-period-lagged serial correlation measure; t-stat and p-value for the AR(1) Beta; and LRAPR – our liquidity-risk-adjusted performance ratio.

We follow Lo (2002) and Getmansky et al. (2004) and use serial correlation of fund returns as the proxy for liquidity risk. Lo (2002) measures serial correlation with the correlation coefficient.

[2] A duplicate fund is a fund that is an individual legal entity but it shares the same manager, the same strategy, and either the same or a very similar portfolio with another fund.

[3] We took the returns for the US equity market benchmark index as well as the risk-free-rate data from the data library of David French (of Fama-French fame) found at the website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Getmansky et al. (2004) do that with the regression-estimated slope coefficient β_{iT} of a simple one period lagged autocorrelation AR(1) process model:

$$r_{it} = \propto_{iT} + \beta_{iT} r_{it-1} + e_{it}, \qquad \text{Eq. (1)}$$

where r_{it} – is a return of a fund i in time t, r_{it-1} is – is a return of a fund i in time t–1, $\propto_{iT}$, β_{iT} are coefficients and e_{it} is a standard error of Eq. (1). Both measures of serial correlation are essentially the same[4]. We modify the standard Sharpe ratio measure such that we divide the standard deviation in the denominator by 1 minus the regression-estimated coefficient for the AR(1) process serial correlation for a single fund. Our "liquidity-risk-adjusted performance ratio" (LRAPR) is calculated as follows:

$$LRAPR = \frac{R_i - R_f}{\sigma_i / \left(1 - \beta_{AR(1)i}\right)}, \qquad \text{Eq. (2)}$$

where $\beta_{AR(1)i}$ is the coefficient β_{iT} from Eq. (1). Further in this article we refer to it as the AR(1) Beta. Our method for adjusting the Sharpe ratio is simple, accessible to the analyst and easy to deploy for practical use. To give an example, consider two hedge funds from our universe for the period 2012 through 2020 (see Table 1).

Table 1
Example of Changed Rank Ordering of Funds under LRAPR versus standard Sharpe Ratio

Fund	Annual Return	Risk free rate	Ann Std Dev	Sharpe Ratio	AR(1) Beta	LRAPR
Hedge Fund A	11.85%	0.65%	8.67%	**1.292**	.257	**.961**
Hedge Fund B	7.31%	0.65%	4.94%	**1.348**	.319	**.915**

Source: own calculations.

Of note in this example is that the ordering of the funds has changed after the inclusion of the heightened liquidity risk in the reward-for-risk formulation. The fund with the higher Sharpe ratio actually records a lower LRAPR; this is due to its higher serial correlation and hence higher likely liquidity risk. When adjusted for the heightened possible liquidity risk, the reward-for-risk measures for both funds are re-stated at lower levels, and the rank ordering is reversed. We investigate whether this is the case in general by comparing the LRA performance ratio with the standard Sharpe ratio and with the Lo (2002) adjusted Sharpe ratio for each individual fund in the whole universe of our 1,186 hedge funds for the 9-year period 2012 through 2020.

Next, we ask a question: "Does a fund's serial correlation that is a proxy for liquidity risk influence a fund's Sharpe ratio?" Or stated somewhat differently, "Does a high Sharpe ratio evidence the fund's superior performance relative to other funds, or is it reflective of compensation for bearing liquidity risk?" We regress in turn the fund LRAPR, standard Sharpe ratio and Lo (2002) adjusted Sharpe ratio on the fund AR(1)Beta to address the question. The regression outputs are presented for the 142 high AR(1)Beta funds and for the 1,186 fund universe as well as for the funds with positive and negative serial correlation – we do the latter in order to confirm our intuition that adjustment to the Sharpe ratio is best confined to funds

[4] The formula for Beta is $Beta = \frac{cov(r_x, r_y)}{\sigma_x * \sigma_x}$, and the formula for correlation is $R = \frac{cov(r_x, r_y)}{\sigma_x * \sigma_x}$. The only difference is in the denominator, replacing σ_x with σ_y. Since x and y are essentially the same data (lagged one month, so that n–1 of the n data points are in common), the standard deviation of x is very close to the standard deviation of y, in cases where neither the first and nor last months of the returns time series is an extreme return.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

with positive serial correlation, and that negative serial correlation funds harbor no particular liquidity risk.

Finally, we examine the association between the AR(1) Beta and a hedge fund Alpha. We are motivated by our findings that Sharpe ratios are overestimated due to exclusion of liquidity risk from the Sharpe ratio risk measure. We hypothesize that a fund Alpha is similarly overstated in the presence of heightened liquidity risk in the case where the Alpha is estimated in an equilibrium model that excludes liquidity risk as an identified and modeled risk factor – which describes just about every equilibrium model in use in hedge fund performance evaluation. We employ our own 7-factor equilibrium model – where liquidity risk is not among the risk factors – to estimate a fund Alpha for all of the 1,186 funds in the universe. We regressed the resulting fund Alpha on fund AR(1) Beta, to assess whether fund Alpha is driven by liquidity risk and is overstated in the presence of liquidity risk. Our 7-factor model is based on the Carhart 4-factor model (Carhart, 1997) and the extensive work of Fung and Hsieh over the many years, that resulted in their identification of useful risk factors (Fung & Hsieh, 2007). Our seven factors are: market risk, size, value, momentum, interest rate risk, credit spread risk, and emerging market equity risk, as represented in this model:

$$\begin{gathered} r_{it} - rf_t = \propto_{iT} + \beta_{1iT}(RMRF)_t + \beta_{2iT}(SMB)_t + \beta_{3iT}(HML)_t + \beta_{4iT}(MOM)_t + \\ + \beta_{5iT}(10yrUTS)_t + \beta_{6iT}(10yrUTS - Baa\ Bonds)_t + \beta_{7iT}(EM\ Equity - R_f)_t + e_{it} \end{gathered} \quad \text{Eq. (2)}$$

The results of all the steps of our research are presented in the next Section.

4. RESULTS

The average values of the standard Sharpe ratios, our LRA performance ratio and the Lo (2002) adjusted Sharpe ratio for the hedge funds in our universe are presented in Table 2. First, we focus on the 142 hedge funds where the AR(1) serial correlation coefficient is positive and significant at the 95% confidence level. The measures of AR(1)Beta for these 142 funds range from a low of 0.188 to a high of 0.878. Looking at the averages for those 142 hedge funds with higher assumed liquidity risk, we see a few noteworthy aspects. First, the hedge funds that have significant serial correlation on average have much higher Sharpe ratios than the average Sharpe ratio across the 1,186 fund universe (1.256 versus 0.699). Second, the 142 hedge funds with higher assumed levels of liquidity risk surrender much more of their Sharpe ratio in descending toward the LRA performance ratio, dropping 36,1% from 1.256 to 0.803; the average fund in the 1,186 fund universe, surrendered 7.7% of its Sharpe ratio in the process of incorporating liquidity risk into the reward-for-performance measure, from 0.699 to 0.645. Third, our LRA performance ratio gives about the same result as the Lo (2002) method, while being simpler to calculate and being easier to understand intuitively.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Table 2
Comparison of Sharpe Ratios and LRAPRs

	1	2	3	4	5	6
9-year period 2012–2020	Average AR(1) serial correlation coefficient	Average Sharpe Ratio value	Average LRAPR value	% decrease from Sharpe Ratio to LRAPR (2 – 3)	Average Lo (2002)-adjusted Sharpe ratio value	% decrease from Sharpe Ratio to Lo (2002)-adjusted (2 – 5)
142 Hedge Funds with positive AR(1)	0.284	1.256	0.803	-36.1%	0.904	-28.1%
1,186 Hedge Funds	0.048	0.699	0.645	-7.7%	0.652	-6.7%

Source: own calculations.

Additionally, whereas the Lo (2002) method allows for an increase in Sharpe ratio for a fund with negative one-period lagged correlation, we do not credit negative AR(1) funds with a higher Sharpe ratio simply for the fact that the fund's AR(1) measure is even lower than a level that already indicates "little or no" liquidity risk. There are interesting phenomena with negative serial correlation funds that merit attention. The AR(1) for the SP500 varies over time, but is sometimes in the range of about 0.10 to 0.15 for long periods of time. While in times of crisis there may be short-lived bouts of poor liquidity or illiquidity in the US large-cap equity market, in general, we would argue that we can classify the US large-cap equity market as "liquid" – and it is certainly liquid when considered in relation to managed funds, such as hedge funds. So, if a hedge fund has an AR(1) Beta measure that is on par with or smaller than the AR(1) Beta measure for the overall US large-cap equity market, we would not suggest that we can infer from the fund's serial correlation measure that fund harbors particular liquidity risk. So, if funds with AR(1) measures of serial correlation of 0.10 to 0.15 can be characterized as "little or no liquidity risk" funds, then what to make of funds with AR(1) measures of zero or -0.10 or -0.25. Is there such a thing as less liquidity risk than zero liquidity risk? We think not. We can consider funds with negative, near-zero, or only very modest levels of positive serial correlation to be funds characterized by mean reverting returns. Funds with negative measures for AR(1) tend to be in areas such as managed futures, global macro trading, etc., with high turnover and high exposure to high-liquidity non-equity and non-bond securities such as FX, futures, and other derivatives. Therefore, we make no adjustment to the Sharpe ratio in the case of funds where the serial correlation is negative.

The results for regressing the fund Sharpe ratio, LRAPR, Lo (2002) Sharpe ratio as well as 7-factor Alpha on the fund AR(1)Beta are summarized in Table 3.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Table 3
Results for Regressing Sharpe Ratio, LRAPR, 7-factor fund Alpha and Lo 2002 Sharpe Ratio on Fund AR(1)Beta

	Dependent Variable	Independent Variable	Universe	R^2	Regression Coefficient (loading)	p-value	Intercept	p-value
1	Sharpe Ratio	AR(1) Beta	142 high liquidity risk funds	.11	8.69	.0000	-1.21	.0462
2		AR(1) Beta	Entire universe 1,186 funds	.04	1.80	.0000	.61	.0000
3		AR(1) Beta	758 AR(1) Beta > 0 funds	.07	3.41	.0000	.35	.0000
4		AR(1) Beta	428 AR(1) Beta < 0 funds	.00	-.31	.4177	.59	.0000
5	LRAPR	AR(1) Beta	142 high liquidity risk funds	.01	1.44	.2389	.40	.2843
6		AR(1) Beta	Entire universe 1,186 funds	.00	.39	.0394	.63	.0000
7		AR(1) Beta	758 AR(1) Beta > 0 funds	.00	.67	.0490	.58	.0000
8		AR(1) Beta	428 AR(1) Beta < 0 funds	.00	-.31	.4177	.59	.0000
9	Lo (2002) Sharpe Ratio	AR(1) Beta	142 high liquidity risk funds	.04	3.35	.0148	-.05	.8969
10		AR(1) Beta	Entire universe 1,186 funds	.00	.47	.0186	.63	.0000
11		AR(1) Beta	758 AR(1) Beta > 0 funds	.02	1.33	.0002	.49	.0000
12		AR(1) Beta	428 AR(1) Beta < 0 funds	.01	-.99	.0172	.59	.0000
13	7-Factor Fund Alpha	AR(1) Beta	142 high liquidity risk funds	.03	.68	.0476	.14	.1731
14		AR(1) Beta	Entire universe 1,186 funds	.05	.76	.0000	.06	.0000
15		AR(1) Beta	758 AR(1) Beta > 0 funds	.06	.99	.0000	.03	.1606
16		AR(1) Beta	428 AR(1) Beta < 0 funds	.01	-.67	.0586	-.04	.0000

Source: own calculations.

These output suggest several findings. First (Table 3, row 1) for the 142 high-liquidity-risk funds in the universe, an increase in AR(1)Beta of, for instance, 0.10 (from, say, 0.45 to 0.55) is associated with an increase in Sharpe ratio of 0.86. And across the entire universe of 1,186 funds (Table 3, row 2), an increase in AR(1)Beta of 0.10 is associated with a 0.18 increase in Sharpe ratio. These findings are consistent with our intuition that standard deviation as a measure of "total risk" – and as the risk measure used in the standard Sharpe ratio – does not capture liquidity risk in that Sharpe ratios are sensitive to and positively related to increases in assumed fund liquidity risk.

Further, we performed similar regressions dividing the universe into two groups: positive serial correlation funds (Table 3, row 3) and negative serial correlation funds (Table 3, row 4). The results seem to confirm our intuition. For negative serial correlation funds, there is no particular association of the Sharpe ratio with the liquidity risk, as indicated by the high p-value for the slope coefficient and by the high level of Significance F for the model overall. As for the positive serial correlation funds (over 80% of which have AR(1) Betas that are not statistically significant at the 95% confidence level), a 0.10 increase in serial correlation is associated with a 0.34 increase in Sharpe ratio. And the model overall is significant.

Our expectation is that the modification to the Sharpe ratios that we performed when we transformed the fund Sharpe ratios into fund LRAPRs will result in a better measure of reward-for-risk than the original Sharpe ratios, which are overstated due to liquidity risk being absent from the Sharpe ratio risk measure. One area in which we can test if the LRAPR is an improvement

DOI: 10.7172/2353-6845.jbfe.2021.2.5

on the Sharpe ratio is to investigate whether or not the LRAPR is systematically and positively associated with liquidity risk, as the Sharpe ratio seems to be. If the LRAPR has successfully adjusted for fund liquidity risk, then we would expect to see little evidence (or at least less evidence) of a positive association between the LRAPR and liquidity risk. We also perform a similar analysis for the Lo (2002) modified Sharpe ratios. Repeating the foregoing regression analysis, substituting LRAPR (and then the Lo 2002 Sharpe ratios) for the Sharpe ratio we get the results presented in Table 3, rows 5 and 9.

For the 142 high liquidity risk funds, the estimated slope coefficient for the LRAPR regression has a p-value of 0.24 (lacking statistical significance), indicating that the LRAPR provides a reward-for-risk measure for hedge funds that accounts for differential liquidity risk as proxied by the serial correlation AR(1) Beta. In the case of the Lo (2002) Sharpe ratios, the estimated slope coefficient is statistically significant (p-value of 0.01), and a loading of 3.35, implying an increase in Lo (2002) Sharpe ratio of 0.33 for every 0.10 increase in AR(1) Beta. While this sensitivity to liquidity is less than the 8.69 slope coefficient in row 1 for the Sharpe ratio correlation, it still shows that variations in liquidity risk across the group of 142 funds explains a good portion of the differences in Lo (2002) Sharpe ratios.

Again, dividing the funds into two groups (positive and negative serial correlation) also shows that the LRAPR seems to be an improvement over the standard Sharpe Ratio, as well as over the Lo (2002) Sharpe ratios, in removing any dependence in the differences in performance ratios among the funds to the level of liquidity risk in those funds. Adding the Lo (2002) Sharpe ratios to this discussion (Table 3, rows 11 and 12), we can see that for the AR(1) Beta > 0 funds the Lo (2002) Sharpe ratios are statistically significantly and positively related to the level of liquidity risk in the funds' portfolios, with a 0.13 increase in Sharpe ratio associated with a 0.10 increase in AR(1) Beta.

As for the AR(1) Beta for the AR(1) Beta < 0 funds, as we have discussed above, we view such funds to be funds with mean reverting yields, not funds with assets that trade "sticky" and may be illiquid. These are likely funds with very liquid portfolios, in strategies such as global macro, managed futures, and the like. By contrast, Lo (2002) adjusts the Sharpe ratio upwards for these AR(1) Beta < 0 funds, resulting in the statistically significant and negative relationship between Lo (2002) Sharpe ratio and AR(1) Beta for this cohort. Finally, we examine the association between the AR(1) Beta and a hedge fund 7-factor Alpha. The finding for the 142 high liquidity risk funds indicates that at a 95% confidence level there seems to be a positive relationship between fund Alpha and the AR(1) Beta measure of fund liquidity risk. More specifically, we can interpret the relationship as follows: the regression slope shows that an increase in monthly Alpha of 0.6799% (or an annual Alpha of 8.16%) is associated with a 1.0 increase in serial correlation. On a more reasonable scale, we can say that a 0.10 increase in serial correlation is associated with a 0.82% increase in estimated annual Alpha in the 7-factor model. This seems to confirm our intuition that a significant portion of fund Alpha is likely compensation for bearing liquidity risk. For the entire universe of 1,186 hedge funds, the relationship is a statistically significant 0.92% increase in annual Alpha for an increase in serial correlation of 0.10.

Again, applying the analysis to positive and negative serial correlation funds as distinct groups gives the following results (Table 3 rows 15 and 16). For funds with positive serial correlation, and some assumed level of liquidity risk, it seems that an increase in AR(1) Beta of 0.10 is associated with an increase in annual Alpha of 1.19%. Stated differently, accounting for liquidity risk in fund performance evaluation may lead the analyst to reduce the evaluation of annual fund Alpha by 1.19% per every 0.10 of serial correlation (above some threshold that the analyst will have to decide upon) identified in the fund returns. Interestingly, again the negative serial correlation funds seem to exhibit a distinct lack of liquidity risk. In fact, the negative slope of -0.6710 seems to imply that funds that are strongly mean reverting (and this could include hedge

DOI: 10.7172/2353-6845.jbfe.2021.2.5

fund strategies such as global macro, managed futures, and the like) are adding value at the rate of 0.81% per annum for every increase of 0.10 of serial correlation in the negative direction.

We conclude this section on research findings with a summary comparison of the Sharpe ratios, Lo (2002) Sharpe ratios, and LRAPRs for the funds in our universe. We calculated each performance measure for each fund in the universe, and we present the findings aggregated by AR(1) Beta decile, in Table 4:

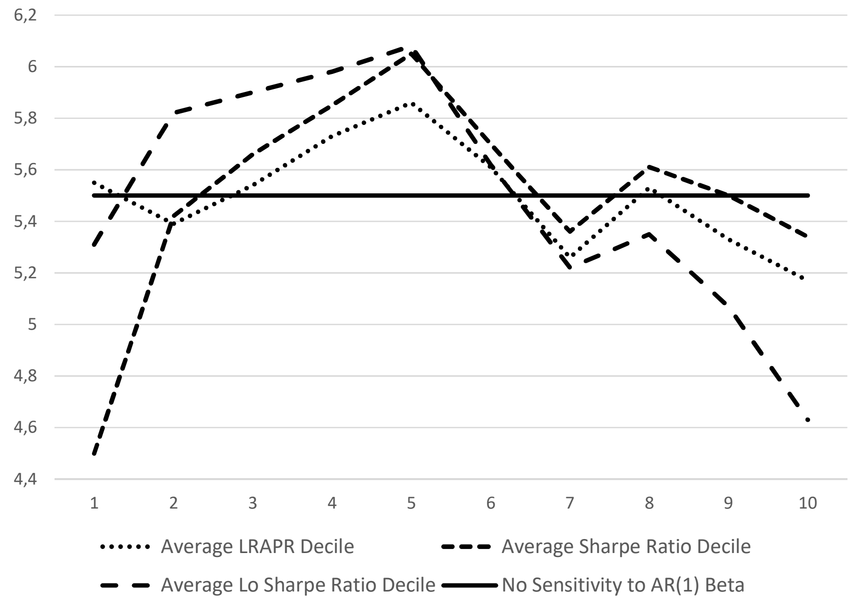
Table 4
Sharpe ratios, Lo (2002) Sharpe ratios and LRAPR by AR(1)Beta Decile

AR(1) Beta Decile	Average AR(1) Beta	Average Sharpe Ratio Decile	Average Sharpe Ratio	Average Lo 2002 Sharpe ratio Decile	Average Lo 2002 Sharpe ratio	Average LRAPR Decile	Average LRAPR
1	.30	4.50	1.34	5.31	.94	5.55	.83
2	.17	5.42	.68	5.82	.58	5.39	.65
3	.11	5.66	.61	5.90	.55	5.54	.61
4	.08	5.85	.59	5.98	.55	5.73	.59
5	.05	6.05	.55	6.08	.52	5.86	.55
6	.02	5.70	.59	5.62	.58	5.61	.59
7	.00	5.36	.77	5.22	.77	5.26	.77
8	-.03	5.61	.62	5.35	.64	5.53	.62
9	-.07	5.50	.59	5.07	.63	5.33	.59
10	-.14	5.34	.65	4.63	.74	5.17	.65
Column Average		5.50		5.50		5.50	
Average of all Funds	.05		.70		.65		.64
Standard Deviation		.42		.46	.	.21	

We placed the deciles sorted by AR(1) Beta along the X-axis. Each decile contains 118 or 119 of the total 1,186 funds in the universe. The first decile, on the left side of the graph, contains the 119 funds with the largest estimates of AR(1) Beta. The tenth decile, on the right, contains the 118 funds with the lowest estimates of AR(1) Beta. For each decile by AR(1) Beta, we calculated three values. As an example, consider the short dashed line with a value of 4.5 for AR(1) Beta decile 1: the 119 funds on the top AR(1) Beta decile, on average, reside the middle of the fourth decile when we sorted the 1,186 funds by Sharpe ratio and placed them into deciles by Sharpe ratio. If there were no relationship at all between AR(1) Beta and the Sharpe ratio – that is a merely random association – then we would expect the Sharpe ratio decile average to be 5.50 in every AR(1) Beta decile. Displaying the decile averages graphically, we find what is seen in Graph 1.

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Graph 1
Average Performance Ratio Decile for Three Different Performance Measures, across the Ten Deciles by AR(1) Beta



Source: own calculation.

The solid horizontal line at the vertical height of 5.50 represents no sensitivity to AR(1) Beta. We can see that the Sharpe Ratio and the Lo (2002) Sharpe ratio performance measures display some sensitivity to the AR(1) Beta. As shown in Table 4, the standard deviation of the decile averages are 0.42 and 0.46, respectively. For the LRAPR, we can see from the chart that the sensitivity to the liquidity risk proxy measure is much less, measured as 0.21 by standard deviation. The regression results in Table 3 tells a similar story. This asserts our belief that the LRAPR may provide the investor or analyst with a better apples-to-apples comparison of risk-reward performance across funds, in that the LRAPR does a better job than original Sharpe ratio of adjusting for varying liquidity risk among funds.

5. CONCLUSIONS

Ideally, a reward-for-risk measure should include all relevant risks, so that the risk is not understated and so that the reward-for-risk is not overstated. The original Sharpe Ratio relies on standard deviation as a measure of "total risk," which measure does not incorporate potential liquidity risk at the fund level. We have borrowed from the literature a proxy measure for potential liquidity risk for hedge funds, and then applied and extended that in the direction of modifying the Sharpe Ratio to create a new liquidity-risk-adjusted performance ratio (LRAPR). Others have dealt with this topic before us, notably Lo (2002). We strived to improve on Lo (2002)'s formulation of a solution by applying it just to funds with liquidity risk (not all funds in the universe) and by proposing a simplified calculation method that yields similarly modified results when compared to the Sharpe ratio, but that is accessible to and implementable nowadays by just about any analyst or investor.

We find a positive and significant relationship between liquidity risk and Sharpe Ratio among funds with statistically significant levels of liquidity risk as measured in an AR(1) process. This indicates that funds with higher liquidity risk are rewarded with higher Sharpe Ratios, even though that higher reward may be the result of compensation for bearing liquidity risk and not from some other source, such as manager skill or some other factor. In contrast to this, we find no significant relationship between our LRAPR and liquidity risk, indicating that we might use

DOI: 10.7172/2353-6845.jbfe.2021.2.5

our LRAPR as a reward-for-risk measure that incorporates or corrects for differences in liquidity across funds.

Further, we find that funds with higher likely liquidity risk exhibit higher levels of fund Alpha as estimated in a 7-factor model that does not account for liquidity risk. This would seem to imply that some of the fund Alpha of higher liquidity risk funds is probably a compensation for bearing liquidity risk rather than returns due to the manager skill or some other source of return. Therefore, all the more important it is for the analyst or investor to have a method to take fund liquidity risk into account and to be able to evaluate fund performance on a more comparable basis. We are hopeful that analysts and investors will find our LRAPR to be accessible, implementable, and helpful in fund analysis: hedge funds, mutual funds or any other type of portfolios that are managed locally or globally but may be exposed to the liquidity risk.

References

Auer, B. R., & Schuhmacher, F. (2013). Performance hypothesis testing with the Sharpe ratio: The case of hedge funds. Finance Research Letters, 10(4), 196–208. https://doi.org/10.1016/j.frl.2013.08.001

Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. In Journal of Finance (Vol. 65, Issue 1). https://doi.org/10.1111/j.1540-6261.2009.01527.x

Billio, M., Getmansky, M., & Pelizzon, L. (2011). Crises and Hedge Fund Risk. UMASS-Amherst Working Paper; Yale ICF Working Paper No. 07-14. https://doi.org/10.2139/ssrn.1130742

Brandon, R. G., & Wang, S. (2013). Liquidity risk, return predictability, and hedge funds' performance: An empirical study. Journal of Financial and Quantitative Analysis, 48(1), 219–244. https://doi.org/10.1017/S0022109012000634

Brooks, C., & Kat, H. M. (2002). The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors. The Journal of Alternative Investments, 5(2), 26–44. https://doi.org/10.3905/JAI.2002.319053

Cao, C., Chen, Y., Liang, B., & Lo, A. W. (2013). Can hedge funds time market liquidity? Journal of Financial Economics, 109(2), 493–516. https://doi.org/10.1016/J.JFINECO.2013.03.009

Carhart, M. M. (1997). On persistence in mutual fund performance. Journal of Finance, 52(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chen, J., Wu, W., & Tindall, M. L. (2016). Hedge Fund Return Prediction and Fund Selection: A Machine-Learning Approach. Financial Industry Studies Department, Dallas Fed, November. https://www.dallasfed.org/banking/fis/~/media/documents/banking/occasional/1604.pdf

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fung, W., & Hsieh, D. A. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. Review of Financial Studies, 14(2), 313–341. https://doi.org/10.1093/rfs/14.2.313

Fung, W., & Hsieh, D. A. (2004). Hedge fund benchmarks: A risk-based approach. In Financial Analysts Journal (Vol. 60, Issue 5, pp. 65–80). CFA Institute. https://doi.org/10.2469/faj.v60.n5.2657

Fung, W., & Hsieh, D. A. (2007). Will Hedge Funds Regress Towards Index-Like Products? Journal of Investment Management, 5(2), 56–80. https://papers.ssrn.com/abstract=989612

Getmansky, M., Lo, A. W., & Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. Journal of Financial Economics, 74(3), 529–609. https://econpapers.repec.org/RePEc:eee:jfinec:v:74:y:2004:i:3:p:529-609

Jensen, M. C. (1969). Risk, The Pricing of Capital Assets, and The Evaluation of Investment Portfolios. The Journal of Business, 42(2), 167–247. https://doi.org/10.1086/295182

Khandani, A. E., & Lo, A. W. (2011). Illiquidity Premia in Asset Returns: An Empirical Analysis of Hedge Funds, Mutual Funds, and US Equity Portfolios. Quarterly Journal of Finance, 1(2), 205–264. https://doi.org/10.1142/S2010139211000080

Kosowski, R., Naik, N. Y., & Teo, M. (2007). Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. Journal of Financial Economics, 84(1), 229–264. https://doi.org/10.1016/j.jfineco.2005.12.009

Li, B., Luo, J., & Tee, K. H. (2017). The Market Liquidity Timing Skills of Debt-oriented Hedge Funds. European Financial Management, 23(1), 32–54. https://doi.org/10.1111/EUFM.12090

Li, C., Li, B., & Tee, K. H. (2020). Are hedge funds active market liquidity timers? International Review of Financial Analysis, 67, 101415. https://doi.org/10.1016/j.irfa.2019.101415

Liang, B. (1999). On the Performance of Hedge Funds. Financial Analysts Journal, 55(4), 72–85. https://doi.org/10.2469/faj.v55.n4.2287

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Lo, A. W. (2002). The Statistics of Sharpe Ratios. Financial Analysts Journal, 58(4), 36–52. https://doi.org/10.2469/faj.v58.n4.2453
Luo, J., Tee, K. H., & Li, B. (2017). Timing liquidity in the foreign exchange market: Did hedge funds do it? Journal of Multinational Financial Management, 40, 47–62. https://doi.org/10.1016/j.mulfin.2017.04.001
Mahdavi, M. (2004). Risk-Adjusted Return When Returns Are Not Normally Distributed. The Journal of Alternative Investments, 6(4), 47–57. https://doi.org/10.3905/JAI.2004.391063
Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111(3), 642–685. https://doi.org/10.1086/374184
Sadka, R. (2010). Liquidity risk and the cross-section of hedge-fund returns. Journal of Financial Economics, 98(1), 54–71. https://doi.org/10.1016/j.jfineco.2010.05.001
Sadka, R. (2012). Hedge-Fund Performance and Liquidity Risk. Journal of Investment Management, 10, 60–72. https://papers.ssrn.com/abstract=2072774
Schuhmacher, F., & Eling, M. (2011). Sufficient conditions for expected utility to imply drawdown-based performance rankings. Journal of Banking and Finance, 35(9), 2311–2318. https://doi.org/10.1016/j.jbankfin.2011.01.031
Schuhmacher, F., & Eling, M. (2012). A decision-theoretic foundation for reward-to-risk performance measures. Journal of Banking & Finance, 36(7), 2077–2082. https://doi.org/10.1016/J.JBANKFIN.2012.03.013
Sharpe, W. F. (1966). Mutual Fund Performance. The Journal of Business, 39(S1), 119. https://doi.org/10.1086/294846
Sharpe, W. F. (1994). The Sharpe Ratio. The Journal of Portfolio Management, 21(1), 49–58. https://doi.org/10.3905/jpm.1994.409501
Siegmann, A., & Stefanova, D. (2017). The evolving beta-liquidity relationship of hedge funds. Journal of Empirical Finance, 44, 286–303. https://doi.org/10.1016/j.jempfin.2017.04.002
Stulz, R. M. (2007). Hedge funds: Past, present, and future. Journal of Economic Perspectives, 21(2), 175–194. https://doi.org/10.1257/jep.21.2.175
Zakamouline, V., & Koekebakker, S. (2009). Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. Journal of Banking & Finance, 33(7), 1242–1254. https://doi.org/10.1016/J.JBANKFIN.2009.01.005

DOI: 10.7172/2353-6845.jbfe.2021.2.5

Conventionalists, Pioneers and Criminals Choosing Between a National Currency and a Global Currency

Guizhou Wang
Faculty of Science and Technology
University of Stavanger
guizhou.wang@uis.no

Kjell Hausken[1]
Faculty of Science and Technology
University of Stavanger
kjell.hausken@uis.no
https://orcid.org/0000-0001-7319-3876

Received: 27 September 2021 / Revised: 24 November 2021 / Accepted: 3 December 2021 / Published online: 30 December 2021

ABSTRACT

The article analyzes how conventionalists, pioneers and criminals choose between a national currency (e.g. a central bank digital currency) and a global currency (e.g. a cryptocurrency such as Bitcoin) that both have specific characteristics in an economy. Conventionalists favor what is traditional and historically common. They tend to prefer the national currency. Pioneers (early adopters) tend to break away from tradition, and criminals prefer not to get caught. They both tend to prefer the global currency. Each player has a Cobb-Douglas utility with one output elasticity for each of the two currencies, comprised of backing, convenience, confidentiality, transaction efficiency, financial stability, and security. The replicator equation is used to illustrate the evolution of the fractions of the three kinds of players through time, and how they choose among the two currencies. Each player's expected utility is inverse U-shaped in the volume fraction of transactions in each currency, skewed towards the national currency for conventionalists, and towards the global currency for pioneers and criminals. Conventionalists on the one hand typically compete against pioneers and criminals on the other hand. Fifteen parameter values are altered to illustrate sensitivity. For parameter values where conventionalists go extinct, pioneers and criminals compete directly with each other. Players choose volume fractions of each currency and which kind of player to be. Conventionalists go extinct when criminals gain more from criminal behavior, and when the parameter values in the conventionalists' expected utility are unfavorable, causing competition between pioneers and criminals.

JEL Classification: C60; E50

Keywords: Bitcoin, digital currencies, currency competition, money, evolution, replicator dynamics, cryptocurrencies, central bank digital currencies.

[1] Corresponding author: Kjell Hausken – Faculty of Science and Technology, University of Stavanger, 4036 Stavanger, Norway, kjell.hausken@uis.no, Tel.: +47 51831632, Fax: +47 51831550.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

1. INTRODUCTION

1.1. Background

This article considers a national currency operational within a country, and a global currency operational within the same country and also outside the country. We do not model the characteristics of more than one country, but do model the characteristics of the global currency assumed operational beyond the country under analysis. We require the two currencies to operate as media of exchange (means of payment). We do not specify whether the two currencies are non-digital or digital, paper currencies combined with physical coins, etc. The comparison of a national currency and a global currency has become more relevant with the emergence of digital currencies. At the time of writing this article most countries still allow paper currencies. For some countries most transactions are digital, conducted e.g. through debit and credit cards, electronic funds transfers, etc. We expect currencies to become increasingly digital in the future, to transform the financial system in ways that are still unclear, but with more competitors. Most central banks are in the process of launching CBDCs (central bank digital currencies), e.g. the People's Bank of China, the European Central Bank, the Bank of England, and the US Federal Reserve. The transformation is partly impacted by the emergence of blockchain technology and the cryptocurrency Bitcoin, with a genesis block mined[2] on January 3, 2009 at 18:15:05 UTC. Bitcoin is increasingly considered to have value (Kelleher, 2021). On November 22, 2021, 14,641 cryptocurrencies contribute to a marketcap of $2.5 trillion. Among these, 1,039 are coins (not tokens) which are our main interest in this article (coinmarket.com).

When the global currency is conceptualized as a cryptocurrency such as Bitcoin, which allows 5–7 transactions per second, we account for the presence of Layer 2 solutions for scaling such as the lightning network where transactions are faster, less costly and more readily confirmed (Frankenfield, 2021).[3] The lightning network introduces off-ledger transactions, and disintermediates central institutions such as banks. The off-ledger transactions are updated on the main blockchain on the base Layer 1 only when two parties open and close a payment channel on the lightning network. Two examples of Bitcoin payments on the lightning network are the El Salvador Chivo wallet, which on October 16, 2021 recorded 24,076 remittance requests, which added up to $3,069,761.05 in one day (Sarkar, 2021), and Twitter tipping applying various third party operators such as the Strike Bitcoin lightning wallet service (Rodriguez, 2021). El Salvador's acceptance of Bitcoin as legal tender, and Tesla's on-and-off acceptance of Bitcoin for car payments (Zainab Hussain & Balu, 2021) means that goods and services in principle can be priced in Bitcoin. Hence, to the extent the global currency is a cryptocurrency combined with a Layer 2 solution, the global currency functions as a medium of exchange and a unit of account. It may also function as a store of value and a standard of deferred payments, which are beyond the scope of this article.

A plethora of different kinds of digital currencies emerge, tentatively classified into CBDCs, cryptocurrencies, digital currencies issued by private companies such as Meta's Diem, which is a stablecoin, digital currencies issued by political jurisdictions such as Miami's MiamiCoin, etc. As digital currencies become more common, these can be expected to compete with each other and with non-digital currencies. Hence it becomes relevant to assess which factors affect the market share of each currency over time, the implications of different market shares, and which

[2] Mining is how new Bitcoins enter circulation and how transactions are confirmed by the network on the blockchain ledger. Bitcoins are awarded through mining to the first computer to solve mathematical problems to verify blocks of transactions, applying hardware and energy known as "proof of work" (Hong, 2021).

[3] The Bitcoin base Layer 1 requires "proof of work" to ensure decentralization, which costs energy. See Willms (2021) regarding energy consumption. Bitcoin mining enables locating stranded energy sources, favorable technology, politically favorable jurisdictions, and financially favorable circumstances; grows its network optimally, and operates optimally through space and time. Layer 2 usually does not require proof, which causes more centralization.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

kinds of users apply the various currencies. Each currency‘s market share may depend on various factors such as backing, convenience, confidentiality, transaction efficiency, financial stability, and security, as perceived by users, contributors, regulators, governments, etc., and as elaborated upon in this article.

Competition between currencies implies different market shares for the various currencies. The implications of changes in the shares of the various currencies, from an economic point of view, are that the various actors involved in the various currencies benefit differently and incur different costs depending on the success of each currency. Examples of actors are currency producers, users, borrowers, lenders, stakers, and miners.

For example, central banks and their associated governments can expect to benefit from the success of CBDCs. Users may benefit if the CBDC is stable with low transaction costs, but may experience a cost if they value privacy and all their transactions get centrally recorded. The success of a cryptocurrency such as Bitcoin can be expected to benefit libertarians and actors preferring decentralized currencies less controlled by central actors, and not to benefit middlemen such as banks and others enabling, facilitating and negotiating transactions. The success of Meta's Diem can be expected to benefit Meta's stakeholders and users. The success of Miami's MiamiCoin can be expected to benefit Miami.

1.2. Contribution

This article considers an economy with a national currency and a global currency. The national currency offers the most common usage, such as buying goods, paying taxes, etc. A global currency may offer more limited usage, e.g. for buying goods and paying taxes, but may offer other opportunities such as tax evasion, user autonomy, etc. Three kinds of players are assumed, i.e. conventionalists, pioneers, and criminals. These are believed, first, to represent all societal players and, second, to have different preferences for the national currency and a global currency. Conventionalists favor what is traditional and historically common, which is often the national currency. Pioneers (early adopters) tend to depart from tradition and search for new ways of transacting, which may involve a global currency. Criminals search for currencies ensuring that they do not get detected and caught, which may also involve a global currency. Conventionalists typically compete against pioneers and criminals. When conditions for conventionalists are unfavorable causing their extinction, pioneers and criminals compete more directly with each other. All the three kinds of players can in principle choose some degree of criminal behavior, but criminals are assumed to have preferences explicitly focused on criminal behavior. The three groups are assumed to be mutually exclusive and jointly exhaustive to represent all possible kinds of market participants. If a player is empirically determined to fall somewhere between two kinds of players, a choice has to be made one way or the other. A player can over time choose to change from being of one kind to being of another kind.

Each player has a Cobb-Douglas utility with one output elasticity for each of the two currencies, split into backing, convenience, confidentiality, transaction efficiency, financial stability, and security, as perceived by the player. Factors such as usability and technological potential are assumed present in most of these six subelasticities, perhaps especially in convenience and transaction efficiency.[4] These six subelasticities are assumed to comprise the main concerns relevant for each player's preferences regarding which of two currencies to choose. Each player makes two strategic simultaneous choices to maximize its expected utility which is shown to be inverse U-shaped in the volume fraction of transactions in each currency. The first choice is the volume fraction of its transactions in each currency. This choice depends on what kind of player the player is, but does not depend on how many players exist of this player's kind, and hence does

[4] A factor such as investment profitability is more relevant for the function of a cryptocurrency as a store of value rather than a medium of exchange and a unit of account.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

not depend on time. Each player's second choice is which kind of player it should be at each point in time. Hence this second choice depends on time, through replicator dynamics.

Applying replicator dynamics, the research questions are how the volume fractions of the two currencies and the fractions of the three kinds of players evolve through time, and are sensitive to various characteristics. A further research question is to determine society's expected utility to account for welfare at the societal level. Scenarios are illustrated where the output elasticities and other characteristics cause some of the three kinds of players to become dominant or inferior over time. For the stationary solution after sufficiently much time has elapsed, sensitivity analysis is conducted to show how the fractions of the three kinds of players depend on variation in parameter values relative to a benchmark. Applying credible specific functional forms, an exact analytical solution is produced for the fraction of each player's transactions in the national currency , and replicator dynamics becomes applicable to determine the fractions of how the three kinds of players evolve.[5]

The world population is 7.9 billion, of which 74% is above 15 years old (Szmigiera, 2021) and 66.8% is above 20 years old (Ang, 2021). Assume that 69.7% is above 18 years old, i.e. 5.5 billion. The World Bank (2017) estimates that 1.7 billion adults lack a bank account, which is subtracted from 5.5 billion to give 3.8 billion adults with a bank account. Howarth (2021) estimates 300 million cryptocurrency users on October 25, 2021, i.e. 5.5% of adults and 7.9% of adults with a bank account. The authors expect these percentages to increase in the future. Without knowing which digital currencies may succeed as global currencies, the authors believe that players may increasingly sort themselves into conventionalists, pioneers, and criminals.

1.3. Literature

Limited literature exists on this topic. The following literature review is intended to cover and extend beyond this article's topic, usefully divided into four groups as an overview, i.e. competition between fiat currencies and cryptocurrencies, CBDC and cryptocurrencies, the cryptocurrency market, and game theoretic analyses.

1.3.1. Competition between fiat currencies and cryptocurrencies

The following articles that have been identified are the closest relative to the current article and somehow consider competition between fiat currencies and cryptocurrencies, with various implications. Schilling and Uhlig (2019) enable agents to choose between two kinds of currencies, i.e. a cryptocurrency and a fiat currency. They explore how asymmetry in transaction costs and exchange fees decreases currency substitution. This exploration corresponds to the generally different transaction efficiencies considered for the national and global currencies in the current article. For payments of certain goods, cryptocurrencies are more suitable or cost less than fiat money, due to censorship resistance, tax evasion and anonymity. However, exchanging cryptocurrencies to fiat money is costly, and some goods are more easily purchased using fiat money. The condition under which agents are indifferent between purchasing with Bitcoin or US dollars depends on the amount of the value-added tax and transaction fees to miners. These assessments correspond to some extent to different backing, convenience, confidentiality, financial stability, and security for the national and global currencies in the current article.

Fernández-Villaverde and Sanches (2019) build a model of competition among privately issued fiat currencies. Based on the Lagos-Wright environment, they identify a price stable equilibrium for multiple currencies, comparable to two coexisting currencies in the current article,

[5] In return for sacrificing generality, a successful specification through functional forms demonstrates internal consistency and is illuminating. For example, the Cobb-Douglas function has enhanced our understanding of consumer preferences. Functional forms facilitate determining ranges of parameter values within which solutions are possible.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

and various less desirable equilibria. In the current article society's expected utility is a weighted sum, by the fraction of players of each kind, of each player's expected utility.

Almosova (2018) extends her model by assuming that the circulation of private currencies involves costs, i.e. verification of transactions, mining costs, etc. She points out that cryptocurrency competition will not cause price stability. But when the costs of private currency circulation are sufficiently low, competition will impose a downward pressure on the inflation of the public currency.

Rahman (2018) applies the Friedman rule to investigate the implications of digital and fiat currency competition for monetary policy. He finds that a monetary equilibrium with a purely private arrangement of digital currencies cannot deliver a socially efficient allocation. Rahman's (2018) article is linked to the current article, which considers society's expected utility as a weighted sum of the three kinds of players' expected utilities.

Benigno, Schilling, and Uhlig (2019) consider a two-country economy with complete markets, two national currencies and a global cryptocurrency. They propose that the deviation from interest rate equality implies the risk of approaching the zero lower bound or the abandonment of the national currency, which they call Crypto-Enforced Monetary Policy Synchronization (CEMPS). Consequently, the impossibility of simultaneously ensuring a fixed exchange rate, free capital flows and an independent monetary policy (the classic Impossible Trinity) becomes even less reconcilable.

Verdier (2021) examines how issuing a digital currency impacts competition in the deposit and lending markets. She assumes that a digital currency can be issued or managed by a central bank, a regulated entity, or a non-bank operator, and that a digital currency issued by a non-bank operator does not enable offering loans to individuals. This assumption gradually seems ready for revision as decentralized finance increasingly allows loans, e.g. of cryptocurrencies, to individuals. Verdier (2021) assumes that depositors decide how much money to store in a bank account or in a digital currency account. Thus, issuing a digital currency generates a crowding-out effect on commercial deposits. The author concludes that the lending rate of banks increases when a digital currency crowds out a higher amount of bank deposits.

1.3.2. CBDCs and cryptocurrencies

The following articles that have been identified are the closest relative to the current article and compare CBDCs and cryptocurrencies, where we interpret CBDC as the national currency and cryptocurrencies as the global currency. Caginalp and Caginalp (2019) determine Nash equilibria for how players divide their assets between a home currency and a cryptocurrency, similarly to the focus in the current article. Additionally they assume that the government seizes fractions of the players' assets with certain probabilities.

Blakstad and Allen (2018) review opportunities for central banks and individuals presented by cryptocurrencies for central banks and individuals, together with the risks. They assess possible impacts on financial systems and structures which may challenge CBDC issuance.

Masciandaro (2018) proposes a function of a store of information for cryptocurrencies and central bank digital currencies as new media of payments emerge over the next years, supplementing a medium of exchange and a store of value. Thus, the evolution of the different media of payments may depend on individual preferences.

Benigno (2021) points out that the presence of multiple currencies can jeopardize the primary function of central banking. In addition, in a world of multiple competing currencies issued by profit-maximizing agents, the nominal interest rate and inflation are both determined by structural factors, i.e. the intertemporal discount factor, the exit rate and the fixed cost of entry, and are thus not subject to manipulation.

Asimakopoulos, Lorusso, and Ravazzolo (2019) present a Dynamic Stochastic General Equilibrium (DSGE) model to evaluate the economic repercussions of cryptocurrencies. They

DOI: 10.7172/2353-6845.jbfe.2021.2.6

estimate the model with Bayesian techniques. They document a sturdy substitution effect between the real balances of government currency and cryptocurrencies, in response to technology, preferences and monetary policy shocks. Similarly, the current article shows how the three kinds of players strike balances between the two currencies.

1.3.3. The cryptocurrency market

The following articles analyze multiple currencies in the cryptocurrency market, which relates to the current article since the two currencies may also be two cryptocurrencies which evolve over time with fluctuating volume fractions of transactions. ElBahrawy, Alessandretti, Kandler, Pastor-Satorras, and Baronchelli (2017) assess the evolutionary dynamics of the cryptocurrency market. They illustrate the fluctuating market shares of 1,469 cryptocurrencies between April 2013 and May 2017, akin to fluctuations.

Caporale, Gil-Alana, and Plastun (2018) implement a rescaled range analysis and a fractional integration method to analyze the persistence in the cryptocurrency market. They identify a positive correlation between cryptocurrencies' past and future values.

ElBahrawy, Alessandretti, and Baronchelli (2019) investigate the relationship between online attention to digital currencies on Wikipedia and market dynamics across multiple digital currencies.

White (2014) points out, based on empirical observation, that as a first-mover monopolist in the market for cryptocurrencies, Bitcoin is surrounded by effective competitors. The introduction of various altcoins, if successful, decreases Bitcoin's market share. The current article similarly shows how the market share of two currencies may change over time.

Sapkota and Grobys (2021) analyze the top ten cryptocurrencies ranked by market capitalization in 2016–2018. They find that the submarket equilibria of privacy coins and the submarket equilibria of non-privacy coins are unrelated. This contrasts with the current article where players strike balances between which currencies to choose, and what kind of player to be.

Milunovich (2018) applies Granger causality tests to five popular cryptocurrencies and six major asset classes. He estimates weak connectedness between the two groups and strong connectedness within each group. A few exceptions exist. Out of 80 cross-pairs, six statistically significant relations are shown from non-digital to digital assets (e.g. from Monero to US$), and two statistically significant relations are shown from digital to non-digital assets (e.g. from the SPGSCI commodity index to Litecoin).

Gandal and Halaburda (2016) explore how network effects impact competition in the cryptocurrency market. They identify no winner-take-all effects in the early stages, but strong network effects and winner-take-all dynamics more recently. Similarly, the current article shows how two currencies and three kinds of players may coexist, and also that one kind of players, e.g. conventionalists, may go extinct.

1.3.4. Game theoretic analyses

The following articles are game theoretic analyses, which are linked to this group since the three kinds of players, while choosing among two currencies, interact with each other through time modeled by game theory and replicator dynamics. Imhof and Nowak (2006) propose that a frequency dependent, stochastic Wright-Fisher process can be used to describe the evolutionary game dynamics in finite populations to determine which of two strategies survives. This article similarly determines how the fractions of the three kinds of players, and the volume fraction of transactions in each currency, evolve over time.

Lewenberg, Bachrach, Sompolinsky, Zohar, and Rosenschein (2015) develop a cooperative game theoretic model to explore the dynamics of pooled Bitcoin mining and rewards. They show that it is difficult or even impossible to distribute rewards in a stable way. Players are always

DOI: 10.7172/2353-6845.jbfe.2021.2.6

incentivized to switch between pools. This is partly linked to the current article where players switch between which of three kinds of players to be, and which volume fraction of transactions in each currency to choose.

1.4. Article Organization

Section 2 presents the model. Section 3 analyzes the model. Section 4 explains the implications of the results. Section 5 concludes.

2. THE MODEL

2.1. Nomenclature

Parameters

j Currency of kind j, $j = n, g$
n National currency
g Global currency
i Player of kind i, $i = x, y, z$
x Conventionalist player
y Pioneer player
z Criminal player
b_{ij} Output subelasticity for backing of currency j at time t as perceived by player i, $b_{ij} \geq 0$
c_{ij} Output subelasticity for convenience of currency j at time t as perceived by player i, $c_{ij} \geq 0$
d_{ij} Output subelasticity for confidentiality of currency j at time t as perceived by player i, $d_{ij} \geq 0$
e_{ij} Output subelasticity for transactional efficiency for currency j at time t as perceived by player i, $e_{ij} \geq 0$
f_{ij} Output subelasticity for financial stability of currency j at time t as perceived by player i, $f_{ij} \geq 0$
s_{ij} Output subelasticity for security of currency j at time t as perceived by player i, $s_{ij} \geq 0$
w_i Fraction of player i's transactions which is criminal, $0 \leq w_i \leq 1$
k_i Scaling exponent for what player i retains after criminal behavior, $k_i \geq 0$
ω_i Probability that the government detects and prosecutes player i's criminal behavior, $0 \leq \omega_i \leq 1$
m_i Scaling exponent for how player i gets increased/decreased expected utility, $-\infty \leq m_i \leq \infty$
μ_i Scaling proportionality parameter for how player i gets increased expected utility, $\mu_i \geq 0$
α_i Parameter for the rapidity of change or sensitivity of the replicator equation, $\alpha_i > 0$
t Time, $t \geq 0$

Free choice variables

p_i Volume fraction of player i's transactions in currency n, $0 \leq p_i \leq 1$, $i = x, y, z$
$1-p_i$ Volume fraction of player i's transactions in currency g, $0 \leq 1 - p_i \leq 1$
p Volume fraction of all players' transactions in currency n, $0 \leq p \leq 1$
$1-p$ Volume fraction of all players' transactions in currency g, $0 \leq 1 - p \leq 1$
q_i Fraction of players of kind i, $0 \leq q_i \leq 1$, $i = x, y, z$, $q_x + q_y + q_z = 1$
q_x Fraction of conventionalists
q_y Fraction of pioneers
q_z Fraction of criminals, $q_z = 1 - q_x - q_y$

Dependent variables

$U_i(p_i, q_i)$ Player i's expected utility, $i = x, y, z$
U Society's expected utility

DOI: 10.7172/2353-6845.jbfe.2021.2.6

2.2. Two Currencies *n* and *g*

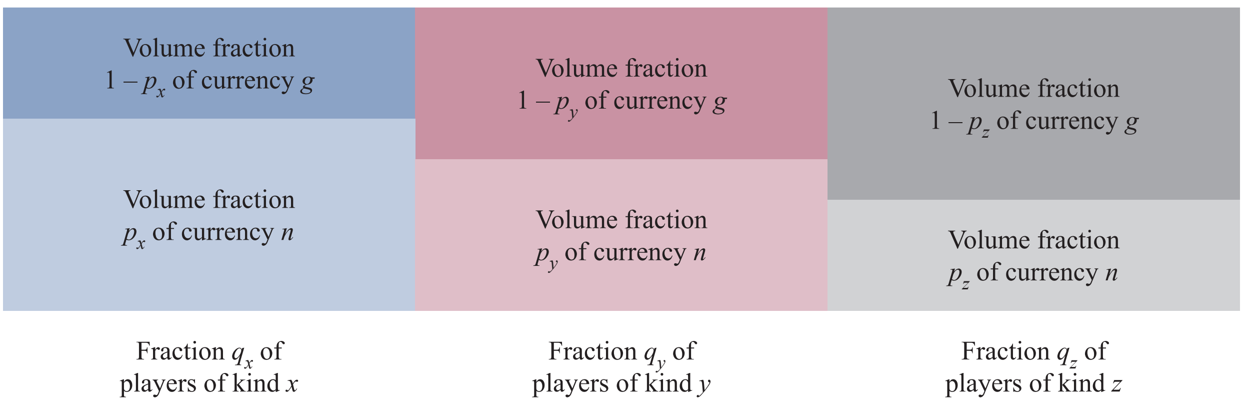
Consider an economy with two available currencies. The first currency n is national and offers the most common usage, and especially legal usage, within the economy. Examples of usage are to make various purchases or pay taxes. For simplicity, we can think of this currency as a CBDC (central bank digital currency). The second currency g is a global currency which on the one hand offers more limited usage (e.g. cannot be used for all kinds of purchases), but on the other hand offers other opportunities, e.g. tax evasion, payment on the black market, user autonomy, discretion, peer-to-peer focus, no banking fees, low transaction fees. For simplicity, we can think of this currency as a cryptocurrency such as Bitcoin or Monero, a privately issued currency such as Meta's Diem, or some future hypothetical currency operating globally.

2.3. Three Kinds of Players *x*, *y*, *z*

Assume three kinds of players which we can think of as households, referred to as player i, $i = x, y, z$. We can think of the three kinds of players as conventionalists, pioneers and criminals, respectively. Conventionalists tend to do what is traditional and historically common, and tend to prefer the national currency n more than the global currency g. Pioneers (early adopters) tend to break away from tradition and prefer the global currency g more than the national currency n. Criminals prefer not to get caught and tend to prefer the global currency g more than the national currency n if the global currency g offers confidentiality and user autonomy, e.g. through a privacy coin such as Monero. Assume that q_i, $0 \leq q_i \leq 1$ is the fraction of players of kind i. We assume that q_x is the fraction of conventionalists, that q_y is the fraction of pioneers, and that $q_z = 1 - q_x - q_y$ is the fraction of criminals. As time progresses, what used to be conventional may become old-fashioned, and what pioneers do may become conventional. Hence q_x and q_y may change over time. All players of the same kind i are equivalent. Player i (i.e. player of kind i) conducts a volume fraction p_i, $0 \leq p_i \leq 1$ of its transactions in currency n, and the remaining volume fraction $1 - p_i$ of its transactions in currency g, as shown in Figure 1 which assumes $p_x > p_y > p_z$, but generally $0 \leq p_i \leq 1$, $i = x, y, z$.

Figure 1

Three kinds of players. Player i (i.e. player of kind i), $i = x, y, z$, conducts a volume fraction p_i of its transactions in currency n, and the remaining volume fraction $1 - p_i$ of its transactions in currency g, $0 \leq q_i \leq 1$, $q_x + q_y + q_z = 1$. The illustration assumes $p_x > p_y > p_z$, but generally $0 \leq p_i \leq 1$, $i = x, y, z$.



DOI: 10.7172/2353-6845.jbfe.2021.2.6

2.4. Volume Fraction *p* of All Players' Transactions in Currency *n*

The volume fraction p of all players' transactions in currency n is the weighted sum of each player i's volume fraction in currency n, weighted by the fraction of each kind of player i, $i = x, y, z$, i.e.

$$p = \sum_{i=x,y,z} p_i q_i. \tag{1}$$

2.5. Cobb-Douglas Utility With Two Output Elasticities

Assume that player i has a risk-neutral Cobb-Douglas utility in net terms, hereafter referred to as utility, described by

$$U_{iCD}(p_i) = p_i^{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}}(1-p_i)^{b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}} \tag{2}$$

with one output elasticity $b_{in} + c_{in} + d_{in} + e_{in} + f_{in} + s_{in}$ for the national currency n, and one corresponding output elasticity $b_{ig} + c_{ig} + d_{ig} + e_{ig} + f_{ig} + s_{ig}$ for the global currency g. Player i's Cobb-Douglas utility $U_{iCD}(p_i)$ in (2) is multiplied with a penalty described in the next section 2.6 if player i's criminal behavior is detected and prosecuted by the government, and multiplied with the impact of the fractions q_x, q_y, q_z of the three kinds of players in the subsequent section 2.7. When $S = b_{in} + c_{in} + d_{in} + e_{in} + f_{in} + s_{in} + b_{ig} + c_{ig} + d_{ig} + e_{ig} + f_{ig} + s_{ig} = 1$, $S > 1$, $S < 1$, (2) expresses constant, increasing, and decreasing returns to scale, respectively. The 12 output subelasticities a_{ij}, $a_{ij} = b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij}, s_{ij}$ in (2), for currency j, $j = n, g$, at time t as perceived by player i, $i = x, y, z$, are as follows:

First, b_{ij} expresses how currency j has various forms of backing from actors, systems or characteristics that users of currency j respect and trust, as perceived by player i. Examples of backing for currency j are central banks for CBDCs, and various decentralized characteristics such as a distributed ledger technology for cryptocurrencies. The variable b_{ij} is not objective, but depends on player i's subjective judgment. The parameter b_{ij} expresses the weighted average backing of currency j by its users, i.e. within each of the three kinds x, y, z of players. For example, legitimate lawful users preferring transparency and allegiance to a certain country, may back the CBDC (central bank digital currency) of that country, which may be currency n, whereas illegitimate users may not back that currency, but back the global currency g instead. Criminal users may, for example, back a privacy cryptocurrency such as Monero, which may also be backed by many legitimate users. Currently, after the gold standard collapse (June 5, 1933 in the US), no fiat currency is backed by gold. The extent to which a player backs currency j may depend on a variety of factors. For example, a central bank may back its CBDC in the hope of obtaining a broader tax base, reduced tax evasion, a backstop to the private sector which may fail, and enhanced financial inclusion.

Second, c_{ij} expresses the convenience of using currency j as perceived by player i. One example of convenience is ease of use, e.g. few and easily comprehensible operations when purchasing at the supermarket or online, when transferring funds nationally or globally, or when incurring and paying back a loan. Other or related examples are how electronic wallets operate, how transfers between one's own and other wallets operate, and how offline transactions are processed when offline and getting back online. Furthermore, for some digital currencies users may not need to open a bank account with required identifications, but may instead install a digital currency wallet, and transact and pay via a digital currency address.

Third, d_{ij} expresses the confidentiality of using currency j, as perceived by player i, which expresses well-known balances to be struck between privacy, availability or accessibility for

DOI: 10.7172/2353-6845.jbfe.2021.2.6

oneself and various other players, and discrimination. For example, privacy cryptocurrencies such as Monero, Dash, and Zcash[6] offer enhanced privacy for users since transactions are harder to track, which also may make it harder to rectify, correct, or reverse undesirable transactions. For example, paying ransom money in Monero may preserve the anonymity of the recipient and the provider, but may make it harder for law enforcement to reverse or prosecute the transaction. A CBDC, properly designed, may offer confidentiality for player i with respect to many other players if the central bank can be trusted, but may not offer confidentiality for player i if the central bank cannot be trusted, or a court orders the confidentiality to be broken. The output subelasticity d_{ij} thus also expresses discrimination regarding in what sense and for whom and towards whom confidentiality is honored.

Fourth, e_{ij} expresses the transaction efficiency of currency j, as perceived by player i, operationalized as low cost, fast speed, affordability, and finality. Fast speed refers to how quickly the transaction is executed, which for cryptocurrencies is impacted by how many confirmations are needed for execution and how quickly the miners can mine blocks. Wire transfers have historically had a certain speed, and may be held up over weekends. Affordability refers to a fee or cost of executing the transaction, which is usually positively correlated with how quickly the transaction is executed. Finality refers to the extent to which the transaction is final, or can somehow be reversed or negotiated. Cryptocurrency transactions are usually irreversible, which is the common logic of smart contracts on the blockchain. Non-cryptocurrency transactions, exemplified by traditional wire transfers are usually reversible, e.g. if a court of law determines that the transaction was illegal. Costs of transactions have historically varied substantially across different kinds of transactions. Affordability may depend on size, recipient, sender, whether the transaction is recurring, etc. Costs may range from the common no costs, e.g. for grocery purchases, to high costs for international money transfers. Costs of transacting cryptocurrencies have usually been low, and often beneficial when transacting high amounts, with variation across different cryptocurrencies. Speed of transfers also vary. At the time of writing, the speed of CBDC transactions is unknown. For Bitcoin the average time for mining one block is 10 minutes. For two confirmations, the transaction may take 20 minutes. The initiator of a cryptocurrency transaction is usually requested to specify a transaction fee (e.g., low, medium, high), which impacts how quickly it gets processed by the miners. For Ethereum the average time for mining one block is 10–15 seconds, which may cause one transaction after two confirmations to require 20–30 seconds. In 2019 Bitcoin processes ca 4.6 transactions per second, while Visa processes ca 1700 transactions per second. The lightning network may speed up the transaction time for Bitcoin. Credit card transactions typically require around 48 hours to settle. The finality of transactions also pertains to efficiency. Some cryptocurrency exchanges may require three confirmations, six confirmations for large transactions, and 60 confirmations for very large transactions. Different central banks may develop different procedures for finality and confirmations depending on the characteristics of transactions, senders, recipients, etc., which impacts the efficiency e_{ij}.

Fifth, f_{ij} expresses the financial stability of currency j, as perceived by player i. The financial stability of the national currency n depends on the conditions in the given country. A variety of indicators exist for the financial stability of countries and currencies. Some currencies such as the Swiss franc, the Japanese yen, and the Norwegian krone are relatively stable (Protska, 2021b), while some, such as the Venezuelan bolivar, the Iranian ria and the Vietnamese dong (Protska, 2021a) can be more unstable than many cryptocurrencies. For CBDCs the central bank adjusts interest rates (which can be negative for digital currencies), and can be expected to be able to adjust a variety of factors to adjust the financial stability of currency j, within the constraints of the country's conditions. One hypothetical possibility is to adjust the tax rate for households or individuals depending on their characteristics (e.g. in understanding with tax authorities and

[6] https://www.investopedia.com/tech/five-most-private-cryptocurrencies/, retrieved November 22, 2021.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

others) to ensure financial stability. Fast response time when faced with crises, and activities to curtail or prevent money laundering and terrorist financing may impact the financial stability of currency j. Most cryptocurrencies, and especially altcoins, have traditionally varied substantially in value, caused partly by their novelty and limited usage, but also by the absence of a governing authority. One exception is stablecoins, e.g. Tether, USD Coin, TrueUSD, Dai, Paxos Standard, Binance USD, which have the stated purpose of being stable in some sense. The top ten list of countries adopting Bitcoin typically contains countries in the western world, but also countries which struggle to ensure financial stability, e.g. Venezuela (Lanz, 2020).

Sixth, s_{ij} expresses the security of currency j, as perceived by player i. A variety of security possibilities exist for digital currencies, see e.g. Allen et al. (2020) and Kiff et al. (2020). The security of the blockchain supporting Bitcoin has not collapsed since the first block was mined on January 3, 2009 at 18:15:05, although controversies and forks have occurred. Considering that 7,594 cryptocurrencies exist (https://coinmarketcap.com), 51% attacks are relatively rare.[7]

Each of the two output elasticities consists of six summed subelasticities as expressed above. Each of the six output subelasticities for the national currency n is of the form $p_i^{a_{in}}$, where p_i is the volume fraction of player i's transactions in the national currency n. Each of the six corresponding output subelasticities for the global currency g is of the form $(1-p_i)^{a_{ig}}$, where $1 - p_i$ is the volume fraction of player i's transactions in the global currency g. The parameter a_{ij}, $a_{ij} = b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij}, s_{ij}$ is the output subelasticity in the Cobb-Douglas function, $0 \le a_{ij} \le 1$, which is a characteristic of currency j, $j = n, g$, as perceived by player i. The output subelasticity a_{ij} may sometimes be objectively specified, and may occasionally be mutually agreed upon by the players x, y, z, allowing the removal of the subscript i from a_{ij}. Since objective specification, and mutual agreement, may not be generally possible, and player i may perceive the output subelasticity a_{ij} subjectively, we keep the subscript i on a_{ij}.

2.6. Detection and Prosecution of Criminal Behavior

Examples of criminal behavior are tax evasion, money laundering, theft, terrorist financing, corruption, and financial crimes. Although we expect criminals to be more criminal than conventionalists and pioneers, all these three kinds of players can in principle engage in criminal behavior, through both the national currency n and the global currency g. This reflects that in our societies no groups of citizens can be expected to be 100% non-criminal. We thus assume that a fraction w_i, $0 \le w_i \le 1$ of player i's transactions is criminal and is detected and prosecuted by the government with probability ω_i, $0 \le \omega_i \le 1$. The product $\omega_i w_i$ multiplies player i's fraction w_i of criminal behavior with its detection and prosecution probability ω_i. Hence $1 - \omega_i w_i$ expresses the joint probability of neither engaging in criminal behavior nor being detected and prosecuted. We introduce a scaling exponent k_i, $k_i \ge 0$, on the fraction w_i and express player i's expected utility as

$$U_{iC} = 1 - \omega_i w_i^{k_i} \tag{3}$$

which is a fraction between 0 and 1. When $k_i = 1$, player i's expected utility U_{iC} decreases linearly in the fraction w_i of player i's transactions which is criminal. When $k_i > 1$, U_{iC} decreases concavely in w_i, which economically means that a higher fraction w_i (compared with when $k_i = 1$) of player i's criminal transactions is needed in order to decrease player i's expected utility U_{iC}. In contrast, when $0 < k_i < 1$, U_{iC} decreases convexly in w_i, which economically means that a lower fraction w_i (compared with when $k_i = 1$) of player i's criminal transactions is sufficient in order to decrease

[7] The most well-known 51% attacks among cryptocurrencies occurred for Verge, Ethereum Classic, Bitcoin Gold, Feathercoin, and Vertcoin (Attah, 2019). A 51% attack means that a majority of miners impact mining to their advantage, including preventing other miners from completing blocks, and channeling funds from each block to themselves. Changing historical blocks is difficult due to the hard coding of past transactions into the Bitcoin software.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

player i's expected utility U_{iC}. When $k_i = 1$, $U_{iC} = 1 - \omega_i$ is independent of w_i. Player i's expected utility U_{iC} in (3) expresses what is probabilistically retained for potential criminal behavior, and is multiplied with player i's Cobb-Douglas utility $U_{iCD}(p_i)$ in (2) to determine what player i keeps of its utility when accounting for criminal behavior being probabilistically detected and prosecuted.

2.7. How a Fraction q_i of Players of Kind i Impacts Expected Utilities

Players of kind i may get increased or decreased expected utility if their fraction q_i increases or decreases. We operationalize this with the term $1 + \mu_i q_i^{m_i}$, where μ_i, $\mu_i \geq 0$ is a scaling proportionality parameter, and m_i is a scaling exponent. The term $1 + \mu_i q_i^{m_i}$ is multiplied with the Cobb-Douglas utility and what is probabilistically retained for potential criminal behavior.

Conventionalists prefer to do what others do and what is common, which gives them increased expected utility. Hence conventionalists get increased expected utility if the fraction q_x of conventionalists increases, i.e. $m_x \geq 0$. The positive exponent m_x scales the strength of how conventionalists get multiplicatively increased expected utility when the fraction q_x increases.

In contrast, pioneers prefer to do what others do not do, what is uncommon, and what breaks ground beyond what is conventional, which gives them increased expected utility. When pioneers become a majority, they are no longer pioneers, but conventionalists. Hence pioneers get decreased expected utility if the fraction q_y of pioneers increases, i.e. $m_y \leq 0$. The negative exponent m_y scales the strength of how pioneers get multiplicatively decreased expected utility when the fraction q_y increases.

Criminals focus on what is criminally lucrative, what they can get away with, and what does not get detected and prosecuted. Whether what they do is common or uncommon may be irrelevant. What criminals have in common with pioneers is that they prefer to be few so that they can operate under the radar. As criminals become more numerous, the benefits for each in most stable and relatively lawful societies can be expected to decrease since they compete with each other, and non-criminals adapt to defending against them. Exceptions, such as the Italian mafia in Italy, or the cartels in Colombia, operate according to another logic not considered in this article, where subsections of societies follow different norms. At the extreme, a society with only criminals will not function since everyone will prey on everyone causing breakdown. Hence criminals, just as pioneers, get decreased expected utility if the fraction q_z of criminals increases, i.e. $m_z \leq 0$. The negative exponent m_z scales the strength of how criminals get multiplicatively decreased expected utility when the fraction q_z increases.

The three paragraphs above enable us to operationalize player i's expected utility as

$$U_{iF}(q_i) = 1 + \mu_i q_i^{m_i} \tag{4}$$

which is multiplied with player i's Cobb-Douglas utility $U_{iCD}(p_i)$ in (2) and player i's expected utility U_{iC} in (3). When $m_i = 1$, player i's expected utility $U_{iF}(q_i)$ increases linearly in the fraction q_i of players of kind i. When $m_i > 1$, $U_{iF}(q_i)$ increases convexly in q_i, which economically means that a higher fraction q_i (compared with when $m_i = 1$) of players of kind i is needed in order to increase player i's expected utility $U_{iF}(q_i)$. In contrast, when $0 < m_i < 1$, $U_{iF}(q_i)$ increases concavely in q_i, which economically means that a lower fraction q_i (compared with when $m_i = 1$) of players of kind i is sufficient in order to increase player i's expected utility $U_{iF}(q_i)$. When $m_i = 0$, $U_{iF}(q_i) = 1 + \mu_i$ is independent of q_i.

Equation (4) means that player i's expected utility $U_{iF}(q_i)$ depends explicitly on the fraction q_i of players of kind i, $i = x, y, z$, which is a measure of the number of players of kind i. This dependence of $U_{iF}(q_i)$ on q_i implicitly means that $U_{iF}(q_i)$ depends on the fraction $1 - q_i$ of players which is not of kind i, since $q_x + q_y + q_z = 1$. That is, more players of one kind mean fewer players of the two other kinds. In the next section 3 on the replicator equation the interdependence of

DOI: 10.7172/2353-6845.jbfe.2021.2.6

the numbers of players of each kind, and thus the interaction between the three kinds of players, becomes clearer.

2.8. The Players' Expected Utilities

This section combines multiplicatively player i's expected utilities $U_{iCD}(p_i)$ in (2), U_{iC} in (3), and $U_{iF}(q_i)$ in (4), which gives player i's expected utility

$$\begin{aligned} U_i &= U_i(p_i, q_i) = U_{iCD}(p_i) U_{iC} U_{iF}(q_i) \\ &= p_i^{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}} (1-p_i)^{b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}} \left(1-\omega_i w_i^{k_i}\right)\left(1+\mu_i q_i^{m_i}\right). \end{aligned} \tag{5}$$

Equation (5) assumes that player i is risk neutral and abstracts away other factors such as player i's consumption preferences concerning goods, and player i's preference for work versus leisure, which are beyond the scope of this article. Such factors are to some extent implicitly or indirectly present in (5). For example, player i's convenience c_{ij} of using currency j and transaction efficiency e_{ij} of currency j may play different roles for different goods, and may impact player i's preference for work versus leisure.

2.9. Society's Expected Utility

Society's expected utility $U(p_x, p_y, p_z, q_x, q_y)$ is the weighted sum of each player's expected utility $U_i(p_i, q_i)$, weighted by the fraction of players of kind i, $i = x, y, z$, i.e.

$$U = U(p_x, p_y, p_z, q_x, q_y) = \sum_{i=x,y,z} q_i U_i(p_i, q_i),\ q_z = 1 - q_x - q_y. \tag{6}$$

2.10. The Players' Strategic Choices

Assume that player i at time t makes two strategic simultaneous choices to maximize its expected utility $U_i(p_i, q_i)$ in (5). First, it chooses its volume fraction p_i of its transactions in currency n, causing the remaining volume fraction $1 - p_i$ of its transactions to be in currency g. Player i's choice of p_i to maximize $U_i(p_i, q_i)$ in (5) does not depend on time t, and does not depend on the fraction q_i of player i in the population, since $1 + \mu_i q_i^{m_i}$ appears proportionally in (5), without impacting the shape of $U_i(p_i, q_i)$ as a function of p_i, and without impacting which value of p_i causes $U_i(p_i, q_i)$ to have its maximum. Hence no dynamic considerations for player i's choice of volume fraction p_i of its transactions in currency n are needed. Second, player i chooses which kind i of player it should be, $i = x, y, z$. That choice depends strongly on time t, as described by the replicator equation in the next section. When player i switches from being of one kind to another kind, $i = x, y, z$, its first choice of the optimal volume fraction p_i of its transactions in currency n also changes. In other words, as long as player i remains of a specific kind, its optimal volume fraction p_i does not depend on time t, which reflects real life, but if it switches to be of another kind according to the replicator equation described in the next section, then it also changes its optimal volume fraction p_i at time t to what is optimal for this new kind i, $i = x, y, z$.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

2.11. The Replicator Equation

To determine the evolution of the fraction q_i of players of kind i, $i = x, y, z$, we consider the replicator equation (Taylor & Jonker, 1978; Weibull, 1997)

$$\frac{\partial q_i}{\partial t} = \alpha_i q_i \left(U_i(p_i, q_i) - U(p_x, p_y, p_z, q_x, q_y)\right)$$
$$\Leftrightarrow \begin{bmatrix} \frac{\partial q_x}{\partial t} \\ \frac{\partial q_y}{\partial t} \end{bmatrix} = \begin{bmatrix} \alpha_x\left(U_x(p_x, q_x) - U(p_x, p_y, p_z, q_x, q_y)\right) & 0 \\ 0 & \alpha_y\left(U_y(p_y, q_y) - U(p_x, p_y, p_z, q_x, q_y)\right) \end{bmatrix} \begin{bmatrix} q_x \\ q_y \end{bmatrix} \tag{7}$$

where α_i, $\alpha_i > 0$, is the rapidity of change or sensitivity of the process. The process is stable when α_i is intermediate. If α_i is high, the process changes rapidly. If α_i is low, a negligible change occurs. The right hand side of (7) multiplies the fraction q_i of players of kind i with the difference $U_i(p_i, q_i) - U$ between player i's expected utility $U_i(p_i, q_i)$ and the average expected utility U of the three kinds $i = x, y, z$ of players. If the right hand side of (7) is positive (negative), player i's expected utility $U_i(p_i, q_i)$ is higher (lower) than the average expected utility U, which causes the fraction q_i of players of kind i to increase (decrease).

The economic interpretation of (7) is that the three kinds of players over time continuously move towards becoming the kind of player where the expected utility U_i, i.e. U_x, U_y, U_z, is highest. In doing so, player i accounts for both the income effect (i.e., the absolute value of player i's expected utility U_i) and the substitution effect (i.e., which kind of player is optimal for player i to be or become). As a player changes from being of one kind to becoming of another kind, the fraction q_i of players of kind i, i.e. the fractions q_x, q_y, $q_z = 1 - q_x - q_y$, change. The prominent presence of q_i in (7) on the left hand side, multiplicatively on the right hand side, and in $U_i(p_i, q_i)$ and $U(p_x, p_y, p_z, q_x, q_y)$, means that the replicator equation is quite sensitive to changes in q_i. The expected utilities $U_i(p_i, q_i)$ and $U(p_x, p_y, p_z, q_x, q_y)$ also depend on the volume fractions p_i and $1 - p_i$ of player i's transactions in the currencies n and g, respectively. Hence the replicator equation reflects how the three kinds of players perceive the two currencies n and g as they choose which kind of player they want to be to maximize their expected utility $U_i(p_i, q_i)$.

The limiting behavior (the evolutionary outcome) of the replicator equation in (7) is a Nash equilibrium. We determine a pure-strategy Nash equilibrium where each player i, $i = x, y, z$, maximizes its expected utility $U_i(p_i, q_i)$. This equilibrium is a set of strategies q_i^* for the three players, $i = x, y, z$, such that

$$U_i(p_i, q_i^*) \geq U_i(p_i, q_i) \forall 0 \leq q_i \leq 1, i = x, y, z; q_z = 1 - q_x - q_y. \tag{8}$$

For research on the equilibrium properties of replicator dynamics see (Duong & Han, 2020) and the references therein.

If $\alpha_i\left(U_i(p_i, q_i) - U(p_x, p_y, p_z, q_x, q_y)\right)$ in (7) had been constant, (7) would have been a linear time-invariant system for which well-known techniques illustrated by Khalil (2002, p. 46), or Laplace and Fourier transforms, are applicable. Since $\alpha_i\left(U_i(p_i, q_i) - U(p_x, p_y, p_z, q_x, q_y)\right)$ is not constant, (7) is a time-variant system which is more challenging to analyze theoretically. We thus proceed over to the next sections to analyze (7) with simulations.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

3. ANALYZING THE MODEL

3.1. Analyzing As a Function of p_i When q_i Is Exogenously Fixed

This section assumes that the fraction q_i of players of kind i is fixed, and analyzes how player i chooses its volume fraction p_i of currency n, implying volume fraction $1 - p_i$ for currency g. Differentiating player i's expected utility $U_i(p_i, q_i)$ in (5) with respect to p_i and equating with zero gives

$$\frac{\partial U_i(p_i,q_i)}{\partial p_i} = \left(\frac{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}}{p_i} - \frac{b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}}{1-p_i}\right)p_i^{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}}(1-p_i)^{b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}}\left(1-\omega_i w_i^{k_i}\right)\left(1+\mu_i q_i^{m_i}\right) = 0 \tag{9}$$

which is solved to yield

$$p_i = p_{iopt} = \frac{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}}{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}+b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}}. \tag{10}$$

Property 1. $\partial p_{iopt}/\partial a_{in} \geq 0$, $\partial p_{iopt}/\partial a_{ig} \leq 0$, $a_{ij} = b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij}, s_{ij}, j = n, g$.

Proof. Follows from differentiating (10).

Property 1 states that the optimal fraction p_{iopt} of player i's transactions in currency n increases in the six subelasticities a_{in} for currency n, and decreases in the six subelasticities a_{ig} for currency g.

Inserting $p_i = p_{iopt}$ into the second order derivative gives

$$\left.\frac{\partial^2 U_i(p_i,q_i)}{\partial p_i^2}\right|_{p_i=p_{iopt}} = -\left(b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}\right)p_{iopt}^{b_{in}+c_{in}+d_{in}+e_{in}+f_{in}+s_{in}-1}(1-p_{iopt})^{b_{ig}+c_{ig}+d_{ig}+e_{ig}+f_{ig}+s_{ig}-2}\left(1-\omega_i w_i^{k_i}\right)\left(1+\mu_i q_i^{m_i}\right) < 0 \tag{11}$$

which is satisfied as negative, and hence $p_i = p_{iopt}$ is a maximum.

To illustrate the model, the following plausible benchmark parameter values are chosen. If the 12 output subelasticities a_{ij}, $a_{ij} = b_{ij}, c_{ij}, d_{ij}, e_{ij}, f_{ij}, s_{ij}$, for player i, $i = x, y, z$, for currency j, $j = n, g$, were to be given equal weight, assuming constant returns to scale as specified after (2), each output subelasticity would get weight $a_{ij} = x, y, z = 1/12$.[8] Table 1a shows 36 output subelasticities a_{ij}, which all satisfy the requirement $a_{ij} \geq 0$, for player i, $i = x, y, z$, for currency j, $j = n, g$.

[8] Since we have no evidence to justify increasing or decreasing returns to scale, we make the simplest and common assumption of constant returns to scale.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Table 1
Output subelasticities a_{ij} in three panels a,b,c for currency j, $j = n, g$, as perceived by player i, $i = x, y, z$.

Player i	$i = x$		$i = y$		$i = z$	
Currency j	$j = n$	$j = g$	$j = n$	$j = g$	$j = n$	$j = g$
Panel a						
b_{ij}	1/4	0	0	1/4	0	1/12
c_{ij}	1/12	0	0	1/12	0	1/12
d_{ij}	1/12	1/12	1/12	1/12	1/12	1/4
e_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
f_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
s_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
Panel b						
b_{ij}	1/3	0	0	1/3	0	1/12
c_{ij}	1/12	0	0	1/12	0	1/12
d_{ij}	1/12	0	0	1/12	0	1/3
e_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
f_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
s_{ij}	1/12	1/12	1/12	1/12	1/12	1/12
Panel c						
b_{ij}	1/2	0	0	1/2	0	1/12
c_{ij}	1/12	0	0	1/12	0	1/12
d_{ij}	1/12	0	0	1/12	0	1/2
e_{ij}	1/12	0	0	1/12	0	1/12
f_{ij}	1/12	0	0	1/12	0	1/12
s_{ij}	1/12	1/12	1/12	1/12	1/12	1/12

Table 1a assumes that player x as a conventionalist prefers at least output subelasticity $a_{ij} = 1/12$ for all the six output subelasticities backing, convenience, confidentiality, transaction efficiency, stability, and security for the national currency n, and three times higher output subelasticity $b_{xn} = 1/4$ for the backing of currency n, which it respects and trusts, and justifies player x as a conventionalist. Table 1a further assumes that player x prefers at most output subelasticity $a_{ij} = 1/12$ for the six output subelasticities for the global currency g, and zero output subelasticity for the backing $b_{xg} = 0$ and convenience $c_{xg} = 0$ of currency g, which also justifies player x as a conventionalist. Table 1a assumes that player y as a pioneer has the opposite preference of player x, i.e. at least output subelasticity $a_{ij} = 1/12$ for all the six output subelasticities for the global currency g, and three times higher output subelasticity $b_{yg} = 1/4$ for the backing of currency g, at most output subelasticity $a_{ij} = 1/12$ for the six output subelasticities for the national currency n, and zero output subelasticity for the backing $b_{yn} = 0$ and convenience $c_{yn} = 0$ of currency n. Table 1a assumes that player z as a criminal has the same preference as the pioneer player y, except that its three times higher preference is for output subelasticity $d_{zg} = 1/4$ for the confidentiality of currency g. Hence it prefers output subelasticity $b_{zg} = 1/12$ for the backing of currency g.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Table 1b assumes that the three kinds of players have higher preferences $b_{xn} = b_{yg} = d_{zg} = 1/3$ for their preferred output subelasticities, i.e. backing of currencies n and g for players x and y, and confidentiality of currency g for player z. They compensate for these higher preferences by having no preferences $d_{xg} = d_{yn} = d_{zn} = 0$ for confidentiality, i.e. of currency g for player x and of currency n for players y and z.

Table 1c assumes that the three kinds of players have even higher preferences $b_{xn} = b_{yg} = d_{zg} = 1/2$ for their preferred output subelasticities, i.e. backing of currencies n and g for players x and y, and confidentiality of currency g for player z. They compensate for these higher preferences by having no preferences $e_{xg} = e_{yn} = e_{zn} = f_{xg} = f_{yn} = f_{zn} = 0$ for transaction efficiency and financial stability, i.e. of currency g for player x and of currency n for players y and z. We alternate between applying Table 1 panels a, b, c, and combinations of these for players x, y, z, as our benchmark, as we proceed.

The benchmark furthermore assumes that the conventionalist player x and pioneer player y choose a zero fraction $w_i = 0$ of its transactions to be criminal, $i = x, y$, which may be a good approximation for many countries, while the criminal player z chooses a positive fraction $w_z = 0.5$ of its transactions to be criminal, assumed as a focal intermediate between $w_z = 0.5$ and $w_z = 1$. The government is assumed to detect and prosecute criminal behavior with probability $\omega_i = 0.5$, also assumed as a focal intermediate between $w_z = 0.5$ and $w_z = 1$. We assume scaling exponent $k_i = 1$ for what player i retains after criminal behavior, which in (3) means that player i's expected utility decreases linearly in the fraction w_i of player i's transactions which is criminal. The authors believe that a linear decrease is more plausible than a convex or concave decrease. Unitary values, also assumed below to the extent possible, are assumed plausible focal points when no particular evidence seems suitable for non-unitary values.

The scaling exponent for how player i gets increased or decreased expected utility depending on the fraction q_i of players of kind i is assumed to be positive and unitary, $m_x = 1$, for conventionalists, and negative and unitary, $m_y = m_z = -1$, for pioneers and criminals.

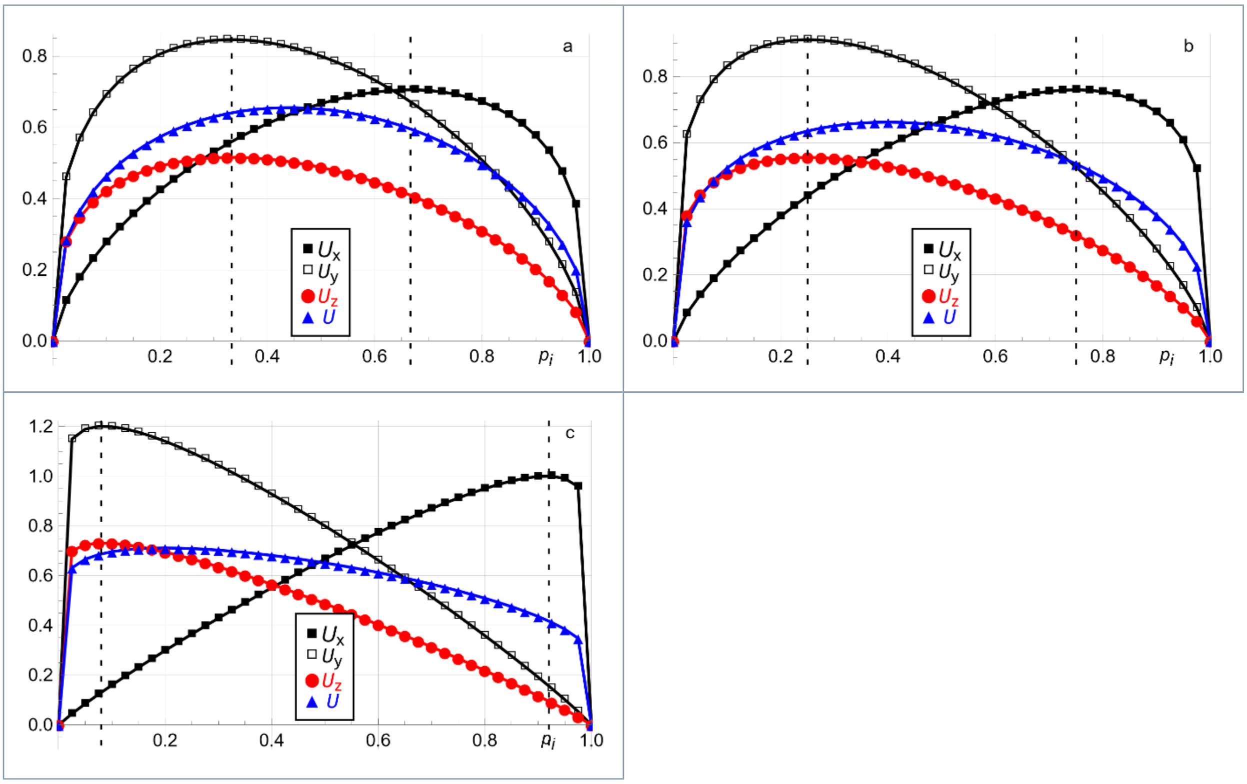
The scaling proportionality parameter μ_i for how player i gets increased or decreased expected utility depending on the fraction q_i of players of kind i, $i = x, y, z$, impacts the analysis crucially. We assume the unitary $\mu_x = 1$ as a benchmark for conventionalists, which in (4) causes $U_{xF}(q_x)$ to vary between $U_{xF}(q_x) = 1$ when $q_x = 0$ and $U_{xF}(q_x) = 2$ when $q_x = 1$. For pioneers and criminals we assume $\mu_i < 1$, since $U_{iF}(q_i)$ in (4) varies between $U_{iF}(q_i) = \infty$ when $q_i = 0$ and $U_{iF}(q_i) = 1 + \mu_i$ when $q_i = 1$, $i = x, y$, since $m_y = m_z = -1$. More specifically, we assume the five times lower $\mu_y = 0.2$ for pioneers and the ten times lower $\mu_z = 0.1$ for criminals.

In this section, where the fraction q_i of players of kind i is exogenous, we assume equally large fractions $q_i = 1/3$ of the three kinds of players, $i = x, y, z$, thus not giving eminence to one kind of player over another kind. The values $q_i = 1/3$ are needed to determine player i's expected utility $U_i(p_i, q_i)$ in (5), due to the last proportional term $1 + \mu_i q_i^{m_i}$, but do not impact the shape of $U_i(p_i, q_i)$ as a function of p_i and for which value of p_i that $U_i(p_i, q_i)$ has its maximum.

Figure 2 applies the above benchmark, including the exogenous $q_i = 1/3$, and plots player i's expected utility U_i in (5) and society's expected utility U in (6) as functions of player i's volume fraction p_i of currency n, $i = x, y, z$. The Mathematica software (www.wolfram.com) is used for plotting. Panel k assumes the output subelasticities a_{ij} in Table 1k, k = a, b, c. The two dashed vertical lines in each panel show the values of p_i where at least one expected utility U_i has its maximum value, i.e. $p_x = 2/3$ and $p_y = p_z = 1/3$ in panel a, $p_x = 3/4$ and $p_y = p_z = 1/4$ in panel b, and $p_x = 11/12$ and $p_y = p_z = 1/12$ in panel c. In panel a, society's expected utility U reaches its maximum at $p_i = 4/9$ which is the weighted sum of the p_i's across the three kinds of players. If the weights change from q_i =1/3, e.g. such that q_z increases and q_x and q_y decrease, the value p_i changes from $p_i = 4/9 \approx 0.44$ towards $p_i = 2/3$. In panels b and c, society's expected utility U reaches their maxima at $p_i = 5/12 \approx 0.42$ and $p_i = 9/25 = 0.36$, calculated analogously.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Figure 2
Player i's expected utility U_i as a function of its volume fraction p_i of currency n when $q_i = 1/3$, $i = x, y, z$. Panel k assumes the output subelasticities a_{ij} in Table 1k, k = a, b, c.



In all the three panels in Figure 2 the conventionalist player x's inverse U-shaped expected utility U_x is skewed towards the right since it values the national currency n more than the global currency g. When the volume fraction p_x of the conventionalist player x's transactions in the national currency n is low, the conventionalist player x's expected utility U_x is intuitively low. As the fraction p_x increases, its expected utility U_x increases to its maximum when $p_x = 2/3$, $p_x = 3/4$, $p_x = 11/12$, in panels a, b, c, and thereafter decreases, as player x also assigns some, although low, output subelasticities to currency g.

In contrast, in all the three panels in Figure 2 the pioneer player y's and criminal player z's inverse U-shaped expected utilities U_i are skewed towards the left since they value the global currency g more than the national currency n, and thus prefer $p_i < 1/2$. As the fraction p_i increases, its expected utility U_i increases to its maximum when $p_i = 1/3$, $p_i = 1/4$, $p_i = 1/12$, in panels a, b, c, respectively, $i = x, y$. As p_i increases further, U_i decreases. The criminal's expected utility U_z is lower than the pioneer's expected utility U_y since its fraction $w_z = 0.5$ of transactions is criminal, detected and prosecuted by the government with probability $\omega_i = 0.5$.

3.2. Analysis Applying the Replicator Equation

This section applies the replicator equation in (7) to determine the fraction q_i of players of kind i endogenously, while player i determines the volume fraction p_i of currency n by maximizing its expected utility U_i in (5), $i = x, y, z$. Figure 3 applies the output subelasticities in Table 1 and the benchmark parameter values in section 3.1, i.e. $w_x = w_y = 0$, $w_z = 0.5$, $\omega_i = 0.5$, $k_i = 1$, $m_x = 1$, $m_y = m_z = -1$, $\mu_x = 1$, $\mu_y = 0.2$, $\mu_z = 0.1$, $i = x, y, z$. Player i chooses its volume fraction p_i of currency n optimally to maximize its expected utility U_i, $i = x, y, z$. Assuming rapidity $\alpha_i = 1$ of change or sensitivity of the replicator equation, $i = x, y, z$, (7) is used to determine the fraction q_i of players of kind i, $i = x, y, z$. Figure 3 plots these fractions q_x, q_y, $q_z = 1 - q_x - q_y$, and the volume fraction p of all players' transactions in the national currency n from (1), as functions of time t.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Figure 3
Fraction q_i of players of kind i, $i = x, y, z$, and the volume fraction p of all players' transactions in currency n, as a function of time t for the benchmark parameter values in Table 1, $w_x = w_y = 0$, $w_z = 0.5$, $\omega_i = 0.5$, $k_i = 1$, $m_x = 1$, $m_y = m_z = -1$, $\mu_x = 1$, $\mu_y = 0.2$, $\mu_z = 0.1$, $\alpha_i = 1$, $i = x, y, z$. Panel a: Table 1a. Panel b: Table 1b. Panel c: Table 1c. Panel d: Table 1a for player x and Table 1c for players y and z. Panel e: Table 1c for player x and Table 1a for players y and z.

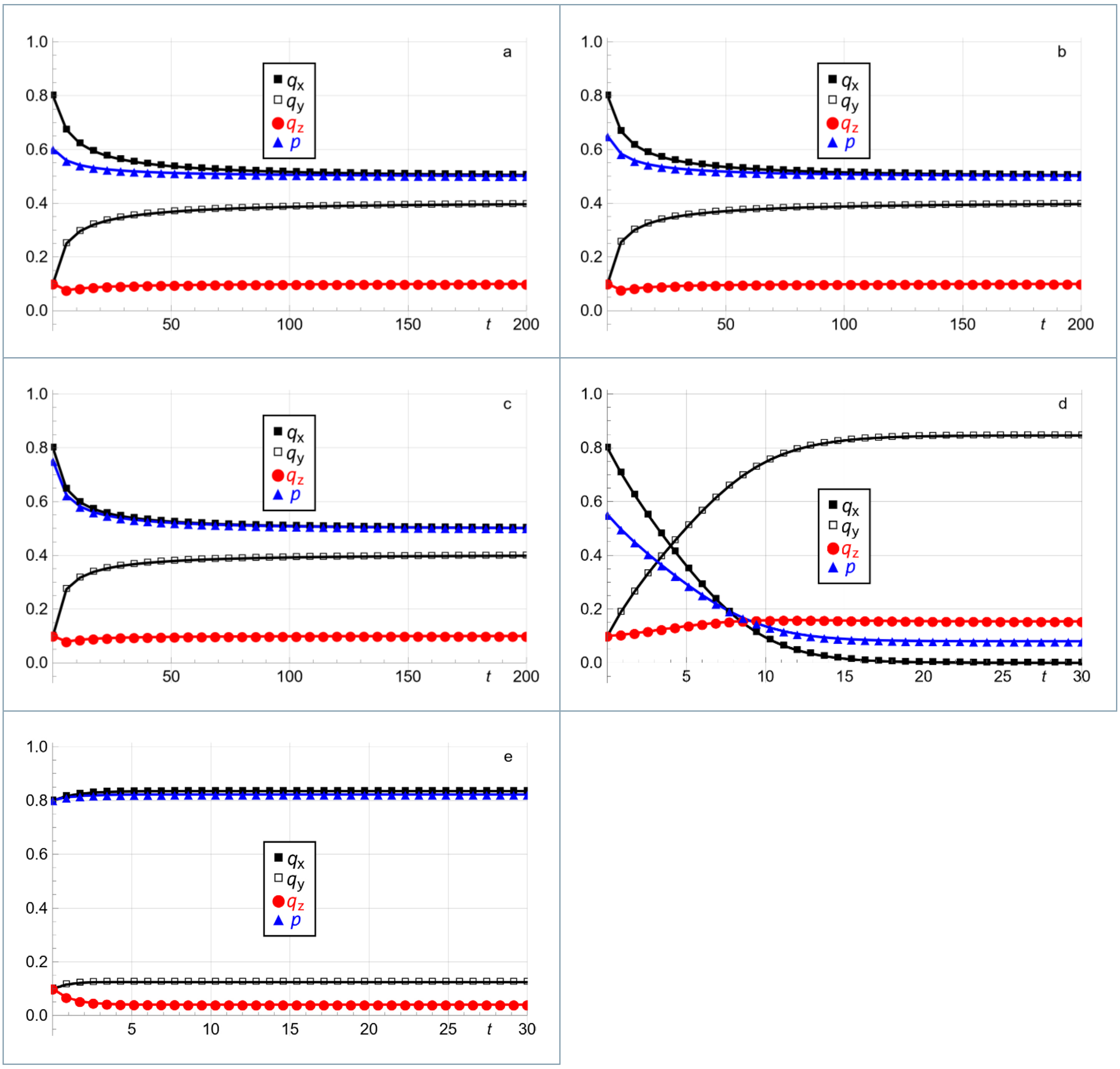


Figure 3 assumes initial conditions at time $t = 0$ equal to $q_x(0) = 0.8$ and $q_y(0) = q_z(0) = 0.1$, which means that conventionalists initially are in the majority at 80%, while pioneers and criminals are in the minority, each at 10%.

Figure 3a assumes the 36 output subelasticities in Table 1a, which according to Figure 2a gives the optimal volume fractions $p_x = 2/3$ for conventionalists and $p_y = p_z = 1/3$ for pioneers and criminals, for player i's transactions in currency n. The fraction q_x of conventionalists decreases convexly from $q_x(0) = 0.8$ to $\lim_{t\to\infty} q_x = 0.5$, hereafter referred to as the stationary solution, after sufficiently much time t has elapsed. All limit values are determined numerically. The fraction q_y of pioneers increases concavely from $q_y(0) = 0.1$ to $\lim_{t\to\infty} q_y = 0.4$. The fraction q_z of criminals first decreases marginally and briefly from $q_z(0) = 0.1$, as the fraction q_y of pioneers increases rapidly. Thereafter q_z increases concavely back up towards $\lim_{t\to\infty} q_z = 0.1$. Hence the volume fraction p of all players' transactions in the national currency n decreases towards $\lim_{t\to\infty} p = 0.5$.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Figure 3b assumes the 36 output subelasticities in Table 1b, which according to Figure 2b gives the higher optimal volume fractions $p_x = 0.75$ for conventionalists and the lower $p_y = p_z = 0.25$ for pioneers and criminals, for player i's transactions in currency n. The evolution of the fractions q_x, q_y, q_z is qualitatively similar to Figure 3a, with the same limit values $\lim_{t\to\infty} q_x = \lim_{t\to\infty} p = 0.5$, $\lim_{t\to\infty} q_y = 0.4$, $\lim_{t\to\infty} q_z = 0.1$. The reason for the similar result is that the increase in the optimum from $p_x = 2/3$ to $p_x = 3/4$ for conventionalists equals the decrease in the optimum from $p_y = p_z = 1/3$ to $p_y = p_z = 1/4$ for pioneers and criminals. These changes are in the opposite direction and equal $3/4 - 2/3 = 1/3 - 1/4 = 1/12$. Furthermore, at the limit when $t \to \infty$, the fraction q_x of conventionalists equals the sum of the fractions q_y and q_z of pioneers and criminals, i.e. $\lim_{t\to\infty} q_x = 0.5 = \lim_{t\to\infty} q_y = 0.4 + \lim_{t\to\infty} q_z = 0.1$, which means that the impact in the opposite direction when determining q_x, q_y, q_z in (7) is equally strong.

Figure 3c assumes the 36 output subelasticities in Table 1c, which according to Figure 2c gives the higher optimal volume fractions $p_x = 0.92$ for conventionalists and the lower $p_y = p_z = 0.08$ for pioneers and criminals, for player i's transactions in currency n. Also here the evolution of the fractions q_x, q_y, q_z is qualitatively similar to Figure 3a and Figure 3b, with the same limit values $\lim_{t\to\infty} q_x = \lim_{t\to\infty} p = 0.5$, $\lim_{t\to\infty} q_y = 0.4$, $\lim_{t\to\infty} q_z = 0.1$. The reason for the similar result is again that the increase in the optimum from $p_x = 2/3$ to $p_x = 11/12$ for conventionalists equals the decrease in the optimum from $p_y = p_z = 1/3$ to $p_y = p_z = 0.08$ for pioneers and criminals. These changes are in the opposite direction and equal $11/12 - 2/3 = 1/3 - 1/12 = 1/4$. At the limit when $t \to \infty$, the fraction q_x of conventionalists equals the sum of the fractions q_y and q_z of pioneers and criminals, i.e. $\lim_{t\to\infty} q_x = 0.5 = \lim_{t\to\infty} q_y + \lim_{t\to\infty} q_z$, which means that the impact in the opposite direction when determining q_x, q_y, q_z in (7) is equally strong.

To illustrate results different from Figure 3a, b, c, we consider two extreme combinations of output subelasticities from Table 1, one favoring pioneers and criminals, and one favoring conventionalists. Figure 3d assumes the 12 output subelasticities in Table 1a for the conventionalist player x, which gives the minimum optimal volume fraction $p_x = 2/3$, and assumes the 24 output subelasticities in Table 1c for the pioneer and criminal players y and z, which gives the minimum optimal volume fractions $p_y = p_z = 1/12$. That both $p_x = 2/3$ and $p_y = p_z = 1/12$ are minimum optimum values for the respective players, among the alternatives in Table 1, chosen by the three kinds of players maximizing their expected utilities U_x, U_y, U_z in (5), means that all the three kinds of players choose currency n with minimum volume fractions p_x, p_y, p_z. That favors pioneers and criminals, who to a lower extent back and favor currency n. Consequently, the fractions q_y and q_z of pioneers and criminals increase concavely and quickly from $q_y(0) = q_z(0) = 0.1$ toward $\lim_{t\to\infty} q_y = 0.85$ and $\lim_{t\to\infty} q_z = 0.15$, while the fraction q_x of conventionalist decreases convexly and quickly from $q_x(0) = 0.8$ toward $\lim_{t\to\infty} q_x = 0$, thus going extinct. This shows how a change in the output subelasticities among the alternatives in Table 1 may tilt the balance from emphasis on the national currency n towards emphasis on the global currency g. Hence the volume fraction p of all players' transactions in the national currency n decreases towards $\lim_{t\to\infty} p = 1/12$.

Figure 3e assumes the 12 output subelasticities in Table 1c for the conventionalist player x, which gives the maximum optimal volume fraction $p_x = 11/12$, and assumes the 24 output subelasticities in Table 1a for the pioneer and criminal players y and z, which gives the maximum optimal volume fractions $p_y = p_z = 1/3$. That both $p_x = 11/12$ and $p_y = p_z = 1/3$ are maximum optimum values for the respective players, among the alternatives in Table 1, means that all the three kinds of players choose currency n with maximum volume fractions p_x, p_y, p_z. That favors conventionalists, who to a higher extent back and favor currency n. Consequently, the fraction q_x of conventionalists increases concavely, quickly and marginally from $q_x(0) = 0.8$ toward $\lim_{t\to\infty} q_x = 0.835$. The fraction q_y of pioneers increases concavely, quickly and marginally from $q_y(0) = 0.1$ toward $\lim_{t\to\infty} q_y = 0.125$. The fraction q_z of criminals decreases convexly and quickly from $q_z(0) = 0.1$ toward $\lim_{t\to\infty} q_z = 0.040$. This shows how a different change in the output subelasticities among the alternatives in Table 1 may preserve the emphasis on the

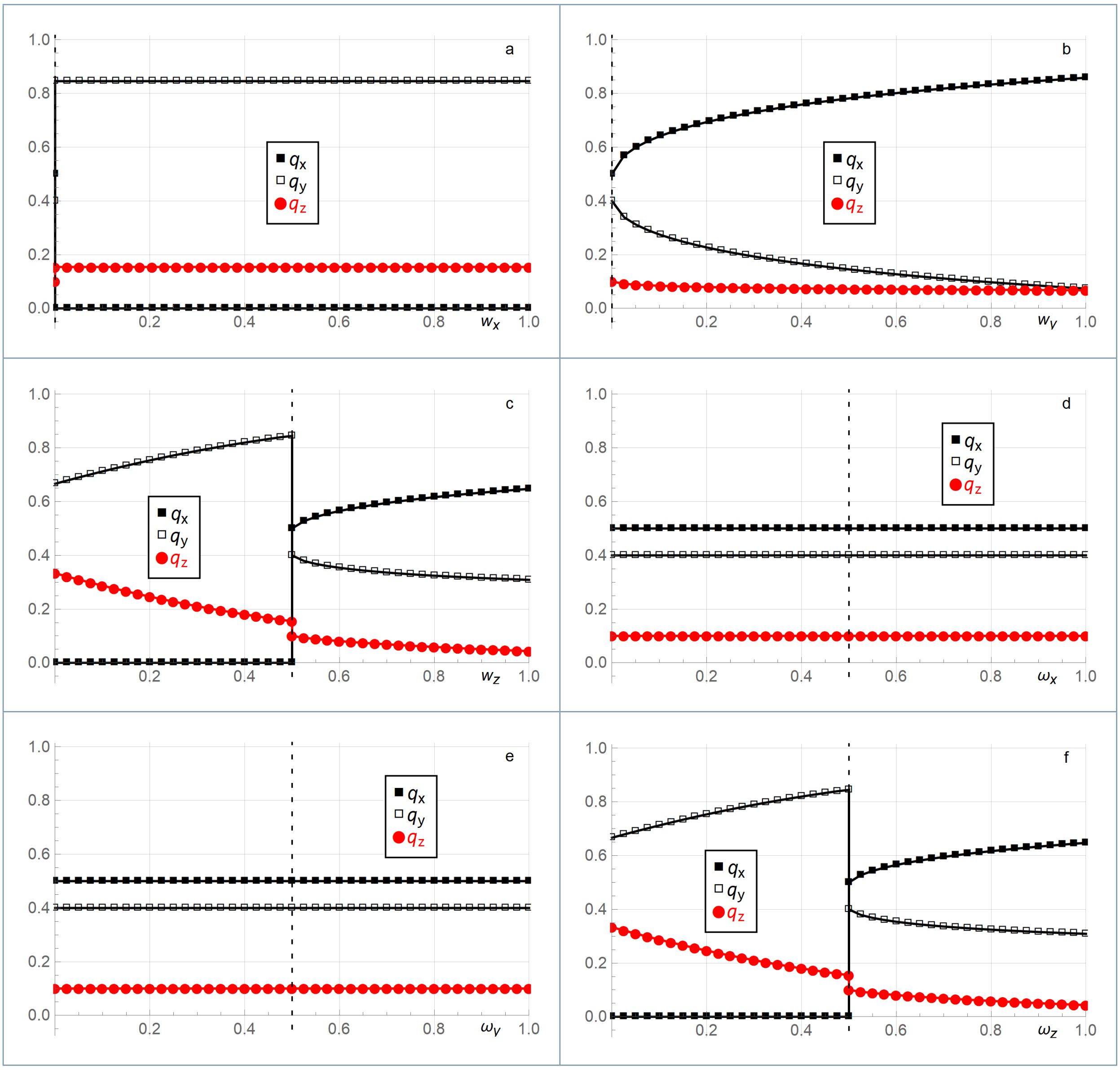
DOI: 10.7172/2353-6845.jbfe.2021.2.6

national currency n, rather than tilting the balance towards the global currency g. The volume fraction p of all players' transactions in the national currency n increases marginally towards $\lim_{t\to\infty} p = 0.820$.

3.3. Sensitivity Analysis

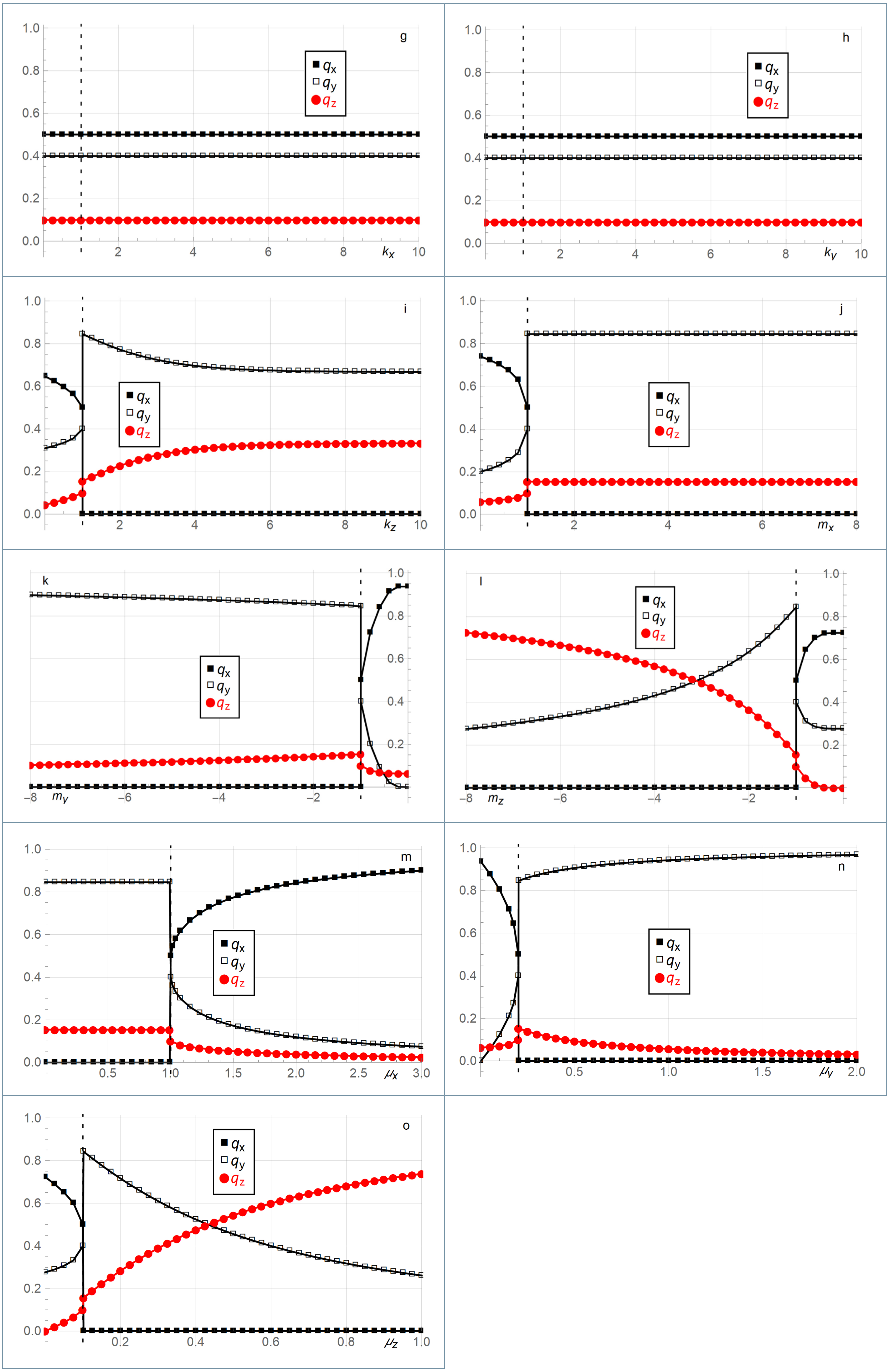
The previous section 3.2 implies a stationary solution after sufficiently much time t has elapsed, i.e. at the limit when $t \to \infty$. This section 3.3 determines the sensitivity of that stationary solution relative to the output subelasticities in Table 1b and the 15 benchmark parameter values in section 3.1, i.e. $w_x = w_y = 0$, $w_z = 0.5$, $\omega_i = 0.5$, $k_i = 1$, $m_x = 1$, $m_y = m_z = -1$, $\mu_x = 1$, $\mu_y = 0.2$, $\mu_z = 0.1$, $i = x, y, z$. We choose Table 1b which has intermediate, compared with Table 1 panels a and c, optimal volume fractions $p_x = 0.75$ for conventionalists and $p_y = p_z = 0.25$ for pioneers and criminals, for player i's transactions in currency n. In Figure 4 each of the 15 parameter values is altered from its benchmark, while the other 14 parameter values are kept at their benchmarks.

Figure 4
Fraction q_i of players of kind i, $i = x, y, z$, as a function of the 15 parameters w_x, w_y, w_z, ω_i, k_i, m_x, m_y, m_z, μ_x, μ_y, μ_z, relative to the benchmark parameter values in Table 1b, $w_x = w_y = 0$, $w_z = 0.5$, $\omega_i = 0.5$, $k_i = 1$, $m_x = 1$, $m_y = m_z = -1$, $\mu_x = 1$, $\mu_y = 0.2$, $\mu_z = 0.1$, $i = x, y, z$, assuming the stationary solution, i.e. after sufficiently much time t has elapsed, in section 3.2.



DOI: 10.7172/2353-6845.jbfe.2021.2.6

Figure 4 (cont.)



DOI: 10.7172/2353-6845.jbfe.2021.2.6

In our benchmark from the previous section 3.2, Figure 3b based on Table 1b determines the stationary solution $\lim_{t\to\infty} q_x = 0.5$ for conventionalists, $\lim_{t\to\infty} q_y = 0.4$ for pioneers, and $\lim_{t\to\infty} q_z = 0.1$ for criminals, after sufficiently much time t has elapsed, depicted with a dashed vertical line in the 15 panels in Figure 4. As each parameter value varies, the stationary solution, hereafter for simplicity referred to as q_x, q_y, q_z, varies from $q_x = 0.5$, $q_y = 0.4$, $q_z = 0.1$ to some other values.

In Figure 4a, as the fraction w_x of conventionalists' transactions which is criminal increases above the benchmark $w_x = 0$, causing conventionalists to risk detection and prosecution if transacting criminally, the fraction q_x of conventionalists decreases from $q_x = 0.5$ to $q_y = 0$, which means extinction, due to lower expected utility. Pioneers and criminals benefit from increasing w_x. As w_x increases above $w_x = 0$, the fraction q_x of pioneers increases from $q_y = 0.4$ to $q_y = 0.85$, and the fraction q_z of criminals increases from $q_z = 0.1$ to $q_z = 0.15$, due to higher expected utilities. The fractions q_x, q_y, q_z, remain constant for $0 < w_x \leq 1$ since w_x impacts only conventionalists' expected utility, and not pioneers' and criminals' expected utilities.

In Figure 4b, as the fraction w_y of pioneers' transactions which is criminal increases above the benchmark $w_y = 0$, causing pioneers to risk detection and prosecution if transacting criminally, the fraction q_y of pioneers decreases convexly from $q_y = 0.4$ to $q_y = 0.07$ when $w_y = 1$, while the fraction q_z of criminals decreases marginally and convexly from $q_z = 0.1$ to $q_z = 0.07$ when $w_y = 1$. Conventionalists benefit from increasing w_y. As w_y increases above $w_y = 0$, the fraction q_x of conventionalists increases concavely from $q_x = 0.5$ to $q_x = 0.86$ when $w_y = 1$.

In Figure 4c, as the fraction w_z of criminals' transactions which is criminal increases above the benchmark $w_z = 0.5$, the fraction q_z of criminals decreases convexly from $q_z = 0.1$ to $q_z = 0.04$ when $w_z = 1$, while the fraction q_y of pioneers decreases convexly from $q_y = 0.4$ to $q_y = 0.31$ when $w_y = 1$. That is because criminals and pioneers do not benefit when they or their criminal transactions become more numerous, cf (4) when $m_y = m_z = -1$ and $m_x = 1$. Conventionalists benefit from increasing w_z, while criminals and pioneers do not. As w_z increases above $w_z = 0.5$, the fraction q_x of conventionalists increases concavely from $q_x = 0.5$ to $q_x = 0.65$ when $w_z = 1$. In contrast, as w_z decreases below $w_z = 0.5$, criminals benefit from their criminal transactions becoming less numerous. That causes the expected utility U_x for conventionalists to be lower than U_y and U_z for pioneers and criminals, $U_x < U_y$ and $U_x < U_z$, regardless of the fraction q_x of conventionalists. That is economically detrimental for conventionalists. In such circumstances no one wants to be a conventionalist. Hence $q_x = 0$ when $w_z < 0.5$. That gives a sudden downward jump in q_x, and hence upward jumps in q_y and q_z as all the three kinds of players adapt to the disappearance of conventionalists who cannot justify their low expected utility U_x. Hence, when $w_z < 0.5$, the replicator equation in (7) strikes a balance between the fractions q_y and q_z of pioneers and criminals, which are $q_y = 0.85$ and $q_z = 0.15$ when $w_z = 0.5 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small but positive, thus excluding conventionalists. As w_z decreases below $w_z = 0.5$, the fraction q_z of criminals increases convexly from $q_z = 0.15$ to $q_z = 0.33$ when $w_z = 0$, while the fraction q_z of pioneers decreases concavely from $q_y = 0.85$ to $q_y = 0.67$ when $w_z = 0$.

In Figure 4d, as the probability ω_x that the government detects and prosecutes conventionalists' criminal behavior changes from the benchmark $\omega_x = 0.5$, the fractions $q_x = 0.5$, $q_y = 0.4$, $q_z = 0.1$ of conventionalists, pioneers and criminals remain constant and unchanged since ω_x in (5) is multiplied with the benchmark fraction $w_x = 0$ of conventionalists' transactions which is criminal. Since $w_x = 0$, ω_x has no impact.

In Figure 4e, analogously, as the probability ω_y that the government detects and prosecutes pioneers' criminal behavior changes from the benchmark $\omega_y = 0.5$, the fractions $q_x = 0.5$, $q_y = 0.4$, $q_z = 0.1$ of conventionalists, pioneers and criminals remain constant and unchanged since ω_y in (5) is multiplied with the benchmark fraction $w_y = 0$ of pioneers' transactions which is criminal. Since $w_y = 0$, ω_y has no impact.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Figure 4f, where the probability ω_z that the government detects and prosecutes the criminals' criminal behavior varies, is equivalent to Figure 4c since $k_z = 1$ in (5), and thus varying ω_z has the same impact as varying the fraction w_z of the criminals' transactions which is criminal, acknowledging that both parameters are restricted to the same interval, $0 \le \omega_z, w_z \le 1$ and have the same benchmark values $\omega_z = w_z = 0.5$. As in Figure 4c, as $w_z < 0.5$ so that the fraction w_z of the criminals' transactions which is criminal decreases below the benchmark $w_z = 0.5$, conventionalists cannot justify their existence due to their low utility $U_x < U_y$ and $U_x < U_z$, and hence $q_x = 0$.

In Figure 4g, as the scaling exponent k_x for what conventionalists retain after criminal behavior changes from the benchmark $k_x = 1$, the fractions $q_x = 0.5$, $q_y = 0.4$, $q_z = 0.1$ of conventionalists, pioneers and criminals remain constant and unchanged since k_x in (5) is an exponent where the base $w_x = 0$ of the conventionalists' transactions which is criminal. Since $w_x = 0$, k_x has no impact.

In Figure 4h, as the scaling exponent k_y for what pioneers retain after criminal behavior changes from the benchmark $k_y = 1$, the fractions $q_x = 0.5$, $q_y = 0.4$, $q_z = 0.1$ of conventionalists, pioneers and criminals remain constant and unchanged since k_y in (5) is an exponent with base $w_y = 0$ which expresses the fraction of the pioneers' transactions which is criminal. That is, since $w_y = 0$, k_y has no impact.

In Figure 4i, as the scaling exponent k_z for what criminals retain after criminal behavior increases above the benchmark $k_z = 1$, the expected utility U_x for conventionalists becomes lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $k_z > 1$. Hence conventionalists cannot justify their existence due to $U_x < U_y$ and $U_x < U_z$, just as when $w_z < 0.5$ in Figure 4c and Figure 4f. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals. As k_z increases, the fraction q_y of pioneers increases from $q_y = 0.4$ when $k_z = 1$ to $q_y = 0.85$ when $k_z > 1$, and thereafter decreases convexly towards the same value as when $w_z = 0$ in Figure 4c, or when $\omega_z = 0$ in Figure 4f, i.e. $\lim_{\substack{t\to\infty,\\ k_z\to\infty}} q_y = 0.67$. The fraction q_z of criminals increases from $q_z = 0.1$ when $k_z = 1$ to $q_z = 0.15$ when $k_z > 1$, due to the disappearance of conventionalists, and thereafter increases concavely, due to successful competition with pioneers as k_z increases, eventually reaching the same value as when $w_z = 0$ in Figure 4c, or when $\omega_z = 0$ in Figure 4f, in accordance with the term $\omega_z w_z^{k_z}$ in (5), $\lim_{\substack{t\to\infty,\\ k_z\to\infty}} q_z = 0.33$. In contrast, as k_z decreases below $k_z = 1$, the fraction q_x of conventionalists increases concavely, competing successfully against pioneers and criminals, eventually reaching $q_z = 0.65$ when $k_z = 0$. As k_z decreases below $k_z = 1$, the fractions q_y and q_z of pioneers and criminals decrease convexly towards $q_y = 0.31$ and $q_z = 0.04$ when $k_z = 0$.

In Figure 4j, as the scaling exponent m_x for how conventionalists get increased (since $m_x \ge 0$) expected utility increases above the benchmark $m_x = 1$, the expected utility U_x for conventionalists becomes lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $m_x = 1$. Hence conventionalists cannot justify their existence, just as when $w_z < 0.5$ in Figure 4c and Figure 4f and $k_z > 1$ in Figure 4i. This follows mathematically from (5) where $q_x^{m_x}$ decreases as m_x increases when $0 < q_x < 1$. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals. Since m_x does not impact that balance, the fractions q_y and q_z of pioneers and criminals are constant at $q_y = 0.95$ and $q_z = 0.15$ when $m_x > 1$. In contrast, as m_x decreases below $m_x = 1$, the fraction q_x of conventionalists increases concavely, competing successfully against pioneers and criminals, eventually reaching $q_x = 0.74$ when $m_x = 0$. This also follows mathematically from (5) where $q_x^{m_x}$ increases as m_x decreases when $0 < q_x < 1$. As m_x decreases below $m_x = 1$, the fractions q_y and q_z of pioneers and criminals decrease convexly, eventually reaching, $q_y = 0.2$ and $q_z = 0.06$ when $m_x = 0$.

In Figure 4k, as the scaling exponent m_y for how pioneers get decreased (since $m_y \le 0$) expected utility increases above the benchmark $m_y = -1$, the fraction q_y of pioneers decreases convexly, eventually going extinct, i.e. $q_y = 0$ when $m_y = 0$. This follows mathematically from

DOI: 10.7172/2353-6845.jbfe.2021.2.6

(5) where $q_y^{m_y}$ decreases as m_y increases when $0 < q_y < 1$. As m_y increases above $m_y = -1$, the fraction q_x of conventionalists increases concavely, competing successfully with pioneers and criminals, eventually reaching $q_x = 0.94$ when $m_y = 0$, while the fraction q_z of criminals decreases convexly, eventually reaching $q_z = 0.06$ when $m_y = 0$. In contrast, as m_y decreases below $m_y = -1$, the expected utility U_x for conventionalists is lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $m_y < -1$. Conventionalists then vanish, as in several of the panels above. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals, which are $q_y = 0.85$ and $q_z = 0.15$ when $m_y = -1 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small but positive. As m_y decreases below $m_y = -1 - \varepsilon$, the fraction q_y of pioneers increases concavely, eventually outcompeting criminals, i.e. $\lim_{t\to\infty,\, m_y\to-\infty} q_y = 1$, while the fraction q_z of criminals decreases convexly, eventually going extinct, i.e. $\lim_{t\to\infty,\, m_y\to-\infty} q_z = 0$. This follows mathematically from (5) where $q_y^{m_y}$ increases without bounds as m_y decreases towards minus infinity when $0 < q_y < 1$.

In Figure 4l, as the scaling exponent m_z for how criminals get decreased (since $m_z \le 0$) expected utility increases above the benchmark $m_z = -1$, the fraction q_z of criminals decreases convexly, eventually going extinct, i.e. $q_z = 0$ when $m_z = 0$. This follows mathematically from (5) where $q_z^{m_z}$ decreases as m_z increases when $0 < q_z < 1$. As m_z increases above $m_z = -1$, the fraction q_x of conventionalists increases concavely, competing successfully with pioneers and criminals, eventually reaching $q_x = 0.72$ when $m_z = 0$, while the fraction q_y of pioneers decreases convexly, eventually reaching $q_y = 0.28$ when $m_z = 0$. In contrast, as m_z decreases below $m_z = -1$, the expected utility U_x for conventionalists is lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $m_z < -1$. Conventionalists then vanish, as in several of the panels above. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals, which are $q_y = 0.85$ and $q_z = 0.15$ when $m_z = -1 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small but positive. As m_z decreases below $m_z = -1 - \varepsilon$, the fraction q_z of criminals increases concavely, eventually outcompeting pioneers, i.e. $\lim_{t\to\infty,\, m_z\to-\infty} q_z = 1$, while the fraction q_y of pioneers decreases convexly, eventually going extinct, i.e. $\lim_{t\to\infty,\, m_z\to-\infty} q_y = 0$. This follows mathematically from (5) where $q_z^{m_z}$ increases without bounds as m_z decreases towards minus infinity when $0 < q_z < 1$.

In Figure 4m, as the scaling proportionality parameter μ_x for how conventionalists get increased (since $m_x = 1$) expected utility increases above the benchmark $\mu_x = 1$, the fraction q_x of conventionalists increases concavely, eventually outcompeting pioneers and criminals, i.e. $\lim_{t\to\infty,\, \mu_x\to\infty} q_x = 1$. Thus the fractions q_y and q_z decrease concavely, $\lim_{t\to\infty,\, \mu_x\to\infty} q_y = \lim_{t\to\infty,\, \mu_x\to\infty} q_z = 0$. In contrast, as μ_x decreases below $\mu_x = 1$, the expected utility U_x for conventionalists is lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $\mu_x < 1$. Conventionalists then vanish, as in several of the panels above. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals, which are $q_y = 0.85$ and $q_z = 0.15$ when $\mu_x < 1$.

In Figure 4n, as the scaling proportionality parameter μ_y for how pioneers get decreased (since $m_y = -1$) expected utility increases above the benchmark $\mu_y = 0.2$, the expected utility U_x for conventionalists becomes lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $\mu_y > 0.2$. Conventionalists then vanish, as in several of the panels above. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals. As μ_y increases, the fraction q_y of pioneers increases from $q_y = 0.4$ when $\mu_y = 0.2$ to $q_y = 0.85$ when $\mu_y > 0.2$, and thereafter increases concavely, eventually outcompeting criminals, $\lim_{t\to\infty,\, \mu_y\to\infty} q_y = 1$. The fraction q_z of criminals increases

DOI: 10.7172/2353-6845.jbfe.2021.2.6

from $q_z = 0.1$ when $\mu_y = 0.2$ to $q_z = 0.15$ when $\mu_y > 0.2$, due to the disappearance of conventionalists, and thereafter decreases convexly, due to unsuccessful competition with pioneers, eventually going extinct, $\lim_{\substack{t\to\infty,\\ \mu_y\to\infty}} q_z = 0$. In contrast, as μ_y decreases below $\mu_y = 0.2$, the fraction q_x of conventionalists increases concavely, competing successfully against pioneers and criminals, eventually reaching $q_y = 0.94$ when $\mu_y = 0$. As μ_y decreases below $\mu_y = 0.2$, the fractions q_y and q_z of pioneers and criminals decrease convexly, pioneers eventually going extinct, $q_y = 0$ when $\mu_y = 0$, while criminals enjoy some presence, i.e. $q_z = 0.06$ when $\mu_y = 0$.

In Figure 4o, as the scaling proportionality parameter μ_z for how criminals get decreased (since $m_z = -1$) expected utility increases above the benchmark $\mu_z = 0.1$, the expected utility U_x for conventionalists becomes lower than U_y and U_z for pioneers and criminals, regardless of the fraction q_x of conventionalists, and hence $q_x = 0$ when $\mu_z > 0.1$. Conventionalists then vanish, as in several of the panels above. That causes the replicator equation in (7) to strike a balance between the fractions q_y and q_z of pioneers and criminals. As μ_z increases, the fraction q_y of pioneers increases from $q_y = 0.4$ when $\mu_z = 0.1$ to $q_y = 0.85$ when $\mu_z > 0.1$, and thereafter decreases convexly, eventually being outcompeted by criminals and going extinct, $\lim_{\substack{t\to\infty,\\ \mu_z\to\infty}} q_y = 0$. The fraction q_z of criminals increases from $q_z = 0.1$ when $\mu_z = 0.1$ to $q_z = 0.15$ when $\mu_z > 0.1$, due to the disappearance of conventionalists, and thereafter increases concavely, due to successful competition with pioneers, eventually becoming dominant and excluding pioneers, $\lim_{\substack{t\to\infty,\\ \mu_z\to\infty}} q_z = 1$. In contrast, as μ_z decreases below $\mu_z = 0.1$, the fraction q_x of conventionalists increases concavely, competing successfully against pioneers and criminals, eventually reaching $q_z = 0.72$ when $\mu_z = 0$. As μ_z decreases below $\mu_z = 0.1$, the fractions q_y and q_z of pioneers and criminals decrease convexly, criminals eventually going extinct, $q_z = 0$ when $\mu_z = 0$, while pioneers are present at $q_y = 0.28$ when $\mu_z = 0$.

4. EXPLAINING THE IMPLICATIONS OF THE RESULTS

With the emergence of new currencies, each player's first choice of which volume fractions of its transactions should be in the national currency and the global currency can be expected to become more significant. The player's choice impacts both its utility, society's utility, which currencies gain traction, and which institutions and parts of society benefit from which currencies gain traction. These factors in turn can be expected to impact finance, business, markets and probably monetary policy, especially if no single currency is or becomes dominant within a given country.

Each player's second choice of whether to be a conventionalist, pioneer or criminal also impacts its utility, and impacts how society becomes composed of these three kinds of players. If conventionalists become less numerous, as illustrated for several combinations of parameter values in the previous section, society may evolve to become less conventional, with competition between pioneers and criminals.

The finding that each player's expected utility is inverse U-shaped as a function of the volume fraction of its transactions in each currency challenges each player to assess its identity as a conventionalist, pioneer or criminal. Each player is furthermore challenged to determine the impact of the subelasticities labeled as backing, convenience, confidentiality, transaction efficiency, financial stability, and security on in its Cobb-Douglas expected utility for the two currencies. This amounts to determining whether the inverse U-shape is skewed with a maximum towards the left or the right, and hence which currency should be chosen for the highest fraction of transactions, which may give fluctuations in currency markets.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

5. CONCLUSION

This article analyzes conventionalists, pioneers and criminals choosing between a national currency, e.g. a CBDC (central bank digital currency) or another currency common within a nation, and a global currency, e.g. Bitcoin or Meta's Diem, which may have limited usage within a nation (e.g. for purchases and tax payments), but may offer other possibilities such as application across nations and user autonomy. Conventionalists tend to prefer the national currency, pioneers (early adopters) tend to prefer the global currency, and criminals tend to prefer the global currency if it contributes (e.g. through confidentiality) to not getting caught.

Each player has a Cobb-Douglas utility with one output elasticity for each of the two currencies. Each output elasticity is comprised of six subelasticities, i.e. which kind of backing a currency has from trustworthy actors or systems (e.g. central banks for CBDCs and distributed ledger technology for cryptocurrencies), convenience (e.g. user friendliness), confidentiality (balancing privacy, availability, accessibility, and discrimination), transaction efficiency (low cost, fast speed, affordability, finality), financial stability (e.g. resilience during crises and shocks), and security (e.g. whether funds are safe and not subject to 51% attacks). Each player's expected utility is expanded to account negatively for detection and prosecution of criminal behavior, and accounts for the fractions of the three kinds of players. Conventionalists benefit from the presence of many conventionalists. Pioneers and criminals benefit from the presence of few pioneers and criminals, respectively.

Each player makes two strategic choices to maximize its expected utility, i.e. which volume fraction of its transactions should be in the national currency (causing the remaining fraction to be in the global currency), and what kind of player it should be, i.e. a conventionalist, pioneer or criminal. The first choice becomes increasingly relevant in today's world as we expect players to have easier access to more than one currency. Hence the market share of two currencies may change over time, as illustrated in this article. The first choice depends on which kind of player the player is, but does not depend on the number of players of this kind, and hence does not depend on time. Each player's second choice is what kind of player it should be through time. Hence this second choice depends on time, through replicator dynamics.

Each player's expected utility is inverse U-shaped as a function of the volume fraction of its transactions in the national currency. Hence each player prefers not to rely exclusively on one currency. The expected utility is skewed towards the right (high fraction) for conventionalists, who prefer the national currency, and more so if the conventionalists' six output subelasticities for the national currency are high. The expected utility is skewed towards the left (low fraction) for pioneers and criminals, who prefer the global currency, and more so if the pioneers' and criminals' six output subelasticities for the global currency are high. Three examples are considered for the degree of skewness towards the right and left. Today's financial system increasingly seems to require players to assess whether the various available currencies are characterized by inverse U-shaped expected utilities skewed towards the right or the left. Players more able to assess these inverse U-shapes as functions of volume fractions, and more able to assess whether they are conventionalists, pioneers and criminals, can expect to earn higher expected utilities. Society's expected utility is the weighted sum of each player's expected utility weighted by the fraction of players of each kind.

The replicator equation is used to illustrate the evolution of the fractions of the three kinds of players through time, assuming initial conditions with conventionalists in the majority, and pioneers and criminals in the minority. We illustrate how conventionalists may become more dominant and criminals less dominant through time if all the three kinds of players' expected utilities are skewed towards the right (i.e. prefer the national currency). In contrast, pioneers and criminals may become more dominant and conventionalists may go extinct if all the three kinds of players' expected utilities are skewed towards the left (i.e. prefer the global currency).

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Considering the stationary solution after sufficiently much time has elapsed, the model's sensitivity with respect to 15 parameter values is analyzed. The analysis shows that, typically, conventionalists (which prefer to be in the majority) tend to compete against pioneers and criminals (which prefer to be in the minority). Hence if a change in a parameter value causes the fraction of conventionalists to increase (decrease), the fractions of both pioneers and criminals may decrease (increase). The exception is, of course, when conventionalists are extinct, which is caused by their expected utility being too low, in which case pioneers and criminals compete directly with each other, so an increasing (decreasing) fraction of pioneers causes a decreasing (increasing) fraction of criminals.

As the fraction of a player's transactions which is criminal, or the probability that the government detects and prosecutes the player's criminal behavior, increases, the fraction of that kind of players in the population decreases, causing the fraction of at least one of the other kinds of players to increase. Each player thus responds to incentives, ceasing to be a kind of player with many criminal transactions, and ceasing criminal transactions if these are detected and prosecuted.

As the scaling exponent for what criminals retain after criminal behavior increases, their fraction in the population increases. That also causes the fraction of pioneers to increase, and the fraction of conventionalists to decrease, except when conventionalists are extinct, which occurs when the scaling exponent is high, in which case the fraction of pioneers decreases due to competition with criminals.

As the positive scaling exponent for how the conventionalists get increased expected utility increases, their expected utility decreases causing their fraction in the population to decrease and eventually go extinct. That causes the fractions of pioneers and criminals to increase. As the negative scaling exponents for how pioneers and criminals get decreased expected utilities increase, their expected utilities decrease causing their fractions in the population to decrease and eventually go extinct. That causes the fraction of conventionalists to transition from extinction to increase. This illustrates how economic incentives for conventionalists can make them more numerous.

As the scaling proportionality parameter for how conventionalists get increased expected utility increases, their fraction increases, as they respond to economic incentives, causing the fractions of pioneers and criminals to decrease. As the scaling proportionality parameters for how pioneers and criminals get increased expected utility increase, both their fractions increase, also responding to economic incentives, causing the fraction of conventionalists to decrease. Eventually, conventionalists go extinct, causing more pioneers and fewer criminals if the pioneers' scaling proportionality parameter increases, and more criminals and fewer pioneers if the criminals' scaling proportionality parameter increases.

Future research should compile and assess empirical support for the six kinds of output subelasticities for national and global currencies, the relevance of each output subelasticity, whether other output subelasticities can be envisioned, or whether the focus should be on fewer output subelasticities. Such empirical support should be assessed against which volume fractions players choose for national and global currencies, and which fractions of players choose to be conventionalists, pioneers, and criminals. These assessments should be made over various time periods to determine which factors impact which national and global currencies spread and become dominant, and which currencies decline in relevance and go extinct. For a more extensive dynamic analysis, the parameters such as the 12 output subelasticities may be allowed to depend on time. Various alternatives to the players' expected utilities may be evaluated, with different risk attitudes, and more than three kinds of players may be modeled. Each kind may have different time horizons and different exchange and trading strategies, e.g. many exchanges per day versus few exchanges per decade. More than one national currency may be analyzed, with competition between multiple national and global currencies which may be generalized to national and global assets (e.g. cryptoassets). The impact of competition on inflation, interest rates, etc., may be assessed, and other players such as regulators and governments may be incorporated.

DOI: 10.7172/2353-6845.jbfe.2021.2.6

References

Allen, S., Čapkun, S., Eyal, I., Fanti, G., Ford, B. A., Grimmelmann, J., . . . Zhang, F. (2020). Design choices for central bank digital currency: Policy and technical considerations (National Bureau of Economic Research Working Paper. No.w27634). Cambridge: National Bureau of Economic Research. https://doi.org/10.3386/w27634

Almosova, A. (2018). A note on cryptocurrencies and currency competition (International Research Training Group 1792 Discussion Paper No. 2018-006). Berlin: Technical University Berlin.

Ang, C. (2021). Visualizing the world's population by age group. Retrieved from https://www.visualcapitalist.com/the-worlds-population-2020-by-age/

Asimakopoulos, S., Lorusso, M., & Ravazzolo, F. (2019). A new economic framework: A DSGE model with cryptocurrency (Centre for Applied Macro- and Petroleum Economics Working Paper No. 07/2019). Oslo: BI Norwegian Business School.

Attah, E. (2019). Five most prolific 51% attacks in crypto: Verge, Ethereum Classic, Bitcoin Gold, Feathercoin, Vertcoin. Retrieved on November 5, 2020 from https://cryptoslate.com/prolific-51-attacks-crypto-verge-ethereum-classic-bitcoin-gold-feathercoin-vertcoin/

Benigno, P. (2021). Monetary policy in a world of cryptocurrencies (Centre for Economic Policy Research Discussion Paper No. DP13517). Roma: Luiss Guido Carli University.

Benigno, P., Schilling, L. M., & Uhlig, H. (2019). Cryptocurrencies, currency competition, and the impossible trinity. National Bureau of Economic Research Working Paper Series, (w26214). Cambridge: National Bureau of Economic Research. https://doi.org/10.3386/w26214

Blakstad, S., & Allen, R. (2018). Central bank digital currencies and cryptocurrencies. In FinTech Revolution (pp. 87–112). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-76014-8_5

Caginalp, C., & Caginalp, G. (2019). Establishing cryptocurrency equilibria through game theory. AIMS Mathematics, 4(3), 420–436. https://doi.org/10.3934/math.2019.3.420

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

Duong, M. H., & Han, T. A. (2020). On equilibrium properties of the replicator-mutator equation in deterministic and random games. Dynamic Games and Applications, 10(3), 641–663. https://doi.org/10.1007/s13235-019-00338-8

ElBahrawy, A., Alessandretti, L., & Baronchelli, A. (2019). Wikipedia and cryptocurrencies: Interplay between collective attention and market performance. Frontiers in Blockchain, 2(12). https://doi.org/10.3389/fbloc.2019.00012

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., & Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. Royal Society Open Science, 4(11), https://doi.org/10.1098/rsos.170623

Fernández-Villaverde, J., & Sanches, D. (2019). Can currency competition work?. Journal of Monetary Economics, 106, 1–15. https://doi.org/10.1016/j.jmoneco.2019.07.003

Frankenfield, J. (2021). Lightning network. Retrieved from https://www.investopedia.com/terms/l/lightning-network.asp

Gandal, N., & Halaburda, H. (2016). Can we predict the winner in a market with network effects? Competition in cryptocurrency market. Games, 7(3), 16. https://doi.org/10.3390/g7030016

Hong, E. (2021). How does Bitcoin mining work?. Retrieved from https://www.investopedia.com/tech/how-does-bitcoin-mining-work/

Howarth, J. (2021). How many cryptocurrencies are there in 2021?. Retrieved from https://explodingtopics.com/blog/number-of-cryptocurrencies

Imhof, L. A., & Nowak, M. A. (2006). Evolutionary game dynamics in a Wright-Fisher process. Journal of Banking Regulation, 52(5), 667–681. https://doi.org/10.1007/s00285-005-0369-8

Kelleher, J. P. (2021). Why do Bitcoins have value?. Retrieved from https://www.investopedia.com/ask/answers/100314/why-do-bitcoins-have-value.asp

Khalil, H. K. (2002). Nonlinear systems (3rd ed.). Upper Saddle River, N.J.: Prentice Hall.

Kiff, J., Alwazir, J., Davidovic, S., Farias, A., Khan, A., Khiaonarong, T., . . . Zhou, P. (2020). A survey of research on retail central bank digital currency (International Monetary Fund Working Paper No. 20/104). Washington: International Monetary Fund. https://doi.org/10.5089/9781513547787.001

Lanz, J. A. (2020). These 10 countries lead the world in Bitcoin adoption. Retrieved on November 5, 2020 from https://decrypt.co/41254/these-10-countries-lead-world-bitcoin-adoption

Lewenberg, Y., Bachrach, Y., Sompolinsky, Y., Zohar, A., & Rosenschein, J. S. (2015). Bitcoin mining pools: A cooperative game theoretic analysis. Paper presented at the Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey.

Masciandaro, D. (2018). Central bank digital cash and cryptocurrencies: Insights from a new Baumol-Friedman demand for money. Australian Economic Review, 51(4), 540–550. https://doi.org/10.1111/1467-8462.12304

DOI: 10.7172/2353-6845.jbfe.2021.2.6

Milunovich, G. (2018). Cryptocurrencies, mainstream asset classes and risk factors: A study of connectedness. Australian Economic Review, 51(4), 551–563. https://doi.org/10.1111/1467-8462.12303
Protska, O. (2021a). TOP 10 – The lowest world currencies in 2021. Retrieved from https://fxssi.com/top-10-of-the-weakest-world-currencies-in-current-year
Protska, O. (2021b). TOP 10 – The most stable currencies in the world in 2021. Retrieved from https://fxssi.com/top-10-world-most-stable-currencies
Rahman, A. J. (2018). Deflationary policy under digital and fiat currency competition. Research in Economics, 72(2), 171–180. https://doi.org/10.1016/j.rie.2018.04.004
Rodriguez, S. (2021). You can now get paid in Bitcoin to use Twitter. Retrieved from https://www.cnbc.com/2021/09/23/you-can-now-get-paid-in-bitcoin-to-use-twitter.html
Sapkota, N., & Grobys, K. (2021). Asset market equilibria in cryptocurrency markets: Evidence from a study of privacy and non-privacy coins. Journal of International Financial Markets, Institutions and Money, 74, Article 101402. https://doi.org/10.2139/ssrn.3407300
Sarkar, A. (2021). Salvadorans are now selling 'way more' US dollars to buy Bitcoin. Retrieved from https://cointelegraph.com/news/salvadoreans-are-now-selling-way-more-us-dollars-to-buy-bitcoin
Schilling, L. M., & Uhlig, H. (2019). Currency substitution under transaction costs. AEA Papers and Proceedings, 109, 83–87. https://doi.org/10.1257/pandp.20191017
Szmigiera, M. (2021). World population by age and region 2021. Retrieved from https://www.statista.com/statistics/265759/world-population-by-age-and-region/
Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. Mathematical Biosciences, 40(1), 145–156. https://doi.org/10.1016/0025-5564(78)90077-9
Verdier, M. (2021). Digital currencies and bank competition (Manuscript. 10.2139/ssrn.3673958). Paris: Université Panthéon-Assas Paris 2. https://doi.org/10.2139/ssrn.3673958
Weibull, J. W. (1997). Evolutionary game theory. Cambridge, MA: MIT Press.
White, L. H. (2014). The market for cryptocurrencies. Cato Journal, 35(2), 383–402. https://doi.org/10.2139/ssrn.2538290
Willms, J. (2021). Michael Saylor's Bitcoin Mining Council's first quarterly report. Retrieved from https://www.nasdaq.com/articles/michael-saylors-bitcoin-mining-councils-first-quarterly-report-2021-07-02
World Bank. (2017). The unbanked. Retrieved from https://globalfindex.worldbank.org/sites/globalfindex/files/chapters/2017%20Findex%20full%20report_chapter2.pdf
Zainab Hussain, N., & Balu, N. (2021). Tesla will 'most likely' restart accepting Bitcoin as payments, says Musk. Retrieved from https://www.reuters.com/business/autos-transportation/tesla-will-most-likely-restart-accepting-bitcoin-payments-says-musk-2021-07-21/

DOI: 10.7172/2353-6845.jbfe.2021.2.6