MANAGERIAL ECONOMICS

SEMIANNUAL
vol. 23 no. 1 ■ 2022

Wydział Zarządzania Akademii Górniczo-Hutniczej

EKONOMIA
MENEDŻERSKA

PÓŁROCZNIK

tom 23 nr 1 ■ 2022

WYDAWNICTWA AGH

KRAKÓW 2022

Faculty of Management
AGH University of Science and Technology

MANAGERIAL ECONOMICS

SEMIANNUAL
vol. 23 no. 1 ■ 2022

AGH UNIVERSITY PRESS

KRAKOW 2022

Editor-in-Chief of AGH University of Science and Technology Press: *Jan Sas*

Managerial Economics Periodical Editorial Board

Editor-in-Chief:
Henryk Gurgul (AGH University of Science and Technology in Krakow, Poland)

Associate Editors:
Joanna Duda (AGH University of Science and Technology in Krakow, Poland)
management
Stefan Schleicher (University of Graz, Austria)
quantitative methods in economics and management

Scientific Editor:
Łukasz Lach (AGH University of Science and Technology in Krakow, Poland)

Technical Editor:
Paweł Zając (AGH University of Science and Technology in Krakow, Poland)

Editorial Board:
Stefan Bojnec (University of Primorska, Koper, Slovenia)
Gerrit Brösel (University of Hagen, Germany)
Karl Farmer (University of Graz, Austria)
Piotr Górski (AGH University of Science and Technology in Krakow, Poland)
Jessica Hastenteufel (IUBH University of Applied Sciences, Saarland University, Germany)
Thomas Hering (University of Hagen, Germany)
Ingo Klein (University of Erlangen-Nuremberg, Germany)
Mariusz Kudełko (AGH University of Science and Technology in Krakow, Poland)
Rafał Kusa (AGH University of Science and Technology in Krakow, Poland)
Ulrike Leopold-Wildburger (University of Graz, Austria)
Manfred Jürgen Matschke (University of Greifswald, Germany)
Xenia Matschke (University of Trier, Germany)
Roland Mestel (University of Graz, Austria)
Stefan Palan (University of Graz, Austria)
Peter Steiner (University of Graz, Austria)
Marcin Suder (AGH University of Science and Technology in Krakow, Poland)

Articles published in the semiannual Managerial Economics have been peer reviewed by reviewers appointed by the Editorial Board. The procedure for reviewing articles is described at the following web address: http://www.managerial.zarz.agh.edu.pl

Language Editor: *Aeddan Shaw*

Statistical Editor: *Anna Barańska*

Editorial support: *Magdalena Grzech*

Cover and title page design: *Zofia Łucka*
DOI: https://doi.org/10.7494/manage

© Wydawnictwa AGH, Kraków 2022,
ISSN 1898-1143 (paper version), ISSN 2353-3617 (on-line version)

Number of copies 40. The printed version of the journal is the primary one.

Wydawnictwa AGH (AGH University of Science and Technology Press)
al. A. Mickiewicza 30, 30-059 Kraków
tel. 12 617 32 28, 12 636 40 38
e-mail: redakcja@wydawnictwoagh.pl; http://www.wydawnictwa.agh.edu.pl

CONTENTS

Managerial Economics
2022, Vol. 23, No. 1, pp. 7–26
https://doi.org/10.7494/manage.2022.23.1.7
©2022 Authors. Creative Commons CC-BY 4.0

Philipp Blumenstein*, Robert C. Schmidt**,
Jessica Hastenteufel***

Development of key performance indicators of capital market-oriented entities in the Prime Standard since the introduction of DRS 20

1. Introduction

The group management report is one of the most important publicly accessible instruments for assessing the corporate governance of capital market-oriented businesses (Müller et al. 2012, p. 281). The German Accounting Standard 20 – Group Management Report (DRS 20) was published on 2nd November 2012 and contains the requirements for consolidated management reporting. The DRS 20 substitutes the standards DRS 5 – Risk Reporting, DRS 5-10 – Risk Reporting of Credit and Financial Services Institutions, DRS 5-20 – Risk Reporting of Insurance Companies, and DRS 15 – Management Reporting (Deutsches Rechnungslegungs Standards Committee e. V, 2012, pp. 38–39). According to § 315 of the German Commercial Code (HGB), the DRS 20 applies to all companies that have to prepare a group management report. In this context, the application of this standard to the management report in accordance with § 289 HGB is recommended (Deutsches Rechnungslegungs Standards Committee e. V, 2012, p. 6).

Since the financial year 2013, capital market-oriented companies listed in the Prime Standard have been obliged to present their management system and the performance indicators used in accordance with the requirements of DRS 20 in their management reports. In addition, significant changes in the management

* European Air Transport Leipzig GmbH (Germany), e-mail: philipp.blumenstein@icloud.com
** International University of Applied Sciences (Germany), e-mail: robert.schmidt@iu.org
*** International University of Applied Sciences (Germany), e-mail: jessica.hastenteufel@iu.org, ORCID ID: 0000-0002-4386-5771

ratios must be presented and explained. The purpose of this study is to empirically analyse the use of KPIs of Prime Standard enterprises since the introduction of DRS 20 and to provide important insights into changes regarding the use of key performance indicators. Thus, a total of 1,176 individual annual reports from 168 companies between 2013 and 2019 are incorporated into our analysis.

A similar study by Göck/Dresp analysed annual reports of 145 capital market-oriented companies regarding the key performance indicators used in corporate management but limited itself to one financial year only (Göck, Dresp, 2017, pp. 8–12). Several other studies have mainly dealt with value-based key performance indicators in the annual reports of DAX companies in the past. In contrast to this, this paper focuses – in addition to an empirical presentation of the use of key performance indicators – primarily on the development or modification of key performance indicators and their use. Thus, we focus on the following questions:

- How did the use of key performance indicators of companies listed in the Prime Standard change since the introduction of DRS 20?
- Which key performance indicators are changed most frequently?
- How often do the analysed companies adjust their key performance indicators?

To answer these questions, we examine a large sample of corporate publications to determine, by means of descriptive statistics, a comprehensive picture of possible changes in the management systems regarding the key performance indicators used and to illustrate their development.

2. Methodology

2.1. Scope of the study

The study is based on the publicly available information of Deutsche Börse AG with all companies listed in the Prime Standard. This list contains information from 302 companies that fulfil the requirements of the Prime Standard (as of 1st December 2020).

The selection of the enterprises that are considered for our study consists of two steps. First, a rough distinction of the companies is made based on three criteria. In a second step, more enterprises are excluded if their inclusion would have created significant limitations for the overall analysis; for example, due to missing annual reports or due to an indistinct description of their management systems and key performance indicators.

For the preliminary screening of potentially relevant entities, only companies listed in the Prime Standard are considered, for which the following conditions apply:

- only companies located in Germany,
- that do not belong to the financial sector (i.e., banks or insurance companies),
- that have been constantly listed in the Prime Standard since the year 2013.

These limitations are made to ensure the comparability of the analysed entities. This is especially important with regard to the exclusion of businesses that belong to the financial sector as they have to fulfil additional regulatory requirements that would lower the comparability of the reports.

As a result, 112 companies had to be excluded from the analyses. These include 22 companies headquartered abroad, 45 companies from the financial sector, 42 companies with IPOs after 2013, and three companies that have not been consistently listed in the Prime Standard since the year 2013.

Factors such as unclear descriptions of management systems or key performance indicators significantly determine the second step in our screening process. For this purpose, the publicly available annual reports are considered. To be able to collect data on the performance indicators from these reports, the following criteria must be met:

- continuous presentation of the key performance indicators since the year 2013,
- sufficient description of the management system including the performance indicators.

In this step, 22 more enterprises had to be excluded. In total, the selection process led to the exclusion of 45 percent of all companies listed in the Prime Standard, so that in the end only 168 companies fulfilled all of the relevant criteria and were selected for the evaluation.

With almost 30 percent, the industrial sector is the largest sector, followed by the software industry with nearly 14 percent of the companies analysed. The pharmaceutical and healthcare sector represents the third largest industry with close to 12 percent. The technology-, utilities-, consumer-, chemicals- and automobile sectors are almost equally represented in the Prime Standard with six to seven percent of all companies (Fig. 1). All other sectors comprise less than ten companies and thus represent only a minor percentage.

To improve comparability, companies are grouped by size into four equally sized categories. Group 1 companies are the smallest companies and Group 4 companies are the largest ones. Thus, for each business year, the three criteria “market capitalisation”, “revenue”, and “number of employees” are taken from the annual reports.

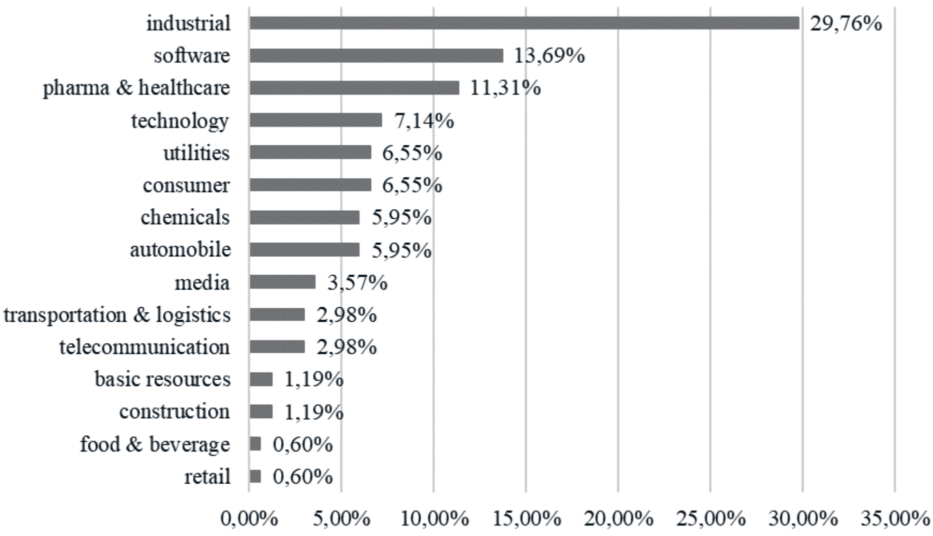

Figure 1. Selected companies classified according to industry sector

For each criterion, the first step is to compile a ranking for each year. These rankings are then averaged across the three characteristics in such a way that each criterion is considered equally important for the final ranking. For example, a company that has the highest market capitalisation in the year 2013 (and is therefore ranked first according to this criterion) but is only ranked fourth and seventh for turnover and number of employees, receives a fourth rank on average for the year 2013. By combining the three criteria and averaging them, we can better compare the three characteristics and thus the size of the organisations. Based on this, rankings in terms of company size are created for the years 2013 and 2019. Moreover, an average is calculated for the entire study period. In this way, a consistent presentation of the research results is ensured for the entire period under review.

2.2. Identifying key performance indicators in management reports

Analysing management reports requires a considerable amount of time and effort, as all data must be analysed manually. Another complicating factor is that there are no uniform guidelines on how a management system and performance indicators must be disclosed. As a result, different presentations of management systems must be evaluated. The differences range from companies that report about their management system only in a short paragraph (e.g., TELES AG Informationstechnologien, 2019, p. 10) to businesses that describe their management system in detail on several pages (e.g., Fraport AG 2019, 67–72). In most cases, the reports are copied from previous years and are only adjusted in the event of changes, so that the reports on the management systems generally resemble each other over the years.

First, all management-relevant performance indicators are compiled. The focus is always on the KPIs that are mentioned in the management system of a company. Basically, a KPI is always attributed to the respective business year in which it is mentioned. For enterprises where the financial year differs from the calendar year, the performance indicator is allocated to the subsequent year. The performance indicators of the business year 2013/2014 are thus assigned to the year 2014. To standardise different descriptions for the same key figures and to make them comparable for statistical purposes, the KPIs will initially be summarised. For example, the key indicators "amortisation period" and "payback period" are summarised as payback period.

Throughout the evaluation we noticed that many companies prioritise their key performance indicators and that some indicators seem to be more important than others. Delticom AG, for example, declares two indicators as key performance indicators and explains that in addition to these KPIs subsequent performance indicators are used (Delticom AG, 2014, p. 20). For this reason, all key performance indicators are categorised as main performance indicators or additional performance indicators. This provides the necessary differentiation and draws a complete picture of the performance indicators used. A key performance indicator must be identified in the management report as a company's main performance indicator. Any indicator that is not identified or explicitly declared as a main performance indicator is an additional performance indicator. Conjunctive adverbs such as "in addition", "furthermore" or "moreover" are interpreted as a sign that a KPI is not among a company's most significant performance indicators. If the performance indicators are described without any differentiation, then the indicators are automatically classified as main performance indicators.

2.3. Analysing changes in the management systems

Additionally, changes in the key performance indicators are documented. In this study the annual reports of the business year 2013 (or 2012/2013) are neglected. Usually, a comparison of the management report to the previous year is necessary to identify changes in the use of key performance indicators. Only in some cases changes are mentioned in the reports of the 2013 (or 2012/2013) financial year (e.g., Evonik Industries AG, 2014, p. 41). Thus, the identification of changes in the year 2013 is not possible in most cases, as these changes require the comparison of management reports from the years 2012 and 2013. Due to the introduction of DRS 20 in November 2012, the obligation to describe the management system, including the key performance indicators used, only applies to annual reports from the years 2013 onwards. For this reason, the period from 2014 to 2019 is observed.

The documentation of each adjustment includes the changed performance indicator, the year in which the adjustment was made, the form of the modification, the reason for the modification (if any reason is mentioned), the priority of the performance indicator and its classification as a financial or non-financial indicator. The year in which the adjustment was made is the year in which the change is documented in the annual report. If an annual report explicitly mentions that a change will only take place from the next year onwards, then their change is only documented for the following year. For the sake of transparency, all modifications are classified into one of five categories:

1) the performance indicator is added to the management system,
2) the performance indicator is removed from the management system,
3) the performance indicator is now a main KPI,
4) the performance indicator is no longer a main KPI,
5) the performance indicator is adjusted.

Category 1 includes indicators that were not previously part of the management system or indicators that are mentioned again in the management system after at least one year without being mentioned. Category 2 includes all indicators that are no longer listed in the management system as main or additional performance indicators. Category 3 contains only indicators that were previously listed as additional performance indicators and have since become main performance indicators. Category 4 deals with all key performance indicators that have become less important for the enterprises over time and are thus no longer main key performance indicators. Finally, category 5 includes all adjustments of performance indicators without a shift in prioritisation or an addition to or exclusion from the management system. This mainly includes adjustments due to new regulations under the International Financial Reporting Standards (IFRS) or adjustments due to changed framework conditions.

Moreover, it is documented for the two categories 1 and 2 which priority an introduced or removed performance indicator has for a business. The prioritisation is based on the procedure already described. The following distinctions are made:

- the performance indicator will be or was a main performance indicator,
- the performance indicator will be or was an additional performance indicator.

The most important classification for the following analysis is the assessment of whether a documented change is a verifiable modification of a company's key performance indicators. For this purpose, the change is classified into four categories to answer the question of whether an entity clearly communicates such an adjustment. In this context, every apparent change is checked for plausibility in the annual reports. This procedure is intended to remove unclear changes from the analysis to minimise the bias in the results of the study. It should be stressed that according to the regulations of DRS 20, no justification must be given for changes regarding performance indicators. According to DRS 20.K47, only significant changes in the management system used in a group compared to the previous year must be disclosed. Nevertheless, a distinction should be made between changes with justification and changes without justification but with prior notice (Tab. 1). This is reasonable in the context of evaluating the investor relations of the assessed enterprises.

Table 1

Change categories and essential characteristics of each category

	Change category	Characteristics of the category
Adjustments	justification of the change	a justification for the change is given
	announcement of the change	there is no justification for the change, but the company communicates that a change was made
	plausible change	there is no justification and no company announcement for the implemented change, however, according to the described plausibility check, it can be assumed that a change was made
No adjustments	unclear	there has been a change in the annual report, but this change cannot be verified without further investigation

Furthermore, we will explain how additional changes are identified from the management reports that are not actively communicated by the respective company by an explanation or announcement. It is necessary to form these categories, as otherwise too many changes in the key performance indicators could remain undetected.

However, before a change in a key performance indicator can be assigned to the category “plausible change”, a plausibility check must be performed in addition to a change identified by comparing information from management reports. For this purpose, several parameters must be checked:

- Is the change maintained in the following management reports?
- What information is communicated about the modified performance indicator in the annual reports before and after the change?
- Is there any information about this change in another part of the annual report?
- Does the enterprise justify changes for other key performance indicators?

Depending on the case, additional, individual, and specific data must be examined. Any information that is considered in a follow-up evaluation and that is not verified may lead to an unclear change in classification.

However, if a single indicator is changed several times within the period under review between the years 2014 and 2019 without any justification or announcement, this does not mean that this change is automatically considered ‘unclear’. Here, too, more information must be reviewed, and a subjective decision must be made as to whether these modifications should be included in the evaluation. However, it cannot be excluded that despite this careful review, changes are identified as plausible changes although they are not, or vice versa. This must be kept in mind especially when considering absolute values. For this reason, we focus on relative results when considering changes. Due to the large number of changes that could be analysed, we consider the results to be reliable despite the described limitations and that they allow additional analyses to be carried out.

3. Results of the empirical study

3.1. Introductory overview

The assessment of the key performance indicators is performed in two steps. In the first step, the performance indicators of the examined Prime Standard companies from the years 2014 and 2019 are compared to identify possible changes since the introduction of the DRS 20. For this purpose, the main features of the management systems and the key performance indicators of these years are identified. These include the number of performance indicators in the management systems as well as other statistical characteristics of the use of performance indicators. The most frequent KPIs in the years 2014 and 2019 are then compared to each other. Finally, the adjusted and non-adjusted indicators are aggregated. The management systems of the years 2015 to 2018 are not described in detail, as changes in these management systems can be shown better by presenting the actual modifications. This is because if, for example, the KPI "EBIT" is no longer used by one company in one year but is newly introduced by another enterprise in that same year, then the number of EBIT key figures used would not change, although there is a change in two companies. The comparison of management systems from the years 2014 and 2019, however, could reveal fundamental changes. In addition, the status quo of the key performance indicators used by companies listed in the Prime Standard will be described. In a second step, the changes in the performance indicators will be evaluated to gain a detailed insight into the development of the performance indicators of the businesses examined. This is to identify changes that cannot be observed by comparing single performance indicators directly.

3.2. Management systems of the years 2014 and 2019

3.2.1. Introduction

In total, 482 different indicators are used by the 168 companies examined. However, despite the different descriptions, some of these indicators measure the exact same thing and can therefore be aggregated to 267 different indicators in total. The differences between the individual companies in the number of KPIs used are huge, and ranges from one (Bastei Lübbe AG 2014) to 25 used key performance indicators (adidas AG 2020).

3.2.2. Essential characteristics of the management systems

The comparison of the main structural characteristics of the management systems in the years 2014 and 2019 does not show any significant changes (Tab. 2).

The management systems became marginally larger, i.e., on average three percent more key performance indicators were used in 2019 than in 2014. The number of companies using additional performance indicators also increased from 48 percent to 58 percent. Among other things, this could result from the small increase in the number of key performance indicators overall, which also leads to more companies making a distinction between main and additional key performance indicators. The number of companies using non-financial performance indicators increased from 37.5 percent in 2014 to 41.7 percent in 2019.

However, the average number of non-financial performance indicators used by these companies remained almost the same.

Additionally, no significant differences could be identified in the individual sectors regarding the development of the use of indicators. The problem in presenting the sectors is the usually small number of companies in the individual sectors, which means that a comparison of the sectors is only of limited use.

Table 2

Comparison of the key characteristics of the key performance indicators of the years 2014 and 2019

		2014	2019	Trend
All KPIs (main + additional)	**average, total**	**6.7**	**6.9**	**+**
	median, total	**6**	**6**	**o**
	average, financial KPI	5.6	5.7	+
	median, financial KPI	5	5	o
	average, non-financial KPI	1.1	1.2	+
	average, non-financial KPI, adjusted[1]	2.9	2.8	–
	median, non-financial KPI, adjusted	2	2	o
Main KPIs	**average, total**	**4.6**	**4.4**	**–**
	median, total	**4**	**4**	**o**
	average, financial KPI	4.1	3.8	–
	median, financial KPI	4	3	–
	average, non-financial KPI	0.5	0.6	o
	average, non-financial KPI, adjusted	2.7	2.7	o
	median, non-financial KPI, adjusted	2	2	o
Additional KPIs	average, total	2.1	2.5	+
	average, financial KPI	1.5	1.9	+

[1] The addition “adjusted” indicates that the values only relate to those companies to which the corresponding classification applies.

Table 2 cont.

	average, non-financial KPI	0.6	0.6	o
	companies using additional KPIs	**47.6%**	**57.7%**	+
	average, total, adjusted	**4.4**	**4.5**	+
	median, total, adjusted	**3**	**3**	o
	average, financial KPI, adjusted	3.5	3.7	+
	average, non-financial KPI, adjusted	2.9	2.8	–
Companies using non-financial KPIs	**total**	**37.5%**	**41.7%**	+
	companies using non-financial KPIs as main KPIs	**20.2%**	**22.0%**	+
	companies using non-financial KPIs as additional KPIs	**20.2%**	**21.4%**	+
	companies using **only** non-financial KPIs as main KPIs	17.3%	20.2%	+
	companies using **only** non-financial KPIs as additional KPIs	17.9%	19.7%	+
	companies using non-financial KPIs as main **and** additional KPIs	2.3%	1.8%	–

3.2.3. Comparison of the key performance indicators

The comparison of the most frequent performance indicators of the years 2014 and 2019 provides some insights that were examined by analysing the changes. There are changes regarding the relative use of the five most common key performance indicators in the years 2014 and 2019. Overall, the KPI “revenue” is used approximately five percent more frequently, the KPI “free cash flow” ten percent more frequently, the KPI “ROCE” three percent more frequently and the KPI “EBITDA” five percent more frequently. Only the use of the KPI “EBIT” decreases by five percentage points. When analysing the individual groups of companies according to size, we noticed that the key figures “ROCE” and “free cash flow” are mainly used by larger capital market-oriented enterprises. For companies in group 4, the use of ROCE has increased by nine percentage points since the year 2014. Finally, Figure 2 illustrates the most common main and additional key performance indicators for the years 2014 and 2019.

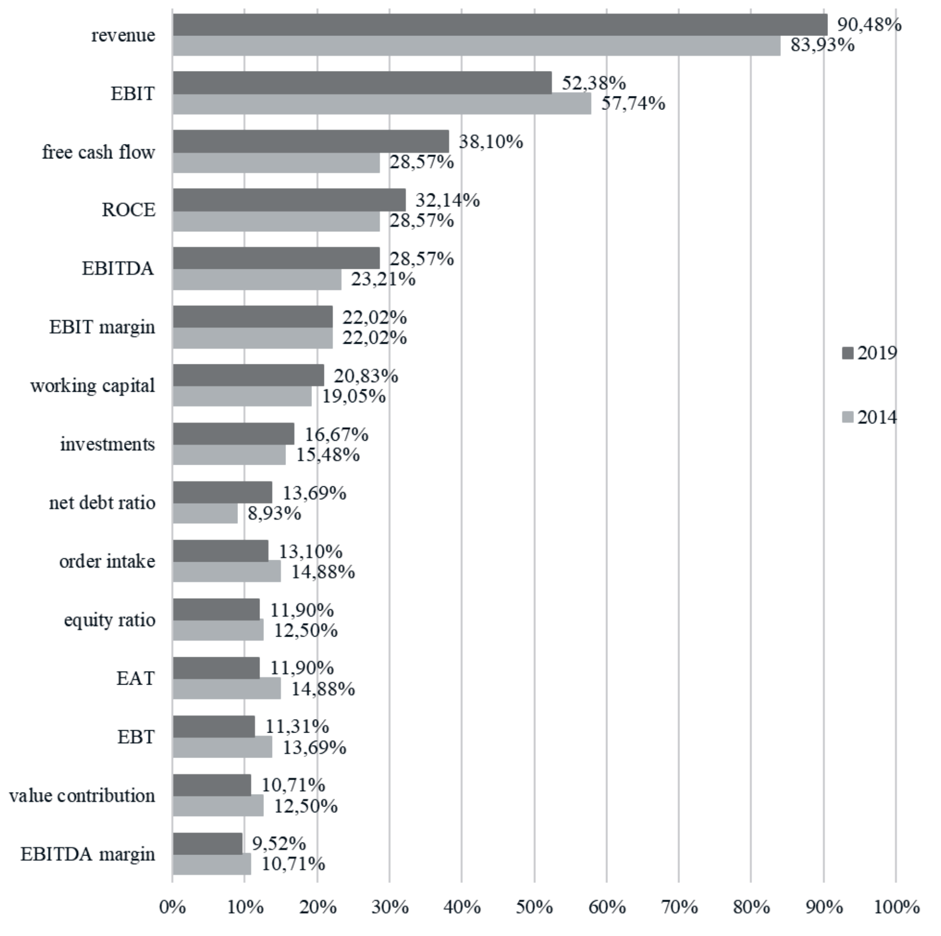

Figure 2. Illustration of the main and additional financial KPIs of the management systems in the years 2014 and 2019

3.3. Changes in the key performance indicators

3.3.1. Essential characteristics of the modifications

For 145 out of 168 companies, we were able to identify a total of 804 potential changes in the key performance indicators, of which only 557 changes from 125 enterprises are plausible. Consequently, the plausibility check leads to the exclusion of the modifications in 30.7 percent of all identified modifications and thus they are not taken into consideration for the analysis. 36 companies have intentionally or unintentionally disclosed unclear changes in their management reports. In this context, 20 companies use a purely qualitative presentation of their management systems and 16 use a mix of tables and qualitative presentation. It can therefore not be stated without further reflection that a qualitative method of presentation favours the ambiguities in the modifications.

During the evaluation, a rate of change is defined, which represents the average number of changed indicators per company and year. In relation to the 125 companies whose key performance indicator changes are evaluated, this value is 0.74. It shows the dynamics in the change of key performance indicators. 34 percent of the companies modified their performance indicators between one and three times in the period under review, and about 21 percent of the businesses changed their management system between four and six times during this period. Only 20 percent of the companies analysed changed their performance indicators more than six times.

Furthermore, it cannot be confirmed that companies with larger management systems change their KPIs more frequently than those with smaller management systems. The annual change rates of the companies studied are presented over the average size of the respective management systems. The average size of a management system is the average number of key performance indicators of the individual companies from the years 2014 to 2019. All explicit changes, both for main and additional key performance indicators, are considered. When analysing the graphs, no correlation can be found in the data. This is confirmed by the calculation of the Pearson's correlation coefficient. It is 0.31 and thus shows a weak linear correlation.

Figure 3 shows how the modifications are categorised. Only the three categories of changes that can be identified as such according to the plausibility check are included. 54 percent of these changes were not communicated by the companies, which means they were neither justified nor announced. For 29 percent of the modifications there was a justification and for 17 percent there was at least an announcement by the company, but without a corresponding justification. In this context, 88 percent of the changes were made to financial KPIs. Only 12 percent of the modifications related to non-financial performance indicators.

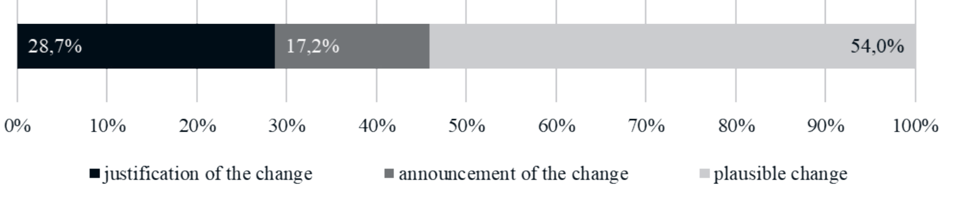

Figure 3. Change categories without "ambiguous" changes

Furthermore, the type of change in the KPIs is presented. A total of 42 percent of the changes relate to performance indicators that were added to the management systems. In contrast, 34 percent of the changes relate to performance indicators that were removed from the management systems. In total, 76 percent of the modifications analysed relate to performance indicators that were either added to or removed from the management systems. The remaining 4 percent of the changes relate to indicators that have already been part of the management systems. Modifications of the indicators occurred in 11 percent of the amendments. In 8 percent of the cases, key performance indicators that were previously used as main KPIs were downgraded to additional KPIs. However, these indicators remain part of the management system, but with a lower priority for the respective companies. Moreover, 5 percent of the modifications involve upgrading an additional performance indicator to a main KPI (Fig. 4).

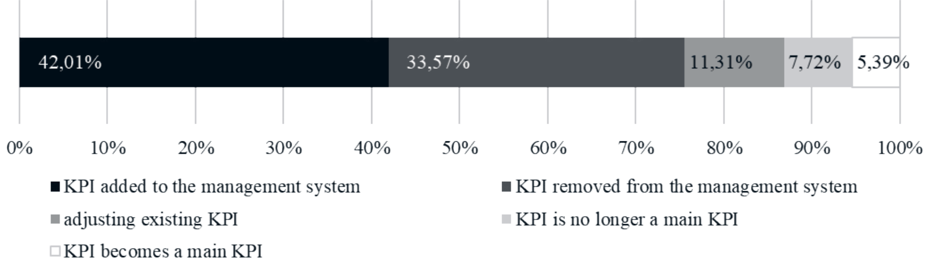

Figure 4. Classification of changes in the KPIs

In addition, the changes will be analysed more closely where key performance indicators were either added to or removed from the management systems. This is the case for about 76 percent of all changes, as shown in Figure 4. Figure 5 shows that slightly more key performance indicators were added to the management systems than were removed from them, with 39 percent of the changes affected. The percentage of KPIs that were added to or removed from the management systems as additional KPIs is about 12 percent.

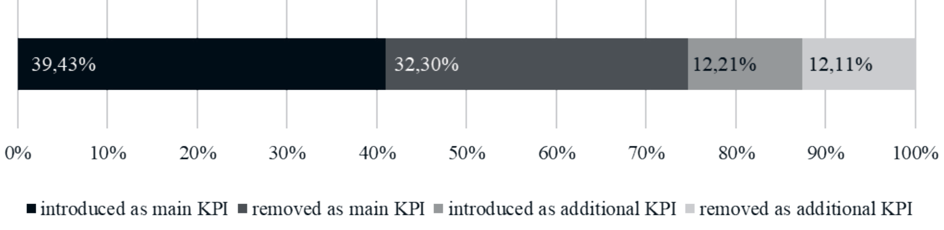

Figure 5. Classification of modifications to KPIs removed or added to the management system

Finally, we analysed the changes in relation to the size of a company. Figure 6 shows the percentage of justified or announced changes as well as the percentage of plausible changes in the groups 1–4 sorted by size. While 61 percent of the 181 documented changes in group 4 were disclosed, only 21 percent of the 106 documented changes in group 1 were disclosed. The number of companies that continuously communicate their changes increased steadily with the size of the enterprise. Group 2 companies report on implemented changes in about 43 percent of the changes examined. For companies in Group 3, the percentage is somewhat higher at 48 percent of the changes communicated.

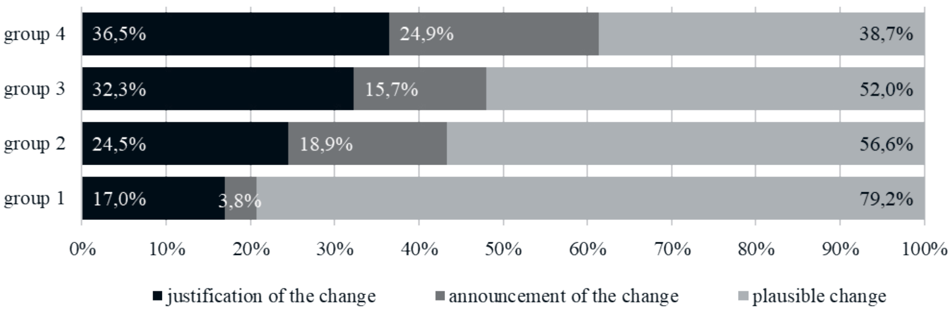

Figure 6. Change categories by company size

3.3.2. The most frequently changed KPIs

In total, 136 different KPIs were changed by 125 companies between the years 2014 and 2019. The analysis of the modifications shows that EBIT, EBITDA, ROCE, and free cash flow are the most frequently changed key performance indicators. About every fourth reported adjustment relates to one of these four ratios. However, revenue was also changed often. As in the presentation of the management systems, adjusted and non-adjusted KPIs are considered together.

A detailed analysis of the changes of the key figures “EBIT”, “EBITDA”, “ROCE”, and “free cash flow” confirms the identified changes by comparing the key performance indicators. Out of these ratios, EBIT is the only one companies were using less. We noticed that EBIT was mostly replaced by other earnings ratios. Depending on the current framework conditions and investment goals, enterprises seem to switch between different earnings ratios in order to be able to present the current business situation as advantageously as possible with the respective ratios.

For the KPIs “EBITDA”, “ROCE” and “free cash flow”, the trend observed by comparing the performance indicators can also be confirmed. All three KPIs were used more frequently in 2019 when compared to 2014. The KPI “EBITDA” showed certain parallels to the KPI “EBIT”, as it usually replaces other earnings figures or is replaced by them, too. The ROCE continued to have a strong influence on the management systems of the companies studied and was used more frequently, especially by group 4 companies. The free cash flow frequently replaced other cash flow figures and was also used more frequently by smaller companies in 2019 than in 2014. The changes in revenue were not evaluated due to the focus of the analysis (Fig. 7).

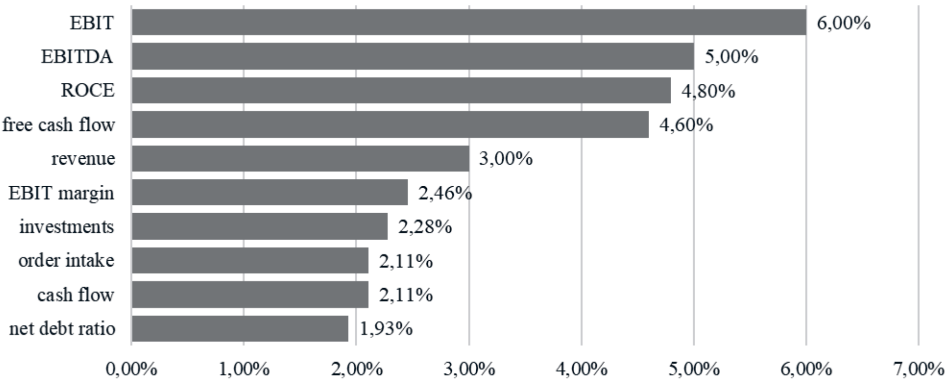

Figure 7. Most frequently changed KPIs between 2014 and 2019

4. Recommendations

The assessment of the annual reports and the extraction of the key performance indicators as well as their changes were faced with difficulties and obstacles. Due to the different ways of disclosing information about the management systems and the sometimes considerable differences in quality between the individual companies, the analysis was time-consuming and is subject to a few assumptions. The overall very low percentage of enterprises with a distinct communication of modifications confirms the urgent need for a standardised presentation of key performance indicators and their modifications, which will now be discussed in detail.

To present the performance indicators and their changes more transparently, we develop a re- commendation for action for a standardised method of presentation. Due to the high number of companies that make a distinction between main and additional performance indicators, there should first be a standardised, table-based presentation of both the main performance indicators and possible additional performance indicators. The results have shown that this prioritisation of key performance indicators is mostly implemented by companies with management systems of above-average size. The management systems of the companies that disclose additional key performance indicators are, with an average of 8.4 KPIs, about 1.7 KPIs larger than the average. Due to the large number of indicators, it makes sense for businesses to subdivide the indicators into main and additional indicators to keep the management systems as clear as possible. However, this distinction is not regulated in DRS 20, which means that there is currently no obligation to implement it.

For standardisation, it would generally be helpful if such a differentiation were to become mandatory. Each company could decide for itself which indicators are to be classified as more or less important or whether this differentiation is necessary at all. The identified and classified key performance indicators could be presented in a table, for example, to minimise the scope for interpretation.

Furthermore, it is recommended that the changes themselves should also be presented in a table. Currently, many businesses, especially smaller ones, do not comply with the obligations to present changes as regulated in DRS 20.K47 or cause confusion with a non-transparent presentation of the management system. By presenting the key performance indicators in a table, including a description of the changes, the information asymmetry between a company and its stakeholders could be reduced. A qualitative description of management systems should by no means be omitted, but the representation in the form of a table could be seen as a mandatory supplement.

DRS 20.K47 refers to the disclosure of so-called "significant changes". It is recommended to replace the wording with "any changes in the key performance indicators in the management system" to reduce the scope for interpretation at this point as well. These changes in the interpretation of DRS 20 could lead to changes in the management systems being communicated more transparently. Justifications for changes can still be communicated by the enterprises but should not be mandatory and should not be included in DRS 20.K47. Finally, Table 3 illustrates a possible way of presenting the KPIs and their changes in management reports.

Table 3

Potential, standardised presentation of key performance indicators in a management report

<table>
<tr><th>Main KPIs</th><th>Additional KPIs</th></tr>
<tr><td>– revenue
– EBIT
– ROCE</td><td>– employee engagement
– investments
– free cash flow</td></tr>
<tr><td colspan="2">Changes from the previous year</td></tr>
<tr><td colspan="2">– The free cash flow is no longer a main key performance indicator, but it is still used as an additional key performance indicator.
– The performance indicator ROCE is added to the management system and is a new main key performance indicator.</td></tr>
</table>

5. Conclusion

This paper provides insights into the management systems of companies listed in the Prime Standard with regard to the development of its key performance indicators since the introduction of DRS 20. They also show that there are various weaknesses in the implementation of DRS 20 with regard to the presentation of key performance indicators and the communication of their modifications. In total, only 46 percent of the changes in the performance indicators assessed are communicated accordingly by the companies. The larger the company, the more frequently changes in the performance indicators are communicated directly.

Based on a detailed plausibility check of observed changes in KPIs, recommendations for action are presented for an adjustment of DRS 20 aimed at a standardised presentation of the key performance indicators and their changes. Such a presentation would reduce the information asymmetry between a company and its stakeholders and improve the transparency of group management reports.

References

[1] adidas AG (2020) 2019 Annual Report, [Online], Available: https://report.adidas-group.com/2019/en/servicepages/downloads/files/adidas_annual_report_2019.pdf [22 Feb 2022].

[2] Bastei Lübbe AG (2014) Business Report 2013/2014, [Online], Available: https://www.luebbe.com/web/downloads/module/download/2604150/Annual_Report_Fiscal_Year_2013_2014.pdf [22 Feb 2022].

[3] Delticom AG (2014) Annual Report 2013, [Online], Available: https://www.delti.com/Investor_Relations/Delticom_AnnualReport_2013.pdf [22 Feb 2022].

[4] Deutsche Börse AG (2020) Listed companies, [Online], Available: https://www.deutsche-boerse-cash-market.com/dbcm-en/instruments-statistics/statistics/listes-companies [22 Feb 2022].

[5] Deutsches Rechnungslegungs Standards Committee e. V. (2012) Deutscher Rechnungslegungs Standard Nr. 20 (DRS 20) – Konzernlagebericht, [Online], Available: https://www.drsc.de/app/uploads/2017/02/120928_DRS_20_near-final.pdf [22 Feb 2022].

[6] Evonik Industries AG (2014) Annual Report 2013, [Online], Available: https://corporate.evonik.de/downloads/corporate/evonik-annual-report-2013.pdf [22 Feb 2022].

[7] Fraport AG (2019) Annual Report 2018, [Online], Available: https://www.fraport.com/content/dam/fraport-company/documents/investoren/eng/publications/annual-reports/annual-report-2018.pdf/_jcr_content/renditions/original.media_file.download_attachment.file/annual-report-2018.pdf [22 Feb 2022].

[8] Gitt, N., Völl, W. and Kettenring, T. (2013) 'Anwendung wertorientierter Steuerungskennzahlen in deutschen HDAX-Unternehmen: Aktueller Stand und Entwicklungen', *Controlling – Zeitschrift für erfolgsorientiere Unternehmenssteuerung*, vol. 25, pp. 100–108.
[9] Göck, M. and Dresp, M. (2017) 'Steuerung mit Kennzahlen: die Qual der Wahl im Zahlendschungel', *Corporate Finance*, vol. 8, pp. 8–12.
[10] Müller, S., Juchler, D. and Ergün, I. (2012) 'Risikoberichterstattung im Konzernlagebericht nach DRS 20: Verbesserungspotenziale und weiter bestehende Qualitätsrisiken', *Zeitschrift für Corporate Governance*, vol. 7, pp. 281–287.
[11] Rupp, R. and Haberstumpf, E. (2018) 'Verbreitung und Ausgestaltung der wertorientierten Berichterstattung', *Der Betrieb*, vol. 71, pp. 2129–2134.
[12] TELES AG Informationstechnologien (2019) Annual Financial Statement 2018, [Online], Available: https://www.teles.com/uploads/media/TELES_-_Annual_Financial_Statement_2018.pdf [22 Feb 2022].

Summary

This paper examines the performance indicators of 168 Prime Standard entities since the introduction of DRS 20, focusing on the core question of how the use of performance indicators has changed over time. For this purpose, we compare the published key performance indicators from various companies in different years to point out existing differences. Furthermore, we examine which KPIs are changed most often and how frequently businesses adjust their performance indicators. The companies examined are differentiated according to size and sector.

JEL classification: G00, G34, M41

Keywords: *DRS 20, Prime Standard, key performance indicators*

Managerial Economics
2022, Vol. 23, No. 1, pp. 27–48
https://doi.org/10.7494/manage.2022.23.1.27
©2022 Authors. Creative Commons CC-BY 4.0

Aneta Kosztowniak*

Changes in the quality of bank loan portfolios in EU countries – with the particular case of Poland[1]

1. Introduction

The size of non-performing loans (NPLs) plays a key role in the stability of banking sector of every country. Rising NPLs are often referred to as a failure of banks to manage credit policy and bank losses. After global crises, NPLs are of interest to banks in connection with asset management as they are considered failures and crises of the banking system (Gosh, 2015). A growing level of non-performing loans in the longer term will affect commercial banks first and then the financial situation of a country's economy (Souza, Feijó, 2011).

According to Handley (2010) and Ivanovic (2016), NPLs affect a country's economic growth by reducing credit development. Low NPLs indicate a strong monetary system, while high NPLs suggest a weak financial situation. The negative impact of NPLs manifests itself in a decline in banking efficiency, causing banking crises (Vouldis and Louzis, 2018). NPLs block interest income, limit new investments, cause liquidity crises in the financial system, resulting in bankruptcy problems and lower economic welfare. For these reasons, it is necessary to identify the factors that influence NPLs so that they do not compromise financial stability (Stijepović, 2014).

In EU countries, including Poland, where the main place of obtaining capital is the banking system, the supervision of NPLs is particularly important (Moradi et al., 2016). NPL statistics confirm the problem in European countries, although

* SGH Warsaw School of Economics, Collegium of Management and Finance, Department of Applied Economics, e-mail: akoszt@sgh.waw.pl, and NBP. ORCID ID: 0000-0001-6088-1899.
[1] The views expressed in this study are the views of the author and do not necessarily reflect those of NBP. Paper prepared under the research project no. SGH/KZiF/S22:1:26 (project title: Financial system versus economic growth and development. Stage 1. Variability of the system and the financial crises; title in Polish: System finansowy a wzrost i rozwój gospodarczy. Etap 1. Zmienność uwarunkowań systemu i kryzysy finansowe).

its scale varies. The share of household loans in the euro area countries accounts for over 60% of total loans and, including over 40% for Poland, and their value is approx. 35% of GDP in 2021 (BIS, 2021). The ratio of outstanding bank loans to total gross loans according to the World Bank (2021) shows significant differences in the banking sectors of the EU countries (e.g., 27.0% Greece, 15.0% Cyprus, 5.8% Bulgaria, 4.9% Portugal, 3.7% Poland, and 1.1% Germany in 2020).

The main aim of this chapter is to identify changes in the quality of bank loan portfolios in the EU countries in 2009–2021 using the example of the Visegrad Group (Czech Republic, Poland, Slovakia, Hungary, V4) as well as France and Germany. Considering the fact that the share of loans to households in EU portfolios is approximately 60%, it has a significant impact on the share of non-performing loans (NPL) in a bank's entire portfolio. Therefore, it is important to identify macroeconomic determinants influencing the creditworthiness of households and their loan servicing capacity.

The specific aims are, first, to present the differences in NPLs, debt servicing costs, and the structure of loan portfolios in the selected EU countries. Second, to identify countries with high-quality portfolios and those undertaking restructuring. Thirdly, to examine the determinants of NPL for household loans based on the example of Poland, i.e., a country considered representative in terms of the average level of NPL and the portfolio structure in the group of countries studied.

The present chapter identifies several macroeconomic factors influencing the NPLs rate in the Polish banking system. We concentrate on macro-level factors but the quality of a loans portfolio also depends on the specificity of each bank and its customers.

2. Literature and research review

The increase in NPLs over the past decade has caught the attention of many academics around the world who have tried to explain the phenomenon. The main reasons for high NPLs are poor credit procedures, weak credit specialists, high margins, low credit rules, and the lack of a borrower monitoring policy. Most authors study changes in NPLs for entire loan portfolios and analyze various factors. However, we can define two main groups of macroeconomic and banking factors.

The following macroeconomic factors are commonly studied: real GDP growth, value of GDP/GDP per capita, the exchange rate, interest rates and the level of inflation. The results confirm that real GDP growth usually translates into a higher level of income, improving the financial standing of borrowers and decreasing NPLs. When an economy is below normal conditions or in a recession, NPL levels may rise due to the ensuing growth in unemployment, and borrowers face severe debt repayment difficulties (Salas, Suarina, 2002; Ranjan, Dhal, 2003; Fofack, 2005;

Jiménez, Saurina, 2005; Thalassinos et al., 2015, Kosztowniak, 2020; 2021). Exchange rate fluctuations may have a negative impact on the quality of assets, especially in countries with a large amount of foreign currency loans. The same applies to interest rate increases, particularly in the case of loans with flexible interest rates (Louzis et al., 2012; Zaman, Meunier, 2017). However, on the one hand, higher inflation may ease debt compensation by affecting the real value of unpaid credit, while on the other hand it may also reduce the real income of unprotected borrowers. In countries where credit rates are flexible, higher inflation may lead to higher rates resulting from monetary policy actions to fight inflation (Nkusu, 2011).

The research by Klein (2013) for NPLs in Central, Eastern and South-Eastern European countries (CESEE) in 1998–2011 confirmed that NPLs responded to macroeconomic conditions, i.e., unemployment, GDP growth and inflation, and that high NPLs in these countries have a negative effect on economic recovery. According to Mazreku et al. (2018) for 10 transition countries (Central and Eastern Europe) in 2006 and 2016, dynamic panel estimates show that GDP growth and inflation are both negatively and significantly correlated with the level of NPLs, while unemployment is positively related to NPLs. Export growth shows largely insignificant results, indicating that NPLs in the sample are mainly influenced by domestic conditions rather than external economic shocks. Vogiazas and Nikolaidou (2011) investigate the determinants of nonperforming creditors in the Romanian banking sector during the Greek crises (2001–2010) and find that inflation and external GDP information influenced the credit risks of the banking system in the country. According to Hada et al. (2020, pp. 1–19), the exchange rates (mainly EUR, USD and CHF), unemployment rate and inflation rate had a significant impact on NPLs in the Romanian banking system in the period 2009–2019.

Among the banking variables that define NPLs, research tends to focus on return on assets (ROA), bank efficiency, and bank capital. However, the specificity of each bank and its customers are very important for NPL changes. For example, Godlewski (2008) investigates the association between NPLs and return on assets (ROA) and states that the lower the rate of ROA, the higher the NPLs and vice versa. Boudriga et al. (2010) confirm from their study that there is a negative association between ROA and NPLs. They conclude that when the ROA decreases, then a bank starts to make investments in high-risk projects and as a result the level of NPLs rises. Dimitrios et al. (2016) investigate the various determinants of NPLs in the euro banking system and conclude that ROA has a significant impact upon NPLs. An insufficient control of the loan portfolio (including short-term loans) increases risk and NPLs. Fiordelisi et al. (2011) examine the various factors that increase the risk level in the EU banks and conclude that a declining efficiency hikes the risk level of banks in future. Furthermore, efficiency and performance factors had an influence on NPLs in the Greek banking sector (Louzis et al., 2012). Rachman et al. (2018) state that operating efficiency does not influence NPLs.

The effect of bank capital on NPLs works in the opposite direction. For one part, incentivised managers of low capitalized banks tend to get involved in high-risk investments and give loans that are issued without proper credit rating and monitoring (Keeton, 1999). For another part, banks with a high level of capital tend to give loans easily as they know that owing to these loans banks are not going to be bankrupt and fail; therefore, banks are highly engaged with these kinds of risky credit activities suggesting a positive association between capital and NPLs (Rajan, 1994). Moreover, the capital adequacy ratio (CAR) shows the ability of an organization to face abnormal losses and to survive that situation. Makri et al. (2014) also state that there is a negative association between CAR and NPLs. Constant and Ngomsi (2012) claim that NPLs and CAR have a positive association with each other. Bank profitability and sustainability can only be provided through a proper flow of interest income generated through the lending function. However, since banks are no longer able to generate enough interest income through classical safe credit and are required to maintain reserves in the form of provisions to cover for eventual loan losses, bank capital decreases together with their health, which is becoming fragile, increasing the trend of NPLs. Therefore, banks are required to take proactive action to deal with the phenomenon of the poor choices of borrowers, mainly by identifying and understanding the macroeconomic factors that contribute to the rise of classified credit in the banking system (Anjom, Karim, 2016).

The European Commission (EC) (Kasinger et al., 2021) has announced strategies to combat non-performing loans. The first plan was unveiled by the ECOFIN Council in July 2017. It was then extended with a new package of measures in March 2018 and a capital market recovery package in July 2020. The outbreak of the COVID-19 pandemic may additionally adversely affect household incomes and, consequently, the growth of NPL. Therefore, it is important to identify the main NPL determinants of the household loan portfolio, that is, the variable income and cost that determine the serviceability of loans. In Staeher's and Uusküla (2020) opinion, estimations show that many macroeconomic and macro-financial variables are the leading indicators for non-performing loans in the EU countries, even years ahead. Higher GDP growth, lower inflation and lower debt are robust leading indicators of a lower ratio of non-performing loans in the future.

3. Changes of NPLs in selected EU countries

Non-performing loans, with the exception of Hungary, showed relatively stable levels, with an average deviation of 1–2 pp, in the countries of the Visegrad Group (V4), as well as in Germany and France, in 2009–2020. In the case of Hungary, the financial crisis of 2007/2008 had a negative impact on the deterioration of banks' loan portfolios, escalating the growth of non-performing loans to 18.8% in 2013. It took ten years to restore the portfolio quality to its previous levels (3.2% in 2008, 2.5% in 2018). The lowest average level of NPL was maintained by Germany (2.1%) and France (3.6%), it was comparable in Poland, Slovakia and the Czech Republic (slightly above 4%), and the highest in Hungary (8.6%). The example of Hungary shows that allowing a deterioration in the loan portfolio is difficult to repair and sometimes takes a long time (around a decade). Therefore, the supervision and prevention of a deterioration of the loan portfolio should be a permanent responsibility of banks. Moreover, the data for 2020 indicate that the COVID-19 pandemic has not yet affected NPL changes in the group of analyzed countries. They remained similar to 2019 levels. Taking into account the continued demand for loans and the lack of growth in non-performing loans, this indicates the positive impact of government assistance programs (Fig. 1).

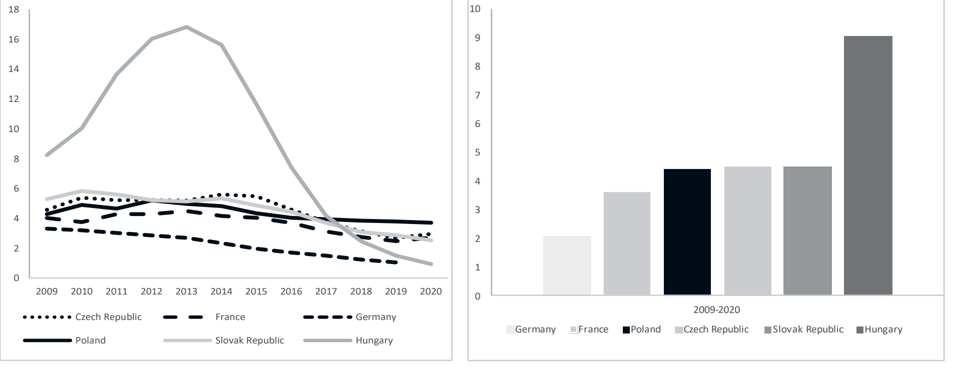

Figure 1. Bank NPLs to total gross loans in selected countries in 2008–2020 [%]

Source: The author`s compilation based on WDI (2021)

There were significant differences in the amount of debt servicing costs between the analyzed countries. While the average level of these costs for the Czech Republic and Poland was just over 7.0%, they were higher by over 4 pp in Germany and Hungary and by 11 pp for France. In 2009–2020, the Czech Republic and Poland maintained a stable level of debt servicing costs. Germany slightly decreased (by 1 pp) their level. In Hungary, along with the restructuring of loan portfolios, these costs fell significantly from Q1 2009 to Q4 2019 (by 12.3 pp). It is worth noting that while NPLs did not show changes as a result of the COVID-19 pandemic in 2020, debt servicing costs exhibited such a reaction. They climbed in France, Germany and Hungary. The increase in these costs may further affect the deterioration of the quality of the loan portfolio and the growth of NPLs in the coming year (Fig. 2).

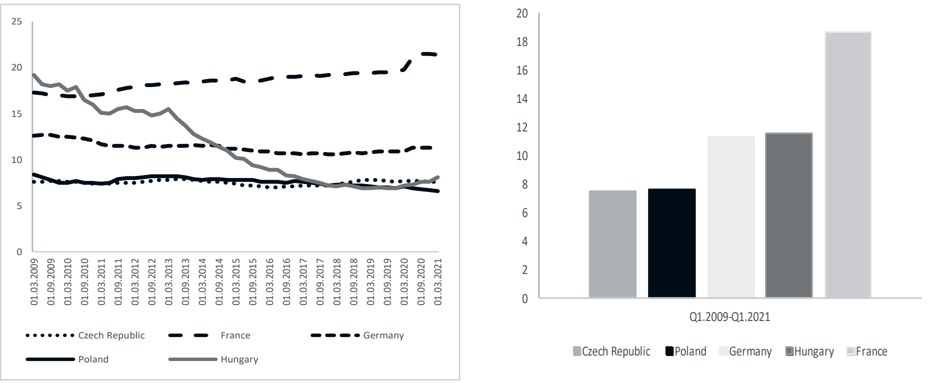

Figure 2. Debt service ratio for the private non-financial sector in selected countries in 2009–2021 (quarters, %)

Source: The author`s compilation based on BIS.org (2021)

The share of loans to households in the total loans of the analyzed countries ranges from 23.6% in Hungary to over 40% in Poland and Germany. The value of these loans accounts for nearly 30% of GDP in Hungary to around 60% of GDP in the euro area countries, including Germany and France. Thus, changes in the financial situation of households significantly affect the quality of the entire banking sector loan portfolio and the possibilities of economic growth, requiring the monitoring of the determinants of this situation (Fig. 3).

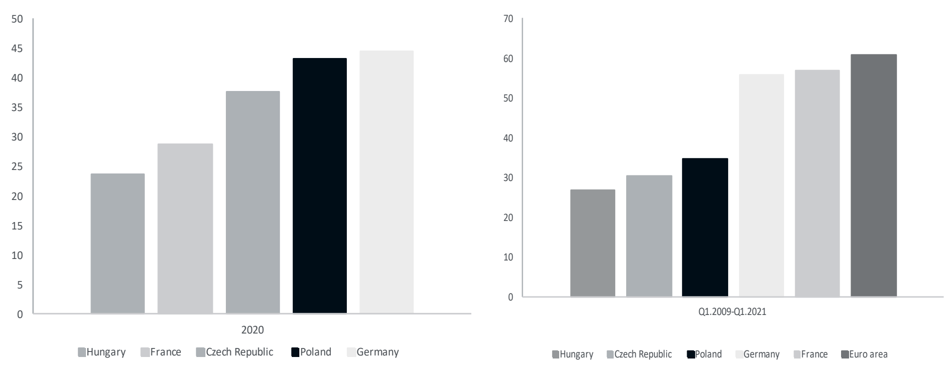

Figure 3. Household loans in selected countries (left panel, % of total loans, right panel, % GDP)

Source: The author`s compilation based onBIS.org (2021)

4. Results and discussion

4.1. Data and methodology

The National Bank of Poland (NBP) and other institutions, e.g., the International Monetary Fund (IMF), state that loans would be considered NPLs if they do not produce interest and principal amount for a minimum of 90 days. The NPL rate is calculated as the ratio of non-performing loans (impaired loans) and advances to the gross value of total loans and advances (NBP, 2021).

Poland is selected for the analysis of NPL determinants for household loans because the amount of NPLs and the structure of the loan portfolio in this country remain average among the analyzed countries. To specify the determinants of NPL for household loans (which in Poland account for 40% of total loans), it was decided to carry out research for this loan portfolio, not for the entire loan portfolio. Attention is paid to the variables determining the creditworthiness of households, i.e., mainly real income and loan servicing costs. Thus, the results of the study fill a gap in this area.

In the methodological approach used by the NBP (2021), household loans are available to: private persons, individual entrepreneurs, individual farmers, and non-commercial institutions operating for the benefit of households. The article attempts to assess the quality of the portfolio of loans granted to households, therefore, respectively, impaired loans and total loans granted to these households (included in the so-called phase III, portfolio B) are considered.

The time series of the model variables are presented in Figure 4.

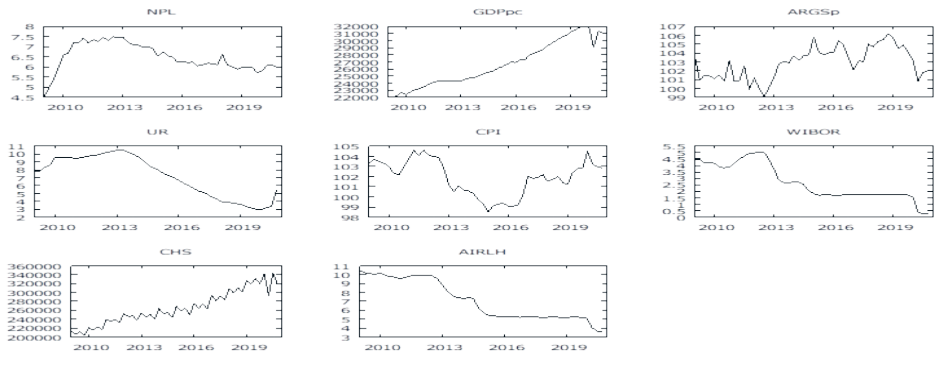

Figure 4. The time series of the model variables

Source: The author's own calculations, GRETL program

The research is based on statistics from the NBP, Central Statistics Office (CSO), Organization for Economic Co-operation, Development (OECD Internet databases), and Eurostat. EViews is employed for the purposes of calculations.

The specificity of the base equation is developed as a formula:

$$lnNPL_t = a_1 + a_2 \ln GDPpc_t + a_3 \ln ARGSp_t + a_4 \ln UR_t + a_5 \ln CPI_t + \\ + a_6 \ln WIBOR_t + a_7 \ln CHS_t + a_8 \ln AIRLH + u_t$$

where the explained variable: NPL_t – non-performed loan ratio

Explanatory variables:

GDPpc – gross domestic product per capita (*GDPpc*, fixed PPPs, seasonally adjusted, US dollars),
ARGSp – average monthly real gross salary (analogous period of the previous year = 100),
UR – unemployment rate [%],
CPI – consumer Price Index [%],
WIBOR – Warsaw Interest Board Rate [%],
CHS – consumption in the household sector [PLN million],
AIRLH – average interest rate on loans to households and non profit institution serving households [%],
ln – natural logarithm,
u – random factor,
t – period.

The methodology of changes in the quality of the loan portfolio corresponds to the methodologies used by central banks, e.g., by NBP and IMF (2003), Matthewes et al. (2007), Maggi and Guida (2010), Mazreku et al. (2018). The study period includes quarterly data for the period Q1.2009–Q2.2021 (Tab. 2).

Methods are used known from literature on international economics and international finance and econometric methods like the VECM model (*Vector Error Correction Method*) including the impulse response functions and the decomposition of variance. The expected influence of the explanatory variables on the explained variable (NPLs) is presented in Table 1.

The model data is verified on the basis of tests for unit roots, e.g., Augmented Dickey–Fuller (ADF) test, and cointegration is tested using the Johanson test and the Engle Granger test. The results confirm the applicability of the VECM model.

The sources of a changing quality of the loans portfolio are explained by means of the following methodology: (NBP, 2020), (IMF, 2003) and e.g. (Matthewes et al., 2007), (Maggi and Guida, 2010). The study period includes 50 quarterly data for the period Q1.2009–Q2.2021. All variables are smoothed by simple moving averages.

Table 1
Model variables

No.	Variables	Data source	Expected impact on the NPLs
1	*NPL*	NPB	"–"
2	*GDPpc*	OECD	"–"
3	*ARGSp*	CSO	"–"
4	*UR*	CSO	"–"
5	*CPI*	CSO	"+"
6	*WIBOR*	Eurostat	"+"
7	*CHS*	CSO	"+ / –"
8	*AIRLH*	NBP	"+"

Source: The author's own preparation

To verify the stationarity of the analyzed time series, the Augmented Dickey–Fuller (ADF) test is used, estimated by means of the regression equation in the following form:

$$\Delta y_t = \mu + \delta_{t-1} + \sum_{i=1}^{k} \delta_i y_{t-1} + \epsilon_t$$

The value of the test statistic is calculated by:

$$ADF = \frac{\tilde{\delta}}{S_{\tilde{\delta}}}$$

where δ means the parameter evaluation and s_δ is the parameter estimate error.

All the analyzed variables are found to lack the stationarity of time series, but a unit root a = 1 occurred at process I(1). A comparison between test τ statistics and critical values of these statistics shows that in the case of basic variables, the series are non-cointegrated and variables are non-stationary because the test probabilities are above 0.05. On the other hand, in the case of first differences, variables are mostly stationary and the series are co-integrated to the order of 1 (Tab. 2).

Table 2

Augmented Dickey–Fuller (ADF) test

Variable	Null hypothesis: unit root appears	With absolute term (const)	
		test statistic: $\tau\_ct(1)$	asymptotic p-value
l_NPL	a = 1; process I (1)	−1.62283	0.4708
l_GDPpc		−0.94158	0.7755
l_ARGSp		−1.61224	0.4763
l_UR		−1.88842	0.3381
l_CPI		−1.48913	0.5394
l_WIBOR		0.52368	0.9876
l_CHS		−0.94661	0.7738
l_AIRLH		−0.17732	0.9390

Source: The author's own calculations

To verify the conclusions drawn on the basis of the ADF test, the KPSS (Kwiatkowski–Philips–Schmidt–Shin) stationarity test is carried out, where the null hypothesis assumes a sequence stationarity, whereas the alternative hypothesis assumes the occurrence of the unit root. The initial test model can take the following form:

$$\gamma t = \beta t + rt + \xi t$$

where: $r_t = r_t - 1 + u_t$, where ξ_t and u_t are a stationary and a white-noise random component, respectively. On the other hand, the KPSS test statistic is calculated with the use of the formula:

$$KPSS = T^{-2} \sum_{t=1}^{T} \left(\sum\nolimits_{t=1}^{t} e_i \right) / \hat{\delta}^2$$

where e_i enotes residuals and $\hat{\delta}^2$ is a long-term variance estimator (Kufel, 2011).

An ultimate confirmation of stationarity requires an additional test, e.g., KPSS (Tab. 3).

The lag order for the VAR/VECM model is determined on the basis of estimation of the following information criteria: the Akaike information criterion (AIC), Schwartz–Bayesian information criterion (BIC), and Hannan–Quinn information criterion (HQC). According to these criteria, the best, that is, minimal values of the respective information criteria are: AIC = 2, BIC = 2 and HQC = 2, with the maximum lag order 4. Ultimately, the lag order 2 is accepted.

Table 3

KPSS stationarity test results (lag truncation = 4)

Specification		l_NPL	l_GDP-pc	l_AR-GSp	l_UR	l_CPI	l_WI-BOR	l_CHS	l_AIR-LH
Include a trend	test statis-tic	0.17234	0.12428	0.115756	0.22196	0.203683	0.085267	0.10575	0.101386
	critical value of the test	0.121 (10%); 0149 (5%); 0.213 (1%)							
Interpolated *p*-value		0.035	0.095	0.01	0.01	0.016	0.10	0.10	0.10

Source: The author's own calculations

In order to analyze the stability of the VAR model, a unit root test is applied. The test indicates that in the analyzed model equation roots in respect of the module are lower than one, which means that the model is stable and may be used for further analyses (Fig. 5.).

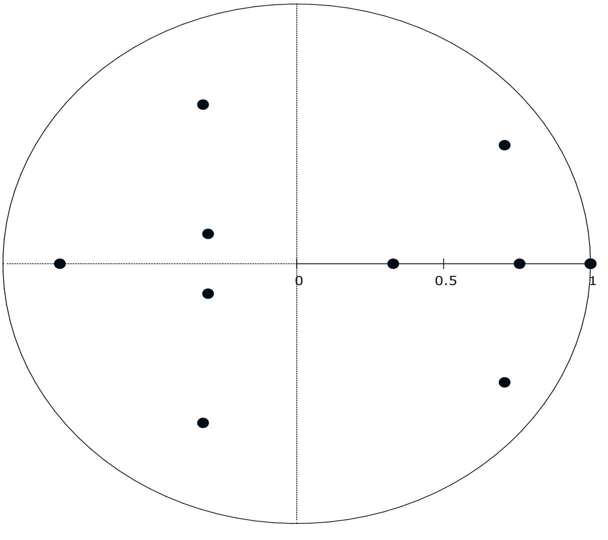

Figure 5. VAR inverse roots in relation to the unit circle

Source: The author's own calculations

Co-integration is verified using two tests: the Engle–Granger and Johansen tests (Johansen 1991, 1992, 1995). Their results comprehensively confirm co-integration for lag 1. This is proved by the values of the test statistic τ_e which are lower than critical values $\tau_{critical}$, the levels of asymptotic *p*-values and integrated processes a = 1 and I(1), at a significance level α = 0.05 (Tab. 4).

Table 4

Results of the Engle–Granger co-integration test

Specification	l_NPL	l_GDPpc	l_UR	l_CPI	l_WIBOR	l_CHS	l_AIRLH
Unit root appears	a = 1, process I (1)						
Test statistic τ_c τ_e (asymptotic *p*-value)	−1.62283 (0.4708)	−0.94158 (0.7755)	−1.88842 (0.3381)	−1.48913 (0.5394)	0.52368 (0.9876)	−0.94661 (0.7738)	−0.17732 (0.9390)

Source: The author's own calculations

Testing cointegration is designed to find a long-term relationship between variables. Using the strong testing methods of Johansen Cointegration and cointegration relationship variables, it can be concluded there is a long-term relationship between variables. The results of the Johansen test (including trace and eigenvalue) show that at the significance level of 0.05, co-integration to the order of one occurs (Tab. 5).

Table 5

Johansen test, lag order = 4, estimation period: 2010:1–2021:2

<table>
<tr><th>Rank</th><th>Eigenvalue</th><th>Trace test [p-value]</th><th>Lmax test [p-value]</th><th rowspan="7">0.74557</th></tr>
<tr><td>0</td><td>0.93506</td><td>492.78 [0.0000]</td><td>120.31 [0.0000]</td></tr>
<tr><td>1</td><td>0.91863</td><td>372.47 [0.0000]</td><td>110.39 [0.0000]</td></tr>
<tr><td>2</td><td>0.90316</td><td>262.08 [0.0000]</td><td>102.73 [0.0000]</td></tr>
<tr><td>3</td><td>0.74557</td><td>159.36 [0.0000]</td><td>60.223 [0.0000]</td></tr>
<tr><td>4</td><td>0.68691</td><td>99.133 [0.0000]</td><td>51.095 [0.0000]</td></tr>
<tr><td>5</td><td>0.45126</td><td>48.038 [0.0001]</td><td>26.406 [0.0066]</td></tr>
</table>

Table 5 cont.

6	0.28865	21.632 [0.0044]	14.986 [0.0363]	0.74557
7	0.14020	6.6462 [0.0099]	6.6462 [0.0099]	
eigenvalue	0.93506	0.91863	0.90316	

Source: The author's own calculations

Due to the occurrence of a unit element in all the time series and the existence of cointegration between the model variables, it is possible to extend and transform the model into vector error correction models (VECM).

4.2. VECM model and results

Co-integration is verified, thus justifying the use of the VECM model for the lag order 2 and co-integration of the order 1. In accordance with the Granger representation theorem, if variables y_t and x_t are integrated to the order of I (1) and are co-integrated, the relationship between them can be represented as a vector error correction model (VECM) (Piłatowska, 2003).

The general form of the VECM can be written as:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \ldots \Gamma_{k-1} \Delta Y_{t-k+1} + \pi Y_{t-k} + \varepsilon_t = \\ = \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \pi Y_{t-k} + \varepsilon_t,$$

where:

$$\Gamma_i = \sum_{j=1}^{i} A_j - \mathrm{I}, \qquad i = 1, 2, \ldots, k-1, \qquad \Gamma_k = \pi = -\pi(1) = -\left(\mathrm{I} - \sum_{i=1}^{k} A_i\right)$$

and I is a unit matrix.

The results of the beta index of the VECM model indicate that the variables can be treated as the variables of long-term effect on NPL. The parameters of the alpha vector suggest that the highest rates of adaptation show their own changes in NPL, then in UR and WIBOR.

The EC1 index (containing the evaluation of the error correction index) confirms that the strongest correction of deviation from the long-term equilibrium occurs in the case of the NPL equation. Here, 14.3% of the imbalance from the long-term growth path is corrected by the short-term adjustment process. The results of the determination coefficient (R2) indicate a moderately good adjustment of the VECM model equations to the empirical data. The results of the Durbin–Watson (DW) test do not confirm the existence of a significant residual autocorrelation (Tab. 6).

Table 6

The VECM model

VECM system, lag order 2, observations 2009:3–2021:2 (T = 48) Cointegration rank = 1, Case 3: Unrestricted constant			
β (cointegrating vectors, standard errors in parenthes)			α (adjustmentvectors)
1_NPL	1.0000	(0.0000)	−0.1439
1_GDPpc	−2.6484	(1.4716)	−0.0171
l_ARGSp	4.4294	(2.1068)	−0.0158
l_UR	0.1064	(0.1649)	−0.1236
l_CPI	−2.1136	(2.5833)	−0.0036
l_WIBOR	0.5392	(0.1844)	−0.1207
l_CHS	2.7568	(0.8638)	−0.0785
1_AIRLH	−1.1109	(0.4019)	−0.0018

Specification	EC1	R2	DW
1_NPL	−0.14387	0.666641	1.972752
1_GDPpc	−0.01708	0.317993	1.692827
l_ARGSp	−0.01579	0.224475	2.189725
l_UR	−0.12361	0.855231	1.956559
l_CPI	−0.00362	0.151093	1.928489
l_WIBOR	−0.12067	0.324845	1.841010
l_CHS	−0.07852	0.694566	1.845236
1_AIRLH	−0.00186	0.268971	1.838926

Source: The author's own calculations, GRETL program

In order to verify the correctness of the VECM model results, two tests are carried out verifying the occurrence of autocorrelation, i.e.: autocorrelation Ljung-Box Q′ test, lag order for test = 2, and ARCH test = lag order for test = 2. Ljung–Box tests (LMF, LM, Q) verify autocorrelation for the lag order 4. The verifying statistic using the autocorrelation coefficient function (ACF) in the form Q′ and empirical *p*-value levels higher than the nominal $\alpha = 0.05$ let us conclude that there is no autocorrelation in the residual process (Kufel, 2011). The ARCH test results indicate the ARCH effect is not observed in the examined model of the residual-based process (four variables), because LM test statistics are lower than the levels of χ^2. This means that there is no autoregressive changeability of the conditional variance and there is no need to estimate model parameters by means of the weighted least squares method. Thus, the results of both the tests confirm credibility of the VECM model and allow for conclusions drawn on their basis.

The results presented in the article are consistent with those reported by such authors as: Salas and Suarina (2002), Ranjan and Dhal (2003), Fofack (2005), Jiménez and Saurina (2005), Djiogap and Ngomasi (2012), Thalassinos et al. (2015), Mazreku et al. (2018).

4.3. Impulse response functions

The analysis of the NPL response to impulses from the explanatory variables confirms that the strength of the influence of these impulses increased over time. About 5–7 quarters of the forecast, the impact of explanatory variables on NPLs showed a stabilization (constant).

The NPLs showed increasing trends in response to change impulses from own NPL (3%), CPI (2%), and AIRLH (1%). Earlier changes in NPL (problems with servicing loans) translate into future changes, inflation lowers the purchasing power of disposable income with rising consumer prices, and changes in the interest rate of loans raise the interest due. After about 2.5 years, the NPL also shows a weak increase due to the influence of GDPpc, which may indicate a rising demand for credit accompanied by a GDPpc growth.

The NPLs show diminishing trends in response to the changes of: CHS, ARGSp, UR, and WIBOR. The increase in consumption expenditure proves that creditworthiness (the repayment of loan costs) is maintained with arise in real wages, which contribute to a reduction of NPL. The weak but negative impact of the unemployment rate and WIBOR on NPL can be explained by compliance with the requirements of creditworthiness assessment, a loan application may be rejected as it deteriorates (Fig. 6).

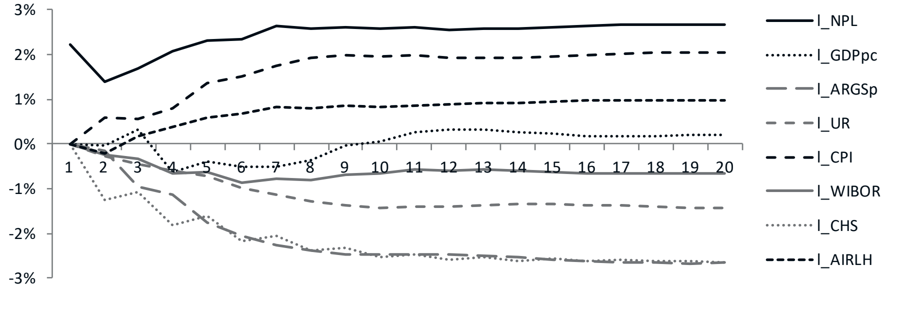

Figure 6. Responses of NPL to a one-standard error shock coming from variables

Source: The author's own calculations

To sum up, the quality of the household loan portfolio deteriorates as a result of previous unfavorable changes in this portfolio, increased inflation and interest on loans to households. The importance of the impact of inflation on NPLs implies the important role played by monetary policy and the legitimacy of monitoring the level of inflation, the increase of which may affect the quality of the household loan portfolio by approx. 2%.

4.4. The decomposition of variance

The results of the variance decomposition indicate that the previous NPL changes as well as CHS and ARGSpc have the highest share in explaining changes in NPL, deciding about 87% and 75% of changes in the 1st and 5th year of the forecast. Over time, the impact of own changes diminishes from 100% in Q1 to 28% in Q20, while the importance of CHS rises from 17.5% to 23.6% and of ARGSp from 0.3% to 23.0%, respectively. An increasing degree of explanation of NPL changes by CPI, from 3.8% to 13.7%, and UR, from 0.9% to 6.5%, is notable. Other explanatory variables (GDPpc, WIBOR and AIRLH) are significant, too, however, their influence does not exceed 6% in total (Fig. 7).

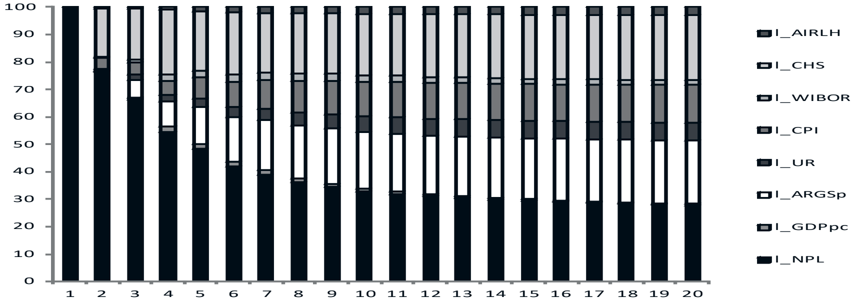

Figure 7. Forecast variance decomposition for I_NPL

Source: The author's own calculations

The results of the decomposition confirm the results of the analysis of the impulse response function, indicating three pillars of NPL changes, i.e., own changes of NPL, CHS and ARGSp.

5. Discussion

The VECM model, the impulse response function and the variance decomposition confirm the importance of the main determinants of household creditworthiness, i.e., income (relative wages) and expenditure (consumption demand) for changes in the quality of the loan portfolio in the Polish banking sector.

The results of the research corroborate a growing influence of macroeconomic conditions, including the CPI and the unemployment rate. These two indicators have a key impact on the amount of relative household income as well as the ability to earn. Thus, they play an important role in the monetary policy pursued by the central bank and in the economic policy of the government. The importance of other variables, such as interest rates on loans to households, is less important than the aforementioned relative wages and expenses. The study is consistent with the results of other authors analyzing changes in the portfolio of total non-performing loans, which emphasize the important role of borrowers' financial conditions.

As the research results presented in the article focus on one group of borrowers (households), these results additionally specify the portfolio quality determinants for this group. Thus, they constitute the author's contribution to research into the quality of the loan portfolio. Moreover, these findings may constitute proposals for extending the assessment of the creditworthiness of borrowers, in this case of households, to include market conditions.

6. Conclusion

As a deterioration in loan quality may destabilize the situation in the banking sector and spread to the entire economy, it is important to monitor the determinants of NPL change. Compliance with macroprudential regulations in the banking sectors of EU countries reduces non-performing loans, which is confirmed by the NPL data presented for the V4 countries, France, and Germany.

The empirical data show that, first, in the V4 countries, as well as in Germany and France, it was possible to improve the quality of loans in 2009–2020. However, the greatest restructuring effort was undertaken by Hungary, which reduced the level of NPLs from 16.4% (2013) to 0.93% (2020). Secondly, the highest quality of the loan portfolio (with the lowest NPL) was maintained by Germany, France and Poland (with a stable NPL level). In 2009–2017, Hungary had the gravest problems with non-performing loans, yet managed to restructure them in 2018–2020. Thirdly, the model analysis of the VECM and the function of response to impulses and variance decomposition for Poland in the period 2009–2021 allows for the identification of the main determinants of the quality of the household loan portfolio. The NPLs showed increasing trends in response to change impulses from own changes of NPL, CPI, and average interest rate on loans to households (AIRLH). Earlier changes in NPL (problems with servicing loans) translate into future changes, inflation lowers the purchasing power of disposable income with rising consumer prices, and changes in the interest rate of loans raise the interest due. The NPLs showed declining trends in response to the changes of: consumption in the household sector (CHS), average monthly real gross salary (ARGSp), unemployment rate (UR), and WIBOR. The results of the variance decomposition indicate that previous own NPL changes as well as CHS and ARGSpc have the highest share in explaining changes in NPL, deciding about 87% and 75% of changes in the 1st and 5th year of the forecast. The shrinking degree of explanation of NPL changes by CPI and UR is worth underlining. Other explanatory variables (GDPpc, WIBOR and AIRLH) are significant, however, their influence does not exceed 6% in total.

In the context of asset quality management, constant monitoring of NPLs is important, as a deterioration in loan service produces effects in subsequent periods. The significant impact of CHS and ARGSpc on NPLs proves the importance of changes in demand (expenditure) and real wages (income) of households, i.e., the pillars of creditworthiness. The interest rate on loans influences the NPL, however, it is weaker than in the case of expenses and income. The growing degree of explanation of changes in NPL on the part of UR and CPI indicates the importance of macroeconomic conditions determining the real incomes of households.

In the beginning of Q2.2021, the impact of the COVID-19 pandemic has not affected the growth of the NPL yet, although the costs of debt servicing have already shown an uptick in e.g., France, Germany, or Hungary. In the following year, an increase in these costs may additionally affect the deterioration of the quality of the loan portfolio in the EU countries. As the share of loans to households in total loans in the analyzed countries ranges from 24% in Hungary to over 40% in Poland and Germany and their value ranges from nearly 30% of GDP in Hungary to around 60% of GDP in the euro area, it is important to study the determinants of changes in this loan portfolio.

In summary, in the years 2009–2021 the quality of the loan portfolio improved, as evidenced by the decrease in the NPL ratio in the analyzed EU countries. Household loans are important in the structure of this portfolio. The results of the model analysis for Poland confirm the importance of demand (expenditure) and income conditions for the improvement of the quality of this portfolio as well as of changes in UR and CPI affecting these conditions. Although the NPL data in 2020 do not show a deterioration in loan quality, an increase in servicing costs found in some countries in early 2021 may affect its changes in subsequent periods. In practical terms, the conclusions from the research for Poland can be used by other EU countries, including mainly the Czech Republic and Germany (with a similar structure of the loan portfolio, i.e., 40% household loans) or Slovakia (a similar 4% NPLs level). In the case of Hungary, although they have managed to restructure the loan portfolio, the challenge is to preserve the achieved portfolio quality in the future. Moreover, the added value of the article consists in drawing attention to the importance of the structure of the loan portfolio, including other determinants influencing the NPL of households than of enterprises or public sector institutions. The practical aspect of the study means that the results can be used to manage the portfolio of loans to households and forecast changes in banking risk.

References

[1] Anjom, W. and Karim, A.M. (2016) 'Relationship between Non-Performing Loans and Macroeconomic Factors with Specific Factors: A Case Study on Loan Portfolios – SAARC Countries Perspective', *Asia Pacific Journals of Finance and Risk Management*, vol. 15(3), pp. 84–103.

[2] BIS (2021) Statistics Explorer, [Online], Available: http://stats.bis.org:8089/statx/srs/table/ f3.1?f=xlsx [29 Sep 2022].

[3] Boudriga, A., Taktak, N.B. and Jellouli, S. (2010) 'Bank Specific, Business and Institutional Environment Determinants of Banks Nonperforming Loans: Evidence from MENA Countries', Economic Research Forum, Working Paper, pp. 1–28.

[4] Djiogap, C.F. and Ngomsi, A. (2012) 'Determinants of Bank Long-term Lending Behavior in the Central African Economic and Monetary Community (CEMAC)', *Review of Economics and Finance, Better Advances Press*, vol. 2, pp. 107–114.

[5] Dimitrios, A., Helen, L. and Mike, T. (2016) 'Determinants of Non-Performing Loans: Evidence from Euro-Area Countries', *Finance Research Letters*, vol. 18, pp. 116–119, doi: 10.1016/j.frl.2016.04.008.

[6] Fiordelisi, F., Marques-Ibanez, D. and Molyneux, P. (2011) 'Efficiency and risk in European banking', *Journal of Banking and Finance*, vol. 35, No. 5, pp. 1315–1326.

[7] Fofack, H. (2005) 'Non-performing Loans in Sub-Saharan Africa: Causal Analysis and Macroeconomic Implications', World Bank Policy Research Working Paper No. 3769.

[8] Godlewski, C.J. (2008) 'Bank capital and credit risk taking in emerging market economies', *Journal of Banking Regulation*, vol. 6, No. 2, pp. 128–145, doi: 10.1057/palgrave.jbr.2340187.

[9] Gosh, A. (2015) 'Banking-Industry Specific and Regional Economic Determinants of Non-Performing Loans: Evidence from US States', *Journal of Financial Stability*, vol. 20, pp. 93–104, doi: 10.1016/j.jfs.2015.08.004.

[10] Hada T., Bărbuță-Mişu, N., Iuga, I.C. and Wainberg, D. (2020) 'Macroeconomic Determinants of Nonperforming Loans of Romanian Banks', *Sustainability*, vol. 12(18), 7533, doi: 10.3390/su12187533.

[11] Handley, N.J. (2010) 'Growth in a time of debt', *CFA Digest*, vol. 40, No. 3, pp. 19–20, doi: 10.2469/dig.v40.n3.19.

[12] Ivanović, M. (2016) 'Determinants of Credit Growth: The Case of Montenegro', *Journal of Central Banking Theory and Practice*, vol. 5, No. 2, pp. 101–118, doi: 10.1515/jcbtp-2016-0013.

[13] Jiménez, G. and Saurina, J. (2005) 'Credit Cycles, Credit Risk, And Prudential Regulation', *International Journal of Central Banking*, vol. 2(2), pp. 1–34.

[14] Kasinger, J., Krahnen, J.P., Ongena, S., Pelizzon, L., Schmeling, M. and Wahrenburg, M. (2021) 'Non-Performing Loans – New Risks and Policies? NPL Resolution After COVID-19: Main Differences To Previous Crises', European Parliament, Economic Governance Support Unit (EGOV) Directorate-General for Internal Policies PE 651.387, March, [Online], Available: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/651387/IPOL_STU(2021)651387_EN.pdf [29 Sep 2022].

[15] Keeton, W.R. (1999) 'Does faster loan growth lead to higher loan losses?', *Economic Review*, Federal Reserve Bank of Kansas City, vol. 84, pp. 57–76.

[16] Kosztowniak, A. (2020) 'Changes in The Quality of The Loan Portfolio in The Polish Banking System', *Journal of Economics, Business & Organization Research*, CEBOR Academy Ltd., Italy, [Online], Available: https://dergipark.org.tr/en/download/article-file/1470997 [29 Sep 2022].

[17] Kosztowniak A. (2021) 'The modelling of the Main Determinants of Non-Performing Loans in the Polish Banking Sector', in Bukowski, S. and Lament, M. (ed.) Prospects and challenges to development of businesses in the world economy, Warsaw: CeDeWu, pp. 89–108.

[18] Klein, N. (2013) 'Non-Performing Loans in CESEE: Determinants and Impact on Macroeconomic Performance', IMF Working Paper European Department, No. 13/72, pp. 1–26.

[19] Louzis, D.P., Vouldis, A.T. and Metaxas, V.L. (2012) 'Macroeconomic and Bank-Specific Determinants of Non-Performing Loans in Greece: A Comparative Study of Mortgage, Business And Consumer Loan Portfolios', *Journal of Banking and Finance*, vol. 36, No. 4, pp. 1012–1027, doi: 10.1016/j.jbankfin.2011.10.012.

[20] Maggi, B. and Guida, M. (2010) 'Modelling non-performing loans probability in the commercial banking system: efficiency and effectiveness related to credit risk in Italy', *Empirical Economics*, vol. 41, pp. 269–291, https://link.springer.com/article/ 10.1007/s00181-010-0379-2 [22 Sep 2022].

[21] Makri, V., Tsagkanos, A. and Bellas, A. (2014) 'Determinants of Non-Performing Loans: The Case of Eurozone', *Panoeconomicus*, vol. 61, No. 2, pp. 193–206.

[22] Mazreku, I., Morina, F., Misiri, V., Spiteri, J.V. and Grima, S. (2018) 'Determinants of the Level of Non-Performing Loans in Commercial Banks of Transition Countries', *European Research Studies Journal*, vol. XXI, Issue 3, pp. 3–13.

[23] Moradi, Z.S., Mirzaeenejad, M. and Geraeenejad, G. (2016) 'Effect of bank-based or market-based financial systems on income distribution in selected countries', *Procedia Economics and Finance*, vol. 36, No. 16, pp. 510–521, doi: 10.1016/s2212-5671(16)30067-3.

[24] NBP (2021) 'Assets And Liabilities of Monetary Financial Institutions', [Online], Available: https://www.nbp.pl/homen.aspx?f=/en/statystyka/zobowiazania.html [29 Sep 2022].
[25] Nkusu, M. (2011) 'Nonperforming Loans and Macrofinancial Vulnerabilities in Advanced Economies', IMF Working Papers, 161.
[26] Piłatowska, M. (2003) Modelowanie niestacjonarnych procesów ekonomicznych. Studium metodologiczne [Modelling of Non-Stationary Economic Processes. A Methodological Study], Toruń: Wydawnictwo UMK.
[27] Rachman, R.A., Kadarusman, Y.B., Anggriono, K. and Setiadi, R. (2018) 'Bank-Specific Factors Affecting Non-Performing Loans in Developing Countries: Case Study of Indonesia', *The Journal of Asian Finance, Economics and Business (JAFEB),* vol. 5, No. 2, pp. 35–42.
[28] Rajan, R. (1994) 'Why bank credit policies fluctuate', *The Quarterly Journal of Economics*, vol. 2, No. 109, pp. 399–441.
[29] Ranjan, R. and Dhal, S.Ch. (2003) 'Non-Performing Loans and Terms of Credit of Public Sector Banks in India: An Empirical Assessment', *Reserve Bank of India Occasional Papers*, vol. 24, No. 3, pp. 1–41.
[30] Salas, V. and Saurina, J. (2002) 'Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks', *Journal of Financial Services Research*, No. 22, pp. 203–224.
[31] Staehr K. and Uusküla L. (2021) 'Macroeconomic and Macro-Financial Factors as Leading Indicators of Non-Performing Loans: Evidence from the EU countries', *Journal of Economic Studies*, vol. 48, https://www.emerald.com/insight/content/doi/10.1108/JES-03-2019-0107/full/html.
[32] Souza G.J.D.G. and Feijó C.A. (2011) 'Credit risk and macroeconomic interactions: empirical evidence from the Brazilian banking system', *Modern Economy*, vol. 2, No. 5, pp. 910–929, doi: 10.4236/me.2011.25102.
[33] Staeher K. and Uusküla L. (2020), 'Macroeconomic and macro-financial factors as leading indicators of non-prforming loans: Evidence from the EU countries', *Journal of Economic Studies* (ahead-of-print), *Journal of Economic Studies*, vol. 48, doi: 10.1108/JES-03-2019-0107.
[34] Thalassinos, I.E., Stamatopoulos, D.T. and Thalassinos, E.P. (2015) 'The European Sovereign Debt Crisis and the Role of Credit Swaps', in Ziemba, W.T. and Malliaris, A.G. (eds), The World Scientific Handbook of Futures Markets, World Scientific Handbook in Financial Economic Series, vol. 5, pp. 605–639, doi: 10.1142/9789814566926_0020.
[35] Vogiazas, S. and Nikolaidou, E. (2011) 'Investigating the Determinants of Nonperforming Loans in the Romanian Banking System: An Empirical Study with Reference to the Greek Crisis', *Economics Research International*, vol. 2011, pp. 1–13.

[36] Vouldis, A.T. and Louzis, D.P. (2018) 'Leading indicators of non-performing loans in Greece: the information content of macro-, micro- and bank-specific variables', *Empirical Economics,* vol. 54, No. 3, pp. 1187–1214, doi: 10.1007/s00181-017-1247-0.
[37] Zaman C. and Meunier B. (2017), 'A Decade of EU Membership: Evolution of Competitiveness in Romania', *European Research Studies Journal*, vol. 20, pp. 224–236.

Summary

As non-performing loans (NPLs) can cause monetary crises that may turn into financial crises affecting an entire economy, monitoring them is very important. If NPLs are not identified and recognized efficiently, both in terms of speed and scope, NPL resolution effectiveness is undermined, which in turn will have negative effects on the banking sector and ultimately on GDP growth.

The main aim of this article is to identify changes in the quality of bank loan portfolios in European Union (EU) countries in 2009–2021, using an example of the Visegrad Group (Czech Republic, Poland, Slovakia, Hungary) as well as France and Germany. Keeping in mind the fact that the share of loans to households in EU portfolios is approximately 60%, it has a significant impact on the share of non-performing loans (NPL) in a bank's entire portfolio. Therefore, it is important to identify macroeconomic determinants influencing the creditworthiness of households and their loan servicing capacity.

The specific aims are, first, to present the differences in NPLs, debt servicing costs, and the structure of loan portfolios in the selected EU countries. Second, to identify countries with high-quality portfolios and those undertaking restructuring. Thirdly, to examine the determinants of NPL for household loans based on the example of Poland, i.e., a country considered representative in terms of the average level of NPL and the portfolio structure in the group of countries studied.

This chapter presents the changes of NPLs, debt service ratio, and household loans in selected EU countries in 2009–2021. Moreover, an NPLs econometric model for Poland is constructed, which considers the main factors determining the creditworthiness of households, i.e., macroeconomic factors, financial standing, and debt servicing costs. Tools such as the VECM model, the variance decomposition and the impulse response functions are used. The results for Poland confirm that the NPLs ratio for households was the strongest explanation of previous changes in own NPL, consumption and real wages in the household sector in 2009–2021.

JEL codes: E32, E44, G21, G26, N10, N20

Keywords: *loan portfolio quality, non-performing loans (NPL), households, credit risk, EU, Poland*

Managerial Economics
2022, Vol. 23, No. 1, pp. 49–76
https://doi.org/10.7494/manage.2022.23.1.49
©2022 Authors. Creative Commons CC-BY 4.0

Hannah Martensen*, Malte Martensen*,
Johannes Asanger*

Gamified Recruitment: A Way to Win the Talent of Tomorrow?

1. Introduction

Talent is the key to the success and survival of an organization. Whereas the lack of the right talent may result in poor performance, the lack of new talent may lead to problems such as unfilled vacancies, too little innovation, and limited growth (Scully et al., 2014). Thus, organizations must boost both their effectiveness and efficiency in acquiring new talent so as not to fall behind competitors. Recruitment activities that worked well in the past, however, may no longer be in tune with modern requirements: The transformation in technology and target group preferences calls for a change in recruitment (Gilch, Sieweke, 2021). Since its emergence as an academic topic in 2010, gamification has remarkably found its way into both private and professional environments (Koivisto, Hamari, 2019). Given its crucial role in the *War for Talent* (Michaels et al., 2001), HR recruitment appears to be an excellent choice to further investigate the potential and limitations of gamification and has not yet been given much attention in the academic literature. After all, if gamified recruitment processes could help acquire the desired talent more effectively, it may create and sustain a competitive advantage for the employer.

The aim of this study is to investigate the potential and limitations of gamified recruitment processes. This should assist companies in implementing gamification in their recruitment processes, understanding the motivation of the participants and successfully develop recruitment applications.

In detail, this means we seek to answer three research questions (see Table 1). With recruitment involving both the organization and the individual (Barber,

* IU International University of Applied Sciences, Berlin, Germany. H.M. — ORCID ID: 0000-0001-9260-5469, M.M. — ORCID ID: 0000-0002-6538-2957, J.A. — ORCID ID: 0000-0003-2724-1480.

1998), our research questions represent both perspectives. Moreover, we focus on design elements aiming to shed light on acceptance drivers of gamified recruitment. Graduate job seekers form the target group of our study. The rationale behind this choice is twofold: 1) job seekers are the natural target group of recruitment activities and 2) graduates comprise the largest group of job seekers thus showing the highest likelihood of coming across and experiencing gamified recruitment processes.

Table 1

Research questions

Research Questions		Perspective
RQ1	What are the motivations of graduate job seekers to engage with gamified recruitment?	individual
RQ2	What organizational goals can be pursued through gamified recruitment?	organizational
RQ3	What design elements foster the acceptance of gamified recruitment?	design

The findings of our study contribute to a better understanding of the functionality, benefits, and target group preferences of gamified recruitment. This should enable organizations to create applications that more adequately match the needs of graduate job seekers and initiate an effective and entertaining way of recruitment to attract the best talent. Furthermore, we want to advance research in the field of gamification by obtaining insights into the design of successful applications.

By means of a comprehensive literature review and two 90-minute focus groups, we investigate and elaborate individual motivations, organizational strategies, and design elements. By applying the *Unified Theory of Acceptance and Use of Technology 2*, we derive a comprehensive model incorporating these three dimensions.

The remainder of this paper is structured as follows. After an extensive literature review in section 2, the research methodology is presented in section 3 and next, the data analysis and results are addressed in section 4. The article concludes with a discussion of the results and an outlook on gamified recruitment in section 5.

2. Literature review and theoretical foundations

Several recruitment trends such as employer branding (Trost, 2009), employee referrals (Breaugh, 2008), or e-recruitment (Strohmeier, Kabst, 2009; Doherty, 2010;

Holm, 2012) have been well-researched. And while there is a growing interest in digital HRM (Strohmeier, 2020), gamification in recruitment has only recently received attention from researchers (Murawski, 2020). In their literature review, Koivisto and Hamari (2019) discerned a growing interest in gamification in general as an academic topic and report the positive effects of its application. Still, only few of the reviewed studies focus on gamified recruitment. This gap is surprising given that practitioner outlets and business magazines such as *HBR*, *Forbes*, and *The New York Times* have long praised the potential of gamified recruitment (Rampell, 2014; Chamorro-Premuzic, 2015; Maycotte, 2015).

2.1. Recruitment of graduate talent

In times of talent shortage, recruitment has evolved into one of today's most critical business functions: When Michaels et al. (2001) coined the term *War for Talent*, they suggested that "a company's ability to attract, develop, and retain talent will be a major competitive advantage far into the future" (p. 2). In fact, more than 70% of CEOs call for a skilled, educated and adaptable workforce as a business priority and at the same time worry about the availability of key skills in their organizations (PricewaterhouseCoopers, 2016).

In the organizational recruitment process, potential candidates have two contact points with the organization. They experience the recruitment activity (e.g., a job ad or a gamified recruitment application) and decide whether or not to apply (Breaugh et al., 2008). The experience ultimately influences the recruitment results (e.g., the number of hires). Of course, several intervening variables can shape the success of a recruitment activity, too. For example, if a company's recruitment activity intends to convey that its organizational culture matches the needs of job seekers, applicants need to see a consistent match between their actual needs and their interpretation of the organizational culture (Breaugh, 2008). Otherwise, the desired recruitment results will not be achieved, that is, too many unsuitable candidates apply and too many suitable candidates refrain from doing so.

As an organization's recruitment process can only be effective if it attracts the right quantity *and* quality of talent, HR professionals need to understand the drivers that contribute to such effectiveness. During their job search, candidates will be more likely to submit an application if they (1) are aware that an organization exists, (2) have a positive perception toward it, (3) consider it an attractive employer, and (4) find the job appealing (Trost, 2009).

Perceived fit is a significant factor during job search (Chapman et al., 2005), as it enables candidates to evaluate if their career goals and preferences correspond with the employer's organizational culture and the requirements of the job position.

In a meta-analysis about the drivers of applicant attraction, Uggerslev et al. (2012) identify perceived fit as the strongest predictor, suggesting that "organizations should direct their initial recruitment resources at fostering applicants' perceptions of fit" (p. 637). As gamification helps candidates assess their *person-organization (P-O)* and *person-job (P-J) fit* (Diercks, 2013), fewer suitable candidates may drop out at the early recruitment stages.

The ability to attract young graduate talent in particular is crucial to the long-term success of an organization and may become even more apparent in aging societies like Japan or Germany, whose labor force is likely to shrink significantly due to low birth rates and a growing share of retirees. Any organization seeking to recruit graduate talent should thus embrace this target group's main characteristics to better address their needs in the recruitment process. Younger generations are argued to have an affinity towards technology and (video) games (Thomas, 2011). Hence, Nair and Sadasivan (2019) suggest that gamification would generally be most effective for targeting individuals that are graduate job seekers due to their preference for technology.

2.2. Gamification

In recent years, gamification has received widespread attention from both academics and practitioners (Koivisto, Hamari, 2019; Murawski, 2020; Bina et al., 2021; Machado Leitão et al., 2021). Marketsandmarkets (2020) predict the global gamification market to grow from US $9.1 billion in 2020 to US $30.7 billion in 2025. Technological advances and the increasing diversity of video games are key drivers of this development (Hamari, Keronen, 2017).

Gamification, in its broadest but also most popular definition, is defined as the use of game design elements in non-game contexts (Deterding et al., 2011). Huotari and Hamari (2017) emphasize that value creation and behavioral change in users are the main objectives of gamification. In contrast to traditional games, which primarily seek to entertain the players, gamified applications follow additional objectives, such as boosting motivation (Alsawaier, 2019), engagement (Hamari, 2017), job satisfaction (Oprescu et al., 2014), learning (Zainuddin et al., 2020), collaboration (Raftopoulos, Walz, 2013), or recruitment effectiveness (Georgiou et al., 2019).

2.3. Gamification of recruitment processes

Even though many studies have emphasized the potential of gamification for HR recruitment there is a lack of empirical research on gamified recruitment

(Langer et al., 2018). Only recently studies have been published rigorously investigating the effects of gamified recruitment on both individual candidates (e.g. Collmus, Landers, 2019; Buil et al., 2020; Georgiou, Nikolaou, 2020) as well as on organizations (e.g., Georgiou et al., 2019).

Getting a better understanding of the individual motivations of candidates participating in gamified recruitment can significantly influence the recruiting success for candidates and organizations alike. Van der Heijden (2004) claims that any information system (IS) either targets hedonic or utilitarian purposes. Whereas an IS that follows utilitarian purposes aims to increase productivity, an IS that follows hedonic purposes seeks to provide self-fulfillment to its users (e.g., in the form of fun experiences). Hamari and Koivisto (2015) point out the unique character of gamification, as it combines both parts. On the one hand, gamification affords gameful experiences, which intrinsically motivate users to play. On the other hand, it helps achieve additional goals of extrinsic nature or provides external rewards (e.g., badges, ranks, or reputation points). When designing an application, it is therefore important to consider the motivational drivers of the desired user behavior in order to leverage both intrinsic and extrinsic motives (Blohm, Leimeister, 2013). Thus, we aim to better understand the motivations of individual candidates to engage with gamified recruitment (RQ1).

Although several literature reviews detected positive effects of gamification in general (Bina et al., 2021; Koivisto, Hamari, 2019; Murawski, 2020; Woods et al., 2020), its use is not free from criticism. Bogost (2014) questions the effectiveness of gamification, calling it an *exploitation ware* which tries to manipulate people and only helps marketers make quick profits. Callan et al. (2014) warn that gamification is prone to failure if rewards do not contain deeper meaning for the user or if a gamified activity is not aligned with the overall business strategy. Given this potential downside of gamification, how can organizations benefit from its adoption (RQ2)?

According to Stephan et al. (2017), recruitment has become a digital experience in which gamification can be a major tool for organizations to attract new talent, especially when combined with other media such as video or social networks. Georgiou et al. (2019) develop a gamification selection method that increases the recruitment effectiveness for organizations in terms of candidates' soft skills assessment. Chow and Chapman (2013) suggest that those organizations that implement such processes may be perceived as "technologically advanced, trendy and innovative" (p. 93). If an application is well-designed and enjoyable, candidates may be more willing to use it and even share it with their friends on social media. Ideally, this would lead to higher brand awareness and more potential applicants for the organization.

At the same time, organizations should not abuse gamification to create false expectations, because unrealistic job previews may increase attrition rates among newly hired employees, thereby countervailing the alleged benefit of more applications and hires (Armstrong et al., 2016). To harness this promising potential, each application should be designed in accordance with the individual needs of an organization.

Diercks (2013) suggests four different design possibilities, which are divided by two dimensions: the *objective* as well as the *methodology* of an application. Whereas *objective* distinguishes applications that orient users regarding their *P-O* or *P-J fit*, *methodology* refers to self-assessments that are based on *diagnostics* or *simulations*. An application can take a survey-like form, in which the most relevant constructs are operationalized and answers evaluated via algorithms. At the end, users receive condensed, automatically generated feedback on their match (i.e., diagnostics). Alternatively, the application can be designed in a way that enables users to experience and comprehend the relevant aspects of a job or organization by themselves (i.e., simulations). Here, the feedback is less decisive than the process in which candidates discover their abilities, skills, and passions while playing. In other words, the application motivates users to reflect if they find the kind of job or organization displayed attractive, instead of directly telling them about their suitability. Simulations may hence be more time-consuming and costly during development. An understanding of these design possibilities enables organizations to create applications that match both their recruitment goals and financial budgets (Diercks, 2013).

2.4. Technology acceptance

Technology acceptance research is considered one of today's most developed fields in information systems research (Venkatesh et al., 2007). There are several traditional models in the acceptance research (e.g. TRA and TAM) which are united by Venkatesh et al. (2003) in their *Unified Theory of Acceptance and Use of Technology* (UTAUT). The UTAUT originally targets IS use in organizations. Users' behavioral intentions are determined by their *performance expectancy* ('the degree to which using a technology will provide benefits to consumers in performing certain activities'), *effort expectancy* ('the degree of ease associated with consumers' use of technology'), *social influence* ('the extent to which consumers perceive that important others [e.g., family and friends] believe they should use a particular technology'), and *facilitating conditions* ('consumers' perceptions of the resources and support available to perform a behavior') (Venkatesh et al., 2012, p. 159). The model also takes the moderating variables of gender, age, experience, and voluntariness into account. In a bibliometric analysis of 450 citations and 43 studies about UTAUT, Williams et al. (2012) conclude that it offers 'a useful tool by which to evaluate the potential for success of new technology initiation, and helps identify factors likely to influence adoption of technology' (p. 58).

Hamari and Koivisto (2015) and Lowry et al. (2013) further refine the well-established UTAUT. As illustrated in Figure 1, the consumer-oriented UTAUT2 adds *hedonic motivation* ('the fun or pleasure derived from using a technology'), *price value* ('consumers' cognitive tradeoff between the perceived benefits of the applications and the monetary cost for using them') and *habit* ('perceptual construct that reflects the results of prior experiences') as three novel factors that determine the behavioral intention (Venkatesh et al., 2012, p. 161).

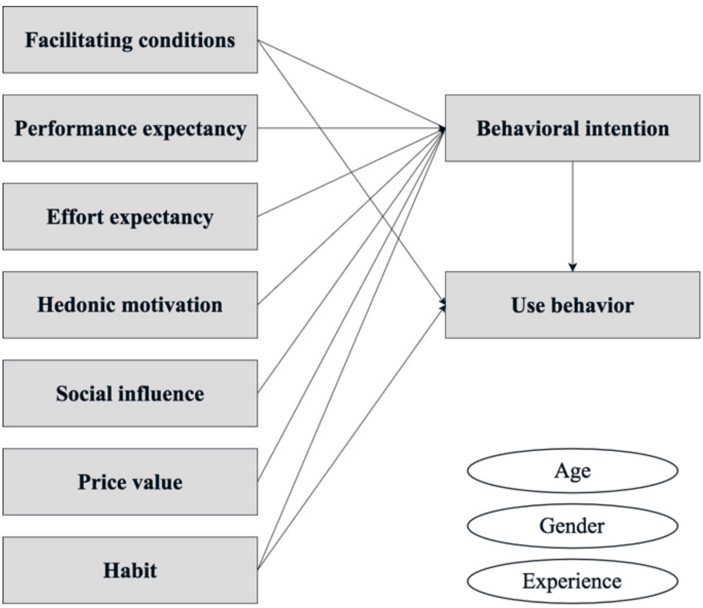

Figure 1. Unified Theory of Acceptance and Use of Technology 2

Source: Venkatesh et al., 2012

In terms of gamification, Hamari and Koivisto (2015) emphasize the dearth of acceptance research and underline the necessity of further studies in this field. Previous research has focused on the acceptance of gamification in general (i.e., non-HR and non-recruitment related) contexts. Despite the growing interest in gamification research for business processes (Machado Leitão et al., 2021), little is known about the acceptance of gamified recruitment and its corresponding design elements.

Some studies investigated the effects of personality (Codish, Ravid, 2014) or demographics (Koivisto, Hamari, 2014) on the acceptance of gamification in non-recruiting contexts. Baptista and Oliveira (2017) applied a modified UTAUT2 to examine the acceptance of gamification in mobile banking services. The authors emphasize the importance of proper application design to enable the benefits of implementation, such as increased customer acceptance and satisfaction. Laumer et al. (2012) tested the TAM in a serious game recruiting context and suggest that the use of game elements may increase the acceptance of traditional self-assessments among applicants. In a second step, such gamified self-assessments are more likely to be used if they display a variety of job aspects, such as tasks or skill requirements, and can easily be accessed and played on the career website. A requirement to download the application first might prevent a significant share of potential candidates from actually using the application. However, measuring the acceptance of gamification with traditional models (e.g., TRA, TAM, or UTAUT) can be quite difficult as they are specifically tailored for utilitarian information systems (Codish, Ravid, 2014).

In this paper, we apply UTAUT2 as our theoretical framework for three reasons. First, the model is robust. Venkatesh et al. (2012) document that that it explains 74% of the variance in behavioral intention and 52% in technology use, respectively. Second, the explanatory power of the unified approach is higher than those of its integrated theories, as shown by Bradley (2012). Third, the consumer-oriented approach of UTAUT2 better matches the needs of our research than the original UTAUT. Candidates can voluntarily choose which organization they want to apply for and are hence not required (by an organizational mandate) to use gamified recruitment processes.

To summarize, as gamified recruitment is a relatively new field of study, we validate, expand, and adapt UTAUT2 for the purpose of this study (RQ 3).

3. Methods

To answer our three research questions, we conducted focus groups as our data collection method. Focus groups allow for a discussion and exchange of ideas and thereby support participants in developing opinions about a novel topic like gamification (Finch, Lewis, 2003). We have analyzed the acquired data by means of qualitative content analysis (Mayring, 2014). In the following, we discuss the stimuli we used during the focus groups as well as our data collection and analysis.

3.1. Stimuli

To give participants a better feeling of the different design possibilities and ultimately increase the depth of their answers during the focus groups, we employed two P-O fit assessment applications as stimulus: *Heineken's Go Places* as an example of a *simulation* and *Air New Zealand's Be the next* as a *diagnostics-based* application. Both applications foster a variety of positive and negative aspects to discuss.

In *Heineken's Go Places*, potential candidates can check their fit with Heineken's organizational culture in a playful, multimedia-based job interview. After answering twelve dual choice questions (e.g. 'Would you rather be a) world famous or b) have strong roots?'), candidates immediately receive feedback in the form of a personality profile. Although this approach sounds rather like a diagnostics-based self-assessment, *Go Places* does not explicitly calculate the match between user and company values. Instead, each of the eight possible profiles (e.g., achiever, pioneer, or enthusiast) is formulated in a comparably positive way regarding the fit with the organization. Therefore, it is less about the specific profile category and more about the process of getting to know *Heineken* as an employer and one's P/O fit. In between the twelve questions, candidates receive additional information about Heineken's values, brands, activities, locations, and employees. The whole process is video-based, i.e., both the questions and the company characteristics are presented via film clips. The scenes are connected by the underlying narrative frame, in which a fictitious interviewer guides the user through the self-assessment.

In *Air New Zealand*'s gamified recruitment process *Be the next*, users assess their fit with the airline's organizational culture. The application follows a diagnostics-based approach, as candidates answer 16 questions and receive condensed feedback on their match at the end of the process. The questions are illustrated by a comic design and most of them must be answered on a ten-point scale. For example, when asked 'You've [*sic*] got a tricky problem to solve at work. What do you do first?', the potential candidate can choose to 'Dive into research' or 'Throw ideas around with your team'. The comic design changes dynamically as the user moves the cursor to either end of the scale. Whereas on one end of the scale, an avatar sits in front of a computer next to bookshelves and inside a quiet room, on the other end the same avatar discusses with her colleagues and clips ideas to a whiteboard.

3.2. Data collection and sample

Our data collection via focus groups followed a structured process (Morgan, 1996). Comparability between the focus groups was ensured by an interview guide that was derived from the theoretical foundations and previous research. We categorized the interview questions along five different themes of gamified recruitment structured from broad to narrow so that participants could first become comfortable talking about the topic in general and then share more specific and detailed insights later (Krueger, Casey, 2015). The interview guide consisted of open questions only. The order of questions was flexible to maintain a natural flow of the discussion. Both stimuli and the interview guide were pretested for clarity and logical structure.

We recruited ten participants for the study by sending invitations via instant messenger or email. In dividing these ten participants into two groups of five, we adopted purposive sampling to ensure that the sample was homogeneous enough for participants to effectively exchange their viewpoints, but as diverse as possible to fully saturate the topic (Morgan, 1996). All ten participants were graduate job seekers who had recently finished one of two postgraduate study programs. Therefore, the sample was representative for the purpose of the study. The sample was further characterized by six different nationalities, an age range from 23 to 32 years, different study backgrounds, and an equal representation of women and men.

The focus groups were conducted in the classrooms of an international university. After a brief introduction to the topic, agenda, and rules, each focus group participant could independently experiment with both stimuli applications for around 20 minutes. After the experimentation phase, participants went on to discuss their experiences and perceptions. For most participants, it was actually their first encounter with gamified recruitment.

To facilitate a more focused discussion, participants were asked to picture themselves in a real job search process. Candidate behavior tends to be comparable in simulated and real job seeking situations (Chapman et al., 2005). To minimize order biases, the first focus group started to discuss *Go Places*, while the second focus group began with *Be the next*. The interviewer controlled the influence of more dominant speakers by explicitly seeking out the opinions of others (Litosseliti, 2003). The total duration of the discussion was 64 minutes in the first and 94 minutes in the second focus group, respectively. Both focus groups reached saturation, and even after repeated enquiring, no additional ideas came up in the discussions (Krueger, Casey, 2015).

After conducting the focus groups, the audio recordings were transcribed using the analysis software MAXQDA12. We chose a word-by-word transcription to guarantee the integrity of the data and minimize the risk of premature interpretations (Poland, 2003). Paragraphs and participants received individual reference numbers so that statements could be precisely retrieved and subsequently interpreted.

3.3. Data analysis

Building upon the theoretical foundations of gamification research (Cho, Lee, 2014), we used Mayring's (2014) qualitative content analysis to analyze the data. This standardized approach serves to maximize the objectivity and reliability of our study, given its transparent documentation of the research process.

Before coding and analyzing the data, we specified the three different perspectives from the research questions: the individual, the organizational, and the designer perspective. For each perspective, we developed a comprehensive category system. The *individual perspective* included the categories *motivation* to use gamified recruitment and *evaluation* of such applications compared to traditional recruitment. The *organizational perspective* involved *types and goals* of companies that offer such applications and the *employer image* (i.e., the perception of these organizations as potential employers). The *designer perspective* consisted of two categories, *design elements* and *design and implementation guidelines*. Design elements referred to positive (*acceptance drivers*) and negative elements (*acceptance barriers*) of gamification. Design and implementation guidelines were derived from the recommendations elaborated in the theoretical foundations and an analysis of five best practice examples. The guidelines included the following ten common characteristics: *easy IT access*, *intuitive gameplay*, *support functions*, *hedonic aspects*, *duration*, *appropriate design*, *feedback*, *sociality*, *job search relevance*, and *transparency*.

Based on the category system, we developed *coding guidelines* to ensure intersubjectivity. We explicitly defined each category, supported it with an anchor example from the transcripts, and set clear rules for the correct use of the codes (Mayring, 2014). Whenever a participant's statement described one of the categories, it was coded as such. After coding the first 20 percent of the material in this manner, the category system was revised, and the process was repeated for a second time after a complete run-through. As a result, the original categories and subcategories were systematically refined so that the category system covered all relevant aspects mentioned by the participants and distinguished more precisely between categories. Specifically, *easy IT access* was adapted to *IT requirements*, *intuitive gameplay* to *gameplay* (absorbing *hedonic aspects*), and *appropriate design* to *interface design.* For a better understanding of the positive and negative elements of gamified recruitment, each coding of an *acceptance driver* or *acceptance barrier* was subject to an additional coding among nine subcategories in the area of *design and implementation guidelines.* Table 2 illustrates the final category system and the respective coding distributions. In total, 394 codings were employed during the data analysis process.

Table 2
Category system with number of codings per category and focus group

Perspectives	Categories	Subcategories	Focus Group 1	Focus Group 2	Total
Individual	motivation	–	6	13	19
	evaluation		17	23	40
Organiza-tional	types and goals	–	11	31	42
	employer image		12	14	26
Designer	design elements	acceptance drivers	25	41	66
		acceptance barriers	18	25	43
	design and implementation guidelines	IT requirements	6	8	14
		gameplay	19	26	45
		support functions	2	6	8
		interface design	7	9	16
		feedback	11	7	18
		duration	4	5	9
		job search relevance	8	17	25
		transparency	6	11	17
		sociality	4	2	6
Total			156	238	394

4. Findings and discussion

4.1. Individual perspective

At the individual level, our analysis identified three different motivations to engage in gamified recruitment processes among graduate job seekers (RQ1): 1) to find out more about the organization, 2) to find out more about themselves, and 3) to enjoy the process of playing.

First, the participants stated that they had learned interesting facts about the products, brand values (P7-87), career paths, talent development opportunities (P9-70), business activities, and innovativeness of a company through gamified recruitment (P6-292). Such learnings can be useful for a candidate because 'if you think this company suits you, you can apply here' (P7-61). Second, it may be beneficial for applicants to find out more about themselves, as explained by P10: 'for me, it was a really good experience to know about what kind of personality I have and what the other people expect from me' (P10-17). Third, the use of gamified recruitment may become even more motivating the longer a candidate has been actively looking for a job, as these applications may provide an 'enjoyable experience' (P6-117) or a 'fun pause' (P5-55) from conventional methods. During a naturally rather tedious job search, this fun factor might be a key incentive for job seekers to use gamification.

The motivation to use gamified applications, however, may significantly depend on the availability of attractive jobs. Both P6 and P7 said that they would only use such applications after encountering attractive vacancies in the organization (P6-191-193; P7-187). Companies may hence provide links to their current job postings before, during, and after candidates use the gamified application. This transparent connection between the self-assessment and open positions might give candidates more confidence in using it, because they could see that the organization is hiring and that the whole process is not 'just for fun' (P7-187). After all, they 'expect that using [a] self-assessment will support them in […] simplifying their decision as to whether to apply for a job or not and improve the chances of getting hired' (Laumer et al., 2012, p. 234). In case there are currently no attractive vacancies available, candidates could at least be encouraged to submit unsolicited applications.

In comparison to more traditional recruitment methods, the participants described the gamification approach as 'a cool process of [job] application' (P5-17) and 'a positive [candidate] experience' (P7-288), which differs from the 'generally boring kind of things which every recruiter do [*sic*] these days' (P6-117).

4.2. Organizational perspective

Our analysis identified three different goals which organizations aim to achieve by means of gamified recruitment (RQ2): authenticity, uniqueness, and attractiveness. Concerning authenticity and uniqueness, organizations could use gamification to genuinely differentiate from competitors. Gamification can sharpen the employer brand and thereby generate an edge over organizations that compete for the same talent. A well-defined employer brand may prevent less fitting candidates from applying and engage more fitting applicants (P7-128). Different design possibilities of gamified recruitment allow for a variety of application designs that highlight the individual characteristics of an organization, as Diercks (2013) indicates.

Concerning attractiveness, it is necessary to appeal to the target group. Several participants speculated that gamified recruitment may be used to attract primarily millennials, as the following statement indicates: 'So the whole thing which we did right now was for the young generation' (P1-143). In line with this perception, Nair and Sadasivan (2019) claim that the affinity for technology among young generations contributes to the potential of gamified applications.

The participants questioned, however, whether gamified recruitment could be a successful tool for all types of organizations. They reasoned that the appropriateness of such applications may depend on the industry (P6-234) and the organization's type of product offerings (P9-263). Another critical factor could be the employer brand. P9 argued that '[…] if the company has this sophisticated, traditional brand, probably it will contradict [to use such applications] because the essence of the gamification is, like, to make the process funky' (P9-263). P8, on the contrary, emphasized that any organization could implement gamified recruitment as long as it effectively conveys the employer brand (P8-260). Gamified recruitment activities of an investment bank, for example, should probably be designed differently than *Heineken's Go Places* (perceived as pushy, P6-138) or *Air New Zealand's Be the Next* (described as childish, P2-87).

In line with Chow and Chapman's (2013) suggestion of positive image effects through gamified recruitment, the participants characterized such organizations as 'modern' (P8-63), 'cool' (P4-32), and 'young, fun, innovative' (P2-52). P4 described these companies as 'more human, more approachable' (P4-30) because gamified self-assessments could effectively reduce the perceived distance between successful organizations and graduates.

4.3. Designer perspective

The acceptance of any application among the target group is crucial for the application's success. Our analysis of the participants' feedback allowed us to derive *design and implementation guidelines* (RQ3). These guidelines cover the following aspects, which we discuss in the following: IT requirements, gameplay, duration, interface design, job search relevance, feedback, transparency, support functions, and sociality.

IT requirements. While testing the applications, the participants experienced several technical errors. As a result, they experienced negative feelings of confusion (P8-140), frustration (P7-9), desperation (P3-121), or anger (P1-90). After finishing Air New Zealand's application, for example, P8 uploaded her CV, but was unable to receive the final feedback no matter how often she hit the 'submit' button. When asked about the experience, she called it 'a waste of time' (P8-110). Extensive pretests are clearly a prerequisite for reducing the risk of technical failures that spoil the user experience. Organizations should hence consider optimizing their applications for all devices to prevent triggering negative emotions of users. This corresponds with Laumer et al. (2012) who underline the importance of easy accessibility.

Gameplay. The second major pillar of gamified recruitment is the gameplay of an application, which was (with 45 codings) the most intensely discussed design element in the focus groups. Participants found that the application should be intuitive, entertaining, engaging, and complete in terms of user experience. P2 highlighted the importance of intuitive gameplay as follows: '[…] I do not want to look into directions, I just want to go for it' (P2-167). To avoid confusion among users, P8 suggested 'to keep the questions very clear and simple' (P8-176). Unsurprisingly, many participants called intuitive gameplay one of the most important design elements (P1-185; P3-171; P7-187; P9-206).

A second key component of gameplay is entertainment. That is, the application should induce fun or surprise experiences in users compared to traditional recruitment. When describing her emotions while testing the applications, P1 said: 'I felt happy through [*sic*] the whole time. Especially, I remember, once or twice, that I was laughing because the thing was really funny' (P1-3). The next level of successful gameplay would be an application that is not only entertaining but also engaging. Engagement goes beyond providing fun and novel experiences, because it seeks to gain the users' full attention and maximize their involvement. In *Go Places*, the narrative frame created by the interview situation managed to engage users (P3-26; P10-17). P2 expressed that she 'was positively surprised […] because of this process of the guy leading you through something, and you could kind of influence on where he is going' (P2-87). According to Sailer et al. (2013), a narrative frame can make it easier for users to grasp the situational context of gamified processes. Giving users control over how the storyline unfolds may hence increase their acceptance.

To further increase engagement, gamified self-assessments could also involve challenges users need to solve in order to advance in the process. P4 suggested that, for each question, there should be different answers to choose from, and that picking the 'wrong' option (e.g., in terms of cultural values or desired behavior) could lead to a classic game-over scenario like in video games (P4-102; P4-203). Regardless of potentially challenging definitions of 'wrong', this approach could increase the user engagement because they feel something is at stake. The application, however, has to deny a restart of the same user to inhibit strategical lying. An organization would otherwise get a wrong impression of the candidate.

Duration. One aspect that influences how well the gameplay of an application is received by its users is its duration. A gamified self-assessment should not take too long. According to P6, '[it] should be exciting to me till the end, otherwise I will just leave in between and [I] will never go back' (P6-179). Users might become increasingly impatient as more and more organizations implement gamified recruitment. Koivisto and Hamari (2014), for example, warn that novelty effects tend to wear off with longer exposure time.

Interface design. Whereas gameplay describes the *feel* of an application, interface design refers to its *look*. The interface should be clear and professional so that candidates do not question the legitimacy and seriousness of the application. P8 pointed out that '[…] if it looks like a child game, then I cannot take it seriously' (P8-202), which supports Dale's (2014) argument that the application style should be consistent with the corporate identity. It is essential to design an interface purposefully, in a way that depicts the organizational culture and makes the application easy to use, and economical in the sense that it supports the gameplay without causing information or effect overload on users. Participants also suggested a multimedia-based approach, combining images (P2-87), videos (P5-74), and audio (P4-98). The application should appeal to both auditory and visual senses, thereby contributing to a deeper and more memorable user experience.

Job search relevance. Another major design element in gamified recruitment is its relevance for the users' job search. This relevance is primarily driven by the clarity of both the application tasks, job openings, and characteristics of a job/organization. Several participants stated that they did not see the purpose of the application task in *Be the next*, requiring users to click on as many moving candies as possible in 60 seconds (P3-5; P8-14). Instead of requiring users to fulfill similarly hedonistic, yet unrelated game tasks, each application task should demonstrate a clear connection to the superordinate goal of orienting users regarding their P-O/P-J fit. In line with this argument, Callan et al. (2014) postulate that successful gamification should always contain deeper meaning for the user.

Job openings should also be prominently and transparently displayed to boost candidate attraction. In the case of *Heineken's Go Places*, P2 and P5 said that they had no idea for what kind of jobs they were being assessed (P2-20-22; P5-21-23). P1 argued that, after finishing the application, 'there should come the jobs which […] fit for your psychology' (P1-65). By playing the application, the users could find out if they like the job/organization or not while suitable and interested candidates could be attracted even more by presenting them with those job openings that best match their profile.

Furthermore, gamified recruitment should clearly highlight the characteristics of a job/organization. According to the participants, this should include the organizations' expectations (P7-212; P10-17), brand values (P7-87; P9-135), or career opportunities (P7-87; P9-70). Such an approach would debilitate Bogost's (2014) criticism of gamification as being manipulative and ineffective.

Feedback. Whereas job search relevance refers to offering users suitable job openings, feedback means explaining the reasons why a candidate would be a fitting match for any of these positions or the organization as a whole. The feedback should be meaningful and specific, as consistently suggested by the participants. For example, P5 wondered if each of the eight final character profiles in *Go Places* (equally) qualified for working at *Heineken* (P5-15). After all, 'you do not know which one they want' (P2-16). The participants also suggested to formulate the feedback positively to avoid disappointment in users (P1-67; P4-34). P5 described her feelings with the final evaluation in *Go Places* as follows: '[…] it is really flattering […] I really like it. I love when I can read some nice feedback about myself' (P5-116). For ill-fitting candidates, the application could suggest that the organization may not yet be the perfect fit, but that those candidates could learn more about its organizational values, for example, by inviting them to get in touch through social media to clarify mutual expectations.

Transparency. In terms of transparency, the participants expressed mixed opinions about how their data may be used by the organizations. Both applications failed to clearly communicate their actual purpose to the participants, i.e., to anonymously help them assess their individual P-O fit. According to Diercks (2013), it is crucial to educate candidates about why to use a gamified recruitment process in the first place to increase its acceptance.

Support functions. Given the importance of intuitive gameplay, support functions may not be the most critical design element. Yet, they may improve the acceptance of an application by eliminating users' minor uncertainties and issues. By providing audio output, for example, users may not have to read all the information displayed, which can increase usability (P2-52). To guarantee a smooth game start, the application could offer a demo question (P7-220) or visual aid (P7-228), introducing the gameplay.

Sociality. According to the participants, the least important design element in gamified recruitment seems to be sociality (e.g., P8-204). Sociality generated the least entries (6) among all categories. P1 even called it a negative element 'because I do not want anyone to see my results' (P1-185-189). This result is interesting because social features can be considered a key motivational driver of gamified processes in other environments (Hamari, Koivisto, 2015). P4 suggested that sociality could help spread the application, with users trying to beat their friends' scores (P4-194).

4.4. Acceptance of gamified recruitment elements

Every design element discussed can be either an *acceptance driver* or an *acceptance barrier*, depending on how effectively it is implemented in an application. Here, effectiveness refers to the extent to which designers adhere to the results of the *design and implementation guidelines*. A clear, professional, and multimedia-based interface, for example, is likely to improve the acceptance of an application. Another interface that looks rather childish and has not been optimized for usability, on the other hand, may have the opposite effect. All design elements must not be considered standalone but in interaction with each other. We thus consolidate our findings in a *model for acceptance of gamified recruitment elements* (Fig. 2).

In section 4.3, we identified four fundamental design elements in gamified recruitment: *IT requirements*, *gameplay*, *interface design*, and *job search relevance*. They hence serve as the basis of our model, each subsuming several of the other design elements. As even the most entertaining activity may become boring after some playtime, *duration* was incorporated into *gameplay*. *Job search relevance* includes *transparency* because an open communication about purpose and data usage can tell candidates how the application may help them in their job search process. The components of *feedback* were distributed among *gameplay* (i.e., the display of the progress until completion) and *job search relevance* (i.e., the final evaluation of the users regarding their P-O/P-J fit). To improve the look and feel of an application, *support functions* were introduced as the fifth element of the model. Ideally, they may not be necessary in an application with perfect gameplay and interface design, but they can help prevent minor issues. Last, we integrated *sociality* as a *support function* due to its potential to raise awareness for the application among the target group.

The model does not only offer a more concise and condensed version of each guideline. It also manages to connect our findings to the well-established constructs of UTAUT2. *IT requirements* are related to the *facilitating conditions* in our model, because without the necessary technical infrastructure, the usability of an application is likely to be limited (with negative effects on its acceptance). With users seeking advantages for their job search when using gamified recruitment processes, *job search relevance* resembles *performance expectancy* of UTAUT2. Regarding the look and feel of an application, the users' acceptance may depend on how easy and enjoyable it is to use. As a result, both *gameplay* and *interface design* can be linked to *effort expectancy* and *hedonic motivation*. If the gamified recruitment process offers a sharing function for social media, there could also be a touch point between the *support functions* and *social influence*. However, such a sharing function is likely to be less decisive for the success of the application than the other four factors. The two remaining constructs of UTAUT2 (*price value*, *habit*) may not be relevant for the acceptance of gamified recruitment, as applications are naturally free of charge and still rarely used by organizations.

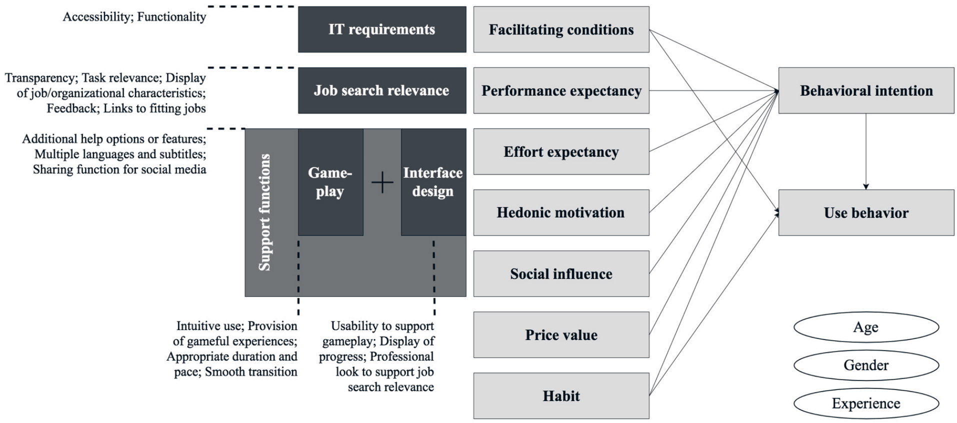

Figure 2. Model for acceptance of gamified recruitment processes

5. Conclusion and outlook

5.1. Theoretical implications

Our research represents one of the first empirical studies about gamification in recruitment processes. It sheds light on the triggers of successful implementation of gamification in this area. More specifically, the theoretical implications of our research are threefold. First, we have identified motivations to engage with gamified recruitment (RQ1). We have closed the missing conceptual and empirical link between recruitment, user motivation, and gamification on an individual perspective.

Second, we have unveiled strategies to tackle organizational challenges imposed by demographic changes and talent shortage. Our findings indicate the relevance of organizational goals to transport authenticity, uniqueness, and attractiveness in the context of gamified recruitment.

Third, we have gained insight into the drivers of gamified recruitment acceptance: IT requirements, gameplay, interface design, job search relevance, and support functions. Our findings enable the deduction of clear, concise, and comprehensive guidelines for effective design and implementation of gamified recruitment elements. We provide the first study to connect the requirements of gamified recruitment for graduate job seekers with the postulates of acceptance research. These guidelines can now be validated through quantitative follow-up studies.

Furthermore, our study represents one of the first attempts to apply UTAUT2 to the field of gamified recruitment, an expansion from more traditional information systems. The results of our qualitative content analysis suggest that five of the seven constructs in UTAUT2 may influence the behavioral intention of users to engage in gamified recruitment: *performance expectancy*, *effort expectancy*, *hedonic motivation*, *facilitating conditions*, and *social influence*. Venkatesh et al. (2012) suggest that in non-organizational settings hedonic motivation plays a more decisive role for the behavioral intention than performance expectancy. With users of gamified self-assessments primarily trying to find a suitable employer, however, one could assume that performance expectancy is the most important driver in a job search context.

5.2. Practical implications

The results of our study are particularly relevant to HR and IT professionals who need to facilitate the successful design and implementation of gamified recruitment elements. The gamification market is expected to grow massively over the next years and the dearth of research still makes it difficult to explain how gamified recruitment works and why. Organizations should have an essential, evidence-based understanding of the topic.

Our findings help designers to tackle the four fundamental design elements of their implementation projects. These four elements provide concrete applicable insights and thus a check-mark for future developments. Consequently, more attractive applications can be developed which allow to win top talents. These targeted applications facilitate more effective recruiting activities and may hence lead to cost savings in the recruitment process.

Despite its fun nature, gamification should be treated with the same diligence as other business activities. The success of gamified recruitment may significantly depend on the support and commitment of the organization and its stakeholders. HR professionals should not consider gamified recruitment a panacea in winning the war for young talent. After all, graduate job seekers are a highly heterogeneous population, and not all of them may show the same positive attitude toward gamified recruitment as the participants of the focus groups. Accordingly, gamified recruitment is a supplement to, rather than a replacement of, traditional recruitment methods. A combination of gamified and traditional recruitment can be a viable strategy and beneficial to organizations for three reasons. First, media richness has a positive effect on the candidate attraction because the various channels provide job seekers with more relevant employer information (Allen et al., 2004). Second, outside-the-box recruitment activities can make it easier for candidates to grasp and potentially identify with the organizational characteristics of an employer (Cable, 2007). This is important because, third, it seems to be a widespread fear of graduate job seekers to not find a job that corresponds with their personalities (Universum, 2014).

5.3. Limitations and future research

Our study is subject to several potential limitations. First, the sampling of our focus groups ensured a high degree of diversity in attitudes and ideas generated. Although both focus groups reached saturation (i.e., no new aspects were discussed at the end of the data gathering) our focus group members were exclusively graduate students, who are about to enter the labor market. A focus group of young professionals who already have work experience and are looking for new jobs might bring forth different opinions about gamified recruitment. The same holds true for more experienced middle managers and executives.

Second, the two applications we have used as stimuli may have influenced the results as well. Other gamified recruitment applications might have triggered different ideas and discussions. We have chosen both applications due to their wide array of positive and negative characteristics. Selecting two best practice examples instead, might have led participants to overgeneralize the benefits of gamified recruitment, whereas two poorly designed applications might have had the opposite effect.

The timing of the study may be another relevant factor. After all, as gamified recruitment applications are still in their infancy, user attitudes may change accordingly. There might be novelty effects in the adoption of gamification. Longitudinal research would help explore whether habituation or fatigue impact the revealed motivations. It would also be exciting to see how the gamification of recruitment processes may develop over time. Finally, the factors of our model for acceptance of gamified recruitment processes need to be operationalized and hence there is a need for a quantitative analysis of validity and robustness. Nonetheless, with this study we have been able to demonstrate the benefits of gamified recruitment for organizations as well as for individuals and shed light on the factors that must be considered in the design process.

References

[1] Allen, D.G., Van Scotter, J.R. and Otondo, R.F. (2004) 'Recruitment communication media: Impact on prehire outcomes', *Personnel Psychology,* vol. 57, pp. 143–171.

[2] Alsawaier, R.S. (2018) 'The effect of gamification on motivation and engagement', *The International Journal of Information and Learning Technology,* vol. 35, No. 1, pp. 56–79.

[3] Armstrong, M.B., Landers, R.N. and Collmus, A.B. (2016) 'Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management', in Gangadharbatla, H. and Davis, D.Z. (ed.) Emerging research and trends in gamification, Hershey: Information Science Reference.

[4] Baptista, G. and Oliveira, T. (2017) 'Why so serious? Gamification impact in the acceptance of mobile banking services', *Internet Research,* vol. 27, No. 1, pp. 118–139.

[5] Barber, A.E. (1998) Recruiting employees: Individual and organizational perspectives, Thousand Oaks: SAGE.

[6] Bina, S., Mullins, J. and Petter, S. (2021) 'Examining Game-based Approaches in Human Resources Recruitment and Selection: A Literature Review and Research Agenda, Proceedings of the 54th Hawaii International Conference on System Sciences, pp. 1325–1334.

[7] Blohm, I. and Leimeister, J.M. (2013) 'Gamification: Design of IT-based enhancing services for motivational support and behavioral change', *Business & Information Systems Engineering,* vol. 5, No. 4, pp. 275–278.

[8] Bogost, I. (2014) 'Why gamification is bullshit', in Walz, S.P. and Deterding, S. (ed.) The gameful world: Approaches, issues, applications, Cambridge: MIT Press.

[9] Bradley, J. (2012) 'If we build it they will come? The Technology Acceptance Model' in Dwivedi, Y.K., Wade, M.R. and Schneberger, S.L. (ed.) Information systems theory: Explaining and predicting our digital society, vol. 1, New York: Springer.

[10] Breaugh, J.A. (2008) 'Employee recruitment: Current knowledge and important areas for future research', *Human Resource Management Review,* vol. 18, pp. 103–118.

[11] Breaugh, J.A., Macan, T.H. and Grambow, D.M. (2008) 'Employee recruitment: Current knowledge and directions for future research', in Hodgkinson, G.P. and Ford, J.K. (ed.) International review of industrial and organizational psychology, Chichester: Wiley & Sons.

[12] Buil, I., Catalán, S. and Martínez, E., (2020) 'Understanding applicants' reactions to gamified recruitment', *Journal of Business Research*, vol. 110, pp. 41–50.

[13] Cable, D.M. (2007) Change to strange: Create a great organization by building a strange workforce, Upper Saddle River: Pearson Education.

[14] Callan, R.C., Bauer, K.N. and Landers, R.N. (2014) 'How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences', in Reiners, T. and Wood, L.C. (ed.) Gamification in education and business, Cham: Springer.
[15] Chamorro-Premuzic, T. (2015) 3 emerging alternatives to traditional hiring methods, [Online], Available: https://hbr.org/2015/06/3-emerging-alternatives-to-traditional-hiring-methods [2 Feb 2022].
[16] Chapman, D.S., Uggerslev, K.L., Carroll, S.A., Piasentin, K.A. and Jones, D.A. (2005) 'Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes', *Journal of Applied Psychology,* vol. 90, No. 5, pp. 928–944.
[17] Cho, J.Y. and Lee, E. (2014) 'Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences', *The Qualitative Report*, vol. 19, no. 32, pp. 1–20.
[18] Chow, S. and Chapman, D. (2013) 'Gamifying the employee recruitment process', Proceedings of the First International Conference on Gameful Design, Research, and Applications (Gamification 2013), Stratford, pp. 91–94.
[19] Codish, D. and Ravid, G. (2014) 'Personality based gamification: How different personalities perceive gamification', Proceedings of the European Conference on Information Systems (ECIS 2014), Tel Aviv.
[20] Dale, S. (2014) 'Gamification: Making work fun, or making fun of work?', *Business Information Review*, vol. 31, No. 2, pp. 82–90.
[21] Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011) From Game Design Elements to Gamefulness: Defining 'Gamification', Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, Tampere, Finland, pp. 9–15.
[22] Diercks, J. (2013) 'Warum Personalauswahl ein beidseitiger Prozess ist – die Verbesserung der Selbstauswahl durch Self-Assessment Verfahren und Berufsorientierungsspiele' in Diercks, J. and Kupka, K. (ed.) Recrutainment: Spielerische Ansätze in Personalmarketing und -auswahl, Wiesbaden: Springer Gabler.
[23] Doherty, R. (2010) 'Getting social with recruitment', *Strategic HR Review,* vol. 9, No. 6, pp. 11–15.
[24] Finch, H. and Lewis, J. (2003) 'Focus groups', in Ritchie, J. and Lewis, J. (ed.) Qualitative research practice: A guide for social science students and researchers, London: SAGE Publications.
[25] Georgiou, K. and Nikolaou, I. (2020) 'Are Applicants in Favor of Traditional or Gamified Assessment Methods? Exploring Applicant Reactions Towards a Gamified Selection Method', *Computers in Human Behavior*, vol. 109, 106356.
[26] Gilch, P.M. and Sieweke, J. (2021) 'Recruiting digital talent: The strategic role of recruitment in organisations' digital transformation', *German Journal of Human Resource Management: Zeitschrift für Personalforschung*, vol. 35, No. 1, pp. 53–82.

[27] Hamari, J. (2017) 'Do badges increase user activity? A field experiment on the effects of gamification', *Computers in Human Behavior,* vol. 71, pp. 469–478.

[28] Hamari, J. and Keronen, L. (2017) 'Why do people play games? A meta-analysis', *International Journal of Information Management,* vol. 37, No. 3, pp. 125–141.

[29] Hamari, J. and Koivisto, J. (2015) 'Why do people use gamification services?', *International Journal of Information Management,* vol. 35, pp. 419–431.

[30] Holm, A.B. (2012) 'E-recruitment: towards an ubiquitous recruitment process and candidate relationship management', *German Journal of Human Management: Zeitschrift für Personalforschung*, vol. 26, no 3, pp. 241–259

[31] Huotari, K. and Hamari, J. (2017) 'A definition for gamification: Anchoring gamification in the service marketing literature', *Electronic Markets,* vol. 27, No. 1, pp. 21–31.

[32] Koivisto, J. and Hamari, J. (2014) 'Demographic differences in perceived benefits from gamification', *Computers in Human Behavior,* vol. 35, pp. 179–188.

[33] Koivisto, J. and Hamari, J. (2019) 'The rise of motivational information systems: a review of gamification research', *International Journal of Information Management*, vol. 45, pp. 191–210.

[34] Krueger, R.A. and Casey, M.A. (2015) Focus groups: A practical guide for applied research, Thousand Oaks: SAGE Publications.

[35] Langer, M., König, C., and Fitili, A. (2018) 'Information as a double-edge sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection', *Computers in Human Behavior*, vol. 81, pp. 19–30.

[36] Laumer, S., Eckhardt, A. and Weitzel, T. (2012) 'Online gaming to find a new job: Examining job seekers' intention to use serious games as a self-assessment tool', *German Journal of Human Management: Zeitschrift für Personalforschung*, vol. 26, Issue 3, pp. 218–240.

[37] Litosseliti, L. (2003) Using focus groups in research, London: Continuum.

[38] Lowry, P.B., Gaskin, J., Twyman, N., Hammer, B. and Roberts, T.L. (2013) 'Taking "Fun and Games" Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM) to increase understanding of adoption of hedonically motivated systems', *Journal of the Association for Information Systems,* vol. 14, No. 11, pp. 617–671.

[39] Machado Leitão, T., Lima Navarro, L.L., Flórido Cameira, R. and Silva, E.R. (2021) 'Serious games in business process management: a systematic literature review', *Business Process Management Journal*, vol. 27, No. 3, pp. 685-721.

[40] MarketsandMarkets (2020) Gamification Market by Component: Global forecast to 2025, [Online], Available: https://www.marketsandmarkets.com/Market-Reports/gamification-market-991.html [2 Feb 2022].

[41] Maycotte, H.O. (2015) How companies can improve recruitment and engagement with gamification, [Online], Available: https://www.forbes.com/sites/homaycotte/2015/07/07/how-companies-can-improve-recruitment-and-engagement-with-gamification/2/#3a9ad2f1486d [2 Feb 2022].

[42] Mayring, P. (2014) Qualitative content analysis: Theoretical foundation, basic procedures and software solution, Klagenfurt: Beltz.

[43] Michaels, E., Handfield-Jones, H. and Axelrod, B. (2001) The war for talent, Boston: Harvard Business School Publishing.

[44] Morgan, D.L. (1996) 'Focus groups', *Annual Review of Sociology*, vol. 22, pp. 129–152.

[45] Murawski, L. (2020) 'Gamification in human resource management. Status quo and quo vadis', *German Journal of Human Management: Zeitschrift für Personalforschung*, vol. 35, Issue 3, pp. 1–19.

[46] Nair, A. and Sadasivan, R. (2019) 'Winning the talent game: HR gamification experience for Generation Z', *International Journal on Leadership*, vol. 7, No. 1, pp. 44–49.

[47] Oprescu, F., Jones, C. and Katsikitis, M. (2014) 'I play at work – ten principles for transforming work processes through gamification' *Frontiers in Psychology,* vol. 5, [Online], Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00014/full [2 Feb 2022].

[48] Poland, B.D. (2003). 'Transcription quality' in Holstein, J. and Gubrium, J.F. (ed.) Inside interviewing: New lenses, new concerns, Thousand Oaks: SAGE Publications.

[49] PricewaterhouseCoopers (2016) Redefining business success in a changing world: CEO Survey, [Online], Available: https://www.pwc.com/gx/en/financial-services/publications/assets/key-talent-findings-in-the-financial-services-industry-feb-2016.pdf [2 Feb 2022].

[50] Raftopoulos, M. and Walz, S.P. (2013) 'Designing events as gameful and playful experiences', Conference Paper presented at the CHI 2013, Paris.

[51] Rampell, C. (2014) Your next job application could involve a video game, [Online], Available: https://www.nytimes.com/2014/01/26/magazine/your-next-job-application-could-involve-a-video-game.html [2 Feb 2022].

[52] Sailer, M., Hense, J., Mandl, H. and Klevers, M. (2013) 'Psychological perspectives on motivation through gamification', *Interaction Design and Architecture(s) Journal*, vol. 19, pp. 28–37.

[53] Scully, J., Turner, P. and Gregson, M. (2014) 'Workforce intelligence planning', in Crawshaw, J.R., Budhwar, P. and Davis, A. (ed.), Human resource management: Strategic and international perspectives, London: SAGE Publications.

[54] Stephan, M., Brown, D. and Erickson, R. (2017) 'Talent acquisition: Enter the cognitive recruiter', in Pelster, B. and Schwartz, J., Rewriting the rules for the digital age. 2017 Deloitte Global Human Capital Trends, Deloitte University Press.

[55] Strohmeier, S. (2020) 'Digital human resource management: A conceptual clarification', *German Journal of Human Resource Management: Zeitschrift für Personalforschung*, vol. 34, No. 3, pp. 345–365.
[56] Strohmeier, S. and Kabst, R. (2009) 'Organizational adoption of e-HRM in Europe: An empirical exploration of major adoption factors', *Journal of Managerial Psychology*, vol. 24, No. 6, pp. 482–501.
[57] Thomas, M. (2011) Deconstructing Digital Natives: Young People, Technology and the New Literacies, New York: Routledge.
[58] Trost, A. (ed.) (2009) Employer Branding, Köln: Luchterhand.
[59] Uggerslev, K.L., Fassina, N.E. and Kraichy, D. (2012) 'Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process', *Personnel Psychology,* vol. 65, No. 3, pp. 597–660.
[60] Universum (2014) A Brave new workplace: A look at how Generations X, Y and Z are reshaping the nature of work, [Online], Available: https://universumglobal.com/blog/brave-new-workplace-gen-x-y-z/ [2 Feb 2022].
[61] Van der Heijden, H. (2004) 'User acceptance of hedonic information systems', *MIS Quarterly*, vol. 28, No. 4, pp. 695–704.
[62] Venkatesh, V., Davis, F.D. and Morris, M.G. (2007) 'Dead or alive? The development, trajectory and future of technology adoption research', *Journal of the Association for Information Systems*, vol. 8, No. 4, pp. 268–286.
[63] Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003) 'User acceptance of information technology: Toward a unified view', *MIS Quarterly*, vol. 27, No. 3, pp. 425–478.
[64] Venkatesh, V., Thong, J.Y. and Xu, X. (2012) 'Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology', *MIS Quarterly,* vol. 36, No. 1, pp. 157–178.
[65] Williams, M.D., Rana, N.P. and Dwivedi, Y.K. (2012) 'A bibliometric analysis of articles citing the Unified Theory of Acceptance and Use of Technology', in Dwivedi, Y.K., Wade, M.R. and Schneberger, S.L. (ed.), Information Systems Theory: Explaining and Predicting Our Digital Society, New York: Springer, pp. 37–62.
[66] Woods, S.A., Ahmed, S., Nikolaou, I., Costa, A.C., and Anderson, N.R. (2020) 'Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges.'*European Journal of Work and Organizational Psychology*, vol. 29, No. 1, pp. 64–77.
[67] Zainuddin, Z., Chu, S.K.W., Shujahat, M. and Perera, C.J. (2020) 'The impact of gamification on learning and instruction: A systematic review of empirical evidence', *Educational Research Review*, vol. 30, 100326.

Summary

Organizations are faced with increased competition in the war for talent, and their sustained competitive advantage may depend on the ability to attract suitable candidates. The gamification of HR recruitment processes can be one solution, as it creates employer brand awareness and enables candidates to better assess their fit with the organizational culture and job requirements. Based on a comprehensive literature review and through focus groups and qualitative content analysis, we develop guidelines for effective application design and implementation. Our findings are mirrored against UTAUT 2 theory and consolidated in a Model for Acceptance of Gamified Recruitment Elements. Results suggest that gamified recruitment is an effective option to support traditional recruitment processes in orienting candidates and companies about the individual employer fit, ultimately increasing the quality of applications and strengthening organizations' talent pools. From the results, we derive guidelines on how to effectively implement design features.

JEL codes: M12, M15, M51

Keywords: *gamification, human resource management, recruitment, gamified recruitment, technology acceptance*

Managerial Economics
2022, Vol. 23, No. 1, pp. 77–98
https://doi.org/10.7494/manage.2022.23.1.77
©2022 Authors. Creative Commons CC-BY 4.0

Tomasz Wójtowicz*

How to define macroeconomic announcement surprises? An example of the impact of US macroeconomic news on stock prices on the Warsaw Stock Exchange

1. Introduction

In an efficient financial market, only unexpected information leads to significant price changes. This is also the case with the publication of important macroeconomic data. Therefore, the unanticipated component of the announced macroeconomic data is crucial in the analysis of the impact of macroeconomic news on financial markets (e.g., stock or bond markets). This unexpected news (surprise) is usually defined as the difference between the observed and expected value of the published indicator:

$$Surp_{0,i} = A_i - E_i \tag{1}$$

where A_i is the value of the i-th announcement of the indicator, while E_i is the market expected value of the indicator. The more the released indicator value differs from market expectations, the higher the value of $Surp_{0,i}$ is. Thus, $Surp_{0,i}$ measures the size of the news surprise.

The vast majority of studies on the impact of macroeconomic data announcements on financial markets are based on the analysis and estimation of appropriate models in which there are dummy variables corresponding to an unexpected part of the announced news. This approach assumes a linear dependence of returns of

* AGH University of Science and Technology in Cracow, Department of Applications of Mathematics in Economics, e-mail: twojtow@agh.edu.pl. ORCID ID: 0000-0002-9742-3493

the analyzed assets on the size of the information surprise, i.e. on the difference $A_i - E_i$ (see, for example: Balduzzi et al., 2001; Andersen et al., 2003, 2007; Będowska-Sójka 2010; Harju, Hussain, 2011; Kočenda, Moravcova, 2018; Kurov et al., 2019).

However, the question arises as to whether $Surp_0$ is a good measure of news surprises, i.e. whether it accurately reflects how surprising the published value of the macroeconomic indicator is for investors. For this reason, the properties of the differences $A_i - E_i$ should be examined. In particular, it should be verified whether this measure is robust to the occurrence of outliers. Additionally, the nature of the relationship between information surprises $Surp_0$ and returns should be carefully studied in order to determine the strength of the linear relationships between them. In addition, it is worth considering other alternative ways of defining and measuring the size of the unexpected part of news that will have the desired properties.

The aim of this paper is to analyze the properties of various surprises measures for the announcements of US macroeconomic news. The relationship between surprises and returns will also be studied based on the impact of announcements of macroeconomic indicators describing the US economy on the 5-minute returns of the WIG20 index (the main index of the Warsaw Stock Exchange). The analysis presented in this paper is based on data on announcements of 15 US macroeconomic indicators and 5-minute returns of WIG20 from January 2001 to February 2021.

Recent studies (for example, Będowska-Sójka, 2010; Suliga, Wójtowicz, 2013; Gurgul, Wójtowicz, 2014, 2020; Gurgul et al., 2016; Gurgul et al., 2021) indicate a significant and very strong impact of the publication of macroeconomic data from the US on the share prices of companies listed on the Warsaw Stock Exchange. Among the various indicators that describe the state of the US economy, the Nonfarm Payrolls (NFP) stands out. It is also one of the most important macroeconomic indicators for the US economy (Andersen, Bollerslev, 1998; Andersen et al., 2007). Therefore, a more detailed analysis presented in the first part of the paper is carried out on the example of NFP announcements. The application of NFP data also ensures that the results are not distorted by the impact of the publication of the other indicators.

In addition to the analysis of the properties of the $Surp_0$ distribution and distribution of the other measures of surprises, the linear relationship between surprises and WIG20 returns in the first five minutes after the announcements is examined. The analysis of correlation coefficients allows one to determine the strength of the linear relationship between surprises and returns. However, the analysis of such individual relationships may not lead to correct conclusions because sometimes more than one US macroeconomic indicator is released at the same time. Therefore, a study considering the possibility of the impact of the publication of individual indicators overlapping is also carried out with the use of an appropriate model in which surprises are explanatory variables.

The rest of the paper is organized as follows. In the next section, we describe the data used in the paper. In Section 3, we present the results of the analysis of macroeconomic news surprises distribution for NFP announcements. This study is extended in Section 4 to other US macroeconomic indicators. In Section 5, we briefly analyse the properties of linear models explaining WIG20 returns by news surprises. The final section concludes the paper.

2. Data

The analysis presented in this paper is based on data from the announcements of 15 macroeconomic indicators from the US economy and on 5-minute returns of the WIG20 index from January 2001 to February 2021. These indicators are: the Consumer Confidence Index (CCI), the Consumer Price Index (CPI), the Durable Goods Orders (DGO), the Existing Home Sales (EHS), the Real GDP (GDP), the Housing Starts (HS), the Initial Jobless Claims (IJC), the Industrial Production (IP), the ISM Manufacturing Index (ISM), Leading Indicators (LI), the New Home Sales (NHS), the Nonfarm Payrolls (NFP), the Philadelphia Fed Business Outlook Survey (PFBO), the Personal Income (PI), and the Retail Sales (RS). Almost all of them are released on a monthly basis and describe the economic situation in the US in the previous (or even in the current) month. The only exception is IJC, which is announced weekly and describes the labor market in the previous week, and GDP, which is released monthly but describes the GDP in the previous quarter.

As shown, for example, by the analysis of Będowska-Sójka (2010), Gurgul et al. (2021), Gurgul et al. (2016), Gurgul and Wójtowicz (2014, 2020), Suliga and Wójtowicz (2013) carried out with the use of various methods, there is a significant and very strong impact of the publication of US macroeconomic data on share prices of companies listed on the Warsaw Stock Exchange. The strongest reaction of investors is observed after the announcements of unexpected values of NFP, which is one of the most important American indicators.

NFP is one of the indicators published in the Employment Report of the Bureau of Labor Statistics. The others are: the Unemployment Rate, Average Hourly Earnings and Average Workweek. Each of them describes different aspect of the employment situation and its changes can lead to different investor reactions.

As Employment Reports are usually published on the first Friday of the month, the information they contain is one of the first macroeconomic data to describe the US economy in a given month. In addition, the reports describe the labor market, which is a very important part of the economy. As a result, information contained in Employment Reports is closely followed by investors around the world and has a very strong impact on bonds, exchange rates, and stock prices (Carnes, Slifer, 1991; Andersen, Bollerslev, 1998; Andersen et al., 2007). The NFP announcements are also one of the most important American data for investors on the Warsaw Stock Exchange (Suliga, Wójtowicz, 2013; Gurgul, Wójtowicz, 2014, 2020; Gurgul et al., 2016).

It should also be emphasized that, since the NFP is one of the first US macroeconomic indicators to be published during the month, its impact is not distorted by the announcements of other indicators, which usually are released a few days later. Therefore, the results of the examination of the properties of the various measures of unexpected part of NFP announcements are not distorted by the impact of other important information from the American economy. Hence, in the first part of the paper, particular attention will be paid to the analysis of the properties of surprise measures on the example of NFP announcements.

Determining the expected value of the announced macroeconomic indicator is also important for studying the relationships between surprise measures and WIG20 returns. The expected value E_i of an indicator is usually defined in two ways. First, it may be estimated from the previous values of the indicator with the use of an appropriate econometric model, for example an ARMA model. The second way of defining the expected value of the macroeconomic fundamentals is based on surveys. According to it, E_i is proxied by the median response (consensus) of managers and professional financial analysts. From these two approaches, the survey-based definition of the expected component of macroeconomic news announcement is more common in the literature (see, for example: Almeida et al., 1998; Balduzzi et al., 2001; Andersen et al., 2003, 2007; Będowska-Sójka, 2010; Harju, Hussain, 2011; Gurgul, Wójtowicz, 2014, 2015; Kurov et al., 2019). As shown by Pearce and Roley (1985), the application of surveys to the forecast announced value of macroeconomic fundamentals outperforms any forecast based on their historical values.

Forecasts of the announced value of the macroeconomic indicator obtained on the basis of surveys are quite easily available, because they are published by most economic data platforms a few days before the announcement date (for example, by Bloomberg, Yahoo, Trading Economics, Investing, koyfin, DeltaStock, Econoday, etc.). Furthermore, as shown by Pearce and Roley (1985), most of such forecasts are unbiased (i.e. in most cases the expected value of the difference $A_i - E_i$ is equal to zero) and have smaller mean squared errors than forecasts based on autoregressive models.

The vast majority of on-line financial data and analytics platforms provide only the announced value of the indicator and the consensus[1]. However, some of them (e.g., Econoday) also report additional information about the survey results. For example, they provide the smallest and highest values specified in the surveys. It allows one to calculate the range of analysts' forecasts. The range can be seen as a measure of analyst uncertainty about the future value of the indicator.

Professional users, for example, in the Bloomberg Terminal, have access to more detailed data on the survey statistics for each indicator. Before each announcement of an important macroeconomic indicator, Bloomberg Terminal not only provides the value of consensus, but also provides standard deviation of the survey results from which the consensus was calculated. This is a more precise measure of uncertainty, and, on this basis (after the announcement), the surprise value is defined. The surprise, which is equal to the difference between the announced and predicted value of the indicator divided by the standard deviation of the surveys, shows how large the surprise value when is compared to the variability of the forecasts.

The news surprises considered in this paper are calculated based on the reported value of the indicator and the survey median (consensus). We also take into account the lowest and highest values of the forecasts, as well as standard deviations of the surveys. All these values come from Bloomberg database.

The basic measure that describes how much the announced value of the indicator differs from the market expectations is surprise $Surp_0$ defined for each i-th announcement in (1) as the difference between the announced value A_i of the indicator and the survey expectation E_i. This difference is a natural measure of a news surprise and is of great importance to investors. This is because most analysts describing and interpreting macroeconomic data releases make two comparisons of the announced value: with its previous values from the last few months or with its market expectation. The problem with interpreting the $Surp_0$ difference is that it does not take into account the uncertainty about the true value of the indicator and the variation of the forecasts of the true indicator value among analysts. Furthermore, when measuring the size of the news surprise related to the current publication of the indicator, one should also take into account how the values of the surprise have changed in the past. Therefore, in this paper, we additionally consider the following measures of the magnitude of news surprises:

$$Surp_{1i} = \frac{A_i - E_i}{S_i} \tag{2}$$

[1] As shown by Wójtowicz (2015), application of data from various news websites leads to very similar conclusions regarding the impact of NFP publications on stock prices on the Warsaw Stock Exchange.

$$Surp_{2i} = \frac{A_i - E_i}{S_{12,i}} \tag{3}$$

$$Surp_{3i} = \frac{A_i - E_i}{H_i - L_i} \tag{4}$$

where S_i is a standard deviation of surveys from which the consensus E_i was calculated, H_i and L_i are the maximum and minimum values of the surveys, and $S_{12,i}$ is the standard deviation of surprises from the given announcement and the 11 previous announcements.

The above surprise measures differ in the way they take into account additional information that allows for a comparison of the difference $A_i - E_i$ with analysts' uncertainty ($Surp_1$ and $Surp_3$) or with the previous surprises ($Surp_2$). $Surp_1$ relates the difference $A_i - E_i$ to standard deviation of market forecasts. In this way, even a very large value of the difference between the announced and expected value of the indicator may be of little importance if the analysts were very heterogeneous in their forecasts. On the other hand, sometimes a small difference $Surp_0$ can become significant if the analysts' forecasts were very consistent and were characterized by a very small standard deviation. The same idea is behind the definition of $Surp_3$. In this case, however, the measure of the heterogeneity of market expectations is the difference between the largest and the smallest forecast value of the indicator. The range has some serious disadvantages because it depends only on two values of the data (maximum and minimum), and thus it is very sensitive to outliers. Standard deviation is also sensitive to outliers, but their impact on it is much weaker than the impact of minimum and maximum on the range. However, the obvious advantage of the range is its simplicity: the max and min values of surveys are generally available in some financial datasets, while it is much more difficult to find the values of the standard deviation of analysts' forecasts.

In the absence of more detailed information on the survey results, the earlier values of the difference $A_i - E_i$ may be used to assess the size of the surprise. It is quite a natural approach to measure the surprise because usually investors compare various current data with historical values looking for repeating patterns or trends. The large value of the difference is more important for investors if it is preceded by much lower values in the previous months. On the other hand, the second or third very high value of $A_i - E_i$ in a row does not lead to an equally strong reaction from investors.

When using the measures $Surp_1$, $Surp_2$ and $Surp_3$, it is important that they are standardized and that their values (which are calculated for different macroeconomic indicators) can be compared. Usually, this cannot be done when comparing the value of the $Surp_0$ measure itself, as different macroeconomic indicators are expressed in different units or have values of very distant levels.

3. NFP announcements

When analyzing reactions of investors on the Warsaw Stock Exchange to unexpected information contained in NFP announcements, it is worth paying attention to the values of this indicator itself. Comparing the NFP values published in the subsequent months of the period under consideration, the extreme values announced between April and August 2020 (that is, from the initial stage of the COVID-19 pandemic) clearly stand out. For this reason, in Figure 1, the published NFP values are presented in two graphs: before April 2020 (left graph) and after that time (right graph). These graphs also show the differences between the published NFP value and the market expectation measured by consensus. To illustrate the differences between the NFP and the surprise values before and after April 2020, both graphs show the values for the announcement on April 3, 2020 (the last value on the left graph and the first value on the right graph). To supplement this information, Table 1 presents more detailed data on selected NFP announcements in 2020.

a) b)

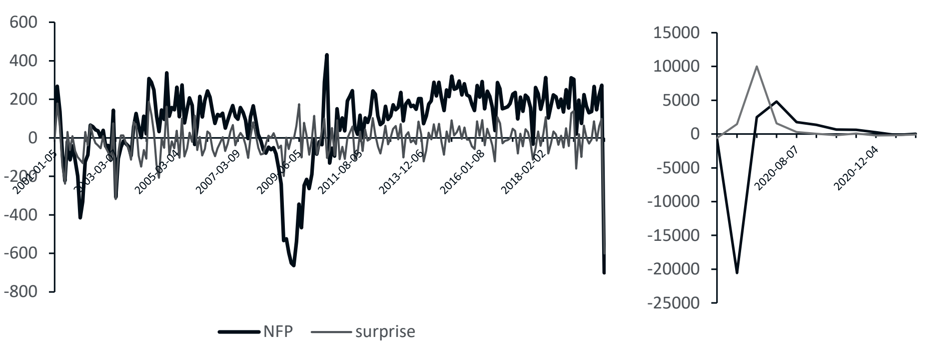

Figure 1. Announced values of NFP and values of surprises $Surp_0$ in the period 2001–2021

Note: This Figure presents announced NFP values (thick line) along with surprises $Surp_0$ values (thin line) form January 2001 to April 2020 (a) and from April 2020 to February 2021 (b).

Table 1

Details of NFP announcements in the initial phase of the COVID-19 pandemic

Date	Actual	Consensus	Survey High	Survey Low	Survey Std. Dev.
2020-03-06	273	175	249	132	21.72
2020-04-03	–701	–100	100	–4000	626.36
2020-05-08	–20 537	–22 000	–8600	–30 000	2928.31
2020-06-05	2509	–7500	–800	–12 000	2384.01
2020-07-02	4800	3230	9000	500	1493.19
2020-08-07	1763	1480	3210	–600	819.97

Note: This table presents announced values of NFP and some basic survey statistics for published between March and August 2020.

From January 2001 to March 2020, the published NFP values ranged from −663 (April 2009) to 431 (June 2010). NFP value announced on April 3, was –701. Then, in May 2020, the NFP fell to −20537. Analysis of Figure 1 shows that fluctuations of the $Surp_0$ values are much smaller than changes in NFP and the values of the difference $A_i - E_i$ are only to some extent related to the size of the published NFP. Before the COVID-19 pandemic $Surp_0$ ranged from −318 (March 2003) to 188 (April 2004). However, extreme values of NFP announced in 2020 associated with very high uncertainty of analysts led to extreme values of $Surp_0$ in the following months. $Surp_0$ in April 2020 (equal to −601) is the lowest value in the entire period. Similarly, the differences in the next four months are the four highest values over the whole period 2001–2021. These extremely high values significantly distort the distribution of $Surp_0$ values and have negative consequences for the analysis of the impact of the unexpected NFP value announcements on the stock prices on the Warsaw Stock Exchange. This negative effect is probably also observed in other markets.

At this point, it is worth commenting on the impact of the employment situation reports on the prices of shares listed on the Warsaw Stock Exchange in the first months of the COVID-19 pandemic, i.e., between April and August 2020. First, the investor reaction to the announcements in April and May was inadequate to the sign of the surprise. On April 3, we observe an increase in the WIG20 value by about 0.375% in the first 5 minutes after the announcement of the value of NFP lower than expected by 601. Similarly, the very high value of the surprise released on May 8 ($Surp_0$ = 1463) was followed by negative change in the WIG20 index ($R_t \approx -0.232\%$). Second, the changes in the WIG20 in the first 5 minutes after the NFP releases in the following months are not as large as the surprise values would suggest. This is evidenced by the comparison of the ranks of surprises and returns reported in parentheses in the respective columns in Table 2. The reason for these discrepancies may be the COVID-19 pandemic itself and the fact that at that time investors probably paid much more attention to information on the development of the pandemic in Poland and in other countries, in particular to information on the introduced restrictions and their possible impact on economies. On the other hand, the returns in the first 5 minutes after the NPF releases from April to August 2020 are significant when compared to their values in the three hours prior to the announcements[2].

Table 2

Values of surprise measures for NFP announcements in the initial phase of the COVID-19 pandemic

Date	$Surp_0$		$Surp_1$		$Surp_2$		$Surp_3$		R_t	
2020-03-06	98	(209)	4.51	(221)	1.70	(220)	0.84	(223)	0.183%	(170)
2020-04-03	−601	(1)	−0.96	(91)	−3.17	(2)	−0.15	(97)	0.375%	(204)
2020-05-08	1463	(225)	0.5	(148)	3.10	(226)	0.07	(143)	−0.232%	(37)
2020-06-05	10009	(227)	4.2	(219)	3.45	(227)	0.89	(224)	0.415%	(209)
2020-07-02	1570	(226)	1.05	(166)	0.54	(171)	0.18	(167)	0.265%	(187)
2020-08-07	283	(224)	0.35	(140)	0.10	(136)	0.07	(145)	0.149%	(164)

Note: This table presents values of surprise measures for NFP announcements published between March and August 2020. In the last column, values of WIG20 returns from the first 5 minutes after the announcements are reported. The numbers in parentheses indicate the rank of a given value of the surprise measure (or returns) in the entire sample.

The differences between the WIG20 returns and the values of $Surp_0$ can be partially explained by the values of the remaining surprise measures. In particular, the $Surp_1$ values indicate that if we take into account the heterogeneity of analysts' forecasts, only the NFP value published on June 5 was clearly different from market expectations. The analysis of the $Surp_3$ value leads to the same conclusion. Values of $Surp_2$ are very high for NFP announcements in April, May, and June 2020. This, in turn, is due to the fact that the absolute values of the differences $A_i - E_i$ increased in the following months. As $Surp_2$ compares the value of the difference from the given month with its historical values up to 11 months, the first large value of the difference $Surp_0$ appearing after a period of low values leads to a very large value of $Surp_2$. Successive, greater and greater values of $Surp_0$ also

[2] To verify the significance of the first 5-minute returns after a news announcement, we checked that they are within a 90% confidence interval constructed from 36 returns in the 3-hour period prior to the announcement.

imply high values of $Surp_2$. It is worth noting, however, that the announcement on July 2, 2020, although very different from expectations, does not caused such a high value of $Surp_2$ because it was preceded by a few even stronger surprises.

Table 3

Descriptive statistics of macroeconomic news surprises

	$Surp_0$	$Surp_0$ before April 2020	$Surp_1$	$Surp_2$	$Surp_3$
Mean	42.3	−11.8	−0.29	−0.13	−0.06
Std. dev.	685.5	74.9	2.50	1.04	0.47
Min	−601	−318	−8.19	−3.27	−1.82
1st quartile	−58	−57.3	−1.84	−0.87	−0.35
Median	−8	−8.5	−0.38	−0.07	−0.06
3rd quartile	33	32.3	1.17	0.54	0.21
Max	10 009	188	8.69	3.45	1.97
Skewness	13.8	−0.28	0.14	0.05	0.10
Kurtosis	203.3	4.12	3.38	3.65	4.74

Note: This table presents descriptive statistics of macroeconomic news surprises under study computed for NFP announcements in the period from January 2001 to February 2021. Due to extreme values of $Surp_0$ in 2020, its distribution before the COVID-19 pandemic period is described in the separate column ('$Surp_0$ before April 2020').

For a more detailed analysis of the values and properties of the surprise measures under study, several basic descriptive statistics of their distributions are presented in each column of Table 3. Due to the described-above extreme values of the differences $Surp_0$ between announced and expected values of NFP in the initial period of the COVID-19 pandemic, the characteristics of the $Surp_0$ distribution before April 2020 are presented in the separate column. The values of order statistics (confirmed also by the value of the skewness coefficient) indicate that distribution of $Surp_1$, $Surp_2$ and $Surp_3$ are symmetric. Moreover, the values of kurtosis close to 3 suggest that the $Surp_1$, $Surp_2$, and $Surp_3$ can be described by a normal distribution. As a confirmation of this conjecture, the Shapiro test does not reject normality of $Surp_1$ and $Surp_2$ at the 5% significance level, and of $Surp_3$ at the level of 1%. The skewness of $Surp_0$ before April 2020 also is very close to zero and kurtosis is close to 3. However, the extreme values of $Surp_0$ after April 2020 disturb the values of both measures suggesting a very strong asymmetry of the $Surp_0$ distribution.

Most of the models used to analyze the impact of announcements on stock prices considers linear relationship between returns and news surprises. Hence, it is worth analyzing in detail the strength of the Pearson correlation between surprise measures and the changes of the WIG20 in the first 5 minutes after the announcements of unexpected NFP values. Results of this analysis are summarized in Table 4. As before, we separately analyze correlations with $Surp_0$ before April 2020. The values in the first row ('Correlation') are calculated on the basis of data from the entire dataset. They show a similar strength of the dependence of the WIG20 returns on the size of macroeconomic surprises for $Surp_1$, $Surp_2$, and $Surp_3$. However, as can be concluded from the comparison of the results in the first two columns, the extreme values of $Surp_0$ from 2020 have a very strong negative impact on this linear dependence. They lower the value of Pearson correlation coefficient from 0.475 (before April 2020) to 0.119 when data from the whole period are analyzed together.

Investors do not always react to the releases of macroeconomic data. This applies in particular to those announcements that are not surprising enough, i.e., when the published value of the indicator differs little from the market expectations. In such a situation, the observed changes in WIG20 right after the announcement may seem random. Therefore, to describe the impact of only very unexpected news, the following rows of Table 4 report the values of the Pearson correlation coefficients calculated for announcements with a respective surprise measure greater in absolute value than the breakpoint indicated. In the second row ('Correlation – 1st Q'), 25% of announcements with the weakest surprises are removed. In the third row ('Correlation – me') we present correlations computed for the half of the strongest surprises. Analogously, the last row in Table 4 contains the correlation between the surprise measures and the WIG20 returns computed for the 25% strongest unexpected news about NFP.

Table 4

Correlations between macroeconomic news surprises and WIG20 returns

	$\boldsymbol{Surp_0}$	$\boldsymbol{Surp_0}$ before April 2020	$\boldsymbol{Surp_1}$	$\boldsymbol{Surp_2}$	$\boldsymbol{Surp_3}$
Correlation	0.119	0.475	0.454	0.449	0.463
Correlation – 1st Q	0.133	0.523	0.499	0.494	0.513
Correlation – me	0.158	0.557	0.548	0.511	0.566
Correlation – 3rd Q	0.186	0.589	0.591	0.567	0.555

Note: This table presents values of Pearson correlation coefficients between surprise values and 5-minute WIG20 returns immediately after news announcements. The second, third and fourth rows report correlations computed only for announcements with an absolute value of surprises greater than the indicated breakpoints (1st quartile, median, and 3rd quartile, respectively).

From the comparison of the values in Table 4, it can be noticed that the restriction of the analysis to stronger surprises increases the correlation coefficients between the surprise measures and the 5-minute returns of the WIG20 right after the announcements. Moreover, correlations in each row are similar to each other except $Surp_0$ for the entire period. This means that the use of relative measures of macroeconomic news surprises (i.e. $Surp_1$, $Surp_2$, and $Surp_3$) gives very similar results about the strength of the linear relationship with WIG20 returns, regardless of whether we relate the difference $A_i - E_i$ to the dispersion of analysts' forecasts or to the variability of the previous surprises. However, the differences $Surp_0$ themselves, are very sensitive to the occurrence of outliers, which adversely affect the measurement of the strength of the linear relationship between $Surp_0$ and the WIG20 returns right after news announcements.

The period under study (2001–2021) is very long and includes both bull and bear market time periods. In particular, it covers two large financial crises: the global financial crisis in 2007–2009 (which originated in the USA) and the European debt crisis in 2010–2014. Therefore, it is important to analyze changes in the strength of the relationship between surprises and WIG20 returns in subsequent years. For this purpose, the correlation coefficients between the surprises and WIG20 returns are calculated in windows with a length of 48 months, shifted by one month. The first such window covers data from January 2001 to December 2004. The next one: from February 2001 to January 2005, etc. Figure 2 shows the Pearson correlation values in such windows for each of the surprise measures considered. The ends of the windows are marked on the X-axis.

As can be seen in Figure 2, for the greater part of the period 2001–2021, the values of the correlation coefficients for the surprise measures are very close. However, there are periods when they differ noticeably from each other. The largest discrepancy (before the pandemic period) is visible for data covering the beginning of the global financial crisis, i.e. from January 2004 to January 2008. For that period, the difference between the correlation coefficients is over 0.1 (correlation for $Surp_0$ is equal 0.5, whereas the correlation for $Surp_3$ equals 0.61). A large spread in the correlation values is also observed in the last few years before the COVID-19 pandemic. However, in that period, the lowest correlations are observed for $Surp_3$.

The most cases of the highest values of correlation coefficients occur for $Surp_0$, and the least for $Surp_1$. However, when we take into account the observed changes in the correlations and small differences between correlations computed for different surprises, it is difficult to unambiguously select the best measure of macroeconomic news surprises. Hence, when selecting the appropriate surprise measure, one can follow its simplicity. In this respect $Surp_0$ is the best. However, its sensitivity to outliers means that the models built on it may lead to inconsistent or erroneous conclusions about returns. For this reason, a more stable measure of surprises should be used for modeling purposes.

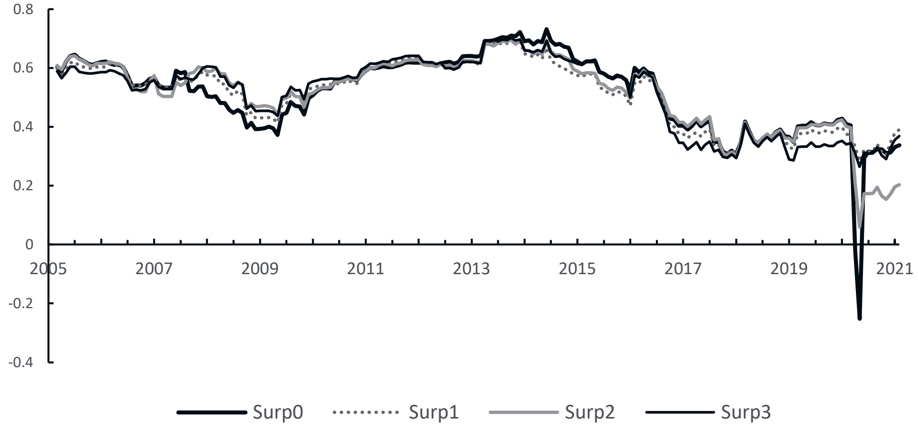

Figure 2. Pearson correlation coefficients between surprise measures and 5-minute WIG20 returns in 4-year windows

Note: The ends of the windows are marked on the X-axis.

4. US macroeconomic news announcements

A comparison of surprise measures on the example of announcements of only one (even the best) macroeconomic indicator may not give a complete picture of their properties. For this reason, the analysis of the distribution of surprise measures will also be carried out on the basis of the announcements of 15 macroeconomic indicators describing the US economy. In order to facilitate the comparison of the distributions of individual measures, in Figures 3–6 we present boxplots for these measures calculated for each of the indicators[3].

In general, distributions of surprise values computed for different indicators have a similar range for each of the surprise measures (except $Surp_0$). However, there are some outliers. The smallest number of outliers and the most symmetric distributions can be observed in the case of $Surp_2$, i.e. when the difference $Surp_0$

[3] As the indicators are expressed in different units, in order to compare $Surp_0$ for different indicators, all values of $Surp_0$ for each indicator are divided by their sample standard deviation.

is compared with its past values. On the other hand, the least stable are the $Surp_0$ distributions. In the case of some indicators (like DGO, IJC, NFP, PI, and RS), distributions of $Surp_0$ show a strong positive asymmetry caused by extremely large positive surprise values. A slight asymmetry can also be observed in the case of some distributions of $Surp_1$ and $Surp_3$, however in these cases the effect of outliers is much weaker than in the case of $Surp_0$. The analysis of the boxplots in Figures 3–6 confirms the conclusions about the distributions of surprises drawn on the basis of the analysis carried out previously on the basis of the NFP announcements only.

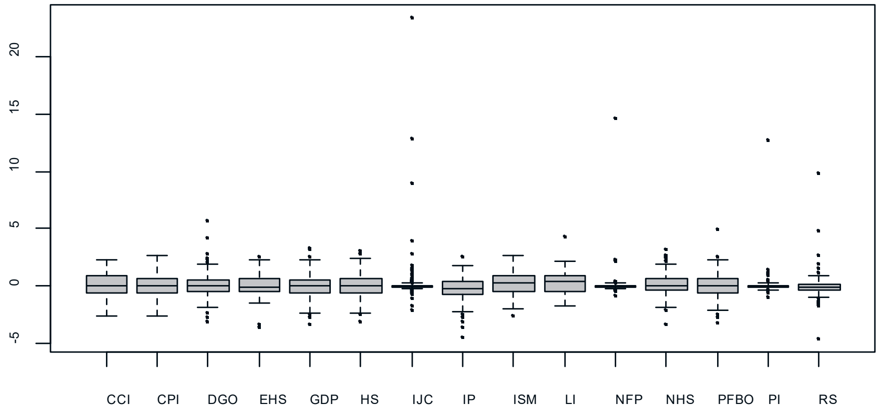

Figure 3. Boxplots of the $Surp_0$ values computed for the announcements of various US macroeconomic indicators

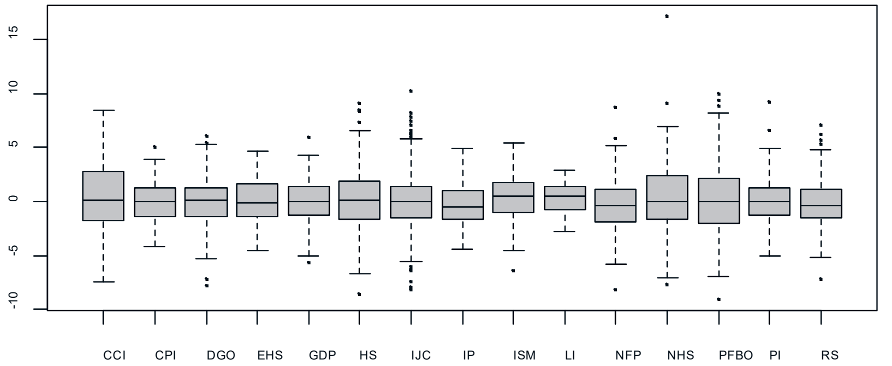

Figure 4. Boxplots of the $Surp_1$ values computed for the announcements of various US macroeconomic indicators

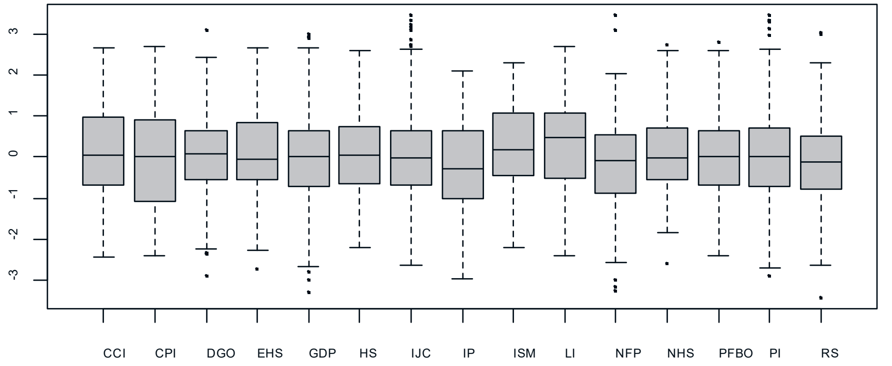

Figure 5. Boxplots of the $Surp_2$ values computed for the announcements of various US macroeconomic indicators

In contrast to the NFP, most of the considered indicators are published around the middle of the month or even later. Moreover, very often they are released at the same time. Therefore, the study of the Pearson correlation between the values of individual surprise measures and the WIG20 returns in the first minutes immediately after news releases may not give a true picture of the impact of unexpected information contained in the announcements of the indicators on stock prices on the WSE. As the impact of the publication of various indicators may overlap, the relationships between returns and surprises will be examined on the basis of appropriate models.

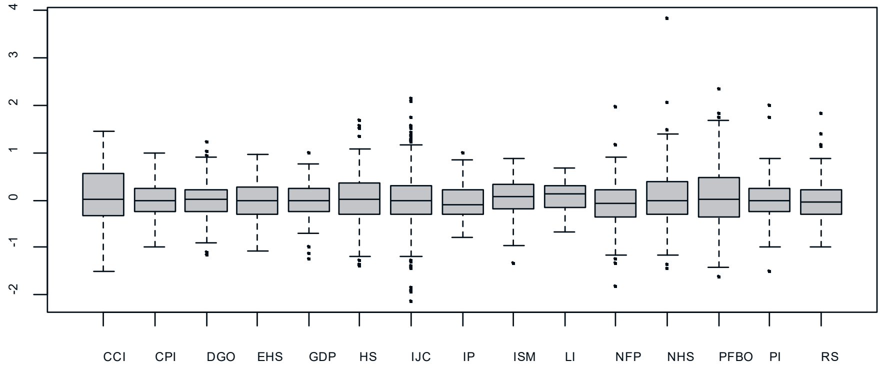

Figure 6. Boxplots of the $Surp_3$ values computed for the announcements of various US macroeconomic indicators

5. Linear models for WIG20 returns

Various models to describe the impact of unexpected macroeconomic news announcements on stock prices, futures, or bonds are considered in the literature. These are mainly various versions of linear models (VAR, ARMA, ARFIMA or AR) with dummy variables added (see, for example: Balduzzi et al., 2001; Andersen et al., 2003; Andersen et al., 2007; Hanousek, Kočenda, 2011; Harju, Hussain, 2011; Będowska-Sójka 2013; Kurov et al., 2019). Due to the possible heteroskedasticity of the residuals and the seasonal patterns observed in intraday volatility (for example, Harju, Hussain, 2011; Gurgul, Wójtowicz, 2020), residual variance in these models is described in an additional equation. Despite their diversity, all the above models assume a linear relationship between news surprises and returns (or their volatility).

To analyze how various definition of news surprises affect the results of such linear models we apply to the WIG20 returns the model presented by Andersen et al. (2007). In this model the conditional mean of the 5-minute returns is a linear function of their I lags and J lags of each of the K news announcements. This model is given by the formula:

$$R_t = \beta_0 + \sum_{i=1}^{I} \beta_i R_{t-i} + \sum_{k=1}^{K} \sum_{j=0}^{J} \beta_{kj} S_{k,t-j} + \varepsilon_t \tag{5}$$

where R_t are 5-minute WIG20 returns, $S_{k,t}$ are the news surprise for k-th indicator at time t ($K = 15$). To take into account only data strongly related to the announcements under consideration, estimates of the model parameters are based only on observations from days when the announcements were released[4]. More precisely, for each day of the announcement, we analyze fifteen 5-minute returns before the announcement and eighteen observations after it. The number of lags I and J are determined based on Schwarz information criterion.

Despite the fact that each of the surprise measures has similar values for different indicators, the range of values differs between the measures. To ensure the comparability of the regression results obtained for different surprise measures, we divide the values of each measure calculated for a given indicator by its sample standard deviation. This is a procedure similar to that applied to compute standardized news in Andersen et al. (2003) (and also in Andersen et al., 2007; Harju, Hussain, 2011; Kurov et al., 2019).

[4] Due to the fact that very often more than one indicator is released at the same time, in the entire period from January 2001 to February 2021 there are 2365 days on which the value of at least one of the analyzed indicators was published. As a result, the model (5)-(6) is estimated on the basis of 78045 5-min WIG20 returns.

Following Andersen et al. (2007), to improve the efficiency of the estimates of model (5), we use the weighted least squares estimation procedure with time-varying volatility approximated with the following model:

$$|\hat{\varepsilon}_t| = \sum_{i=1}^{I'} \delta_i |\hat{\varepsilon}_{t-i}| + \sum_{d=1}^{D} \gamma_d D_d + \sum_{k=1}^{K'} \sum_{j=0}^{J'} \gamma_{kj} D_{k,t-j} + u_t \tag{6}$$

where the first term (with $I' = 9$) takes into account the ARCH effect in the residuals. The second term accounts for the seasonal pattern in intraday volatility, and it contains dummy variables D_d for each of the 5-minute intraday intervals of data included in the model. The last term in the above regression models possible impact of news announcements on intraday volatility in a one-hour period after news release. For announcements of the k-th indicator, it contains dummy variables $D_{k,t-j}$ up to a lag $J' = 12$ (i.e. up to one hour).

Parameters of the model (5)–(6) are estimated separately for each surprise measure under study on the basis of data from the whole period January 2001 – February 2021. The results of these estimations are reported in Table 5 where, for simplicity, we present only the values of the model (5)[5]. The Schwarz information criterion indicates $I = 4$ (significant autocorrelation of 5-minute returns up to 20 minutes) and $J = 1$ (significant impact of unexpected news only in the first 5 minutes after the announcements) in model (5) for each surprise measure. These values are in line with the previous results from the literature indicating the very fast reaction of investors on the WSE to the announcements of US macroeconomic data (see, for example, Gurgul, Wójtowicz, 2014; 2020). As the vast majority of the news surprise variables $S_{k,t-j}$ are insignificant for $j = 0$, in Table 5 we report only the values of parameter estimates for $S_{k,t-1}$ describing the impact of news surprises right after news announcements. Additionally, we present the values of t statistics in the significance test.

A comparison of the results presented in Table 5 leads to the conclusion that in most cases the estimated parameters and significance of the parameters are similar for different surprise measures. The impact of unexpected news about CCI, CPI, DGO, GDP, HS, IP, ISM and RS on WIG20 returns in the first 5 minutes after news releases is significant at the 1% level in the case of each surprise measure. On the other hand, announcements of LI, PFBO, and PI do not lead to significant changes in stock prices on the WSE. The significance tests give mixed results about the impact of releases of unexpected values of IJC, NFP, and NHS. In most of these unclear cases, the difference is due to the results of the linear model estimation for $Surp_0$. The distinct behavior of this measure is most evident

[5] We do not present estimates of the intercept because they are insignificant and very close to zero in each case.

when unexpected IJC or NFP values are analyzed. When $Surp_0$ is applied in the model, the model parameters for IJC and NFP announcements are insignificant, whereas for the other news surprise measures, the values of t statistics strongly reject the null hypothesis about insignificance of IJC and NFP. These ambiguous results are due to very extreme positive values of $Surp_0$ reported in Figure 3 for both IJC and NFP.

Table 5

Response of WIG20 returns to US macroeconomic news surprises

Variable	$Surp_0$		$Surp_1$		$Surp_2$		$Surp_3$	
	Estimate	t statistics	Estimate	t statistics	Estimate	t statistics	Estimate	t statistics
R_{t-1}	-0.009^{**}	(−2.09)	-0.010^{**}	(−2.24)	-0.009^{**}	(−2.17)	-0.010^{**}	(−2.24)
R_{t-2}	-0.020^{***}	(−4.67)	-0.020^{***}	(−4.73)	-0.020^{***}	(−4.68)	-0.020^{***}	(−4.70)
R_{t-3}	-0.008^{**}	(−1.97)	-0.008^{*}	(−1.85)	-0.008^{*}	(−1.91)	-0.008^{*}	(−1.87)
R_{t-4}	0.014^{***}	(3.51)	0.014^{***}	(3.53)	0.014^{***}	(3.40)	0.014^{***}	(3.50)
CCC_1	0.056^{***}	(5.17)	0.050^{***}	(4.75)	0.049^{***}	(4.55)	0.048^{***}	(4.62)
CPI_1	-0.082^{***}	(−3.92)	-0.080^{***}	(−4.05)	-0.068^{***}	(−3.37)	-0.075^{***}	(−3.83)
DGO_1	0.092^{***}	(6.91)	0.120^{***}	(9.38)	0.098^{***}	(7.27)	0.120^{***}	(9.39)
EHS_1	0.024^{*}	(1.89)	0.025^{**}	(2.12)	0.021^{*}	(1.68)	0.024^{*}	(1.95)
GDP_1	0.113^{***}	(5.89)	0.091^{***}	(4.72)	0.106^{***}	(5.74)	0.090^{***}	(4.69)
HS_1	0.040^{***}	(4.16)	0.039^{***}	(4.34)	0.038^{***}	(4.17)	0.037^{***}	(4.08)
IJC_1	−0.006	(−1.01)	-0.041^{***}	(−7.64)	-0.042^{***}	(−7.75)	-0.039^{***}	(−7.4)
IP_1	0.048^{***}	(3.81)	0.040^{***}	(3.47)	0.039^{***}	(3.40)	0.040^{***}	(3.47)
ISM_1	0.079^{***}	(5.15)	0.078^{***}	(5.15)	0.074^{***}	(4.71)	0.077^{***}	(5.06)
LI_1	0.025^{*}	(1.70)	0.023	(1.62)	0.025^{*}	(1.77)	0.022	(1.59)
NFP_1	0.047	(1.61)	0.175^{***}	(6.89)	0.173^{***}	(6.87)	0.184^{***}	(7.30)
NHS_1	0.022^{**}	(2.08)	0.028^{***}	(2.70)	0.031^{***}	(2.89)	0.025^{**}	(2.36)
$PFBO_1$	0.016	(1.20)	0.025^{*}	(1.87)	0.018	(1.36)	0.026^{*}	(1.96)
PI_1	-0.007	(-0.52)	0.006	(0.70)	0.004	(0.50)	0.007	(0.73)
RS_1	0.089^{***}	(5.64)	0.098^{***}	(6.28)	0.108^{***}	(7.04)	0.097^{***}	(6.19)

Notes: This table presents the parameter estimates of model (5). Parameters are estimated by weighted least squares with residual volatility modelled by (6). *, **, *** indicate significance of a mean at 10%, 5%, and 1%, respectively.

It should be noted here that the differences in the results of the significance tests for various surprise measures lead to different conclusions about the impact of macroeconomic news announcements on stock prices. When $Surp_0$ is applied, the model suggests insignificant reaction of the WIG20 returns to IJC and NFP announcements. However, this conclusion is inconsistent with the results from the literature, which indicate a very high importance of information from the US labor market for investors in various stock markets.

A comparison of the parameter estimates leads to an interesting observation: results for $Surp_1$ and $Surp_3$ are very similar. This should come as no surprise as both measures of unexpected information follow a similar structure: the difference between the published and the expected value of the indicator is divided by a measure of variability of analysts' forecasts. For the same reason, the estimation results for $Surp_0$ are mainly similar to the results obtained with in the model with $Surp_2$ as an explanatory variable. Here, it is worth noting that $Surp_2$ does not have disadvantages that can be seen when using only the $Surp_0$ differences.

In addition to the analysis of the similarities and differences between the measures of surprise considered, the results in Table 5 also allow us to compare the strength of the impact of the US macroeconomic indicators announcements under consideration on stock prices. Investors on the WSE are most strongly affected by unexpected news contained in the monthly publications of the Bureau of Labor Statistics. This is definitely evidenced by the highest values of the NFP coefficients for 3 of 4 analyzed surprise measures. The GDP and DGO publications also have a very strong impact on the stock market in Poland, although the assessment of the strength of the impact depends on the measure of information surprises used.

6. Conclusions

In this paper, we analyzed the properties of the various measures of unexpected part of the announcements of macroeconomic news. The study was carried out based on data on announcements of 15 American macroeconomic indicators from January 2001 to February 2021. The most commonly used measure of news surprise is the difference between the announced and expected value of the indicator. However, it allows for the occurrence of extremely positive or extremely negative values, which distort its distribution. This, in turn, causes a noticeable weakening of the linear relationship between surprises and returns in the stock markets. For this reason, the difference between the announced and expected values of the indicator should not be used in linear models that describe returns. Therefore, we analyzed other surprise measures that took into account the heterogeneity of analysts' forecasts or the variability of previous surprises. The distributions of these measures are robust to the announcements of values that are far from market expectations. Additionally, each of these enhanced surprise measures is characterized by a similar strength of the linear relationship with returns. The choice of a specific surprise measure depends on the availability of the data. However, most can be calculated even on the basis of freely available data.

Acknowledgements

The author would like to thank two anonymous referees for their valuable comments on an earlier version of the paper.

References

[1] Almeida, A., Goodhart, C. and Payne, R. (1998) 'The Effects of Macroeconomic News on High Frequency Exchange Rate Behavior', *The Journal of Financial and Quantitative Analysis*, vol. 33(3), pp. 383–408.

[2] Andersen, T.G. and Bollerslev, T. (1998) 'Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies', *Journal of Finance*, vol. 53, pp. 219–265.

[3] Andersen, T.G., Bollerslev, T., Diebold, F. and Vega, C. (2003) 'Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange', *American Economic Review*, vol. 93, pp. 38–62.

[4] Andersen T.G., Bollerslev, T., Diebold, F.X. and Vega, C. (2007) 'Real-time price discovery in global stock, bond and foreign exchange markets', *Journal of International Economics*, vol. 73, pp. 251–277.

[5] Balduzzi, P., Elton, E.J. and Green, T.C. (2001) 'Economic News and Bond Prices: Evidence from the U.S. Treasury Market', *Journal of Financial and Quantitative Analysis*, vol. 36, pp. 523–543.

[6] Będowska-Sójka, B. (2010) 'Intraday CAC40, DAX and WIG20 returns when the American macro news is announced', *Bank i Kredyt*, vol. 41(2), pp. 7–20.

[7] Carnes, W.S. and Slifer, S.D. (1991) The Atlas of Economic Indicators: A Visual Guide to Market Forces and The Federal Reserve, Washington, DC: Harper Collins Publishers.

[8] Gurgul, H., Hastenteufel, J. and Wójtowicz, T. (2021) 'Changes in the impact of US macroeconomic news on financial markets the example of the Warsaw Stock Exchange', *Statistics in Transition: New Series*, vol. 22(4), pp. 41–58.

[9] Gurgul, H., Lach, Ł. and Wójtowicz, T. (2016) 'Impact of US macroeconomic news announcements on intraday causalities on selected European stock markets', *Finance a úvěr – Czech Journal of Economics and Finance*, vol. 66(5), pp. 405–425.

[10] Gurgul, H. and Wójtowicz, T. (2015) 'The response of intraday ATX returns to U.S. macroeconomic news', *Finance a úvěr – Czech Journal of Economics and Finance*, vol. 65(3), pp. 230–253.
[11] Gurgul, H. and Wójtowicz, T. (2014) 'The impact of US macroeconomic news on the Polish stock market: the importance of company size to information flow', *Central European Journal of Operations Research*, vol. 22(4), pp. 795–817.
[12] Gurgul, H. and Wójtowicz, T. (2020) Wpływ informacji makroekonomicznych na transakcje na rynkach akcji, Warszawa: Wydawnictwo C.H. Beck.
[13] Hanousek, J. and Kočenda, E. (2011) 'Foreign News and Spillovers in Emerging European Stock Markets', *Review of International Economics*, vol. 19(1), pp. 170–188.
[14] Harju, K. and Hussain, S.M. (2011) 'Intraday seasonalities and macroeconomic news announcements', *European Financial Management*, vol. 17, pp. 367–390.
[15] Kočenda, E. and Moravcová, M. (2018) 'Intraday effect of news on emerging European forex markets: An event study analysis', *Economic Systems*, vol. 42(4), pp. 597–615.
[16] Kurov, A., Sancetta, A., Strasser, G. and Halova Wolfe, M. (2019) 'Price Drift Before U.S. Macroeconomic News: Private Information about Public Announcements?', *Journal of Financial and Quantitative Analysis*, vol. 54(1), pp. 449–479.
[17] Pearce, D.K. and Roley, V.V. (1985) 'Stock Prices and Economic News', *The Journal of Business*, vol. 58, pp. 49–67.
[18] Suliga, M. and Wójtowicz, T. (2013) 'The reaction of the WSE to U.S. employment news announcements', *Managerial Economics*, vol. 14, pp. 165–176.
[19] Wójtowicz, T. (2015) 'Macroeconomic indicators forecasts accuracy and reaction of investors on the WSE', *Quantitative Methods in Economics*, vol. 16(2), pp. 142–151.

Summary

The definition of a news surprise plays a crucial role in the analysis of the impact of unexpected macroeconomic news announcements. In this paper, we study the properties of the most commonly used measure of news surprise, defined as the difference between the announced and expected value of the indicator. Due to the high vulnerability of this measure to outliers, we consider alternative definitions of macroeconomic surprises. Based on the analysis of announcements of 15 American macroeconomic indicators, we show that taking into account the heterogeneity of analysts' forecasts or the variability of the previous surprises, noticeably improves the properties of the distribution of surprise measures. An additional study performed with the use of a dynamic model proves a strong linear relationship between surprise measures and WIG20 returns in the first five minutes after news announcements.

JEL classification: G14, E44

Keywords: *unexpected news, macroeconomic announcements, intraday data, Warsaw Stock Exchange*

Instruction for authors

Before submitting the paper we encourage the authors to use English language editing support.

Papers which are to be published in *Managerial Economics* should be prepared according to the following guidelines.

All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.

Title page should include a footnote, giving the author(s) affiliation(s) (including postal and e-mail addresses of all authors).

Figures must be prepared in a form suitable for direct reproduction. Digital artwork at least 300 dpi resolution is accepted. Photographs, on glossy paper (9 by 13 cm or larger), should display sharp contrast. Figures, tables and photographs should be numbered according to their reference in text.

Illustrations should be edited in CorelDraw (*.CDR), DrawPerfect (*.WPG) or in any other vector graphics form e.g. HPGL, Encapsulated PostScript (*.EPS), Computer Graphics Metafile *.CGM) or bitmaps (*.TIF, *.PCX).

Mathematical equations within the text should be written in separate lines, numbered consecutively (numbers within round brackets) on the right-hand side. Greek characters must be written out clearly.

Summary and 3–5 keywords should be submitted in separate file containing the name of the author, title of the paper with the heading “Summary”.

Authors using Word are requested to employ, as far as possible, text form of mathematical symbols leaving graphic form for the equations longer than single line.

Reference style

In general, the authors should use the Harvard style of referencing. References to literature within the text should be given in the form: the name of the author(s) and the year of publication (in parentheses), e.g. "Smith (1990) underlines…", "As shown in Smith (1990)…". In case of more than two authors of the cited publication the "et al." shortcut should be used.

Lists of references should be written in alphabetical-chronological order, numbered and follow the rules:

- JOURNAL ARTICLE

Muller, V. (1994) 'Trapped in the body: Transsexualism, the law, sexual identity', *The Australian Feminist Law Journal*, vol. 3, August, pp. 103–107.

- BOOKS

Book with one author

Adair, J. (1988) Effective time management: How to save time and spend it wisely, London: Pan Books.

Book with two authors

McCarthy, P. and Hatcher, C. (1996) Speaking persuasively: Making the most of your presentations, Sydney: Allen and Unwin.

Book with three or more authors

Fisher, R., Ury, W. and Patton, B. (1991) Getting to yes: Negotiating an agreement without giving in, 2nd edition, London: Century Business.

Book – second or later edition

Barnes, R. (1995) Successful study for degrees, 2nd edition, London: Routledge.

Book by same author in the same year

Napier, A. (1993a) Fatal storm, Sydney: Allen and Unwin.
Napier, A. (1993b) Survival at sea, Sydney: Allen and Unwin.

Book with an editor

Danaher, P. (ed.) (1998) Beyond the ferris wheel, Rockhampton: CQU Press.

A chapter in a book

Byrne, J. (1995) 'Disabilities in tertiary education', in Rowan, L. and McNamee, J. (ed.) Voices of a Margin, Rockhampton: CQU Press.

- WORLD WIDE WEB PAGE

Young, C. (2001) English Heritage position statement on the Valletta Convention, [Online], Available: http://www.archaeol.freeuk.com/EHPostionStatement.htm [24 Aug 2001].

- CONFERENCE PAPERS

Hart, G., Albrecht, M., Bull, R. and Marshall, L. (1992) 'Peer consultation: A professional development opportunity for nurses employed in rural settings', Infront Outback – Conference Proceedings, Australian Rural Health Conference, Toowoomba, pp. 143–148.

- NEWSPAPER ARTICLES

Cumming, F. (1999) 'Tax-free savings push', Sunday Mail, 4 April, p. 1.

All the items cited in the main text, and no other items, must be placed in the list of references.

Authors should include 2–3 JEL codes with manuscript during submission. For more details on the JEL classification system CLICK HERE.

Information about the journal and the deadlines for submitting articles for next issues are presented at

http://www.managerial.zarz.agh.edu.pl

Double blind peer review procedure

1. In order to assess a quality of submitted publication the Editorial Board consults at least two outside referees (not affiliated to any of the authors' institutions) which are recognized experts in the specific field.

2. At least one of the referees must represent a foreign institution (i.e. an institution located in other country than the home institution of each author).

3. The journal uses double blind peer review policy, i.e. neither the author nor the referee knows the identity of the other. In addition, each referee signs a declaration of no conflict of interests, whereby the conflict of interest is defined as a direct personal relationship (kinship to the second degree, legal relationships, marriage) or a professional scientific cooperation between the referee and the author which took place at least once in the past two years preceding the year of accepting the invitation to review.

4. Only written referee reports are considered (journal does not accept face-to-face or phone-call- based reports). Each report should clearly express the referee's final recommendation (i.e. whether the article should be accepted for publication, revised or rejected). The referees are kindly requested to fill the review form which can be found in "For reviewers" section. In general, the referees are asked to:

– assess:

- the scientific importance of the submission's topic,
- the quality of research;

– verity whether:

  - the Abstract is concise and informative,
  - the facts and interpretations are satisfactorily separated in the text,
  - the interpretations and conclusions follow from the data,
  - the length and structure of the paper is appropriate,
  - the paper can be shortened without loss of quality,

- all the tables and figures are necessary,
- the diagrams and photographs are of good quality,
- there are all essential figures that should be prepared,
- all the references are exact,
- the manuscript requires proof reading by native speaker,
- there is sufficient attention given to previous research.

5. The names of the referees of particular articles are classified.

6. Once a year the journal publishes the complete list of referees.