UNIVERSITATEA BABEŞ-BOLYAI
BABEŞ-BOLYAI TUDOMÁNYEGYETEM
BABEŞ-BOLYAI UNIVERSITÄT
BABEŞ-BOLYAI UNIVERSITY
TRADITIO ET EXCELLENTIA

STUDIA UNIVERSITATIS
BABEŞ-BOLYAI

OECONOMICA

1/2025

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
OECONOMICA

1/2025

April

EDITORIAL OFFICE OF OECONOMICA: Teodor Mihali str. no. 58-60, s. 251, 418655 Cluj-Napoca,
Phone: 0040-264-41.86.52, oeconomica@econ.ubbcluj.ro, https://studia.reviste.ubbcluj.ro/index.php/subboeconomica

YEAR **(Volume 70) 2025**
MONTH **April**
ISSUE **1**

Studia Universitatis Babeş-Bolyai Oeconomica

1

EDITORIAL OFFICE of **Studia Universitatis Babeş-Bolyai Oeconomica**
Teodor Mihali str. no. 58-60, s. 251, 400591 Cluj-Napoca,
Phone: 0040-264-41.86.52, studiaoeconomica@econ.ubbcluj.ro,
https://studia.reviste.ubbcluj.ro/index.php/subboeconomica/

SUMAR – SOMMAIRE – CONTENTS – INHALT

STUDIA UNIVERSITATIS BABEȘ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 1-32
DOI: 10.2478/subboec-2025-0001

IS ACCOUNTING EDUCATION KEEPING PACE WITH CONTEMPORARY CHALLENGES? A COMPARATIVE STUDY OF PUBLICATIONS

Camelia-Ancuța MÜLLER*
Babeș-Bolyai University, Romania

Abstract: This study provides a comparative analysis of the accounting education literature using articles published in dedicated journals and Web of Science-indexed journals over a 28-year period (1997-2024). The paper highlights the main trends, emerging themes and geographical contributions in the context of accounting education, with a focus on adapting to the new global challenges of the post-pandemic context and the development of artificial intelligence. Through a combined quantitative and qualitative approach, the study emphasizes the importance of integrating educational technology, the development of digital competences and ethical values in the training of future professionals. The article contributes to a better understanding of developments and challenges in accounting education and provides directions for future research.

JEL classification: I23, M41

Keywords: accounting education, educational technology, literature review, comparative analysis, higher education

1. Introduction

Education is an essential element for the development of society and individuals, providing a framework through which cognitive and vocational skills can be formed and utilized. Hanushek and Woessmann (2008) emphasize that through the development of cognitive skills, education contributes directly to overall economic growth. It is also emphasized that education significantly influences the way individuals work and perform, providing opportunities to acquire knowledge and develop professional skills (Schultz, 1961). Rieckmann (2012) states that education aims to shape a global society by creating new knowledge and developing essential skills.

Accounting is emerging as a universal language, integrated across all economic and social sectors. Ravenscroft and Rebele (2008) point out that accounting plays a crucial role in the interpretation and management of economic information, and adapting to changes in the economic environment requires continuous development

* Corresponding author. Address: Faculty of Economics and Business Administration, Babeș-Bolyai University, Teodor Mihali 58-60 street, RO-400591, Cluj-Napoca, Romania
E-mail: anca.muller@econ.ubbcluj.ro

of expertise for both practitioners and educators. In addition, Helliar (2013) argues that the global dimension of accounting and education is driving the need for accounting education to be as uniform and internationally comparable as possible. In the same vein, Cheng et al. (2023) emphasize that numerous researchers and practitioners have extensively studied ways in which the quality of accounting education can be improved so that students can meet the increasingly rigorous demands of the profession.

From this point of view, accounting education represents the meeting point between complex professional requirements and the role of universities in preparing future professionals. Given Anderson's (2002) (in Paisey and Paisey (2004)) finding of a steady increase in the number of articles exploring publication patterns in international journals, in conjunction with Marriott et al.'s (2014) observation that accounting education is one of the most frequently reviewed areas within the accounting discipline, it becomes necessary to synthesize reviews of the literature in this area. Such an approach is even more warranted as accounting education journals globally constitute a valuable resource (Urbancic, 2009).

The contemporary context, marked by the effects of the COVID-19 pandemic and the rise of the Fourth Industrial Revolution (4IR), is driving new approaches in accounting education (Cao et al. 2024). Tettamanzi et al. (2023) note, based on their analysis, that the period marked by the health crisis has highlighted and continues to highlight the need to rethink accounting education, including the technical and practical skills to be cultivated among students. In view of these changes, a review study in the field of accounting education becomes necessary to identify the trends, the state of the art, and the knowledge needed by future professionals so that they can respond to the growing demands and be prepared for the technological and economic innovations that define the present and the future of the profession.

This article makes significant contributions to the field of accounting education through an extensive comparative literature review of journals devoted exclusively to this field and articles published in Web of Science (WoS)-indexed journals. The study covers a broad period between 1997 and 2024, providing a detailed perspective on long-term developments and trends. A distinctive aspect of the research is the integration of the geographical dimension, highlighting the involvement of authors from different regions of the world, with a particular focus on contributions from Central and Eastern Europe, including Romania. In addition, the article explores the impact of contemporary factors, such as the COVID-19 pandemic and the rise of the Fourth Industrial Revolution, on emerging themes such as the use of educational technologies and digital skills training. By combining a quantitative analysis, focusing on trends and thematic categories, with a qualitative approach to the literature, the paper highlights major changes in the field and provides directions for future research, emphasizing the need for accounting education to adapt to the changing demands of the profession. This contribution supports a better understanding of the field and identifies existing gaps, proposing a more comprehensive and relevant approach to accounting education.

2. Research Design and Methodology

In the scientific approach of the present research on the literature, aimed at obtaining a concrete picture of the evolution, current status and new requirements of accounting education, we propose to analyze on the one hand the series of review studies carried out consistently from 1986 to the present articles based on six specialized

publications, and, on the other hand, to identify and analyze those articles addressing accounting education issues that have been published in WoS-indexed journals (Science Citation Index Expanded-SCIE and Social Sciences Citation Index-SSCI). Studies conducted on scholarly articles published in journals devoted exclusively to accounting education will be the mainstays of this research.

Our study is organized in three main sections. In the first section, we briefly reviewed the literature review articles in the field of accounting education, identifying and analyzing the 19 review articles conducted by teams of American researchers (Rebele et al., Watson et al., Apostolou et al., Churyk et al.) over a 38-year period. However, we also focused our attention on the contributions of other researchers, the most significant works being those of Paisey and Paisey (2004), Urbancic (2009), Jackling et al. (2013), Marriott (2014), Ameen and Guffey (2017) and Cao et al. (2024). Finally, given their significance in specialized research, we mention the work of Rebele and Pierre (2015) as well as that of Apostolou et al. (2017b) which provides a longitudinal quantitative analysis of research in the field of accounting education.

The second section includes a quantitative analysis of the accounting education literature, focusing on articles from two main sources: journals specialized in accounting education (indexed in international databases) and WoS-indexed journals (SCIE and SSCI), which are not specialized in accounting education. In this subsection, we analyze the evolution of the total number of scientific articles in the field, highlighting both the main research topics and the predominant type of research (empirical or descriptive), and, finally, identifying information on the geographical origin of the authors.

In the third section of this paper, we conducted a qualitative analysis of the literature in the field of accounting education, focusing on the seven main research themes that underlie the analysis of review articles on accounting education.

The research questions that emerge from the analysis are as follows:

1. How has the number of articles published on accounting education in specialized journals and WoS-indexed journals evolved between 1997 and 2024, and what factors may explain any fluctuations and trends identified?

2. What are the main thematic directions in accounting education research published in refereed journals and WoS-indexed journals between 1997-2024 and how can these be interpreted in the context of the development of the field?

3. How has the type of research (empirical or descriptive) evolved in the accounting education literature published in peer-reviewed journals compared to that published in WoS-indexed journals?

4. What are the geographical regions of origin of the authors with the highest publication interest in accounting education, and what explains this distribution of publications?

5. What are the most significant differences and similarities between specialized accounting education journals and non-specialized but WoS-indexed journals?

This paper adopts a systematic methodological approach, aiming to analyze the existing literature in accounting education for the period 1986–2024. The first step consisted of a qualitative analysis of literature review journals published between 1986 and 2024, in order to identify the evolution of major themes in accounting education. The methodology is inspired by recent studies (Apostolou et al. (2023), Churyk et al. (2024)) and structured around three main steps: data collection, critical analysis, and in-depth discussion.

Data were collected from two main sources. The first source consists of internationally recognized accounting education journals, namely Accounting Education (AE), Issues in Accounting Education (IAE), Journal of Accounting Education (JAcEd), Advances in Accounting Education (AAE), Accounting Educator's Journal (AEJ), and Global Perspectives on Accounting Education (GPAE). For the period 1997–2023, data were extracted from review articles published annually by teams of American researchers (Rebele et al., Watson et al., Apostolou et al., Churyk et al.). These articles were summarized and presented graphically using Microsoft Excel. For 2024, in the absence of a review article, data were collected manually from the same journals using resources such as WoS, ScienceDirect Freedom Collection, and the publishers' official websites. However, the absence of recent issues was noted for some journals, such as Global Perspectives on AE since 2018 and AEJ for 2024.

The second source includes articles published in WoS-indexed journals (SCIE and SSCI), identified through a specific search for the term "accounting education" for the period 1997–2024. These articles were manually collected, synthesized, and analyzed in terms of their significant contributions to the field of accounting education. The results of this process were centralized in tabular and graphical form, using Microsoft Excel to facilitate the visualization of trends.

The analysis included two dimensions: quantitative and qualitative. The quantitative dimension concerned the evolution of the number of articles published, their distribution by research topic, and their classification according to the type of research (empirical or descriptive). The research topics were grouped into the following seven categories: curricular issues, core competences, teaching and assessment, teaching by content areas, educational technology, students, and teachers, according to a well-established scheme in the literature (Watson et al. 2003). We also considered the geographical distribution of the authors, investigating their origin and regional distribution.

The qualitative dimension involved selecting and analyzing articles with higher visibility (in terms of citations) from WoS-indexed databases for each of the seven identified themes. We considered only WoS-indexed articles, following the practice of regularly published review articles (Apostolou et al., 2023; Churyk et al., 2024) that focus on scientific papers in peer-reviewed journals. These articles were evaluated to capture their key contributions and to identify emerging research directions.

In-depth discussions revealed emerging trends in accounting education, such as the integration of educational technology and the development of digital and soft skills. The pandemic context and adapting to the requirements of the Fourth Industrial Revolution were also addressed, emphasizing the need to reconfigure the curriculum to adequately prepare students for the future.

Limitations of the methodology include the manual collection of data for the entire period under analysis (1997–2024) for WoS-indexed articles and for the year 2024 for scholarly articles, which may introduce errors. Also, the categorization of articles into one of the seven themes can sometimes be subjective, as an article may address several themes simultaneously. In addition, the analysis depends on the data summarized in the review articles for the 1997–2023 interval and focuses exclusively on WoS-indexed journals and accounting education journals, which may exclude other relevant sources. Nonetheless, the proposed approach provides a solid framework for exploring developments and trends in accounting education, helping to form a clear picture of the field.

3. Series of reviews and longitudinal analyses of accounting education research

The accounting education literature has developed in a continuum of reviews begun by Rebele and Tiller in 1986 (in Rebele et al., 1991) and continued to the present, mainly in JAcEd, IAE, AAE, and several other refereed publications. The first reviews (Rebele et al., 1991, 1998a, 1998b) focused mainly on topics related to teachers and, to a lesser extent, on processes and effects on students. As the number of articles increased, new journals were introduced (AAE, Accounting Education in the UK, GPAE), and topics began to be structured in appendices and tables, progressively highlighting teaching methods, learning resources, student performance, and curricular innovations.

An important development has been the attempt to balance empirical and descriptive research, but reviews have shown that although the share of empirical studies has increased (especially after 2006–2009), descriptive work has continued to be very common. Over time, researchers have pointed to methodological limitations and narrow scope (e.g., focusing on a single course, class, or institution), emphasizing the need to collect more extensive data to increase the generalizability of results.

In addition, corporate scandals (Enron, WorldCom) have generated interest in topics such as ethics, fraud, and forensic accounting, reflecting the field's adaptation to the challenges of the business environment. Other recurrent, relevant topics have been academic integrity, assessment and quality assurance of learning, and the role of students and teachers in improving the educational process.

With the development of digital technologies, a section dedicated to them was initially observed, and then it became an integral part of the general themes, especially after the pandemic period. In the 2020–2022 reviews (Apostolou et al., 2021, 2022, 2023), the total number of articles fluctuated, and educational technology topics progressively merged into other categories, indicating that digital solutions have become ubiquitously integrated into accounting education. The COVID-19 pandemic affected the flow and focus of publications, emphasizing the importance of rigorous experimental (pre/post-test, control groups) research aimed at highlighting pedagogical innovations and effective practices for online, hybrid, and face-to-face learning.

The most recent reviews (Apostolou et al., 2022, 2023; Churyk et al., 2024) attest to a resurgence of interest in empirical studies and content-area-specific instruction after a period of decline. At the same time, new themes such as generative artificial intelligence (e.g., ChatGPT) and data analytics, along with a growing concern for diversity and equity in the curriculum, have also emerged. The need to broaden research methodologies, explore the international context, and connect accounting education to the dynamic demands of the labor market was underscored. Thus, the literature paints a complex picture in which technological innovation, curricular transformations, and the need for methodological rigor define future research directions.

Another impactful contribution to this series of reviews is attributed to Rebele and Pierre (2015), who examined review articles on accounting education published in JAcEd over a 25-year period (1991 to 2015). At the foundation of their study lies the question of whether accounting education research has stagnated, mirroring concerns about accounting research in general in leading accounting journals. In their analysis, the authors noted that the topics covered were relatively limited, typically relying on the same research methods and theories. In addition, most articles are not empirical, do not address topics relevant to accounting practice, and do not employ experimental or quasi-experimental designs. Thus, the authors provided evidence that research in accounting education shows signs of stagnation.

For a better awareness of the importance of accounting education research, Rebele and Pierre (2015) draw a parallel with the medical literature, where there is a preponderance of empirical rather than purely descriptive work or case studies, and where the urgency to find cures for severe diseases tolerates no delay or stagnation. They argue that a similar sense of urgency should apply in accounting education, given the responsibility to provide students with the best possible learning experience.

Exceptionally, in the same year, Apostolou et al. (2017b) published a paper providing an overview of all the analyses carried out over 20 years. This paper incorporates their entire body of review work in the field of accounting education, which aimed to analyze all publications in six journals and identify trends as well as gaps and areas for improvement. Over the 20-year period (1997–2016), 2023 articles were published and analyzed in 299 issues of the six accounting education journals, generally showing an upward trend in the number of articles published. The review study was based on categorizing the published articles into empirical, descriptive, teaching/assessment resources (by content areas), and case studies, the relative weight of which varied over the years. Referring to the five topics used to group the evaluated articles, "students" was the most researched topic during the last 20 years, accounting for 31% of all empirical articles, with an average of 11 articles per year. It was followed by "curricular issues," representing 28% of the empirical articles published in the same period, with an average of 10 articles per year across the six journals.

As geographic location became a topic of increasing interest, between 2010 and 2016 most empirical articles originated in the USA and Canada (65%), followed by Australia and New Zealand (14%), Europe (10%), Asia and Africa (6%), and other geographic regions (4%). The authors stress that geographic information, along with its cultural implications, is important in interpreting research findings, whether to assess the generalizability of those findings or to use as control variables. The most common content areas are financial accounting (n = 29, 30%) and ethical issues (n = 25, 26%).

Apostolou et al. (2017b) speak of a shift from descriptive to empirical work, but they note that a clear theoretical underpinning remains lacking. The authors also emphasize that while replication is crucial for developing a knowledge base in accounting education, studies that merely administer the same survey to a different sample—without examining why the sample and results might differ—are of limited use for advancing research. They conclude that future research should prioritize high-quality projects, theory development or the application of theories from other disciplines, and their validation across multiple populations.

The review series is complemented by a review of the literature published in AE over a decade (1992–2001) by Paisey and Paisey (2004), researchers outside the American group. A study of the progress of AE was also conducted by Wilson (2002), who analyzed several facets over the same period, including the 18 most frequent authors contributing articles to the journal. The study by Paisey and Paisey (2004) does not overlap with Wilson's (2002) study but aims to fill in the gaps of the literature reviews published in AE during its first 10 years.

Taking into account that the series of articles by Rebele et al. (1998a, 1998b), Apostolou et al. (2001), and Watson et al. (2003) covers the period 1991–2002, while Paisey and Paisey (2004) focuses on 1992–2001, an overlap (1992–2001) can be observed. This also allows a comparison with research published in the USA. Thus, to maximize the usefulness of the research and facilitate comparison, the

literature review broadly follows the headings used by the U.S. studies. It also supports researchers in search of topics and/or methodologies that have been less explored so far.

According to Paisey and Paisey (2004), the most frequently published topics were, in order of prevalence: curricular issues, course content and curriculum structures; course delivery and teaching methods; student characteristics; and teacher issues. Areas in which much less material was published include factors influencing students' performance, assessment, and student recruitment and job selection.

In terms of methods found in the literature, the authors observed that the four most common were description and reflection on teaching, literature review with critical analysis, statistical data analysis, and questionnaires. These four methods were employed in 82% of the papers. It was also noted that 61% of the papers mainly used qualitative methods, while 39% relied on quantitative methods.

As suggestions for future research, Paisey and Paisey (2004) welcome the growing interest in education research but underscore the need for a solid educational knowledge base. In addition, the authors highlight an area not considered by U.S. researchers yet notable for articles appearing in AE, namely professional issues. Several articles refer to the needs or educational systems of professional accounting bodies and access to these bodies, including competency standards.

Building on the long stream of reviews appearing in JAcEd, Urbancic (2009) made a significant contribution to accounting education research by emphasizing and refining previous analyses of authors' sources of inspiration, institutional affiliations, and geographical origins. The author examined 868 articles published in the six accounting education journals—also central to our research—over ten years (1998–2007). This period fully coincides with the reviews by Watson et al. (2003, 2007) and partially overlaps with Apostolou et al. (2001) (years 1998 and 1999) and Apostolou et al. (2010) (years 2006 and 2007).

The results of this study indicate that accounting education offers a vehicle for academic dialogue across a very broad segment of the global accounting research community. Evidence of the almost universal importance placed on teaching and learning is reflected in the wide mix of countries that have contributed publications to the field of accounting education.

Jackling et al. (2013) reviewed articles published in AE: An International Journal (AE) over the first 20 years of its existence (1992–2011). The authors focused on a thematic analysis of publications in the second decade, continuing the study by Paisey and Paisey (2004) on the journal's first 10 years. Of the 222 articles reviewed, 196 addressed a single topic, while the remaining 26 dealt with multiple topics. The subjects analyzed, based on their prevalence, were: (1) accounting syllabus (32.7%), (2) course delivery and teaching methods (23.6%), (3) student characteristics (21.8%), (4) faculty issues (9.7%), (5) accounting student recruitment and job selection (6.7%), and (6) professional issues (5.5%). Jackling et al. (2013) suggest, among other points, that future research in accounting education could draw on methods from education and psychology, offering opportunities to test theories and expand research areas.

A far-reaching contribution to the reviews of the research found in the six accounting education journals is made by Marriott et al. (2014), authors from outside the renowned group of American researchers. The research by Marriott et al. (2014) spanned a 5-year period (2005-2009), the period under review thus overlapping with

the review paper conducted between 2006-2009 by Apostolou et al. (2010), and partially (2005 only) with the one conducted by Watson et al. (2007) for the period 2003-2005.

In an attempt to compare and contrast the North American research tradition with that of Europe and the rest of the world, the authors included other (non-specialized) journals that featured accounting education in 70 publications, in addition to the six specialized journals. The choice of alternative journals for publication was based on the list of elite accounting journals compiled by Reinstein and Calderon (2006).

Marriott et al. (2014) assess that the series of reviews carried out by the group of American researchers are generally very similar in design and purpose. While they competently summarize the content of published work, they provide little quantitative categorization and analysis. Moreover, in earlier studies in particular, they tend to ignore or downplay the importance of publications outside North America. The authors conclude that the review studies appearing in JAcEd indicate a gradual and delayed shift toward empirical work and an increasing focus on curricular issues.

Marriott et al. (2014) identified the differences among these six accounting education journals, showing the preferences and trends in terms of published articles for each one. As a result, scholars can use these findings to better decide which of these journals they should target for their work.

Focusing on the three journals that publish the most primary papers, AE (87), IAE (74), and JAcEd (31), Marriott et al. (2014) showed that survey-based research is more likely to be published in a UK journal than in any of the US outlets. The preference for experiments over surveys may be attributed to the greater pressure in the US to make education research look scientific.

Regarding non-specialized journals, Marriott et al. (2014) believe that they should have included many more articles on accounting education; 88 of the 99 journals ranked by Reinstein and Calderon (2006) did not publish any such articles. Thus, the authors reached the conclusion that specialized journals remain the best home for their work on accounting education research. In addition, readers interested in accounting education topics still turn to specialized journals. However, the presence of accounting education articles in other journals could also suggest, in the authors' view, that scholars are being directed to publish in journals considered highly ranked in their respective countries.

Another study reviewing the accounting education literature is by Ameen and Guffey (2017), who examined the first 16 volumes of Advances in Accounting Education (AAE): Teaching and Curriculum Innovations, published between 1998 and 2015. The authors evaluated 195 articles in the journal, focusing on citation and citation-rate analysis to identify the 20 most influential papers and the 30 most influential authors (out of 383 identified). They also compared the research themes and methodologies used in AAE with those published in JAcEd and IAE for the period 2006–2015. While the AAE analysis focused on identifying the articles with the highest impact, the comparison with the other two journals emphasized the most common research themes and methodologies in the published articles, thus offering a broader perspective on the field of accounting education.

The results indicate that, in AAE, the most common themes were classroom pedagogy and student issues, while in JAcEd and IAE, the predominant topics were teacher issues and classroom pedagogy. The dominant methodologies in all three

journals were “persuasive argument” and “persuasive argument with statistical support,” accounting for over 60% of the articles published. The analysis provides authors with useful information about the types of articles and methodologies that had the greatest impact, guiding their choice of the appropriate journal for publication.

This paper complements the existing literature with a citation-oriented approach and an overview of methodological trends, offering valuable insights for researchers interested in accounting education and potential future research directions.

The recent work by Cao et al. (2024) represents a significant advance in the accounting education literature, being a longitudinal review that captures the transition from traditional topics (curriculum, assessment, teaching methods, educational technology, teacher and student issues in accounting) to new challenges introduced by the COVID-19 pandemic and the current Fourth Industrial Revolution. The authors frame their study around two research questions addressing key trends and themes in accounting education, as well as future directions for maintaining relevance in a dynamic global environment.

Their analysis was based on 673 articles published between 2005 and 2023 in 17 top accounting journals (ranked A* and A in the Australian Business Deans Council Journal Quality List—ABDC list), also including some specialized journals such as JAcEd, Accounting Research Journal (ARJ), and Journal of Business Ethics (JBE). Fourteen journals actively published articles on accounting education during the period under analysis; of these, IAE and AE are dominant, comprising 80.16% of all articles published.

The working method involved a multifaceted approach: bibliometrics with HistCite (co-citation analysis), VOS Viewer (bibliographic linkage), CiteSpace (structuring the knowledge map), and the R package Bibliometrix (providing descriptive insights: citation trend, author profile, keyword co-occurrence analysis, and thematic mapping). This complex strategy lends accuracy and objectivity to the findings, capturing thematic patterns in a comprehensive manner.

The three main research streams identified by Cao et al. (2024) are (1) accounting pedagogy, (2) competencies, and (3) ethics. In particular, the accounting pedagogy stream (the most significant) underscores learning approaches such as active learning, blended learning, and flipped classrooms, confirming a sustained interest in modernizing teaching practices. Themes related to competencies focus on the integration of technical skills (digital literacy, data analysis) and soft skills (communication, emotional intelligence), while discussions on ethics emphasize the need to foster professional responsibility among future accountants.

Compared with earlier syntheses (e.g., the periodic studies by Apostolou et al.), Cao et al. (2024) analyze a broader time range (2005–2023) and a wider set of journals, going beyond a strictly defined set of topics. Only the reviews in accounting education by Jackling et al. (2013), Rebele and Pierre (2015), and Apostolou et al. (2017b) cover comparable periods—20 and 18 years, respectively. In examining publications by year, Cao et al. (2024) report an average annual growth rate of about 15% in the number of articles on accounting education, indicating rising interest in this domain. Furthermore, the authors point out that between 2005 and 2023, 1,153 researchers contributed to this field, and 14 of them were notably productive, each publishing more than five articles on accounting education. The study also highlights a significant increase in citations per year, showing a continued push to improve and innovate how accounting is taught.

Table 1: Summary of longitudinal reviews on research in accounting education

Nr.	Author	Journal	Year of publication	Revised diaries	Period analyzed	Nr. articles reviewed	Authors' main suggestions/recommendations
1	Cao et al.	IAE	2024	17 Journals: JAcEd, AE, IAE, plus 14 other Journals classified A*, A and B (ARJ, JBE) from ABDC list	2005-2023	673	Two main research agendas are suggested: (1) strengthening education on the role of the accounting profession in protecting the environment by revising the curriculum and clearly establishing the relevant competencies, and (2) improving student engagement in the new post-pandemic context by harnessing technologies of the Fourth Industrial Revolution (4IR)-from augmented reality and the Internet of Things (IoT) to virtual reality-without neglecting the benefits of face-to-face interactions.
2	Apostolou et al.	JAcEd	2017b	6 Journals: JAcEd, AE, AAE , IAE, GPAE, AEJ	1997-2016	7	It is recommended to develop theories or apply theories developed in other disciplines and validate them on several populations.
3	Ameen and		2017	EFA EFA, JAcEd, IAE	1998-2015 2006-2015	195 661	Authors are provided with useful information about the types of articles and methodologies that have had the greatest impact, making it easier for them to choose the right journal to publish their research.
4	Rebels and Pierre	JAcEd	2015	1 Logbook: JAcEd	1991-2015	9	It is emphasized that improvements in accounting education must keep pace with new demands related to demographic changes in student demographics, the use of technology for the educational curriculum, the importance of information systems for graduates and the growing gap between accounting education and practice.
5	Marriott et al.	BAR	2014	6 Journals: JAcEd, AE, AAE, IAE, GPAE, AEJ plus 11 non-specialized journals	2005-2009	446+70	There is a need for more articles on accounting education in non-specialized journals.
6	Jackling et al.	AE	2013	1 Diary: AE	1992-2011	222	The need to draw on some research methods from the disciplines of education and psychology is emphasized. This would provide an opportunity to test theories and thus expand research areas.
7	Urbancic	AEJ	2009	6 Journals: JAcEd, AE, AAE, IAE, GPAE, AEJ	1998-2007	868	For example, the need for a successful accounting education journal in the Pacific is suggested as an example of the need to attract contributions primarily from countries such as Australia, Hong Kong, India, Japan, Korea, New Zealand, Singapore, Taiwan or Thailand.
8	Paisey and Paisey	AE	2004	1 Diary: AE	1992-2001	209	It is emphasized that research work in the field of accounting education needs to build on a solid educational knowledge base.

The thematic evolution revealed by Cao et al. (2024) begins between 2005 and 2010 with an emphasis on “accounting textbooks” and “business education,” followed by a growing focus on “the accounting profession,” “ethics,” and “general skills” from 2010 onward. Reflecting the push to adapt to technological innovations, there is a deepening discussion on the crucial skills for future professionals (digitization, advanced data analytics, soft skills). Additionally, recent publications focus on the COVID-19 pandemic’s impact, questioning the readiness of both teachers and students for online or blended learning.

Cao et al. (2024) primarily propose two research agendas: (1) strengthening education on the accounting profession’s role in environmental protection by revising the curriculum and clearly defining relevant competencies, and (2) enhancing student engagement in the post-pandemic era through leveraging Fourth Industrial Revolution technologies—from augmented reality and the Internet of Things (IoT) to virtual reality—without overlooking the benefits of face-to-face interaction. This aligns with the need for the accounting profession to serve the public interest and sustainability, while also providing opportunities for developing countries to fortify their accountability. Moreover, the authors observe a shift from teacher-centered to student-centered learning approaches. Consequently, curriculum developers and educators must adapt their teaching strategies, integrate technical competencies and ethical values, and strengthen the overall academic culture.

In Table 1, we have summarized the longitudinal reviews of accounting education research examined in this section, to highlight the main trends, current status, and emerging demands in the field.

4. Quantitative analysis of accounting education literature

4.1 Introductory considerations

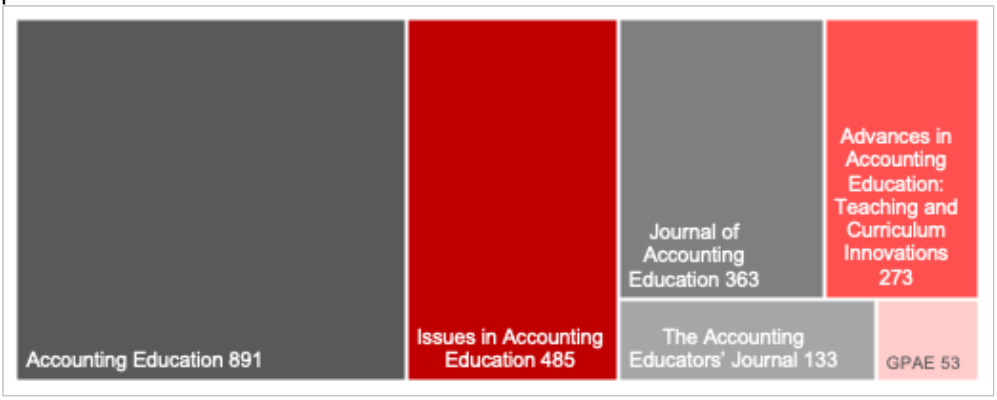
In the quantitative analysis of the literature in the field of accounting education, we focused on articles from two main sources, namely journals specialized in accounting education and WoS-indexed journals that are not specialized in accounting education. Beyond these two main sources for articles on accounting education, there is, of course, a third, broader category of journals that are neither specialized in accounting education nor indexed by WoS. However, we limited our analysis to the first two categories/sources because we considered that, generally, these journals publish higher-quality articles—either because they address an audience with in-depth research interests in accounting education (the first category) or because the review and selection process for WoS-indexed journals is usually regarded as more rigorous than in other publications.

Traditionally, the specialized accounting education journals recognized in the literature and considered in this analysis are: AE, IAE, JAcEd, AAE: Teaching and Curriculum Innovations, The Accounting Educators’ Journal, and GPAE. The first four of these journals are SCOPUS-indexed and are also published by reputable publishers (Routledge, American Accounting Association, Elsevier, and Emerald). They have been active in the field for several decades, publishing both scholarly articles on accounting education, instructional resources, and case studies. In this endeavor, we focused strictly on scholarly articles to maintain relevance and to ensure a

comparable analysis. A total of 2,198 articles were identified in these six journals for the period 1997–2024, the distribution of which is shown in Figure 1.

In terms of WoS-indexed journals, we considered those included in the WoS Core Collection and indexed in SSCI and SCIE. In total, 79 WoS-indexed journals were identified as having published scientific articles on accounting education during the reference period. Of these, 13 journals had at least five articles on this topic, as highlighted in Figure 2.

Fig. 1: Distribution of scholarly articles on specialized accounting education journals in the period 1997-2024



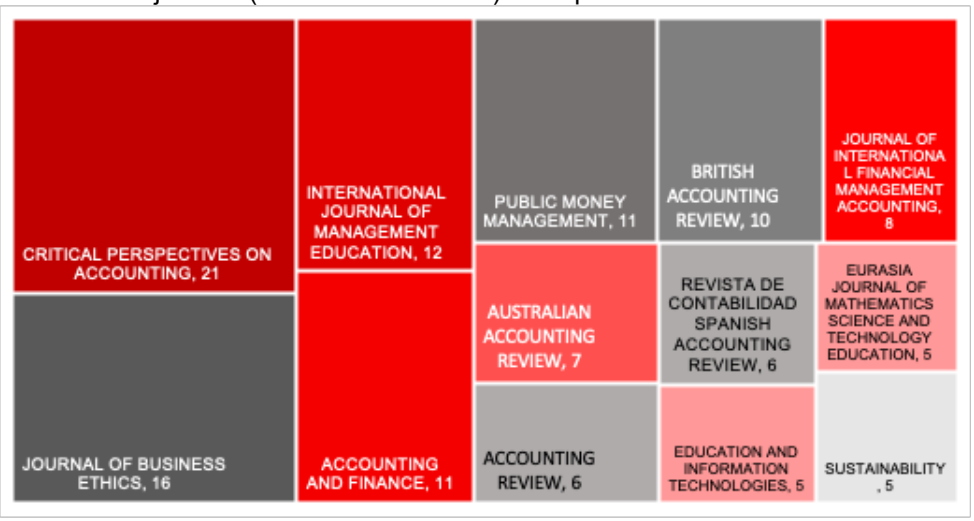
It is noteworthy that, in addition to journals devoted to related areas (e.g., Critical Perspectives on Accounting or Journal of Business Ethics), there are also journals with a broader managerial or educational focus (e.g., International Journal of Management Education) that publish papers relevant to accounting education.

The relevant point in time for the WoS indexing was the publication year of each article. For some journals, indexing is no longer currently valid, but the articles were indexed at the time of publication. Overall, for the WoS-indexed journals category, 224 articles were identified between 1997 and 2024, distributed by topic and type of research, as summarized in Table 3.

The period considered in this analysis covers 1997–2024 for articles from the two categories of sources (specialized journals and WoS-indexed journals). For the purpose of highlighting the evolution, this period was divided into nine sub-intervals: 1997–1999, 2000–2002, 2003–2005, 2006–2009, 2010–2012, 2013–2015, 2016–2018, 2019–2021, and 2022–2024. This segmentation made it possible to observe the dynamics of the number of articles and the main research themes, as well as to assess the type of methodological approach (empirical or descriptive).

The analysis examined the comparative evolution of the total number of articles published in the two categories, as well as their distribution by topic and by type of research (empirical or descriptive). For articles published in WoS-indexed journals, the geographical dimension of the research was also examined, i.e., the authors' country of origin. Quantitative data are presented in Tables 2 and 3 and the associated figures (Figures 1 and 2).

Fig. 2: Distribution of scientific articles in the field of accounting education published in WoS indexed journals (with at least 5 articles) in the period 1997-2024



4.2 Evolution of the total number of scientific articles in the field of accounting education

A first aspect analyzed is the evolution of the total number of scientific articles in the field of accounting education, published both in specialized journals and in WoS-indexed journals. As far as specialized accounting education journals are concerned, several fluctuations can be observed from one interval to another. Thus, between 1997 and 1999 there are 216 articles, followed by 206 in 2000–2002 and 223 in 2003–2005. A significant jump occurs in 2006–2009, when the highest level—330 articles—is reached. After this peak, there is a decrease to 291 articles in 2010–2012, 231 in 2013–2015, and 205 in 2016–2018, respectively. In 2019–2021, 203 articles are recorded, a marginal decrease compared to the previous period. However, the last analyzed interval (2022–2024) shows a notable rebound to 293 articles, marking an increase of about 44% compared to the previous interval. Overall, although some cyclicality may be observed—featuring episodes of growth and decline—the number of publications in specialized journals remains relatively stable overall, and this recent uptick suggests a possible renewed interest in accounting education among these journals.

Table 2 : Evolution of the number of scientific articles on accounting education published in journals specialized in accounting education between 1997-2024

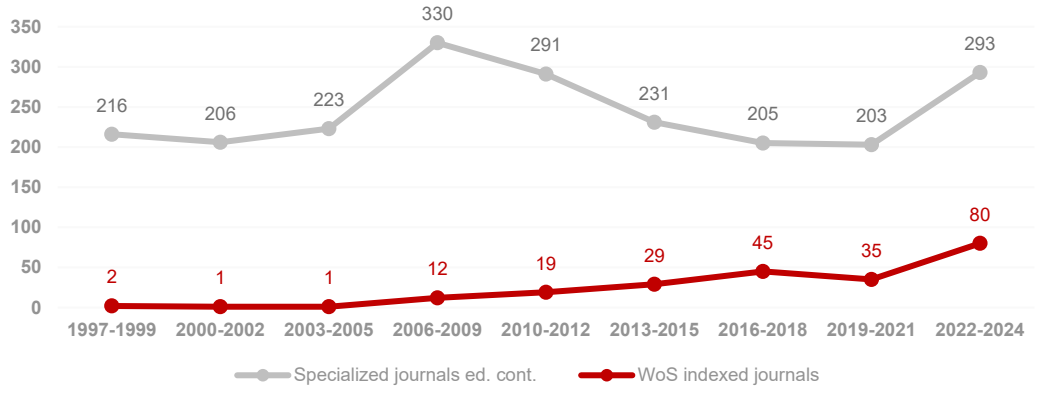
Period	1997-1999			2000-2002			2003-2005			2006-2009			2010-2012			2013-2015			2016-2018			2019-2021			2022-2024			1997-2024		
Research Topic/Type	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	***E***	***D***	***Total***
Curricular aspects	14	6	20	4	12	16	8	26	34	11	9	20	14	14	28	7	9	16	10	7	17	11	10	21	10	17	27	**89**	**110**	**199**
Core competences	14	23	37	5	12	17	6	5	11	5	4	9	6	26	32	9	9	18	11	3	14	20	23	43	7	9	16	**83**	**114**	**197**
Teaching/assessment process	10	12	22	14	17	31	22	20	42	33	21	54	15	16	31	11	14	25	21	11	32	10	13	23	33	25	58	**169**	**149**	**318**
Content area teaching	17	44	61	10	48	58	12	40	52	51	53	104	18	63	81	17	40	57	11	13	24	0	0	0	10	35	45	**146**	**336**	**482**
Educational Technology	2	7	9	11	6	17	8	6	14	3	15	18	17	6	23	17	8	25	17	3	20	11	3	14	20	12	32	**106**	**66**	**172**
Students	37	3	40	25	6	31	33	6	39	9	50	59	32	12	44	42	7	49	45	9	54	49	11	60	52	18	70	**324**	**122**	**446**
Teachers	15	12	27	20	16	36	16	15	31	33	33	66	24	28	52	27	14	41	19	25	44	12	30	42	16	29	45	**182**	**202**	**384**
Total	**109**	**107**	**216**	**89**	**117**	**206**	**105**	**118**	**223**	**145**	**185**	**330**	**126**	**165**	**291**	**130**	**101**	**231**	**134**	**71**	**205**	**113**	**90**	**203**	**148**	**145**	**293**	**1099**	**1099**	**2198**

Table 3 : Evolution of the number of scientific articles on accounting education published in WoS indexed journals between 1997-2024

Period	1997-1999			2000-2002			2003-2005			2006-2009			2010-2012			2013-2015			2016-2018			2019-2021			2022-2024			1997-2024		
Research Topic/Type	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	**E**	**D**	**Total**	***E***	***D***	***Total***
Curricular aspects	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	1	3	4	0	3	3	4	0	4	1	1	2	7	9	16
Core competences	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	1	3	4
Teaching/assessment process	0	1	1	0	0	0	0	0	0	1	3	4	4	2	6	3	10	13	2	3	5	4	4	8	5	6	11	19	29	48
Content area teaching	0	0	0	0	1	1	0	0	0	1	2	3	1	4	5	1	3	4	6	5	11	3	3	6	7	16	23	19	34	53
Educational Technology	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1	3	5	1	6	6	8	14	13	11	24
Students	1	0	1	0	0	0	0	0	0	2	0	2	2	1	3	1	1	2	5	3	8	3	2	5	8	6	14	22	13	35
Teachers	0	0	0	0	0	0	1	0	1	2	0	2	0	2	2	0	5	5	10	5	15	3	3	6	4	9	13	20	24	44
Total	**1**	**1**	**2**	**0**	**1**	**1**	**1**	**0**	**1**	**6**	**6**	**12**	**8**	**11**	**19**	**6**	**23**	**29**	**25**	**20**	**45**	**22**	**13**	**35**	**32**	**48**	**80**	**101**	**123**	**224**

On the other hand, regarding WoS-indexed journals, the data indicate a marked upward trend in articles dealing with accounting education. While in 1997–1999 and 2000–2002 the number was very low (2 and 1 article respectively), and in 2003–2005 it remained at a similar level (1 article), from 2006–2009 there was a noticeable increase to 12 articles. This upward trend is accentuated in the following subintervals, with 19 articles in 2010–2012, 29 in 2013–2015, and 45 in 2016–2018. Even though the next period, 2019–2021, shows a decline to 35 articles, the 2022–2024 interval marks a spectacular increase to 80 articles—more than double the previous interval. This trajectory confirms that interest in accounting education has steadily grown in WoS-indexed journals, which may also signal a qualitative evolution of research on this topic, given the higher publication requirements in such journals.

Fig. 3: Evolution of the total number of scientific articles in the field of accounting education between 1997-20 24



4.3 Main research topics addressed

To delineate the research area in accounting education into several main themes, we started from the categorization proposed by Apostolou et al. (2019), which includes curriculum and instruction, content-area teaching, educational technology, students, and teachers. Given the complexity of the curriculum and instruction theme, it was divided into three narrower subthemes: curricular issues, core competencies, and the teaching/assessment process. Thus, in total, seven distinct research strands emerged: (1) curricular issues, (2) core competencies, (3) teaching/assessment, (4) content-area teaching, (5) educational technology, (6) students, and (7) teachers.

In journals specializing in accounting education, over the entire 1997–2024 period, content-area teaching is the main research topic (22%), followed by students (20%) and teachers (17%), while the teaching/assessment process stands at 14%. Curricular issues and core competencies each register 9%, with educational technology at a lower level of 8%. These figures suggest that, overall, research has focused more on how subject content is structured and on the relationships between

students, teachers, and the teaching process. However, when looking exclusively at the most recent period (2022–2024), significant redistributions among themes emerge: the teaching/assessment process rises to 20%, educational technology to 11%, and student-related studies to 24%. Meanwhile, content-area teaching drops to 15%, core competencies to 5%, and teachers to 15%, while curricular issues maintain their share at 9%. These changes reflect a more intense orientation (in the last period compared to the entire span) towards teaching strategies and the use of technological resources, accompanied by increased interest in student experiences, expectations, or characteristics. The fact that articles focusing on educational technology did not register an even higher share (as might be expected given the current context) can be explained by the gradual integration of technology into the teaching process, thus reducing the need to address it as a distinct field (Apostolou et al., 2021).

Fig. 4a: Share of research topics in accounting education between 1997-2024

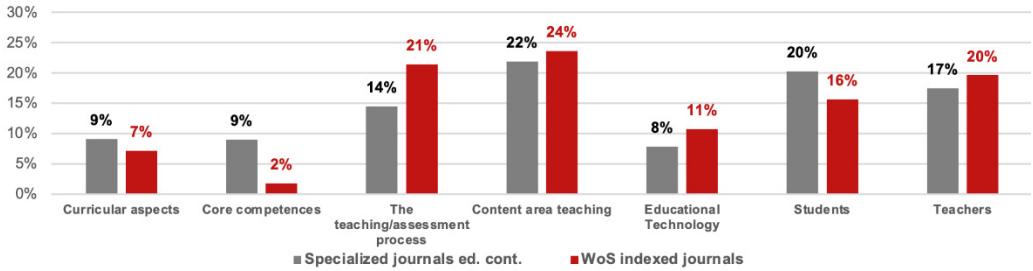
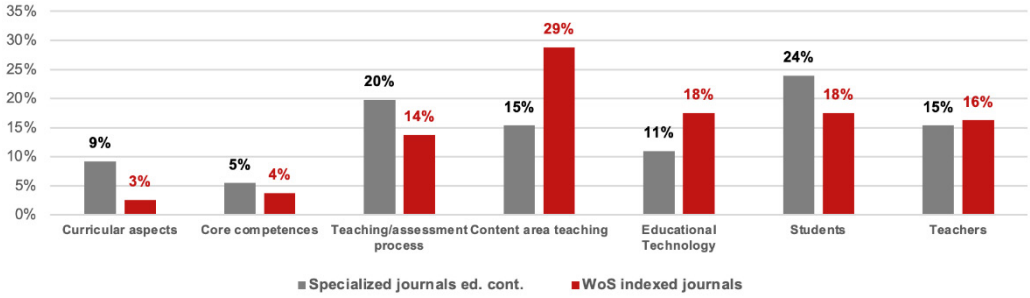


Fig. 5b: Share of research topics in accounting education in the period 2022-2024



In WoS-indexed journals, for the entire 1997–2024 interval, the highest percentages are for content-area teaching (24%) and the teaching/assessment process (21%), followed by teachers (20%) and student topics (16%). Educational technology occupies 11%, while curricular issues and core competencies stand at 7% and 2%, respectively. These values suggest a pronounced emphasis on disciplinary structure and the refinement of teaching/assessment practices, alongside significant research into the role of teachers and student engagement in learning. Examining the recent period (2022–2024), content-area teaching rises to 29%, and both instructional

technology and student topics increase to 18%, reflecting growing interest in the digitization of education and its implications for training future accounting professionals. Meanwhile, the share of teaching/assessment drops to 14%, teachers to 16%, and curricular issues and core competencies to 3% and 4%, respectively, indicating a possible reorientation of research toward innovative approaches in constructing specialized content and assimilating modern technologies.

Based on the above, the analysis of topic evolution and distribution in the literature of accounting education journals shows ongoing dynamism, with substantial variations in attention paid to certain themes. Nonetheless, a more stable increase in research devoted to student issues is evident, indicating that, despite fluctuations in other research areas, the focus on the learning experience and on the profile of new generations remains essential. On the other hand, the progression of topics associated with educational technology has become increasingly visible, especially toward the end of the period under review, signaling both the academic environment's adaptation to new digital realities and a repositioning of researchers toward teaching methods and tools. Furthermore, technology is gradually being integrated into the teaching process, thus reducing the need to address it as a separate field. With higher education transitioning to hybrid and online formats, this focus on educational technology is a natural response to both student needs and the demands of the labor market. In the medium to long term, the combined emphasis on students and the adoption of educational technologies may stimulate a profound reform in the design and delivery of accounting courses, leading to more interactive teaching methods and a learning environment better suited to today's requirements.

4.4 Predominant type of research (empirical or descriptive research)

Traditionally, scientific research is divided into two categories—empirical and descriptive—differentiated by the type of data collected, the nature of the processing, and the way conclusions are formulated. According to Apostolou et al. (2018), empirical articles base their conclusions on data analysis, generally adopting a quantitative approach. In contrast, papers that present strategies, describe innovations, or report the perceptions of groups (e.g., students, teachers) without using statistical analysis are categorized as descriptive, predominantly qualitative in nature.

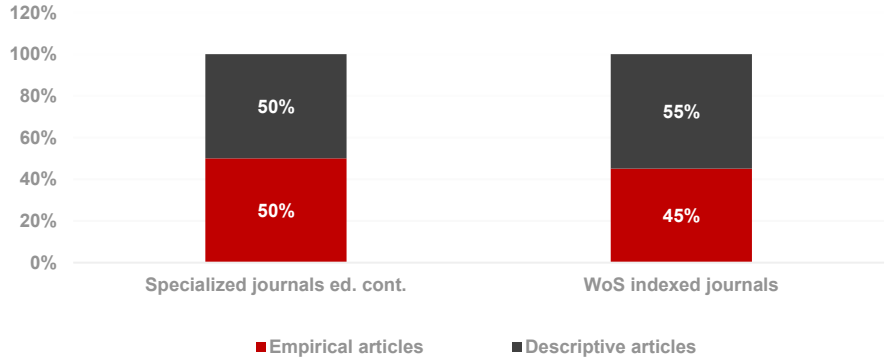
In addition, in the accounting education literature, two other categories of publications are identified: instructional resources and case studies. In our analysis of accounting education articles published in WoS-indexed journals, we could identify only the first two categories (empirical and descriptive articles).

Figure 5 shows a balance between these two types of articles—empirical research (50%) and descriptive research (50%)—published in specialized accounting education journals. This equal proportion indicates a methodological variety, including both statistical analyses and studies focused on describing practices, strategies, or teaching methods. Essentially, it illustrates the need to approach accounting education from both a hypothesis-testing, data-driven perspective and a qualitative one, emphasizing the context and processes involved in teaching.

In WoS-indexed journals, between 1997 and 2024, the proportion is 45% empirical and 55% descriptive articles. Consequently, on this topic, WoS journals appear to publish slightly more descriptive than empirical research, which may reflect

an interest in presenting and discussing educational approaches not necessarily limited to statistical tests or extensive quantitative analysis. This finding does not, of course, preclude methodological rigor but rather highlights the variety of formats through which the academic community shares its insights and innovations in accounting education.

Fig. 5: Share of descriptive empirical research in accounting education between 1997-2024



Overall, it can be observed that although there are some differences in their relative proportions between specialized and WoS-indexed journals, both empirical and descriptive research remain well represented in the accounting education literature. On the one hand, empirical studies strengthen the evidence regarding the effectiveness of certain teaching methods or validate specific hypotheses, while on the other hand, descriptive articles clarify instructional practices and strategies by providing examples and qualitative insights that statistics alone may not fully capture. Looking ahead, maintaining a balance between these two types of research can lead to a more comprehensive understanding of how accounting education evolves and transforms, integrating both empirical arguments and concrete experiences from academia.

4.5 Geographical origin of the authors

Analysis of the geographic location of authors of articles in accounting education has gained increasing importance in synthesis research (Apostolou et al., 2013–2023; Marriott et al., 2014; Churyk et al., 2024), particularly because of the cultural and methodological implications that geographic distribution may have on the generalizability of findings. In line with this perspective, we consolidated data on the country of origin of the authors of articles published in WoS-indexed journals between 1997–2024 (Table 4, Figures 6 and 7), identifying contributions from 49 countries.

Table 1: Geographical origin of authors of articles in the field of accounting education published in WoS indexed journals between 1997-20 24

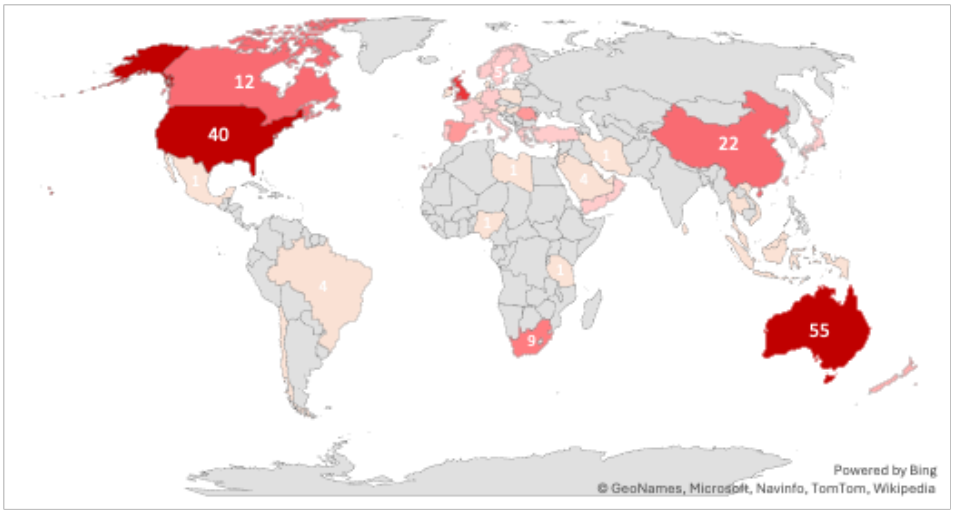
Country	No. of articles	Country	No. of articles	Country	No. of articles
Australia	55	Poland	4	Chile	1
USA	40	Saudi Arabia	4	Denmark	1
United Kingdom	37	Taiwan	4	Hungary	1
China	22	Yemen	4	Iran	1
Canada	12	Finland	3	Ireland	1
Spain	12	Norway	3	Kuwait	1
New Zealand Dollar	9	Portugal	3	Libya	1
South Africa	9	Turkey	4	Mexico	1
Netherlands	5	Cyprus	2	Nigeria	1
Romania	5	Indonesia	2	Qatar	1
Sweden	5	Japan	2	Singapore	1
Brazil	4	Jordan	2	Switzerland	1
France	4	Oman	2	Tanzania	1
Germany	4	Sri Lanka	2	United Arab Emirates	1
Greece	4	Thailand	2	Vietnam	1
Italy	4	Austria	1		
Malaysia	4	Belgium	1		

The ranking is led by Australia (55 articles), followed by the United States (40 articles) and the United Kingdom (37 articles). Next come China (22 articles), Canada and Spain (12 each), New Zealand and South Africa (9 each), and the Netherlands, Romania, and Sweden (5 each), along with other countries contributing fewer publications.

In terms of continents, Oceania and North America are mainly represented by Australia, New Zealand, the United States, and Canada, respectively. Meanwhile, in Europe, the UK stands out, alongside countries such as Spain, Romania, France, Germany, Italy, Sweden, Norway, the Netherlands, and Poland.

There is also a notable presence of Romanian authors in five articles on accounting education (Păcurari and Nechita, 2013; Pitulice et al., 2018; Guș e and Mangiuc, 2022; Ionescu-Feleagă et al., 2022; Bunea and Guinea, 2023), three of which were published in Romanian WoS-indexed journals (Transylvanian Review of Administrative Sciences, Revista de Cercetare și Intervenție Socială, Amfiteatru Economic). Also, authors from Africa (e.g., Nigeria, Libya, Tanzania) and the Middle East (e.g., Saudi Arabia, Qatar, United Arab Emirates) are present, confirming a significant expansion of academic interest in accounting education topics.

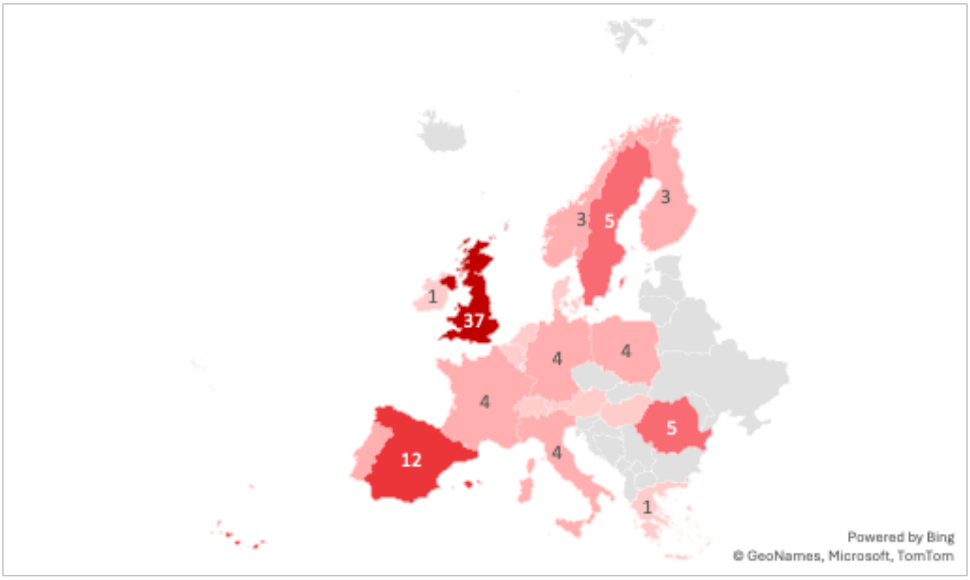
Fig. 6: Geographical map of the worldwide spread of authors of articles in the field of accounting education published in WoS indexed journals (with at least 3 articles) in the period 1997-2024



According to Marriott et al. (2014), in specialized journals, for the period 2005–2009, authors from North America (the United States and Canada) dominated (78%), followed at a considerable distance by European and Australian authors (9%). A more recent study (Churyk et al., 2024), focusing on 2023, shows a decrease in the dominance of North American authors (51%), accompanied by a slight increase in the number of authors from Australia (14%) and Europe (17%). In WoS-indexed journals, contributions come from a wider variety of geographical regions, and the primacy of North American authors (18%) is relatively diminished, surpassed by European countries (36%), Australia and New Zealand (22%), and Asia (19%).

This distribution confirms that, at the WoS level, articles on accounting education are published in an array of journals with global reach, enabling researchers from diverse cultural and institutional contexts to participate. Moreover, the involvement of countries in Central and Eastern Europe (such as Romania, Poland, and Hungary) and in Asia (China, Taiwan, Indonesia) suggests that accounting education topics are increasingly relevant in geographic areas where research in this field has not traditionally enjoyed broad international visibility. Consequently, the geographical origin of authors remains an important factor in evaluating and interpreting findings in accounting education studies, both in terms of the validity of results across different cultural contexts and from the perspective of international collaboration and the exchange of best practices.

Fig. 7: Geographical map of the European spread of authors of articles in the field of accounting education published in WoS indexed journals in the period 1997-2024



5. Qualitative analysis of accounting education literature

In this section, we analyzed, in terms of content, the most cited scientific articles in the field of accounting education published in WoS-indexed journals between 1997 and 2024. We structured this analysis according to the seven main research themes identified in the previous chapter, namely: curricular issues, core competencies, the teaching/assessment process, teaching by content area, educational technology, students, and teachers. For each theme, we selected those articles that generated the highest level of interest among researchers, as measured by the number of citations in WoS-indexed journals.

5.1 Articles on curricular issues

Cheng et al. (2014) produced a highly influential study introducing the concept of integrated reporting, as outlined by the International Integrated Reporting Council (IIRC), and exploring how this idea evolved during the four years following the IIRC's establishment in 2010. This evolution culminated in the IIRC's release of a consultation draft for an integrated reporting framework in March 2013. The authors then examine key issues surrounding the draft—issues the IIRC must address before its planned publication of the integrated reporting framework at the end of 2013. Their analysis is based on concerns identified and communicated to the IIRC by a subcommittee of the International Association for Accounting Education and Research (IAAER), comprising international accounting scholars. Finally, they point to several research questions emerging from the ongoing development and implementation of integrated reporting.

In their 2013 paper, Tweedie et al. advocate for expanding the range of ethical theories covered in accounting instruction. They note that core accounting programs often fail to adequately address non-Western ethical frameworks or newer Western ethical thought, as reflected in the ethical content of foundational accounting textbooks and publications by the International Federation of Accountants. To remedy this shortfall, the authors propose a "thematic" method of incorporating ethics into accounting education. Their approach tackles two potentially competing aims from International Education Standard 4: first, that all accountants receive training to uphold global ethical norms, and second, that ethics instruction incorporate diverse ethical traditions and practices across cultures and countries.

In an article by Bayerlein and Timpson (2017), the authors examine the extent to which accredited undergraduate accounting programs in Australia meet the needs and expectations of the accounting profession. The study focuses on how well the learning outcomes of these programs align with the minimum educational expectations (MEEs) set by professional bodies. The results indicate poor alignment, with only 12.2% of programs meeting five of the six MEE criteria. The research highlights the continued dominance of technical skills at the expense of professional skills such as communication, teamwork, and self-management. This situation is explained by the ongoing use of traditional teaching paradigms, which perpetuates misconceptions about modern accounting practice. From a curricular standpoint, the article proposes adjusting programs to better reflect contemporary professional demands and support the integration of graduates into the labor market.

5.2 Articles on basic skills

Evans et al. (2009) investigate the "Language for Professional Communication in Accounting" (LPCA) project implemented in the Master of Accounting (MAcc) program at Macquarie University, Australia. The study explores the integration of professional communication skills with disciplinary learning through an interdisciplinary collaboration between lecturers and English language teachers. Results show that this approach enhances student performance and improves curricula. Through interviews and quantitative analysis, the researchers demonstrate that LPCA contributes to the development of skills needed for professional practice and reduces the failure rate in the Professional Communication in Accounting (CPA) exams. The proposed model goes beyond traditional approaches by integrating skills directly into disciplinary content, eliminating perceptions of deficits related to linguistic and cultural diversity. The paper highlights the relevance of interdisciplinary collaboration in internationalized education and provides an example of good practice for developing graduate attributes in diverse educational contexts.

In his study, Wilkin (2021) explores the development of critical reflection in accounting education through an integrated approach. The study examines a problem-based learning exercise introduced in a postgraduate course, where students select and evaluate research articles relevant to their disciplinary knowledge. The activity included interactive elements such as individual and collaborative learning, oral and written exercises, along with formative and summative assessments. Results indicate that this approach stimulates critical reflection and an understanding of knowledge relativity, aiding students' transition to

independent or contextual “ways of knowing.” Comparative analysis of student perceptions over two consecutive years underscores the value of written reflection and qualitative feedback in building professional skills, including critical thinking. The article contributes to the literature by demonstrating how incorporating mainstream research into accounting education can develop essential professional competencies.

In a 2022 study, Barbosa et al. investigated how enterprise risk management (ERM) competencies influence accounting students’ perceptions of their readiness for the job market. Their findings indicate that acquiring fundamental skills in risk management—covering recognized ERM frameworks (e.g., ISO 31000 and COSO), targeted methodologies, and interpersonal abilities—significantly boosts students’ sense of professional preparedness. These insights underscore the importance of reinforcing curricula with ERM concepts and tools, alongside communication, strategic thinking, and change management proficiencies. Furthermore, the researchers present a model for evaluating risk management competencies, which universities can employ to pinpoint knowledge gaps. Overall, this work advances the discussion on the essential skills graduates need to succeed in increasingly complex economic environments.

5.3 Articles on the teaching/assessment process

In his 2008 study, Chiou examined whether using concept mapping could improve students’ learning outcomes and engagement. A total of 124 students from two advanced accounting classes at a Taiwanese university’s School of Management participated in the research. The findings led to two main conclusions. First, compared to a traditional expository teaching method, concept mapping resulted in significantly better performance. Second, most students viewed concept mapping positively, noting that it helped them understand, consolidate, and clarify accounting principles while also boosting their interest in the subject. They further acknowledged that the technique could prove beneficial in other academic areas.

Chabrak and Craig (2013) encourage accounting educators to enhance students’ capacity for creativity and critical reflection. Their study focuses on a project assigned to French accounting students, who were asked to analyze the Enron scandal. The submitted work revealed both inventive and unconventional narratives from some students, but also highlighted considerable cognitive dissonance. This tension arose from inconsistencies students observed between capitalist rhetoric and capitalist reality, as well as from the presence of capitalist ideals that were not challenged in the accounting curriculum. The assignment showcased student dissatisfaction with the dominant economic framework and its moral and ethical foundations, and explored various ways in which students coped with their cognitive dissonance. The authors suggest several teaching and curricular strategies, including stronger ties between accounting topics and broader social contexts (to question the assumed normalcy of capitalism), encouraging students to consider different cultures and viewpoints, and prompting them to question dominant ideologies.

In a 2013 article, Hopper examines the reform of accounting pedagogy, highlighting the need to incorporate the socio-economic context of accounting into education and to promote critical thinking. His commentary, inspired by Chabrak and Craig (2013), explores four dimensions: the accounting curriculum, the relationship

between universities and accounting professions, student expectations, and the role of accounting academics. Hopper argues that, despite external and commercial pressures on educational systems, universities and academics bear the responsibility to avoid turning accounting courses into imitations of professional training. Pedagogical reforms, such as those based on critical case studies (e.g., Enron), can foster creative and critical thought but require institutional backing and curriculum changes. The conclusion stresses that integrating ethical, theoretical, and public interest issues into accounting instruction is essential to move beyond a purely technical approach and address both academic and student demands.

5.4 Articles on content-area teaching

In his 2001 paper, McPhail argues that emotion should be introduced into accounting education, particularly by fostering emotional engagement with others. He proposes that one way to achieve this is through business ethics education, suggesting that strengthening ethical engagement with others can help counter accounting's tendency to dehumanize. Although there have been specific studies on ethics and accounting education, an open debate about the objectives of accounting ethics education or the specific techniques that could be used to achieve those objectives remains scarce. The paper posits that accounting has become dangerously dehumanized and that one of the most important goals of any business ethics education should be to cultivate empathy for others. Drawing on developments in the medical, legal, and engineering professions, the work proposes concrete methods that could rehumanize accounting and foster a moral commitment to others.

Liyanarachchi and Newdick (2009) investigated how moral reasoning and the risk of retaliation influence accounting students' likelihood of reporting major irregularities in an educational setting. In an experiment with 51 New Zealand undergraduates, participants completed hypothetical whistle-blowing scenarios and were classified by moral reasoning level and by the severity of retaliation (strong or weak). The findings indicate that both factors significantly affect the decision to blow the whistle, underscoring the importance of ethics education in preparing future accountants. However, no significant interaction emerged between moral reasoning and retaliation. The study highlights the need to incorporate whistle-blower protection into legislation and integrate ethics training into the accounting curriculum, fostering ethical behavior and a safer reporting environment.

Brazel et al. (2016) examine the impact of the outcome effect on auditors' professional skepticism. The study shows how supervisors' evaluations are swayed by their awareness of an investigation's outcome, regardless of the quality of the auditor's decisions. In an educational context, the authors suggest that this bias may undermine auditors' motivation to probe inconsistencies. Employing an experimental approach, the research finds that auditors receive lower evaluations when a skeptical inquiry fails to uncover errors, even if their questioning is professionally justified. Additionally, consultation with supervisors did not mitigate this effect, and corporate managers often regard time spent on fruitless investigations as wasted effort. The study underscores the importance of appraisal systems that promote skepticism and proposes professional training to counteract such bias. These findings are relevant to teaching professional skepticism in accounting, offering concrete examples of cultural barriers to acquiring this essential skill.

5.5 Articles on educational technology

Carenys et al. (2017) set out to compare the impact of video games versus simulations in higher education, focusing on three central factors—attributes, motivation, and learning outcomes—recognized as key contributors to successful game-based digital learning. Their analysis revealed marked distinctions in both attributes and motivation but found no significant difference in learning outcomes. In other words, although each teaching tool was similarly effective in fostering knowledge, video games appeared to provide enhanced motivation and features that, collectively, yield a richer educational experience. Consequently, the findings advocate for the inclusion of video games alongside simulations in higher-level accounting and business courses, proposing a blended approach that combines "the best of both worlds."

Alshurafat et al. (2021) investigate the factors influencing the use of online learning systems by accounting students in public universities in Jordan during the COVID-19 pandemic. Their approach integrates social capital theory, the theory of reasoned action, and the technology acceptance model. The study examines data from 274 students, underscoring the role of social trust in shaping perceptions of both the usefulness and ease of use of technology. Subjective norms and perceived ease of use also positively influence attitudes toward technology adoption, which in turn determines the intention and actual use of online systems.

The results highlight the importance of technological, individual, and social factors in successfully implementing such educational platforms, offering recommendations for optimizing their use in developing countries. The study underlines the need for IT training and user-friendly system design, contributing substantially to the literature on educational technology.

Wisneski et al. (2017) explore the impact of the learning environment on students' academic performance in introductory and advanced accounting courses. Their mixed-methods study examines whether the learning environment (online or face-to-face) in a foundational course affects knowledge transfer and performance in subsequent advanced courses. Findings indicate that students' performance in advanced courses is not significantly influenced by the environment of previous courses, although strategies for dealing with academic challenges differ. Online students rely more on individual study, whereas those in traditional settings seek peer collaboration and tutoring. The study emphasizes the importance of integrating technological resources and consistent support in both modalities. Its major contribution lies in broadening understanding of the cumulative effects of learning in different environments, thereby informing better design choices for online and traditional courses.

5.6 Articles on student issues

In their 2008 study, Ferreira and Santoso note that while earlier research often portrayed students as holding unfavorable views of accounting, more recent discussions suggest that the accountant stereotype may have gained more positive connotations. Because these viewpoints can shape learners' attitudes, they can in turn affect academic performance. The authors examine how perceptions correlate with performance by surveying both undergraduate and graduate students in

managerial accounting courses. The findings indicate that starting the course with negative perceptions of accounting hinders students' achievement, whereas adopting a more positive outlook by the semester's end exerts a beneficial impact on overall performance.

Blackmore et al. (2015) examine the experiences of Chinese accounting graduates from Australia in transitioning to the local labor market, highlighting the challenges faced by international students. Their longitudinal study involved 13 graduates and used Bourdieu's concepts of social, cultural, and symbolic capital to explore how they acquire skills and adapt to market rules. Key factors include English language proficiency, practical experience, and local social networks—all critical to enhancing graduates' "professional habitus." Although universities promote the internationalization of education as a way to increase cultural and social capital, the results point to a disconnect between student expectations and labor market realities. These findings underscore the need for education policies better aligned with market demands, offering practical opportunities and support for integrating international students.

Beatson et al. (2018) investigate how feedback based on partial exam results influences the self-efficacy beliefs of undergraduate accounting students. Their study, conducted at the University of Otago, explores the role of enactive success—demonstrated skills through tasks such as exams—in shaping self-efficacy. Data were gathered via surveys administered before and after the announcement of partial exam results. The results indicate that scores positively affect students' beliefs in their capacity to succeed, including their academic organization and willingness to seek help. The study highlights the significance of non-cognitive factors, such as self-efficacy, in educational success. The authors recommend careful assessment design to support the development of students' confidence in their abilities, offering practical implications for undergraduate education.

5.7 Articles on teacher issues

Kaplan (2011) examines the contributions of accounting research to professional knowledge, identifying substantial gaps in addressing practical challenges. The article stresses that accounting research has overly emphasized statistical methods applied to financial data, while overlooking measurement and evaluation issues central to practice. Kaplan highlights the necessity of fostering innovation, including in teacher education, to support future accounting professionals. He proposes a shift toward field studies and interdisciplinary approaches that tackle issues such as risk measurement, complex asset valuation, and the deployment of new financial instruments. Kaplan also recommends more active engagement of academics in practice development and the integration of technological innovations into educational programs. He urges closer collaboration between researchers and practitioners to produce relevant, applicable solutions.

Glaum et al. (2018) analyze the determinants of goodwill impairment decisions under IFRS using an international sample of firms from 21 countries. The study underscores the impact of regulatory frameworks and institutional investors on the appropriateness of these decisions. The findings show that goodwill impairment

is negatively correlated with economic performance and is shaped by managerial motivations, such as CEO compensation and earnings management preferences. In countries with strict enforcement of accounting regulations, firms exhibit more timely recognition of asset value declines. Conversely, in countries with weaker regulation, managerial discretion plays a significant role. The study also highlights that private monitoring by institutional investors can offset deficiencies in public oversight—especially in countries with less rigorous regulation—thereby emphasizing the institutional framework's importance for high-quality financial reporting.

Boyce and Greer (2013) approach the reform of accounting education from a critical perspective, focusing on the roles of cognitive dissonance, imagination, and critical thinking in the process. They argue that these elements are essential for an emancipatory accounting education centered on the real-world experiences of students and teachers. The study illustrates the obstacles encountered during reform efforts, highlighting students' resistance to information that conflicts with personal values and the capitalist system embedded in education. In this context, the authors propose employing cognitive dissonance as a pedagogical tool to foster critical thinking and social awareness, linking accounting instruction to social and global realities. This approach involves building an educational environment that questions the status quo and cultivates imagination as a catalyst for social change, emphasizing the teacher's role in managing these processes. The article offers valuable insights for transforming accounting education from a purely technical domain into a more social and critical field.

The qualitative analysis of the literature on accounting education underscores, first and foremost, the need for a broader curricular perspective. The reviewed articles indicate that focusing solely on technical skills is insufficient; it is crucial to integrate ethics, integrated reporting, and transversal skills (e.g., communication, strategic thinking) to adequately prepare future accountants.

Second, developing core competencies extends beyond specialized knowledge, encompassing abilities such as critical reflection, self-management, and an appreciation of diversity (including linguistic aspects). Research highlights the importance of interdisciplinary collaboration and integrated approaches, which not only enhance student performance but also increase curricular relevance.

Concerning the teaching/assessment process, there is an emphasis on active and critical methods (e.g., concept mapping) designed to spark creative thinking and address accounting's socio-economic dimensions. Studies propose that including complex case studies (like Enron) and ethical debates can bolster social responsibility and awareness.

Content-area teaching emphasizes the need to "re-humanize" accounting through the cultivation of empathy and moral reasoning. Integrating ethical considerations, including those related to whistleblowing, fosters an accounting culture capable of meeting real-world challenges.

Educational technology plays a major role, but research indicates that while video games and online platforms offer heightened motivation and interactivity, learning outcomes can be comparable between online and face-to-face environments. Social factors—such as trust and collaboration—remain paramount for effectively adopting new platforms.

Student-related issues reveal the effect of perceptions on academic performance and highlight the difficulties international graduates face when transitioning to the job market. Constructive feedback and the enhancement of self-efficacy can reshape attitudes toward accounting, improving results.

Finally, teachers play a vital role in reforming accounting education, requiring both collaboration with the profession and a critical, interdisciplinary outlook. Involving teachers in innovation and aligning content with practical realities helps cultivate a generation of accounting professionals better aligned with contemporary demands.

6. Conclusions and Recommendations

The analysis conducted in this article highlights a number of trends and challenges in the field of accounting education, reflecting on its evolution in a contemporary context marked by significant technological and social changes. One of the main contributions of the study is the integration of a comparative perspective between articles published in journals dedicated to accounting education and those published in broader journals indexed in WoS. This approach made it possible to identify differences and similarities in terms of the topics addressed and the methodologies used.

The results obtained show that, despite a certain cyclicality observed in specialized journals, the total number of scientific articles published in the field of accounting education remains relatively constant globally, with a notable revival in the recent period (2022–2024). In the case of WoS-indexed journals, the analysis even reveals a spectacular increase—up to 80 articles, more than double the previous interval. These findings are in line with trends reported in the literature, which in turn signal increasing concerns for accounting education (Apostolou et al., 2021).

Thematically, the data identified confirm a high level of interest in both the curricular dimension and the refinement of teaching and assessment practices, alongside increased attention to educational technology and student characteristics. These orientations are in line with the results of studies showing a gradual transition of research from traditional topics (curriculum, assessment) to the new demands of the digital environment (Paisey & Paisey, 2004; Cao et al., 2024).

At the same time, the findings indicate a significant emphasis on the integration of ethics and transversal skills in the training of future accountants, as well as the importance of developing critical reflection and self-management skills. These directions are consistent with the findings of other authors, who suggest the need to “rehumanize” accounting and to connect the teaching process to broader socio-economic contexts (Apostolou et al., 2017b).

In terms of research methodologies, the analyses carried out show that empirical and descriptive studies remain well represented, a trend that resonates with the findings formulated in the literature (Rebele & Pierre, 2015; Marriott et al., 2014). Therefore, adopting complementary approaches—statistical and qualitative—seems essential to capture the complexity of the educational process, and integrating perspectives from related fields (pedagogy, psychology) can add scientific rigor (Jackling et al., 2013).

The analysis of the geographical origin of authors of scientific papers in WoS-indexed journals reveals a more diverse worldwide participation. While in the previous literature the dominance of North American authors was much clearer, the identified data point to a balancing of contributions, with a strong increase from Europe, Australia/New Zealand, and Asia. This picture confirms that the relevance of accounting education has long transcended the boundaries of regions traditionally engaged in research, and the new international contributions open up additional prospects for academic partnerships and cross-cultural exchange of best practices.

In conclusion, the results suggest that, in parallel with previous concerns about the efficiency and structure of curricula, accounting education is moving increasingly toward an integrated approach, where digital, ethical, and interpersonal competences are essential to prepare professionals who are able to meet the current demands of the labor market. Educators play a critical role in this process, both through innovative teaching and assessment initiatives and by adopting an interdisciplinary and collaborative attitude aimed at linking theoretical content with practical reality and fostering the responsible development of future accountants.

In the context of global dynamics and accelerating transformations of the accounting profession, strategic directions are needed to strengthen accounting education. Future research should explore in detail how emerging technologies such as generative artificial intelligence, augmented reality, and the Internet of Things (IoT) can be effectively integrated into the curriculum. This integration is essential to ensure the relevance of curricula and to produce market-ready professionals.

There is also a need to pay more attention to the development of transversal competences such as critical thinking, communication, and self-efficacy. Future research could explore the impact of innovative teaching methodologies such as flipped classrooms, gamification, or virtual/augmented reality experiences on the development of these skills.

In view of the globalization of accounting education, it is recommended that international collaborations be intensified to support the exchange of best practices and the adaptation of curricula to different cultural contexts. Studies could investigate how regional specificities influence the effectiveness of teaching strategies and their impact on student performance.

Finally, emphasis should be placed on continuous teacher training to prepare them to use modern pedagogical methods and digital technologies. Future studies could evaluate the effectiveness of teacher training programs in adopting these innovations, thus contributing to the quality of accounting education.

With these conclusions and recommendations, the article contributes to strengthening the understanding of the field of accounting education and provides a solid basis for future research and practice initiatives.

References

Alshurafa'a, H., Al Shbail, M. O., Masadeh, W. M., Dahmash, F., Al-Msiedeen, J. M. (2021) Factors affecting online accounting education during the COVID-19 pandemic: an integrated perspective of social capital theory, the theory of reasoned action and the technology acceptance model. *Education and Information Technologies,* 26, 6995–7013

Ameen, E. C., Guffey, D. M. (2017) A citation analysis and review of research issues and methodologies in Advances in accounting Education: Teaching and Curriculum Innovations. *Advances in Accounting Education: Teaching and Curriculum Innovations,* 21, 1–25

Anderson, M. (2002) An analysis of the first ten volumes of research in Accounting, Business and Financial History. *Accounting, Business and Financial History,* 12(1), 1–24. In Paisey, C., Paisey, N. J. (2004) An analysis of accounting education research in accounting education: an international journal e 1992e2001. *Accounting Education: An International Journal,* 13(1), 69–99

Apostolou, B., Churyk, N. T., Hassell, J. M., Matuszewski, L. (2023) Accounting education literature review (2022). *Journal of Accounting Education,* 63, 100831

Apostolou, B., Dorminey, J. W., Hassell, J. M. (2022) Accounting education literature review (2021). *Journal of Accounting Education,* 59, 100781

Apostolou, B., Dorminey, J. W., Hassell, J. M. (2021) Accounting education literature review (2020). *Journal of Accounting Education,* 55, 100725

Apostolou, B., Dorminey, J. W., Hassell, J. M. (2020) Accounting education literature review (2019). *Journal of Accounting Education,* 51, 100670

Apostolou, B., Dorminey, J. W., Hassell, J. M., Hickey, A. (2019) Accounting education literature review (2018). *Journal of Accounting Education,* 47, 1–27

Apostolou, B., Dorminey, J. W., Hassell, J. M., Rebele, J. E. (2018) Accounting education literature review (2017). *Journal of Accounting Education,* 43, 1–23

Apostolou, B., Dorminey, J. W., Hassell, J. M., Rebele, J. E. (2017a) Accounting education literature review (2016). *Journal of Accounting Education,* 39, 1–31

Apostolou, B., Dorminey, J. W., Hassell, J. M., Rebele, J. E. (2017b) Analysis of trends in the accounting education literature (1997–2016). *Journal of Accounting Education,* 41(C), 1–14

Apostolou, B., Dorminey, J. W., Hassell, J. M., Rebele, J. E. (2016) Accounting education literature review (2015). *Journal of Accounting Education,* 35, 20–55

Apostolou, B., Dorminey, J. W., Hassell, J. M., Rebele, J. E. (2015) Accounting education literature review (2013–2014). *Journal of Accounting Education,* 33(2), 69–127

Apostolou, B., Dorminey, J. W., Hassell, J. M., Watson, S. F. (2013) Accounting education literature review (2010–2012). *Journal of Accounting Education,* 31(2), 107–161

Apostolou, B., Hassell, J. M., Rebele, J. E., Watson, S. F. (2010) Accounting education literature review (2006–2009). *Journal of Accounting Education,* 28(34), 145–197

Apostolou, B., Watson, S. F., Hassell, J. M., Webber, S. A. (2001) Accounting education literature review (1997–1999)

Barbosa, M. W., Martinez Carrasco, S. I., Rodriguez Abarca, P. C. (2022) The effect of enterprise risk management competencies on students' perceptions of their work readiness. *The International Journal of Management Education,* 20(2), 100638

Bayerlein, L., Timpson, M. (2017) Do accredited accounting programs in Australia meet the needs and expectations of the accounting profession? *Education + Training,* 59(3), 305–322

Beatson, N. J., Berg, D. A. G., Smith, J. K. (2018) The impact of mastery feedback on undergraduate students' self-efficacy beliefs. *Studies in Educational Evaluation,* 59, 58-66

Blackmore, J., Gribble, C., Rahimi, M. (2015) International education, the formation of capital and graduate employment: Chinese accounting graduates' experiences of the Australian labour market. *Critical Studies in Education,* 58(1), 69-88

Boyce, G., Greer, S. (2013) More than imagination: Making social and critical accounting real. *Critical Perspectives on Accounting,* 24(2), 105–112
Brazel, J. F., Jackson, S. B., Schaefer, T. J., Stewart, B. W. (2016) The outcome effect and professional skepticism. *The Accounting Review,* 91(6), 1577–1599
Cao, J., Kristanto, A. B., Gu, Z. (2024) Evolution of Research Streams and Future Research Directions in Accounting Education: Quantitative Systematic Literature Review. *Issues in Accounting Education,* 39(4), 19-53
Carenys, J., Moya, S., Perramon, J. (2017) Is it worth it to consider videogames in accounting education? A comparison of a simulation and a videogame in attributes, motivation and learning outcomes. *Revista de Contabilidad,* 20(2), 118–130
Chabrak, N., Craig, R. (2013) Student imaginings, cognitive dissonance and critical thinking. *Critical Perspectives on Accounting,* 24(2), 91–104
Cheng, C. K. C., Chow, E. Y. H., Lam, K. C. K., Lee, J. H. Y. (2023) Participation in internship, professional competition and overseas exchange and accounting students' subsequent academic and job market performance. *The International Journal of Management Education,* 21(100887), 1-15
Cheng, M., Green, W., Conradie, P., Konishi, N., Romi, A. (2014) The International Integrated Reporting Framework: Key Issues and Future Research Opportunities. *Journal of International Financial Management & Accounting,* 25(1), 90-119
Chiou, C. C. (2008) The effect of concept mapping on students' learning achievements and interests. *Innovations in Education and Teaching International,* 45(4), 375-387
Churyk, N. T., Eaton, T. V., Matuszewski, L. (2024) Accounting education literature review (2023). *Journal of Accounting Education,* 67, 100901
Evans, E., Tindale, J., Cable, D., Hamil Mead, S. (2009) Collaborative teaching in a linguistically and culturally diverse higher education setting: a case study of a postgraduate accounting program. *Higher Education Research & Development,* 28(6), 597-613
Ferreira, A., Santoso, A. (2008) Do students' perceptions matter? A study of the effect of students' perceptions on academic performance. *Accounting and Finance,* 48(2), 209-231
Glaum, M., Landsman, W. R., Wyrwa, S. (2018) Goodwill Impairment: The Effects of Public Enforcement and Monitoring by Institutional Investors. *The Accounting Review,* 93(6), 149-180
Hanushek, E. A., Woessmann, L. (2008) The role of cognitive skills in economic development. *Journal of Economic Literature,* 46(3), 607–668
Helliar, C. (2013) The global challenge for accounting education. *Accounting Education,* 22(6), 510–521
Hopper, T. (2013) Making accounting degrees fit for a university. *Critical Perspectives on Accounting,* 24(2), 127–135
Jackling, B., Natoli, R., Nuryanah, S., Ekanayake, D. (2013) Celebrating 20 Years of Publication of Accounting Education: an international journal: 1992–2011. *Accounting Education: An International Journal,* 22(1), 18-43
Kaplan, R. S. (2011) Accounting Scholarship that Advances Professional Knowledge and Practice. *The Accounting Review,* 86(2), 367-383
Liyanarachchi, G. A., Newdick, C. (2009) The impact of moral reasoning and retaliation on whistle-blowing: New Zealand evidence. *Journal of Business Ethics,* 89(1), 37–57
Marriott, N., Stoner, G., Fogarty, T., Sangster, A. (2014) Publishing characteristics, geographic dispersion and research traditions of recent international accounting education research. *The British Accounting Review,* 46, 264-280

McPhail, K. (2001) The other objective of ethics education: Re-humanising the accounting profession—A study of ethics education in law, engineering, medicine and accountancy. *Journal of Business Ethics,* 34(3–4), 279–298

Paisey, C., Paisey, N. J. (2004) An analysis of accounting education research in accounting education: an international journal e 1992e2001. *Accounting Education: An International Journal,* 13(1), 69-99

Ravenscroft, S. P., Rebele, J. E. (2008) The importance of accounting education research. *Journal of Accounting Education,* 26(4), 180–187

Rebele, J. E., Pierre, K. S. (2015) Stagnation in accounting education research. *Journal of Accounting Education,* 33(2), 128–137

Rebele, J. E., Tiller, M. G. (1986) Empirical research in accounting education: A review and evaluation. In Rebele, J. E., Stout, D. E., Hassell, J. M. (1991) A review of empirical research in accounting education: 1985–1991. *Journal of Accounting Education,* 9(2), 167–231

Rebele, J. E., Apostolou, B. A., Buckless, F. A., Hassell, J. M., Paquette, L. R., Stout, D. E. (1998a) Accounting education literature review (1991–1997), part I: Curriculum and instructional approaches. *Journal of Accounting Education,* 16(1), 1–51

Rebele, J. E., Apostolou, B. A., Buckless, F. A., Hassell, J. M., Paquette, L. R., Stout, D. E. (1998b) Accounting education literature review (1991–1997), part II: Students, educational technology, assessment and faculty issues. *Journal of Accounting Education,* 16(2), 179–245

Rebele, J. E., Stout, D. E., Hassell, J. M. (1991) A review of empirical research in accounting education: 1985–1991. *Journal of Accounting Education,* 9(2), 167–231

Reinstein, A., Calderon, T. G. (2006) Examining accounting departments' rankings of the quality of accounting journals. *Critical Perspectives on Accounting,* 17(4), 457–490

Rieckmann, M. (2012) Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures,* 44(2), 127–135

Schultz, T. W. (1961) Investment in human capital. *The American Economic Review,* 51(1), 1–17

Tettamanzi, P., Minutiello, V., Murgolo, M. (2023) Accounting education and digitalization: A new perspective after the pandemic. *The International Journal of Management Education,* 21(3), 100847

Tweedie, D., Dyball, M. C., Hazelton, J., Wright, S. (2013) Teaching global ethical standards: A case and strategy for broadening the accounting ethics curriculum. *Journal of Business Ethics,* 115(1), 1–15

Urbancic, F. (2009) Individual and Institutional Contributors to Research in Accounting Education. *The Accounting Educators' Journal,* 19, 21-44

Watson, S. F., Apostolou, B., Hassell, J. M., Webber, S. A. (2003) Accounting education literature review (2000–2002). *Journal of Accounting Education,* 21(4), 267–325

Watson, S. F., Apostolou, B., Hassell, J. M., Webber, S. A. (2007) Accounting education literature review (2003–2005). *Journal of Accounting Education,* 25(12), 1–58

Wilkin, C. L. (2021) Developing Critical Reflection: An Integrated Approach. *The British Accounting Review,* 53(4), 101043

Wilson, R. M. S. (2002) Accounting education research: a retrospective over ten years with some pointers to the future. *Accounting Education: An International Journal,* 11(4), 295-310

Wisneski, J. E., Ozogul, G., Bichelmeyer, B. A. (2017) Investigating the impact of learning environments on undergraduate students' academic performance in a prerequisite and post-requisite course sequence. *Internet and Higher Education,* 32(2017), 1-10

STUDIA UNIVERSITATIS BABEŞ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 33-50
DOI: 10.2478/subboec-2025-0002

GREEN HRM AND CSR AS ANTECEDENTS OF ORGANISATION FINANCIAL GROWTH

Lavinia-Mihaela BECEA[*]
Babeș-Bolyai University, Romania

Codruta OSOIAN
Babeș-Bolyai University, Romania

Abstract: This paper aims to examine the impact of green HRM and corporate social responsibility on financial performance growth, measured by profit, revenue, ROA, and ROE growth. To test the hypotheses, we employed ordinary least squares (OLS) regression analysis. The findings revealed that green HRM is a significant predictor of growth indicators, including profit, ROA, and ROE. Furthermore, contrary to our expectations, the implementation of CSR practices negatively impacts the firm's potential for economic growth. The empirical evidence shows that CSR is positively associated only with revenue growth. This study contributes to the literature by examining the debated impact of green HRM and CSR on financial performance growth while also offering valuable insights for practitioners and policymakers.

JEL classification: M42, O15, J81, Q56

Keywords: HRM, Green HRM, Corporate social responsibility, Performance growth, OLS

1. Introduction

Businesses have made sustainable production and development a top strategic priority in response to the escalating environmental challenges worldwide. Accordingly, organizations are interested in developing business strategies that can accomplish their goals while also protecting the environment and society (George et al., 2016). Recent studies have emphasized the critical role of human resource management, arguing that a successful environmental management system relies on various HR factors (Bilousko et al., 2024). Furthermore, sustainable HRM is increasingly recognized as a crucial driver of organizational competitiveness and effectiveness, supporting

[*] Corresponding author. Address: Faculty of Economics and Business Administration, Babeş-Bolyai University, 58-60, Teodor Mihali Street, 400591, Cluj-Napoca, Romania.
E-mail: lavinia.becea@econ.ubbcluj.ro

businesses in achieving financial success, community well-being, and positive environmental outcomes (Anlesinya and Susomrith, 2020). As a component of the environmental management system, green human resource management (green HRM) has attracted significant attention from researchers, highlighting the importance of educating managers on integrating environmental considerations into HRM practices (Ren et al., 2022). On the other hand, building on the idea that a company cannot thrive in isolation, it becomes essential for organizations to incorporate CSR principles into their actions (Latif et al., 2020).

While the existing literature provides essential insights into how green HRM and CSR practices affect organizational performance, there are still some significant gaps. *First*, previous research has emphasized the role of green HRM in contributing to environmental performance (Obeidat et al., 2020; Yusoff et al., 2020) or explored this concept in relation to the triple bottom line, which assesses economic, social, and environmental performance (Kramar 2014). However, *the literature is scarce* in addressing how green HRM contributes to non-green outcomes (Shen et al., 2018), such as specific financial returns (Ren et al., 2018). Moreover, there is a *dearth of literature* discussing organizational strategies that can ensure long-term performance. For instance, researchers generally avoid discussing financial firm growth due to the insufficient understanding of its definition, measurement and complex nature (Harb and Ahmed 2019; Demir et al., 2017). *Second,* there is an *inconsistency* in the literature regarding the contribution of green HRM to financial performance. While numerous studies highlight the positive impact of green HRM on financial performance (Nguyen and Nguyen 2024; Al-Abbadi et al., 2023; Longoni et al., 2018), other research suggests that investing in green practices may lead to additional costs and further compromise the financial outcomes (Acquah et al., 2021; Aguilera-Caracuel and Ortiz-de-Mandojana 2013; Triguero-Sánchez et al., 2013). Third, there are also fragmented findings regarding the link between CSR and organizational performance. While several academics highlight the economic benefits a business can gain from engaging in CSR activities (Bernal-Conesa et al., 2017; Orlitzky et al., 2003; Pava and Krausz 1996), others argue that CSR initiatives incur additional costs that reduce the economic gains (Barnett and Salomon 2006; Alexander and Buchholz 1978). Therefore, more research is required on this relationship.

In response to the gaps identified in the literature, this paper aims to propose a comprehensive model that clarifies the connection between green HRM and corporate social responsibilities on financial performance growth. In doing so, the findings offer several contributions to the existing literature. First, it advances understanding of green HRM by addressing the ambiguous question of why businesses adopt green HRM practices (Jackson et al., 2011). Moreover, it provides a thorough model that explains the contribution of green HRM to business economic performance. Second, this paper adds to the existing literature by examining CSR as an antecedent of financial performance growth.

This paper is organized as follows. After the introduction, the next section provides a review of the relevant literature and key concepts. The third section outlines the development of the hypotheses, followed by the fourth section that presents the research methodology. The subsequent section discusses the results, while the final section presents the conclusions, highlights the key findings, discusses implications and suggests directions for future research.

2. Literature review

2.1. Green HRM

Human resource management typically involves a set of practices focused on attaining organizational objectives (Sels et al., 2016). This set of practices usually encompasses procedures for recruitment and selection, training and development, compensation and rewards. However, recent HRM research suggested that *this system should be designed to target specific strategic goals* to better leverage an organization's workforce for long-term success (Lepak et al., 2006). Moreover, the research suggested that *HRM practices should be seen as complementary components* that collaborate to address specific organizational issues, rather than being treated as isolated practices. More specifically, the effectiveness of certain HRM practices can be enhanced when combined with others, leading to synergistic effects (Ichniowski et al., 1995). These two perspectives are particularly relevant for green HRM, which is tailored to support specific strategies meant to attain environmental performance. In a broader context, it can be defined as a set of HRM practices developed to achieve both economic and environmental objectives of the organization (Renwick et al., 2016).

As green HRM is a relatively new concept, the literature has highlighted various aspects of human resource management aimed specifically at enhancing environmental performance. For instance, practices such as green recruitment and selection, green training, green performance management, green pay and reward systems, and green employee involvement are considered key dimensions of green HRM (Tang et al., 2018). Another study (Shah 2019) advocated for a more holistic approach, encompassing green job analysis and design, green health and safety, green employee' relations and green involvement in the previously mentioned practices. However, the most widely used functions for green HRM practices are green recruitment, green selection, green training and development, green performance management, and green pay and reward systems (Yong et al., 2020). These studies reinforce the idea of HRM practices seen as a set of complementary components that need to work together in order to efficiently address organizational green issues. *Studying bundles of green HRM practices*, rather than individual practices, *would provide a clearer understanding* of how they interact to achieve strategic environmental goals (Dyer and Holder 1998; Dyer and Reeves 1995).

The literature widely acknowledges the connection between HRM and organizational performance (Syed and Jamal 2012; Yang and Lin 2009; Katou 2008). However, growing environmental concerns have prompted researchers to increasingly focus on green HRM as a strategic tool for addressing environmental challenges and achieving organizational performance (Kramar 2014). In this context, previous research revealed that green HRM contributes to environmental performance (Obeidat et al., 2020; Arda et al., 2019; Rawashdeh 2018), improved organizational culture (Jabbour et al., 2013), sustainable performance (Shahzad et al., 2023; Zaid et al., 2018), green competitive advantage (Muisyo et al., 2022), organizational citizenship behavior (Pham et al., 2019). Despite the fact that a recent literature review conducted on studies between 2007 and 2019 highlighted that the outcomes of green HRM have been the most explored subject related to this topic, *the relationship between green HRM and objective measurements of economic performance has received limited attention* (Ren et al., 2022).

2.2. Firm financial growth

Organizational performance is a multifaceted concept that reflects the overall success of an organization, encompassing various perspectives to assess progress toward predefined objectives (Zhang et al., 2008). For instance, businesses track their financial success using *financial performance* metrics, such as profitability, revenue growth, return on assets (ROA), and return on investment (ROI) (Salimath et al., 2008) or *operational performance metrics*, such as productivity, input quality, and costs (Koufteros et al., 2014). A recent study proposed a model to measure organizational performance, encompassing four dimensions: profitability, liquidity, growth, and stock market performance. However, they noted that while profitability was commonly used as the primary performance metric in most studies, liquidity, growth, and stock market performance were rarely considered (Hamann and Schiemann 2021).

Researchers generally avoid discussing firm financial growth due to the insufficient understanding of its definition, measurement and complex nature (Demir et al., 2017). The existing literature, however, discusses about the high-growth firm and financial sustainability concept. The high-growth firms provide a unique context for understanding firm growth considering the peculiarities of rapid growth (Delmar et al., 2003), while financial sustainability discuss long-term financial results (Obeidat et al., 2023; Harb and Ahmed 2019). More specifically, financial sustainability refers to the ability of an organization to maintain its financial health by generating consistent income to cover its expenses and obligations, without experiencing financial instability (Leon 2001).

2.3. Corporate Social Responsibility

Corporate social responsibility (CSR) emphasizes the obligation of organizations to make decisions and adopt policies that align with societal values and goals (Bowen 1953). CSR encompasses four key categories of social responsibilities (Carroll 1979, 499). Economic responsibility is the fundament of any business, which involves providing goods and services and selling them to generate profit. Legal responsibility refers to the obligation of organizations to comply with laws and regulations. Ethical responsibility pertains to the expectations society has from the organization, while discretionary responsibility refers to the voluntary actions a company chooses to take beyond the other three mentioned above. Highlighting its role in promoting sustainable development, CSR can be defined as the responsibility of businesses to enhance the overall well-being, focusing on environmental conservation, social justice and economic stability (Maignan and Ferrell 2000; European Commission 2003).

The commitment to CSR principles requires organizations to tailor the five dimensions of CSR (Vilanova et al., 2009) to suit their specific circumstances and needs (Jones and Wicks 2018). Building on the premise that a business cannot thrive in isolation, it is understood that an organization's ability to survive and grow depends on how it treats society. Fair corporate behavior increases the likelihood that societies will evaluate business entities favorably (Husillos et al., 2011). The CSR principles can indeed address both internal and external factors, recognizing that businesses impact and are impacted by various groups. Internally, CSR principles can focus on the well-being of employees by promoting activities to improve the well-being of workers. Externally, CSR may focus on charity work, improving the quality of life, and providing financial support to the community, while also engaging in activities designed to strengthen relationships with stakeholders (Singh and Misra 2021; Freeman 2010).

3. Hypothesis development

3.1. Green HRM and financial performance growth

Notwithstanding the fact that Green HRM is a sustainable strategy that focuses on the achievement of environmental goals, its potential to deliver economic benefits for the organization has been underlined within the framework of sustainable performance (Nguyen and Nguyen 2024). As suggested in the existing literature, companies adopt green practices primarily in response to external pressure (Aust et al., 2020; Shen and Benson 2016), but with a focus on economic benefits (Lu et al., 2023; Mehta and Chugan 2015; Kramar 2014; Margaretha and Saragih 2013). This assertion reflects the growing recognition of sustainability as not just a regulatory concern but also a *strategic opportunity.*

The existing literature *underlines several organizational-level outcomes* of green HRM (Yong et al., 2020). For instance, adopting green HRM is perceived as a strategic tool for gaining competitive advantage and fostering sustainability (Carmona-Moreno et al., 2012). In the context of environmental concerns, green HRM positively influences environmental performance (Obeidat et al., 2020; Yusoff et al., 2020; Arda et al., 2019; Rawashdeh 2018). Expanding on the triple bottom line perspective of sustainability, green HRM affects environmental, social and financial performance (Nguyen and Nguyen 2024; Al-Abbadi et al., 2023; Shahzad et al., 2023; Montalvo-Falcón et al., 2023). Interestingly, only a limited number of studies have explored how green HRM positively impacts specific *financial performanc*e (Luzzini et al., 2014). For instance, small and medium enterprises are encouraged to adopt green HRM practices due to their potential of generating superior financial results (O'Donohue and Torugsa 2016). Green HRM, such as green training and development, along with green performance management, has a significant impact on perceived financial performance (Harb and Ahmed, 2019).

Although the literature on the relationship between green HRM and financial performance is limited, valuable insights have been revealed by previous studies regarding how HRM practices impact financial outcomes. For instance, the implementation of HRM in SMEs is positively associated with *profitability growth* (Nguyen and Bryant 2004). By properly managing the human resource asset, organizations can enhance labor *productivity*, which in turn can result in a quicker and more efficient execution of business processes (Sels et al., 2016; Koch and McGrath, 1996). Human resource management practices typically yield significant *financial returns* (Flamholtz 1999). This statement is reinforced by Triguero-Sánchez et al. (2013), who underlined the positive effects on both subjective and objective financial performance (such as *ROA*) that were found in organizations that implemented HRM practices (Cochran and Wood, 1984). Furthermore, building on the HRM's potential to deliver benefits when aligned with specific objectives (such as innovation performance), Oke et al. (2012) argued that HRM can play a key role in driving *revenue growth.*

Despite the above-listed contribution to this relationship, we identify two main shortcomings. *First*, there is an *inconsistency* in the literature regarding the contribution of green HRM to financial performance. While numerous studies highlight the positive impact of green HRM on financial performance (Nguyen and Nguyen 2024; Al-Abbadi et al., 2023; Longoni et al., 2018), other research suggests that investing in green practices may lead to additional costs and further compromise the financial outcomes

(Acquah et al., 2021; Aguilera-Caracuel and Ortiz-de-Mandojana 2013; Triguero-Sánchez et al., 2013). *Second*, none of these papers analyzed the link between green HRM and the *dynamic view of performance.* Building on the triple bottom line principles, organizations must develop strategies to ensure long-term success regarding economic, social and environmental aspects (Elkington 1998). Thus, monitoring the evolution of financial performance can be crucial in achieving sustainable performance because it enables companies to monitor the potential of their sustainability efforts in yielding financial benefits.

This article employs the resource-based view (RBV) theory (Barney 1991). This management paradigm places a strong emphasis on an organization's internal resources and competencies as a way to improve long-term performance and obtain a competitive edge (Lubis 2022). Therefore, building on the above arguments, we suggest that green HRM practices have the potential to drive positive financial outcomes. Consequently, we propose:

H1a: Green HRM has a positive influence on profit growth.
H1b: Green HRM has a positive influence on revenue growth.
H1c: Green HRM has a positive influence on ROE growth.
H1d: Green HRM has a positive influence on. ROA growth.

3.2 CSR and financial performance growth

The link between CSR and financial performance has been widely discussed in the literature. However, the findings are fragmented. On one hand, many studies agreed with the *positive relation* between socially responsible activities and financial performance (Orlitzky et al., 2003). Businesses that participate in CSR initiatives are more likely to gain financial benefits compared to those that do not (Pava and Krausz 1996). For instance, adopting CSR principles may help organizations minimize risk, improve reputation and organizational identification, reduce costs, strengthen supplier networks, increase quality and productivity, enhance job performance and strengthen customer loyalty (OECD 2011). Indeed, CSR principles can contribute to organizational success through cost reduction, increased sales, lower expenses for capital and increased competitiveness (Bernal-Conesa et al., 2017). *On the other hand*, there are academics that claim a *negative relation* between social responsibility and economic performance. These studies suggest that organizations engaging in socially responsible activities may face disadvantages and incur additional costs compared to those that prioritize profit over social responsibility (Chowdhury et al., 2017; Barnett and Salomon 2006). For example, efforts to prevent pollution, improve employees' quality of life, and engage in fundraising or sponsorship for the community reflect socially responsible behavior, though they come with additional costs (Alexander and Buchholz 1978). Friedman (2007) underlines the idea that generating profit is the primary responsibility of any organization. He argues that there will always be tension between top management and shareholders, as socially responsible activities require resources and funds that could otherwise be allocated to shareholders.

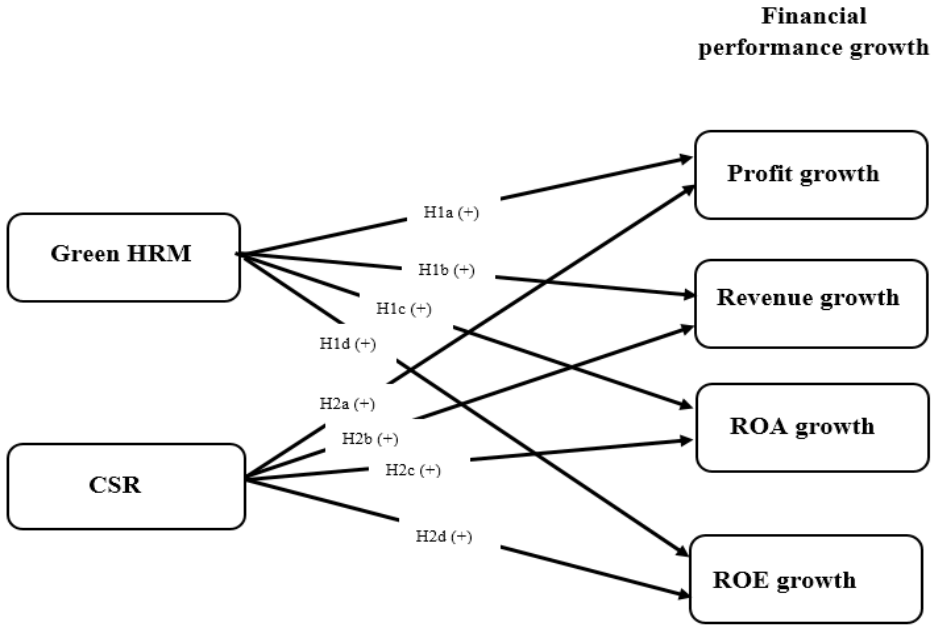
Due to the complex nature of organizational performance, previous research has considered CSR in relation to financial and non-financial performance (Liang et al., 2022; Sánchez and Benito-Hernández 2015). In terms of accounting-based measured, the literature argued that between CSR and profit is a positive relation. Companies usually incorporate CSR practices when they experience favorable economic outcomes

(Wang 2014). Higher levels of CSR implementation are associated with greater economic *profitability* (Hategan et al., 2018). Furthermore, CSR can result in substantial profitability when combined with growth strategies (Maury 2022). Organizations that invest in responsible practices will experience an increase in return on assets *(ROA*) and return on equity *(ROE)*. Therefore, CSR is an important element that directly influences whether a business utilizes all of its resources (Maury 2022; Lin et al., 2021; Cho et al., 2019; Hidayat 2017; Lee and Jung 2016; Raza et al., 2012). A case study analysis revealed a positive correlation between CSR and revenue, with Apple and Nike reporting higher yearly *revenues* following the implementation of CSR (Regonda et al., 2020). In certain situations, business charitable contributions can serve as an effective strategy for boosting revenue growth (Lev et al., 2010).

Despite the inconsistency revealed in previous research, Margolis et al. (2007) and Orlitzky et al. (2003), through their meta-analysis, claimed that a positive relationship between CSR and financial performance is more prevalent. On the basis of previous arguments, we propose the following hypothesis:

H2a: CSR has a positive influence on profit growth.
H2b: CSR has a positive influence on revenue growth.
H2c: CSR has a positive influence on ROA growth.
H2d: CSR has a positive influence on ROE growth.

Figure 1. Research framework



4. Research methodology

This study aims to present a model that links green HRM practices and CSR to the growth of financial performance. The research model is presented in *Figure 1.*

4.1. Research sample and procedure

Adopting a quantitative approach, this study conducted an anonymous online survey. Participants were assured that their responses would be used exclusively for academic research and that their involvement is completely voluntary. Furthermore, the survey guaranteed the confidentiality of respondents' personal information, ensuring it would be managed with the utmost care and discretion. Data from 430 employees is included in the questionnaire.

4.2. Measures - Research instrument

The items in the questionnaire were chosen based on their established validity and reliability, as demonstrated in previous studies, ensuring they accurately measure the relevant constructs. The selected variables were assessed using a seven-point Likert scale.

Green human resource management. This was measured by using a twelve-item scale adopted from Siyambalapitiya et al. (2018) to evaluate employees' perceptions of the green HRM practices implemented within their organization. An example of the survey item was, "In my organization, employees receive financial and non-financial benefits depending on their involvement in environmental protection".

Corporate social responsibility. This was measured by using a three-item scale adopted from Shin et al. (2016) to evaluate workers perception of the CSR practices implemented within their organization. An example of the survey item was, "My organization is concerned with improving the well-being of society".

Firm financial growth. This is our dependent variable, a multidimensional construct consisting of four indicators that capture the dynamic perspective of performance through growth-related metrics (Lee and Tsang 2001). Measures commonly used and acknowledged in the academic community were used to evaluate this construct (Acquah et al., 2021; Triguero-Sánchez et al., 2013). As a result, ROE, ROA, profit, and revenue growth were employed. We used self-reported data from the "listafirme.ro" website, considering the reported figures for profit and revenue. For ROA and ROE, we calculated the two indicators based on the available data: ROA, defined as net income divided by total assets, and ROE, defined as net income divided by shareholders' equity. To capture the dynamic perspective of performance, we calculated the growth for each indicator by dividing data from t1 (2024) to data from t0 (2023).

Control variables. To minimize bias and enhance model accuracy, we included the logarithmic average number of employees and the type of organization (measured as dummy variable, national/multinational) as control variables.

5. Research findings

In this article, we employed ordinary least squares (OLS) regression analysis to examine the relationship between the dependent (firm financial growth measured as profit, revenue, ROA and ROE growth) and independent variables (green HRM and CSR). Four models are presented in *Table 1*, combining the independent variables to predict the dependent variable. Based on the regression equations, the results show that green HRM and CSR significantly contribute to explaining the variance in one or more of the firm's financial performance indicators. A detailed analysis of each of the four models corresponding to the four dependent variables is provided below.

Table 1. Results of ordinary least squares (OLS) regression analysis

Dependent variable	**Model 1** Profit growth		**Model 2** Revenue growth		**Model 3** ROA growth		**Model 4** ROE growth	
	β	t	β	t	β	t	β	t
Green HRM	0.170**	3.116	-0.004	-0.065	0.171**	3.129	0.191***	3.496
CSR	-0.130*	-2.374	0.038	0.697	-0.128**	-2.350	-0.139*	-2.539
Type of organization	-0.077	-1.443	-0.037	-0.704	-0.074	-1.388	-1.016	0.310
Number of employees	0.018	0.347	0.211	3.993	0.009	0.167	0.389	0.697
R^2	0.030		0.042		0.030		0.032	
Adjusted R^2	0.021		0.033		0.021		0.023	
F	3.482		4.911		3.487		3.733	

*statistically significant at .05 level
**statistically significant at .01 level
***statistically significant at .001 level

Model 1-Profit growth

The first model shows that green HRM and CSR are statistically significant predictors of profit growth, with the R^2 value 0.030, which indicates that the model explains 3% of the variance in profit growth. Specifically, green HRM has a positive and significant influence on profit growth (β = 0.170, ρ < 0.01), suggesting that as green HRM practices increase, profit growth also tends to rise. This provides support for *H1a,* which predicted a positive effect of green HRM on profit growth. On the other hand, CSR has a negative but still statistically significant influence on profit growth (β = -0.130, ρ < 0.05). This indicates that while CSR activities are associated with a decrease in profit growth, the relationship is statistically robust, with a p-value below the 0.05 threshold. This does not provide support for *H2a,* showing that CSR negatively influences profit growth.

Model 2- Revenue growth

As shown in *Table 1,* the second model reveals that the effect of green HRM on revenue growth is relatively weak, with a negative relationship between the two variables. Specifically, β=-0.004 indicates that as green HRM practices increase, revenue growth tends to decrease slightly. However, given the small magnitude of the beta coefficient, the effect is minimal, suggesting that while there is a negative association, it is not substantial enough to have a significant impact on revenue growth. The *H1b*, which proposes a positive relationship between green HRM and revenue growth, is not supported. The model also indicates that CSR has a positive but statistically insignificant effect on revenue growth, as evidenced by β = 0.038. For revenue growth R^2= 0.042, which indicates that the model only explains 4.2% of the variation in revenue growth. Thus, *H2b* is supported.

Model 3- ROA growth

For the third model, R^2= 0.030. This suggests that the model explains only 3% of the variance in ROA growth. According to the results, the effect of green HRM on ROA growth is positive and statistically significant (β=0.171, ρ< .01). This means that as the involvement in green HRM practices increases, ROA growth tends to improve, so green HRM is a predictor of financial performance measured by ROA growth. Thus, according to the results, *H1c* is supported. Also, the effect of CSR on ROA growth is negative (β = -0.128) and statistically significant (ρ < 0.01). Thus, the negative beta coefficient indicates that CSR has an inverse relationship with ROA growth, meaning that higher CSR involvement might be associated with lower returns on assets. In this case, *H2c,* which proposes a positive relation between CSR and ROA growth, is not supported.

Model 4- ROE growth

As shown in the fourth model, for ROE growth R^2= 0.032, meaning that this model accounts for 3.2% of the variance in ROE growth. According to the results, green HRM has a moderate positive and statistically significant effect on ROE growth (β=0.191, ρ< .001). This indicates that an increase in green HRM practices adoption is associated with a moderate increase in ROE growth. This finding highlights the potential financial benefits of adopting green HRM practices, which could contribute to enhanced profitability and efficiency for the organization. On the other hand, the results indicate that the effect of CSR on ROA growth is negative (β = -0.139) and statistically significant (ρ < 0.05). This suggests that as CSR activities increase, ROA growth tends to decrease. The β = -0.139 indicates an inverse relationship between CSR and ROA growth, meaning that companies with higher CSR involvement may experience slower growth in their return on assets. While the relationship is statistically significant (ρ < 0.05), the magnitude of the effect suggests that CSR has a moderate negative impact on ROA growth. Therefore, *H1d* is supported, indicating that green HRM positively influences ROE growth. However, *H2d* is not supported, as the results show that CSR has a negative impact on ROE growth.

6. Discussion

The findings of this study are in line with previous research that supports the *positive connection between green HRM and improved financial performance***.** Specifically, the results of this study align with the research of Nguyen and Bryant (2004), which suggests that the implementation of HRM practices in SMEs is positively linked to profitability growth. By effectively managing human resources, organizations can boost labor productivity, leading to a faster and more efficient execution of business processes (Sels et al., 2016; Koch and McGrath 1996). Green HRM is seen as a strategy that businesses can use to increase their long-term profitability (Chowdhury et al., 2017), as reduced expenses for material purchases, energy use, waste management and environmental accident costs are all included in the financial improvement (Younis et al., 2016). However, the finding that green HRM positively influences profit growth while negatively impacting revenue growth is an interesting result. A plausible explanation could be that although increasing efficiency through lowering utility expenses could lead to higher profitability in the long run (Sheikh et al., 2019), firms adopting green HRM practices may not necessarily experience an increase in sales. This is similar to the findings of Acquah et al. (2021) who suggested that the adoption of green HRM

has a limited effect on market performance, as measured by sales. Further, this study is consistent with previous research that proposes that HRM positively influences financial turnover (Flamholtz 1999). This means that positive and significant effects on ROA and ROE are found for businesses that adopt HRM practices (Triguero-Sánchez et al., 2013; Cochran and Wood 1984).

Interestingly, the results of this study revealed that there is *a negative connection between CSR and improved financial performance*. Contrary to the claims of Margolis et al. (2007) and Orlitzky et al. (2003), who argued that a positive relationship between CSR and financial performance is more commonly observed in the literature, the results of this study align with those suggesting that CSR does not contribute to the economic performance growth. That means that organizations involved in socially responsible activities may face disadvantages and have higher costs compared to those that prioritize profit over social responsibility (Chowdhury et al., 2017; Barnett and Salomon, 2006). More specifically, the efforts to prevent pollution, to improve employees' quality of life and to engage in fundraising or sponsorship for the community (which are socially responsible practices) come with additional costs (Alexander and Buchholz 1978). An interesting perspective is presented in the paper by Chowdhury et al. (2017), who argued that CSR practices can lead to long-term success when organizations adopt green HRM as a strategy to gain social acceptance. However, this statement requires further research for validation.

This paper generates several *theoretical contributions* to the existing literature. First, by addressing the outstanding issue of why businesses adopt green HRM practices (Jackson et al., 2011), this paper offers a more nuanced understanding of the consequences of this concept, showing that green HRM implementation contributes to an increase in financial performance growth. Second, this study contributes to the existing literature by providing empirical evidence to support the negative correlation between CSR and firm financial growth. In terms of *practical implications*, our findings provide useful information for practitioners and policymakers. First, organizations can achieve positive financial outcomes by incorporating environmentally friendly principles into their HRM practices. The significance of this finding is evident, as many companies implement green practices to balance financial performance goals with responding to external pressures (Aust et al., 2020; Shen and Benson 2016), often engaging in green initiatives only when they anticipate measurable financial benefits. Second, this study highlights a key consideration for practitioners, emphasizing the need to effectively manage the implementation of CSR practices and closely monitor associated expenses. In the context of sustainable performance, policymakers should consider that adopting CSR practices may initially lead to a decrease in profit due to the additional costs involved (Chowdhury et al., 2017; Barnett and Salomon 2006). However, in the long run, these practices can positively impact business success (Maury 2022).

7. Conclusion

This study investigated the effect of green HRM and CSR on firm financial growth (measured through profit, revenue, ROA, and ROE growth). Using data from 430 respondents and applying the OLS regression model, the results partially support the research hypothesis. Specifically, the study provides evidence on the positive influence of green HRM practices on firm financial growth indicators. Furthermore, contrary to our proposed hypothesis, the results revealed that CSR negatively

impacts three of the firm's financial growth indicators, being positively related only to ROA growth. This study suggests more investigation to confirm these results and test additional mechanisms that might generate different results.

This paper also has several limitations that can be addressed in future research. This model was conducted using data collected from a single source (workers' perceptions) at a single point in time. Future research should consider incorporating the perspective of top management, as they are responsible for implementing strategies that support social and environmental principles (Friedman 2007). Also, employing a longitudinal research model in future studies would be beneficial, as it would allow researchers to track changes over time and observe the long-term effects of green HRM and CSR practices on financial performance growth. Second, we evaluated economic performance (profit, revenue, ROA, ROE) using information provided by the company on the "listafirme.ro" website. This dependence on self-reported data could produce unreliable results and jeopardize the validity of the study's findings. Therefore, rather than relying solely on self-reported data, future research should find alternative ways to gather relevant performance-related data. Third, due to the inconsistent findings in the literature, future research should be conducted in order to determine the circumstances under which the implementation of CSR leads to positive economic results (Maury 2022). Additionally, our model does not include any mediators or moderators. Future research should consider this limitation, as incorporating such variables could yield different results.

References

Acquah, I. S. K., Agyabeng-Mensah, Y., and Afum, E. 2021. "Examining the link among green human resource management practices, green supply chain management practices and performance." *Benchmarking: An International Journal* 28 (1): 267-290. doi:https://doi.org/10.1108/BIJ-05-2020-0205.

Al-Abbadi, L. H., and Abu Rumman, A. R. 2023. "Sustainable performance based on entrepreneurship, innovation, and green HRM in e-Business Firms." *Cogent Business and Management* 10 (1). doi:https://doi.org/10.1080/23311975.2023.2189998.

Alexander, G. J., and Buchholz, R. A. 1978. "Corporate social responsibility and stock market performance." *Academy of Management journal* 21 (3): 479-486. https://doi.org/10.5465/255728.

Anlesinya, A., and Susomrith, P. 2020. "Sustainable human resource management: a systematic review of a developing field." *Journal of Global Responsibility* 11 (3): 295-324. doi:https://doi.org/10.1108/JGR-04-2019-0038.

Arda, O. A., Bayraktar, E., and Tatoglu, E. 2019. "How do integrated quality and environmental management practices affect firm performance? Mediating roles of quality performance and environmental proactivity." *Business strategy and the environment* 28 (1): 64-78. doi:https://doi.org/10.1002/bse.2190.

Aust, I. Matthews, B. Muller-Camen, M. 2020. "Common good HRM: a paradigm shift in sustainable HRM?" *Human Resource Management Review* 30 (3): 100705.

Barnett, M. L., and Salomon, R. M. 2006. "Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance." *Strategic management journal* 27 (11): 1101-1122. https://doi.org/10.1002/smj.557.

Barney, J. 1991. “Firm Resources and Sustained Competitive Advantage.” *Journal of Management.*

Bernal-Conesa, J. A., de Nieves Nieto, C., and Briones-Peñalver, A. J. 2017. “CSR strategy in technology companies: Its influence on performance, competitiveness and sustainability.” *Corporate social responsibility and environmental management* 24 (2): 96-107. https://doi.org/10.1002/csr.1393.

Bilousko, T., Kazanovskyi, A., Mogylevskyi, R., Zasunko, S., and Rud, I. 2024. “The impact of human resource management strategies on the company profitability.” *Financial and Credit Activity: Problems of Theory and Practice* 3 (56). doi:DOI: 10.55643/fcaptp.3.56.2024.4342.

Bowen, H. R. 1953. *Social responsibilities of the businessman.* University of Iowa Press.

Carmona-Moreno, E., Céspedes-Lorente, J., and Martinez-del-Rio, J. 2012. “Environmental human resource management and competitive advantage.” *Management Research: Journal of the Iberoamerican Academy of Management* 10 (2): 125-142. doi:DOI 10.1108/1536-541211251607.

Carroll, A. B. 1979. “A three-dimensional conceptual model of corporate performance.” *Academy of management review* 4 (4): 497-505. https://doi.org/10.5465/amr.1979.4498296.

Cho, S. J., Chung, C. Y., and Young, J. 2019. “Study on the Relationship between CSR and Financial Performance.” *Sustainability* 11 (2): 343. https://doi.org/10.3390/su11020343.

Chowdhury, S. R., Sanju, N. L., and Asaduzzaman, A. K. M. 2017. “Green HRM practices as a means of promoting CSR: Suggestions for garments industry in Bangladesh.” *Global Journal of Management and Business Research: A Administration and Management* 17 (6): 29-35.

Cochran, P. L., and Wood, R. A. 1984. “Corporate social responsibility and financial performance.” *Academy of management Journal* 27 (1): 42-56. https://doi.org/10.5465/255956.

Comission, European. 2003. *Responsible entrepreneurship: a collection of good practice cases among small and medium-sized enterprises across Europe.* Belgium: Brussels: Publications Office.

Delmar, F., Davidsson, P., and Gartner, W. B. 2003. “Arriving at the high-growth firm.” *Journal of business venturing* 18 (2): 189-216. https://doi.org/10.1016/S0883-9026(02)00080-0.

Demir, R., Wennberg, K., and McKelvie, A. 2017. “The strategic management of high-growth firms: A review and theoretical conceptualization.” *Long range planning* 50 (4): 431-456. doi:https://doi.org/10.1016/j.lrp.2016.09.004.

Dyer, L., and Holder, G. W. 1998. *Toward a strategic perspective of human resource management.*

Dyer, L., and Reeves, T. 1995. “Human resource strategies and firm performance: what do we know and where do we need to go?” *International Journal of human resource management* 6 (3): 656-670. doi:https://doi.org/10.1080/09585199500000041.

Elkington, J. 1998. “Partnerships from cannibals with forks: The triple bottom line of 21st-century business.” *Environmental quality management* 8 (1): 37-51. https://doi.org/10.1002/tqem.3310080106.

Flamholtz, E. 1999. *Human resource accounting: Advances in concepts, methods, and applications.* Springer Science and Business Media.

Freeman, R. E. 2010. *Strategic management: A stakeholder approach.* Cambridge University Press.

Friedman, M. 2007. “The social responsibility of business is to increase its profit.” In *Corporate ethics and corporate governance*, 173-178. Berlin, Heidelberg: Springer berlin heidelberg. https://doi.org/10.1007/978-3-540-70818-6_14.

George, G., Howard-Grenville, J., Joshi, A., and Tihanyi, L. 2016. “Understanding and tackling societal grand challenges through management research.” *Academy of management journal* 59 (6): 1880-1895. doi:https://doi.org/10.5465/amj.2016.4007.

Hamann, P. M., and Schiemann, F. 2021. “Organizational performance as a set of four dimensions: An empirical analysis.” *Journal of Business Research* 127: 45-65. https://doi.org/10.1016/j.jbusres.2021.01.012.

Harb, T. H. A. D. N., and Ahmed, S. 2019. “Perceived financial sustainability of tourism enterprises: Do green human resource management practices really matter.” *Journal of Tourism and Hospitality Management* 7: 173-185. doi: https://doi.org/10.15640/jthm.v7n2a15.

Hategan, C. D., Sirghi, N., Curea-Pitorac, R. I., and Hategan, V. P. 2018. “Doing well or doing good: The relationship between corporate social responsibility and profit in Romanian companies.” *Sustainability* 10 (4): 1041. https://doi.org/10.3390/su10041041.

Hidayat, W. W. 2017. “The influence of size, return on equity, and leverage on the disclosure of the corporate social responsibility (CSR) in manufacturing companies.” *International Journal of Education and Research* 5 (8): 57-66.

Husillos, J., González, C. L., and Gil, M. J. Á. 2011. “The emergence of triple bottom line reporting in Spain.” *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad* 40 (150): 195-219. https://doi.org/10.1080/02102412.2011.10779701.

Ichniowski, C., Shaw, K. L., and Prennushi, G. 1995. “The effects of human resource management practices on productivity.” *National Bureau of Economic Research (Working paper).*

Jabbour, C. J. C., de Sousa Jabbour, A. B. L., Govindan, K., Teixeira, A. A., and de Souza Freitas, W. R. 2013. “Environmental management and operational performance in automotive companies in Brazil: the role of human resource management and lean manufacturing.” *Journal of Cleaner Production* 47: 129-140. https://doi.org/10.1016/j.jclepro.2012.07.010.

Jackson, S. E., Renwick, D. W., Jabbour, C. J., and Muller-Camen, M. 2011. “State-of-the-art and future directions for green human resource management: Introduction to the special issue.” *German Journal of Human Resource Management* 25 (2): 99-116. doi:https://doi.org/10.1177/2397002211025002.

Jones, T. M., and Wicks, N. C. 2018. *Convergent stakeholder theory.* Vol. Volumes I and II, in *Business Ethics and Strategy*, 361-376. 361-376.

Katou, A. A. 2008. “Measuring the impact of HRM on organizational performance.” *Journal of industrial engineering and management* 1 (2): 119-142. https://doi.org/10.3926/jiem.v1n2.p119-142.

Koch, M. J., and McGrath, R. G. 1996. “Improving labor productivity: Human resource management policies do matter.” *Strategic management journal* 17 (5): 335-354. https://doi.org/10.1002/(SICI)1097-0266(199605)17:5<335::AID-SMJ814>3.0.CO;2-R.

Koufteros, X., Verghese, A. J., and Lucianetti, L. 2014. “The effect of performance measurement systems on firm performance: A cross-sectional and a longitudinal study.” *Journal of operations Management* 32 (6): 313-336.

Kramar, R. 2014. “Beyond strategic human resource management: is sustainable human resource management the next approach?” *he international journal of human resource management* 25 (8): 1069-1089.

Latif, K. F., A., Sajjad, R., Bashir, M. B., Shaukat, M. B., Khan, and U. F. and Sahibzada. 2020. “Revisiting the relationship between corporate social responsibility and organizational performance: The mediating role of team outcomes.” *Corporate Social Responsibility and Environmental Management* 27 (4): 1630-1641. doi: https://doi.org/10.1002/csr.1911.

Lee, D. Y., and Tsang, E. W. 2001. “The effects of entrepreneurial personality, background and network activities on venture growth.” *Journal of management studies* 38 (4): 583-602. https://doi.org/10.1111/1467-6486.00250.

Lee, S., and Jung, H. 2016. “The effects of corporate social responsibility on profitability: The moderating roles of differentiation and outside investment.” *Management Decision* 54 (6): 1383-1406. https://doi.org/10.1108/MD-07-2015-0268.

Leon, P. 2001. *Four pillars of financial sustainability: resources for success series.* Arlington, Virginia, USA.: The Nature Conservancy.

Lepak, D. P., Liao, H., Chung, Y., and Harden, E. E. 2006. “A conceptual review of human resource management systems in strategic human resource management research.” *Research in personnel and human resources management* 217-271. doi:https://doi.org/10.1016/S0742-7301(06)25006-0.

Lev, B., Petrovits, C., and Radhakrishnan, S. 2010. “Is doing good good for you? How corporate charitable contributions enhance revenue growth.” *Strategic management journal* 31 (2): 182-200. https://doi.org/10.1002/smj.810.

Liang, Y., Cai, C., and Huang, Y. 2022. “The effect of corporate social responsibility on productivity: Firm-level evidence from Chinese listed companies.” *Emerging Markets Finance and Trade* 58 (12): 3589-3607. https://doi.org/10.1080/1540496X.2020.1788537.

Lin, C. W., Tan, W. P., Lee, S. S., and Mao, T. Y. 2021. “Is the improvement of CSR helpful in business performance? Discussion of the interference effects of financial indicators from a financial perspective.” *Complexity* 2021 (1). https://doi.org/10.1155/2021/4610097.

Lu, Y. Zhang, M. M. Yang, M. M. and Wang, Y. 2023. “Sustainable human resource management practices, employee resilience, and employee outcomes: Toward common good values.” *Human Resource Management* 62 (3): 331-353.

Lubis, N. W. 2022. “Resource based view (RBV) in improving company strategic capacity.” *Research Horizon* 2 (6): 587-596. https://doi.org/10.54518/rh.2.6.2022.587-596.

Luzzini, D., Longoni, A., and Guerci, M. 2014. “Green HRM and SCM practices and their effects on environmental and economic performance.” *Academy of Management Proceedings.* Academy of Management. 13491.

Maignan, I., and Ferrell, O. C. 2000. “Measuring corporate citizenship in two countries: The case of the United States and France.” *Journal of business ethics* 23: 283-297. https://doi.org/10.1023/A:1006262325211.

Margaretha, M., and Saragih, S. 2013. “Developing new corporate culture through green human resource practice.” Bangkok - Thailand: International Conference on Business, Economics, and Accounting. 20-23.

Margolis, J. D., Elfenbein, H. A., and Walsh, J. P. 2007. " Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance." *Ann Arbor* 1001 (48109-1234): 1-68.

Maury, B. 2022. "Strategic CSR and firm performance: The role of prospector and growth strategies." *Journal of Economics and Business* 118. https://doi.org/10.1016/j.jeconbus.2021.106031.

Mehta, K., and Chugan, P. K. 2015. "Green HRM in pursuit of environmentally sustainable business." *Universal Journal of Industrial and Business Management*, 74-81. doi:DOI: 10.13189/ujibm.2015.030302.

Montalvo-Falcón, J. V., Sánchez-García, E., Marco-Lajara, B., and Martínez-Falcó, J. 2023. "Green human resource management and economic, social and environmental performance: evidence from the Spanish wine industry." *Heliyon* 9 (10). doi:doi: 10.1016/j.heliyon.2023.e20826.

Muisyo, P. K., Qin, S., Ho, T. H., and Julius, M. M. 2022. "The effect of green HRM practices on green competitive advantage of manufacturing firms." *Journal of Manufacturing Technology Management* 33 (1): 22-40. https://doi.org/10.1108/JMTM-10-2020-0388.

Nguyen, L. Q., and Nguyen, T. N. 2024. "Exploring the relationship of green HRM practices with sustainable performance: the mediating effect of green innovation." *Journal of Hospitality and Tourism Insights.* doi:https://doi.org/10.1108/JHTI-06-2024-0610.

Nguyen, T. V., and Bryant, S. E. 2004. "A study of the formality of human resource management practices in small and medium-size enterprises in Vietnam." *International small business journal* 22 (6): 595-618. https://doi.org/10.1177/0266242604047412.

Obeidat, S. M., Abdalla, S., and Al Bakri, A. A. K. 2023. "Integrating green human resource management and circular economy to enhance sustainable performance: an empirical study from the Qatari service sector." *Employee Relations: The International Journal* 45 (2): 535-563. https://doi.org/10.1108/ER-01-2022-0041.

Obeidat, S. M., Al Bakri, A. A., and Elbanna, S. 2020. "Leveraging "green" human resource practices to enable environmental and organizational performance: Evidence from the Qatari oil and gas industry." *Journal of business ethics* 164: 371-388. https://doi.org/10.1007/s10551-018-4075-z.

O'Donohue, W., and N. Torugsa. 2016. "The moderating effect of 'Green'HRM on the association between proactive environmental management and financial performance in small firms." *The international journal of human resource management* 27 (2): 239-261.

OECD, (Organisation of Economic Cooperation and Development). 2011. *OECD Guidelines for Multinational Enterprises.* Paris: OECD. https://doi.org/10.1787/9789264115415-en.

Oke, A., Walumbwa, F. O., and Myers, A. 2012. "Innovation strategy, human resource policy, and firms' revenue growth: The roles of environmental uncertainty and innovation performance." *Decision sciences* 43 (2): 273-302. https://doi.org/10.1111/j.1540-5915.2011.00350.x.

Orlitzky, M., Schmidt, F. L., and Rynes, S. L. 2003. "Corporate social and financial performance: A meta-analysis." *Organization studies* 24 (3): 403-441. https://doi.org/10.1177/0170840603024003910.

Pava, M. L., and Krausz, J. 1996. "The association between corporate social-responsibility and financial performance: The paradox of social cost." *Journal of business Ethics* 15: 321-357. doi:https://doi.org/10.1007/BF00382958.

Pham, N. T., Tučková, Z., and Jabbour, C. J. C. 2019. “Greening the hospitality industry: How do green human resource management practices influence organizational citizenship behavior in hotels? A mixed-methods study.” *Tourism management* 72: 386-399. https://doi.org/10.1016/j.tourman.2018.12.008.

Rawashdeh, A. 2018. “The impact of green human resource management on organizational environmental performance in Jordanian health service organizations.” *Management Science Letters* 8 (10): 1049-1058.

Raza, A., Ilyas, M. I., Rauf, R., and Qamar, R. 2012. “Relationship between corporate social responsibility (CSR) and corporate financial performance (CFP): Literature review approach.” *Elixir Financial Management* 46 (9): 8404-8409.

Regonda, S., Nguyen, L., Khan, F., and Amine, S. 2020. “The Correlation Between CSR and Revenue.” *The Canadian Science Fair Journal* 3 (3).

Ren, S., Jiang, K., and Tang, G. 2022. “Leveraging green HRM for firm performance: The joint effects of CEO environmental belief and external pollution severity and the mediating role of employee environmental commitment.” *Human Resource Management* 61 (1): 75-90. doi:https://doi.org/10.1002/hrm.22079.

Renwick, D. W., Jabbour, C. J., Muller-Camen, M., Redman, T., and Wilkinson, A. 2016. “Contemporary developments in Green (environmental) HRM scholarship.” *The International Journal of Human Resource Management* (27) 2: 114-128.

Salimath, M. S., and J. B., and Umesh, U. N. Cullen. 2008. “Outsourcing and performance in entrepreneurial firms: contingent relationships with entrepreneurial configurations.” *Decision Sciences* 39 (3): 359-381.

Sánchez, P. E., and Benito-Hernández, S. 2015. “CSR policies: Effects on labour productivity in Spanish micro and small manufacturing companies.” *Journal of Business Ethics* 128: 705-724. https://doi.org/10.1007/s10551-013-1982-x.

Sels, L., De Winne, S., Delmotte, J., Maes, J., Faems, D., and Forrier, A. 2016. “Linking HRM and small business performance: An examination of the impact of HRM intensity on the productivity and financial performance of small businesses.” *Small business economics* 26: 83-101. https://doi.org/10.1007/s11187-004-6488-6.

Shah, M. 2019. “Green human resource management: Development of a valid measurement scale.” *Business Strategy and the Environment* 28 (5): 771-785. doi:https://doi.org/10.1002/bse.2279.

Shahzad, M. A., Jianguo, D., and Junaid, M. 2023. “Impact of green HRM practices on sustainable performance: mediating role of green innovation, green culture, and green employees’ behavior.” *Environmental Science and Pollution Research* 30 (38): 88524-88547. doi:https://doi.org/10.1007/s11356-023-28498-6.

Sheikh, Wahidul, M. Shahedul Islam, and Farhana Rahman. 2019. “Implementing Green Human Resource Management: Cost-Effective Strategies and Tools.” *Journal of Entrepreneurship Organization Management* 8 (1): 1-8.

Shen, J., and Benson, J. 2016. “When CSR is a social norm: How socially responsible human resource management affects employee work behavior.” *Journal of management* 42 (6): 1723-1746. https://doi.org/10.1177/0149206314522300.

Shen, J., Dumont, J., and Deng, X. 2018. “Retracted: Employees’ perceptions of green HRM and non-green employee work outcomes: The social identity and stakeholder perspectives.” *Group and Organization Management* 43 (4): 594-622. https://doi.org/10.1177/1059601116664610.

Shin, I., Hur, W. M., and Kang, S. 2016. “Employees’ perceptions of corporate social responsibility and job performance: A sequential mediation model.” *Sustainability* 8 (5): 493. https://doi.org/10.3390/su8050493.

Singh, K., and Misra, M. 2021. “Linking corporate social responsibility (CSR) and organizational performance: The moderating effect of corporate reputation.” *European Research on Management and Business Economics* 27 (1). doi:https://doi.org/10.1016/j.iedeen.2020.100139.

Siyambalapitiya, J., Zhang, X., and Liu, X. 2018. “Green human resource management: A proposed model in the context of Sri Lanka’s tourism industry.” *Journal of cleaner production* 201: 542-555. doi:https://doi.org/10.1016/j.jclepro.2018.07.305.

Stanaland, A. J., Lwin, M. O., and Murphy, P. E. 2011. “Consumer perceptions of the antecedents and consequences of corporate social responsibility.” *Journal of business ethics* 102: 47-55. https://doi.org/10.1007/s10551-011-0904-z.

Syed, Z. J. W., and Jamal, W. 2012. “Universalistic perspective of HRM and organizational performance: meta-analytical study.” *International Bulletin of Business Administration* 13 (4): 45-57.

Tang, G. Chen, Y. Jiang, Y. Paillé, P. and Jia, J. 2018. “Green human resource management practices: scale development and validity.” *Asia pacific journal of human resources* 56 (1): 31-55.

Triguero-Sánchez, R., C. Peña-Vinces, J., and Sánchez-Apellániz, M. 2013. “Hierarchical distance as a moderator of HRM practices on organizational performance.” *International Journal of Manpower* 34 (7): 794-812. doi:https://doi.org/10.1108/IJM-03-2012-0046.

Vilanova, M., Lozano, J. M., and Arenas, D. 2009. “Exploring the nature of the relationship between CSR and competitiveness.” *Journal of business Ethics* 87: 57-69. https://doi.org/10.1007/s10551-008-9812-2.

Wang, S. 2014. “On the Relationship between CSR and Profit.” *Journal of International Business Ethics* 7 (1): 51-57.

Yang, C. C., and Lin, C. Y. Y. 2009. “Does intellectual capital mediate the relationship between HRM and organizational performance? Perspective of a healthcare industry in Taiwan.” *The International Journal of Human Resource Management* 20 (9): 1965-1984. https://doi.org/10.1080/09585190903142415.

Yong, J. Y., Yusliza, M. Y., and Fawehinmi, O. O. 2020. “Green human resource management: A systematic literature review from 2007 to 2019.” *Benchmarking: An International Journal* 2005-2027. https://doi.org/10.1108/BIJ-12-2018-0438.

Younis, H., Sundarakani, B., and Vel, P. 2016. “The impact of implementing green supply chain management practices on corporate performance.” *Competitiveness Review* 26 (3): 216-245. https://doi.org/10.1108/CR-04-2015-0024.

Yusoff, Y. M., Nejati, M., Kee, D. M. H., and Amran, A. 2020. “Linking green human resource management practices to environmental performance in hotel industry.” *Global Business Review* 21 (3): 663-680. doi:https://doi.org/10.1177/0972150918779294.

Zaid, A. A., Jaaron, A. A., and Bon, A. T. 2018. “The impact of green human resource management and green supply chain management practices on sustainable performance: An empirical study.” *Journal of cleaner production* 204: 965-979. doi:https://doi.org/10.1016/j.jclepro.2018.09.062.

Zhang, Z., Wan, D., and Jia, M. 2008. “Do high-performance human resource practices help corporate entrepreneurship? The mediating role of organizational citizenship behavior.” *The Journal of High Technology Management Research* 19 (2): 128-138.

STUDIA UNIVERSITATIS BABEȘ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 51-66
DOI: 10.2478/subboec-2025-0003

DETERMINANTS OF RENT PRICES: AN ANALYSIS OF EUROPEAN CITIES

Anuța BUIGA*
Babeș-Bolyai University, Romania

Mirela POPA
Babeș-Bolyai University, Romania

Abstract: Using the data of a sample consisting of 99 cities from 37 European countries (2024) this paper discusses the determinants of rent prices. The results indicate that rents are lower in areas with longer commute times and higher cost of living index, while high wages and high house price to income ratio (HPIR) lead to increase rent prices. In this paper, we constructed the Rent Index as average of four variables relating to the rents of one-room and three-room apartments in the city centre and in the suburbs. We identify the factors that explain the variation of the Rent Index using multiple linear regression. The overall conclusions of analysed models are mixed: salary is the strongest predictor in all models, if the city is a university one, we expect an average rent increase between 105.9 and 115.81 euro compared to non-university cities. If the city has a tourist attraction, the rent price will increase on average with a sum between 114.59 and 158.61 euro; the safety index and pollution index have significant effects on Rent Index; an increase of HPIR by one unit will lead to an increase in the Rent Index with 13.6 euros per month.

JEL classification: C13, C51, D60, E31, I31, R21, R31

Keywords: rent price; European cities; quality of life; housing supply, housing supply demand; Rent Index

1. Introduction

The literature shows that the differences between rich and poor people are becoming deeper (OECD, 2020), regardless of the major efforts (aid, compensation, exemptions, etc.) implemented by various governments. The World Bank Annual

* Corresponding author. Address: Babeș-Bolyai University, Faculty of Economics and Business Administration, Department of Statistics Forecasts Mathematics, RO-400083, 68, Avram Iancu St., Cluj-Napoca, Romania, E-mail: anuta.buiga@econ.ubbcluj.ro

Report (2024, p.66) estimates that 314 million extremely poor people live in 39 countries classified as fragile or conflict affected. “Wealth inequality between countries across Europe is very strong. In the EU, the difference in wealth per adult between the countries with the highest and lowest levels exceeds a factor of ten” (Yanatma, 2024). Therefore, more and more people are forced to rent a house.

The relevance of this study is supported by recent statistics from numerous countries, which show that rental prices have increased in recent decades due to several factors. Renting has become a huge burden especially for the poor families that cannot afford to buy their own home, ending up living in indecent conditions.

Our study aims to identify various economic and social factors that influence rent prices in the most important European countries and which, therefore, affect the living conditions, especially of the poorest Europeans.

In Section 2 of the article (regarding the literature review), we include a literature review highlighting the following aspects: the housing supply, regulations of the private rental sector affecting the rent fluctuations and quality of life, the housing demand and the trends of rent prices.

We also reiterate the results of recent studies that highlight the possible causes of rent increases but also various negative aspects identified in many countries, affecting tenants’ quality of life and health; for example, rigidity of housing markets, unsafe and unhealthy housing, poor housing quality (including overcrowding, damp and insufficient housing facilities), precarious housing situations (such as unaffordability, evictions and displacement), the large number of minors living in overcrowded housing, additional costs related to additional monthly expenses and housing-costs-induced energy poverty, air pollution, etc.

In addition, in the 2nd Section we present the results of recent studies that highlight the possible causes of rent increases as well as various negative aspects identified in many countries, affecting tenants’ quality of life and health: for example, rigidity of housing markets, unsafe and unhealthy housing, poor housing quality (including overcrowding, damp and insufficient housing facilities), precarious housing situations (such as unaffordability, evictions and displacement), the large number of minors living in overcrowded housing, additional costs related to additional monthly expenses and housing-costs-induced energy poverty, air pollution, etc.

The data used in our study were collected from a sample of 99 cities from 37 European countries in relation to a series of variables that capture the quality of life. We constructed the Rent Index variable as an average of the four variables.

Thus, this present paper is the first in jointly exploring the effects of six explanatory variables: Salary, Pollution Index, Safety Index, House Price to Income Ratio (HPIR), Traffic Commute Time Index and Cost of Living Index. Other analysed predictor variables refer to three characteristics of European cities (i.e., university cities and tourist cities (i.e., university cities, tourist cities and capital or not). Obviously, the dependent variable is monthly rent price, more precisely Rent Index constructed as an average of the four variables.

Further, this paper is organized as follows: the data, the description of explanatory variables and the 6 research hypotheses are presented in the 3rd and 4th sections, then the empirical results of the models applied are analysed and discussed in the 5th section. Finally, short sections with discussion and conclusions, are provided.

2. Literature review

In the specialized literature, to explain the many reasons/causes/factors that generate very high rent prices in some cities, effects of housing supply and housing demand have been generally studied.

The housing supply can be influenced by raising house prices, by over-regulated housing construction and rentals (legal rules on ownership and tenancy, fiscal policies, rent control, planning and building regulations, cost burden of energy retrofitting etc.), monetized power relations and housing insecurity, housing production more expensive, mortgage availability, the provision of social housing through the local authority system, homeownerships, etc.

"Monetized power relations involving landlords (public and private), creditors and multi-scalar state interventions (federal, municipal and district) play an integral, yet paradoxical, role in low-income rental housing insecurity" (Soederberg, 2018, p.114).

Even between EU countries, many differences can be identified regarding *regulations and rent control affecting the rental sector.*

For example, in Sweden the rent prices are not subject to market mechanisms but are rather strictly regulated, which entailed a smaller rent increase in the last decades due to inflation and rising prices of landlords' expenditures for waste management, energy, and water. "Although renovation costs can lead to rent increases in energy retrofitting, it is often assumed that reductions in energy costs will counterbalance the rent increase" (von Platten et al., 2022, p.1, 3).

Several studies found, for example, that the regulations had a negative impact on rental inflation, that rent control reduced prices, or removal of rent control had little effect on the level of rents, etc. (O'Toole et al., 2021, p.2)

Other studies have highlighted comparative analyses of national regulations of the private rental sector in EU countries. It seems that (a) if there are very strict regulations ('over-regulation') of rents (e.g. Denmark, the Netherlands, France) there are too high risks for private investors and, therefore, the private rental sector cannot be stable and effective; (b) if there are countries with minimal regulations of the private rental sector ('under-regulation'), e.g. Finland and the UK, the share of private rented housing increases, but the private rental sector cannot be stable in the long term (Lux and Sunega, 2010).

From our point of view, it is very important to avoid both 'over-regulation' and 'under-regulation' in the rental sector of the housing market.

An example of 'over-regulation' is the rent price freeze (setting a cap on rental prices), part of a rigorous rent-control policy in Berlin, introduced in 2020 and revoked 13 months later by the German Constitutional Court (Hahn et al., 2022, p.7).

Early' s study (2000, p. 185) estimates the net benefits of rent control to tenants in New York City and suggests that, by controlling for higher prices in the uncontrolled sector, the average benefit to tenants in regulated units is negative.

From our point of view, in each EU country the economic, social, political, cultural particularities, etc. can be analysed, and a middle ground specific to each context can be found, meant to implement a moderate system of regulation of rents, the duration of rental contracts, a balanced protection of tenants and landlords, etc.

In order to improve the standards in the rental sector and to provide a redress system on a par with that part enjoyed by tenants of social landlords, in UK

part of the existing regulations (for example, the Decent Homes Standard and 'Awaab's Law') will also be extended to the private rented sector and a private rented sector database will be created that will be used to provide information for landlords and tenants (Stephens et al., 2024, pp.9).

Kholodilin (2022, p.7) examined several empirical studies published between 1972-2022, studies focused on rent control, and concluded that rent control, although it helps lower rent prices for tenants, has many undesirable effects.

If we consider the state regulator for the Irish private rental sector, studies have found that "overall rent controls had a deflationary impact, with the rental inflation rate dropping by approximately 1–2 percentage points after the introduction of the regulations", but the regulations "have been effective in the areas of Ireland where rent controls were implemented" (O'Toole et al. 2021, p.3, p.14).

The housing demand can be influenced by changes in tenants' preferences, overcrowded households, raising rental prices, the socioeconomic consequences of the crisis, homelessness, immigration, the invasions/war conflicts from neighbouring countries (e.g., the Russian aggression against Ukraine), migration shocks and patterns, weakened rental protections, wage changes, mortgage availability, the quality of housing, unsafe and unhealthy housing, high maintenance costs and low energy efficiency of rented homes, loss of social housing, subsidised housing, mobility, search and moving costs etc.

It is known that rent controls impact tenants' mobility and increase rent prices in the long term (Diamond et al., 2019).

Gete and Reher (2018) show that a contraction of mortgage supply and tighter lending standards (over the 2010–2014 period) after the Great Recession has increased demand for rental housing, leading to higher rents and an increase in rental supply.

The role of growing demand to live in housing-supply U.S. cities was studied by Howard and Liebersohn (2021) and they showed that rents increase in the aggregate, even if a person leaves an elastic city (where rent price will fall only a little) for an inelastic city (where rent price will rise much more). "The location demand channel explains 54% of the rent increase in all U.S. cities and 75% of the rent increase where the consumer price index rents data exists" (Howard and Liebersohn, 2021, pp.1, 13).

Spectacular *increases in rent prices* were observed in two cities in Poland, Krakow and Warsaw, where rent prices increased by 16.5%, respectively 14% in a very short period of time (March and April of 2022), mainly due to the influx of Ukrainian refugees fleeing the war, the refugee shock being caused by the Russian invasion against Ukraine in February 2022 (Trojanek and Gluszak, 2022, p. 4).

In the decade 2013-2023, on average, the share of European renters (paying rent) did not increase significantly (+5.1%). However, the largest increases, in the same period, were in countries such as Lithuania (+43.6%), Malta (+29.7%), Greece (+25.6%) and the Czech Republic (+20.6%) (Horvat and Coupechoux, coord., 2024, p.76).

Studying the issue of apartment rental in some cities in Poland during the COVID-19 pandemic, Źróbek-Różańska (2022) observed a 10-20% decrease in rental prices and noted a marked decline in interest in student housing. The author showed that only the owners of 1–2 room flats in a good location found a tenant quite quickly (in Warsaw, Cracow, Poznań).

The empirical results (after interviewing students from a university city in Poland) show that students do not remain attached to rented apartments and are not willing to pay to reserve an apartment during the COVID-19 pandemic, leading to the negotiation of lower prices and more favourable rental conditions (Źróbek-Różańska, 2022).

Saiz (2007) shows that an immigration flows equal to 1% of the population of a U.S. city is associated with an increase of approximately 1% in average rents and housing prices.

In just 9 years house prices in the EU increased significantly (by 48.5%) while rental prices in the EU increased by 14.6% in 2023 compared to 2015. With the exception of Greece where a small decrease (-2.1%) in rental prices was observed (in the 2015-2023 period), in all other EU countries increases in rental prices were observed (since 2015, the reference year), the largest increases being in Lithuania (68.2%), in Hungary (66.5%) and in Slovenia (60.4%) (Horvat and Coupechoux, ed., 2024, p.72).

European statistics show that in 2023 more renters (24.6% on the private market and 24.4% in subsidised housing) were affected by overcrowding than homeowners: 16.4% of non-mortgaged homeowners and 8.2% of mortgaged homeowners (Horvat and Coupechoux, coord., 2024, p.75).

One issue facing providers forced to increase rent prices is the levels of inflation and other financial aspects related to the tenants' social circumstances and cost of living.

To identify vulnerability factors related to housing-costs-induced energy poverty Karpinska and Śmiech (2023) focus on the profile of energy-poor households in 11 Central and Eastern Europe countries (Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia). The authors explain that renters are usually more vulnerable because tenants do not have the resources to implement thermal-modernization solutions and rigidity of housing markets make renting even less affordable than before (Karpinska and Śmiech, 2023, p.5).

Under these conditions, reviewing energy efficiency policies to reduce rent prices and maintenance costs of apartments, urgent measures and regulations are required.

Despite the existence of regulations designed to protect tenants from the risks of unsafe and unhealthy housing, the private rental sector in several countries requires more attention because in many countries countless complaints and claims from tenants have been registered and have remained unresolved.

Many studies in the specialized literature show that there is a correlation between poor housing quality (including overcrowding, damp and insufficient housing facilities) or precarious housing situations (such as unaffordability, evictions and displacement) and increased anxiety, depression, stress, insomnia and panic attacks, respectively emotional distress (Newton et al., 2022, p.1).

Other studies in the literature have also found that depression in older adults is significantly influenced by the residential environment and residential experience, but the impact may vary by country. For example, the area (rural non-rural) or regional disparities in development do not have a significant impact on the levels of depression of European residents (Kou et al., 2024, p. 6).

In the Ninth Overview of Housing Exclusion in Europe (2024), a Joint Publication by FEANTSA and the Foundation Abbé Pierre, alarming statistics are presented. For example, it was found that "one in four minors in the European Union - and over one in five children under the age of six - were living in overcrowded conditions (Horvat and Coupechoux, coord., 2024, p.41).

Table 1. Minors living in overcrowded housing (2023)

Countries		Countries		Countries	
Malta	1.9%	France	15.8%	Italy	40.9%
Cyprus	3.7%	Slovenia	15.8%	Lithuania	41.1%
Netherlands	4.6%	Germany	18.7%	Greece	41.5%
Ireland	6.3%	Portugal	21.8%	Slovakia	42.3%
Finland	9.3%	Sweden	23.3%	Poland	43.7%
Belgium	9.4%	Austria	23.9%	Croatia	44.4%
Denmark	9.7%	Estonia	25.0%	Latvia	55.6%
Luxembourg	9.9%	Czechia	26.6%	Bulgaria	56.5%
Spain	11.3%	Hungary	28.8%	Romania	61.1%

Source: EUROSTAT (from Horvat and Coupechoux, coord., 2024, p.42).

As it can be seen from Table 1, only in Malta (1.9%), Cyprus (3.7%) and the Netherlands (4.6%) were less than 5% of minors living in overcrowded housing (in 2023). At the opposite pole, over 55% of minors faced poor housing conditions in Romania (61.1%), Bulgaria (56.5%) and Latvia (55.6%).

Under these circumstances, structural changes (i.e. preventive practices and policies, the impact of the short-term rentals, housing innovations, debts advice and legal assistance for tenants, assess the vulnerabilities in the housing and rental market etc.) and regulating the housing market that affect the rental sector, including regulating short-term rentals to prevent rent price increases and urban overcrowding, are important in EU.

Other problems (especially for poor people) are vulnerability, factors related to rental-costs, additional costs related to additional monthly expenses and housing-costs-induced energy poverty, etc.

More than 20% of UK private renters have lost at least £100 income per month during the COVID-19 pandemic (Perry et al., 2021, p. 5). More than that, the Covid-19 pandemic highlighted the impact of poor housing conditions on physical and mental health (Cromarty, 2022. p.9).

"The social and material vulnerabilities associated with poor quality and precarious conditions which, underpinned by powerlessness, resulted in threats that contributed to and exacerbated the mental-ill health of renters at a time of broader structural fragility" (Newton et al., 2022, p.7). Key issues identified for improving housing fitness in the England private rented sector are reform of the Housing, Health and Safety Rating System, ineffective local authority enforcement, power imbalance between tenants and landlords, complex and piecemeal legislation and insufficient data on the private rented sector (Cromarty, 2022, pp. 27-34).

Although many national and international institutions make consistent continuous investments to reduce inequalities and reduce poverty, in 2023, on average half of European households who lived below the poverty threshold were

tenants (49.2%); the highest percentages of tenants in the poor category were in Denmark (70.0%), Austria (75.1%), and Germany (75.3%). In addition, statistics show that many Europeans cannot afford to live in decent conditions.

If we refer to EU citizens in 2023, 16% of them lived in overcrowded households, the highest percentages being in Latvia (40.9%), Romania (40.0%) and Bulgaria (34.9%), and the lowest in Cyprus (2.2%), Malta (2.4%) and the Netherlands (3.8%).

Behera et al. (2024) analysed the factors that contribute to people's happiness or life satisfaction (between 2005 – 2020, in 166 countries) and showed, among other important factors that per capita income positively impacts happiness, while air pollution exposure has a negative impact.

The literature shows that increased air pollution can cause mental health issues (Muresan et al., 2023) and continuous exposure to air pollution can reduce people's happiness or life satisfaction (Behera et al., 2024 p.10).

3. Data

The data refer to a sample of European cities analysed in relation to a series of variables that capture the quality of life. Variables Indices are based on data and use surveys collected by Numbeo. The surveys capture the perceptions and experiences of site visitors on various aspects of quality of life. Each survey entry is assigned a number in the range of -2 to +2, where -2 represents a strongly negative perception and +2 represents a strongly positive perception. Based on these responses, score variables with values between 0 and 100 are constructed.

The sample consists of 99 cities from 37 European countries. Out of these, 36.4% are capital cities, 37.4% are tourist cities and 57.5% are university cities. In terms of geographical distribution of 99 cities in European countries (from our sample), 21.2% are in the north, 21.2% are in the south, 31.3% are in the east and 26.3% are in the west. The data refer to the year 2024.

Table 2. Descriptive statistics for rent prices (euro)

Variables	*Mean*	*Median*	*Std. Deviation*	*Minimum*	*Maximum*
rent_1B_CityCentre	903.62	895.00	428.290	216.81	2033.45
rent_1B_OutsideCentre	681.73	689.32	344.709	142.48	1717.41
rent_3B_CityCentre	1577.19	1519.84	739.436	470.94	3749.64
rent_3B_OutsideCentre	1155.71	1119.57	580.721	282.48	2821.96

Source: authors' own calculations

We collected data related to the rent price for Apartment (1 bedroom) in the City Centre and the rent price for Apartment (3 bedroom) in the City Centre, respectively, the rent price for Apartment (1 bedroom) Outside of Centre and the rent price for Apartment (3 bedroom) Outside of Centre (see Table 2).

Starting from these data we constructed the Rent Index as an average of the four variables. The descriptive statistics related to the Rent Index variable are the following: mean is 1079.56 euro/month, median is 1049.39 euro/month, standard

deviation is 516.19 euro/month, Skewness = 0.41, Kurtosis = -0.26, which allows us to confirm that we have a distribution that can be assimilated to the normal one, the Lilliefors test indicating a p-value = 0.2. The lowest values are found in cities from Ukraine, Russia, Bosnia and Herzegovina, North Macedonia, and the highest values are in cities for Switzerland, United Kingdom, Ireland and Netherlands.

4. Description of explanatory variables and the research hypotheses

This paper focuses on the impact of six explanatory variables (salary, house price to income ratio, pollution index, safety index, traffic commute time index and cost of living index) on monthly rent. The descriptive statistics for these explanatory variables (predictors) are presented in Table 3.

Salary refers to Average Monthly Net Salary.

House Price to Income Ratio (HPIR) is a fundamental measure for apartment purchase affordability, where a lower ratio indicates better affordability. It is typically calculated as the ratio of median apartment prices to median family disposable income, expressed as years of income (numbeo.com).

Pollution Index provides an estimate of overall pollution levels in cities worldwide. Factors such as air and water pollution, garbage disposal, cleanliness, noise and light pollution, green spaces, and comfort in relation to pollution are considered in the calculation of the Pollution Index (numbeo.com).

The safety index measures residents' and visitors' feelings of safety while walking during the day and at night.

Traffic Commute Time Index is a composite measure that considers factors such as commute time, dissatisfaction with time spent in traffic, CO2 emissions, and overall traffic system inefficiencies. It provides insights into the overall traffic conditions in a city (numbeo.com). The lower the value, the better.

Cost of Living Index indicates the relative prices of consumer goods like groceries, restaurants, transportation, and utilities. It excludes accommodation expenses such as rent price or mortgage (numbeo.com).

Table 3. Descriptive statistics for explanatory variables (predictors)

Predictors	*Mean*	*Median*	*Std. Deviation*	*Minimum*	*Maximum*
Salary	2063.36	1706.51	1292.92	347.13	7019.34
House Price to Income Ratio (HPIR)	67.90	70.80	9.96	47.10	80.70
Pollution Index (Pollution)	63.20	64.90	11.31	35.80	79.50
Safety Index (Safety)	11.26	10.80	3.56	4.70	21.40
Traffic Commute Time Index (Traffic)	43.45	39.20	18.74	12.40	87.00
Cost of Living Index (Cost)	51.20	51.10	16.14	25.50	99.30

Source: authors' own calculations

Research hypotheses

In our study, we employ validation of the following six research hypotheses: (1) higher wages lead to higher rents, (2), there is no direct, immediate link between urban pollution and rent prices, (3) higher safety leads to higher rents, (4) a high House Price to Income Ratio (HPIR) is associated with higher rents, (5) rents are lower in areas with longer commute times, (6) there is a direct link between cost of living and rent price.

H1: Higher wages lead to higher rent prices

In general, there is a direct relationship between wages and housing rent. Higher wages allow people to pay higher rents, which often leads to higher rent prices in cities with developed economies. Conversely, low wages limit access to quality housing and influence the increase in demand for renting the smaller or cheaper housing. Hypothesis H1 is validated, meaning that there is a direct and strong relationship (0.863) between wages and rent prices (see Table 4).

H2: There is no direct, immediate link between urban pollution and rent prices

The link between pollution and rent prices is influenced by housing preferences, environmental conditions, and the costs associated with pollution levels. As environmental quality affects housing decisions, demand for housing in polluted areas may decrease and demand for housing in less polluted areas may increase, which will lead to a direct influence on rent prices. In addition, investments in green infrastructure and environmental policies can contribute to increasing rent prices in cleaner and more sustainable areas. The data from our sample (99 cities from 37 European countries) do not support a direct and significant link (0.085) between pollution and rent prices. Air pollution and urban noise can lead to a decrease in quality of life, and some residents may choose to pay less for rent to live in a cleaner area, away from the city centre, even if this means a longer commute or a compromise in terms of accessibility to amenities. In our study we demonstrate the link between Pollution Index and Traffic Commute Time Index to be quite strong (-0.505) (see Table 4).

Table 4. Pearson Correlation

	Salary	*Pollution*	*Safety*	*HPIR*	*Traffic*	*Cost*
Rent Index	0.863***	0.085	-0.342***	0.524***	-0.549***	0.856***
Salary	1	0.202**	-0.531***	0.615***	-0.644***	0.927***
Pollution		1	-0.022	0.254**	-0.505***	0.135
Safety			1	-0.349***	0.34***	-0.433***
HPIR				1	-0.638***	0.597***
Traffic					1	-0.595***

*Note: *Indicates significance at 1% level or less, **significance at 5% level, ***significance at 10% level*
Source: authors' own calculations

H3: Higher safety leads to higher rent prices

The safety of an area/city is an essential factor influencing rent prices. Safe areas are usually more attractive to residents, which leads to a higher housing demand and, implicitly, higher rents. Conversely, in less safe areas, demand decreases and rents can be lower. Thus, the Safety Index has a significant impact on the real estate market, influencing both the housing demand and the rent prices. In our study, hypothesis H3 was not validated because the link between Safety and Rent Index is inversely (-0.342) (see Table 4).

H4: A high House Price to Income Ratio (HPIR) is associated with higher rent prices

The relationship between HPIR and rent prices can be understood in the context of the housing market and housing affordability. The higher the ratio, the more difficult it is to buy a house relative to one's income. In many housing markets, the rent for an apartment is related to the selling house price. In general, rent tends to be lower than the mortgage rate for the same type of apartments, but this difference can vary depending on location and general economic conditions. If the HPIR is high (i.e. house prices are very high relative to income), people may prefer to rent because buying a house becomes unaffordable. In conclusion, if housing prices are too high relative to income, this can increase the demand for renting apartments, thus increasing rents, in parallel with increasing HPIR. Thus, our study shows that the relationship is direct and quite intense (0.524), meaning that increase HPIR will lead to an increase rent price, which is in line with economic theory, meaning that hypothesis H4 is validated (see Table 4).

H5: Rent prices are lower in areas with longer commute times

People who work in urban centres or areas where jobs are concentrated are often willing to pay higher rents to live nearby, saving time and energy without commuting. Thus, housing located close to commercial, or business centres will usually have higher rents. Housing offers that are more affordable in terms of rent price, but located further from the city centre, are attractive to those who are willing to accept a longer commute to save money. So, we find a strong inverse relationship (-0.549) between Traffic Commute Time Index and Rent Index, and Hypothesis H5 is validated (see Table 4)

H6: There is a direct link between Cost of Living and rent price

Areas with high rent prices are often areas with high population density (e.g., in urban centres), which can lead to increase prices for goods and services due to high demand. For example, prices from restaurants, public transportation, and other services may be higher in large cities or in areas with high population density, which affects the overall cost of living. On the other hand, in areas with lower rent prices (e.g., in suburbs or smaller cities), goods and different services' prices may be smaller, which can contribute to a lower cost of living, even if rent price is not directly considered. Our study confirms hypothesis H6, meaning that there is a direct and strong link (0.856) between cost of living and rent price (0.856) (see Table 4).

5. Results of the applied models

Further, we identify the factors that explain the variation of the Rent Index variable using multiple linear regression (MRA). Considering the predictor variables and the correlations between them, we created six models.

The standardized Beta coefficient in a regression model measures the relative impact of each independent variable on the dependent variable, in terms of standard deviations. This helps to compare variables that are measured in different units more easily and allows for a clearer understanding of the influence of each variable in the regression model. A higher standardized Beta coefficient indicates a more significant impact on the dependent variable.

Salary is the strongest predictor in all models, with large Beta standardized coefficients, suggesting a strong, direct and significant relationship between Salary and Rent Index (1st model).

Tourist cities and university cities also have a significant impact on Rent Index, in 2nd and 3rd models. Respectively, if the city is a university city we expect an average rent price increase between 105.9 and 115.81 euro compared to non-university cities. Similarly, if the city has a tourist attraction, the rent price will increase on average with a sum of between 114.59 and 158.61 euro. The Safety Index and Pollution Index have significant effects on Rent Index in models 2 and 3, but with smaller effects compared to Salary.

Overall, the models are statistically significant and explain a large part of the variability of Rent Index, with R^2s between 0.744 and 0.80 (see Table 5).

4th Model: House Price to Income Ratio (HPIR) is significant (p = 0.017) and positive, suggesting that an increase in this indicator by one unit will lead to an increase in the Rent Index with an average of 13.6 euro per month. Safety Index is not significant (p = 0.124), meaning that safety does not have a significant impact on Rent Index in this model. The Traffic Commute Time Index has a significant negative coefficient (p = 0.002), indicating that an increase in traffic will reduce the Rent Index by 9.23 euro per month on average (see Table 6).

Table 5. Summary of estimated linear models 1, 2 and 3 for Rent Index

Predictors variables	*Model 1 Rent Index*		*Model 2 Rent Index*		*Model 3 Rent Index*	
	Coeff. B (p value)	*Coeff. Beta*	*Coeff. B (p value)*	*Coeff. Beta*	*Coeff. B (p value)*	*Coeff. Beta*
Salary	0.344*** (0.000)	0.863	0.323*** (0.000)	0.809	0.36*** (0.000)	0.903
Safety Index	-	-	-	-	19.02** (0.027)	0.131
Pollution Index	-	-	-	-	-4.78** (0.031)	-0.105
University city or not	-	-	105.9* (0.053)	0.102	115.81** (0.029)	0.111
Tourist city or not	-	-	158.61*** (0.003)	0.149	114.59** (0.036)	0.108
Constant	369.02*** (0.000)		292.66*** (0.000)		314.48* (0.068)	
R Square	0.744		0.781		0.80	
F	281.89***		113.5***		74.43***	

*Note: *Indicates significance at 1% level or less, **significance at 5% level, ***significance at 10% level*
Source: authors' own calculations

5th Model: HPIR is significant (p = 0.000), suggesting that it has a strong effect on Rent Index. The Safety Index is significantly negative (p = 0.044), indicating that lower safety leads to a lower Rent Index. The Traffic Commute Time Index also has a significant negative coefficient, like 4th Model. It is significant (p = 0.01) if a city is a university one, indicating that university cities have a higher Rent Index, meaning that rent prices will increase on average by 230.76 euros per month. Tourist cities have a less significant effect (p = 0.065) but still indicate an increase in rent (see Table 6).

6th Model: HPIR and Traffic Commute Time Index remain significant variables (p < 0.05), and Traffic Commute Time Index has a significant negative effect on Rent Index. Safety Index is significantly negative (p = 0.016), suggesting that lower safety reduces Rent Index. Capital has a significant positive effect (p = 0.039), indicating that capital cities have a Rent Index higher by 197.5 euro on average. Tourist city is not significant in this model (p = 0.31), suggesting that it is not an important predictor of Rent Index (see Table 6).

Table 6. Summary of estimated linear models 4, 5 and 6 for Rent Index

Predictors variables	*Model 4 Rent Index*		*Model 5 Rent Index*		*Model 6 Rent Index*	
	Coeff. B (p value)	*Coeff. Beta*	*Coeff. B (p value)*	*Coeff. Beta*	*Coeff. B (p value)*	*Coeff. Beta*
HPIR	13.6 ** (0.017)	0.262	8.03 *** (0.000)	0.155	10.45* (0.062)	0.202
Safety	-19.85 (0.124)	-0.137	-26.26** (0.044)	-0.181	-31.68 ** (0.016)	-0.218
Traffic	-9.23*** (0.002)	-0.335	-8.48*** (0.004)	-0.308	-7.59 *** (0.009)	-0.275
University city or not	-	-	230.76 ** (0.01)	0.222	254.12 *** (0.004)	0.245
Tourist city or not	-	-	170.26 * (0.065)	0.160	97.40 (0.31)	0.092
Capital or not	-	-	-	-	197.45** (0.039)	0.184
Constant	780.68 (0.125)		1002.4 ** (0.04)		801.7 * (0.088)	
R Square	0.368		0.447		0.472	
F	18.47***		15.02***		13..71***	

*Note: *Indicates significance at 1% level or less, **significance at 5% level, ***significance at 10% level*
Source: authors' own calculations

6. Discussion

This paper discusses the impact on monthly rent prices (i.e. Rent Index) of six explanatory variables (Salary, Pollution Index, Safety Index, House Price to Income Ratio (HPIR), Traffic Commute Time Index and Cost of Living Index) and of three dummy variables referring to characteristics of European cities (i.e., university cities, tourist cities and capital or not). The data refer to a sample of 99 European cities analysed in relation to a series of variables that capture the quality of life.

In the last two decades, the role and expansion of the private rented sector have experienced important changes for both landlords and tenants. For example, in 2022, 4.6 million households in England rented their home from a private landlord, representing 19% of all households that have more than doubled since 1997 (Cromarty, 2022. p.8).

The results of our study are relevant because statistics show that in 2023 a significant proportion of European households lived in overcrowded conditions (16.8%) or in housing situated in a particularly polluted area with smoke, dust, unpleasant odours or water pollution on a regular basis (12.2%) (Horvat and Coupechoux, coord., 2024, p.76, p.112). Moreover, the English Housing Survey (2021) estimated that 23% of homes in the private rented sector in England did not provide decent living conditions (Cromarty, 2022. p.8).

Salary is the strongest predictor in 1st, 2nd and 3rd models, suggesting a strong, direct and significant relationship between it and the Rent Index. The 4th, 5th and 6th models show that variables such as HPIR, Safety, Traffic Commute Time Index, university city, and capital have significant effects on Rent Index.

5th and 6th models (with higher R^2) are more strongly explanatory than 4th model, suggesting that adding variables such as university city and capital improves model performance. Traffic Commute Time Index and Safety Index are important variables that negatively influence Rent Index, while HPIR, university city, and capital have positive effects.

This paper's results are consistent with others, for example, Papantonis et al. (2022) state that, on average, across the EU, private rental of a home consumes up to 35% of the disposable income of lower-income households and tenants must bear the constant increase of market-rate rent prices (13% between 2010 and 2019).

The limitations of our study are multiple, as both qualitative and quantitative methods are needed to understand how rental supply and rental demand affect all aspects of the local housing market and the quality of Europeans life and health. In addition, special attention should be paid to exploring and understanding rental market dynamics and the influence of these dynamics on the labour market and local economic, social and cultural development. Another limitation of our study is that we were unable to obtain real data from many other European cities for 2024.

7. Conclusion

The housing rental sector may become a very risky business in the future for various reasons: increasing costs related to the construction and maintenance of apartments, increasingly congested traffic in big cities, increasing inflation at a faster pace than wage growth, acceptance of remote work in more and more fields and very diverse jobs, expansion of online university courses (so that students can work part-time), trend towards reducing globalization, etc.

An efficient and effective rental sector of the housing market can provide numerous benefits at the individual (as tenants and landlords), social (as public socio-legal institutions) and local levels (citizenship and administrative actors), starting from balanced and fair competition (social rental versus private rental), well-being for everyone by assuring housing availability and affordability, reducing the risks of home ownership for those employed on a fixed-term basis, promoting residential mobility, workforce mobility and increasing demand, encouraging the supply of competitive housing and affordable rental prices, transparent taxation, etc.

The results of this study may be useful not only to current and potential landlords and tenants but also to persons who own several homes and need to decide whether it is better to sell them or keep them for rental. The conclusions of this study may also be useful to businesspeople involved in the real estate market, city authorities, urban planners, etc., and finally, to researchers concerned with economic development, the labour market, migration and mobility, etc.

Future research may consider using more predictor variables in order to capture more accurately the determinants of rent prices in many other European cities. A comparative analysis of cities from other continents, separately for developed economies and for developing economies, with very different national regulations of the private rental sector should not be neglected.

References

Behera, D.K., Rahut, D.B., Padmaja, M. and Dash, K.A. (2024) Socioeconomic determinants of happiness: Empirical evidence from developed and developing countries. Journal of Behavioral and Experimental Economics, 109(April), pp. 1-14. https://doi.org/10.1016/j.socec.2024.102187

Cromarty, H., (2022) Housing conditions in the private rented sector (England). House of Commons Library, 7328. https://researchbriefings.files.parliament.uk/documents/CBP-7328/CBP-7328.pdf

Diamond, R., McQuade, T., Qian, F. (2019) The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco. American Economic Review, 109(9). http://www.aeaweb.org/articles?id=10.1257/aer. 20181289

Early, D.W. (2000) Rent Control, Rental Housing Supply, and the Distribution of Tenant Benefits. Journal of Urban Economics, 48(2), 185–204. https://doi.org/10.1006/juec.1999.2163

EUROSTAT.https://ec.europa.eu/eurostat/statistics-explained/index.php?title= Living_ conditions_in_Europe_-_housing [Accessed June 21st, 2024].

Gete, P. and Reher, M. (2018) Mortgage Supply and Housing Rents. The Review of Financial Studies, 31(12), pp. 4884-4911. https://doi.org/10.1093/rfs/hhx145

Hahn, A.M. Kholodilin, K.A., Waltl, S.R. and Fongoni, M., 2022. Forward to the Past: Short-Term Effects of the Rent Freeze in Berlin (Revised Version). German Institute for Economic Research: DIW Berlin. https://www.diw.de/discussionpapers

Horvat, N. and Coupechoux, S. (coordination) (2024) Ninth Overview of Housing Exclusion in Europe. Fondation Abbé Pierre and FEANTSA. https://www.feantsa.org/public/user/Activities/events/2024/9th_overview/Rapport_-_EN.pdf

Howard, G. and Liebersohn, J. (2021) Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities. Journal of Urban Economics, 124 (July), pp. 1-14. https://doi.org/10.1016/j.jue.2021.103369

Karpinska, L. and Śmiech, S. (2023) Multiple faces of poverty. Exploring housing-costs-induced energy poverty in Central and Eastern Europe. Energy Research & Social Science, 105(November). https://doi.org/10.1016/j.erss.2023.103273

Kholodilin, K.A. (2022) Rent control effects through the lens of empirical research. Berlin: Deutsches Institut für Wirtschaftsforschung, 139. https://www.diw.de/documents/publikationen/73/diw_01.c.833177.de/diw_roundup_139_en.pdf

Kou, N., Suo, J., Wu, M. and Song, B. (2024) The influence of residential Environment and residential experience on psychological depression in older adults: Evidence from China and Europe. Health & Place, 88(July). https://doi.org/10.1016/j.healthplace.2024.103264

Lux, M. and Sunega, P. (2010) Private Rental Housing in the Czech Republic: Growth and…"? Czech Sociological Review, 46(3), pp. 349–373. https://www.jstor.org/stable/41132863

Mureșan, G.M., Ciumaș, C. and Achim, M.V. (2020) Can money buy happiness? Evidence for European countries. Applied Research in Quality of Life, 15, pp. 953–970. https://doi.org/10.1007/s11482-019-09714-3

Newton, D., Lucock, M., Armitage, R., Monchuk, L. and Brown, P. (2022) Understanding the mental health impacts of poor quality private-rented housing during the UK's first COVID-19 lockdown. Health and Place, 8 (November). https://doi.org/10.1016/j.healthplace.2022.102898

Numbeo (the world's largest cost of living database), [online] Available at: https://www.numbeo.com/cost-of-living/

OECD (2020) Regions and cities at a glance. France, Paris.

O'Toole, C., Martinez-Cillero, M., Ahrens, A. (2021) Price regulation, inflation, and nominal rigidity in housing rents. Journal of Housing Economics, 52. https://doi.org/10.1016/j.jhe.2021.101769

Papantonis, D., Tzani, D., Burbidge, M., Stavrakas,V., Bouzarovski, S. and Flamos, A. (2022) How to improve energy efficiency policies to address energy poverty? Literature and stakeholder insights for private rented housing in Europe. Energy Research & Social Science, 93 (November). https://doi.org/10.1016/j.erss.2022.102832

Perry, J., Stephens, M., Williams P., Young G., et al. (2021) UK Housing Review Briefing Paper. Chartered Institute of Housing, Pdf free download. https://www.ukhousingreview.org.uk/ukhr21/docs/UKHR-Briefing-2021.pdf

Saiz, A. (2007) Immigration and housing rents in American cities. Journal of Urban Economics, 61(2), pp. 345-371. https://doi.org/10.1016/j.jue.2006.07.004

Soederberg, S. (2018) The rental housing question: Exploitation, eviction and erasures. Geoforum, 89(February), pp. 114-123. https://doi.org/10.1016/j.geoforum.2017.01.007

Stephens, M., Perry, J., Williams P., Young G., et al. (2024) UK Housing Review Briefing Paper. Chartered Institute of Housing, Pdf free download. https://www.cih.org/media/ox2gxc55/ukhr-autumn-briefing-2024.pdf

Trojanek, R. and Gluszak, M. (2022) Short-run impact of the Ukrainian refugee crisis on the housing market in Poland. Finance Research Letters, 50(December). https://doi.org/10.1016/j.frl.2022.103236

von Platten, J., Mangold, M, Johansson, T. and Mjörnell, K. (2022) Energy efficiency at what cost? Unjust burden-sharing of rent increases in extensive energy retrofitting projects in Sweden. Energy Research & Social Science, 92 (October). https://doi.org/10.1016/j.erss.2022.102791

Źróbek-Różańska, A. (2022) Stay or leave the city? Students and rented accommodation during COVID-19 pandemic. Cities, 128(September), pp.1-6. https://doi.org/10.1016/j.cities.2022.103788
World Bank (2024). Annual Report: A Better Bank for A Better World. https://www.worldbank.org/en/about/annual-report
Yanatma, S. (2024). Wealth of nations: Gap between rich and poor in Europe widens. Published on https://www.euronews.com/business/2024/06/04/the-wealth-of-nations-the-widening-gap-between-rich-and-poor-in-europe

STUDIA UNIVERSITATIS BABEŞ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 67-82
DOI: 10.2478/subboec-2025-0004

THE SECOND AXELROD TOURNAMENT: A MONTE CARLO EXPLORATION OF UNCERTAINTY ABOUT THE NUMBER OF ROUNDS IN ITERATED PRISONER'S DILEMMA

Gabriel POP[*]
Babeș-Bolyai University, Romania

Mircea MILENCIANU
Babeș-Bolyai University, Romania

Alexandra POP
Babeș-Bolyai University, Romania

ABSTRACT: Strategic decision-making in multi-agent interactions inside the Iterated Prisoner's Dilemma (IPD) is investigated in this work using Monte Carlo simulations. Building on Axelrod's work, we present a second-generation tournament with stochastic components, including unpredictable game lengths, to evaluate strategy adaptability and resilience. We analyze how uncertainty influences strategic performance by using a comparison between instances with fixed and uncertain times. We identify, using a descriptive approach, methods demonstrating important behavioral differences between deterministic and uncertain settings. The results provide understanding of adaptive learning, response dynamics, and strategic flexibility, so helping to build strong collaborative strategies for artificial intelligence and decision-making systems. Our results highlight the limitations of exclusively deterministic methods and suggest the necessity for adaptive approaches to improve long-term cooperative success.

JEL Classification: C70; C73; C79

Keywords: Prisoner's Dilemma; repeated games; Axelrod second Tournament; agent-based modeling; finite and infinite games

1. Introduction

An essential feature of multi-agent interactions is strategic decision-making, who game theory offers a strong structure for examining such choices. Being much investigated to understand cooperative and competitive behaviors in repeated

[*] Corresponding author. Address: Faculty of Economics and Business Administration, Babeş-Bolyai University, 58-60, Teodor Mihali Street, 400591, Cluj-Napoca, Romania.
E-mail: gabriel.pop@ ubbcluj.ro

interactions, the Iterated Prisoner's Dilemma (IPD) is one of the most well-known models in this field of study (Chong et al., 2007). Robert Axelrod's innovative work revealed a collection of computational competitions evaluating the success of different approaches in the IPD, therefore generating important new perspectives on the development of collaboration.(Axelrod, 1984)

Building on this foundation, the current work investigates a second-generation Axelrod tournament (project, 2015) using improved computational frameworks and uncertainty factors to evaluate the adaptability and robustness of many agent strategies. Unlike the original tournament, which mostly concentrated on deterministic interactions, our work incorporates stochastic aspects, such unpredictable game lengths, to investigate how strategies function in non-deterministic situations.

Using Monte Carlo simulations, we methodically evaluate the performance of a diverse set of strategies under both fixed and uncertain conditions (Thomopoulos, 2013). We examine important indicators including win rates, utility differences, and adaptive resilience to find which tactics hold true when the game environment veers from strict determinism and which ones get vulnerable. The results establish the effects on perpetual success in repeated games of adaptive learning, a response dynamic, and strategic flexibility.

This work wants to develop and investigate a second-generation Axelrod tournament (Axelrod, 1984) including stochastic components, such unpredictable game lengths, so improving the realism of strategic interactions. It makes use of Monte Carlo simulations to evaluate agent adaptability of several approaches, therefore measuring agent resilience in uncertain environments. Finding important performance metrics—including win rates, utility variations, and adaptive resilience—that impact long-term strategic success under both deterministic and non-deterministic settings is a major focus.

Investigating how response dynamics and adaptive learning structures support cooperation or allow exploitation in iterative game environments is another objective. The study also compares conventional and most recently suggested approaches to evaluate their performance in handling uncertainty, therefore revealing important information on strategic resilience and adaptability (Chiong & Jankovic, 2008). Furthermore, it looks at how unpredictability in game parameters affects emergent strategy efficacy and how stochastic perturbations affect the stability and evolution of cooperative behaviors.

By using Monte Carlo methods to improve prediction accuracy and strategy evaluation, the study also seeks to maximize computational models for predicting recurrent interactions under uncertainty. At last, it aims to provide a basis for practical applications in artificial intelligence and decision-making by establishing rules for creating strong cooperative strategies that can resist different and changing strategic environment.

Through following these objectives, we want to improve our understanding of strategic decision-making in complex multi-agent interactions and provide a framework for future research on fostering cooperation in uncertain environments.

2. Theoretical background. Definitions. Notation

2. 1. Normal-form games

Normal-form games represent a fundamental model for studying strategic decisions. A normal-form game is defined as a tuple:

$$\boldsymbol{\Gamma} = (\boldsymbol{N}, \boldsymbol{A}, \boldsymbol{u}) \tag{1}$$

where:

- $\boldsymbol{N}$ is a finite set of players, $|\boldsymbol{N}| = \boldsymbol{n}$, with $\boldsymbol{n} \geq \boldsymbol{2}$.
- $\boldsymbol{A_i}$ is the set of pure strategies for each player $\boldsymbol{i} \in \boldsymbol{N}$.
- $\boldsymbol{A} = \prod_{i \in N} \boldsymbol{A_i}$ is the Cartesian product of all strategies.
- $\boldsymbol{u_i}: \boldsymbol{A} \rightarrow \mathbb{R}$ is the utility function, assigning a numerical outcome to each strategy profile.

Games where players make decisions at the same time without knowing what other players are thinking work well with this structure.(Ben-Porath, 1990)

A Nash equilibrium is a fundamental concept in game theory, representing a strategy profile where no player can gain by unilaterally deviating from their strategy. Formally, a strategy profile (s_i^*, s_{-i}^*) a Nash equilibrium if:

$$\boldsymbol{u_i}(\boldsymbol{s_i^*}, \boldsymbol{s_{-i}^*}) \geq \boldsymbol{u_i}(\boldsymbol{s_i}, \boldsymbol{s_{-i}^*}), \quad \forall \boldsymbol{s_i} \in \boldsymbol{A_i}. \tag{2}$$

In other words, given the strategies chosen by the other players, each player's strategy is optimal, meaning that no player has an incentive to deviate. Nash equilibria exist in all finite games with mixed strategies (Nash, 1950) The concept is applicable to various strategic contexts, ranging from basic two-player games to complex multi-agent systems. In the context of normal-form games, Nash equilibria can be classified as either pure or mixed. A pure strategy Nash equilibrium occurs when all players choose a single, deterministic strategy, while a mixed strategy Nash equilibrium involves players randomizing over their available strategies according to a probability distribution. Finding Nash equilibria in complex games often requires computational tools. One common method involves the use of Nash mappings, which are mathematical constructions that transform the problem of finding equilibria into solving fixed-point equations. Nash mappings are essential for identifying equilibria, particularly in games with infinite or continuous strategy spaces (Zhou et al., 2011). Zhou et al. explain that Nash mappings rely on fixed-point theorems, such as Tychonov's fixed-point theorem, which guarantee the existence of fixed points under certain conditions (Tychonov, 1935). The mapping ensures that for any starting point in the strategy space, there is a corresponding equilibrium point at which no player can independently enhance their payoff. This theoretical framework is essential for examining games with extensive strategy spaces and continuous payoff functions. Algorithms like the Lemke–Howson algorithm and best-response dynamics are frequently used to calculate Nash equilibria. Best-reaction dynamics involve the iterative adjustment of each player's strategy to optimize their response based on the dominant strategies of the other players. This process may converge to a Nash equilibrium in numerous instances, although this outcome is not invariably assured.

The importance of Nash equilibria extends simple theoretical examination. Nash equilibria provide a paradigm for forecasting the results of strategic interactions in practical applications throughout economics, political science, and computer science. They are utilized to simulate competitiveness in the market, negotiation scenarios, and multi-agent decision-making in distributed systems.

2. 2. Dominant strategies and Iterative elimination

Dominant strategies play a crucial role in decision-making within strategic games. A strategy is considered dominant if it provides a player with a higher or equal payoff regardless of the actions of other players (Rothe, 2010) In Axelrod's tournaments, iterative elimination of dominated strategies explains why inflexible strategies, such as constant defection, perform poorly in uncertain environments. Akin explains that zero-determinant (ZDS) strategies allow players to fix the relationship between their payoffs and those of their opponents, which can either stabilize cooperation or enable exploitation (Akin, 2016) ZDS strategies were extended by Press and Dyson (2012) and play a significant role in the evolutionary dynamics of repeated games.

2. 3. The classic example: Prisoner's Dilemma

The Prisoner's Dilemma is a classic example of a normal-form game and is central to studying cooperation and defection between players. Each player must simultaneously choose between cooperation (C) or defection (D), with outcomes depending on the combination of their decisions.

The payoff matrix can be represented as:

Table 1. The payoff matrix

X/Y	C	D
C	R	S
D	T	P

Where:

$\boldsymbol{R}$ is the reward for mutual cooperation;
$\boldsymbol{P}$ is the punishment for mutual defection;
$\boldsymbol{T}$ is the temptation payoff when a player defects while the other cooperates;
$\boldsymbol{S}$ is the sucker's payoff for cooperating when the other defects.

These payoff relationships must satisfy:

$$\boldsymbol{T} > \boldsymbol{R} > \boldsymbol{P} > \boldsymbol{S} \quad \text{and} \quad \boldsymbol{2R} > \boldsymbol{T} + \boldsymbol{S}. \tag{3}$$

Alternatively, the payoff vectors for each player can be expressed as:

$$\boldsymbol{S_X} = (\boldsymbol{R}, \boldsymbol{S}, \boldsymbol{T}, \boldsymbol{P}) \quad \text{and} \quad \boldsymbol{S_Y} = (\boldsymbol{R}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P}). \tag{4}$$

The expected payoffs s_X and s_Y for players X and Y, given a probability distribution v over the four possible outcomes, are:

$$s_X = \langle v \cdot S_X \rangle \quad \text{and} \quad s_Y = \langle v \cdot S_Y \rangle. \tag{5}$$

In the Prisoner's Dilemma, strategy C corresponds to cooperation, where both players receive the reward R when they cooperate. Conversely, strategy D (defection) leads to the temptation payoff T for the defector and the sucker's payoff S for the cooperating player. The mutual defection outcome results in the punishment payoff P for both players.

The condition $2R > T + S$ implies that mutual cooperation provides a payoff higher than splitting the total rewards of outcomes where one player defects and the other cooperates. The cooperative outcome (C, C) is a Pareto optimum, meaning that no other outcome can make one player better off without making the other worse off.

However, players face a dilemma due to the dominance of strategy D. Regardless of what the other player chooses, defection yields a higher payoff for the defector. As a result, both players rationally choose D, leading to the (D, D) outcome with a suboptimal payoff P for each.

In real-world scenarios, additional factors, such as reputation or external enforcement mechanisms, can modify these payoffs to promote cooperation. For instance, if one player defects and causes harm to the other, retaliation may reduce the desirability of defection below R. Anticipation of such consequences may encourage both players to honor agreements to cooperate.

2. 4. Repeated Prisoner's Dilemma

In a repeated version of the Prisoner's Dilemma, players encounter each other multiple times. The game can either have a fixed number of rounds (finite horizon) or continue indefinitely (infinite horizon). The difference in the horizon fundamentally affects the strategies employed by players.

In finitely repeated games, players know when the game will end, which often results in backward induction. Since defection is dominant in the final round, players reason backward to defect in all previous rounds (Myerson, 1991).

When the game is infinitely repeated or has an unknown ending, cooperation can emerge as a rational strategy. The incentive to cooperate arises from the potential future benefits of mutual cooperation. Two main concepts for evaluating payoffs in infinitely repeated games are:

1. Average reward: Given an infinite sequence of payoffs $r_i^1, r_i^2, \ldots$ for player i, the average reward is:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} r_i\,(j). \tag{6}$$

2. Discounted reward: Given an infinite sequence of payoffs $r_i^1, r_i^2, \ldots$ for player i, and a discount factor $0 \leq \beta \leq 1$, the discounted reward is:

$$\lim_{N\to\infty} \sum_{j=1}^{N} \beta^j r_i(j). \quad (7)$$

Strategies like Tit for Tat, which involves initial collaboration followed by imitating the opponent's previous action, effectively maintain cooperation in repeated games (Axelrod & Hamilton, 1981).

Tit for Tat is considered a Nash equilibrium under certain conditions, particularly when the discount factor is high enough, making future payoffs valuable.

The folk theorem provides further insight into repeated games, stating that if players are sufficiently patient, a wide range of outcomes, including mutual cooperation, can be sustained as Nash equilibria (Leyton-Brown & Shoham, 2008).

Good strategies for the iterated Prisoner's Dilemma have been characterized by Akin (2016) as those that stabilize cooperation and ensure Nash equilibria. When both players use a good strategy, neither player can gain by unilaterally changing their strategy. Markov strategies, including memory-one plans, play a central role in this context. For example, Tit for Tat is a memory-one strategy where a player's response depends on the opponent's previous action.

Meng Chen-Lu (2001) highlights that Tit for Tat and its variations, like Generous Tit for Tat, are essential for promoting ongoing cooperation in iterated games, in particular uncertain context. Nash mappings help identify stable strategies in complex situations, providing players with opportunities to optimize their long-term behaviour (Zhou et al., 2011).

Nachbar argues that through progressive learning and feedback-based adjustments, players can converge to Nash equilibria even when their initial strategies are suboptimal (Nachbar, 2005).

2. 5. Strategies in determined vs. uncertain environments

An essential component in analyzing repeated games is the distinction between deterministic and uncertain contexts. In determined environments, where the game's duration is known in advance, dominating strategies could be successful as participants anticipate the conclusion. On the other hand, in uncertain environments, adaptive strategies like Tit for Tat or exploration-exploitation models are generally more robust (Leonardos Stefanos, 2012).

Meng (2001) emphasizes that in complex environments with a large number of players or unpredictable interactions, an agent's performance depends on its ability to adapt. Adaptive algorithms, such as genetic algorithms, are crucial for optimizing behaviour in the face of diverse opponent strategies.

Ben-Porath (1990) highlights the computational complexity of establishing optimal response algorithms in recurrent games with mixed strategies. Computing and verifying optimal responds are frequently non-polynomial problems; however, they can become polynomial when the support size of mixed strategies is restricted.

2. 6. Axelrod's Second Tournament

Axelrod's second tournament extended the scope of his initial experiment by incorporating a larger set of strategies and introducing new conditions such as uncertainty

regarding the number of game rounds. This allowed for a more comprehensive evaluation of strategies in the Iterated Prisoner's Dilemma (IPD) and provided deeper insights into the mechanisms that drive cooperation and competition (Axelrod, 1984) Unlike the first tournament, where Tit-for-Tat dominated due to its simplicity and reciprocity, the second tournament tested both reciprocal and adaptive strategies that adjusted dynamically to their opponents' behavior (Axelrod & Hamilton, 1981)

Key aspects of strategy that contribute to success in frequent experiences are highlighted in theoretical studies of Axelrod's tournaments. Among the key concepts that Axelrod's study has demonstrated are:

(i) **Niceness:** Strategies that avoid initial defection generally provide superior long-term payoffs. (Axelrod, 1984)
(ii) **Provocability:** Effective strategies must respond to defections in a way that discourages exploitation.
(iii) **Forgiveness:** The capacity to reestablish collaboration following the punishment of a defection helps in preventing extensive cycles of payback.
(iv) **Clarity:** Easily interpretable strategies allow opponents to predict the consequences of their actions, stabilizing interactions

In instances with a fixed number of rounds, inflexible strategies like Tit-for-Two-Tats, which delay retaliation until two successive defections occur, shown robust success. However, their inherent a lack in responding to defection made them especially vulnerable to manipulate by adaptive tactics like as Second by Tester. The latter exploited this tolerance by strategically defecting in specific cases, aware that immediate punishment would be missing. This opportunistic behavior enabled adaptive methods to achieve a greater cumulative payoff by exploiting the predictable responses of more rigid adversaries. Similarly, Grudger, which permanently punishes any opponent who defects, struggled in dynamic environments where adversaries could introduce strategic resets that neutralized its punitive approach (Tutzauer, 2007). On the other hand, strategies that adapted to changing conditions, such as Second by Tester and Second by Gladstein, thrived under uncertainty. These strategies adjusted their responses based on game history, allowing them to maximize payoffs in environments where the end-game was unpredictable. Vincent & Fryer (2021) further demonstrated that FSM-based strategies, which evolve through algorithmic optimization, could outperform even established heuristics like Tit-for-Tat by fine-tuning their behavior to different types of opponents.

Axelrod's second tournament findings, along with later theoretical improvements, highlight that adaptation is the essential factor for success in unpredictable situations. Strategies that strictly follow predetermined rules, such as Grudger or Tit-for-Two-Tats, are vulnerable to exploitation by adversaries capable of adapting their behavior (Axelrod, 1984).

3. Our approach to Monte Carlo simulations

3. 1. Definition of a Supergame

A supergame is defined as the repeated execution of N iterations of a simultaneous one-shot game (in our case, the Prisoner's Dilemma). The total utility is computed as the discounted sum of utilities across all one-shot games:

$$\boldsymbol{U}_{\text{total}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\beta}^{\boldsymbol{i-1}}\, \boldsymbol{u_i}, \tag{8}$$

where u_i is the utility at stage i and β is the discount factor with $0 \leq \beta \leq 1$.

Simulation of agent interactions

In a confrontation between two strategies or agents (e.g., Tit-for-Tat vs Tit-for-Two-Tats), the agents play K iterations of a supergame:

$\mathbf{vector_1}$**:** The proportion of wins, losses, and ties for each agent over $\boldsymbol{K}$ supergames.
$\mathbf{vector_2}$**: T**he average payoffs for each agent across $\boldsymbol{K}$ supergames.

To simulate uncertainty, the random variable $\boldsymbol{N}$ is generated from a distribution with mean $\boldsymbol{\mu_N}$ and standard deviation $\boldsymbol{\sigma_N}$. The distribution $\mathcal{D}$ can be:

Normal distribution:

$$\boldsymbol{N} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu_N}, \boldsymbol{\sigma_N}) \tag{9}$$

where $\boldsymbol{\mu_N}$ is the mean and $\boldsymbol{\sigma_N}$ is the standard deviation.

Uniform distribution:

$$N \sim \mathcal{U}a,b \tag{10}$$

where the mean and standard deviation are given by the formulas:

$$\boldsymbol{\mu_N} = \frac{\boldsymbol{a+b}}{\boldsymbol{2}} \tag{11}$$
$$\boldsymbol{\sigma_N} = \frac{\boldsymbol{b-a}}{\sqrt{\boldsymbol{12}}} \tag{12}$$

3. 2. Comparison of fixed and uncertain cases

For each $\boldsymbol{k}$ in $\{\boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{K}\}$, the uncertain supergame is played by randomly selecting $\boldsymbol{N'}$ from the distribution in Equation (2). The corresponding $\mathbf{vector_1}$ and $\mathbf{vector_2}$ are computed and compared to the results from the fixed case.

Experimental setup

We select a set of $\boldsymbol{m}$ agents (strategies), denoted as $\text{Agent}_1, \text{Agent}_2, \ldots, \text{Agent}_{\boldsymbol{m}}$. The results are stored as:

Case I: Fixed $\boldsymbol{N}$. Results are stored in a matrix $\boldsymbol{M}_{\text{case\_I}}$, where each entry above the diagonal represents $\mathbf{vector_1}$ and $\mathbf{vector_2}$ for the bilateral supergame between $\text{Agent}_{\boldsymbol{i}}$ and $\text{Agent}_{\boldsymbol{j}}$.

Case II: Uncertain $\boldsymbol{N}$. Results are stored in a matrix $\boldsymbol{M}_{\text{case\_II}}$ with the same structure as $\boldsymbol{M}_{\text{case\_I}}$.

Calculating distance metrics

The following metrics are used to measure discrepancies between Case I (fixed number of stages) and Case II (uncertain number of stages).

Discrepancy for "won" (victories)

$$\boldsymbol{D}_{\text{won}}(\boldsymbol{i},\boldsymbol{j}) = |\text{won}_{\text{case I}}(\boldsymbol{i},\boldsymbol{j}) - \text{won}_{\text{case II}}(\boldsymbol{i},\boldsymbol{j})|$$

Where:

$\boldsymbol{D}_{\text{won}}(\boldsymbol{i},\boldsymbol{j})$ represents the absolute difference between the number of victories for the pair of agents $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in the two cases.

The results are sorted in descending order to select the pairs of agents with the largest discrepancies, up to a threshold threshold $= \mathbf{10}$.

Discrepancy for "lost" (losses)

$$\boldsymbol{D}_{\text{lost}}(\boldsymbol{i},\boldsymbol{j}) = |\text{lost}_{\text{case I}}(\boldsymbol{i},\boldsymbol{j}) - \text{lost}_{\text{case II}}(\boldsymbol{i},\boldsymbol{j})|$$

Where:

$\boldsymbol{D}_{\text{lost}}(\boldsymbol{i},\boldsymbol{j})$ represents the absolute difference between the number of losses for the pair of agents $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in the two cases.

The results are sorted in descending order to select the pairs of agents with the largest discrepancies, up to a threshold threshold $= \mathbf{10}$.

Notation:

- $\text{won}_{\text{case I}}(\boldsymbol{i},\boldsymbol{j})$: the number of victories for the pair $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in case I (fixed number of stages).
- $\text{won}_{\text{case II}}(\boldsymbol{i},\boldsymbol{j})$: the number of victories for the pair $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in case II (uncertain number of stages).
- $\text{lost}_{\text{case I}}(\boldsymbol{i},\boldsymbol{j})$: the number of losses for the pair $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in case I.
- $\text{lost}_{\text{case II}}(\boldsymbol{i},\boldsymbol{j})$: the number of losses for the pair $\left(\text{Agent}_{\boldsymbol{i}}, \text{Agent}_{\boldsymbol{j}}\right)$ in case II.

Axelrod's tournament framework provides a valuable platform for studying the interactions and strategies of agents in repeated games, particularly the Iterated Prisoner's Dilemma (IPD). This subsection focuses on the qualitative behavior of

specific agent pairs under conditions of uncertainty about the number of game stages. By introducing uncertainty, we investigate whether strategies originally developed for fixed-stage environments demonstrate adaptability or vulnerabilities.

4. Results. Some relevant examples explained.

Second by Tester and Tit For 2 Tats

Behavior: Tit For 2 Tats, known for its leniency in tolerating a single defection, is exploited by the adaptive strategy of Second by Tester. Specifically, Second by Tester leverages the tolerance built into Tit For 2 Tats, which delays its retaliation and allows an opportunistic exploitation pattern.

Cause of discrepancy: The fundamental cause lies in the adaptability of Second by Tester. Unlike Tit For 2 Tats, which applies a rigid, predefined retaliation rule, Second by Tester dynamically adjusts its behavior to maximize payoffs under varying conditions. This adaptability is particularly advantageous in environments where the end of the game is uncertain.

Impact of uncertainty: Under conditions of uncertainty, Second by Tester's performance gains are amplified. This is due to the variability in game length, which increases the opportunities for exploitation. The increased discrepancies highlight the potential for adaptive strategies to thrive in variable conditions while exposing the limitations of rigid, rule-based strategies like Tit For 2 Tats.

Second by Gladstein and Tit For 2 Tats

Behavior: A similar pattern emerges in the interaction between Second by Gladstein and Tit For 2 Tats. Gladstein's strategy, characterized by its adaptive responses, takes advantage of Tit For 2 Tats' predictable tolerance for initial defections.

Observation: The introduction of uncertainty magnifies the discrepancy between these two strategies. Gladstein's ability to adjust to changing game dynamics enhances its effectiveness, while Tit For 2 Tats' static rule set makes it vulnerable.

Impact of uncertainty: The amplified discrepancy underscores the role of adaptability in uncertain environments. While Tit For 2 Tats performs reliably under fixed conditions, its lack of dynamic response mechanisms makes it susceptible to exploitation when the game's duration becomes unpredictable.

Grudger and Second by Colbert

Behavior: Grudger, a highly rigid strategy that perpetually punishes any defection, interacts with Second by Colbert, which introduces resets to destabilize its rigid retaliation mechanism. This interaction highlights the limitations of strict punitive strategies in dynamic environments.

Cause of discrepancy: The rigidity of Grudger is a critical weakness. Its unyielding approach fails to account for strategic resets, which Second by Colbert effectively uses to reset the game's cooperative dynamics. By doing so, Second by Colbert neutralizes Grudger's retaliatory strategy and destabilizes its long-term effectiveness.

Impact of uncertainty: In uncertain environments, Grudger's inability to adapt is further exposed. The resets introduced by Second by Colbert exploit the variability in game length, leading to significant performance discrepancies. This demonstrates that strategies overly reliant on fixed retaliation rules are ill-suited to uncertain conditions.

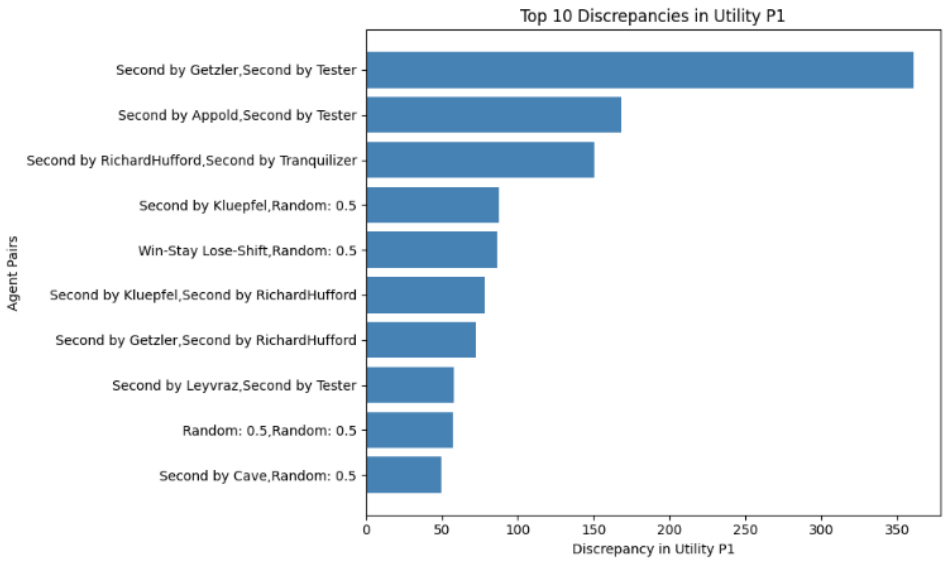
5. Discussions and conclusions

In this section, we analyze and interpret the results of the second-generation Axelrod tournament simulations conducted in environments with both certainty and uncertainty regarding the number of game stages. The main focus lies on understanding how agents adapt (or fail to adapt) to the uncertainty about the game's end stage and how this impacts their performance.

5. 1. Discussions about essential metrics and observations

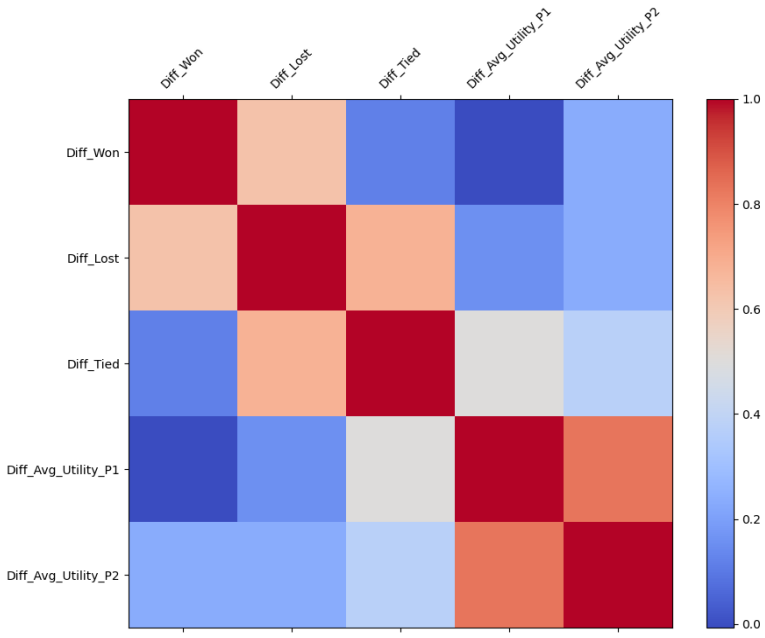
We evaluated agent behavior by computing critical metrics, especially looking at the differences in average utilities—referred to as "Diff P1" and "Diff P2"—between certain environments and those that introduce uncertainty regarding the end stage. These metrics establish a quantitative basis for identifying and evaluating performance differences in these distinct scenarios. The analysis presented in Figure 1 identifies the top 10 significant discrepancies in utility for P1. It demonstrates that certain agents, including "Second by Getzler" and "Second by Tester," show important sensitivity to the introduction of uncertainty, illustrating the influence of environmental conditions on their strategies.

Figure 1. Top 10 Discrepancies in Utility for P1 Across Certain and Uncertain Environments



A significant correlation was identified between the discrepancies in utilities and adaptability metrics, as illustrated in Figure 2. This suggests that agents' ability to adapt plays a pivotal role in their performance across varying environmental conditions. For example, strategies that are heavily dependent on end-game scenarios, such as “Second by Tester,” exhibit pronounced drops in utility when uncertainty is introduced, underscoring their reliance on predictable termination points. In contrast, more resilient strategies, such as “Grudger,” demonstrate remarkable stability and maintain consistent performance regardless of the level of uncertainty, reflecting their inherent robustness and less dependence on end-stage predictability.

Figure 2. Correlation Matrix of Wins, Losses, and Utilities



The difference in average utilities between the two environments is another important result that provides insight on how flexible and resilient various tactics are. Figure 3 illustrates these differences, highlighting specific patterns in performance. Strategies characterized by significant adaptability, such as "Tit for Tat," display minimal fluctuations in their average utility, indicating a reliable and stable reaction to environmental changes. This stability shows their intrinsic resilience and ability to function efficiently in both certain and uncertain conditions. In contrast, techniques characterized by rigid or deterministic decision-making frameworks, such as "Second by Tester," show significant reductions in utility when tested in unpredictable situations. These significant losses highlight their reliance on stable conditions and their restricted capacity to adapt to dynamic changes in the game framework. This disparity emphasizes the importance of flexibility and adaptability in strategy design, particularly in contexts where environmental conditions are unpredictable or volatile.

Figure 3. Average Utility Discrepancies Between Certainty and Uncertainty Environments

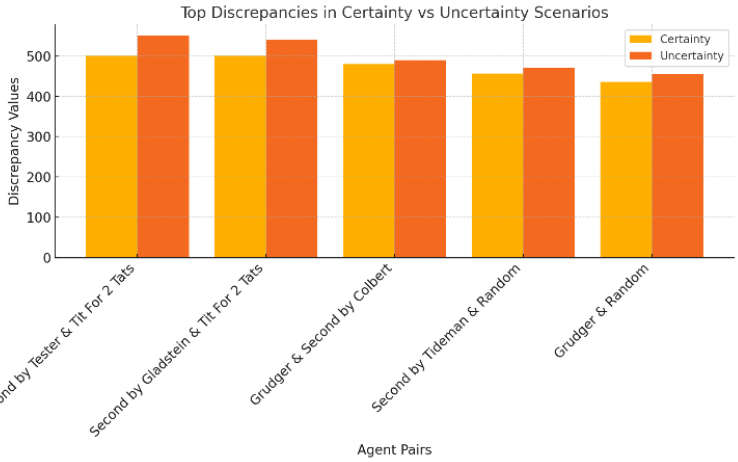


Figure 4 provides a thorough analysis of the distribution of utility values across scenarios defined by certainty and uncertainty, illustrating the fundamental dynamics of strategy performance. Strategies such as "Tit for Tat" show a compact utility distribution, indicating exceptional consistency and durability in their performance. Narrow distributions suggest that these techniques remain mostly unaffected by fluctuations in external variables, rendering them appropriate for both predictable and unexpected situations. In contrast, the "Second by Tester" strategy demonstrates a significantly wider utility distribution under uncertainty, indicating increased variability in its effectiveness. This wider distribution indicates a deficiency in robustness and a greater susceptibility to environmental variability. The difference in value distributions among different solutions highlights the essential importance of adaptability in reducing the negative impacts of uncertainty. Strategies characterized by narrower distributions usually show consistent decision-making frameworks, while those with broader distributions may lack mechanisms to stabilize performance in the face of unpredictable changes. This approach highlights the necessity of including both stability and adaptability in the formulation of strategies for situations characterized by various levels of predictability.

Figure 5 presents a comprehensive examination of the difficulties encountered by particular strategies in adjusting during transitions between environments characterized by certainty and uncertainty. This visualization highlights the differing levels of adaptation across strategies, providing essential insights regarding their performance resilience. The "Second by Tester" strategy demonstrates a substantial decline in efficacy while switching to an uncertain environment, indicating a possible overfitting to deterministic circumstances. This significant decline in performance indicates that the strategy is excessively dependent on predictable termination points and lacks the adaptability to respond to dynamic or less controlled scenarios.

These observations highlight the necessity for finding solutions that provide constant performance under different environmental situations. Strategies that demonstrate a significant decrease in utility under uncertainty may be useful in static

or deterministic environments but are inappropriate for dynamic and unpredictable circumstances. This discrepancy analysis functions as an essential instrument for detecting strategy that are too specialized or restricted in their flexibility.

Figure 4. Utility Distributions in Certainty and Uncertainty Scenarios.

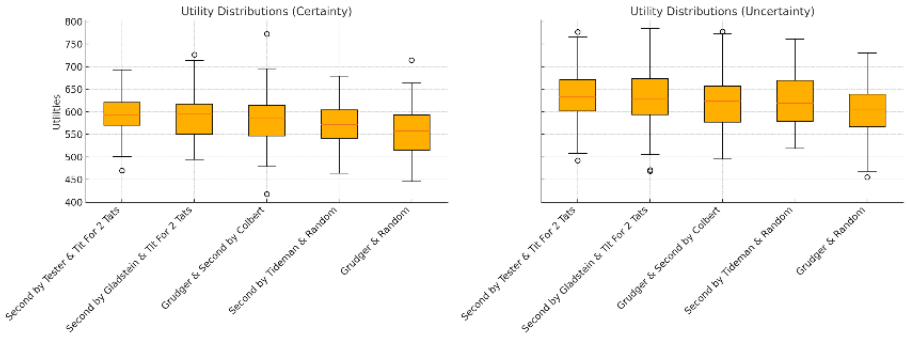
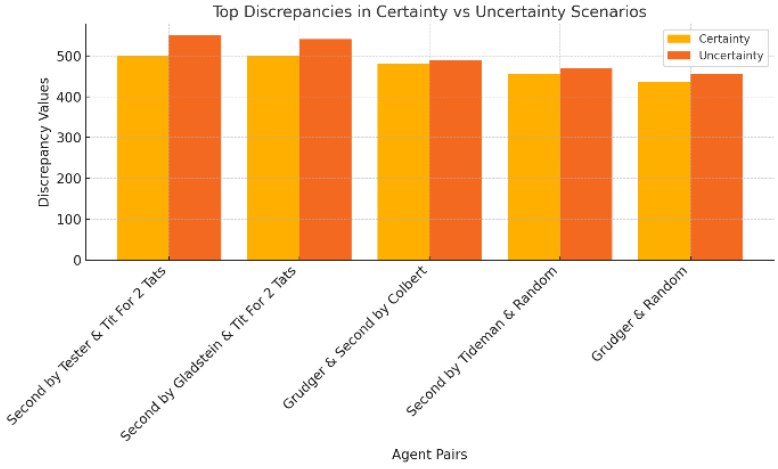


Figure 5. Discrepancy Analysis: Certainty vs. Uncertainty



5. 2. Conclusions

The above visualizations provide important insights into the performance dynamics of agents across different environmental situations, highlighting their adaptation, resilience, and limitations. Strategies like “Tit for Tat” and “Grudger” exemplify robustness, constantly providing stable and reliable performance regardless the volatility in their operating environments. These strategies demonstrate an effective combination of collaboration and flexibility, rendering them especially appropriate for volatile and uncertain situations where stability is important.

In contrast, strategy like “Second by Tester” and “Second by Getzler” reveal considerable weaknesses due to their evident dependence on deterministic frameworks and predictable game characteristics. These tactics demonstrate significant decreases in efficiency under uncertainty, indicating a fundamental inflexibility in their structure. Their incapacity to maintain constant performance under changing settings highlights essential areas for improvement, especially the necessity for reducing dependency on fixed ending points and deterministic frameworks.

An in-depth examination of these inconsistencies reveals opportunities for improving agent algorithms, highlighting the potential for innovation in various critical domains. Integrating probabilistic methods for making decisions may allow agents to navigate environments characterized by inherent randomness more effectively, enhancing adaptability. Moreover, the formulation of hybrid strategies—integrating resilient components from strategies such as “Tit for Tat” with innovative decision-making frameworks—presents a viable option to increase both flexibility and resilience. These improvements could allow agents to operate efficiently in both controlled environments and complex real-world situations marked by uncertainty and dynamic interactions.

Future research should focus on systematically refining these strategies to address their current limitations. This may include using sophisticated computational methods, such as machine learning, to model and improve agent behavior across diverse environmental situations. Furthermore, investigating methods to adapt strategic decision-making in response to real-time feedback might significantly enhance agents' capacity to react to evolving environments. By furthering these domains of investigation, we may improve the broader field of game theory and agent-based modeling, promoting developments that beyond academic study and reach practical applications in economics, artificial intelligence, and other disciplines.

References

Akin, E. (2016). The iterated Prisoner’s Dilemma: good strategies and their dynamics. In A. Idris (Ed.), *Ergodic Theory* (pp. 77-107). De Gruyter. https://doi.org/doi:10.1515/9783110461510-004

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390-1396.

Ben-Porath, E. (1990). The Complexity of Computing a Best Response Automaton in Repeated Games with Mixed Strategies. *Games and Economic Behavior*, *2*, 1–12. https://doi.org/https://doi.org/10.1016/0899-8256(90)90010-R

Chiong, R., & Jankovic, L. (2008). Learning game strategy design through iterated Prisoner's Dilemma. *Int. J. Comput. Appl. Technol.*, *32*, 216-223.

Chong, S. Y., Humble, J., Kendall, G., Li, J., & Yao, X. (2007). Iterated Prisoner's Dilemma and Evolutionary Game Theory. In *The Iterated Prisoners' Dilemma* (Vol. Volume 4, pp. 23-62). WORLD SCIENTIFIC.

Leonardos Stefanos, P. G. (2012). Exploration-Exploitation Tradeoffs in Multi-Agent Systems: Catastrophe theory meets game theory. *Journal of Artificial Intelligence Research*, *304*, 103653. https://doi.org/https://doi.org/10.1016/j.artint.2021.103653

Leyton-Brown, K., & Shoham, Y. (2008). *Essentials of Game Theory: A Concise Multidisciplinary Introduction* (Vol. 2). Morgan and Claypool Publishers.

Meng Chen-Lu, P. R. (2001). The Iterated Prisoner's Dilemma: early experiences with Learning Classifier System-based simple agents. *Decision Support Systems*, *31*(4), 379-403. https://doi.org/https://doi.org/10.1016/S0167-9236(00)00137-8

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict* (H. U. Press, Ed.). Harvard University Press.

Nachbar, J. (2005). Beliefs and Learning in Repeated Games. *Econometrica*, *73*(2), 459-480. http://www.jstor.org/stable/3598794

Nash, J. (1950). Equilibrium Points in n-Person Games. *Proceedings of the National Academy of Sciences*, *36*, 48–49.

Project, T. A. (2015). *Axelrod: second tournament*. https://axelrod.readthedocs.io/en/fix-documentation/reference/overview_of_strategies.html#axelrod-s-second-tournament

Press, W. H. & Dyson, F. J. (2012) Iterated Prisoners Dilemma contains strategies that dominate any evolutionary opponent. Proc. Natl. Acad Sci. USA 109, 10409–10413.

Rothe, J. (2010). *Economics and Computation - An Introduction to Algorithmic Game and Fair Division Theory, Computational Sociial Choice,and Fair Division*. Springer

Thomopoulos, N. (2013). *Essentials of Monte Carlo Simulation: Statistical Methods for Building Simulation Models*. Springer.

Tutzauer, F. (2007). A matrix approach to generalizing Axelrod-type tournaments. *Journal of Mathematical Sociology*, *31*(3), 249-266.

Tychonov, A. (1935). Ein Fixpunktsatz. *Mathematische Annalen*, *111*, 767–776.

Vincent, F. M., & Fryer, D. E. A. (2021). Top score in Axelrod Tournament. *arXiv preprint arXiv:2104.03347*.

Zhou, Y. H., Yu, J., & Wang, L. (2011). A new proof of existence of equilibria in infinite normal form games. *Applied Mathematics Letters*, *24*(3), 253-256. https://doi.org/https://doi.org/10.1016/j.aml.2010.09.014

STUDIA UNIVERSITATIS BABEŞ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 83-111
DOI: 10.2478/subboec-2025-0005

BERTWEETRO: PRE-TRAINED LANGUAGE MODELS FOR ROMANIAN SOCIAL MEDIA CONTENT

Dan Claudiu NEAGU*
Babes-Bolyai University, Romania

Abstract: The introduction of Transformers, like BERT or RoBERTa, have revolutionized NLP due to their ability to better "understand" the meaning of texts. These models are created (pre-trained) in a self-supervised manner on large scale data to predict words in a sentence but can be adjusted (fine-tuned) for other specific NLP applications. Initially, these models were created using literary texts but very quickly the need to process social media content emerged. Social media texts have some problematic characteristics (they are short, informal, filled with typos, etc.) which means that a traditional BERT model will have problems when dealing with this type of input. For this reason, dedicated models need to be pre-trained on microblogging content and many such models have been developed in popular languages like English or Spanish. For under-represented languages, like Romanian, this is more difficult to achieve due to the lack of open-source resources. In this paper we present our efforts in pre-training from scratch 8 BERTweetRO models, based on RoBERTa architecture, with the help of a Romanian tweets corpus. To evaluate our models, we fine-tune them on 2 down-stream tasks, Sentiment Analysis (with 3 classes) and Topic Classification (with 26 classes), and compare them against Multilingual BERT plus a number of other popular classic and deep learning models. We include a commercial solution in this comparison and show that some BERTweetRO variants and almost all models trained on the translated data have a better accuracy than the commercial solution. Our best performing BERTweetRO variants place second after Multilingual BERT in most of our experiments, which is a good result considering that our Romanian corpus used for pre-training is relatively small, containing around 51,000 texts.

JEL classification: C45, C55, C88, O33

Keywords: machine learning; natural language processing; language models; transformers; text classification; under-resourced languages

* Corresponding author. Address: Faculty of Economics and Business Administration, Babeş-Bolyai University, 58-60, Teodor Mihali Street, 400591, Cluj-Napoca, Romania
E-mail: dan.neagu@econ.ubbcluj.ro

1. Introduction

A Transformer model is a deep learning architecture that uses a multi-head attention mechanism to transform texts into numerical representations called tokens. These are then converted into vectors using the word embedding table that is generated in the training stage and each vector is later contextualized with the help of the attention mechanism by using the scope of the context window paired with other (unmasked) tokens. Thus, the "intensity" of the important tokens is amplified but diminished for the less important ones (Vaswavi et al., 2017).

This methodology was proposed by researchers at Google in 2017 and is advantageous because it does not use recurrent units which means that the training times are lower than other architectures such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM). Transformers were developed initially to improve machine translation (Luong et al., 2015) but later they were used on a large scale in many sub fields of Natural Language Processing (NLP), Computer Vision (CV), and robotics.

In the area of NLP, Devlin et al. (2019) introduced Bidirectional Encoder Representations from Transformers (BERT) in 2018 which is based on transformers. This self-supervised learning model was revolutionary because it brought drastic improvements compared to the state-of-the-art models of that time and is now considered to be an early example of a Large Language Model (LLM). BERT is trained for masked token prediction and/or next sentence prediction on huge volumes of unlabeled texts using a deep bidirectional neural network design. The model learns the latent representations of tokens in their textual context bidirectionally allowing it to "understand" more nuanced and complex expressions. The original version of BERT is an "encoder-only" transformer architecture that has 4 modules: tokenizer (used to transform texts into a series of integers), embedding (used to convert sequences of tokens into a vector of real numbers), encoder (Transformer blocks with self-attention), and task head (used to decode the latent representation into token types). One big advantage of this type of model is that, once created, it can be fine-fined in a supervised manner on various downstream tasks such as question answering, document classification, and language translation with higher accuracy and efficiency.

A Robustly Optimized BERT Approach (RoBERTa) was created in 2019 by Liu et al. and is one of the most popular extensions of BERT. It preserves the original architecture but improves upon it by changing key hyperparameters, removing the next sentence prediction task, and by using larger mini-batch sizes in the pretraining process. These adjustments allows RoBERTa to handle complicated variations of language more easily thus, improving the performance that can be achieved in a wide number of applications.

The quality, quantity, and type of data used to create BERT or RoBERTa models from scratch affect their behavior in the field in which they are operated. For example, Raffel et al. (2020) has shown the importance of high-quality datasets by finding a direct correlation between the training sets and the results obtained by the models. The exponential growth of digitally created content and the reduction in computational costs have also made it easier to work with large scale datasets in order to improve state of the art performances in areas like emotion detection or part-of-speech (POS) tagging.

2. Literature review

Text data generated by users on social media platforms like Twitter, Facebook, or TikTok have some specific characteristics that are not usually found in literary texts or other standard documents such as articles, announcements, news, etc. These are referred to as "bad language" in the literature and include slang, jargon, non-standard abbreviations, grammatical mistakes, and an informal tone of expression (Eisenstein et al., 2023). Moreover, these texts are of a short nature due to the limitations imposed by social media platforms (for example Twitter has a limit of 280 characters for each tweet). Dealing with such issues is a mandatory and complex endeavor in any NLP task over social media content (Barriere et al., 2020).

The work of Dat Quoc Nguyen et al. (2020) titled "BERTweet: A pre-trained language model for English Tweets" is one of the biggest contributions in adapting a transformer to the unique characteristics of Twitter. BERTweet was created using public English tweets with the goal of providing better performances for downstream tasks like Sentiment Analysis (SA), Named Entity Recognition (NER), and Topic Modeling (TC). This paper showed how flexible this types of models are while also underscoring the importance of pretraining for unconventional textual contexts.

Other efforts were made to create custom BERT or RoBERTa models that could be used more successfully in niche domains and/or in underrepresented languages as a response to the limitations that come with generic pretrained models. Some of these limitations can be solved by fine tuning different architectures on data belonging to the target domain and by applying a pre-processing methodology that better suits the specialized tasks or linguistic features that are under study. For example, Beltagy et al. (2019) demonstrated that domain specific pretraining yields a better accuracy for custom named entity recognition and relation extraction in the field of biomedical text analysis. In the legal document understanding area, BERT-based models achieved decent results when tasked with parsing and analyzing contracts as shown in the work of Chalkidis et al. (2020). The more recent work of Conneau et al. (2019) introduced XLM-RoBERTa, a crosslingual language model trained on a multilingual corpus, and demonstrated its utility in a number of common NLP tasks in multiple languages.

With these advancements, researchers are now able to better capture the underlying linguistic patterns from texts but, as highlighted in Wei et al. (2021), augmenting the training data might still be required in some cases to address data scarcity and to increase resilience against noise/outliers.

3. Data

Both private institutions and researchers have a different number of options when it comes to the acquisition of data for training BERT models. The most accessible resources are of course the public text corpora which can be accessed and used by any actor to achieve practical or research goals in many domains. As a first example we can mention the Common Crawl[1] dataset which is a collection of web-scraped texts from the digital space and includes a variety of different languages and topics. Another important resource is the OpenAI WebText dataset[2] which is one of the biggest repository

[1] https://registry.opendata.aws/commoncrawl

[2] https://paperswithcode.com/dataset/webtext

of clean data often used for language understanding and generation (Radford et al., 2019). These datasets have been and still are extensively used to tweak language models, leading to a constant flow of improvements for the AI field.

If the available datasets don't meet some preset requirements then one may opt to collect their own dataset and contribute to the field by making it accesible to others. The TweetsCOV19[3] dataset is such an example, created during the COVID-19 pandemic it contains a vast number of Twitter posts that were used to analyze what kind of feelings and opinions people had about this difficult time period (Dimitrov et al., 2020). By doing so researchers are actively increasing the ammount and diversity of resources while also encouraging collaboration and reproducibility in the NLP community.

One may end up having access to proprietary or sensitive data case in which a number of aspects should be taken into account. These datasets are a rich source of information but the people who work with them need to follow and respect ethical guidelines and regulations in order to protect the users. There are certain techniques such as Federated Learning or Differential Privacy framework that can be used to mitigate the risk of privacy violation when working with sensible data (Erlingsson et al. 2019) and such additional efforts are expected to be made to ensure safety for all parties involved in the development life cycle of the models.

3.1. Twitter Stream

Twitter Stream[4], collected by Archive Team, is a valuable public corpus that offers a huge volume of texts that were scrapped from Twitter (now rebranded to "X") and stored in JSON format. This repo represents a testament of social media discourse and can be used for historic or other types of research as it covers all the years starting from 2012 until the middle of 2021, split into 2,900 files that amount to ≈ 6.8 TB of data.

The exact number of tweets in this dataset is not specified, but by considering the long-time frame that it covers plus the size of the documents we can say with a high degree of certainty that Twitter Stream should satisfy a large range of objectives. Researchers, private or public institutions could use this data to analyze trending topics, public sentiments, cultural or social events, and more in real time or in retrospect to answer questions about the dynamics of modern societies. Longitudinal studies can also be done to see how the writing style evolves over time because this archive contains the creation timestamp metadata of each post.

As opposed to other resources this one is not limited to include only tweets in internationally popular languages, such as English or Spanish, because in the web-scraping process the majority of public posts were collected, regardless of their language. Thus, we'll use the Twitter Stream to pretrain our custom BERT models as it captures the evolution of Romanian and the way it is used in a microblogging context.

However, Twitter texts are short and informal in nature, being filled with "bad language" elements such as slang, emojis, URLs, and hashtags which are common in the vast majority of online social platforms. These modern "flavours" in the way people communicate online are also found in Romanian texts and this is another reason for why

[3] https://data.gesis.org/tweetscov19/

[4] https://archive.org/details/twitterstream

we consider Twitter Stream to be suitable for the creation of one or more robust RoBERTa models that can address these issues in order to improve the performance of various down-stream NLP applications like sentiment analysis and topic classification.

3.2. Methodology

Given the size of the entire tweet archive, totaling close to 7 TB of raw data, our limited hardware dictates a need for data selection in order to train multiple versions of RoBERTa models in a reasonable timeframe. With respect to this, we decided to only use a subset of Twitter Stream that encompasses approximately 800 GB of data spanning over the course of one year: July 2020 through June 2021.

Another factor that made us take this action is related to the fine-tuning tasks that are going to be made on the newly created RoBERTa models, namely Sentiment Analysis and Topic Classification, and by acknowledging this constraint we want to state that our aim is to establish a Proof of Concept (POC) that can demonstrate the feasibility of training BERT based models on Romanian social media texts using a relatively small dataset. By doing this we'll most likely achieve lower performances when compared to using the whole archive as the training set, but our ultimate goal is to show that it's possible for researchers to create decent models in cases where there are strong hardware or time limitations. In future iterations, if additional computational resources become available to us, we would like to integrate the rest of the data in the pretraining pipeline to learn even more powerful models.

As a first step we downloaded the data belonging to the target period mentioned in the previous paragraphs after which we performed a manual inspection to familiarize with the structure and nature of it. The data is presented in JSON documents that contain two different types of instances, one denoting the removal of content from the platform and includes the ID of the deletion plus some other metadata but without any other textual information. The other type, referred to as "post", contains a lot of information but of interest to our study are the "text" field, which represents the tweet message, and the "lang" field, which indicates the language of the message.

Next, we selected and examined in more detail 200 random posts from a two-month period. During this we discovered a major problem with "lang": a number of tweets were labeled as Romanian when in reality they weren't. Many were simply misclassified, in some extreme cases as a very different language like Malay, and others were pure "noise" posts that only contained a mix of Twitter mentions, hashtags, URLs, and emojis which makes them unusable for our study. This highlights the problems that can appear when dealing with online user generated content where the informal tone of communication, errors in grammar, and other irregularities are degrading the accuracy of automated language identification tools.

Following this initial investigation we decided to use Python[5] together with *langid*[6] to correctly identify the language of the posts. We selected this library because it has been trained on a large number of languages (currently supporting 97 in total) which makes it a good choice for our multilingual dataset and additionally it offers very fast processing times paired with state-of-the-art results. Another advantage of lanid compared to other algorithms is that it offers a "confidence level" score for each prediction that acts as a measure of reliability.

[5] https://www.python.org/

[6] https://pypi.org /project/langid/

We ran langid on the same subset of 200 tweets using a high threshold approach in which we consider the texts to be Romanian only if the confidence level exceeds 95% to avoid the incorporation of false positives in our corpus. We made a second review of the language classification and saw that most of the texts were labeled correctly this time around but some outliers still persisted.

The overall performance on the raw texts can be considered satisfactory but to have even better results we decided to implement a preprocessing pipeline that includes the automatic identification and removal of URLs, Twitter mentions, Twitter hashtags, and emoticons from the tweets. With this mechanism in place we want to deliver cleaner and more standardized texts to langid in the hope of improving the accuracy.

We ran langid once again but this time on the cleaned data and performed another round of investigations. The results were clearly better which means that the proportion of tweets correctly labeled as Romanian has increased, thus validating our custom language identification framework. Due to this we'll make use of this preprocessing pipeline in the later stages of model development when needed.

Table 1 shows that over a period of 12 months we identified and extracted around 51,000 tweets posted in Romanian, which means that we have ≈ 4,250 tweets for each month on average. Because of preprocessing and language identification the total execution time for this extraction process has very high, totaling to over 72 hours.

Table 1: Comparison of Romanian tweets

Year-Month	Number of texts labeled as Romanian in Twitter Stream	Number of texts labeled as Romanian with our approach	Percent Romanian	Execution Time (Hours)
2020-07	48415	4256	8.8	6.4
2020-08	56292	5100	9.06	7.7
2020-09	59346	4729	7.97	7.3
2020-10	57778	4788	8.27	7.5
2020-11	48867	4406	9.02	5.5
2020-12	52896	4935	9.33	6.1
2021-01	22621	1771	7.83	2.5
2021-02	56163	4621	8.23	6.21
2021-03	57993	5210	8.98	6.7
2021-04	24149	2095	8.68	2.7
2021-05	58576	4702	8.03	7.2
2021-06	52475	4330	8.25	6.3

The size of this data might seem pretty modest, but it does align with the low number of Romanian Twitter users. With Statista[7] as source we found out that the number of Twitter users in Romania was around 600,000 during the time period targeted by us. It is also important to note that not all users make their posts public and additionally some accounts might have privacy settings in place. These aspects together with certain geographical or other restrictions mean that part of the generated content may have been overlooked/skipped during the scraping of Twitter Stream.

[7] https://www.statista.com/forecasts/1143811/twitter-users-in-romania

As a short summary, even though that our dataset has 51,000 tweets and could be viewed as small at a first glance we argue that it's sufficient to provide a relevant snapshot of the activity of Romanian speakers on Twitter. In the next sections we will use this dataset to train from scratch a number of BERT models but other researchers can employ it for any other type of purposes.

4. BERTweetRO

Researchers and private institutions have realized how important linguistic diversity is and the need to have solutions capable of addressing the challenges of less popular languages that face a scarcity of digital resources such as training sets or custom lexicons/dictionaries. Thus, in the last period of time an increasing number of efforts in pretraining transformer based models for underrepresented languages has been observed. There is also the option of creating multilingual models that cover several languages, an approach that brings very good results, but in some use cases it has been found that monolingual models fine-tuned on specific down stream tasks may offer superior performances (Velankar et al., 2022).

For Romanian several studies have tackled the task of creating language models by leveraging the transformer architecture together with large scale datasets to increase the level of automated language understanding and generation. Here we can mention the works of Dumitrescu et al. (2020) who introduced the first purely Romanian transformer-based language model which outperformed Multilingual BERT in the NER task and Masala et al. (2020) who created RoBERT using random texts crawled from the internet and formal texts from Romanian Wikipedia pages.

4.1 Variants

We want to develop 8 distinct RoBERTa variants in total and the motivation behind this is based on the linguistic diversity and complexity of Romanian as well as the varying preprocessing steps that might be needed in some NLP applications. The factors for investigation that we consider to be most important are: text case sensitivity, custom text preprocessing, and the number of tokens. We cover all these aspects in order to increase our chances of finding a model that is truly capable of handling social media texts in real life applications.

BERTweetRO model variants:

- Raw Cased
- Raw Uncased
- PreProcessed (PP) Cased
- PreProcessed (PP) Uncased
- Min Tokens Raw Cased
- Min Tokens Raw Uncased
- Min Tokens PreProcessed (PP) Cased
- Min Tokens PreProcessed (PP) Uncased

The first four variants from the list (Raw Cased, Raw Uncased, PreProcessed Cased, and PreProcessed Uncased) differ from one another in the preprocessing steps and text case handling. Raw Cased preserves the original casing, Raw Uncased

converts all characters to lowercase while the PP Cased and PP Uncased variants transform the data by removing all the URLs, Twitter mentions and hashtags, emoticons, and platform reserved keywords; with the former keeping the original text case and the latter converting to all lowercase. These are the main contenders for our experiments that will allow us to see what impact (if any) case sensitivity and preprocessing has on the models created in this fashion.

The next four variants (Min Tokens Raw Cased, Min Tokens Raw Uncased, Min Tokens PreProcessed Cased, and Min Tokens PreProcessed Uncased) are similar to the first ones, the difference being that in these cases we exclude the tweets that have less than five tokens/words from the dataset. With this filtration we want to remove as many noisy instances as possible from the training set in the hope of improving the predictive power of the models.

In the end we'll compare these variants against each others to find the ones that deliver the best results.

4.2 Tokenizer training

Lexical tokenization is the process in which a text is transformed on a semantic or syntactic basis into a number of meaningful lexical tokens that belong to a predefined category. A common category employed for this is the part of speech which includes nouns, verbs, adjectives, punctuation marks, etc. The tokenization used in the case of transformers or large language models is similar to lexical tokenization but differs in two ways. First of all, lexical tokenization is based usually on a lexical grammar while LLM tokenizers use probability approaches. Secondly, LLM tokenizers include an additional procedure in which textual tokens are transformed to numbers (Alfred et al., 2007).

In other words, tokenization can be seen as the bridge that connects the natural representation of the texts used as inputs and the numerical values that encode the information such that it can be used by machine learning models. Besides breaking down text into individual components, like words or sub-words, tokenizers are also tasked with assigning an unique ID to each token in order to increase processing speeds.

The simplest word based tokenizers are the Bag-Of-Words (BoW) model (Zhang et al., 2010) which uses a representation of the text in the form of a list of unordered tokens meaning that it disregards word order, and thus most of the syntax, but captures multiplicity and Term Frequency–Inverse Document Frequency (TF-IDF) which is an improvement over BoW because this method can measure the importance of a word to a document from a collection of documents adjusted to the fact that some words appear more frequently over the entire corpus (Leskovec et al., 2020). More complex algorithms such as Word2Vec (Goldberg et al., 2014) or fastText (Athiwaratkun et al., 2018) can capture the meaning of words based on the context of other words in their proximity and for these reasons, they use a multidimensional encoding in which each token is represented by a distinct vector, but this adds complexity to model training and interpretation. Each tokenizer has its own advantages and disadvantages in terms of vocabulary size, sub-word granularity, execution speeds, which means that researchers must choose and/or adjust the right method depending on task requirements, text format, and language characteristics.

Another important tokenizer is the Byte Pair Encoding (BPE) algorithm which encodes string of texts into a tabular form and it's commonly used in various downstream modeling tasks (Gage et al., 1994). A modification to the original algorithm was made

allowing it to combine tokens that encode both single characters, including single digits or punctuation marks, and full words (Brown et al., 2020). In this case, all unique characters are considered to be an initial set of 1-character long n-grams. Next, the most frequent adjacent pairs of characters are merged to create new 2-character long n-grams and all instances of previous pairs are replaced by this new token. This process is repeated until a vocabulary of a predetermined size is reached. This version of BPE is very often set as the encoding method of LLMs and transformers. In contrast, the standard BPE doesn't merge the most frequent pair of bytes of data but instead replaces them with a new byte that was not seen in the initial dataset (Paass et al., 2023).

Due to the popularity and effectiveness of BPE we decided to apply it in our work with the help of the ByteLevelBPETokenizer implementation from HuggingFace[8] library.

To train each variant of BERTweetRO Tokenizer we selected the following parameter configuration:

- Vocabulary size of 16,000 tokens
- Minimum frequency threshold of 2
- A set of special tokens containing <s>, <pad>, </s>, <unk>, and <mask>

The special tokens used in the training process have the following meaning: <s> marks the beginning of a sequence and is used when models require a clearly defined starting point for the input sequences, <pad> is a padding token used to ensure that all sequences have the same size without adding any meaningful content and it is necessary because the sequences can have variable sizes but the models expects them to have the same size, </s> marks the end of a sequence and is used when the models require clearly defined ending points for the input sequences, <unk> is used to represent words or subwords that are not in the tokenizer's vocabulary to handle unknown inputs, and <mask> is used in the Masked Language Modeling (MLM) pretraining process where a number of tokens from the sequence are replaced with this value in order to train the model to predict the original token (task also known as "fill in the blanks").

The creation of the tokenizers consists in training them to transform the corpus of Romanian tweets in a number of ways that matches our target data variants: Raw Cased, Raw Uncased, PP Cased, PP Uncased, Min Tokens Raw Cased, Min Tokens Raw Uncased, Min Tokens PP Cased, and Min Tokens PP Uncased. During this process the BPE algorithm discovers and learns statistical patterns based on the input texts and iteratively updates its vocabulary to capture as much information as possible for each subword unit. The resulting 8 tokenizers models were then saved future usage.

4.3 Model training

To successfully learn our RoBERTa models for Romanian text processing we selected an internal configuration that can yield good performances in relation to the training times and we integrated the previously trained tokenizers with the eight variants of RoBERTa in a consistent way to ensure that the hyperparameters and the end-to-end system allows for a fair comparison of performances in the downstream tasks. We decided to use the approach called Masked Language Modeling (MLM),

[8] https://huggingface.co/

implemented with the help of Hugging's RobertaForMaskedLM, which is a pre-training technique that enables transformers to predict masked tokens from input sequences. This is done without the need for labeled data making it an unsupervised learning method and unlike other traditional algorithms that can only predict the next token in a given sequence MLM can use both the previous and following tokens to predict a masked one. Thus, the models that use MLM can better understand the context that surrounds each word and it was found that more diverse training objectives are generally better for overall model behaviour (Tay et al., 2022).

The architectural specifications of our RoBERTa models are as follows:

- Hidden size of 768
- 12 attention heads
- 12 hidden layers
- MLM probability of 15%

This selection of parameters was made in such a way as to balance predictive performance with computational efficiency in the hope that the models can still capture complicated patterns while remaining manageable for running on our hardware. The reason not all tokens are masked is to avoid the dataset shift problem that arises when the distribution of tokens seen during training differs greatly from inference. The vocabulary size is different for each individual model, being set by the associated tokenizer to ensure compatibility.

All variants were trained over 5 epochs as we observed that it's sufficient to lead to an acceptable level of convergence without costing too much in terms of execution time. A larger number of epochs could incrementally improve the performance but we decided against this in order to avoid the risk of overfitting. The masked language model probability of 15% is in line with the recommendations in the literature (He et al., 2020; Levine et al., 2020; Izsak et al., 2021), based on the reasoning that models can't learn good representations when too much text is masked and the training is inefficient when too little is masked. If the training set is extremely big (not the case for our experiments) then higher percentage values should be considered (Wettig et al., 2022).

The models were trained on our GPU with a batch size of 16 and the total execution time for all 8 variants was a little under 4 hours which is decent if we consider the high computational overhead that is expected when creating transformer models from scratch.

5. Fine-tuning

Fine tuning is a transfer learning approach in which the parameters of a pretrained model are adjusted on a new dataset in order to refine or enrich its functionality (Zhang et al., 2021). This can be done on the entire neural network or only on a subset of layers in which case the layers that are not fine-tuned are "frozen" i.e. they are not changed in the backpropagation step. A model can also be augmented with the help of "adapters" that contain a much smaller number of parameters compared to those of the original model and thus it's fine tuned in a more efficient way because the initial weights remain the same (Liu et al., 2022).

For some architectures like CNNs it is common to keep the first layers (the ones closest to the input layer) frozen as they have the role of capturing low level features while the last layers are fine-tuned because these often discern high level features that are more related to the task that the model is trained on (Zeiler et al., 2014). Large scale models that have been pretrained on extensive corpora are usually fine-tuned by reusing the original parameters as a starting point, on top of which task specific layer(s) that are trained from scratch are added. The alternative of fine-tuning the whole system is also an option that usually delivers superior results but it's more computationally demanding due to the larger number of parameters that must be adjusted to the downstream task (Dingliwal et al., 2021).

In the case of BERT, a pretrained model is used as the feature extraction module with the aim of capturing the general linguistic representations while the newly added task specific layers are trained on labeled data to handle the target assignment. In the works of Devlin et al. (2019) and Liu et al. (2019) fine tuning general language models to become specialized in other tasks is highlighted as an easy way of bringing new capabilities to existing models, or improve their current performance, with minimal additional training data. The most common NLP applications targeted by fine tuning are text classification, named entity recognition, part of speech tagging, and machine translation.

5.1. Sentiment Analysis

Sentiment analysis, also known as opinion mining or emotion AI, is the use of natural language processing, text analysis, computational linguistics, and biometrics to automatically identify, extract, quantify, and study affective states and subjective information. It is widely employed in various domains such as reviews and survey responses, online and social media, and healthcare materials. The application of sentiment analysis ranges from marketing and customer service to clinical medicine. With the rise of transformer-based models and deep language models, more difficult data domains can be analyzed, such as texts where authors typically express their opinion/sentiment in a less explicit fashion (Hamborg et al., 2021).

A popular sentiment analysis task is to classify the overall polarity of a given text into representative categories like "negative", "neutral", or "positive". Depending on the context and requirements this can be done at the document, sentence, or feature/aspect level. The advantage of this approach relies on its simplicity and clear categorization process which makes it easy to understand and apply in practice. Its disadvantage is that it cannot capture nuanced emotions and to overcome this limitation the "beyond polarity" sentiment classification can be used. In this case more subjective emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise (Ho et al., 2020) are considered. In some fields, i.e. mental health analysis, this method can provide a deeper and better understanding of the emotions that are expressed in written texts. Due to its complexity and subjectivity this technique works best when advanced models are employed and when richer datasets, which are hard to get, are used for training.

To fine tune for this scenario new layers need to be added on top of pretrained BERT or RoBERTa model after which the entire architecture is trained in a supervised manner on a dataset annotated with sentiments. This allows the model to learn and identify sentiment related features from the data and to make predictions

on never seen before texts based on the learned patterns. The quality of the model depends, among others, on the volume of data, the optimization process, and the number of iterations used during training.

When talking about underrepresented languages, such as Romanian, we can specify the work of Ciobotaru et al. (2023) in which the authors trained a fastText-based model and fine-tuned a standard BERT-based model then compared their performances.

They selected a public dataset which contains COVID-19 related Twitter posts split in 2 categories (negative and positive). Next, they built upon this dataset by adding the "neutral" sentiment class and by adding more text samples to all classes. This new dataset was used as the benchmark in their experiments and the reported results on the test set showed that BERT achieved a macro F1 score of 0.84 while fastText had a worse score of 0.73. An additional temporal study over a period of 21 weeks was made in which the authors observed that the general opinion about COVID-19 vaccination changed from positive to negative. The last half of this time period also generated more debate among users resulting in a serious increase in the number of posts made.

We searched in a number of online platforms including academic databases and NLP repositories (like Kaggle) but couldn't find a dataset that could match our requirements. In this work we want to apply a multinomial sentiment analysis on Romanian social media content with negative, neutral, and positive as the polarity classes. We did find some review datasets that contain sentiments about products (Briciu et al., 2024; Istrati et al. 2021) but these are not suitable for us due to the obvious differences between review data and social media posts. Also it is worth mentioning that most of these works only offer a binary analysis of sentiment (negative vs positive).

Thus, we decided to employ an open-source English dataset, translate it to Romanian using an automated translation service and use it as a "surrogate" resource in our experiments (Neagu et al., 2022).

For this research, we selected the Twitter US Airline Sentiment Tweets dataset[9], collected in 2015 and each tweet was manually labeled by external contributors with its global sentiment polarity (positive, negative and neutral). This data contains approximately 15,000 tweets with a class distribution as follows: 63% negative, 21% neutral, and 16% positive. Each tweet is also accompanied by the contributor's confidence about the annotated sentiment and each negative tweet includes a reason for the assessment.

Next we conducted a series of experiments using the newly translated Romanian dataset together with all of our 8 MLM RoBERTa variants. With Hugging Face's BertClassifier we fine-tuned each model for sentiment analysis using the correct tokenizer as the encoding mechanism. We split the data into training and test sets with the training dataset consisting of approximately 11,000 instances while the testing dataset consisting of the remaining 3,700 instances, thus providing a standard 75-25% train-test split. This separation was made such that the class distribution between the train and test data remained similar. Moreover, the English and Romanian train and test data are identical in the sense that they contain the same set of instances. Depending on each model variant the associated preprocessing pipeline was executed in the same manner it was used in the pre-training stage in order to ensure that a fair comparison between the models can be made later on.

[9] https://www.kaggle.com/crowdflower/twitter-airline-sentiment

We will evaluate the performance of our 8 fine tuned variants of RoBERTa models in a comparative study in which we include a number of traditional classifiers from the area of classic machine learning and deep learning. The classic models are paired with TF-IDF encoding and comprise of Bernoulli Naive Bayes (NB), Support Vector Machine with a linear kernel (Linear SVM), Random Forest (RF), and Logistic Regression (LR). The selected deep neural network architectures are Deep Neural Network (DNN) with TF-IDF encoding, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) network both of them using Word2Vec as the encoding method.

We note here that all the traditional classifiers from our comparative study underwent a rigorous hyperparameter optimization process, with the help of evolutionary algorithms, in order to maximize their performance. Evolutionary optimization was selected because it can reach an adequate combination of values for the hyperparameters very quickly (Pelikan et al., 2002) and it has been show to outperform other approaches like Bayesian optimization (Mori et al., 2005). Another advantage of this technique is that the optimization can be done in all three types of search spaces (continuous, discrete, and categorical) regardless of the classifier on which the optimization is performed. Also there are many research works related to the metaheuristic design of neural networks (Ojha et al., 2017) in which the parameters or even the architectural structure (Bochinski et al., 2017) of deep learning models were identified with the help of genetic algorithms (Tani et al., 2021). Thus, it is important to highlight that the fine tuned models will be compared with these popular classifiers which have achieved peak performance on the selected dataset.

Additionally, we also compare the performance of our variants against Hugging Face's Multilingual BERT model which was fine-tuned on our translated dataset but without hyperparameter optimization due to the high execution times that are required. For this process we selected standard parameters and the associated Multilingual Tokenizer was used as the encoder as it can handle a vast number of languages, including Romanian. With this benchmark we want to offer valuable insights into the performance of our smaller RoBERTa models relative to a large scale pre trained multilingual language model that is widely used by researchers and private companies.

For our custom variants we added an extra sequential layer for classification that can handle the output of the pretrained layers and the expected sentiment class labels. As in the case of Multilingual BERT, our variants were not subjected to the hyperparameter optimization process due to time constraints. Instead, the parameters used for fine-tuning were chosen based on standard industry recommendations but also by considering the initial parameters that were used to create the models: batch size of 32, BERT hidden size of 768, classification hidden size of 75, a max token length of 80, Rectified Linear Unit (ReLU) as the activation function, and categorical cross-entropy as the loss function. The number of epochs, ranging from 2 to 10, that offers a decent level of accuracy was investigated and identified individually for each RoBERTa variant as well as for Multilingual BERT.

The results of our comparative analysis are presented in Table 2 in which the models are evaluated strictly on the test sets using 3 metrics often used in the literature (Macro F1, Weighted F1, and Accuracy) to allow us to assess the performance of the models from different points of view. Accuracy is the most reported score in research works because it shows what percentage of total predictions made are correct. The standard F-measure is more complex, computed as the harmonic mean

between precision and recall. Weighted F1-score is a variation of this metric in which a weight is added to the predictions based on their distribution percentage with the goal of assigning a greater contribution for the classes that have more instances (Raschka et al., 2018). On the other hand, Macro F1 is computed as the arithmetic mean of class wise F-scores thus, it treats all classes as equally important no matter their frequency. In the case of imbalanced class labels Macro F1 can better measure performance because it's more punishing for the models that regularly misclassify under represented instances (Ganganwar et al., 2012).

Given the unbalanced nature of our datasets we set Macro F1 as the main measure of predictive performance. By using doing this we want to make sure that our evaluation treats each class in an equal manner in order to offer a correct interpretation of model effectiveness across all sentiments. Therefore we filtered Table 2 based on the Macro-F1 scores in descending order which means that the best results are at the top of the table.

Table 2: Sentiment analysis performance

Classifier	Encoding	Macro F1	Weighted F1	Accuracy
Multilingual BERT	Multilingual Tokenizer	74.81	80.50	80.99
BERTweetRO Raw Cased	BERTweetRO Tokenizer Raw Cased	72.11	78.40	78.74
BERTweetRO Raw Uncased	BERTweetRO Tokenizer Raw Uncased	72.07	78.33	78.66
Bernoulli NB	TFIDF	71.91	78.20	78.20
BERTweetRO Raw Min Tokens Cased	BERTweetRO Tokenizer Raw Min Tokens Cased	71.67	78.14	78.61
BERTweetRO Raw Min Tokens Uncased	BERTweetRO Tokenizer Raw Min Tokens Uncased	71.58	78.00	78.47
LSTM	Word2Vec	71.39	77.98	78.17
Linear SVM	TFIDF	70.54	77.47	78.36
DNN	Word2Vec	69.19	76.23	77.20
Logistic Regression	TFIDF	69.04	76.45	77.81
CNN	Word2Vec	68.67	76.00	77.69
BERTweetRO PP Min Tokens Cased	BERTweetRO Tokenizer PP Min Tokens Cased	64.21	73.10	72.86
BERTweetRO PP Cased	BERTweetRO Tokenizer PP Cased	43.84	59.40	64.50
BERTweetRO PP Uncased	BERTweetRO Tokenizer PP Uncased	42.35	58.73	64.01
Random Forest	TFIDF	38.17	54.71	65.20
BERTweetRO PP Min Tokens Uncased	BERTweetRO Tokenizer PP Min Tokens Uncased	25.62	47.98	62.42

We can see that Multilingual BERT outperformed all the other classifiers by having higher scores across all considered metrics but it's important to note that our best performing RoBERTa variants, namely BERTweetRO Raw Cased and BERTweetRO

Raw Uncased, achieved a similar performance. The differences between them and Multilingual BERT are fairly small with Macro F1 around 3% lower and the other 2 metrics around 2% lower. This outcome was somewhat expected if we take into account the difference in the scale of data used for pre-training as Multilingual BERT benefited from a much larger volume and diverse data whereas our variants were trained on a considerable smaller dataset (around 51,000 tweets).

Surprisingly, the BERTweetRO variants that were trained on texts with a minimum token constraint (Raw Min Tokens Cased and Uncased) also had competitive results that are only slightly below of those obtained by the 2 variants that used all the instances from the training set. This means that by limiting the data used to train the models, i.e. keeping only the texts with more than five tokens, the predictive performance that can be achieved is not reduced in a significant manner and additionally this can improve to some degree the execution speeds. Bernoulli NB, despite its simplicity, obtained good results placing it in between these 4 variants. With this exception the BERTWeetRO models that did not used the text preprocessing pipeline performed better than all the classic and deep learning models.

Another thing that we want to highlighted is the fact that regardless of the BERTweetRO variant, the ones that were trained on the data with the original text case (a.k.a. the Cased variants) have marginally better results than the equivalent variants that were trained on the data in which all characters have been converted to lower case (a.k.a. the Uncased variants). In the middle of the ranking we have LSTM, Linear SVM, DNN, Logistic Regression, and CNN with decent performances, making them suitable for usage in real scenarios.

At the bottom of the table, where the models with the worst results are placed, we have all the BERTweetRO variants that were paired with our custom text preprocessing module which clearly shows that this process negatively affected their predictive performances. This means that better BERT based models can be developed by simply pretraining and fine tuning them on raw social media data without the need for additional text cleaning or feature engineering and for this reason we want to warn other researchers about the risks of extensive preprocessing in contexts similar to ours. These models together with Random Forest had by far the lowest predictive performance meaning that they cannot be considered for sentiment analysis.

After this study we fine tuned BERTweetRO Raw Cased and BERTweetRO Raw Uncased on the whole US Airline Tweets dataset and saved them for future applications.

5.2 Topic Classification

Discovering abstract topics that occur in a collection of texts or documents could be done with either Topic Classification or Topic Modeling. Topic modeling is an unsupervised technique (Blei et al., 2012; Vayansky et al. 2020) that doesn't require labeled data, while topic classification is a supervised one, where labeled data is needed for model training.

Topic modeling is a widely used statistical tool for extracting latent variables from large datasets, being well suited for textual data. Among the most used methods for topic modeling we can mention Probabilistic Latent Semantic Analysis (PSLA) and Latent Dirichlet Allocation (LDA) which state that a document is a mixture of topics, where a topic is considered to convey some semantic meaning by a set of

correlated words, typically represented as a distribution of words over the vocabulary. Statistical techniques are then used to learn the topic components (topic-to-word distributions) and mixture coefficients (topic proportions) of each document. In essence PSLA, LDA, and other conventional topic models reveal topics within a text corpus by implicitly capturing the document-level word co-occurrence patterns (Boyd et al. 2008).

However, directly applying these models on short texts will suffer from the severe data sparsity problem, i.e. the sparse word co-occurrence patterns found in individual document (Hong et al., 2010). Some workarounds try to alleviate the sparsity problem with Albanese and Feuerstein (2021) aggregating a number of short texts to create a lengthy pseudo-document, its effectiveness being heavily data dependent. The Biterm Topic Model (Cheng et al., 2014) extracts unordered word pairs (i.e. biterms) occuring in short texts and the latent topic components are then modeled using these biterms. This method seems to perform better for short texts compared to other traditional approaches.

The main advantage of topic modeling methods is that they do not require labeled data, thus data collection becomes more accessible and could be done in a fully or partially automated manner. Despite its popularity, topic modeling is prone to serious issues with optimization, noise sensitivity, instability which can result in data which is unreliable (Agrawal et al., 2018), and some techniques are not representative of real-world data relationships (Blei et al., 2006). This is usually due to strong assumptions regarding key parameters in the calculation process and the inefficiency of many optimization methods (which often attempt to overcome uncertainty by performing many time-consuming iterations to determine the best parameters). For example, setting the optimum number of topics to be extracted is not trivial and human intervention is needed in order to set a relevant topic label to each identified topic. This is done based on the representative key words or phrases belonging to each topic.

If labeled training data is available then topic classification can overcome most issues related to the unsupervised nature of topic modeling by categorizing the texts into a number of predefined classes based on the subjects/themes in them. In this case, machine learning algorithms treat topic classification as a regular text classification problem: having a set of training records/instances $D = \{X_1, X_2, \ldots, X_n\}$, where each record X_i represents a data point (i.e. document, paragraph, sentence, word) and is labeled with one of k distinct topic labels, the purpose is to build models that are capable of identifying text patterns based on the training records in order to predict the topics of never seen before texts with reliable accuracy rates. Unlike topic modeling, topic classification is easier to understand and evaluate which makes the assessment and comparison of models more straightforward.

Learning models on a small dataset with around 770 tweets distributed over 18 classes, Lee et al. (2011) achieved an accuracy of ≈65% with the multinomial Naive Bayes classifier and ≈62% with the standard SVM classifier. In another study, Rahman and Akter (2019) worked with 6,000 texts extracted from Amazon's product review corpus[10] distributed over only 6 very specific topics and achieved a very high classification rate of approximately 92% with NB, 82% with k-NN, and 79% with decision trees.

[10] https://jmcauley.ucsd.edu/data/amazon/

Zeng et al. (2018) proposed a hybrid approach that combines topic modeling with topic classification. They first extracted the most relevant latent features with topic modeling and then fed them into supervised algorithms like SVM, CNN, and LSTM. For the experiments they used the Twitter dataset released by TREC2011[11], which contains around 15,000 tweets, semi-automatically labeled into 50 topic classes. The highest accuracy of ≈9.5% was achieved by the CNN architecture and can be considered modest at best. Furthermore, they conclude that the topic modeling component did not improve the learning capabilities of the classifiers in any significant way.

Regarding topic classification for Romanian texts the existing research is more limited. Here we can mention the work of Vasile et al. (2014) who evaluated the capabilities of some classic machine learning models when applied to blog content. The data used in this study was extracted from 219 blogs, each instance being labeled with 1 topic class from a total of 9: "Activism", "Business and Finance", "Art", "Travel", "Gastronomy", "Literature", "Fashion", "Politics", and "Religion and Spirituality". The Sequential Minimal Optimization (SMO) and Complement Naive Bayes (CNB) performed the best, both reaching an accuracy of around 77.8%. A lower accuracy of 73.3% was achieved by k-Nearest Neighbors (k-NN) while the standard Naive Bayes (NB) had the worst accuracy of only 68.9%. Important to note that the authors used a very small dataset in their experiments which is problematic because it's unlikely that these results can be reproduced on larger evaluation sets.

We couldn't find any other relevant research works that target Romanian social media content. An explanation for this might be the unusual traits (Barriere et al., 2020; Eisenstein et al., 2013) of microblogging texts which pose a lot of problems for traditional NLP systems. Additionally, labeled datasets are also missing which means that we'll have to translate a suitable English dataset in order to create the training data needed for our topic classification experiments.

For this reason we selected the News Category Dataset[12] which contains 202,372 news headlines collected between 2012 up to 2018 from HuffPost[13], formerly The Huffington Post until 2017, an American news aggregator and blog with localized and international editions. The site offers news, satire, blogs, original content, and covers a variety of topics like politics, business, entertainment, technology, popular media, and more. Each record of the dataset contains the following attributes: *category* (41 categories), *headline*, *short_description*, *authors*, *date* (of the publication), and *link* (URL link of the article). There are a number of reasons why we selected this dataset as the benchmark for our experiments: (i) it contains short texts similar to those found on social media platforms, (ii) the topics are fairly general and the number of topics is large enough, (iii) the category of each article was manually labeled, (iv) high data volume, and (v) it was relatively recently collected.

For our classification problem we will focus only on the headline and short description attributes of the dataset, ignoring the authors and date of publication. Therefore, we merged the headline and the short description attributes and created a novel attribute named *text_merged*. The vast majority of merged texts contain between 94 and 254 characters, with the mean being ≈174 and the standard deviation almost

[11] http://trec.nist.gov/data/tweets

[12] https://www.kaggle.com/datasets/rmisra/news-category-dataset

[13] https://www.huffpost.com/

80 characters. This proves that the generated texts have the characteristics of short texts similar to those present in social media platforms (i.e. a Twitter tweet is limited to 280 characters).

We did an initial investigation on this data and encountered some problems with the distribution and granularity of the original 41 class labels. The top-3 most popular classes are: “POLITICS” which contains ≈16% of the records, “WELLNESS” which contains ≈9% of the records, and “ENTERTAINMENT” which contains ≈8% of the records. The least most popular 4 classes are: “COLLEGE”, “LATINO VOICES”, “CULTURE \& ARTS”, and “EDUCATION” each containing around 0.5% of the records, meaning that there is a significant class imbalance in the data.

Besides this imbalance, we also noticed that there are two inconsistencies related to the existing topics: a subset of them are overlapping while others are highly granular. For example, the categories “SCIENCE” and “TECH” are too specific and could be represented under a single label called “SCIENCE & TECH” while other classes have different labels but denote the same thing, for example “ARTS & CULTURE” and “CULTURE & ARTS”. To address these issues we decided to refine the labels of the dataset by clustering together overly granular and synonymous categories. At the end of this process the revised dataset contains 26 topics that are truly distinct and no class has less than 1% of record labels, meaning that the least popular class has more than 2,000 records. This should increase the performance of the models that will be trained later but at the same time it ensures consistency and coherence in our topic classification task. The new class feature was named *category_merged*. For additional details about this reconstructing process the readers are referred to (Neagu et al., 2023).

Fine tuning in this context involves the use of an existing pre-trained model and adapting it to classify the inputed texts based on the discussion topics they convey. This is done by adding task specific layers on top of the BERT or RoBERTa model after which the entire architecture is trained in a supervised manner on the annotated dataset. This allows the model to learn and identify topic related features from the data and to make predictions on never seen texts with the help of the learned representations. The ability of the model to recognize the abstract themes in text data depends, among others, on the volume of data, the optimization process, and the number of iterations during training.

Next, similar to Sentiment Analysis fine-tuning, we conducted a series of experiments using the newly translated Romanian dataset together with all 8 variants of our pretrained MLM RoBERTa models. With Hugging Face’s BertClassifier we fine tuned each model for topic classification using the correct tokenizer but before doing this we split the data into training and testing sets. The training set contains 75% of instances while the testing test contains the remaining 25%. This split was made such that the class distribution between training and testing remained similar. Moreover, the English and Romanian train and test data are identical in the sense that they contain the same set of instances. Depending on each model variant the associated preprocessing pipeline was applied in the same manner it was used in the pretraining stage in order to ensure that a fair comparison between the models can be made later on.

We'll evaluate the performance of our eight fine tuned variants of RoBERTa models in a comparative study in which we include a number of traditional classifiers from the area of classic machine learning and deep learning. The classic models are paired with TF-IDF vectors and comprise of Bernoulli Naive Bayes (NB), Support Vector Machine with a linear kernel (Linear SVM), and Random Forest (RF). The selected deep neural network architectures are Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) both of them using Word2Vec embeddings. As in the case of Sentiment Analysis, these models underwent a hyperparameter optimization process with the help of genetic algorithms. This means that our fine tuned variants will be compared against models that have peak performance on the selected dataset.

Additionally, we compare the performance of our variants against Hugging Face's Multilingual BERT model which was fine tuned on our translated dataset but without hyperparameter optimization due to the high execution times that are required. Instead we used standard parameters and the associated Multilingual Tokenizer as the encoding mechanism as it can handle a vast number of languages including Romanian. With this benchmark we want to offer valuable insights into the performance of our smaller RoBERTa variants relative to a large scale pretrained multilingual language model that is widely used today.

For our custom variants we added an extra sequential layer for classification on top of the existing architecture that can handle the output of the pre-trained layers and the expected topic labels. As in the case of Multilingual BERT our variants were not subjected to the hyperparameter optimization process due to time constraints. Instead the parameters used for fine tuning were selected based on industry recommendations but also by considering the initial parameters that were used to create the models from scratch: batch size of 32, BERT hidden size of 768, classification hidden size of 128, a max token length of 120, Rectified Linear Unit (ReLU) as the activation function, and categorical cross-entropy as the loss function. The number of epochs, ranging from 2 to 10, which leads to an acceptable level of accuracy was investigated and identified individually for each RoBERTa variant as well as for Multilingual BERT.

Unlike sentiment analysis, where the goal is to detect a text's global polarity, the difficulty of topic classification resides also in the big number of target classes which often overlap (Gentzkow et al., 2019; Liu et al., 2020). To overcome this issue, some authors (Gupta et al., 2014; Oh et al., 2017) use the Top-K accuracy instead of the standard one. Rather than classifying a text into a single class and comparing it to the a-priori label, the model will predict the *K* most probable classes and if the correct label is among them, we consider the text as being correctly classified. In our work we take this into account and report the standard accuracy (i.e. Top-1), as well as Top-2 and Top-3, evaluated strictly on the test set.

Table 3 shows the results of our comparative study which are filtered using the Top-1 accuracy in descending order meaning that the best models appear at the beginning. Here we can see Multilingual BERT in the first place with impressive Top-1, Top-2, and Top-3 accuracies of 72.63%, 85.56%, and 90.25% respectively. This result was expected if we consider the huge volume of data on which this transformer model was pretrained, allowing it to generate robust initial representations, even in the Romanian language, which are then easily adjusted for topic classification with the help of our translated dataset.

Table 3: Topic classification performance

Classifier	Encoding	Top-1 Acc.	Top-2 Acc.	Top-3 Acc.	Opt. (s)	Train (s)	Test (s)
Multilingual BERT	Multilingual Tokenizer	72.63	85.56	90.25	N/A	7498	157
Linear SVM	TFIDF	66.73	80.05	85.30	11803	45.91	0.042
BERTweetRO Raw Uncased	BERTweetRO Tokenizer Raw Uncased	66.14	79.21	84.93		8436	135
BERTweetRO Raw Min Tokens Uncased	BERTweetRO Tokenizer Raw Min Tokens Uncased	66.07	79.10	84.80		6899	157
BERTweetRO Raw Min Tokens Cased	BERTweetRO Tokenizer Raw Min Tokens Cased	65.78	79.02	84.79		6923	157
BERTweetRO Raw Cased	BERTweetRO Tokenizer Raw Cased	65.63	78.75	84.48		8442	132
Bernoulli NB	TFIDF	62.80	77.70	84.11	398	0.59	0.04
CNN	Word2Vec	61.66	74.05	79.28	36797	56.98	1.65
BERTweetRO PP Min Tokens Cased	BERTweetRO Tokenizer PP Min Tokens Cased	54.55	66.60	72.94		6046	137
LSTM	Word2Vec	53.50	65.59	72.39	63605	119.1	6.16
Random Forest	TFIDF	16.56	28.89	37	845	0.6	0.183
BERTweetRO PP Min Tokens Uncased	BERTweetRO Tokenizer PP Min Tokens Uncased	16.56	28.89	37		6019	135
BERTweetRO PP Cased	BERTweetRO Tokenizer PP Cased	16.56	28.89	37		6027	135
BERTweetRO PP Uncased	BERTweetRO Tokenizer PP Uncased	16.56	28.89	37		6022	135

In the race for the runner up position we have several models with similar scores across all three evaluation metrics, namely Linear SVM and the four Raw BERTweetRO variants. These classifiers reached Top-1 accuracies between ≈65% and ≈66%, which are good enough for real life applications. Note that as previously mentioned Linear SVM benefited from a complex process of hyperparameter optimization, conducted on a high performance computer, while the BERTweetRO variants were fine tuned with default parameters on standard hardware. As in the case of Sentiment Analysis, the BERTweetRO variants that didn't use the custom text preprocessing pipeline obtained better results than the other four variants that did. The difference between our best variants and Multilingual BERT is around 6% but can be considered

decent given the relative small size of the data we used for pretraining (around 51,000 texts) with respect to the much larger corpora that was used to pretrain Multilingual BERT. This means that by employing a richer dataset and by applying hyperparameter optimization we could potentially enhance the performance of the BERTweetRO variants in future iterations.

BERTweetRO Raw Min Tokens Uncased and Cased also have competitive performances, as they had in Sentiment Analysis, which reconfirms that by restricting the data that is used for fine tuning, i.e. only including the texts with more than five tokens, the predictive performance that can be achieved is not downgraded and at the same time this can improve (to some degree) the execution speeds when training and executing the models.

Bernoulli NB and CNN are next, with slightly lower scores when compared to the previous models, both having a similar performance when talking about Top-1 accuracy but in the case of Top-2 and Top-3 CNN lags behind by a pretty noticeable margin. These classifiers should behave more or less the same when predicting the main topic of a text but Bernoulli NB is to be preferred if one considers the second and third most probable topics as being important to their use case.

BERTweetRO PP Min Tokens Cased and LSTM share fourth place in our ranking with modest results across all considered metrics. At the bottom of the table we have Random Forest and the four BERTweetRO variants that incorporated the text preprocessing module, all of them having by far the worst predictive performance, with a Top-1 accuracy of only 16.5%. This once again underscores the effectiveness of simply pretraining and fine tuning BERT models on raw social media text data without the need for any text cleaning or feature engineering. For this reason we want to warn other researchers about the risks of extensive text preprocessing for similar NLP tasks in which microblogging data is involved, as this may lead to worse results. Hence, this group of models are not viable for topic classification in production environments.

After these experiments the BERTweetRO Raw Cased and BERTweetRO Raw Uncased variants were fine tuned on the entire News Category dataset to increase their generalization power and saved for future use.

6. Assessing and comparing Sentiment Analysis performance on real cases

Given that the final purpose of our work is to apply the learned models for inferring the polarity of any Romanian tweet, we manually labeled two small test sets, each one containing 120 distinct tweets. The first one includes tweets specific to the airline industry, comparable with the ones used for training our models, and the second one includes general tweets. We will evaluate on these test sets the best performing models as reported in previously in this work: Multilingual BERT, BERTweetRO Raw Uncased, Bernoulli NB, LSTM, and DNN. Additionally, we'll compare these models against a public third-party sentiment analysis tool for Romanian called Sentimetric[14] to see where we stand in relation to a commercially available solution.

The tweets were manually labeled by five human volunteers who were trained in advance on how this process should be carried out. Each volunteer expressed an opinion about the polarity of the tweet and the final sentiment was set as the one

[14] http://sentimetric.ro/

selected by the majority. Labeling statistics regarding how they assessed the polarity is presented in Table 4. We shall note that the labeling task seemed to be a difficult one even for the volunteers, as for only 43 tweets (35.8%) in the case of airline industry specific dataset and 47 tweets (39.2%) in the case of general tweets all of the 5 contributors reached a unanimous decision. Furthermore, the class distribution of these tweets is significantly different from that of the Twitter US Airline Sentiment Tweets (presented in the last row of the table).

Table 4: Manual labelling statistics of tweet polarity, including the number of tweets and corresponding percentages

Dataset	Negative	Neutral	Positive	Unanimous Annotation
Airline industry tweets	51 (46.5%)	36 (30%)	33 (27.5%)	43 (35.8%)
General tweets	45 (37.5%)	32 (26.6%)	43 (35.8%)	47 (39.2%)
Twitter US Airline Sentiment Tweets	63%	21%	16%	

As in the case of the fine tuning experiments we report Macro F1, Weighted F1, and Accuracy as the evaluation metrics for each classifier but seeing the imbalanced distribution of labels of both dateset we again have to set Macro F1 as the main measure of model performance.

Table 5 contains the predictive performance of our target models on the 120 real life Romanian tweets that relate to the aviation industry. In this case we can see that Bernoulli Naive Bayes (NB) achieved the highest Macro F1 score of 61.18% and in second place, with a marginally lower score, we have Multilingual BERT. This result is a little surprising considering that Multilingual BERT had better results on the evaluation set used in the fine tuning section but the success of NB could be attributed to the hyperparameter optimization process that it went through.

Table 5: Model performance on Romanian aviation industry-specific tweets

Classifier	Encoding	Macro F1	Weighted F1	Accuracy
Bernoulli NB	TFIDF	61.18%	63.11%	65%
Multilingual BERT	Multilingual Tokenizer	60.45%	63.38%	65.83%
BERTweetRO Raw Uncased	BERTweetRO Tokenizer Raw Uncased	54.57%	56.68%	60%
LSTM	Word2Vec	52.71%	55.18%	58.33%
DNN	Word2Vec	52.22%	54.9%	59.17%
Sentimetric		45.72%	46.99%	47.5%

The test data includes a small number of samples but despite this our BERTweetRO Raw Uncased variant managed to secure an honorable third place across all evaluation metrics. Even thought that it failed to surpass Bernoulli NB and Multingual BERT its performance is better than the deep learning LSTM and DNN models. The Macro F1 of 54.5%, which is around 6% lower than the best score, can be considered acceptable given that the humans volunteers also had difficulties when labeling the texts.

The most important thing that we want to highlight here is that all of our models outperformed Sentimetric. This shows the positive impact of using a custom methodology for training and validating ML models when compared to off the shelf solutions. It also confirms the value of domain specific knowledge for obtaining better results in such contexts as we fine tuned our models on tweets from the same domain.

In Table 6 we present the performance of the classifiers on the Romanian general tweets dataset, i.e. tweets that don't belong to a single specific industry. In this case things are a little different as Multilingual BERT achieved the best result with a Macro F1 of 55.22% followed closely by BERTweetRO Raw Uncased with a negligible difference in score of only 1%. In both this assessment and the previous one, the transformer models placed at the top which means that they're more reliable for sentiment analysis in practice.

Table 6: Model performance on Romanian general tweets

Classifier	Encoding	Macro F1	Weighted F1	Accuracy
Multilingual BERT	Multilingual Tokenizer	52.22%	54.17%	55.85%
BERTweetRO Raw Uncased	BERTweetRO Tokenizer Raw Uncased	51.35%	52.39%	54.17%
Bernoulli NB	TFIDF	48.48%	49.42%	48.33%
DNN	Word2Vec	48.16%	49.29%	50.83%
Sentimetric		46.16%	47.3%	49.17%
LSTM	Word2Vec	43.17%	44.29%	45.83%

On the other hand, Bernoulli NB and the DNN architecture have more modest results that place them in the middle of the ranking but more surprising is that LSTM delivered a significantly worse performance in this case, being behind all the other models, including the solution offered by Sentimetric. The reasons to why this happened requires future investigations but it's possible that the complexity of the neural network together with its sensitivity to the shape of input data could have affected its ability to correctly recognize the sentiment patterns from these samples.

An interesting detail that we want to point out is that the overall performance of the models on these generic tweets is lower compared to the aviation industry. This decline, which is more obvious for the classic and deep learning models, is a direct result of domain differences between the texts used for training and the ones used for evaluation. For Multilingual BERT and BERTweetRO the decline is less serious due to the fact that they were pretrained on varied data and thus managed to better adapt in this scenario.

For both domains our models' results are lower than those obtained on the translated test set because now the tweets are real ones, not translated, and their inherent characteristics differ, i.e. from a statistical point of view the sets are extracted from different statistical populations.

Nonetheless, as in the case of the first evaluation set, we note that all our models (with the exception of LSTM) have significantly outperformed the commercial solution that was selected as the benchmark for comparison. This once again validates the importance of custom fine tuning and model optimization in achieving superior results for Romanian sentiment prediction.

7. Discussion and further work

As a first idea that we consider for future work directions is the hyperparameter optimization for our BERTWeetRO models for both Sentiment Analysis and Topic Classification. Although our initial experiments that used the default parameters generated promising results, we could considerably increase the predictive performances on these downstream tasks by applying a well thought optimization process. The parameters that we'd like to explore and adjust are: learning rate, dropout rate, batch size, number of neurons per layer, and number of layers.

Other aspects that might be worth to investigate are different text preprocessing steps and the usage of data augmentation algorithms to better clean and enrich our labelled datasets. By doing this the classifiers would have access to additional examples in the training phase which in turn should increase their capability of understanding the meaning of texts based on the context of the words within them. We could also employ different types of encodings and tokenizers to see how these impact the performance and execution times of the models.

In this paper we presented the fine-tuning methodology on two popular NLP tasks but this can be relatively easily extended to other commonly requested tasks in the industry. A good candidate would be Named Entity Recognition (NER) for Romanian texts, which consists in identifying a set of entities and classifying them into categories such as names of people, names of organizations, locations, calendar dates, etc. By fine-tuning our models for NER we can offer another NLP functionality that has direct or indirect application in various domains/systems like entity linking, information extraction, and semantic search.

And last but not least, we would like to increase the size of our pretraining repository by adding more Romanian tweets that cover a wider range of linguistic patterns, expressions, and discussion topics to allow our transformer models to reason with a higher level of generalization which would directly improve their behaviour on any downstream task. The classifiers' accuracy may also be improved by incorporating new annotated datasets for SA and TC in the training or fine tuning stage and this can be done by simply translating English datasets as demonstrated in our study.

8. Conclusions

In conclusion, our work offers a number of contributions for the NLP of social media content in Romanian, a language which is highly under resourced in this area. First of all, we identified and curated an open source repository that contains public tweets posted from July 2020 up to, and including, June 2021. These tweets can be used by anyone who was an interest for studying the characteristics of Romanian conversations in a microblogging space. Using this data we pretrained from scratch 8 different variants of BERTweetRO models, all based on the RoBERTa MLM architecture, and their corresponding text tokenizers that can be used on either raw and preprocessed texts.

Next, we selected the following NLP downstream tasks to fine tune our models on: (i) Sentiment Analysis, which refers to the classification of texts into 3 polarity classes (negative, neutral, and positive) based on the feelings they express, and (ii) Topic Classification, which refers to the classification of texts into 26 distinct and generic discussion topics. We couldn't find any annotated datasets suitable to our specific research needs so we decided to apply an automated translation service on two English

datasets to create the equivalent resources in Romanian. We then fine-tuned our BERTweetRO variants and the popular Multilingual BERT on the translated datasets and aimed to achieve the highest predictive performance possible by finding the optimal number of epochs for each model.

We implemented a comprehensive test bed in which the transformer models were compared against a number of well-known classic and deep learning ML algorithms that were trained for the same tasks using the same datasets. The results of these experiments show that Multilingual BERT is indeed the best option but some of our BERTWeetRO models achieved comparable performances thus highlighting their potential for improvement in the future with the help of hyperparameter optimization or data augmentation. Bernoulli NB, Linear SVM, and CNN also had good overall results, which means that they can be employed in practice, especially when computational resources are limited.

An important finding we want to emphasize is that the text preprocessing steps had a serious negative impact on the performance of the BERTweetRO variants that used them. Thus, we want to warn others about the risks of extensive preprocessing for similar applications where social media texts are involved.

In the end we collected and manually labeled with sentiment polarity two sample sets of real life Twitter posts, one containing tweets specific to the aviation industry and the other containing generic tweets. Then we executed on these datasets our best performing BERTweetRO variant together with Multilingual BERT and evaluated their predictive performances against a commercial classifier called Sentimetric. In both domains our models delivered better results therefore validating our custom methodology for developing language models.

Our pre-trained BERTweetRO models, together with the variants fine-tuned for Sentiment Analysis and Topic Classification, are open source and can be accessed online[15].

References

Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modelling? And how to fix it using search-based software engineering. *Information and Software Technology*, 98, 74-88.

Albanese, F., & Feuerstein, E. (2021). Improved topic modelling in twitter through community pooling. In *String Processing and Information Retrieval: 28th International Symposium*, SPIRE 2021, Lille, France, October 4–6, 2021, Proceedings 28 (pp. 209-216). Springer International Publishing.

Alfred, V. A., Monica, S. L., & Jeffrey, D. U. (2007). Compilers principles, techniques & tools. *pearson Education*.

Athiwaratkun, B., Wilson, A. G., & Anandkumar, A. (2018). Probabilistic fasttext for multi-sense word embeddings. *arXiv preprint arXiv*:1806.02901.

Barriere, V., & Balahur, A. (2020). Improving sentiment analysis over non-English tweets using multilingual transformers and automatic translation for data-augmentation. *arXiv preprint arXiv*:2010.03486.

[15] https://huggingface.co/dan-neagu

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv*:1903.10676.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Blei, D., & Lafferty, J. (2006). Correlated topic models. *Advances in neural information processing systems*, 18, 147.

Bochinski, E., Senst, T., & Sikora, T. (2017). Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3924-3928). IEEE.

Boyd-Graber, J., & Blei, D. (2008). Syntactic topic models. *Advances in neural information processing systems*, 21.

Briciu, A., Călin, A. D., Miholca, D. L., Moroz-Dubenco, C., Petrașcu, V., & Dascălu, G. (2024). Machine-Learning-Based Approaches for Multi-Level Sentiment Analysis of Romanian Reviews. *Mathematics*, 12(3), 456.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *arXiv preprint arXiv*:2010.02559.

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 2928-2941.

Ciobotaru, A., & Dinu, L. P. (2023). SART & COVIDSentiRo: Datasets for Sentiment Analysis Applied to Analyzing COVID-19 Vaccination Perception in Romanian Tweets. *Procedia Computer Science*, 225, 1331-1339.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv*:1911.02116.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, volume 1 (long and short papers) (pp. 4171-4186).

Dimitrov, D., Baran, E., Fafalios, P., Yu, R., Zhu, X., Zloch, M., & Dietze, S. (2020). Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic. In *Proceedings of the 29th ACM international conference on information & knowledge management* (pp. 2991-2998).

Dingliwal, S., Shenoy, A., Bodapati, S., Gandhe, A., Gadde, R. T., & Kirchhoff, K. (2021). Prompt Tuning GPT-2 language model for parameter-efficient domain adaptation of ASR systems. *arXiv preprint arXiv*:2112.08718.

Dumitrescu, S. D., Avram, A. M., & Pyysalo, S. (2020). The birth of Romanian BERT. *arXiv preprint arXiv*:2009.08712.

Eisenstein, J. (2013). What to do about bad language on the internet. In *Proceedings of the 2013 conference of the North American Chapter of the association for computational linguistics: Human language technologies* (pp. 359-369).

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., & Thakurta, A. (2019). Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 2468-2479). Society for Industrial and Applied Mathematics.

Gage, P. (1994). A new algorithm for data compression. *The C Users Journal*, 12(2), 23-38.

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, 2(4), 42-47.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535-574.

Goldberg, Y., & Levy, O. (2014). word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv*:1402.3722.

Gupta, M. R., Bengio, S., & Weston, J. (2014). Training highly multiclass classifiers. *The Journal of Machine Learning Research*, 15(1), 1461-1492.

Hamborg, F., Donnay, K., & Merlo, P. (2021). NewsMTSC: a dataset for (multi-) target-dependent sentiment classification in political news articles. *Association for Computational Linguistics (ACL)*.

He, P., Liu, X., Gao, J., & Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv*:2006.03654.

Ho, V. A., Nguyen, D. H. C., Nguyen, D. H., Pham, L. T. V., Nguyen, D. V., Nguyen, K. V., & Nguyen, N. L. T. (2020). Emotion recognition for vietnamese social media text. In *Computational Linguistics: 16th International Conference of the Pacific Association for Computational Linguistics*, PACLING 2019, Hanoi, Vietnam, October 11–13, 2019, Revised Selected Papers 16 (pp. 319-333). Springer Singapore.

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88).

Istrati, L., & Ciobotaru, A. (2022). Automatic monitoring and analysis of brands using data extracted from twitter in Romanian. In Intelligent Systems and Applications: *Proceedings of the 2021 Intelligent Systems Conference* (IntelliSys) Volume 3 (pp. 55-75). Springer International Publishing.

Izsak, P., Berchansky, M., & Levy, O. (2021). How to train BERT with an academic budget. *arXiv preprint arXiv*:2104.07705.

Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011). Twitter trending topic classification. In *2011 IEEE 11th international conference on data mining workshops* (pp. 251-258). IEEE.

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of massive data sets. *Cambridge university press*.

Levine, Y., Lenz, B., Lieber, O., Abend, O., Leyton-Brown, K., Tennenholtz, M., & Shoham, Y. (2020). Pmi-masking: Principled masking of correlated spans. *arXiv preprint arXiv*:2010.01825.

Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., & Raffel, C. A. (2022). Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems*, 35, 1950-1965.

Liu, X., He, P., Chen, W., & Gao, J. (2019). Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *arXiv preprint arXiv*:1904.09482.

Masala, M., Ruseti, S., & Dascalu, M. (2020). Robert–a romanian bert model. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6626-6637).

Mori, N., Takeda, M., & Matsumoto, K. (2005). A comparison study between genetic algorithms and bayesian optimize algorithms by novel indices. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation* (pp. 1485-1492).

Neagu, D. C., Rus, A. B., Grec, M., Boroianu, M. A., Bogdan, N., & Gal, A. (2022). Towards sentiment analysis for romanian twitter content. *Algorithms*, 15(10), 357.

Neagu, D. C., Rus, A. B., Grec, M., Boroianu, M., & Silaghi, G. C. (2022). Topic Classification for Short Texts. In *International Conference on Information Systems Development* (pp. 207-222). Cham: Springer International Publishing.

Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020). BERTweet: A pre-trained language model for English Tweets. *arXiv preprint arXiv*:2005.10200.

Oh, S. (2017). Top-k hierarchical classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Ojha, V. K., Abraham, A., & Snášel, V. (2017). Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, 60, 97-116.

Paaß, G., & Giesselbach, S. (2023). Pre-trained Language Models. In *Foundation Models for Natural Language Processing: Pre-trained Language Models Integrating Media* (pp. 19-78). Cham: Springer International Publishing.

Pelikan, M., Goldberg, D. E., & Lobo, F. G. (2002). A survey of optimization by building and using probabilistic models. *Computational optimization and applications*, 21, 5-20.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140), 1-67.

Rahman, M. A., & Akter, Y. A. (2019). Topic classification from text using decision tree, K-NN and multinomial naïve bayes. In *2019 1st international conference on advances in science, engineering and robotics technology* (ICASERT) (pp. 1-4). IEEE.

Raschka, S. (2021). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv 2018. arXiv preprint arXiv*:1811.12808.

Tani, L., Rand, D., Veelken, C., & Kadastik, M. (2021). Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics. The European Physical Journal C, 81, 1-9.

Tay, Y., Dehghani, M., Tran, V. Q., Garcia, X., Wei, J., Wang, X., ... & Metzler, D. (2022). Ul2: Unifying language learning paradigms. *arXiv preprint arXiv*:2205.05131.

Vasile, A., Rădulescu, R., & Păvăloiu, I. B. (2014). Topic classification in Romanian blogosphere. In *12th Symposium on Neural Network Applications in Electrical Engineering* (NEUREL) (pp. 131-134). IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.

Velankar, A., Patil, H., & Joshi, R. (2022). Mono vs multilingual bert for hate speech detection and text classification: A case study in marathi. In *IAPR workshop on artificial neural networks in pattern recognition* (pp. 121-128). Cham: Springer International Publishing.

Wei, J., Garrette, D., Linzen, T., & Pavlick, E. (2021). Frequency effects on syntactic rule learning in transformers. *arXiv preprint arXiv*:2109.07020.

Wettig, A., Gao, T., Zhong, Z., & Chen, D. (2022). Should you mask 15% in masked language modeling?. *arXiv preprint arXiv*:2202.08005.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference*, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 (pp. 818-833). Springer International Publishing.

Zeng, J., Li, J., Song, Y., Gao, C., Lyu, M. R., & King, I. (2018). Topic memory networks for short text classification. *arXiv preprint arXiv*:1809.03664.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). Dive into deep learning. *Cambridge University Press*.

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International journal of machine learning and cybernetics*, 1, 43-52.

Zhao, J., Liu, K., & Xu, L. (2016). Sentiment analysis: Mining opinions, sentiments, and emotions.

STUDIA UNIVERSITATIS BABEȘ-BOLYAI OECONOMICA
VOLUME 70, ISSUE 1, 2025, pp. 112-125
DOI: 10.2478/subboec-2025-0006

WEALTH EFFECTS IN THE CEE EMERGING ECONOMIES: A LONG-RUN PANEL APPROACH

Florina BURDET[a,b]*
[a] Babeș-Bolyai University, Romania
[b] Université Clermont Auvergne, France

Abstract: This paper explores the relationship between actual individual consumption, housing wealth, and stock wealth in a panel of emerging European Union economies. Using the pooled mean group estimator (PMG) and a crisis dummy variable, the analysis captures the effects of the 2008–2010 financial crisis. Results indicate that both housing and stock wealth positively influence consumption, with housing wealth having a slightly stronger impact. Consumption is also sensitive to long-term changes in the income from wages and salaries. Particularly during the crisis, when real estate market downturns significantly affected households in Central and Eastern Europe (CEE), the effect of changes in asset prices raised in magnitude. The findings offer important policy implications for managing asset price effects on household consumption, particularly in emerging markets.

JEL classification: D12, E21, E52

Keywords: actual individual consumption, housing wealth, stock wealth, panel cointegration analysis

1. Introduction

The link between consumption and wealth has been a key area of study for decades. This relationship gained prominence through J. M. Keynes's General Theory of Employment, Interest and Money in 1936. Subsequent theories, such as Friedman's Permanent Income Theory (1957) and the Life Cycle Theory proposed by Modigliani and Brumberg in 1954, and later by Ando and Modigliani in 1963, further developed this area. In summary, these theories suggest that changes in wealth, if they permanently alter household resources, should lead to similar changes in consumption, although of a lesser magnitude, compared to changes in income.

* Corresponding author. Address: Faculty of Economics and Business Administration, Babes-Bolyai University, 58-60 Teodor Mihali Street, 400591 Cluj-Napoca, Romania, E-mail burdet.florina@gmail.com and Université Clermont Auvergne, Laboratoire d'Economie d'Orléans (LEO-UCA), 26 Avenue Léon Blum, 63000 Clermont-Ferrand, France.

The wealth effect is one of the several ways in which a financial crisis may influence the real economy. It refers to changes in consumption that arise when households perceive their wealth to have increased or decreased due to fluctuations in asset prices. When households feel wealthier, they tend to spend more, thus affecting aggregate demand and GDP.

This wealth effect is often viewed as a psychological response: rising asset prices encourage higher spending. Studies by Poterba (2000) and Cheng and Fung (2008) indicate that capital market fluctuations can directly influence consumption by affecting individuals' confidence and expectations about future economic conditions. Furthermore, uncertainty in capital markets can significantly impact consumption trends. However, this study focuses solely on the effects of changes in wealth without considering expectations about future wealth.

It is important to recognise that not all types of wealth impact consumption similarly. Economic theory suggests that more liquid assets lead to stronger consumption responses when their value increases, as households feel more confident in raising their spending levels (Poterba, 2000; Jappelli & Pistaferri, 2014).

This paper provides new empirical evidence on the relationship between household consumption, employee compensation, and wealth while also considering the impact of the global financial crisis. The analysis focuses on emerging countries in Central and Eastern Europe (CEE), examining wealth effects in a disaggregated manner (financial and non-financial wealth). The results challenge much of the existing literature. This study is particularly relevant for policy, as asset price booms and busts in CEE countries have been more pronounced over the past two decades than in developed economies (Ahec-Sonje, 2012; Posedel and Vizek, 2009).

The contributions of this paper lie in (i) studying the specific sample of CEE emerging economies over a more extensive time span and (ii) incorporating a crisis dummy variable for the period 2008-2010 to assess the effects of the 2008 Lehman Brothers collapse. This collapse significantly impacted households' consumption, income, and wealth, raising inquiries about households' financial resilience and implications for monetary policy. The crisis dummy period also encompasses the year 2020 to consider the effects of the COVID-19 pandemic, which ultimately resulted in a stock market crash (Baker et al., 2020).

The paper is organized as follows: Section 2 provides a review of the literature on the consumption-wealth relationship in emerging economies. Sections 3 and 4 outline the data and methodology used. Section 5 presents the main findings, and Section 6 concludes with policy implications and suggestions for further research.

2. Literature review

Several empirical studies have explored the relationship between wealth and consumption using various approaches and datasets. A significant portion of these studies focus on macroeconomic data and measure wealth effects using time series methods, especially co-integration techniques (Lettau and Ludvigson, 2001, 2004; Catte et al., 2004). For example, Lettau and Ludvigson (2001, 2004) showed that co-integration can yield consistent estimates of parameters linking consumption, labour income, and wealth in a linear framework.

Most empirical research focuses on the effects of total or financial wealth on US consumer spending, largely because of the available detailed databases and the predominant stock wealth in American households. Funke (2004) was the first study to examine emerging economies (including 16 countries in Asia, Latin America, and Africa). The study found a small but statistically significant wealth effect, though it focused only on stock market wealth and excluded European countries. Compared with this study, more recent works (i.e. Apergis et al., 2018; Singh, 2022) make use of at least two distinctive types of wealth, namely stock wealth and housing wealth, distinction that will be clearly followed in the empirical exercise of this paper.

A recurring finding in the literature is the positive long-term relationship between wealth and consumption, though the strength and nature of this relationship vary depending on asset type, measures utilised as proxies for wealth and methods of estimation (Catte et al., 2004, Jawadi and Sousa, 2014; Apergis et al., 2018;). Ciarlone (2011), for instance, showed that both financial and housing wealth positively affect household consumption in emerging economies in Asia and Central and Eastern Europe (CEE), with housing wealth having a larger elasticity, employing a cointegration-based method, however, over a period that ended in 2009. Similarly, Peltonen et al. (2012) found significant wealth effects, with financial wealth playing a more dominant role in countries with higher market capitalization. In the latter case, it is imperative to mention that the sample was comprised of 14 non-European emerging economies, which faced a different financial market development compared to their European counterparts, which benefited more through integration with Western Europe (Nardo et al, 2022; Bakaert et al., 2023).

Vizek (2011) observed similar long-term consumption responses to stock price changes in Bulgaria, Croatia, and the Czech Republic, though housing price shocks led to divergent consumption reactions across these countries. Rosenberg (2015), focusing on Estonia, which saw dramatic housing market changes during the global financial crisis, concluded that real estate price shocks had a positive and lasting impact on private consumption.

More recent work by Ceh-Casni (2018) on European countries found that housing wealth has a stronger effect on consumption than financial wealth, especially before the global financial crisis. Nicolau (2020) reported similar findings for a panel of CEE countries, noting that the housing wealth effect is more pronounced.

However, despite the general trend of housing wealth having a stronger effect in bank-based economies, the subject has yet to be explored, accounting for other economic factors, besides their economic status (Ahec-Sonje and Ceh-Casni, 2014; Singh, 2022). This may stem from difficulties in finding accurate proxies for different types of wealth, making comparisons between emerging and developed economies challenging. For instance, Rodil and Menezes (2016) found that financial wealth had a more significant effect than housing wealth in 10 Eurozone countries, with financial wealth having a positive and significant effect during the global financial crisis, while housing wealth had a negative effect during the same period. This motivates the inclusion of a crisis dummy in our model as a robustness check.

It is important to note that changes in residential property prices and consumption seem to have become more closely linked in the last decade, particularly following the 2008 financial crisis (Ciarlone, 2011). More recent evidence suggests that this linkage has strengthened further, especially in the post-crisis period, with housing wealth playing a more substantial role in shaping household

spending decisions (Angrisani et al., 2018; Cloyne et al., 2018). For many emerging countries, residential property represents households' largest asset (Rosenberg, 2015), making the wealth effect of house price changes more substantial than other asset price changes.

Recent studies indicate that the elasticity of consumption to house price fluctuations has increased, with consumers now spending more out of housing wealth than in previous decades (Berger et al., 2017). Although Buiter (2008) argues that house price changes may influence consumption through the credit channel by easing collateral constraints, Muellbauer (2008) cautions that these effects may simply reflect wealth redistribution and may not significantly alter aggregate wealth. For instance, Christelis et al. (2021) suggest that the consumption response to house price changes is more pronounced in economies with higher mortgage market penetration and greater financial inclusion, reinforcing the role of credit conditions in shaping the wealth effect.

3. Data

To model the relationship between consumption, income, housing wealth, and stock wealth, we examine a panel of 11 emerging economies from Central and Eastern Europe (CEE), members of the European Union: Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia. The sample was selected based on similarities among these countries, such as the high proportion of fully owned homes, a key factor in our analysis (Kaas et al., 2019). We use quarterly data covering the period from 1998Q1 to 2020Q4, yielding up to 84 observations per country. Details on the available data, variables, and descriptive statistics in logarithmic form are presented in Table 1 below, as well as in Tables A.1. and A.2. in the Appendix.

Table 1: Descriptive statistics for variables in logarithmic form

Variable	Observations	Mean	Std. Dev.	Min.	Max.
Actual Consumption	924	4.516	0.235	3.712	5.024
Wages and salaries	924	4.499	0.302	3.567	5.228
House price index	605	4.699	0.163	4.188	5.231
Stock price index	863	4.615	0.461	3.079	6.149

Source: Author's estimations in STATA 17.
Note: The values displayed are descriptive statistics for the series in initial form.

As a dependent variable, we used actual individual consumption without distinguishing between durable and non-durable consumption. To measure the immediate income effect more accurately, we use income from wages and salaries, which reflect more accurately labour market conditions and individual earning capacity (Blundell et al, 2008) since disposable income also encompasses non-labour income (see, for instance, Vizek, 2011; De Bondt et al., 2019). Moreover, the data on disposable income were unavailable for several countries in the sample and at the needed frequency. The literature is still relatively limited in terms of using income at a more disaggregated level, one of the causes being the data availability.

For housing and stock wealth, consistent measures are lacking for emerging economies, as it is the case for data series on net value of housing and other types of assets owner by households, which are not commonly collected. This inconvenience makes houses prices and stock prices indices the most suitable proxies for housing and stock wealth, respectively. We define stock wealth more intuitively since financial wealth comprises a larger variety of financial assets apart from stocks. While asset price indices do not capture wealth-level changes, potentially leading to inconsistent results over time (Rodil and Menezes, 2016), they are commonly used due to data availability and frequency. Asset prices remain a key element in assessing wealth evolution. To ensure comparability with previous studies (Ciarlone, 2011; Ceh-Casni, 2017, 2018; Nicolau, 2020), we use house price indices instead of house values, which helps avoid measurement errors across countries.

Although some studies include the interest rate as an exogenous variable and find a negative and significant effect on consumption (Rodil and Menezes, 2016; Nicolau, 2020), we were not able to include it due to inconsistent data across the countries in our panel. However, we acknowledge that interest rates could affect consumption by making credit more expensive or by increasing financial returns for households (Rodil and Menezes, 2016).

We include a dummy variable for the global financial crisis, with a value of 1 for the period 2008Q1–2010Q4 and 0 for the pre and post-crisis periods (namely, 2000Q1–2007Q4 and 2011Q1–2019Q4). Additionally, we extend the dummy to include the COVID-19 stock market crash (2020Q1–2020Q4) and test the impact of the COVID-19 crisis on consumption. This variable captures the impact of the crisis on household consumption, reflecting the changes in economic, social, and political contexts, increased uncertainty, and credit restrictions. Additionally, this dummy acts as a time-fixed effect.

All series, except house price indices (deflated at the source), are deflated using the Harmonized Index of Consumer Prices (HICP, base year 2015=100) to express them in real terms. The X-12-ARIMA method is used for seasonally adjusting the data, except for the series already adjusted at the source, and are further transformed into logarithmic form. All series are expressed per capita. Data on household consumption, employee compensation, house prices, and HICP were sourced from Eurostat, while stock market indices were obtained from Thomson Reuters Eikon and investing.com.

4. Methodology

Recent studies on consumption and wealth effects (Rodil and Menezes, 2016; Singh, 2022) have shown growing interest in dynamic panel estimation models, particularly when both the time dimension (T) and the cross-sectional dimension (N) are relatively large. This type of panel data is well-suited for cross-sectional analyses, especially when seeking to estimate long-term effects and the speed of adjustment of consumption to long-term equilibrium.

In this study, we use the pooled mean group (PMG) estimator, introduced by Pesaran et al. (1999), which offers a flexible approach to panel data analysis. The PMG estimator is suited for pooling long-term parameters across countries while allowing short-term dynamics (such as adjustment speed) to vary at the country level. This flexibility makes it ideal for studying the heterogeneity in short-term

consumption responses across economies. The PMG estimator is built on the Auto-Regressive Distributed Lag (ARDL) model, which accounts for both serial-correlated residuals and endogeneity in the regressors by selecting appropriate lags for both dependent and independent variables. Another advantage of this method is that it allows to work with a combination of variables that can be I(1) and I(0), respectively.

To define the link between consumption, income and wealth, we specify a long-term consumption function that is identical for all countries and which, for $\boldsymbol{i} = \overline{\boldsymbol{1,N}}$ and $\boldsymbol{t} = \overline{\boldsymbol{1,T}}$, can be written as follows:

$$\boldsymbol{AIC_{i,t}} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{i,1}I_{i,t}} + \boldsymbol{\beta_{i,2}HW_{i,t}} + \boldsymbol{\beta_{i,3}SW_{i,t}} + \boldsymbol{\varepsilon_{i,t}} \quad (1)$$

where *AIC* is the actual individual consumption, *I* represents the real per capita income from wages and salaries, *HW* refers to the real per capita house prices, *FW* represents the real per capita stock prices, and $\boldsymbol{\varepsilon_{i,t}}$ represents the term error, which incorporates the effects of unexpected shocks on consumption. The notations i and t represent the country and the unit of time, respectively.

The next step is to give the previously introduced representation a dynamic structure. This approach is advised for various reasons, such as persistent habits, adjustment costs, or liquidity constraints, which support the immediate adjustment of consumption to a change in its main determinants.

The literature usually proposes estimating ARDL-type models to capture changes in consumption caused by income and wealth by introducing lags in the model. Equation (1) can be generated accordingly by indicating the deterministic variables, an auto-regressive term for the outcome variable and distributed lags for the explanatory variables. The optimal number of lags for each country in the sample is selected based on the Schwarz Information Criterion (BIC). We identify the ARDL (1, 1, 0, 0) specification for the entire country sample, and we will proceed to estimate the PMG estimator based on it, considering the first lags for consumption and income, while housing and financial wealth will be considered at their current values:

$$\boldsymbol{AIC_{i,t}} = \boldsymbol{\alpha_i} + \boldsymbol{\gamma_i AIC_{i,t-1}} + \boldsymbol{\beta_{i,10}I_{i,t}} + \boldsymbol{\beta_{i,11}I_{i,t-1}} + \boldsymbol{\beta_{i,20}HW_{i,t}} + \boldsymbol{\beta_{i,30}SW_{i,t}} + \boldsymbol{\varepsilon_{i,t}} \quad (2)$$

By re-parameterising equation (2), the model becomes:

$$\Delta\boldsymbol{AIC_{i,t}} = \boldsymbol{\alpha_i} + \boldsymbol{\varphi}\left(\boldsymbol{AIC_{i,t-1}} - \boldsymbol{\alpha_{i,1}I_{i,t}} - \boldsymbol{\alpha_{i,2}HW_{i,t}} - \boldsymbol{\alpha_{i,3}SW_{i,t}}\right) + \boldsymbol{\beta_{i,10}}\Delta\boldsymbol{I_{i,t}} + \boldsymbol{\varepsilon_{i,t}} \quad (3)$$

For the PMG estimator, the null hypothesis, tested using the Hausman test, states that the long-run relationship between the dependent and independent variables is identical across all cross-sectional units. In case the hypothesis is rejected after computing the Hausman test, it is suggested that the mean group (MG) estimator is more appropriate than the PMG estimator. The MG estimator, introduced by Pesaran and Shin (1997), provides consistent estimates of long-term coefficients; however, it may prove inefficient should the homogeneity assumption hold. In contrast, the PMG estimator assumes homogeneity in the long-term coefficients while allowing for heterogeneity in short-term adjustments. The Hausman test evaluates the difference between the PMG and MG estimators to determine whether the homogeneity assumption is valid. If the test fails to reject homogeneity, the PMG estimator is preferred due to its efficiency.

As noted by Pesaran et al. (1999), while pooled estimators are often used without testing for constraints, cross-country analyses frequently reject equality of error variances and long- and short-term coefficients at conventional significance levels. Hence, applying the Hausman test ensures the most appropriate model specification for our analysis.

5. Results and Discussion

To ensure a consistent estimation of the long-term relationship between consumption, the income from wages and salaries, and wealth, the variables must either be stationary or co-integrated in the long run.

We begin by applying several first-generation unit root tests: Levin, Lin and Chu (LLC) test by Levin et al. (2002), Im-Pesaran-Shin (IPS) test by Im et al. (2003), Hadri (2000), Breitung (2000), and the Cross-sectionally augmented Im-Pesaran-Shin (CIPS) test by Pesaran (2007), which accounts for heterogeneity in the panel.

Table 2: First- and second-generation panel unit-root test results

Test	Actual individual consumption	Wages and salaries	House price index	Stock price index
LLC	0.559	4.108	5.012	-2.658
	(0.712)	(1.000)	(1.000)	(0.004)
IPS	0.335	0.921	3.705	-0.967
	(0.631)	(0.821)	(0.999)	(0.166)
ADF-Fisher	23.081	12.918	17.491	35.621**
	(0.397)	(0.935)	(0.735)	(0.033)
ADF-Choi Z-stat	-0.817	2.017	1.158	-1.262
	(0.207)	(0.978)	(0.876)	(0.103)
Hadri Z-stat	82.407***	87.779***	62.016***	59.521***
	(0.000)	(0.000)	(0.000)	(0.000)
Breitung	0.118	3.709	0.659	0.148
	(0.547)	(0.999)	(0.745)	(0.558)
CIPS	-1.512	-1.646	-2.599	-2.205

Source: Author's estimations in STATA 17.
Note: The values displayed are the results of t-tests, and the values in parentheses represent the corresponding p-values. The tests were applied to the series in logarithmic form. H_0 is rejected for a p-value < 0.05.

The results, shown in Table 2, generally support the presence of unit roots in most variables. The stock price results may initially seem unclear but are aligned with the efficient market hypothesis, which treats stock prices as a random walk. The Hadri test further confirms non-stationarity in stock prices.

For cointegration testing, we utilize the Westerlund (2007) tests, which are suitable for shorter time periods. Table 3 indicates that the tests reject the null hypothesis of no cointegration, providing evidence of a long-term relationship between the variables.

Table 3: Westerlund (2007) panel cointegration test results

Statistic	t-stat
G_τ	-2.93***
	(0.000)
G_α	-11.692**
	(0.020)
P_τ	-8.721***
	(0.000)
G_α	-9.622***
	(0.003)

Source: Authors' estimations in STATA 17.
Note: The values displayed are the results of t-tests. Standard errors are reported in brackets. ***, **, * denote statistical significance at the 1%, 5%, and 10% respectively. The tests were applied to the series in logarithmic form. H0 is rejected for a p-value < 0.05.

Table 4 summarizes the regression results. The negative adjustment coefficients indicate a proper adjustment mechanism, ensuring a long-run balance between the variables. Both housing and stock wealth positively and significantly impact consumption, with the effect of housing wealth being marginally larger than stock wealth. As expected, the income from wages and salaries has the strongest effect on consumption due to its liquidity.

Table 4: The results of the ARDL (1, 1, 0, 0) specification

Variables	**PMG**	**MG**
Adjustment coefficient φ	-0.470***	- 0.626***
	(0.000)	(0.000)
Long-run coefficients		
Wages and salaries (lnI)	0.556***	0.617***
	(0.000)	(0.000)
House price index (lnHW)	0.066***	0.034
	(0.001)	(0.641)
Stock price index (lnSW)	0.062***	0.043
	(0.000)	(0.122)
Hausman test	3.80	
	(0.284)	
Number of countries	11	11
Number of observations	594	594
Log Likelihood	1406.495	
Short-run coefficients		
ΔlnI	0.656***	0.716***
	(0.000)	(0.000)
ΔlnHW	0.005	0.027
	(0.927)	(0.816)
ΔlnFW	-0.019	-0.021
	(0.130)	(0.175)
Constant	0.681***	0.847***
	(0.000)	(0.000)

Source: Author's estimations in STATA 17.
Note: Standard errors presented in parentheses.

The PMG results show an elasticity of consumption to housing wealth of 0.066 and to stock wealth of 0.062. While the difference is small, these results confirm most findings for emerging economies, which typically show a larger effect of housing wealth. The result may be explained by the borrowing capacity derived from housing assets. Although housing wealth is less liquid than stocks, it is typically used as collateral when accessing credit, which is then used by households to finance their consumption. This practice is widespread, especially in the CEE emerging economies, where housing is the main asset in households' portfolios.

The result of the Hausman test suggests that the homogeneity of long-term parameters holds, supporting the preference for the PMG estimator over the MG estimator. The main takeaway is that both housing and stock wealth positively affect long-term consumption, with stock wealth having a slightly lower impact. These results are robust and consistent with prior literature, where housing wealth elasticities are typically above financial wealth elasticities (Casni, 2018; Singh, 2022).

Table 5: Robustness checks - the results of the ARDL (1, 1, 0, 0) specification with crisis dummy

Variables	**Model 1**	**Model 2**
Adjustment coefficient φ	-0.470***	- 0.493***
	(0.000)	(0.000)
Long-run coefficients		
Wages and salaries (lnI)	0.556***	0.575***
	(0.000)	(0.000)
House price index (lnHW)	0.066***	0.0735***
	(0.001)	(0.000)
Stock price index (lnFW)	0.062***	0.0732***
	(0.000)	(0.000)
Crisis Dummy	-	-0.011**
		(0.028)
Hausman test	3.80	6.44
	(0.284)	(0.168)
Number of countries	11	11
Number of observations	594	594
Log Likelihood	1406.495	1714.46
Short-run coefficients		
ΔlnI	0.656***	0.390***
	(0.000)	(0.000)
ΔlnHW	0.005	-0.031
	(0.927)	(0.637)
ΔlnFW	-0.019	-0.020
	(0.130)	(0.119)
Crisis Dummy	-	-0.001
		(0.857)
Constant	0.681***	0.683***
	(0.000)	(0.000)

Source: Author's estimations in STATA 17.
Note: Standard errors are reported in brackets. ***, **, * denote statistical significance at the 1%, 5%, and 10% respectively.

Short-run coefficients for housing and stock wealth are not statistically significant, possibly due to delayed wealth effects stemming from sticky expectations or habit formation (Slacalek, 2009; Carroll, 2011).

Our results are in line with previous studies conducted by Ciarlone (2011), Ceh-Casni (2018), and Nicolau (2020), which also found a slightly stronger housing wealth effect. These findings confirm that actual individual consumption data could be a reliable alternative measure to private household consumption expenditures.

In addition, the model includes a crisis dummy variable (see Model 2 in Table 5) to account for the impact of the global financial crisis on consumption, income, and wealth. The crisis context creates negative expectations about the future evolution of asset prices, which may lead households to modify their consumption behaviour in anticipation of a potential decrease in their financial and housing wealth. The findings are presented in Table 5, alongside the results from the previous model, denoted as Model 1, to facilitate comparison.

After including the crisis dummy variable in Model 2, the coefficients for housing and financial wealth maintain their sign, with a slightly larger magnitude, of 0.0735 for housing wealth (compared to 0.066 in Model 1), and 0.0732 for stock wealth (compared to 0.062 in Model 1). The coefficient of the dummy variable is negative and statistically significant at a 5% level, which indicates that households adjust their consumption downwards during the period of economic distress. In the short run, the only significant coefficient in both models is the coefficient of the income from wages and salaries, while the coefficients for the two types of wealth become negative but not statistically significant.

Table 6: Alternative methods of estimation for the two specifications

Variables	**PMG**		**FMOLS**		**DOLS**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Wages and salaries	0.556***	0.575***	0.556***	0.551***	0.548***	0.546***
(lnI)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
House price index	0.066***	0.0735***	0.042**	0.053**	0.051**	0.054**
(lnHW)	(0.001)	(0.000)	(0.043)	(0.014)	(0.022)	(0.018)
Stock price index	0.062***	0.0732***	0.023**	0.023**	0.026**	0.036***
(lnFW)	(0.000)	(0.000)	(0.012)	(0.015)	(0.018)	(0.001)
Crisis Dummy	-	-0.011**	-	-0.009	-	-0.012*
		(0.028)		(0.171)		(0.092)
Number of countries	11	11	11	11	11	10
Number of observations	594	594	594	594	583	542

Source: Author's estimations in STATA 17.
Note: Standard errors are reported in brackets. ***, **, * denote statistical significance at the 1%, 5%, and 10% respectively.

We also specified the consumption-wealth relationship differently while assessing for the effect of interactions with the crisis dummy of the housing and stock prices series, respectively. However, the results proved to be not statistically significant and, as a

consequence, were not presented in the results section. The estimation results, including the interactions, are available upon request.

Both models used in our study are further subject to other two robustness checks regarding the estimation methods. We also estimated the wealth-consumption relationship through Dynamic Ordinary Least Squares (DOLS) and Fully Modified Ordinary Least Squares (FMOLS) methods to account for the long-run wealth effects. The results provided in Table 6 confirm the results from the PMG method: alternative estimation methods indicate a stronger positive housing wealth effect on consumption compared to the stock wealth effect.

It is important to notice that all estimated coefficients for each variable have the same sign, regardless of the estimation method: the income coefficients, house price coefficients and stock market coefficients are significant and positive.

6. Conclusions

This study explored the effect of changes in housing and stock wealth on consumption for a panel of 11 CEE countries that are members of the European Union. While alternative measures for wealth exist in developed countries, such as occupancy rates and market capitalisation, we relied on house price indices and stock indices as proxies, based on data availability.

Our analysis revealed that consumption, the income from wages and salaries, and both types of wealth are non-stationary and co-integrated. By employing modern econometric technique like the PMG procedure, we obtained a clear understanding of both the short-run and long-run relationships among these variables.

The results indicate a wealth effect from both the real estate and stock markets, with the impact of stock wealth being slightly smaller than that of housing wealth. These findings align with previous research on emerging economies. Notably, when incorporating time-fixed effects for the global financial crisis (2008–2010), the coefficients for housing and stock wealth keep their sign, with a slightly larger magnitude. However, direct comparisons between the results obtained in this study and the estimates found in previous literature might be challenging due to significant differences in timeframes, data frequency, country samples, variable definitions, and robustness checks.

Most existing studies suggest that changes in housing wealth, particularly house prices, constitute the primary channel through which economic cycles influence emerging economies, especially during periods of economic expansion. This raises important policy questions regarding the most effective tools to mitigate risks associated with housing market fluctuations in both boom-and-bust phases. During economic downturns, liquidity constraints become more binding, which may justify policies aimed at improving housing affordability for lower-income households, either through direct ownership support or mortgage assistance schemes. Conversely, in periods of economic expansion, rising house prices—particularly in economies with high homeownership rates, such as those in Central and Eastern Europe (CEE)—can lead to increased household consumption. While this wealth effect can stimulate economic growth, it may also contribute to inflationary pressures, which are undesirable from a macroeconomic stability perspective. Addressing these dynamics requires a combination of financial literacy initiatives and well-calibrated monetary, fiscal, and

macroprudential policies. However, no single policy framework applies universally, and identifying optimal policy responses remains a persistent challenge (Crowe et al., 2011).

For future research, alternative methodologies could be explored to examine the wealth-consumption relationship, which may exhibit instability, asymmetry, and nonlinearity depending on economic conditions. Factors such as temporal variations in wealth, household habit formation, differences in utility functions, and loss aversion may contribute to these nonlinear effects. Moreover, households often perceive changes in wealth as transitory rather than permanent, which may influence consumption behaviour. Another promising avenue for further investigation involves quantitative studies on the impact of homeownership rates on private consumption across different economic contexts.

Acknowledgments

The author expresses sincere gratitude to Professors Dorina Lazăr, Cristian Litan, Alexandru Minea, Monica Pop Silaghi and Marcel Voia for their insightful comments at various stages of the writing process.

Statements and Declarations

Florina Burdet received partial financial support from the French Ministry of Foreign Affairs through the Eiffel Scholarship for conducting research in French universities. The author has no other competing interests to declare that are relevant to the content of this article.

References

Ahec-Šonje, A., Čeh-Časni, A., & Vizek, M. (2014). The effect of housing and stock market wealth on consumption in emerging and developed countries. Economic Systems, 38(3), 1–22. https://doi.org/10.1016/j.ecosys.2014.03.001

Ahec-Šonje, A., Ceh-Časni, A., & Vizek, M. (2012). Does housing wealth affect private consumption in European post-transition countries? Evidence from linear and threshold models. Post-Communist Economies, 24(1), 73–85. https://doi.org/10.1080/14631377.2012.647629

Ando, A., & Modigliani, F. (1963). The life cycle hypothesis of saving: Aggregate implications and tests. American Economic Review, 53, 55–84. https://doi.org/10.7551/mitpress/1923.003.0005

Angrisani, M., Hurd, M., & Rohwedder, S. (2019). The effect of housing wealth losses on spending in the great recession. Economic Inquiry, 57(2), 972–996. https://doi.org/10.1111/ecin.12753

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. The Review of Asset Pricing Studies, 10(4), 742–758. https://doi.org/10.1093/rapstu/raaa008

Bekaert, G., Harvey, C. R., & Mondino, T. (2023). Emerging equity markets in a globalized world. Emerging Markets Review, 56, 101034. https://doi.org/10.1016/j.ememar.2023.101034

Berger, D., Guerrieri, V., Lorenzoni, G., & Vavra, J. (2018). House prices and consumer spending. The Review of Economic Studies, 85(3), 1502–1542. https://doi.org/10.1093/restud/rdx060

Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. American Economic Review, 98(5), 1887–1921. https://doi.org/10.1257/aer.98.5.1887

Breitung, J. (2000). The local power of some unit root tests for panel data. In Non-Stationary Panels, Panel Cointegration, and Dynamic Panels: Advanced Econometrics (pp. 161–178). Amsterdam: JAI Press. https://doi.org/10.1016/S0731-9053(00)15006-6

Buiter, W. H. (2008). Housing wealth isn't wealth. National Bureau of Economic Research, No. W14204. https://doi.org/10.5018/economics-ejournal.ja.2010-22

Carroll, C. D., Otsuka, M., & Slacalek, J. (2011). How large are housing and financial wealth effects? A new approach. Journal of Money, Credit and Banking, 43(1), 55–79. https://doi.org/10.1111/j.1538-4616.2010.00365.x

Catte, P., Girouard, N., & André, C. (2004). Housing markets, wealth, and business cycle. OECD Economics Department Working Papers. Paris: OECD Publishing. https://doi.org/10.1787/18151973

Čeh-Časni, A. (2017). A note on housing wealth effect in selected Western Balkan countries. In Proceedings of the 14th International Symposium on Operational Research SOR'17: Slovenian Society Informatika, Section for Operational Research (pp. 21–26). Available at: https://www.croris.hr/crosbi/publikacija/prilog-skup/652514

Čeh-Časni, A. (2018). The housing wealth effect before and after the 2008 financial crisis: Panel vector autoregression approach. Conference Paper. Available at: https://www.croris.hr/crosbi/publikacija/prilog-skup/661732

Cheng, A., & Fung, M. (2008). Financial market and housing wealth effects on consumption: A permanent income approach. Applied Economics, 40(23), 3029–3038. https://doi.org/10.1080/00036840600994021

Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., & Van Rooij, M. (2021). Heterogeneous wealth effects. European Economic Review, 137, 103805. https://doi.org/10.1016/j.euroecorev.2021.103805

Ciarlone, A. (2011). Housing wealth effect in emerging economies. Emerging Markets Review, 12, 399–417. https://doi.org/10.1016/j.ememar.2011.06.003

Cloyne, J., Ferreira, C., & Surico, P. (2020). Monetary policy when households have debt: New evidence on the transmission mechanism. The Review of Economic Studies, 87(1), 102–129. https://doi.org/10.1093/restud/rdy074

Crowe, C., Dell'Ariccia, G., Igan, D., & Rabanal, P. (2013). How to deal with real estate booms: Lessons from country experiences. Journal of Financial Stability, 9(3), 300–319. https://doi.org/10.1016/j.jfs.2013.05.003

De Bondt, G. J., Gleseck, A., & Zekalte, Z. (2019). Thick modelling income and wealth effects: A forecast application to Euro area private consumption. Empirical Economics, 58(1), 257–286. https://doi.org/10.1007/s00181-019-01738-w

Friedman, M. (1957). A theory of the consumption function. Princeton University Press.

Funke, N. (2004). Is there a stock market wealth effect in emerging markets? Economics Letters, 83, 417–421. https://doi.org/10.1016/j.econlet.2003.12.002

Keynes, J. M. (1936). The general theory of employment, interest and money. Harcourt, Brace.

Jappelli, T., & Pistaferri, L. (2014). Fiscal policy and MPC heterogeneity. American Economic Journal: Macroeconomics, 6(4), 107–136. https://doi.org/10.1257/mac.6.4.107

Levin, A., Lin, C. F., & Chu, C. S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics, 108, 1–24. https://doi.org/10.1016/S0304-4076(01)00098-7

Mehra, Y.P., 2001. The wealth effect in empirical life-cycle aggregate consumption equations. Economic Quarterly, 87(2), pp.45–68. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2182655 [accessed 01.08.2024].

Nardo, M., Ossola, E., & Papanagiotou, E. (2022). Financial integration in the EU28 equity markets: Measures and drivers. Journal of Financial Markets, 57, 100633. https://doi.org/10.1016/j.finmar.2021.100633

Nicolau, M., 2020. Interest rate and the financial and housing wealth effects in ten CEEC. Economics Bulletin, 40(2), pp.1132–1145. Available at: https://www.accessecon.com/Pubs/EB/2020/Volume40/EB-20-V40-I2-P98.pdf [accessed 02.08.2024].

Peltonen, T.A., Sousa, R.M. and Vansteenkiste, I.S., 2012. Wealth effects in emerging market economies. International Review of Economics and Finance, 24, pp.155–166. https://doi.org/10.1016/j.iref.2012.01.006

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics, 22, pp.265–312. https://doi.org/10.1002/jae.951

Pesaran, M.H. and Shin, Y., 1997. An auto-regressive distributed lag modelling approach to co-integration analysis. In: Econometrics and economic theory in the 20th century: The Ragnar Frisch Centennial Symposium. Cambridge: Cambridge University Press, pp.371–413.

Pesaran, M.H., Shin, Y. and Smith, R.P., 1999. Pooled mean estimation of dynamic heterogeneous panels. Journal of the American Statistical Association, 94, pp.621–634. https://doi.org/10.1080/01621459.1999.10474156

Posedel, P. and Vizek, M., 2009. House price determinants in transition and EU-15 countries. Post-Communist Economies, 21(3), pp.327–343. https://doi.org/10.1080/14631370903090640

Poterba, J.M., 2000. Stock market wealth and consumption. Journal of Economic Perspectives, 14, pp.99–118. https://doi.org/10.1257/jep.14.2.99

Rodil-Marzábal, Ó. and Menezes-Ferreira-Junior, V., 2016. The wealth effect in the Eurozone. Panoeconomicus, 63(1), pp.87–112. https://doi.org/10.2298/PAN1601087R

Rosenberg, S., 2016. The impact of a change in real estate value on private consumption in Estonia. Research in Economics and Business: Central and Eastern Europe, 7(2), pp.5–26. Available at: https://econpapers.repec.org/RePEc:ttu:rebcee:78 [accessed 20.08.2023].

Slacalek, J., 2009. What drives personal consumption? The role of housing and financial wealth. The B.E. Journal of Macroeconomics, 9(1), pp.1–37. https://doi.org/10.2202/1935-1690.1555

Vizek, M., 2011. The influence of stock market and housing wealth on consumption expenditures in transition countries. Research in Economics and Business: Central and Eastern Europe, 3(1), pp.50–63. Available at: https://econpapers.repec.org/RePEc:ttu:rebcee:8 [accessed 20.08.2023].

Westerlund, J., 2007. Testing for error-correction in panel data. Oxford Bulletin of Economics and Statistics, 69, pp.709–748. https://doi.org/10.1111/j.1468-0084.2007.00477.x.