ARGUMENTA OECONOMICA

1 (50) • 2023

Recommended papers

Magdalena Daszkiewicz, Edyta Mazurek, Anetta Pukas
CITY INTEGRATED MARKETING COMMUNICATION – IDENTIFICATION AND MEASUREMENT FRAMEWORK

Li Sheng, Yechang Yin, Anning Zhang, Jiwei Wu, Ziqing Yang
MODELLING CASINO HOSPITALITY BUSINESS CYCLES

Erik Šoltés, Silvia Komara, Tatiana Šoltésová, Martin Mišút
ANALYSIS OF WORK INTENSITY IN SLOVAKIA USING TESTING AND ESTIMATION OF LINEAR COMBINATIONS OF GLM PARAMETERS

ISSN 1233-5835
e-ISSN 2720-5088

Publishing House of Wroclaw University of Economics and Business

Wroclaw University of Economics and Business
Argumenta Oeconomica – Editorial Office:
Komandorska 118/120, 53-345 Wrocław

Editorial Committee:

International Scientific Council:

ARGUMENTA OECONOMICA

1 (50) • 2023

Publishing House of Wroclaw University of Economics and Business
Wroclaw 2023

Editors
Ewa Knichnicka
Anita Makowska

Verified by
Elżbieta Macauley, Tim Macauley and Ewa Kania

Graphic design
Maciej Szłapka

Argumenta Oeconomica is covered in **Clarivate Analytics** services:

Social Sciences Citation Index®
Social Scisearch®
Journal Citation Reports/ Social Sciences Edition

and in **Elsevier** service Scopus

TABLE OF CONTENTS

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Magdalena Daszkiewicz**, *Edyta Mazurek***, *Anetta Pukas****

City integrated marketing communication – identification and measurement framework

New challenges resulting from dynamic changes observed in the twenty-first century are driving the evolution of approaches to city marketing communication. This article adapts the integrated marketing communication (IMC) concept for cities and develops a measurement framework and a theoretically consistent, valid, and reliable measurement tool for assessing city integrated marketing communication (CIMC). A literature review and previous qualitative studies provided the basis for conceptualising and identifying the specific constructs of CIMC, namely strategic consistency, interactivity, and stakeholder--centred focus. The research developed a theoretically consistent, valid, and reliable measurement tool for assessing CIMC. Empirical validation of the CIMC scale was conducted on data collected from a survey completed by representatives of municipal offices responsible for marketing communication in 279 Polish cities. The value and originality of this article derive from the development of the measurement framework and the new scale for assessing CIMC, which provide the foundation for further research on model solutions in this area. The measurement tool also contains subscales that can be used in research on specific dimensions of city marketing communication. The CIMC scale will assist practitioners in their decision-making processes and facilitate comparisons of cities in a local and international context.

Keywords: city integrated marketing communication (CIMC), city marketing, integrated marketing communication (IMC), place marketing, measurement framework, scale development

JEL Classification: M31, M39
DOI: 10.15611/aoe.2023.1.01

Quote as: Daszkiewicz, M., Mazurek, E., Pukas, A. (2023). City integrated marketing communication – identification and measurement framework. *Argumenta Oeconomica*, 1(50), 5-32.

* Department of Marketing Research, Wroclaw University of Economics and Business, Poland. ORCID: 0000-0003-1466-2454.
** Department of Statistics, Wroclaw University of Economics and Business, Poland. ORCID: 0000-0001-7410-1638.
*** Department of Marketing, Wroclaw University of Economics and Business, Poland. ORCID: 0000-0001-6318-2516.

1. Introduction

The contemporary practices of a large number of cities indicate that their authorities are pursuing new management solutions to enhance the effectiveness and integration of marketing communication. This article addresses the challenges involved in developing such an approach and responds to the call for a "wholesale or partial adoption of the IMC concept by different sectors" (Kitchen and Schultz, 2009).

Over the past two decades, integrated marketing communication (IMC) has been the topic of a considerable volume of scientific publications in the field of marketing management (Kitchen and Schultz, 2009; Kliatchko, 2008; Munoz-Leiva et al., 2015; Schultz, Kerr, Kim, and Patti, 2007; Schultz, Kim, and Kang, 2014). However, empirical research on IMC has tended to focus on business entities (Kliatchko, 2008; Lee and Park, 2007; Luxton, Reid, and Mavondo, 2015; Munoz-Leiva et al., 2015; Schultz et al., 2014; Šerić, 2018). Although some studies on IMC have focused on the non-commercial sector (Edmiston-Strasser, 2009; Foroudi, Dinnie, Kitchen, Melewar, and Foroudi, 2017; Kerr and Kelly, 2017; Šerić 2018), a neglected area is city communication. Previous research findings have indicated the feasibility of integrating city marketing communication (Anonymous, 2019). The limitations resulting from the specifics of public sector management have not excluded undertaking work on adapting the IMC concept to cities.

Whilst conducting a literature review, the authors identified inadequate conceptualisation and adaptation of the IMC concept for cities (an exploratory gap), and a lack of research guidelines for the identification and evaluation of city integrated marketing communication (a methodological gap). This study sought to fill these gaps by adapting the IMC concept for cities and developing a measurement framework and scale for assessing the integration of city marketing communication.

The research objectives addressed two areas that currently present fundamental challenges for managing city marketing communication. The first objective (theoretical objective) was to conceptualise and identify specific constructs of city integrated marketing communication (CIMC). To achieve this, adaptive work was required due to the differences between business management and public sector management. The second objective (methodological objective) was to propose a framework for the measurement of CIMC. This was achieved through the development of a research tool that required the construction of a novel measurement scale – which is linked to the need to adapt and operationalise variables – tailored to the specifics of cities.

The research design employed both qualitative and quantitative approaches. The authors first conducted a review of the literature identified through a search in the Scopus and Web of Science databases. The selection of literature was restricted to publications over the past two decades (since 2000) that have been crucial for developing knowledge on integrated marketing communication, city marketing,

place marketing. city branding, and place branding. Key to inferences made regarding the state of knowledge and research on IMC were reviews of theoretical and empirical studies by Kliatchko (2008), Schultz, Kim and Kang (2014), Munoz-Leiva, Porcu and Del Barrio-Garcıa (2015), Porcu, Del Barrio-García and Kitchen (2017), and Šerić (2018).

The literature review was supported by an analysis of the content of strategic documents derived from Polish cities and the results of studies on city marketing communication conducted by the authors in previous years. The qualitative study provided the basis for conceptualising CIMC and the construction of a proposed measurement tool (CIMC scale). Empirical validation of the measurement scale was performed on data collected from a survey conducted with representatives of municipal offices responsible for marketing communication in 279 Polish cities.

The value and originality of the paper derive from the development of integrated marketing communication theory and its adaptation in the area of city management, as well as the proposed measurement framework and new scale for the assessment of CIMC.

2. Theoretical background

2.1. The rationale for adapting the IMC concept to cities

According to Kerr and Patti (2015), IMC gained worldwide acceptance as a result of "evolving in its understanding and definition, with the exploration of key constructs, and some initial work on measurement". The literature on IMC indicates that since its inception as a field of study, theoretical foundations, definitional issues, and the development of the concept have been foremost concerns among scholars (Kliatchko, 2008; Schultz, Kim and Kang, 2014; Munoz-Leiva, Porcu and Del Barrio-Garcıa, 2015; Porcu, Del Barrio-García and Kitchen, 2017; Šerić, 2018). As part of the ongoing academic debate, doubts have been voiced as to the merits and possibilities of the full practical implementation of the concept (Pettegrew, 2001; McGrath, 2005). As early as 2009, Kitchen and Schultz, among other challenges involved in developing the IMC concept, indicated the need for "evidence of wholesale or partial adoption by different sectors". This inspired the subsequent discussion of the rationale of applying the IMC concept to cities.

The literature review demonstrated that the IMC concept has evolved from simplified communication system management (Duncan and Everett, 1993; Duncan and Caywood, 1996) to a stakeholder-centred, interactive, and cross-functional process (Duncan and Mulhern, 2004; Porcu, Del Barrio-García and Kitchen, 2017). Porcu, Del Barrio-García and Kitchen (2017), based on a comprehensive review of conceptualisations and theoretical models, defined IMC as "the stakeholder-centred interactive process of cross-functional planning and alignment of organisational, analytical and communication processes that allows for the possibility of continuous

dialogue by conveying consistent and transparent messages via all media to foster long-term profitable relationships that create value" (Porcu, Del Barrio-García and Kitchen, 2017, p. 694). This definition of IMC is broad enough to be applied to companies and other entities, including cities.

The basic premise underpinning the implementation of the IMC concept for cities is the development of city marketing and city branding (Lucarelli and Olof Berg, 2011; Acharya and Rahman, 2016; Vuignier, 2017). Previous research findings, supported by numerous studies demonstrating the links between city branding and corporate branding, have indicated the feasibility of IMC implementation in cities (Anonymous, 2019). The adaptation of the IMC concept for cities can therefore be based on scientific achievements in the field of city branding and city brand communication (Parkerson and Saunders, 2005; Kavaratzis and Ashworth, 2006; Hankinson, 2007; Kavaratzis, 2009; Kavaratzis and Hatch, 2013; Zenker, 2011; Merrilees et al., 2012), citizen engagement and participation (Braun, Kavaratzis, and Zenker, 2013; Hereźniak, 2017; Kavaratzis and Kalandides, 2015; Molinillo, Anaya-Sánchez, Morrison, and Coca-Stefaniak, 2019; Rehmet and Dinnie, 2013), city brand partnerships, relationships, networks, and co-branding (Hankinson, 2004; Kavaratzis and Ashworth, 2007; Harrison-Walker, 2013; Merrilees, Miller and Halliday, 2016; Lucarelli, 2018), and interactive city branding (Florek, 2011; van Gelder, 2011; Merrilees, Miller and Halliday, 2016).

Previous research has also identified multiple limitations associated with the direct implementation of IMC solutions in cities. These result from the collective character of cities (various stakeholders, numerous organisational units, diverse objectives and needs); limited control over communication processes; insufficient resources and marketing communication capabilities; and lack of flexibility and limited possibilities for the long-term implementation of strategic assumptions (Anholt, 2004; Skinner, 2005; Moilanen and Rainisto, 2009a; Moilanen, 2015; Anonymous, 2019). However, these limitations do not exclude undertaking work on the adaptation of the IMC approach to cities.

Although the research identified a few studies on the IMC of places, these referred only to tourist destinations. Skinner (2005) focused on the problems with integrating place marketing communication and stated that true integration might be impossible to achieve. Pike (2008) identified the opportunities and challenges related to the integration of marketing communication. His considerations covered five areas: customer relationships, stakeholder relationships, cross-functional processes, stimulation of dialogue with customers, and message synergy. Wang, Wu and Yuan (2009) explored the role of IMC in visitors' destination choices. However, their study was limited to measuring the declared impact of specific communication tools on destination decisions. Šerić and Vernuccio (2019) empirically tested a model capturing the relationships between consistency and interactivity (two dimensions of IMC), and city reputation and consumer brand engagement. Their study focused on cities, but only in the tourism sector. Castañeda-García et al. (2020) examined the

role of consistent messaging via various communication tools in forming consumer-based destination brand equity. This notable study was limited to one dimension of IMC in the context of tourism.

Cities represent multi-level organisms with diverse stakeholders, different objectives, and communication needs. Therefore, the theoretical challenges associated with the conceptualisation of city integrated marketing communication (CIMC) necessitate the adoption of a full organisational perspective. The study proposes to define CIMC as an interactive process involving planning and conducting strategically consistent communication activities and ongoing dialogue via various media to maintain long-term relations that support city branding and create value for city stakeholders. Scientific achievements in the field of city marketing and city branding facilitate the development of measuring instruments for their application in cities. Notwithstanding the necessary adaptation work, developing a measurement scale for CIMC is essential for evaluating and demonstrating the implementation value of IMC.

2.2. IMC measurement frameworks

Over the years, scientists have increasingly come to recognise that the key to a more widespread adoption of IMC is a clear definition of constructs, operationalisation and effective empirical measurement and evaluation (Eagle, Kitchen and Bulmer, 2007; Kitchen, Kim and Schultz, 2008; Kitchen and Schultz, 2009; Schultz and Patti, 2009; Kerr and Patti, 2015; Kitchen and Burgmann, 2015b; Luxton, Reid and Mavondo, 2017). Although measurement has often been considered a fundamental weakness of IMC (Kitchen, Kim and Schultz, 2008, Šerić, 2018), several proposed measurement frameworks for integrated marketing communications have emerged. Table 1 provides an overview of the IMC construct dimensions and scales used in prior research over the past two decades.

Low (2000) proposed a three-item scale to measure IMC in a cross-sectional sample of companies. The IMC construct consisted of integration (1), strategic consistency (2), and message consistency (3). By integration, the author intended the extent to which the same manager plans all marketing communication tools. The second dimension determined the extent of strategic coherence of various elements of the marketing communications programmes. The message consistency dimension reflected the focus of various marketing communications tools on a shared message.

Based on the previous measurement proposition presented by Phelps and Johnson (1996), Ewing, De Bussy and Caruana (2000) developed a four-dimensional construct: one voice (1), direct marketing (2), increased responsibility (3), and response goals (4). One voice appealed to the central themes in previous theoretical approaches: consistency, integration, and synergy. Direct marketing reflected the role of one-to-one, direct response, and database marketing (expressed as interactivity in conceptualisations of IMC proposed by other authors). The distinction of the

increased responsibility dimension resulted from the recognition of the shift in communication focus from traditional to below-the-line advertising. By defining the dimension of the response goals, the authors related their research to the pragmatic and retail-related aspects.

Table 1

IMC construct dimensions and measurement scales (since 2000)

Author/s (year)	IMC Construct Dimensions	Measurement scales (items, type of scale)
Low (2000)	1. Integration (integrated planning) 2. Strategic consistency 3. Message consistency	• 3 items • 9-point Likert scale from “strongly disagree” (1) to “strongly agree” (9)
Ewing, De Bussy and Caruana (2000)	1. One voice 2. Direct marketing 3. Increased responsibility (for below-the--line functions) 4. Response goals	• 14 items • 7-point Likert scale from “strongly disagree” (1) to “strongly agree' (7)
Reid (2005)	1. Interactivity 2. Mission marketing 3. Cross-functional strategic planning	• 15 items • 7-point Likert scale from “not at all” (1) to “to a great extent” (7)
Hočevar, Žabkar and Mumel (2007)	1. Interactivity 2. Strategic organisation 3. Mission communications 4. Planning	• 15 items • 7-point Likert scale from “strongly disagree” (1) to “strongly agree” (7)
Lee and Park (2007)	1. Unified communications for consistent message and image 2. Differentiated communications to multiple customer groups 3. Database-centred communications for tangible results 4. Relationship fostering communications with existing customers	• 18 items • 5-point Likert scale from “strongly disagree” (1) to “strongly agree” (5)
Wang, Wu and Yuan (2009)	1. Public relations 2. Advertising 3. Direct sales and promotion	• 21 items • 5-point Likert scale from “no influence” (1) to “high influence” (5)
Luxton, Reid and Mavondo (2015, 2017)	1. IMC capability	• 12 items • 7-point Likert scale from “not at all” (1) to “to a great extent” (7)
Porcu, Del Barrio-García and Kitchen (2017) Porcu, del Barrio--García, Kitchen, et al. (2019)	1. Message consistency 2. Interactivity 3. Stakeholder-centred strategic focus 4. Organisational alignment	• 25 items • 7-point Likert scale from “strongly disagree” (1) to “strongly agree” (7)

Source: own elaboration.

Several researchers have tested or adapted the scale proposed by Duncan and Moriarty (1997), which is composed of five dimensions: interactivity, mission marketing, strategic consistency, organisational infrastructure, and planning and evaluation. Reid (2005) then reduced this to three dimensions: interactivity (1), mission marketing (2), and cross-functional strategic planning (3). The interactivity measure addresses whether the customer's voice is the basis of brand-communication planning. The mission marketing measure determines the extent to which brand communication planning focus derives from the organisation's mission. The cross--functional strategic planning construct derives from three primary dimensions: organisational infrastructure, strategic consistency, and planning and evaluation. This dimension incorporates involvement from various functions and departments, the level of managerial skills and capabilities in brand communications, an understanding of customer contact points, the consistency of brand communication with brand positioning, and applying SWOT analysis in planning and strategy development.

Hočevar, Žabkar and Mumel (2007) proposed limiting the IMC measurement to four dimensions: interactivity (1), strategic organisation (2), mission communications (3) and planning (4). The interactivity dimension refers to collecting and using information gained through dialogue with customers via various brand contact points. The second dimension addresses the use of a long-term and common strategy to coordinate marketing communications. The third dimension refers to communicating the mission to consumers and its consistency with the company's messages. The IMC planning dimension concerns setting measurable goals for the entire marketing communication process and its tools.

In 2007, Lee and Park proposed the IMC measurement tool based on a four--dimensional conceptualisation of integrated marketing communication: unified communications for a consistent message and image (1), differentiated communications to multiple customer groups (2), database-centred communications for tangible results (3), and relationships fostering communications with existing customers (4). The first dimension refers to delivering a consistent message to create a single brand identity via various channels. The second dimension addresses the importance of developing different marketing communications campaigns for customers in different phases of the purchasing process. The third dimension represents the communicational role of databases in achieving tangible results. The fourth dimension underlines the importance of marketing communications in building long-lasting relationships with customers.

As the concept of IMC has evolved, there has been an increasing shift away from measurements based solely on tool-based approaches. However, some researchers have still based the measurement on marketing communication tools. An example is the research of Wang, Wu and Yuan (2009), who used three dimensions: public relations, advertising, and direct sale and promotion. The proposed scale was quite simple and the items represented marketing communication tools and channels.

Over the past two decades, a considerable part of the research on IMC has adopted a broader organisational perspective. Measures used in previous research by Duncan and Moriarty (1997) and by Low (2000) and Reid (2005) became the basis for the IMC capability construct proposed by Luxton, Reid and Mavondo (2015, 2017). The construct was reflected in a 12-item scale examining organisations' ability to undertake marketing communication activities that affect their brands. In their measurement, they referred to the use of a set of communication tools to achieve overall brand goals, the capacity of a creative theme to use in campaigns aimed at different stakeholders, linking the strategy and its objectives to building relationships with key stakeholders, planning based on target market insights and SWOT analysis, understanding all brand touchpoints, evaluation of all campaigns, maintaining highly skilled personnel, and devoting adequate time to manage brand communications.

Porcu, Del Barrio-García and Kitchen (2017) developed a conceptual framework and operationalisation of IMC by elaborating and validating the new IMC scale. The authors introduced a four-dimensional IMC construct: message consistency (1), interactivity (2), stakeholder-centred strategic focus (3), and organisational alignment (4). Their proposition responds to the need for a broad organisational approach and provides the basis for the adaptation of the measurement framework to entities representing various sectors.

The message consistency (1) dimension refers to strategic and coherent positioning, consistency in visual communication, and the coordination of all messages originated by all departments and functions (Porcu, Del Barrio-García and Kitchen, 2017). Such coordination is intended to lead to communication efficiency based on synergy (Kitchen, Brignell, Li and Jones, 2004; Lee and Park, 2007; Moriarty and Schultz, 2012). The value of message consistency is challenging to test in the marketplace (Moriarty and Schultz, 2012). Therefore, measurements within this dimension focus primarily on the coordination and coherence of various communication tools and channels.

Interactivity (2) involves collecting stakeholders' information, monitoring stakeholder-generated messages, and adapting throughout the organisation a responsive attitude towards establishing a trust-based and ongoing dialogue (Porcu, Del Barrio-García and Kitchen, 2017). The rise of interactivity in integrated marketing communication is strongly linked to the development of digital technology and media (Peltier, Schibrowsky and Schultz, 2003; Peltier et al., 2006; Manser Payne, Peltier and Barger, 2017). The use of ICTs enhances the ability to obtain consumer data and monitor consumer insight (Mulhern, 2009), increasing the chance of achieving a better mutual understanding and strengthening relations.

The stakeholder-centred strategic focus dimension (3) concerns working towards the strategic goals of creating value for stakeholders. Since its inception, numerous researchers have viewed ongoing relations between organisations and various stakeholder groups as the foundation of IMC (Duncan and Moriarty, 1998; Reid, 2005; Finne and Grönroos, 2009; Mulhern, 2009; Tafesse and Kitchen, 2017).

Measuring IMC should therefore address the strategic use of stakeholder insight in providing and achieving high-value stakeholder-centred solutions. Stakeholders should be knowledgeable about the organisation's strategic direction, which requires raising their awareness of its mission (Porcu, Del Barrio-García and Kitchen, 2017).

The organisational alignment dimension (4) refers to the horizontal and vertical communication processes, Developing and deploying IMC capability requires the sharing of corporate values and goals, and firm-wide commitment and cooperation across various functional areas (Abratt and Kleyn, 2012; Zahay et al., 2014; Luxton, Reid and Mavondo, 2017). The horizontal and vertical cooperation and coordination of communicational activities are crucial because all organisational units and partners affect the corporate reputation (Porcu, Del Barrio-García and Kitchen, 2017).

3. Methodology

This study was conducted to develop a framework and new measurement scale for CIMC. The research design employed both qualitative and quantitative approaches. The literature review was supported by an analysis of the content of strategic documents obtained from Polish cities and previous studies on city marketing communication conducted by the authors in the period 2015-2017[1]. The qualitative study formed the basis for the conceptualisation of CIMC and the construction of a measurement tool proposal (CIMC scale). Questionnaire items were translated and adapted in Polish. The scale items were discussed with experts (scholars and practitioners) with experience in city marketing communication and research on city marketing and branding. The purpose of this stage was to review the adapted constructs and items and ensure they were appropriate for research regarding cities. The study also addressed concerns regarding the respondents' ability to understand questionnaire items and provide relevant answers.This stage resulted in revisions to several items and the questionnaire layout.

To empirically validate the proposed measurement scale, the authors conducted a survey on a representative sample of Polish cities[2]. According to Statistics Poland (Polish National Statistical Office), there are 398 Polish territorial units with city rights that have more than 10,000 inhabitants (as of December 2020). The respondents were representatives of municipal offices responsible for decisions in the field of city

[1] These studies include a case study of the city of Wrocław, based on an analysis of the content of strategic documents of the city published in 1998-2017, research reports and materials related to communication activities of the city, and in-depth interviews with representatives of Wrocław City Council and municipal companies responsible for communication activities, city promotion, the implementation of large city events, and coordination of CRM projects. In previous research, the authors also conducted a comparative analysis of the communication activities of selected large European cities (Amsterdam, Copenhagen, Berlin, Edinburgh) and three Polish cities (Wrocław, Szczecin, Poznań).

[2] The definition of a city varies from country to country and is based on various criteria (population size, population density, functional and historical criteria).

marketing communication. Due to the respondents' availability, it was impossible to survey representatives of all municipal offices in 398 cities. The initial sample consisted of 361 cities, but the analysis only included responses from cities with a unit responsible for marketing communication. The respondents were representatives for whom marketing communication was a core competency[3]. Thus, the final sample consisted of 279 cities.

Table 2

The sample and population by number of inhabitants

Polish territorial units with city rights	Sample		Population	
	n	%	n	%
Number of inhabitants				
10,000 to 19,999	131	47	180	45
20,000 to 99,999	118	42	180	45
100,000 and more	30	11	38	10
Total	279	100	398	100

Source: Statistics Poland, 2020.

The sample was selected from three categories of territorial units with city rights defined by the number of inhabitants (based on the Statistics Poland classification system). To a large extent, the sample structure reflects the population structure (Table 2). A description of the sample, including other characteristics, is presented in Appendix 1.

The study used triangulation of research techniques (CATI, CAWI), which resulted from the different preferences of the respondents regarding the form of contact and interviewing, and the need to ensure an appropriate response rate. The survey was administered by a research agency with experience in research on public entities. A preliminary pilot study on representatives of 10 cities was conducted to ascertain the clarity and acceptability of the questions. After ensuring that the respondents had no problems answering the questions, the survey was conducted on a larger sample.

The data gathered via the survey enabled validation of the proposed CIMC scale through the application of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The sample size is crucial for EFA and CFA methods, although researchers represent different approaches in this regard. Tabachnick and Fidell (2007) suggested a sample with at least 300 cases, Cattell (1966) – 250, Hair et al. (1998) – 100, Sapnas and Zeller (2002) – even 50 objects. MacCallum et al.

[3] The study excluded the responses of persons performing various administrative functions, or whose scope of duties included tasks related to marketing communication but did not constitute their primary scope of responsibilities.

(1999) proved that the necessary sample size depends on several specific aspects of a given study, most notably the level of commonality. When commonalities are over 0.6, the factors are well determined, and computations converge to a proper solution, even for the smaller sample size.

Luxton, Reid and Mavondo (2017), defining directions for research on IMC, suggested that future studies should adopt a specific lens, such as examining different organisation sizes. Based on this, the authors assumed it would be reasonable to conduct research and analyses for subgroups according to city size. Therefore, it was decided to validate the scale regarding separate categories of territorial units.

4. CIMC measurement framework and scale development

4.1. CIMC measurement framework

Based on the literature review, supported by an analysis of the content of strategic documents obtained from Polish cities and the results of previous studies on city branding and city marketing communication, the authors assumed that the construct dimensions proposed by Porcu, Del Barrio-García and Kitchen (2017) would form the basis for developing a measurement framework for the assessment of CIMC. However, the development of constructs and measurement scales required adaptive work due to differences between business management and public sector management. On the basis of qualitative studies, the study proposed a measurement framework and operationalisation of the CIMC construct based on three dimensions: strategic consistency (SC), interactivity (IN), and stakeholder-centred focus (SF).

Strategic consistency. The two constructs, "message consistency" and "organisational alignment" proposed by Porcu, Del Barrio-García and Kitchen (2017), were linked into one overarching definition as "strategic consistency". This is a logical connection grounded in IMC theory. The notion that messages and their multiple sources must be strategically coordinated ('one sight, one sound', 'one view') has its roots in the origins of the IMC concept (Duncan and Everett, 1993; Moriarty, 1996; Thorson and Moore, 1996; Caywood et al., 1991). By "strategic consistency", Duncan and Moriarty (1998) referred to the consistent presentation of organisational values and the coherent brand positioning.

Therefore, strategic consistency refers to sharing common brand meaning and content among multiple communication channels and tools (Duncan and Moriarty, 1998; Navarro, Delgado and Sicilia, 2010; Šerić and Vernuccio, 2019). This highlights the need for consistent marketing communication between various sources and media (online and offline) and achieving intra-channel and cross-channel synergy (Batra and Keller, 2016; Pauwels et al., 2016). From a strategic point of view, consistency involves sharing values and goals within all organisational units and requires the alignment of horizontal and vertical communication processes.

To ensure consistent positioning, it is essential to coordinate messages reaching diverse audiences through various channels. Hence, it is helpful to establish a clear key message as this will facilitate communication with different stakeholders from a long-term perspective and establish coherent communication programmes. Hankinson (2007) suggested that one of the fundamental shortcomings of place marketing is the failure to communicate a consistent message across the various target groups. However, as previous case study-based research shows, there are numerous examples of cities conducting communication activities based on a consistent message derived from the city brand strategy (Anonymous, 2016). Strategic consistency allows the continuation of the adopted path of marketing communication activities, even in the face of problems that stem from the terms of office of the city authorities.

Consistency also refers to coherency in the use of brand identity elements. As Govers (2013) observed, city marketing communication efforts often focus on spending time and money on slogans and logos. While these identity elements are essential in the marketing communications of commercial brands, multiple authors have repeatedly claimed their relative insignificance in place branding (Govers, 2013). Conversely, place marketing researchers indicate the importance of symbols in forming local identity and place image (Mueller and Schade, 2012). The consistent use of city identity elements is also facilitated by fostering commitment, engagement, and a sense of belonging among stakeholders and internal audiences (Mueller and Schade, 2012; Govers, 2013).

An important aspect related to strategic consistency is organisational alignment. This includes solutions and mechanisms for the communication and cooperation of various city units in ways that are consistent with the city's values and support its long-term goals. In city marketing, maintaining consistency is difficult if communication programmes are run independently by multiple entities with differing objectives, resources, and capabilities (Skinner, 2005; Moilanen and Rainisto, 2009a). To address the problems with multi-organisational coordination, the municipal office should create opportunities to increase the strategic consistency of city communication and initiate long-term cooperation programmes for various entities operating in the city (Moilanen and Rainisto, 2009b; van Gelder, 2011; Anonymous, 2016).

Interactivity. Interactivity is an essential dimension in contemporary constructs of IMC (Reid, 2005; Moriarty and Schultz, 2012; Luxton, Reid and Mavondo, 2015; Porcu et al., 2017; Porcu, del Barrio-García, Alcántara-Pilar et al., 2019). This refers to ongoing dialogue involving various stakeholders in which the key is reciprocity, speed of response and responsiveness (Johnson, Bruner and Kumar, 2006; Moriarty and Schultz, 2012; Porcu, del Barrio-García, Alcántara-Pilar et al., 2019). Interactivity relies on exchanging messages and co-creating shared meaning by marketers and audiences (Kliatchko, 2008; Finne and Grönroos, 2009; Tafesse and Kitchen, 2017). By encouraging mutually beneficial dialogue, interactivity is considered crucial for initiating, maintaining, and nurturing stakeholder relations (Reid, 2005; Lucia Porcu, del Barrio-García and Kitchen, 2012; Tafesse and Kitchen, 2017).

Recently, researchers involved in marketing and place branding have exhibited an increased interest in interactive communication. Several researchers reported the role of interactive tools and channels of marketing communication in place brand co-creation (Florek, 2011; Kavaratzis, 2012; Hanna and Rowley, 2015; Merrilees, Miller and Halliday, 2016) and building place reputation (Šerić and Vernuccio, 2019).

In the context of cities, interactivity is about finding and providing different ways to conduct and stimulate dialogue with city stakeholders, respond to their inquiries and comments, and collect and monitor information generated by stakeholders. Interactivity is now linked strongly to the use of digital media, which enhance the capability to engage in dialogue, acquire stakeholder data, and monitor stakeholder-generated messages. According to Kavaratzis (2012), user-generated content is the critical change in modern communication. Therefore, it is important to systematically monitor the information generated by city stakeholders in their online word-of-mouth communications. Moreover, an increasing number of cities are engaged in the intensive development and use of interactive tools (platforms, applications, etc.), allowing for dialogue between various stakeholders.

Stakeholder-centred focus. Putting stakeholders at the centre of marketing communication processes to create shared value and build long-term and profitable relations is seen by many researchers as crucial to the implementation of IMC (Duncan and Moriarty, 1998; Reid, 2005; Kliatchko, 2008; Finne and Grönroos, 2009; Mulhern, 2009; Tafesse and Kitchen, 2017).

City marketing communication requires thinking about the complex system of diverse stakeholders, where the actions of some have an impact on others (Jamal and Getz, 1995; Moilanen, 2015). In addition, the city's brand depends not only on the actions of the authorities but also (and in particular) on the activities and communication of the city's stakeholders (Braun et al., 2018). According to Moilanen (2015), modern cities face problems such as conflicting stakeholders' interests, lack of leadership and insufficiently effective communication within loosely structured networks. This implies the need to disseminate knowledge about the strategic directions of city development among stakeholders and involving them in the planning and evaluation processes.

The basis for creating value for the city and its stakeholders is the identification of multiple recipients and understanding their needs and preferences. Stakeholder insights are used to plan marketing communication and other city activities. Stakeholders should be knowledgeable about the city's strategic direction (vision, mission) and may be included in the planning and evaluating of its activities through systematic surveys. Furthermore, the engagement of city staff and an efficient information flow between city units are essential in developing and implementing stakeholder-focused solutions.

The proposed three dimensions (strategic consistency, interactivity, stakeholder-centred focus) were the conceptual basis for the development of a scale for the measurement of CIMC.

4.2. Proposed CIMC measurement scale

The authors proposed a CIMC measurement tool consisting of 20 items: eight items for "strategic consistency", six items for "interactivity", and other six for "stakeholder-centred focus" (Table 3).

Table 3

City Integrated Marketing Communication Scale (CIMC scale)

The CIMC scale
Strategic consistency (SC)
SC_1. We coordinate the messages of the city marketing communication to maintain consistency of city positioning and image.
SC_2. We maintain consistency in all visual components of city communication.
SC_3. We care for the consistency of marketing communication in traditional and modern media.
SC_4. Communication with various recipient groups share a key message.
SC_5. We periodically analyse our city communication messages to assess their consistency level.
SC_6. Due to internal communication, employees of various organisational city units are aware of the values and goals resulting from the city's strategy.
SC_7. Well-developed mechanisms of cooperation and teamwork within various organisational units that serve the implementation of strategic goals exist in the city.
SC_8. The municipal office is the initiator of long-term cooperation programmes for various entities operating in the city.
Interactivity (IN)
IN_1. Our city offers various ways to facilitate dialogue with city stakeholders (obtaining information, submitting comments etc.).
IN_2. In our city, we systematically collect information from stakeholders in a database used by various organisational units.
IN_3. Our city uses information and communication technologies intensively to respond quickly to inquiries and comments submitted by stakeholders.
IN_4. Our city proactively uses social media to engage in dialogue with stakeholders.
IN_5. Our city is intensely developing interactive tools that allow for dialogue between various stakeholders (platforms, applications, etc.).
IN_6. In our city, we systematically monitor the information generated by stakeholders through their online word-of-mouth communications.
Stakeholder-centred focus (SF)
SF_1. The city vision/mission is actively promoted among stakeholders, particularly among residents and entities operating in the city.
SF_2. The insights of marketing research and analysis of stakeholders' needs are the basis for planning city marketing communication.
SF_3. The choice of city communication tools and channels is based on an analysis of the needs and preferences of our stakeholders.
SF_4. The staff of all city organisational units actively pursue stakeholder-centred solutions.
SF_5. In our city, we systematically conduct stakeholder surveys to evaluate and plan the city's activities.
SF_6. The adopted organisational solutions affect the efficient flow of information between various organisational units of the city.

Note: The respondents were asked to indicate their level of agreement with each statement on a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5).

Source: developed based on the firm-wide IMC scale proposed by Porcu, Del Barrio-García and Kitchen (2017).

With regard to specifying the number of points on the Likert scale, the study considered that each respondent must have a relatively precise and stable understanding of the meaning of each point (Krosnick and Presser, 2010). Thus, it was decided to use a 5-point Likert scale (from "1, strongly disagree" to "5, strongly agree") to make it easier for the respondents to read and understand the options and facilitate completion of the questionnaire. The authors also acknowledged that it was quite simple to read out the list of scale descriptors when using the CATI technique. In a study comparing data obtained from using 5-point, 7-point and 10-point scales, Dawes (2008) found that the three options are comparable for analytical tools such as CFA or structural equation models. Furthermore, both 5- and 7-point scales can be easily rescaled, and the resulting data will be reasonably comparable.

5. Empirical CIMC scale validation

The CIMC scale composed of 20 items was then subjected to empirical validation. EFA was used to verify the latent factors of the set of items using the STATISTICA 13.0 statistical programme (StatSoft Power Solutions. Inc.). CFA was applied to verify the underlying factor structure derived from the EFA. Descriptive statistics and a correlation analysis of the different items and Cronbach's alpha were then calculated. Composite reliability (CR) and average variance extracted (AVE) were estimated to ensure reliability of the CFA results following the suggestion of Hair et al. (2014). To ensure the feasibility of the factor analysis, the correlation matrix was evaluated. Bartlett's sphericity test ($\chi^2 = 441.24$; $df = 190$; $p < 0.001$) affirmed that the matrix was not an identity matrix, while the Kaiser-Meyer-Olkin sampling adequacy measure (KMO = 0.765) yielded a value of almost 0.8. Both results indicate that it is possible to extract factors from the matrix of observed correlations.

The statistical analysis was then conducted for cities with 100,000 or more inhabitants. EFA was applied to the data to determine the dimensionality of the IMC scale. It revealed three factors explaining 71.4% of the total variance. Four questions with values lower than 0.6 were excluded from the analysis (SC_3; IN_2; SF_1, SF_3). To test the dimensionality of the scale, CFA was applied. Following Porcu, Del Barrio-García and Kitchen (2017), three diagnostic measures were implemented to estimate construct reliability and assess the degree of consistency between multiple measurements of a variable. The item-to-total correlation and the inter-item correlation of each dimension were analysed and both exceeded the suggested cut-off values (0.5 and 0.3, respectively).

Internal consistency coefficients were then assessed to check the reliability of the three factors. Cronbach's scores were 0.91 for SC; 0.90 for IN; and 0.87 for SF. All these values exceeded the most conservative threshold of 0.8 (Table 4) recommended for purified scales. In addition, the computed average variance extracted (AVE) and the composite reliability (CR) exceeded the recommended thresholds of 0.5 and 0.7,

respectively (Fornell and Larcker, 1981a; Hu and Bentler, 1999). As indicated in Table 4, the composite reliability of every factor ranged from 0.87 to 0.91, whereas the statistical value of AVE for every loaded construct extended from 0.6 to 0.64.

Table 4

Results of the confirmatory factor analysis for cities with 100,000 or more inhabitants

Cities with 100,000 and more inhabitants									
Items	Standardised coefficients	t-value	p-value	R^2	M	SD	α	AVE	CR
Construct I (SC)									
SC_1	0.78	9.65	0.000	0.61	4.37	0.61	0.91	0.60	0.91
SC_3	0.76	8.89	0.000	0.58	4.23	0.77			
SC_4	0.68	6.37	0.000	0.47	4.20	0.76			
SC_5	0.74	7.91	0.000	0.54	4.20	0.61			
SC_6	0.77	9.21	0.000	0.60	4.23	0.68			
SC_7	0.83	11.99	0.000	0.68	3.87	0.68			
SC_8	0.84	12.75	0.000	0.70	4.20	0.71			
Construct II (IN)									
IN_1	0.87	15.20	0.000	0.76	4.30	0.70	0.90	0.64	0.90
IN_3	0.71	6.96	0.000	0.50	4.10	0.66			
IN_4	0.73	7.52	0.000	0.53	4.27	0.69			
IN_5	0.83	11.89	0.000	0.68	4.23	0.63			
IN_6	0.86	14.78	0.000	0.75	4.07	0.74			
Construct III (SF)									
SF_2	0.83	11.21	0.000	0.69	3.97	0.76	0.87	0.64	0.87
SF_4	0.77	8.40	0.000	0.59	4.07	0.74			
SF_5	0.75	7.97	0.000	0.57	4.10	0.66			
SF_6	0.84	11.49	0.000	0.70	4.13	0.63			
Estimation model									
CIMC→SC	0.84	1.98	0.048	0.70				0.65	0.85
CIMC→IN	0.82	2.14	0.032	0.67					
CIMC→SF	0.76	2.68	0.007	0.58					

Notes: M – mean, SD – standard deviation, α – Cronbach's alpha, AVE – average variance extracted, CR – composite reliability.

Source: own elaboration.

Three widely applied model-fit indices are used in this article, all of which are based on a fit function given a specific estimation method, such as: the root mean square error of approximation (RMSEA; Steiger and Lind, 1984; Steiger, 1990), comparative fit index (CFI; Bentler, 1990), and Tucker–Lewis index (TLI; Tucker

and Lewis, 1973; Bentler and Bonett, 1980). The results provided evidence for the adequate dimensionality of the IMC scale and thus the 16 items could be considered appropriate. The results indicated that the model ($\frac{\chi^2}{df} = 1.30$; RMSEA = 0.04; TLI = 0.87; CFI = 0.90) provided an acceptable overall goodness-of-fit. Standardised parameter estimates are provided in Table 4 and all are statistically significant at the alpha level of 0.001. The R^2 values indicate the amount of variance of the items explained by the respective constructs.

To assess discriminant validity, the criterion suggested by Fornell and Larcker (1981b) was applied by calculating the square root of the AVE. The authors suggested that the AVE of each factor should be greater than the squared correlation between that factor and any other. This means that the square root of the AVE should be greater than the correlation shared among the constructs. The results (Table 5) indicated that the correlation coefficients (off-diagonal) between each pair of constructs were less than the square root of each AVE in the diagonal; hence, this measurement model supports discriminant validity between the constructs.

Table 5

Square root of AVE and correlations between constructs for cities with 100,000 or more inhabitants

CONSTRUCT	SC	IN	SF
SC	**0.77**		
IN	0.75	**0.80**	
SF	0.69	0.73	**0.80**

Notes: Values in bold are square roots of AVE showing discriminant validity.

Source: own elaboration.

Based on these findings, the proposed measurement scale demonstrated adequate reliability for cities with 100,000 or more inhabitants. Therefore, a useful tool for assessing CIMC had been obtained.

An analogous analysis was then performed for cities with fewer than 100,000 inhabitants. The results are presented in Table 6, which contains all the measures described earlier. Bartlett's sphericity test for cities with 20,000 to 99,999 inhabitants ($\chi^2 = 979.11$; $df = 190$; $p < 0.000$), and the Kaiser-Meyer-Olkin sampling adequacy measure (KMO = 0.87) for cities with 10,000 to 19,999 inhabitants: ($\chi^2 = 1189.70$; $df = 190$; $p < 0.000$); KMO = 0.92, indicated that it was possible to extract factors from the matrix of observed correlations. The EFA indicated two factors, which were then validated by confirmatory analysis.

Finally, the constructs included items for which the item-to-total correlation and the inter-item correlation of each dimension exceeded the values of 0.5 and 0.3, respectively. Based on Hair et al. (2010), all standardised coefficients were statistically significant ($p < 0.0001$) and greater than 0.7. The results indicated that the model for

cities with 20,000 to 99,999 inhabitants ($\frac{\chi^2}{df} = 1.19$; RMSEA = 0.04; TLI = 0.99; CFI = 0.99) and for cities with 10,000 to 19,999 inhabitants ($\frac{\chi^2}{df} = 1.10$; RMSEA = 0.02; TLI = 0.99; CFI = 0.99) provided an acceptable overall goodness-of-fit.

Table 6

Confirmatory factor analysis for cities with fewer than 100,000 inhabitants

Cities with population of 20,000 to 99,999 inhabitants									
Items	Standardised coefficient	t-value	p-value	R^2	M	SD	α	AVE	CR
Construct II (IN)									
IN_1	0.77	14.41	0.000	0.59	3.80	0.75	0.79	0.55	0.79
IN_2	0.74	13.23	0.000	0.55	3.68	0.75			
IN_5	0.72	12.30	0.000	0.52	3.80	0.79			
Construct III (SF)									
SF_1	0.703	11.68	0.000	0.49	3.81	0.65	0.78	0.54	0.78
SF_2	0.721	12.32	0.000	0.52	3.67	0.83			
SF_5	0.782	14.74	0.000	0.61	3.67	0.80			
Estimation model									
CIMC→IN	0.71	12.33	0.000	0.51				0.65	0.76
CIMC→SF	0.86	10.52	0.000	0.74					
Cities with population of 10,000 to 19,999 inhabitants									
Items	Standardised coefficient	t-value	p-value	R^2	M	SD	α	AVE	CR
Construct II (IN)									
IN_1	0.66	8.91	0.000	0.43	3.85	0.74	0.72	0.46	0.71
IN_2	0.68	9.39	0.000	0.47	3.73	0.65			
IN_3	0.69	9.45	0.000	0.47	3.66	0.72			
Construct III (SF)									
SF_3	0.70	10.52	0.000	0.49	3.534	0.70	0.75	0.50	0.75
SF_4	0.69	10.20	0.000	0.47	3.649	0.74			
SF_5	0.72	10.89	0.000	0.52	3.618	0.72			
Estimation model									
CIMC→IN	0.81	4.58	0.000					0.65	0.73
CIMC→SF	0.71	15.00	0.000						

Notes: M – mean, SD – standard deviation, α – Cronbach's alpha. AVE – average variance extracted, CR – composite reliability.

Source: own elaboration.

An EFA made it possible to hypothesise the existence of up to three dimensions representing CMIC:

- strategic consistency (SC),
- interactivity (IN),
- stakeholder-centred focus (SF).

According to an assumption related to the possibility of measurement differences for subgroups of cities, the number of dimensions depends on the size of the city. The three dimensions listed above comprise the CIMC for cities with 100,000 and more inhabitants. For cities with fewer than 100,000 inhabitants, CIMC is described by two unobserved factors: interactivity (IN) and stakeholder-centred focus (SF).

CFA was performed to corroborate the factor structure obtained following EFA of the scale. The results indicated the existence of adequate values higher than 0.87 in the CFI and TLI indices, as well as values between 0.02 and 0.04 in RMSEA and RMR. Concerning the $\frac{\chi^2}{df}$ ratio, a value of 1.30 for cities with 100,000 and more inhabitants (1.19 for cities with 20,000 to 99,999 inhabitants and 1.10 for cities with 10,000 to 19,999 inhabitants) was obtained, which is considered good as it was lower than three (values below 2.0 are deemed to be indicators of a very good model fit). The scale's internal consistency was excellent, with Cronbach's values between 0.72 and 0.91. To reaffirm the internal consistency, the CR of all the factors was higher than 0.8 (see Tables 4 and 6), exceeding the minimum values stipulated by Hair et al. (2014), who set the value of the statistic at 0.7. For the AVE, all the values exceeded the minimum requirement of 0.5.

The statistical analysis also identified differences in the composition of CIMC scale items for sub-categories of cities. For cities with the largest population (100,000 and more inhabitants), after validation the measurement scale included 16 items encompassing three constructs: seven items for "strategic consistency" (SC), five items for "interactivity" (IN) and four items for "stakeholder-centred focus" (SF) (Table 4). Initially, the proposed scale consisted of 20 items (Table 3), four of which were eliminated after the validation process. All items except one regarding the consistency of visual identity (SC_2) formed the strategic consistency construct. This aligns with the views of authors who have already raised their relatively low importance in place branding (Govers, 2013). The interactivity construct lacked one item of the initial scale related to the collective usage of stakeholder information collected systematically in the database (IN_2). In the stakeholder-centred focus construct, two items were eliminated after the validation process. These concerned city vision/mission promotion among stakeholders (SF_1) and the choice of city communication tools and channels based on an analysis of the needs and preferences of our stakeholders (SF_3).

For cities with fewer than 100,000 inhabitants, CIMC was based on two constructs: interactivity (IN) and stakeholder-centred focus (SF). The surprising aspect of the statistical analysis was that the strategic consistency construct was not distinguished

for the two sub-categories of smaller cities. After validation, CIMC was based on six items: three items for "interactivity" (IN) and other three for "stakeholder-centred focus" (SF). Notably, there were also differences in the composition of items forming the distinct constructs (Table 5).

For the two sub-categories of cities, the interactivity construct includes items concerning various ways to facilitate dialogue with stakeholders (IN_1) and related to the collective usage of stakeholder information collected systematically in the database(IN_2). Within interactivity, the set of scale items differed by one item. Regarding cities with 20,000 to 99,999 inhabitants, the third item concerned the development of interactive tools (platforms, applications) to enable dialogue with stakeholders (IN_5). In the context of cities with fewer than 20,000 inhabitants, there is an item relating to the intensive use of ICTs to respond to stakeholders' inquiries and comments (IN_3). More considerable differences relate to the composition of scale items for "stakeholder-centred focus", because the scale positions in the two categories of smaller cities did not coincide.

Discussion and conclusions

This article addressed the challenges of developing the IMC concept for cities and responds to the call for the "wholesale or partial adoption of the IMC concept by different sectors" (Kitchen and Schultz, 2009). The study adapted the IMC concept for cities based on the company-wide conceptualisation of IMC (Porcu, Del Barrio-García and Kitchen, 2017) and by considering the differences between business and public sector management. This study provided a conceptualisation and measurement framework of CIMC.

The value and uniqueness of this article lies in the development of a conceptual and measurement framework to assess the integration of city marketing communication. The authors proposed a new measurement framework and operationalisation of the CIMC construct based on three dimensions: strategic consistency (SC), interactivity (IN), and stakeholder-centred focus (SF).

Strategic consistency (SC) refers to sharing brand meaning and content among different communication channels and tools used by cities to achieve intra-channel and cross-channel synergy. It requires the alignment of horizontal and vertical communication processes and the existence of mechanisms to facilitate the communication and cooperation of various city units in ways that are consistent with the city's values and support its long-term goals. Strategic consistency also refers to the multi-organisational coordination and long-term cooperation programmes for various entities operating in the city.

Interactivity (IN) concerns the ongoing dialogue that is crucial for building mutually beneficial relations with and between city stakeholders. It is about providing different methods of contact, responding to inquiries and comments, and collecting and monitoring information generated by stakeholders. The ability to engage in

dialogue, obtain stakeholder data, and monitor stakeholder-generated messages is enhanced by the use of ICTs and digital media.

Stakeholder-centred focus (SF) places city stakeholders at the centre of marketing communication processes to create shared value and build long-term and profitable relations. This dimension implies the need to disseminate knowledge about the strategic directions of city development among stakeholders and involving them in the planning and evaluation processes. It also requires the engagement of city staff and efficient information flow between city units in developing and implementing stakeholder-focused solutions.

Initially, the authors proposed a CIMC measurement tool consisting of 20 items: eight items for "strategic consistency", six items for "interactivity", and other six for "stakeholder-centred focus". The scale was validated according to the adopted assumptions for subcategories of cities of different sizes (measured by the number of inhabitants).

The statistical analysis for cities with 100,000 or more inhabitants provided evidence for the adequate dimensionality of the IMC scale and enabled to classify the 16 items as appropriate. Based on the findings, the proposed measurement scale demonstrated adequate reliability. Hence, a useful tool for assessing CIMC has been developed. For cities with the largest populations (100,000 and more inhabitants) after validation, the measurement scale included 16 items spanning three constructs: seven items for "strategic consistency" (SC), five items for "interactivity" (IN), and four items for "stakeholder-centred focus" (SF).

For cities with fewer than 100,000 inhabitants, CIMC is described by two dimensions: interactivity (IN) and stakeholder-centred focus (SF). Isolating these two dimensions for all cities has a strong theoretical basis, because IMC is increasingly seen as the stakeholder-centred interactive process of brand communication that leads to mutual profitable relations (Duncan and Mulhern, 2004; L. Porcu, del Barrio-García and Kitchen, 2012; Porcu et al., 2017).

The surprising aspect of the statistical analysis is that the strategic consistency (SC) construct was not distinguished for the two subcategories of smaller cities. The reasons for this may be found in the scale of the activities or competencies within marketing communications. However, empirical research is needed to move beyond mere speculation. Notably, this study identified differences in the composition of items forming the distinct constructs of IMC for various city subcategories, the explanation of which may pose a topic for further research.

The utilitarian aspect of the achievement of constructing an innovative measurement scale can also be pointed out. It forms the basis for the development of model solutions that are relevant to the further implementation of the concept of integrated marketing communications in the field of city marketing. The CIMC scale can also assist city practitioners in their decision-making processes and enable them to detect areas that require support or adjustments. The proposed measurement tool also provides subscales that can be employed in research on CIMC and specific

dimensions of marketing communications. Managers can use the subscales to assess the degree of implementation of each of the three CIMC dimensions. The new measurement scale may also facilitate comparisons of cities in a local and international context.

There are, nevertheless, some limitations to this study, and these offer scope for future research. For instance, the research was conducted exclusively in Poland, thus further studies are necessary to test the proposed scale in other geographical contexts to enhance its generalisability and strengthen the theoretical basis of CIMC research.

The proposed CIMC scale resulted from adapting a measurement framework suitable for business entities to cities; however, researchers may wish to work on expanding dimensions and modifying scale items for cities or their subcategories. The research focused on the marketing communications of municipal offices and did not take into account the point of view of other city entities. Hence, a particular direction for developing CIMC research may be a broader consideration of co-creation processes and stakeholder engagement.

Notwithstanding these limitations, the present research makes a significant contribution to IMC knowledge by developing a measurement framework essential for evaluating and demonstrating the implementation value of CIMC. This study focused on developing and testing a new CIMC measurement scale. The next challenge is to link such a measure to the antecedents supporting CIMC implementation and its effects.The proposed measurement framework also provides a basis for further research, development, and empirical verification of conceptual model solutions in the area of CIMC.

References

Abratt, R., and Kleyn, N. (2012). Corporate identity, corporate branding and corporate reputations: Reconciliation and integration. *European Journal of Marketing*, 46(7), 1048–1063.

Acharya, A., and Rahman, Z. (2016). Place branding research: a thematic review and future research agenda. *International Review on Public and Nonprofit Marketing*. International Review on Public and Nonprofit Marketing, 13(3), 289–317.

Anholt, S. (2004). Editor's foreword to the first issue. *Place Branding*, 1(1), 4–11.

Batra, R., and Keller, K. L. (2016). Integrating marketing communications: New findings, new lessons, and new ideas. *Journal of Marketing*, 80(6), 122–145.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606.

Braun, E., Klijn, H.E., Eshuis, J., and Zenker, S. (2018). Improving place reputation: Do an open place brand process and an identity-image match pay off? *Cities*. Elsevier, 80(June 2017), 22–28.

Braun, E., Kavaratzis, M., and Zenker, S. (2013). My city – my brand: The different roles of residents in place branding. *Journal of Place Management and Development*, 6(1), 18–28.

Castañeda-García, J. A., Frías-Jamilena, M.D., Rodríguez-Molina, A., and Del Barrio-García (2020). The Effect of Message Consistency and Destination-Positioning Brand Strategy Type on Consumer--Based Destination Brand Equity. *Journal of Travel Research*, 59(8), 1447–1463.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*. Routledge, 1(2), 245–276.

Daszkiewicz, M., and Pukas, A. (2016). City marketing communication – challenges and inspirations. In Budzanowska-Drzewiecka, M., and Jedynak, P. (eds.) *Innovations in Certain Areas of Organization Management and Marketing*. Jagiellonian University Press (Contemporary Management and Marketing Dilemmas), 51–65.

Daszkiewicz, M., and Pukas, A. (2019). Implementation of the IMC Concept for Cities: Feasibility and Limitations. In Soliman, K. S. (ed.) *Vision 2025: Education Excellence and Management of Innovations through Sustainable Economic Competitive Advantage. Proceedings of the 34th International Business Information Management Association Conference*. International Business Information Management Association, 10891–10897.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales'. *International Journal of Market Research*, 50(1), 61–77.

Duncan, T., and Caywood, C. (1996). The Concept, Process, and Evolution of Integrated Marketing Communication. In Thorson, E., and Moore, J. (eds.) *Integrated Communication Synergy of Persuasive Voices*. Mahwah, NJ: Lawrence Erlbaum Associates, 13–34.

Duncan, T., and Everett, S. (1993). Client perceptions of integrated marketing communications. *Journal of Advertising Research*, 33(3), 30–39.

Duncan, T., and Moriarty, S. (1997). The IMC audit: Testing the fabric of integration. *Integrated Marketing Communications Research Journal*, 3(1), 3–10.

Duncan, T. R., and Mulhern, F. (2004). *A white paper on the status, scope and future of IMC., IMC Symposium Sponsored by the IMC Programs, Northwestern University and University of Denver.* New York: McGraw Hill.

Duncan, T. R. R., and Moriarty, S. E. (1998). A communication-based marketing model for managing relationships. *Journal of Marketing*, 62(2), 1–13.

Edmiston-Strasser, D. M. (2009). An examination of integrated marketing communication in US public institutions of higher education. *Journal of Marketing for Higher Education*, 19(2), 142–165.

Ewing, M. T., De Bussy, N. M., and Caruana, A. (2000). Perceived Agency Politics and Conflicts of Interest as Potential Barriers to IMC Orientation. *Journal of Marketing Communications*, 6(2), 107–119.

Finne, Å., and Grönroos, C. (2009). Rethinking marketing communication: From integrated marketing communication to relationship communication. *Journal of Marketing Communications*, 15(2–3), 179–195.

Florek, M. (2011). Online City Branding', in Dinnie, K. (ed.) *City Branding: Theory and Cases*. London: Palgrave Macmillan, 82–90.

Fornell, C., and Larcker, D. F. (1981a). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50.

Fornell, C., and Larcker, D. F. (1981b). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 18(3), 382–388.

Foroudi, P., Kitchen, J.P., and Dinnie, K. (2017). IMC antecedents and the consequences of planned brand identity in higher education. *European Journal of Marketing*, 51(3), 528–550.

van Gelder, S. (2011). City Brand Partnerships. In Dinnie, K. (ed.) *City Branding: Theory and Cases*. London: Palgrave Macmillan, 36–43.

Govers, R. (2013). Why place branding is not about logos and slogans. *Place Branding and Public Diplomacy*, 9(2), 71–75.

Hair, J. F., Anderson, R. E., Tatham, R., and Black, W. C. (1998). Multivariate Data Analysis. Englewood Cliffs: Prentice-Hall Inc.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2014). Multivariate Data Analysis. 7th ed. Upper Saddle River. NJ.: Prentice Hall.

Hankinson, G. (2004). Relational network brands: Towards a conceptual model of place brands. *Journal of Vacation Marketing*, 10(2), 109–121.

Hankinson, G. (2007). The management of destination brands: Five guiding principles based on recent developments in corporate branding theory. *Journal of Brand Management*, 14(3), 240–254.

Hanna, S. A., and Rowley, J. (2015). Rethinking Strategic Place Branding in the Digital Age. In Kavaratzis, M., Warnaby, G., and Ashworth, G. J. (eds.) *Rethinking Place Branding: Comprehensive Brand Development for Cities and Regions*. Heidelberg. Heidelberg, New York, Dordrecht, London: Springer International Publishing Switzerland, 85–100.

Harrison-Walker, J. L. (2013). Place brands and the relational branding communication process. *Academy of Marketing Studies Journal*, 16 (Special Issue), 51–59.

Hereźniak, M. (2017). Place Branding and Citizen Involvement: Participatory Approach to Building and Managing City Brands. *International Studies. Interdisciplinary Political and Cultural Journal*, 19(1), 129–141.

Hočevar, N., Žabkar, V., and Mumel, D. (2007). Effects of Top Management Involvement in Integrated Marketing Communications. *Trziste*, 19(2), 159–171.

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. Routledge, 6(1), 1–55.

Jamal, T. B., and Getz, D. (1995). Collaboration theory and community tourism planning. *Annals of Tourism Research*, 22(1), 186–204.

Johnson, G. J., Bruner, G. C., and Kumar, A. (2006). Interactivity and its facets revisited: theory and empirical test. *Journal of Advertising*, 35(4), 35–52.

Kavaratzis, M. (2009). Cities and their brands: Lessons from corporate branding. *Place Branding and Public Diplomacy*, 5(1), 26–37.

Kavaratzis, M. (2012). From "necessary evil" to necessity: Stakeholders' involvement in place branding. *Journal of Place Management and Development*, 5(1), 7–19.

Kavaratzis, M., and Ashworth, G. J. (2007). Partners in coffeeshops, canals and commerce: Marketing the city of Amsterdam. *Cities*, 24(1), 16–25.

Kavaratzis, M., and Hatch, M. J. (2013). The dynamics of place brands: An identity-based approach to place branding theory. *Marketing Theory*, 13(1), 69–86.

Kavaratzis, M., and Kalandides, A. (2015). Rethinking the place brand: the interactive formation of place brands and the role of participatory place branding. *Environment and Planning A*, 47(6), 1368–1382.

Kerr, G., and Kelly, L. (2017). IMC education and digital disruption. *European Journal of Marketing*, 51(3), 406–420.

Kerr, G., and Patti, C. (2015). Strategic IMC: From abstract concept to marketing management tool. *Journal of Marketing Communications*. Taylor & Francis, 317–339.

Kitchen, P. J., Brignell, J., Li, T., and Jones, G. S. (2004). The emergence of IMC: A theoretical perspective. *Journal of Advertising Research*, 44(1), 19–30.

Kitchen, P. J., Kim, I., and Schultz, D. E. (2008). Integrated Marketing Communications: Practice Leads Theory. *Journal of Advertising Research*, 48(4), 531–546.

Kitchen, P. J., and Schultz, D. E. (2009). IMC: New horizon/false dawn for a marketplace in turmoil? *Journal of Marketing Communications*, 15(2–3), 197–204.

Kliatchko, J. (2008). Revisiting the IMC construct A revised definition and four pillars. *International Journal of Advertising*, 27(1), 133–160.

Krosnick, J. A., and Presser, S. (2010). Question and Questionnaire Design. In Wright, J. D. and Marsden, P. V. (eds.) *Handbook of Survey Research.* 2nd ed. Emerald Publishing, 263–313.

Lee, D. H., and Park, C. W. (2007). Conceptualization and measurement of multidimensionality of integrated marketing communications. *Journal of Advertising Research*, 47(3), 222–235.

Low, G. S. (2000). Correlates of integrated marketing communications. *Journal of Advertising Research*, 40(1/2), 27–39.

Lucarelli, A. (2018). Co-branding public place brands: Towards an alternative approach to place branding. *Place Branding and Public Diplomacy*, 14(4), 260–271.

Lucarelli, A., and Olof Berg, P. (2011). City branding: A state-of-the-art review of the research domain. *Journal of Place Management and Development*, 4(1), 9–27.

Luxton, S., Reid, M., and Mavondo, F. (2015). Integrated marketing communication capability and brand performance. *Journal of Advertising*, 44(1), 37–46.

Luxton, S., Reid, M., and Mavondo, F. (2017). IMC capability: Antecedents and implications for brand performance. *European Journal of Marketing*, 51(3), 421–444.

MacCallum, R. C., and Widaman, K. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99.

Manser Payne, E., Peltier, J. W., and Barger, V. A. (2017). Omni-channel marketing, integrated marketing communications and consumer engagement: A research agenda. *Journal of Research in Interactive Marketing*, 11(2), 185–197.

McGrath, J. M. (2005). A pilot study testing aspects of the integrated marketing communications concept. *Journal of Marketing Communications*, 11(3), 191–214.

Merrilees, B., Miller, D., and Halliday, S. (2016). Brand-meaning co-creation by stakeholders: An interactive city brand. *UH Business School Working Paper*. University of Hertfordshire, 1–25. Available at: https://uhra.herts.ac.uk/dspace/handle/2299/5549%5Cnwww.mbsportal.bl.uk.

Merrilees, B., Miller, D., and Herington, C. (2012). Multiple stakeholders and multiple city brand meanings. *European Journal of Marketing*, 46(7), 1032–1047.

Moilanen, T. (2015). Challenges of city branding: A comparative study of 10 European cities. *Place Branding and Public Diplomacy*. Nature Publishing Group, 11(3), 216–225.

Moilanen, T., and Rainisto, S. (2009a). City and Destination Branding. In *How to Brand Nations, Cities and Destinations*. London: Palgrave Macmillan, 77–146.

Moilanen, T., and Rainisto, S. (2009b). *How to Brand Nations, Cities and Destinations*. London: Palgrave Macmillan.

Molinillo, S., Morrison, M.A., Anaya-Sánchez, R., and Coca-Stefaniak, A.J. (2019). Smart city communication via social media: Analysing residents' and visitors' engagement. *Cities*. Elsevier, 94, 247–255.

Moriarty, S. E. (1996). The Circle of Synergy: Synergy Theoretical Perspectives and an Evolving IMC Research Agenda. In Thorson, E., and Moore, J. (eds.). *Integrated communication: Synergy of persuasive voices*. Mahwah, NJ: Lawrence Erlbaum Associates, 333–353.

Moriarty, S. E., and Schultz, D. E. (2012). Four theories of how IMC works. In Rodgers, S., and Thorson, E. (eds.) *Advertising theory*. New York, NY.: Routledge, 491–505.

Mueller, A., and Schade, M. (2012). Symbols and place identity: A semiotic approach to internal place branding – case study Bremen (Germany). *Journal of Place Management and Development*, 5(1), 81–92.

Mulhern, F. J. (2009). Integrated marketing communications: From media channels to digital connectivity. *Journal of Marketing Communications*, 15(2–3), 85–101.

Munoz-Leiva, F., Porcu, L., and Del Barrio-Garcıa, S. (2015). Discovering prominent themes in integrated marketing communication research from 1991 to 2012: A co-word analytic approach. *International Journal of Advertising*. 34(4), 678–701.

Navarro, M. A., Delgado, E., and Sicilia, M. (2010). Integrated Marketing Communications: A Test for Different Levels of Strategic Consistency. In Terlutter, R., Diehl, S., and Okazaki, S. (eds.). *Advances in Advertising Research (Vol. 1): Cutting Edge International Research*. Wiesbaden: Gabler, 3–20.

Pauwels, K., Pauwels, K., Demirci, C., Yildirim, G, and Srinivasan, S. (2016). The impact of Brand Familiarity on Online And Offline Media Synergy. *International Journal of Research in Marketing*, 33(4), 739–753.

Peltier, J. W., Schibrowsky, J., Zahay, D., and Schultz, D.E.(2006). Interactive IMC: The relational--transactional continuum and the synergistic use of customer data. *Journal of Advertising Research*, 46(2), 146–159.

Peltier, J. W., Schibrowsky, J. A., and Schultz, D. E. (2003). Interactive integrated marketing communication: Combining the power of IMC, the new media and database marketing. *International Journal of Advertising*, 22(1), 93–115.

Pettegrew, L. S. (2001). If IMC is so good, why isn't it being implemented? Barriers to IMC adoption in corporate America. *Journal of Integrated Communications*, 29(9), 29–37.

Phelps, J. E., and Johnson, E. (1996). Entering the quagmire: examining the "meaning" of integrated marketing communications Communications No 3. *Journal of Marketing*, 2(3), 159–172.

Pike, S. (2008). *Destination marketing. An Integrated marketing communication approach*. Oxford: Butterworth.

Porcu, L., del Barrio-Garcia, S., Alcántara-Pilar, J. M., and Crespo, E. (2017). Do adhocracy and market cultures facilitate firm-wide integrated marketing communication (Imc)? *International Journal of Advertising*, 36(1), 121–141.

Porcu, L., del Barrio-García, S., Alcántara-Pilar, J. M., and Crespo, E. (2019). Analyzing the influence of firm-wide integrated marketing communication on market performance in the hospitality industry. *International Journal of Hospitality Management*. Elsevier, 80(November 2017), 13–24.

Porcu, L., del Barrio-García, S., Kitchen, P. J., and Tourky, M. (2019). The antecedent role of a collaborative vs a controlling corporate culture on firm-wide integrated marketing communication and brand performance. *Journal of Business Research*. Elsevier, (October), 0–1.

Porcu, L., del Barrio-García, S., and Kitchen, P. J. (2012). How does Integrated Marketing Communications (IMC) work? A theoretical review and an analysis of its main drivers and effects. *Comunicacion y Sociedad*, 25(1), 313–348.

Porcu, L., Del Barrio-García, S., and Kitchen, P. J. (2017). Measuring integrated marketing communication by taking a broad organisational approach: The firm-wide IMC scale. *European Journal of Marketing*, 51(3), 692–718.

Porcu, Lucia, del Barrio-García, S., and Kitchen, P. J. (2012). How does integrated marketing communications (IMC) work? A theoretical review and an analysis of its main drivers and effects. *Comunicación y sociedad/ Communication & Society*, 25(1), 313–348.

Rehmet, J., and Dinnie, K. (2013). Citizen brand ambassadors: Motivations and perceived effects. *Journal of Destination Marketing and Management*. Elsevier, 2(1), 31–38.

Reid, M. (2005). Performance auditing of integrated marketing communication (IMC) actions and outcomes. *Journal of Advertising*, 34(4), 41–54.

Sapnas, K. G., and Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement*, 10(2), 135–153.

Schultz, D. E. (1996). The inevitability of integrated communications. *Journal of Business Research*, 37(3), 139–146.

Schultz, D. E., Kim, I., Kerr, G., and Patti, Ch. (2007). In search of a theory of integrated marketing communication. *Journal of Advertising Education*, 11(2), 21–31.

Schultz, D. E., Kim, I., and Kang, K. (2014). Integrated Marketing Communication Research: Its Limited Past and Huge Potential. In Cheng, H. (ed.) *The Handbook of International Advertising Research*. Hoboken, NJ: Wiley, 457–483.

Schultz, D. E., and Patti, C. H. (2009). The evolution of IMC: IMC in a customer-driven marketplace. *Journal of Marketing Communications*, 15(2–3), 75–84.

Šerić, M. (2018). Content analysis of the empirical research on IMC from 2000 to 2015. *Journal of Marketing Communications*. Routledge, 24(7), 647–685.

Šerić, M., and Vernuccio, M. (2019). The impact of IMC consistency and interactivity on city reputation and consumer brand engagement: The moderating effects of gender. *Current Issues in Tourism*. Routledge, 0(0), 1–19.

Skinner, H. (2005). Wish you were here? Some problems associated with integrating marketing communications when promoting place brands. *Place Branding*, 1(3), 299–315.

Statistics Poland (2020). Area and population in the territorial profile in 2020. Available at: https://stat.gov.pl/en/topics/population/population/area-and-population-in-the-territorial-profile-in-2020,4,14.html.

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*. Routledge, 25(2), 173–180.

Steiger, J. H., and Lind, C. (1984). Statistically based tests for the number of common factors. In *Annual meeting of the Psychometric Society*. Iowa City, IA.

Tabachnick, B. G., and Fidell, L. S. (2007). Using Multivariate Statistics. Boston: Pearson Education Inc.

Tafesse, W., and Kitchen, P. J. (2017). IMC - An integrative review. *International Journal of Advertising*. Taylor & Francis, 36(2), 210–226.

Thorson, E., and Moore, J. (eds) (1996). Integrated communication: Synergy of persuasive voices. Hillsdale NJ: Lawrence Erlbaum.

Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.

Vuignier, R. (2017). Place branding & place marketing 1976–2016: A multidisciplinary literature review. *International Review on Public and Nonprofit Marketing*. International Review on Public and Nonprofit Marketing, 14(4), 447–473.

Wang, Y. J., Wu, C., and Yuan, J. (2009). The Role of Integrated Marketing Communications (IMC) on Heritage Destination Visitations. *Journal of Quality Assurance in Hospitality & Tourism*, 10(3), 218–231.

Zahay, D., Krishen, A., Peltier, J. W., and Schultz, D. E. (2014) Organisational processes for B2B services IMC data quality. *Journal of Business and Industrial Marketing*, 29(1), 63–74.

Zenker, S. (2011). How to catch a city? The concept and measurement of place brands. *Journal of Place Management and Development*, 4(1), 40–52.

Received: June 2022, revised: October 2022

Acknowledgment: *The project was financed by the Ministry of Science and Higher Education in Poland under the programme Regional Initiative of Excellence 2019–2022, project number 015/RID/2018/19, total funding amount 10,721,040 PLN.*

Appendix 1

Table A1.

Characteristics of the sample and the population

Polish territorial units with city rights	Sample		Population	
	n	%	n	%
Number of inhabitants				
10,000 to 19,999	131	47	180	45
20,000 to 99,999	118	42	180	45
100,000 and more	30	11	38	10
Total	279	100	398	100
Key city functions(each city has an administrative function) Respondents could select 1-3 options				
Tourist	142	51	n.a.	n.a.
Business	33	12		
Industrial	96	34		
Cultural	111	40		
Academic	8	3		
Sport	4	1		
Trade and service	86	31		
Total	279	100		
Organisational unit responsible for city marketing communication				
Marketing communication/promotion department	260	93	n.a.	n.a.
Marketing department	19	7		
Total	279	100		
Document containing guidelines for city marketing communication				
Separate document on marketing communication/ brand strategy/promotion	84	30	n.a.	n.a.
Part of the city development strategy	156	56		
No document in this regard	39	14		
Total	100	100		

Note: n.a. = not available

Source: own elaboration.

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Li Sheng**, *Yechang Yin***, *Anning Zhang****, *Jiwei Wu*****, *Ziqing Yang******

Modelling casino hospitality business cycles

This study decomposes the casino hospitality business cycles of Las Vegas and Macao into high-growth states (HGS) and low-growth states (LGS) using a Markov switching model. The casino gaming sector in Macao experiences greater fluctuations than the sector in Las Vegas due to more volatility in tourism flows; that is, Macao has a slightly higher HGS and a considerably lower LGS than Las Vegas. Las Vegas's hospitality cycle appears to be more robust than Macao's, although both hospitality cycles are desirably asymmetric. Various factors, including external business cycles and supply-side factors, affect local hospitality cycles. In terms of policy suggestions, the study's results suggest that promotional marketing must be strengthened in Las Vegas, and Macao must diversify its industrial base.

Keywords: business cycle, hospitality cycle, casino gaming

JEL Classification: R11
DOI: 10.15611/aoe.2023.1.02

Quote as: Sheng, L., Yin, Y., Zhang, A., Wu, J., Yang, Z. (2023). Modelling casino hospitality business cycles. *Argumenta Oeconomica*, 1(50), 33-42.

1. Introduction

Tourism and hospitality are heavily influenced by business cycles. Business cycles, which consist in fluctuations around the growth trend in gross domestic product (GDP), are primarily triggered by changes in aggregate demand, although changes in technology and investment also play a role. Tourism destination tourist flows may be affected by business cycles in the destination's source markets because demand in the tourism sector is determined by income levels. It is no coincidence that hospitality cycles in tourism destinations follow business cycles in tourism

* School of Political Science and Public Administration of Shandong University, China. ORCID: 0000-0002-2700-9975.
** Faculty of Social Sciences, University of Macau, Macao. ORCID: 0000-0002-6058-2773.
*** Faculty of Social Sciences, University of Macau, Macao. ORCID: 0000-0003-4240-8576.
**** Faculty of Social Sciences, University of Macau, Macao. ORCID: 0000-0002-6730-7176.
***** Faculty of Social Sciences, University of Macau, Macao. ORCID: 0000-0003-4269-9481.

source markets and are mediated by tourism flows. Typical travel destinations, as small open economies, are prone to growth volatility. In this study, the authors examine two of the world's largest gaming and tourism cities: Macao and Las Vegas. As demonstrated by periods of rapid expansion and sharp contraction, the casino gaming industry is characterised by wild volatility in both regions. With the aim of promoting sustainable tourism development in these two small open economies, the study attempted to identify effective policy solutions to facilitate the creation of stable growth in the hospitality sector.

The hospitality cycle in Las Vegas is a notable example of how business cycles affect the hospitality sector. Excluding the current restrictions due to the Covid-19 pandemic, casino tourism in Las Vegas is a long-term success story. However, the city has dealt with two major disruptions. Nevada's rapid growth in tourism and hospitality made it one of the top three fastest growing states in the US throughout the last four decades of the 20th century (Eadington, 1999). Until 11 September 2001, the industry was booming in Las Vegas and elsewhere in Nevada. Yet, despite once being considered recession-proof, the tourism and hospitality industries were vulnerable to this adverse event, with casinos becoming quiet and hotel rooms largely vacant. It took until 2005 for the city to regain its footing as the nation's favourite playground thanks to efforts by the state and local governments (Schwartz, 2006). The terrorist attack caused the strong recession, which lingered for years, and during this time net losses by their companies forced many employees to take early retirement. Numerous casino hospitality projects were cancelled or postponed, and some that continued faced financial difficulties. In 2010, the local economy improved, and this growth accelerated after 2013 as the number of visitors from international sources grew, especially visitors from China and other Asian countries (Sheng and Zhao, 2016).

Macao is a successful example of the casino hospitality industry in China, however it has also had bitter experiences with the pitfalls inherent in gambling tourism (Hao et al., 2017). In 2003, the Free Travel Policy (FTP) of mainland China (MC) began to spur economic growth in Macao. The FTP entitles mainland Chinese to visit Macao to partake in gambling (which is prohibited in the rest of the country). Over a three-year period, Macao became the top casino destination in the world, surpassing Las Vegas in terms of gross gaming revenue (GGR). A steep GGR growth of 28.2% per year throughout the 2002–2013 period drove Macao's GDP growth to 13.0% per year on average. At the time, the GGR of Macao was seven times that of Las Vegas (US$ 45.2 billion). Consequently, Macao's GDP per capita rose to within the top three globally and was more than twice that of Hong Kong (Sheng, 2014). However, a 49.4% drop in GGR and a 28.9% decline in GDP occurred between 2013Q4 and 2016Q2, resulting in massive unemployment among migrant workers and large wage reductions among resident workers (Sheng, et al., 2017). This dependence on the casino hospitality industry makes Macao susceptible to external

shocks. Several events demonstrate this, such as the tightening of the FTP during the 2008–2009 period and the anti-corruption campaign during the 2014–2016 period under Xi Jinping's administration. Two severe economic recessions followed these exogenous tourism shocks, with the second being far worse than the first (Li and Sheng, 2018). Despite the economic recovery that has been ongoing since 2016Q3, the instability of Macao's economy continues to be a concern, given the vulnerability of the city's gaming industry to external shocks.

This study used Macao and Las Vegas, which are fast growing but volatile economies, as case studies of hospitality cycles. While Markov switching models (MSM) are widely used in macroeconomics to study business cycles, few studies examine the effect of a single industry on an economy. In particular, there is a lack of research on the tourism industry and hospitality cycle and a dearth of knowledge about the gaming industry cycle. This paper discusses Las Vegas and Macao from a new perspective (Deng et al., 2020). Most other studies, such as Moore and Whitehall (2005) and Merida et al. (2016), applied vector autoregressive (AR) methods to examine local and external cycles. Using data simulations, these theoretical studies investigated the effect of business cycle fluctuations in source markets on hospitality cycles in destination markets. In contrast, this study examines the specific effects of multiple factors characterising external business cycles on local hospitality cycles.

Some studies employ this two-stage treatment to detect the channels, such as international trade and exchange rates, through which external factors affect the local hospitality industry (Sheng, 2011). By incorporating both external and internal factors, the authors identified more factors that drive local hospitality cycles, the model used in this study also accounts for cyclical factors. In the panel data regression of gaming hospitality cycles, system estimation rather than a fixed-effect model was used, which allowed to compare the two destinations in a heterogeneous manner even though they share some customers.

The gambling and hospitality sectors in Macao and Las Vegas constitute a large share of their regions' GDPs, making them ideal settings for the analysis of business cycles. This study should help to fill the gaps in the literature. A thorough analysis of the hospitality cycles and the potential determinants of gaming hospitality cycles in the two largest gaming hospitality destinations in the world provides local tourism bureaus and hospitality managers with much needed evidence of the effects of specific strategies. The empirical results will be useful to local governments seeking to address the vulnerability of their tourism sectors. Economic volatility can be studied in other industries using the same method (Sheng and Wan, 2017).

In Section 2, the authors briefly review the relevant literature. The MSM analysis is presented in Section 3. The paper concludes by providing policy recommendations in Section 4.

2. Literature review

After briefly reviewing the literature on hospitality and economic cycles, the relevant studies in tourism and hospitality (Croes and Ridderstaat, 2017) and economics (Sheng and Nascimento, 2021) are examined in more detail. Although the trend is largely stable, policymakers and macroeconomic theorists are often concerned by the fluctuations, which are often characterised by business booms and recessions. Approaches to business cycles can be divided into two broad categories. Heterodox economists distrust smooth growth trends and consider them to be prone to large and frequent shocks that permanently alter GDP. Traditional and neo-Keynesian models emphasise supply shocks (Sheng, 2012), whereas real business cycle models focus on demand shocks because of their propagating effects.

AR integrated moving averages and Kalman filter models (Campbell and Mankiw, 1987). Despite their computational convenience, both approaches fail to fully capture the observed asymmetrical characteristics of business cycles due to their use of linear modelling. MSM can address this challenge of time-series analysis (Hamilton, 1989) using regime switching regression. Non-linear filters can be used to estimate non-stationary series with AR processes. AR can traverse through unknown states and change over time according to the state in which it occurs. It is assumed that the time it takes to switch from one state to another and the duration of each state are both determined by a Markov probability matrix. MSM are used to study asymmetric behaviour in economics, finance, social science, and political science.

Due to their lack of importance to large economies, hospitality cycles have attracted limited research attention. Prior to 2010, relatively few studies explored the business cycles of national economies (e.g. Gonzalez and Morales, 1996; Greenidge, 2001; Sheng, 2018). More recently, travel destinations have been treated as small open economies and a growing number of studies examined their cyclical movements. In these economies, local hospitality development drives growth, but cross-border tourism promotion also affects business cycles due to the cyclical nature of the economy in the source market (Sheng, Li, and Gao, 2019; Sheng, Yin and Zhang 2022). This study focuses on the links between a local tourism and hospitality industry and movements in its source market.

A growing number of studies discuss the cyclical nature of tourist flows (Guizzardi and Mazzocchi, 2010; Papatheodorou et al., 2010). Most of these studies focus on the tourism demand in specific destinations (Smeral, 2012; 2016). This kind of demand is driven largely by tourists' income, which is influenced by business cycles in their home markets. Furthermore, given the slow reaction of the travel cycle to the business cycle, the past value of the source country's business cycle might provide a useful guide when forecasting the current value of tourism at tourist sites. According to Smeral (2012), fluctuations in economic activity can be viewed as the cause of fluctuations in travel demand.

Empirical studies show that business cycles and tourism demand cycles are strongly correlated. Guizzardi and Mazzocchi (2010) argued that the broader economic cycle significantly influences the tourism cycle. As they noted, there is always a recurring gap between the turning point of the business cycle and the cycle of tourism demand. More recent studies (Kozic, 2014, Sheng, Gu and Guo in press) looked at the lagged effects of hospitality cycles (asymmetry, amplitude, and diffusion). These patterns allow the managers of tourism destinations to benefit from global economic slowdowns if they predict them in due time. A tourism location can experience cycles resulting from various factors (such as maturity). The life cycle theory provides a more comprehensive explanation of these phenomena (Moore and Whitehall, 2005). Considering such findings, policy makers might be able to introduce countercyclical policies to mitigate the effects of external shocks on local economies. Several of these studies are empirical but need further development in terms of the model specifications and the techniques used for estimation. Moreover, theoretical models of the underlying economics are necessary for such investigations.

3. MSM discussion

This section presents a brief overview of Hamilton's (1989) two-state MSM. As it is commonly assumed that empirical data sets contain random errors, the analysis of a time series of financial data is usually nondeterministic. This phenomenon may be influenced by observed and unobserved variables, the latter are referred to as state variables. Institutional transition dynamics are caused by the transformation of related states over time according to deterministic or stochastic laws. The discrete state-space model, which considers the Markov transformation of a specific financial framework, is the only model used in this study. This analysis is intended to evaluate the stochastic growth process from a rough time series acquired over period *t*. MSM can provide reliable projections of the economic state under a business cycle and yield the growth rate associated with expansions and contractions. Among the many advantages of MSM is their ability to identify patterns over time and to guarantee regime permanence and their flexible implementation, all of which mitigate the difficulties associated with using linear AR models to forecast and perform specification tests.

In this paper, the time-series properties of a gaming hospitality cycle are studied using an econometric model. *TR*, the total revenue generated from operations in a casino resort, is seen as a comprehensive indicator of the state of the business cycle of the industry. It is a key variable, as it represents the market interactions between the supply and demand forces in cross-border gaming markets.

The selected economic model also explains how shocks originating in source markets are transmitted through tourist flows into destination markets in the form of tourism revenues. The study's findings suggest that external business cycles

propagate so strongly (or weakly) that the shock to the external cycle can affect local business cycles and have strong or long-lasting (or slight and transient) effects.

A two-state MSM is developed as growth and revenue *TR* in the first stage of the estimation:

$$Z_t = \mu_{S_t} + \sum_{k=1}^{q} \Theta_k L^k Z_t + \sigma_{S_t} \varepsilon_t, \tag{1}$$

where ϕ_k's are coefficients of an order-q autoregressive process $AR(q)$, L is the lag operator, $S_t = \{0,1\}$ denotes two unobserved states of the hospitality cycle, and $\varepsilon_t \sim$ i.i.d. $N(0,1)$ is the white noise.

Two statistical rules are used to select a time lag length that is appropriate for the $AR(q)$ process of Equation (1). To test whether a regime shift has taken place, H_0: $\mu_0 = \mu_1$ is derived from the Wald statistic:

$$w = \left(\hat{\mu}_0 - \hat{\mu}_1\right)^2 / \left[\hat{v}ar\left(\hat{\mu}_0\right) + \hat{v}ar\left(\hat{\mu}_1\right) + 2\hat{c}ov\left(\hat{\mu}_0, \hat{\mu}_1\right)\right] \sim x^2. \tag{2}$$

Conclusion

The authors studied the effects of business cycles on the casino tourism industry in the two largest gambling cities in the world. Using a two-state MSM, the study first estimated hospitality cycles, and then examined the determinants of hospitality cycles using a structural regression model. The findings are consistent with the literature on this point. Similarly, small tourism-dependent economies show evidence of linked source-destination cycles. This study examines Las Vegas and Macao as distinct but somewhat integrated markets. The findings have implications for policy makers seeking to mitigate tourism vulnerability and economic volatility in both cities. However, local pull factors, such as service offerings, also play a key role. To maintain a high-growth state, Las Vegas must provide marketing campaigns that target Chinese and other Asian customers. Since Macao's economy would otherwise continue to suffer large fluctuations leading to an eventual long-term low growth state, it must diversify its industrial base and pursue sustainable tourism development.

The study provides the following main policy recommendations for the benefit of local tourism authorities and hospitality business managers. Although Las Vegas and Macao have similar hospitality cycles, in particular baccarat players constitute the core of customers in both locations, boosting the casino gaming industry in Las Vegas requires different approaches that boosting the industry in Macao due to three structural differences. First, the Las Vegas hospitality cycle is less volatile, and its two-state asymmetry is more favourable than Macao's; gambling receipts represent only 37% of casino revenue in Las Vegas compared to more than 90% in Macao. Second, in Las Vegas, slot machines generate revenue, whereas Macao relies on table games for its success (Sheng and Gao, 2018). Third, even though Las Vegas

baccarat games made up 47% of table game revenue in 2013 (up from 13% in 1985), this proportion remains lower than that in Macao. During domestic business-cycle downturns, Las Vegas might need to increase its marketing efforts (albeit at a cost) around the globe to smooth gaming hospitality cycles. This should be accomplished by focusing promotion efforts on Chinese investors and new Asian markets. Macao must learn from Las Vegas about mass-market gaming and economic diversification, whereas Las Vegas might benefit from implementing something akin to Macao's offshore tourism marketing programme for VIP hospitality operations (Sheng, 2017a; Sheng et al., 2022).

The lack of natural surroundings and amenities in Macao prevents it from becoming a destination for outdoor/adventure tourism. As such, developing high value-added industries that use a minimal amount of land is a realistic solution to industrial diversification. There are three industries that Macao could take advantage of: (i) meetings, incentives, conferences/conventions, and events/exhibits/expositions (MICE), (ii) sea-boat tourism, and (iii) offshore RMB finance. The city has been building hospitality facilities to allow MICE industry activities, making Macao competitive with nearby cities. Local casino hotels will also benefit from increasing MICE industry activity. South China's Greater Bay Area development plan is likely to generate new MICE business (Sheng, 2016). The Chinese government granted Macao 85 square kilometres of sea territory in 2015 (more than twice its land area); setting up sea-boat hospitality services would be a profitable way to utilise this new natural resource to attract curious tourists. There have been several riverboat casino licences issued by different states in the US to enable gambling, however there are a limited number of regions where this industry is feasible. Although VIP gambling could be extended to sea-boat casino hotels, Macao's sea-boat tourism industry could be expanded beyond traditional gambling tourism (Sheng, 2017b). In addition, RMB transactions outside of MC are now permitted under offshore financing rules. In addition to attracting tourists to Macao for recreational and financial reasons, this non-gaming business could attract many tourists. The fact that Macao has accumulated a huge amount of public savings suggests the need to offer tax incentives to promote the growth of the three industries discussed above, effectively diversifying the economy. To increase the overall stability of the local economy, it will be necessary to develop industries that are less susceptible to adverse conditions, thereby enabling growth to become less volatile and more sustained.

References

Campbell, J. Y. and Mankiw, N. G. (1987). Permanent and transitory components in macroeconomic fluctuations. *American Economic Review, Papers and Proceedings*, 77, 111-117.

Croes, R. and Ridderstaat, J. (2017). The effects of business cycles on tourism demand flows in small island destinations. *Tourism Economics*, 23(7), 1451-1475.

Deng, Q., Gu, X., Law, R. and Lian, Z. (2020). A comparative study for determinants of gaming performance in Macao and Las Vegas. *Tourism Management*, 77, 103964.

DICJ (Direcção de Inspecção e Coordenação de Jogos) (2018). *Information, Gaming Statistics, and Quarterly Data*. http://www.dicj.gov.mo/web/en/information/index.html.

Eadington, W. R. (1999). The economics of casino gambling. *Journal of Economic Perspectives*, 13, 173-192.

Engel, C. and Hamilton, J. D. (1990). Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review,* 80, 689-713.

Gonzalez, P. and Morales, P. (1996). Analysis of tourism trends in Spain. *Annals of Tourism Research*, 23(4), 739-754.

Greenidge, K. (2001). Forecasting tourism demand: An STM approach. *Annals of Tourism Research*, 23(4), 739-754.

Guizzardi, A. and Mazzocchi, M. (2010). Tourism demand for Italy and the business cycle. *Tourism Management*, 31(3), 367-377.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357-384.

Hao, Y, Sheng, L. and Pan, G. (2017). *Political Economy of Macao: Dilemma of Its Success*. Palgrave Macmillan: London.

Kozic, I. (2014). Detecting international tourism demand growth cycles. *Current Issues in Tourism*, 17(5), 309-403.

Li, T. and Sheng, L. (2018). Corruption and travel: Effects of China's anti-graft campaign on Macao. *Die Erde*, 149(1), 44-51.

Mayers, S. and Jackman, M. (2011). Investigating the business cycle properties of tourist flows to Barbados. *Journal of Public Sector Policy Analysis*, 5, 3-21.

Merida, A. and Golpe, A. A. (2016). Tourism-led growth revisited for Spain: Causality, business cycles and structural breaks. *International Journal of Tourism Research*, 18, 39-51.

Moore, M. and Whitehall, P. (2005). The tourism area lifecycle and regime switching models. *Annals of Tourism Research*, 32(1), 112-126.

Schwartz, D. (2006). Roll the bones: The history of gambling. *Gotham Books*, University of Nevada Las Vegas Center for Gaming Research.

Sheng, L. (2011). Foreign investors versus local businesses: An urban economics model for tourist cities. *International Journal of Tourism Research*, 13(1), 32-40.

Sheng, L. (2012). Dealing with financial risks of international capital flows: A theoretical framework. *Cambridge Review of International Affairs* 25(3), 463-474.

Sheng, L. (2014). Effects of foreign expansion on local growth: The case of Macao. *European Planning Studies*, 22(8), 1735-1743.

Sheng, L. (2016). Explaining the transformation of urban island politics: The case of Macau. *Island Studies Journal*, 11(2), 521-536.

Sheng, L. (2017a). Explaining urban economic governance: The city of Macao. *Cities*, 61(2017), 96-108.

Sheng, L. (2017b). Factors determining the success or failure of a tourism tax: A theoretical model. *Tourism Review*, 72(3), 274-287.

Sheng, L. (2018). Hong Kong and Shanghai: A tale of two cities in China. *Tempo Social*, 30(3), 171-190.

Sheng, L., Li, T and Gao, J. (2019). Effects of foreign capital on tourism: Evidence from Macao and Hong Kong. *Ekonomista* 2/2019, 243-254.

Sheng, L. Li, T. and Wang, J. (2017). Tourism and externalities in an urban context: Theoretical model and empirical evidence. *Cities*, 70(2017), 40-45.

Sheng, L., and Gao, J. (2018). Foreign investors versus host communities: An urban political economy model for tourist cities. *Argumenta Oeconomica*, 41(2), 257-275.

Sheng, L., Gu, X. H., and Guo, H. Z. (in press). Business cycles of casino cities: Theoretical model, empirical evidence and policy implications. *Journal of Urban Affairs*.

Sheng, L., and Nascimento, D. F. (2021). *Love and Trade War – China and the U.S. in Historical Context*. London: Palgrave Macmillan.

Sheng, L. and Wan, P. (2017). Explaining urban governance in the midst of political transformation: The city of Macao. *Asia-Pacific Viewpoint*, 58(3), 289-300.

Sheng, L., Yin, Y. and Zhang, A. (2022). Theorising on tourist flows and business cycles in casino cities. *Argumenta Oeconomica*, 48, 187-199.

Sheng, L., Yin, Y., Zhang, A. and Wu, J. (2022). Modelling economic structure: The perspective of tourist cities. *Argumenta Oeconomica*, 49(2), 137-150.

Sheng, L., and Zhao, W. B. (2016). Strategic destination management in the face of foreign competition: The case of Macao. *Journal of Travel & Tourism Marketing*, 33(2), 263-278.

Shu, L. F. (2014). Casino tourism and hospitality in Macao facing the storm of anticorruption in Mainland China. *New Fortune Magazine*, December 28. http://www.xcf.cn/tt2/201412/t20141218_701126.htm.

Smeral, E. (2012). International tourism demand and the business cycle. *Annals of Tourism Research*, 39(1), 379-400.

Smeral E. (2016). Tourism forecasting performance considering the instability of demand elasticities. *Journal of Travel Research*, 55(2), 190-204.

Received: August 2020, revised: December 2021

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Erik Šoltés**, *Silvia Komara***, *Tatiana Šoltésová****,
*Martin Mišút*****

Analysis of work intensity in Slovakia using testing and estimation of linear combinations of GLM parameters

Not only unemployment itself but also the reduced work intensity of a household has a major impact on the social exclusion of a person. The work intensity of households is currently being monitored in Europe mainly for purposes of identifying those people or households that are excluded from the labour market. The households' work intensity directly affects the inclusion or exclusion from the labour market, which is one of the three social exclusion dimensions. Moreover, it also, as confirmed by several studies, fundamentally affects the other two dimensions of social exclusion, namely income poverty and material deprivation. The aim of the paper was to assess which factors in interaction with the economic activity status of a person significantly affect the household's work intensity and, depending on these factors, to estimate the household's work intensity. For this purpose, the general linear model and the associated analysis of marginal means and the contrast analysis were used. The analyses are based on a database EU-SILC 2020 for the Slovak Republic and performed in the SAS Enterprise Guide and by means of PROC GLM in the SAS programming language using CONTRAST and ESTIMATE statements. The article examines between which levels of significant factors there is a significant difference in terms of a household's work intensity and in particular provides estimates of work intensity depending on the household type, educational attainment level and the age of a person. At the same time, in all three cases households are broken down by the economic activity status of the person. The presented analyses revealed categories of persons that are the most and the least threatened by labour market exclusion from the point of view of the considered factors.

Keywords: work intensity, exclusion from the labour market, poverty and social exclusion, general linear model (GLM), least square means, contrast analysis

JEL Classification: C12, C21, E24
DOI: 10.15611/aoe.2023.1.03

Quote as: Šoltés, E., Komara, S., Šoltésová, T., Mišút, M. (2023). Analysis of work intensity in Slovakia using testing and estimation of linear combinations of GLM parameters. *Argumenta Oeconomica*, 1(50), s. 43-66.

* Department of Statistics, University of Economics in Bratislava, Slovakia. ORCID: 0000-0001-8570-6536.
** Department of Statistics, University of Economics in Bratislava, Slovakia. ORCID: 0000-0001-6641-7456.
*** Department of Mathematics and Actuarial Science, University of Economics in Bratislava, Slovakia. ORCID: 0000-0002-0953-2519.
**** Department of Applied Informatics, University of Economics in Bratislava, Slovakia. ORCID: 0000-0002-5545-2624.

1. Introduction

The work intensity (WI) of households is an indicator that is monitored within the framework of sustainable development strategies in the EU, e.g. in the Europe 2030 strategy. Households with very low work intensity are referred to as (quasi-)jobless households and members of such households are considered to be excluded from the labour market. The active participation in the labour market plays an important role in the fight against poverty and social exclusion, hence (quasi)joblessness is one of the three components of the composite indicator AROPE (at risk of poverty or social exclusion), which monitors the EU's progress on social inclusion. Although in 2019 the share of persons living in (quasi-)jobless households in Slovakia (4.8%) was slightly lower than in the EU (6.1%), it is important to identify factors that increase the risk of exclusion from the labour market, also regarding the unfavourable events in the early 2020s. It is a realistic assumption that the COVID-19 pandemic, the energy crisis, inflation, and the war in Ukraine will increase the share of the population in Slovakia (as in other EU countries) that will have to face social exclusion (including exclusion from the labour market).

The above-mentioned facts and presumptions motivated the author to analyse the work intensity in the Slovak Republic. This paper is not limited only to the very low work intensity but focuses on the labour intensity index. This index is a target continuous numerical variable, which was analysed depending on various factors via the analysis of marginal means and contrast analysis, which are based on a general linear model. The aim of the paper was to assess the influence of the most fundamental socio-economic and socio-demographic factors on work intensity, while assessing the influence of other relevant factors. The following research tasks are also oriented on the most relevant factors:

- to assess whether the impact of factors on work intensity is different or the same for different statuses of economic activity,
- for each factor, identify categories between which there are no significant differences and identify those categories or clusters of categories between which there are demonstrable differences in work intensity,
- to quantify the mean of work intensity for individual groups of persons and identify risk groups of persons in terms of exclusion from the labour market.

2. Literature review

Exclusion from the labour market significantly increases the risk of material deprivation and income poverty. De Graaf-Zijl and Nolan (2011) stated that the dependence between these three dimensions does not have a consistent pattern in groups of countries classified together in terms of welfare regime or geographically. García-Gómez et al. (2021) found that this dependence increased significantly in the countries most affected by the economic crisis in the period 2008-2014. Duiella and

Turrini (2014) came to a similar conclusion, and also identified a positive relationship among them, which became stronger after 2010 in countries most severely hit by the crisis. Based on the above, it was supposed that this relationship will also intensify due to the COVID-19 pandemic and the deteriorating economic situation caused by the energy crisis, inflation, and the war in Ukraine. Verbunt and Guio (2019) confirmed that work intensity is very effective in explaining within-country differences in the risk of income poverty/material deprivation in some CEE countries (including Slovakia). In Slovakia, after the period of financial and economic crisis, unemployed persons living in households with a high and medium level of work intensity had markedly higher chances to move to employment, compared to the unemployed in households with low work intensity (Gerbery and Miklošovic, 2020). Fabrizi and Mussida (2020) found that living in a work-poor household is associated with living in consistent poverty (people at consistent poverty are those who are both at risk of poverty and simultaneously experiencing enforced deprivation).

In addition, labour market exclusion also has a negative impact on the population of children and young people and on their social exclusion in the future. Guio and Vandenbroucke (2019) stated that (quasi)joblessness is an important driver of child deprivation in Belgium, even when income is controlled for. Analyses by Cantó et al. (2022) revealed that other household members' employment levels and economic difficulties have strong effects on youth economic outcomes.

This paper does not deal only with low (or very low) work intensity, as other degrees of work intensity can also be associated with poverty and social exclusion. For example, Kis and Gábos (2016) showed that in the new member states of the EU not only low and very low household work intensity is positively associated with a higher risk of consistent poverty, but also medium work intensity. Naturally, higher work intensity is positively correlated with social inclusion. Fabrizi and Mussida (2020) showed that higher work intensity of Italian households with dependent children significantly reduces the probability of falling into poverty and social exclusion. Although having a job is not a sufficient condition to avoid poverty, either in terms of (monetary) objective or subjective poverty (Filandri et al., 2020), low work intensity is a crucial micro-determinant of in-work poverty. Hick and Lanau (2017) stated that work intensity of the household is a very strong predictor of in-work poverty. Colombarolli (2021) confirmed that work intensity is negatively associated with in-work poverty, but the relation is stronger with the objective in-work poverty rather than subjective one. Research of work intensity is therefore of great importance also in terms of in-work poverty, the prevention of which is very important for raising living standards and ensuring its convergence in the EU member states.

Poverty and social exclusion analyses use statistical modeling intensively, using different types of generalised linear models. Among the most commonly used are binomial logit models applied by e.g. Ćwiek and Ulman (2019), Šoltés and Ulman (2015), Mysíková et al. (2019), and multinomial logistic models applied by e.g.

Sanchez-Sellero and Garcia-Carro (2020), Calegari et al. (2021), Verbunt and Guio (2019). This study used a general linear model, which is also a special case of a generalised linear model. Unlike other analyses in the field of poverty and social exclusion, the paper focuses on the analysis of marginal means and the contrast analysis.

In selecting the explanatory factors, the author relied on the results of the previous research and the work of other researchers. These factors include the status of economic activity, education, type of household, age, marital status, health condition, region, and degree of urbanisation. The study focused especially on the impact of the first four factors on work intensity, whose significant impact on poverty and social exclusion was confirmed by, among others, Nieuwenhuis and Maldonado (2018) and Peña-Casas et al. (2019).

3. Method

This paper proceeded from the general linear model GLM (Littell et al., 2010), based on which the influence of categorical factors and their interactions on a continuous numerical response variable characterising work intensity were assessed. In terms of interpreting the results, it is important to note that in the research used factors with fixed effects (Searle and Gruber, 2017), and for categorical factors, the author used indicator (dummy) coding (Darlington and Hayes, 2016). The interaction was based on the crossed classification structure (Littell et al., 2010).

The general linear model can be written in matrix form as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

Matrix $\mathbf{X}$ is not full rank, and a generalised inverse method is used to estimate the vector of parameters $\boldsymbol{\beta}$, the result of which is an estimate

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \tag{2}$$

where matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}$ is a generalized inverse matrix that must satisfy at least the first of the Penrose conditions (Searle and Gruber, 2017). The estimation of the vector of parameters $\boldsymbol{\beta}$ obtained by the generalised inverse method is not unique, but there is a group of linear functions of the parameters, referred to as estimable functions $\mathbf{L}\boldsymbol{\beta}$ (Elswick et al., 1991), for which there is a single solution (for more detail see e.g. Agresti, 2015 and Littell et al., 2010).

As the aim of the paper was, among other things, to assess between which categories of relevant factors there is a significant difference, the subject of interest was the testing of general linear hypotheses. To test the general linear hypothesis H_0: $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ (cf. McFarquhar, 2016, Poline et al., 2007, Searle and Gruber, 2017), the following test statistic was used

$$F = \frac{\dfrac{\left(\mathbf{Lb}\right)^{\mathrm{T}} \cdot \left[\mathbf{L}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-}\mathbf{L}^{\mathrm{T}}\right]^{-1} \cdot \left(\mathbf{Lb}\right)}{l}}{\dfrac{SSE}{n-p}}, \tag{3}$$

where l is the number of independent rows of matrix **L**, *SSE* is the sum of the squared residuals, n is the sample size, and p is the number of parameters of the GLM. Thus the null hypothesis is rejected if the value of the test statistic satisfies the inequality

$$F > F_{1-\alpha}(l;\, n-p). \tag{4}$$

The above test was used to test simple hypotheses (if $l = 1$) and to simultaneously test multiple hypotheses (if $l \geq 2$). To test simple hypotheses, a *t*-test was also used, or alternatively, an interval estimate was also constructed (cf. Kuznetsova et al., 2017, Littell et al., 2010, and Westfall and Tobias, 2007).

The analyses presented in the paper were based on unbalanced data, while the author assessed the impact of several effects. In such a situation, group arithmetic means do not provide an adequate picture of the response of the target variable for the particular factor because they do not take into account other effects, which may lead to the Simpson paradox (Wang et al., 2018). Cai (2014) stated that if the data are unbalanced, arithmetic means are not appropriate because they do not consider that not all factors have the same chance of influencing the target variable. In such cases, it is appropriate to estimate the marginal means, which are based on the model (in this case on the GLM). The marginal mean is also referred to as the LS-mean (*Least Squares mean*; Goodnight and Harvey, 1997) or the EM-mean (*Estimated Marginal mean*; Searle et al., 1980). The estimated marginal means or least squares means are predicted means that are calculated from the fitted model and are adjusted appropriately for any other variable (Suzuki et al., 2019).

This study employed marginal mean analysis using the LSMEANS statement and contrast analysis (Dean et al., 2017, Kim and Timm, 2006, Schad et al., 2020) using the CONTRAST and ESTIMATE statements within PROC GLM (SAS Institute Inc., 2018b) and PROC GENMOD (SAS Institute Inc., 2018a) in the SAS programming language. The procedures in SAS presented in this paper are largely universal and are also used in other software and open-source systems (Lenth, 2016, Tabachnick and Fidell, 2013).

4. Database

The author analysed work intensity (WI) in Slovakia using a general linear model with explanatory variables listed in Table 1. The analyses were based on the EU-SILC 2020 database (with the reference year 2019) provided by the Statistical Office of the Slovak Republic. The statistical unit is the person to whom the WI of the household in which this person lives is assigned.

Table 1

Description of input explanatory variables

Original variables (EU-SILC) – categories and description	Names of new dummy variables
1	2
RB210 – Economic activity status	**EAS**
Other inactive person	IP
Unemployed person	UP
Employed person	EP
HT – Household type	**HT**
Single-person household	1A_0Ch
Single parent household with at least 1 dependent child	1A_1^{+}Ch
2-adult household, at least 1 aged 65+	2A(1^{+}R)
2-adult household without dependent children	2A_0Ch
2-adult household with 1 dependent child	2A_1Ch
2-adult household with 3+ dependent children	2A_3^{+}Ch
Other households without dependent children	Other_0Ch
Other households with dependent children	Other_1^{+}Ch
2-adult household with 2 dependent children	2A_2Ch
PE040 – The highest level of education achieved (ISCED)	**Education**
Pre-primary (ISCED 0)	ISCED 0-2
Primary (ISCED 1)	
Lower secondary (ISCED 2)	
Upper secondary (ISCED 3)	ISCED 3-5
Post-secondary (not tertiary) (ISCED 4)	
Short cycle of tertiary education (ISCED 5)	
Bachelor or equivalent (ISCED 6)	ISCED 6-8
Master‘s or equivalent (ISCED 7)	
Doctorate or equivalent (ISCED 8)	
RX010 – Age	**Age**
Age at the end of income reference period	<30
	30-40
	40-50
	50-60
PH010 – General health	**Health**
Very bad	Bad
Bad	
Fair	Fair

1	2
Good	Good
Very good	
PB190 – Marital Status	**Marital status**
Single	Single
Married	Married
Widowed	Widowed
Divorced	Divorced
Region	**Region**
Banská Bystrica	BB
Prešov	PO
Košice	KE
Žilina	ZA
Trenčín	TN
Trnava	TT
Nitra	NR
Bratislava	BA
DB100 – Degree of urbanisation	**Urbanisation**
Thinly populated area	Sparse
Intermediate area	intermediate
Densely populated area	Dense

Source: own elaboration based on EU-SILC data.

The definition of the target variable WI is given by the methodology used by Eurostat to monitor exclusion from the labour market. For the EU-SILC 2020 database, the definition used in the Europe 2020 Strategy was applied. Based on this, the household work intensity was defined as the proportion of the total number of months during which in the course of the income reference year all members of the productive-age household worked, and the total number of months that the same household members could theoretically work, under state legislation, during the same period. A person of productive age means a person aged 18-59 with the exclusion of students in the 18-24 age group. Note that from 2021 onwards, a modified definition was applied, as used in the Europe 2030 Strategy (Eurostat, 2022). As by definition, the WI is not assigned to some persons, information from 7,424 persons was entered in the analysis, although 13,800 persons were included in the EU-SILC 2020 survey.

In the analyses, a continuous numerical variable WI was applied in the sense of the above definition, but also interpreting the results of the analyses with respect to the degrees of work intensity. According to the Eurostat methodology, for work

intensity of households from intervals $\langle 0\%; 20\%)$, $\langle 20\%; 45\%)$, $\langle 45\%; 55\%\rangle$, $(55\%; 85\%\rangle$, and $(85\%; 100\%\rangle$, the degrees of very low work intensity (VLWI), low work intensity (LWI), medium work intensity (MWI), high work intensity (HWI), and very high work intensity (VHWI), respectively, were assigned.

5. General linear model for work intensity

5.1. Regressor selection

Using the stepwise regression method (Agresti, 2015), the regressors listed in Table 2 were included in the model. Naturally, the WI is fundamentally influenced by economic activity. The EAS variable alone explains more than 1/4 of the WI variability (i.e. 28.09%). The impact of the other variables listed in Table 1 (excluding the urbanisation variable) also proved to be significant. Originally, these variables were considered separately (not in interaction), in which case the model explained the WI variability to about one-third. In fact, the above variables affect the WI differently for different statuses of economic activity, which confirms the significance of the individual interactions (Table 2). Thanks to the consideration of interactions, it was possible to substantially increase the explained variability of the WI to more than 50% (i.e. 50.45%).

Table 2

Verification of statistical significance of the model and the influence of factors on WI

Source	DF	Sum of squares	Mean square	F value	Pr > F
Model	77	309.633	4.021	97.12	<.0001
Error	7 346	304.163	0.041		
Corrected total	7 423	613.796			

R-square	Coeff var	Root MSE	WI mean
0.5045	25.121	0.203	0.810

Source	DF	Partial R-square	Model R-square	Type III SS	Mean square	F value	Pr > F
EAS	2	0.2809	0.2809	33.113	16.556	399.86	<.0001
EAS*HT	24	0.1271	0.4080	19.040	0.793	19.16	<.0001
EAS*AGE	9	0.0462	0.4543	8.979	0.998	24.10	<.0001
EAS*EDUCATION	6	0.0267	0.4809	13.564	2.261	54.60	<.0001
EAS*REGION	21	0.0120	0.4929	6.657	0.317	7.66	<.0001
EAS*HEALTH	6	0.0062	0.4991	4.043	0.674	16.27	<.0001
EAS*MS	9	0.0053	0.5045	3.258	0.362	8.74	<.0001

Source: EU-SILC 2020, own processing in SAS EG.

In Table 2, the regressors are ranked/classified/sorted according to their contribution to explaining the variability of the WI variable. After the EAS variable, the WI is most strongly influenced by household type (HT), age and education, which in interaction with economic activity contributed to the explanation of the WI variability at 12.71%, 4.62%, and 2.67%, respectively. The other three interactions had a significant effect on the WI, but their contribution to the explanation of the WI variability was less than 2%.

This paper, in addition to the impact of economic activity itself, focused on quantifying the impact of the other three most important factors, namely the type of household, age and education.

5.2. Analysis of LS means and contrast analysis in GLM

In this section, the author applied the analysis of marginal means, in order to find out between which pairs of categories of individual factors (HT, age, and education) there is a significant difference in terms of WI, when assessing the influence of other factors. Given the interaction of these factors with economic activity, this comparison was made separately for each economic activity. Contrast analysis served to create clusters of several categories of the relevant factor, so that from the WI perspective, there was no significant difference between the categories belonging to the cluster, and at the same time there was a demonstrable difference between the clusters. For such clusters of categories, the study estimated the mean of WI, providing a picture of the impact of individual factors on the WI and allowing to identify the most vulnerable persons in terms of exclusion from the labour market.

5.2.1. Analysis of the impact of interaction *EAS*×*HT*

Since the type of household has the greatest influence on the WI of the assessed factors, it was possible to illustrate the procedure with the example of this factor. In Table 3, there are estimated LS-means of the WI for individual types of households, especially for three statuses of economic activity (since the WI is assigned only to persons under 60 years of age, the status of economic activity Retired was not considered).

The p-values matrices for LS-means equality tests show between which types of households there is no significant difference from the WI point of view. Originally, it was one matrix, which, due to its size, was divided into three submatrices belonging to the individual statuses of economic activity. Therefore, the author did not report test results between individual statuses, but confirmed that the WI means for employees (EP) were significantly higher ($p < 0.0001$) than for the other two statuses, and this applied to all types of households. Next, the study looked at the influence of HT only within the individual statuses of economic activity. While for inactive persons (IP), the differences were insignificant only between pairs of household types, in the case of the other two statuses of economic activity, the similarity

Table 3

Comparison of LS-means of WI for effect *EAS × HT*

EAS = IP			Least Squares means for effect EAS*HT Pr > |t| for H_0: LSMean(i) = LSMean(j) Dependent variable: WI (EAS = IP)								
HT	**WI LSMEAN**	**i**	**i/j**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1A_0Ch	0.0542	1	1		<.0001	**0.5901**	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
1A_1+Ch	0.5560	2	2	<.0001		<.0001	<.0001	0.0065	0.0011	<.0001	<.0001	**0.7009**
2A(1+R)	0.0266	3	3	**0.5901**	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2A_0Ch	0.2936	4	4	<.0001	<.0001	<.0001		<.0001	<.0001	**0.0538**	<.0001	<.0001
2A_1Ch	0.4881	5	5	<.0001	0.0065	<.0001	<.0001		0.2628	<.0001	0.0022	<.0001
2A_3+Ch	0.4618	6	6	<.0001	0.0011	<.0001	<.0001	**0.2628**		<.0001	**0.2605**	<.0001
Other_0Ch	0.3445	7	7	<.0001	<.0001	<.0001	**0.0538**	<.0001	<.0001		<.0001	<.0001
Other_1+Ch	0.4373	8	8	<.0001	<.0001	<.0001	<.0001	0.0022	**0.2605**	<.0001		<.0001
z_2A_2Ch	0.5653	9	9	<.0001	**0.7009**	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

EAS = UP			Least Squares means for effect EAS*HT Pr > |t| for H_0: LSMean(i) = LSMean(j) Dependent variable: WI (EAS = UP)								
HT	**WI LSMEAN**	**i**	**i/j**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1A_0Ch	0.4063	1	1		0.0049	<.0001	**0.9565**	**0.6438**	0.3735	0.4153	0.7073	0.0826
1A_1+Ch	0.0475	2	2	0.0049		**0.4108**	0.0035	0.0020	0.0015	0.0011	0.0021	0.0003
2A(1+R)	0.1531	3	3	<.0001	**0.4108**		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2A_0Ch	0.4091	4	4	**0.9565**	0.0035	<.0001		**0.5994**	**0.3437**	**0.2903**	**0.6606**	0.0363
2A_1Ch	0.4313	5	5	**0.6438**	0.0020	<.0001	**0.5994**		0.5557	0.7248	0.8369	0.0929
2A_3+Ch	0.4676	6	6	**0.3735**	0.0015	<.0001	**0.3437**	**0.5557**		0.6860	0.4389	0.5067
Other_0Ch	0.4442	7	7	**0.4153**	0.0011	<.0001	**0.2903**	**0.7248**	**0.6860**		0.4460	0.1274
Other_1+Ch	0.4239	8	8	**0.7073**	0.0021	<.0001	**0.6606**	**0.8369**	**0.4389**	**0.4460**		0.0426
z_2A_2Ch	0.5120	9	9	**0.0826**	0.0003	<.0001	0.0363	**0.0929**	**0.5067**	**0.1274**	0.0426	

EAS = EP			Least Squares means for effect EAS*HT Pr > |t| for H_0: LSMean(i) = LSMean(j) Dependent variable: WI (EAS = EP)								
HT	**WI LSMEAN**	**i**	**i/j**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1A_0Ch	0.8923	1	1		**0.4188**	**0.3803**	0.0151	<.0001	<.0001	<.0001	<.0001	<.0001
1A_1+Ch	0.8738	2	2	0.4188		**0.1748**	**0.5274**	0.0412	<.0001	**0.0914**	0.0002	0.0165
2A(1+R)	0.9098	3	3	0.3803	0.1748		0.0063	<.0001	<.0001	<.0001	<.0001	<.0001
2A_0Ch	0.8600	4	4	0.0151	**0.5274**	0.0063		0.0032	<.0001	0.0092	<.0001	0.0003
2A_1Ch	0.8277	5	5	<.0001	0.0412	<.0001	0.0032		<.0001	**0.3757**	0.0007	**0.4833**
2A_3+Ch	0.7389	6	6	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	0.0037	<.0001
Other_0Ch	0.8371	7	7	<.0001	**0.0914**	<.0001	0.0092	**0.3757**	<.0001		<.0001	**0.1033**
Other_1+Ch	0.7913	8	8	<.0001	0.0002	<.0001	<.0001	0.0007	0.0037	<.0001		0.0071
z_2A_2Ch	0.8196	9	9	<.0001	0.0165	<.0001	0.0003	**0.4833**	<.0001	**0.1033**	0.0071	

Source: EU-SILC 2020, own processing in SAS EG.

of several types of households appeared. Therefore, to assess the impact of HT on the WI, a more comprehensive analysis for unemployed persons (UP) and employed persons (EP) was required, using a contrast analysis.

A comparison of pairs of marginal WI means for unemployed persons (UP) shows that there was no statistically significant difference ($p = 0.4108$) between the types of 1A_1^{+}Ch and 2A(1^{+}R) and these two types of households had a statistically significantly lower WI at the significance level of 0.05 as the other types of households. There was no significant difference between the other pairs of household types in terms of the WI at significance level of 0.01, so the study verified whether the WI means in these other household types (except 1A_1^{+}Ch and 2A(1^{+}R)) could be considered the same.

Denoting the WI means for the UP economic activity status by μ_{2i}, where index 2 expresses the 2nd category of the EAS factor (UP) and $i = 1, 2, \ldots, 9$ determines the type of household (see Table 1), then the subject of interest was the hypothesis:

$$H_0 : \mu_{21} = \mu_{24} = \mu_{25} = \mu_{26} = \mu_{27} = \mu_{28} = \mu_{29}\,.$$

To test it, a simultaneous test of six null hypotheses was used:

$$H_0 : \mu_{21} = \mu_{24} \;\wedge\; H_0 : \mu(\mu_{21}, \mu_{24}) = \mu_{25} \;\wedge\; H_0 : \mu(\mu_{21}, \mu_{24}, \mu_{25}) = \mu_{26} \;\wedge$$
$$H_0 : \mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}) = \mu_{27} \;\wedge\; H_0 : \mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}, \mu_{27}) = \mu_{28} \;\wedge$$
$$H_0 : \mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}, \mu_{27}, \mu_{28}) = \mu_{29}\,,$$

which was then rewritten into linear combinations

$$H_0 : \mu_{21} - \mu_{24} = 0\,,$$
$$H_0 : \tfrac{1}{2}\mu_{21} + \tfrac{1}{2}\mu_{24} - \mu_{25} = 0\,,$$
$$H_0 : \tfrac{1}{3}\mu_{21} + \tfrac{1}{3}\mu_{24} + \tfrac{1}{3}\mu_{25} - \mu_{26} = 0\,,$$
$$H_0 : \tfrac{1}{4}\mu_{21} + \tfrac{1}{4}\mu_{24} + \tfrac{1}{4}\mu_{25} + \tfrac{1}{4}\mu_{26} - \mu_{27} = 0\,,$$
$$H_0 : \tfrac{1}{5}\mu_{21} + \tfrac{1}{5}\mu_{24} + \tfrac{1}{5}\mu_{25} + \tfrac{1}{5}\mu_{26} + \tfrac{1}{5}\mu_{27} - \mu_{28} = 0\,,$$
$$H_0 : \tfrac{1}{6}\mu_{21} + \tfrac{1}{6}\mu_{24} + \tfrac{1}{6}\mu_{25} + \tfrac{1}{6}\mu_{26} + \tfrac{1}{6}\mu_{27} + \tfrac{1}{6}\mu_{28} - \mu_{29} = 0\,.$$

The coefficients of linear combinations given in the null hypotheses were used to simultaneously test these hypotheses using the CONTRAST statement. The coefficients needed to test the last (sixth) partial hypothesis are shown in Table 4.

In the 'total' row in Table 4, there are coefficients for the HT factor, not used in the statement because the variable HT itself is not in the model (see Table 2). The coefficients for the EAS factor are zero, so the EAS factor itself did not enter the statement. The coefficients for interaction listed in the field of Table 4 were used. Similarly, one could determine the coefficients for the other five partial hypotheses. The coefficients determined in this way were used in the CONTRAST statement,

Table 4

Coefficients for the CONTRAST statement to test the null hypothesis $H_0 : \mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}, \mu_{27}, \mu_{28}) = \mu_{29}$ for the *EAS*×*HT* interaction

EAS	*HT*									Sum
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
1	0	0	0	0	0	0	0	0	0	**0**
2	$\frac{1}{6}$	0	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	-1	**0**
3	0	0	0	0	0	0	0	0	0	**0**
Sum	$\frac{1}{6}$	**0**	**0**	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	**-1**	**0**

Source: own processing.

while the relevant variable and its associated coefficients for the individual partial hypotheses of simultaneous testing were separated by a comma:

```
CONTRAST '21=24=25=26=27=28=29'
         EAS*HT 0 0 0 0 0 0 0 0 0 1 0 0 -1,
         EAS*HT 0 0 0 0 0 0 0 0 0 0.5 0 0 0.5 -1,
         EAS*HT 0 0 0 0 0 0 0 0 0 0.3333 0 0 0.3333 0.3333 -1,
         EAS*HT 0 0 0 0 0 0 0 0 0 0.25 0 0 0.25 0.25 0.25 -1,
         EAS*HT 0 0 0 0 0 0 0 0 0 0.2 0 0 0.2 0.2 0.2 0.2 -1,
         EAS*HT 0 0 0 0 0 0 0 0 0 0.16666 0 0 0.16666 0.16666
                0.16666 0.16666 0.16666 -1;
```

Running this statement within PROC GLM generates the first line in Table 5.

Table 5

Simultaneous equality tests for LS-Means of WI for unemployed persons from selected types of households

Contrast	DF	Contrast SS	Mean square	F value	Pr > F
21=24=25=26=27=28=29	6	0.2496	0.0416	1.00	0.4200
22 vs 23	1	0.0280	0.0280	0.68	0.4108
21-24-25-26-27-28-29 vs 22-23	1	1.0583	1.0583	25.56	<.0001

Source: EU-SILC 2020, own processing in SAS EG.

Based on the result of a simultaneous test of six null hypotheses ($DF = 6$) at a significance level of 0.05, the null hypothesis ($p = 0.4200$) was not rejected, which means that in the case of unemployed persons there was not enough evidence to assume a different WI mean in households of type 1A_0Ch, 2A_0Ch, 2A_1Ch, 2A_2Ch, 2A_3$^+$Ch, Other_0Ch, and Other_1$^+$Ch. At the same time, there was also no significant difference ($p = 0.4108$) between the other two types of households:

1A_1+Ch and 2A(1+R), which were tested both with the CONTRAST statement (2nd row in Table 5) and the analysis of marginal means (Table 3). Thus, the study created two clusters of household types for unemployed persons, while there were no significant differences between the types of households belonging to the common cluster, but there were significant differences between the clusters ($p < 0.0001$; 3rd row in Table 5).

To estimate the WI mean in these two household clusters, the author used the ESTIMATE statement, illustrated in the estimate for the cluster of household types 1A_0Ch, 2A_0Ch, 2A_1Ch, 2A_2Ch, 2A_3+Ch, Other_0Ch, and Other_1+Ch. The coefficients for these types of households are shown in Table 6. As with the CONTRAST statement, the coefficients in the total row that apply to the HT variable were not used because this variable does not appear separately in the model. However, this time there were non-zero coefficients for the EAS variable (column sum) and a non-zero coefficient for the intercept (total sum of coefficients in the bottom right corner), which were written in the ESTIMATE statement. Since the intercept can only be counted once, the DIVISOR option with constant 7 was used.

Table 6

Coefficients for the ESTIMATE statement to estimate $\mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}, \mu_{27}, \mu_{28}, \mu_{29})$ for the *EAS*×*HT* interaction

EAS	***HT***									**Sum**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
1	0	0	0	0	0	0	0	0	0	**0**
2	1	0	0	1	1	1	1	1	1	**7**
3	0	0	0	0	0	0	0	0	0	**0**
Sum	**1**	**0**	0	1	1	1	1	1	1	7

Source: own processing.

After running the statement

```
ESTIMATE '21-24-25-26-27-28-29' intercept 7 EAS 0 7 0
         EAS*HT 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 1 1 /divisor = 7;
```

within PROC GLM, one obtains the first row in Table 7.

Table 7

The estimate of $\mu(\mu_{21}, \mu_{24}, \mu_{25}, \mu_{26}, \mu_{27}, \mu_{28}, \mu_{29})$ and $\mu(\mu_{22}, \mu_{23})$ for the *EAS*×*HT* interaction

Parameter	**Estimate**	**Standard error**	**t value**	**Pr > |t|**
21-24-25-26-27-28-29	0.4420	0.0247	17.87	<.0001
22-23	0.1003	0.0687	1.46	0.1443

Source: EU-SILC 2020, own processing in SAS EG.

In the case of an unemployed, there were no significant differences in terms of the WI mean person between household types 1A_0Ch, 2A_0Ch, 2A_1Ch, 2A_2Ch, 2A_3^{+}Ch, Other_0Ch, and Other_1^{+}Ch. In the cluster of these types of households, the author estimated the WI mean at 44.20%, and with a probability of 0.95 it was estimated in the interval (0.4420–1.9603×0.0247; 0.4420+1.9603×0.0247), i.e. (0.3936; 0.4904), using the quantile of the Student's distribution $t_{0.975}(7346) = 1.9603$. Similarly, the study estimated the WI mean for a cluster of household types 1A_1^{+}Ch and 2A(1^{+}R) (10.03%; 2nd row in Table 7). For unemployed persons, the WI mean across these two types of households was not significantly different from 0 ($p = 0.1443$) and with a risk of 0.075 did not exceed 20%, which is the limit determining a very low work intensity. In other words, a person who has the status of unemployed and lives in a household of type 1A_1^{+}Ch or 2A(1^{+}R), had in 2019 up to a 92.5% confidence level of showing a very low work intensity over the entire reference period.

If a person is employed, the riskiest types of households, in terms of exclusion from the labour market, are 2A_3^{+}Ch and Other_1^{+}Ch, between which there is a significant difference ($p = 0.0037$; to the detriment of 2A_3^{+}Ch). However, both types of households have a visibly lower mean of the WI than other types of households (Table 2; $p < 0.001$). This is followed by a cluster of household types 2A_1Ch, Other_0Ch, and 2A_2Ch, where there is no significant difference between the pairs of these types ($p = 0.3757$, $p = 0.4833$ and $p = 0.1033$) and based on the CONTRAST statement, we found out that there is also no significant difference between all 3 types of households ($p = 0.2577$; Table 8).

Table 8

Simultaneous test of equality of marginal WI means for employed persons from selected types of households

Contrast	DF	Contrast SS	Mean square	F value	Pr > F
35=37=39	2	0.1123	0.0561	1.36	0.2577

Source: EU-SILC 2020, own processing in SAS EG.

Furthermore, the combination of household types 1A_1^{+}Ch and 2A_0Ch ($p = 0.5274$) and household types 1A_0Ch and 2A(1^{+}R) ($p = 0.3803$) proved to be reasonable. For employed persons from the above five clusters of household types, the estimates of LS-means of the WI are presented in Table 9.

For all the five clusters of household types for employed persons, the WI mean was significantly different from 0 ($p < 0.0001$). In addition, there was a statistically significantly different WI mean between all pairs of these five clusters. The smallest

difference was between persons in the household cluster 1A_1^{+}Ch, 2A_0Ch and persons in household cluster 1A_0Ch, 2A(1^{+}R). The author estimated this difference in the WI mean at 3.4%, but this was also significantly different from 0 at the significance level of 0.05, $p = 0.0196$.

Table 9

The estimate of μ_{36}, μ_{38}, $\mu(\mu_{35}, \mu_{37}, \mu_{39})$, $\mu(\mu_{32}, \mu_{34})$ and $\mu(\mu_{31}, \mu_{33})$ for the *EAS*×*HT* interaction

Parameter	Estimate	Standard error	T value	Pr>|t|
2A_3^{+}Ch	0.7389	0.0191	38.75	<.0001
Other_1^{+}Ch	0.7913	0.0109	72.73	<.0001
2A_1Ch, Other_0Ch, 2A_2Ch	0.8281	0.0099	83.72	<.0001
1A_1^{+}Ch, 2A_0Ch	0.8669	0.0134	64.70	<.0001
1A_0Ch, 2A(1^{+}R)	0.9010	0.0135	66.80	<.0001

Source: EU-SILC 2020, own processing in SAS EG.

Following the above procedure, the study created five household clusters for other inactive persons (IP), two household clusters for unemployed persons (UP), five household clusters for employed persons (EP); these clusters are listed in Table 10.

Table 10

Household clusters for individual statuses of economic activity

EAS	Cluster	Household types
Inactive person	**IP 1**	1A_0Ch, 2A(1^{+}R)
	IP 2	2A_0Ch, Other_0Ch
	IP 3	Other_1^{+}Ch
	IP 4	2A_1Ch, 2A_3^{+}Ch
	IP 5	1A_1^{+}Ch, 2A_2Ch
Unemployed	**UP 1**	1A_1^{+}Ch, 2A_1Ch
	UP 2	1A_0Ch, 2A(1^{+}R), 2A_0Ch, 2A_2Ch, 2A_3^{+}Ch, Other_0Ch, Other_1^{+}Ch
Employed	**EP 1**	2A_3^{+}Ch
	EP 2	Other_1^{+}Ch
	EP 3	2A_1Ch, Other_0Ch, 2A_2Ch
	EP 4	1A_1^{+}Ch, 2A_0Ch
	EP 5	1A_0Ch, 2A(1^{+}R)

Source: own processing.

Point and interval (95%) estimates of the WI mean for persons from individual household clusters are shown in Figure 1.

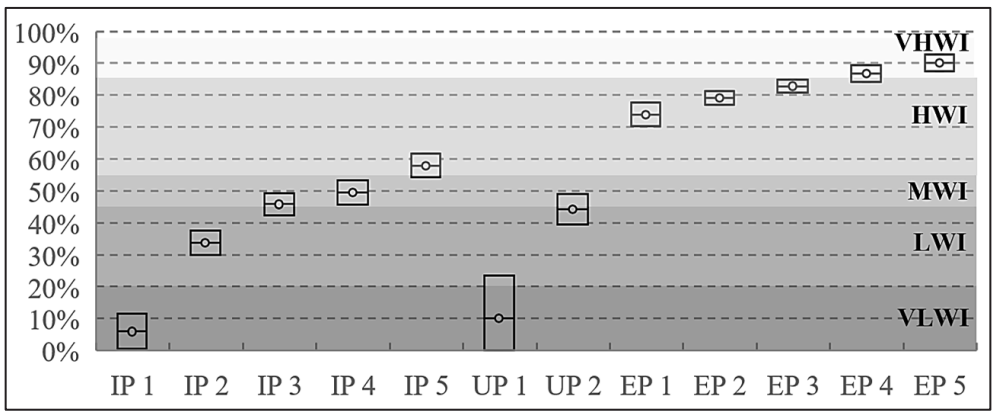


Fig. 1. Interval estimates (95%) of LS-means of WI for *EAS*×*HT* interaction

Source: EU-SILC 2020, own processing in SAS EG.

Households without dependent children (cluster IP 1 and IP 2) are the riskiest for other inactive persons in terms of exclusion from the labour market and within them, households with at most one person of working age (IP 1) are especially at risk. Inactive persons from households in cluster IP 1 (1A_0Ch, 2A(1^{+}R)) have the WI mean much below the upper limit for identifying a very low work intensity (below 20%). On the contrary, for inactive persons, the highest mean of WI was for persons from households of cluster IP 5 (1A_1^{+}Ch, 2A_2Ch), with the identified high work intensity. Other inactive persons from households belonging to other clusters (IP 2 to IP 4) showed the WI mean at the low or medium level.

The HT factor does not show consistent results for individual statuses of economic activity. While, e.g. inactive persons from households of type 1A_1^{+}Ch were not at risk in terms of exclusion from the labour market, the unemployed persons from this type of household were at high risk. Unemployed persons from households 1A_1^{+}Ch were put together with unemployed persons from households 2A_1Ch included in the common cluster UP 1, in which the WI mean was at the level of very low work intensity (with a low probability of reaching the level of low work intensity). The cluster of unemployed persons from other types of households (UP 2) was at the level of low to medium labour intensity. Unlike other inactive persons, unemployed persons in no type of household showed WI mean at a level higher than the medium, while employed persons had WI mean at a high to a very high level in all types of households. With 95% confidence, employed persons from cluster EP 5 (1A_0Ch, 2A(1^{+}R)), i.e. from households with no dependent children

and at most one adult of working age, showed a WI mean at the very high level. It should be noted that persons from these types of households in the case of another inactive person, were in the riskiest position and their WI mean identified their (quasi)joblessness.

5.2.2. Analysis of the impact of interaction *EAS*×*Age*

As with the influence of the type of households, in the case of the influence of age on the WI, it was confirmed that employed persons have a significantly higher WI mean than unemployed persons or other inactive persons ($p < 0.0001$). There was a significant difference between unemployed persons and other inactive persons at the significance level of 0.05 only in the age group up to 30 years ($p = 0.0203$), to the detriment of unemployed persons. For other inactive persons, the lowest WI mean was in the age groups 30-40 and 40-50, between which there was no significant difference ($p = 0.9115$). Across these two categories (persons aged 30-50), the WI mean with 95% confidence level was in the range of 26.8-33.2% (IP 30-50 in Figure 2). Insignificant differences were also confirmed for unemployed persons in the age groups 30-40, 40-50, and 50-60 ($DF = 2$; $p = 0.6795$). Unemployed persons aged 30-60 years showed the WI mean with 95% confidence level from the interval 30.4-40.3% (UP 30-60 in Figure 2). There were significant differences between the other age categories for the individual statuses of economic activity in terms of the WI, and the point and interval estimates of the WI means are shown in Figure 2.

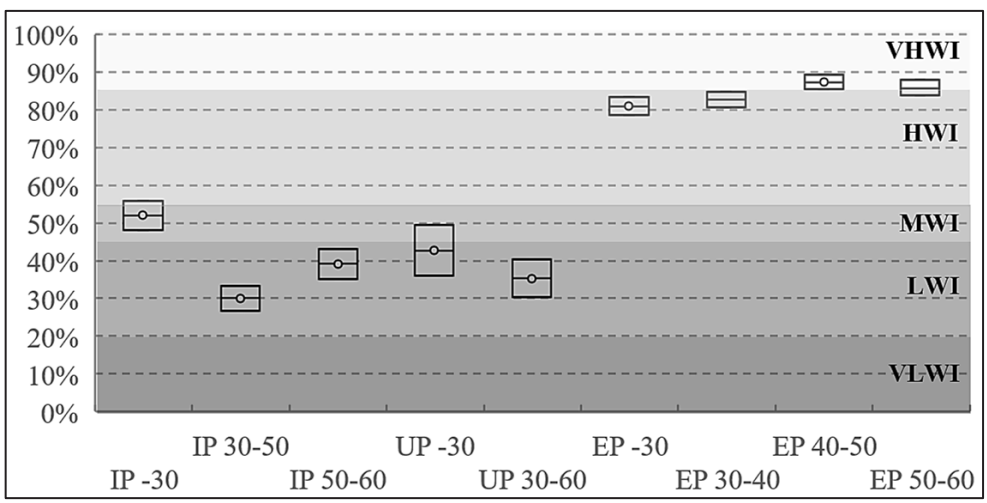


Fig. 2. Interval estimates (95%) of LS-Means of WI for *EAS*×*Age* interaction

Source: EU-SILC 2020, own processing in SAS EG.

Other inactive persons presented a demonstrably highest mean of the WI in the age group under 30 years, and this mean was at the medium level, while for older persons it was at the low level. The author observed a similar phenomenon for the unemployed, however for persons under the age of 30 one cannot convincingly assume the WI mean at the medium level, but at the low to medium level (Figure 2). While the age group of under 30 was the least at risk for other inactive people and unemployed people, this category (EP-30) showed the highest risk for employed people. Employed persons under the age of 30 may have the WI mean below 80% with a risk of 0.05, which was not the case for older employed persons. Despite this finding, employed persons under the age of 30 and also those aged 30-40, had the WI mean at the same level, specifically the high level. Employed persons aged 50-60 showed the WI mean at the level of high to very high, and employed persons aged 40-50 even a very high level (with a reliability of 0.95 from the interval 85.3-89.3%).

5.2.3. Analysis of the impact of interaction *EAS*×*Education*

Not surprisingly, higher WI is associated with higher education, and this applies to all statuses of economic activity (Figure 3). The differences in the WI mean between the ISCED 3-5 and ISCED 6-8 educational groups were not as large and convincing ($p=0.0084$ for IP, $p=0.0329$ for UP, $p=0.0386$ for EP as the differences between these two educational groups from ISCED 0-2 ($p < 0.0001$). For other inactive persons (IP), there were maller differences between the educational groups in the

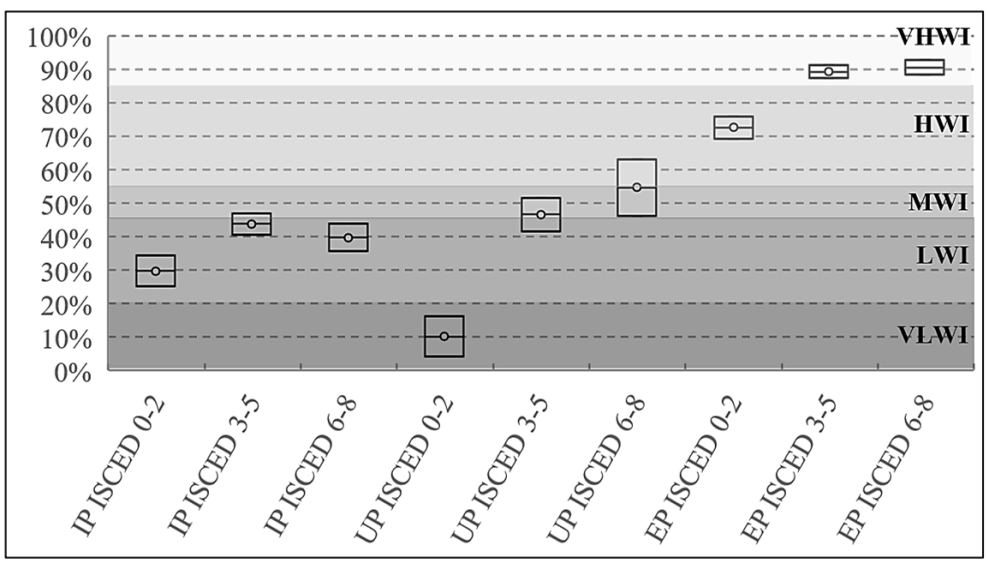


Fig. 3. Interval estimates (95%) of LS-Means of WI for *EAS*×*Education* interaction

Source: EU-SILC 2020, own processing in SAS EG.

WI mean than in the other two statuses. Although other inactive persons with education at ISCED 0-2 level showed a significantly lower WI mean than persons with higher education, all the educational groups for this status of economic activity had the WI mean at the low level. The largest exclusion from the labour market naturally concerns the unemployed with low education (ISCED 0-2), for whom the WI mean was estimated at a very low level.

As unemployed persons with higher education showed the WI mean at a low to medium (ISCED 3-5) or medium to a high level (ISCED 6-8), they seem to remain unemployed for a shorter time and/or have another adult in a household (spouse or partner) who is employed. However, the status of economic activity – Other inactive or Unemployed – had such a significant effect on the WI that education cannot sufficiently compensate for this and therefore persons who are unemployed or other inactive and whose education is tertiary generally live in households with significantly lower WI ($p < 0.0001$) than employed persons. This applies even to employed persons with ISCED 0-2 education, for whom the average WI was at a high level. Employed persons with higher education have the WI mean even at a very high level.

Conclusion

The type of household in which a person lives, education, age, health conditions, marital status, as well as the region in which a person lives, are factors that significantly affect the WI, but this impact varies for different statuses of economic activity. These factors are considered very important by other researchers (Horemans, 2018; Verbunt and Guio, 2019) when assessing employment/unemployment and atypical employment in relation to social exclusion.

The general linear model, in which the study considered the interactions of these relevant factors with the status of economic activity of the assessed persons, explained the variability of the WI to more than 50%. The status of economic activity, type of household, age, and education participated the most in this explained variability (the participation of all the four factors was 48% in total). In addition to economic activity, van der Zwan and de Beer (2021) also used the above factors (age, education, and type of household) as control variables, applying them to the assessment of employment for persons with disability and those without disability. Horemans (2016) discussed the importance of the impact of age and education on non-standard employment, while also describing the need to transform these factors at the individual level to the household level, through the degree of educational and age homogamy.

The paper focuses on the analysis of the influence of the above four factors (status of economic activity, type of household, age, and education) on the WI, to which the

LS-means and contrast analysis within the estimated GLM was used. Through the equality tests of LS-means of WI and simultaneous testing of WI means, the author identified for each economic activity status between which types of households, age categories, and educational categories, there was no significant difference from the WI perspective, and between which there were demonstrable differences. For household types, age groups, and educational groups for which the author did not have sufficient evidence to be able to demonstrably confirm differences in the mean of WI, such clusters were created between which there were statistically significant differences. Individually for employed persons, unemployed persons, and other inactive persons, the mean of WI was estimated for individual types of households, age groups, and educational categories (or their groups). Based on interval estimates of the WI mean, these groups of persons were assigned a level of WI in accordance with the Eurostat methodology.

For other inactive people, the study mainly identified very low and low levels of WI. The WI mean for other inactive persons can be assumed to be above 50% only exceptionally, and this applies to persons under the age of 30, persons from complete households with dependent children, and persons from households with 1 adult and at least 1 child. The study identified very low work intensity ((quasi)joblessness) most often for unemployed persons. For the unemployed, the author estimated the work intensity to be below 50% for all types of households and all age groups. Unemployed people are highly likely to live in households that use more than 50% of their employment potential only if they have a tertiary education. For employed persons, broken down separately by type of household, age, and education, the study revealed the WI mean at least at a high level.

In terms of household type, other inactive persons from households without dependent children, in which there is at most one person of working age, and unemployed persons from households of 1 adult with at least 1 dependent child and households of 2 adults with 1 child, have the greatest risk of exclusion from the labour market. For these groups of persons, the study quantified the WI mean at a very low level. Persons from households without dependent children, in which there is at most one person of working age, were most at risk if they had the status of economic activity other than inactive, but in the case of the status of employed they reached the highest WI mean, which was at a very high level.

The conducted analyses revealed that of all age groups, the under-30 age group had the highest WI mean when considering unemployed persons (UP) or other inactive persons (IP), for whom the mean WI was at a low to medium level (for UP) and at a medium level (for IP). Older unemployed and other inactive persons generally showed low work intensity. Employed persons had the WI mean significantly higher than the other two statuses of economic activity, and this was also confirmed in the case of a breakdown of persons according to their age.

Depending on their age, employed persons had a high to a very high WI mean (up to 40 years – high, 40-50 years – very high, and 50-60 years – high to very high).

With the increase in education, the WI was also growing. Unemployed persons with low education (ISCED 0-2) generally showed (quasi)joblessness, while unemployed people with tertiary education were 50% more likely to have a medium WI, and 50% to even have a high WI. In this way, the study revealed that education plays a crucial role for unemployed persons in terms of work intensity. These results are consistent with the assumption that persons with higher education are less likely to remain unemployed for a longer period (Núñez and Livanos, 2010) and that they usually have a person with higher education (educational homogamy) at their side, whose threat of exclusion from the labour market is lower compared to a less educated person. Education for other inactive persons does not cause such large differences in WI as for the unemployed. Other inactive persons in all educational groups fall into the low level according to the WI mean. Employed persons have the WI mean in terms of education at a high level (for ISCED education 0-2) or a very high level (for ISCED education 3-8).

The author is convinced that the paper fills a gap in research that mostly focuses only on labour market exclusion identified on the basis of very low work intensity. The comprehensive work intensity analysis provided in this paper is important because also other degrees of work intensity may be associated with poverty and social exclusion (Kis and Gábos, 2016, Kalinowski, 2018). Social policies should then be targeted at persons with reduced work intensity, as confirmed by Blatná (2018). Based on the analysis of the share of people living in households with very low work intensity in the Czech Republic in the period 2005-2016, that study found that the growth in social benefits and increase in the proportion of people in lifelong education leads to a reduction of the proportion of people living in (quasi)jobless households.

In conclusion, it should be emphasised that the paper provides an empirical analysis for Slovakia, and although many conclusions apply at least to the CEE countries, this needs to be validated by further research. In particular, the influence of household type on the WI can be significantly different in other countries. According to Atkinson et al. (2017), there is a great deal of cross-country variation in the composition of the (quasi)jobless population by household type. The results of the analysis have their limitations, mainly related to the methodology of measuring work intensity and its levels (specifically very low work intensity). Ward and Özdemir (2013) noted a problem in the definition of work potential (the denominator of the WI indicator), which does not include persons older than 59, as well as in the threshold for identifying (quasi)jobless households that is set at 20%, whereas they advocate its raising to 30%.

References

Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. New York: John Wiley & Sons.

Atkinson, T., Guio, A. C., and Marlier, E. (eds.) (2017). Monitoring social inclusion in Europe. *Statistical Books Eurostat*. Luxembourg: Publications Office of the European Union.

Blatná, D. (2018). Analysis of the relationship between the share of people living in households with very low work intensity and selected socio-economic indicators in the Czech Republic in the period 2005-2016. 12th International Days of Statistics and Economics, Prague.

Cai. W. (2014). Making Comparisons Fair: How LS-Means Unify the Analysis of Linear Models. *SAS Institute Inc. Paper SA*, S060–2014.

Calegari, E., Fabrizi, E., and Mussida, C. (2021). Disability and work intensity in Italian households. *Review of Economics of the Household*, 1–20.

Cantó, O., Cebrián, I., and Moreno, G. (2022). Youth living arrangements and household employment deprivation: Evidence from Spain. *Journal of Family Research. Early View*, 1–33.

Colombarolli, C. (2021). In-work poverty and regional disparities. An analysis of the relationship between work intensity and the probability of being and feeling poor across Italian territories. *Sociologia del lavoro*, 161, 76–93.

Ćwiek, M. T., and Ulman, P. (2019). Income and poverty in households in selected European countries. *Acta Universitatis Lodziensis. Folia Oeconomica*, 345(6), 9–34.

Darlington, R. B., and Hayes, A. F. (2016). Regression Analysis and Linear Models: Concepts, Applications, and Implementation. Guilford Publications.

de Graaf-Zijl, M., and Nolan, B. (2011). Household joblessness and its impact on poverty and deprivation in Europe. *Journal of European Social Policy*, 21(5), 413–431.

Dean, A., Voss, D., and Draguljić, D. (2017). Design and Analysis of Experiments. Springer.

Duiella, M., and Turrini, A. (2014). Poverty developments in the EU after the crisis: A look at main drivers. *ECFIN Economic Brief*, 31, 1–10.

Elswick Jr, R. K., Gennings, C., Chinchilli, V. M., and Dawson, K. S. (1991). A simple approach for finding estimable functions in linear models. *The American Statistician*, 45(1), 51–53.

Eurostat, Glossary: At risk of poverty or social exclusion (AROPE), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At_risk_of_poverty_or_social_exclusion_(AROPE), accessed on 6.2.2022, 2022.

Fabrizi, E., and Mussida, C. (2020). Assessing poverty persistence in households with children. *The Journal of Economic Inequality*, 18(4), 551–569.

Filandri, M., Pasqua, S., and Struffolino, E. (2020). Being working poor or feeling working poor? The role of work intensity and job stability for subjective poverty. *Social Indicators Research*, 147(3), 781-803.

García-Gómez, C., Pérez, A., and Prieto-Alaiz, M. (2021). Copula-based analysis of multivariate dependence patterns between dimensions of poverty in Europe. *Review of Income and Wealth*, 67(1), 165–195.

Gerbery, D., and Miklošovic, T. (2020). Labour market transitions and their determinants in Slovakia: Path from crisis to recovery 1. *Ekonomický časopis*, 68(7), 651–676.

Goodnight, J. H., and Harvey, W. R. (1997). SAS Technical Report R-103. Least Squares Means in The Fixed Effects General Model. Cary, NC: SAS Institute Inc.

Guio, A.-C., Vandenbroucke, F. (2019). Poverty and Child Deprivation in Belgium. A Comparison of Risk Factors in the Three Regions and Neighbouring Countries, A Comparison of Risk Factors in the Three Regions and Neighbouring Countries (March 1, 2019). King Baudouin Foundation.

Hick, R., and Lanau, A. In-work poverty in the UK: Problem, policy analysis and platform for action, Final Report, Cardiff University, https://orca.cf.ac.uk/103013/1/Hick%20and%20Lanau%20_%20InWork%20Poverty%20in%20the%20UK.pdf, accessed on 5.3.2022, 2017.

Horemans, J. (2016). Polarisation of non-standard employment in Europe: Exploring a missing piece of the inequality puzzle. *Social Indicators Research*, 125(1), 171–189.

Horemans, J. (2018). Atypical employment and in-work poverty. In *Handbook on In-Work Poverty*. Edward Elgar Publishing.

Kalinowski, S. (2018). The Working Poor in the European Union. Conference: International Scientific Days 2018, „Towards Productive, Sustainable and Resilient Global Agriculture and Food Systems".

Kim, K., and Timm, N. (2006). Univariate and Multivariate General Linear Models: Theory and Applications with SAS. Chapman and Hall/CRC.

Kis, B., A., and Gábos, A. (2016). Consistent poverty across the EU. *Corvinus Journal of Sociology and Social Policy*, 7(2), 3–27.

Kuznetsova. A., Brockhoff. P. B., and Christensen. R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).

Lenth, R., V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33.

Littell, R. C., Stroup, W. W., and Freund, R. J. (2010). SAS for Linear Models. 4th ed. Cary, NC: SAS Institute Inc.

McFarquhar, M. (2016). Testable hypotheses for unbalanced neuroimaging data. *Frontiers in Neuroscience*, 10, 270.

Mysíková, M., Želinský, T., Garner, T. I., and Večerník, J. (2019). Subjective perceptions of poverty and objective economic conditions: Czechia and Slovakia a quarter century after the dissolution of Czechoslovakia. *Social Indicators Research*, 145(2), 523–550.

Nieuwenhuis, R., and Maldonado, L. (2018). The triple bind of single-parent families: Resources, employment and policies to improve well-being. *Policy Press*.

Núñez, I., and Livanos, I. (2010). Higher education and unemployment in Europe: An analysis of the academic subject and national effects. *Higher Education*, 59(4), 475–487.

Peña-Casas, R., Ghailani, D., Spasova, S., and Vanhercke, B. (2019). In-work poverty in Europe. A study of national policies. European Social Policy Network. Brussels: European Commission.

Poline, J. B., Kherif, F., Pallier, C., and Penny, W. (2007). Contrasts and classical inference. *Statistical Parametric Mapping: the Analysis of Functional Brain Images*, 1, 126–139.

SAS Institute Inc., SAS/STAT® 15.1 User's Guide. The GENMOD Procedure, Cary SAS/STAT® 15.1, NC: SAS Institute Inc., 2018a.

SAS Institute Inc., SAS/STAT® 15.1 User's Guide. The GLM Procedure. Cary SAS/STAT® 15.1, NC: SAS Institute Inc., 2018b.

Sánchez-Sellero, M. C., and Garcia-Carro, B. (2020). Which groups have a greater risk of poverty in Spain? *Revija za socijalnu politiku*, 27(1), 36–36.

Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038.

Searle, S. R., and Gruber, M. H. J. (2017). Linear Models, 2nd ed. John Wiley & Sons.

Searle, S. R., Speed, F. M., and Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician*, 34(4), 216–221.

Suzuki, M., Taniguchi, T., Furihata, R., Yoshita, K., Arai, Y., Yoshiike, N., and Uchiyama, M. (2019). Seasonal changes in sleep duration and sleep problems: A prospective study in Japanese community residents. *PLoS One*, 14(4), e0215345.

Šoltés, E., and Ulman, P. (2015). Material deprivation in Poland and Slovakia – A comparative analysis. *Cracow Review of Economics and Management*, 11(947), 19–36.

Tabachnick, B. G., and Fidell, L. S. (2013). Using Multivariate Statistics, 6th ed. Boston, MA: Pearson.

van der Zwan, R., and de Beer, P. (2021). The disability employment gap in European countries: What is the role of labour market policy? *Journal of European Social Policy*, 31(4), 473–486.

Verbunt, P., and Guio, A.-C. (2019). Explaining differences within and between countries in the risk of income poverty and severe material deprivation: Comparing single and multilevel analyses. *Social Indicators Research*, 144(2), 827–868.

Wang, B., Wu, P., Kwan, B., Tu, M. X., and Feng, Ch. (2018). Simpson's paradox: Examples. *Shanghai Archives of Psychiatry*, 30(2), 139.

Ward, T., and Ozdemir, E. (2013). Measuring low work intensity – An analysis of the indicator. *ImPRovE Discussion Paper*, 13/09, Antwerp.

Westfall, P. H., and Tobias, R. D. (2007). Multiple testing of general contrasts: Truncated closure and the extended Shaffer-Royen method. *Journal of the American Statistical Association*, 102(478), 487–494.

Received: June 2022

Acknowledgments: *This work was supported by the VEGA projects The impact of the COVID-19 crisis on business demography and employment in the Slovak Republic and the EU (No. 1/0561/21) and Poverty and Social Exclusion in Slovakia and the EU in Times of the COVID-19 Pandemic and Energy Crisis (No. 1/0038/23).*

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Zunia Saif Tirmazee**

Unequal pay for equal education! A case of gender wage gap from Punjab, Pakistan[1]

This study aimed to quantify the returns to tertiary educational attainment and measures the extent to which these returns differ for men and women. The article provides new empirical evidence of the returns to tertiary education literature by introducing a unique instrument, namely the supply of education to deal with endogeneity. The analysis is implemented using a pooled cross-section of five rounds of the Pakistan Social and Living Standards survey with 10,000 observations. The results show that the marginal returns to acquiring one extra year of education beyond matriculation are higher for women than for men. This result could partially explain the reversal of the gender gap in enrolments from secondary and lower to the post-secondary level of education in Punjab. The first stage results highlight the significance of investing in physical infrastructure for the greater accumulation of human capital.

Keywords: gender wage gap, returns, supply of education, tertiary education.

JEL Classification: I26, J16, J24, J31, J71
DOI: 10.15611/aoe.2023.1.04

Quote as: Tirmazee, Z. S. (2023). Unequal pay for equal education! A case of gender wage gap from Punjab, Pakistan. *Argumenta Oeconomica*, 1(50), 67-94.

1. Introduction

The gender gaps in primary and lower secondary enrolment in Punjab are smaller, but still, boys outnumber girls at both levels. At higher secondary (intermediate) level the gender gap shrinks, and at BA/BS/postgraduate level female students outnumber males (PSLM, 2014). This paper looks at the advantages in the labour

* Lahore School of Economics, Pakistan. ORCID: 0000-0002-6158-8130.

[1] The author would like to thank Kate Vyborny, Naved Hamid, Mahreen Mahmud, Farah Said, Waqar Wadho, Azam Chaudhry for their helpful suggestions and comments, and also the audience at the LUMS/LSE/DevLab-Duke University Joint Workshop 2020 and the Applied Development Conference 2018 for their valuable comments.

market to the increased higher educational attainment and the extent to which these returns differ for males and females, to see whether these differences can to some extent explain the reversal in the gender gaps in education/enrolment. The returns to any level of education broadly fall into three categories (i) private financial, (ii) private non-financial returns such as availability of both better jobs and better working conditions, and (iii) social ones. In this paper, however, the author focused only on the private financial returns by examining their effect of higher education attainment on the wages and the gender gap.

To estimate the causal link between the returns to tertiary education in Punjab and the gender gap in these returns the study used the instrumental variable (IV) technique. Making use of the exogenous variation in the supply of higher education institutes the author employed the total number of available tertiary educational institutes at the district level in Punjab as an instrument. The steady increase in tertiary education institutions reflects improved access to tertiary education for male and female students for two reasons. Firstly, the expansion of tertiary education facilities exogenously decreases the costs associated with attaining more education, by making access to college education cheaper for individuals when a new college is constructed in the local area. Secondly, having a college in one's area also may reduce the mobility concerns which are an important hurdle, especially for females in attaining education (Cheema et al., 2019). In addition to IV, the study also used the region-and-time fixed effects in the first stage to control for region-specific and time-varying unobserved factors that may cause omitted variable bias if not accounted for.

The first stage of this analysis was a regression of educational attainment on the number of tertiary education institutes available in a district. In estimating the first stage, the two identifying assumptions were: a) the relationship between changes in college availability and changes in educational attainment is not reflective of changes in development in general and b) the exact timing of college opening in a given district is not driven by demand for education.

To show that the changes in college availability and in educational attainment are related regardless of the level of development, the author showed that first stage results are robust to controlling for the development of a region using various community-level indicators of development. Additionally, the responsiveness of years of education to variation in the supply of tertiary educational institutes was evident only for the relevant age cohorts, i.e. from 16 to 32 years of age. The results of the first stage were null for sample observations that lie just above and just below the relevant age cohort. This again showed that the first stage results are not reflective of development in general, because if they were there should be a significant positive relationship between the two regardless of the age brackets due to the confounding effect of regional development.

The main findings of the analysis are that there is a positive significant relationship between the estimated years of education from the first stage and the income levels.

Another important result is that males on average earn significantly more than females regardless of the level of education, however an extra year of education brings higher returns to women compared to men. This implies that the gender earnings gap tends to fall with rising education distribution. Whatever the reason for the differential in the earnings of two genders, discrimination or difference in their respective productivities, the results show that as the years of education attained increase, the earnings differential narrows.

The first stage results show that the greater availability of colleges at district level is significantly associated with higher educational attainment at individual level. Moreover, the impact of the increased supply of tertiary education institutions on tertiary education attainment is higher in low Human Development Index (HDI)[2] districts than in the high HDI districts of Punjab. This is an important result from a policy perspective as it shows that investing in the physical infrastructure in less developed regions yields the greatest returns.

This work is a contribution to the study of the labour market in Pakistan, as tertiary education is still an understudied area there. To that end, the analysis makes a significant contribution to the literature on tertiary education by introducing a unique instrument, the district level supply of education, i.e. the number of Arts and Science Intermediate, Degree, and Post-Graduate Colleges for male and female students in a given district at a given point in time in Punjab. A related contribution of this study is that a pooled cross-section with a very large number of observations has not been used to study the dynamics of returns, particularly to tertiary education for Pakistan. This was achieved by making use of five rounds (2006, 2008, 2010, 2012, and 2014) of the household level survey, Pakistan Social and Living Standards Measurement (PSLM), covering a decade. This provided a very large number of observations, approximately 10,000 in this case, and sufficient data points to study this research question.

This analysis derives its significance from the important lessons it bears for policy. For instance, the returns to tertiary education as projected over the life cycle reflect the expectations that influence current student decisions to participate in higher education. If the returns increase with years of education, then there is a positive signal from the labour market, which should effectively lead to greater investment in human capital accumulation. The literature shows that households do respond to information regarding returns to education (Jensen, 2012; Attanasio and Kaufmann, 2009). The results regarding the gender gap in returns to tertiary education can be taken as confirmation of this hypothesis as higher marginal returns could be a reason why female enrolment has been increasing in tertiary education over the past decade (Table 1).

---

[2] The "Human Development Index measures development by quantifying three dimensions of human life – education, health, and standard of living" (Pakistan Human Development Index report 2017) http://hdr.undp.org/sites/default/files/reports/2847/pk-nhdr.pdf

The first stage of this analysis also has at least two very important policy implications. Firstly, the results show the importance of investing in physical capital for the accumulation of human capital. A concerted effort to plan the expansion of the supply of education, especially in areas where there is a dearth of tertiary educational institutions, may allow for the possibility to accumulate greater years of education for individuals who are on the margins. Secondly, the increased availability of tertiary education institutions could also have substantial positive spillover effects when females with tertiary education enter the labour force as school teachers facilitating the supply of more low-cost private schools (Andrabi et al., 2008).

The rest of this paper is organised as follows. Section 2 provides a review of the literature. Section 3 gives a detailed account of the methodology. Section 4 presents the data and descriptive statistics. Section 5 presents the results of the empirical analysis, Section 6 discusses the robustness check, and finally, Section 7 concludes.

2. Literature review

This review of literature sheds light on three issues. Firstly, the issue of gender inequality in the labour market outcomes, secondly, the household's decision to invest in education, and thirdly, the approaches adopted in various studies to establish a causal link between educational attainment and the financial returns to education.

The labour force participation rate of female graduates in Punjab between the age of 25-35 is only 32% compared to that of males at 96% (Labour Force Survey, 2018). The wages of women with higher education are about 68% of the wages of equally qualified men (PSLM, 2014). In this study, the author probed the likely causes of the gender wage gap by breaking it down into explained and unexplained gaps. The main finding was that almost one-third of this gap can be explained by the difference in the human capital of men and women and their nature of work. The remaining two-thirds of this gap remains unexplained and can be attributed to either discrimination or omitted variables (Tirmazee, 2021). Numerous explanations, such as occupational segregation (Levanon et al., 2009; Blau and Kahn, 1992), work interruptions (Epstein, 1988; Neumark and Korenman, 1992), education and training (Blau and Kahn, 2017; Becker, 2010; Mincer, 1962, 1974), temporal flexibility (Goldin, 2014) or unionization (ILO, 2018), have been advanced in the literature for the gender pay gap. In the context of Pakistan, it has been argued that the gender pay gap is because most women either work as unpaid family workers or if in paid employment they are often employed in low-skilled, low-paid jobs (Khan, 2017).

Human capital accumulation, an important determinant of labour market outcomes, reflects the preferences of the demand side and the capacity of the supply side to meet the demand for education. On the demand side, the parents' decision to enroll their children in another year of schooling involves comparing the cost and

benefits of this additional investment, where the benefits may include better wages, healthcare, higher standard of living, etc., and the costs may include tuition fees, the opportunity cost of the child's time, transportation, etc. The simple economic model of comparing the cost of schooling with the benefits implies that if the benefits of schooling such as monetary and non-monetary returns of acquiring education rise, the optimal investment in schooling may increase.

For instance, according to Banerjee et al. (2013), underinvestment in education can be due to the inadequate availability of resources to fund this, or due to the inadequate information that leads households to underestimate the returns to education. If the problem is inadequate resources to fund investment, then there is a need for governments to subsidise education but if the problem is that of inadequate information then there is a need to inform households of the true benefits of education.

The literature does suggest that households respond to information regarding returns to education. For instance, Jensen (2012) provided evidence of how increasing awareness regarding potential job opportunities owing to the rapid expansion of the Business Process Outsourcing (BPOs) industry in India led to a significant rise in investment in the education of younger girls by households. Similarly, Attanasio and Kaufmann (2009) provided evidence on the significance of individual perceptions regarding future returns to schooling using data for Mexico. In their analysis to model college or school choice, they found that mothers' expectations and individual's own expectations matter for college enrolment. In other settings where different instruments were used to provide information such as the author's own calculated returns to education (Jensen, 2010) or a short video showcasing the ways of acquiring financial resources to fund education (Dinkelman and Martinez, 2014), it was seen that households do update their beliefs and react accordingly. Given how households react to information on returns to education,the author believes that this analysis is crucial as it directly yields information on these returns and also reflects the expectations that influence parents' and students' decisions regarding investing time, money, and effort in education.

Since the main objective in this paper was to estimate the causal link between labour market returns and human capital accumulation, the review of various approaches for estimating this causal link follows. The general approach in the literature for tackling the question of private returns to schooling has been an estimation of the Mincerian wage function (Mincer, 1974), which is a simple regression linking schooling with the wages earned. Below there is a simple Mincerian wage equation:

$$Ln\left(Earnings\right)_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + e_i \tag{1}$$

where $Ln(Earnings)_i$ is the log of yearly earnings of person i and X_i is a vector of individual i's characteristics, and S_i is the accumulated years of education. Ordinary least squares (OLS) yields biased estimates of the parameters of the above equation due to unobserved heterogeneity. In addition, the classic Mincerian wage equation

does not allow to test for the heterogeneity of effects, i.e. how the observed relationship between years of education and the returns to education differs across various subsets of the population. Some possible solutions suggested in the literature to deal with the shortcomings of the Mincerian wage equation are as follows.

Firstly, studies have directly tried to account for the 'ability bias' by including appropriate measures that in some way are a proxy for unobserved ability, such as IQ level or various test scores. However, there are always concerns regarding the extent to which these proxies accurately measure ability, as a multitude of measures for ability has resulted in the past in inconsistent signs for these variables (Dickens and Lang, 1993). One popular method to control for innate ability has been the use of siblings (twins) (Ashenfelter and Zimmerman, 1997; Bingley et al., 2009; Bonjour et al., 2002; Isacsson, 1999; Miller et al., 1995), under the assumption that using twins or siblings allows to differentiate the innate ability since a lot of what determines an individual's ability is common across members of the same household, especially twins. In this way, by eliminating unobserved individual ability by differencing first, one can obtain an unbiased estimator of the return to education by exploiting the differences between education levels and earnings of siblings (Krueger and Ashenfelter, 1992). However, studies regarding twins are often criticised because between-twin differences in schooling are not randomly assigned, but instead are endogenously chosen especially when they depend upon the individual's own aptitude and ability or parental preferences regarding the allocation of expenditure between different children.

A natural experiment is yet another interesting way of tackling the ability bias, where an exogenous event is taken as instrument for the level of education. Some popular natural experiments have been minimum school leaving laws (Harmon and Walker, 1995; Dickson and Smith, 1995), the month of birth (Angrist and Keueger, 1991), and proximity to the school (Card, 1993a, 1999), where the probability of acquiring an extra year of schooling increases/decreases due to the random occurrence of an event that is completely independent of unobserved individual characteristics.

There is another strand of literature that involves identifying an exogenous variable (instrument) that must be correlated with the education level, but is not correlated with the returns to a particular level of education and unobserved ability. In this respect, family background variables such as parental education, spouse's education (Aslam, 2009; Söderbom et al., 2006; Trostel et al., 2002), average education level of the household and/or birth order of an individual (Bertoni and Brunello, 2016; Kantarevic and Mechoulan, 2006) have been used as instruments in the literature. However, the issue remains that because of intergenerational transmission of ability, family background does not completely assure there being no correlation between unobserved ability and the family background variable at hand. Thus the popularly used demand-side family background variables as exogenous determinants of the level of education are often criticised by labour economists as only partially attenuating the ability bias.

As demand-side instruments such as family background, are now widely criticized (Dickson and Smith, 1995), the focus has shifted to the sources of variation in schooling from the supply side such as school-leaving laws or proximity to schools, etc. in search of identifying the source of exogenous variations in education attainment. Using physical capital as an instrument for educational attainment is another way that the literature addresses the endogeneity problem in estimating the returns to education (Duflo, 2001; Maluccio et al., 1998; Card, 1993b). This approach also allows for answer a relevant policy question, namely whether increases in physical infrastructure create opportunities to increase human capital or not. Using physical capital as an instrument is also important from a policy point of view, as increases in human capital ultimately affect the lives and living conditions of citizens, thereby reducing poverty. There is evidence in the literature to suggest that the availability of schools positively affects school enrolment rates owing to the increased and easier access to opportunities to attain education (Khan, 2021; Mazumder et al., 2019; Lavy, 1996; Lillard and Willis, 1994). The availability of schools additionally is linked to improving socio-economic conditions (Carneiro et al., 2013; Case and Deaton, 1999; Currie and Moretti, 2003).

Moreover, Valero and Van Reenen (2019) show that human capital accumulation as well as innovation is an important mediating factor between universities and regional growth.

3. Methodology

In this paper, the author estimates the earnings function (1) using the instrumental variables (henceforth IV methodology). The first stage of the IV procedure is as follows:

$$S_i = \pi_0 + \pi_1 Z_i + \varepsilon_i, \tag{2}$$

where the null hypothesis to be tested is to confirm the instrument (Z_i) relevance $\pi_1 = 0$. The author explains instrument (Z_i) and the first stage (2) in the next section.

3.1. Identification strategy: IV estimation

This study made use of a supply-side IV, i.e. the number of arts and science intermediate, degree and post-graduate colleges for males and females per 10,000 individuals in a district in a given year in Punjab. The region has recently seen tremendous growth in the number of colleges, both private and public, with an increasing number of both genders graduating from these colleges, as shown in Table 1 below. This paper aimed to use this expansion in tertiary education as a means of improving access to tertiary education for male and female students. Moreover, these colleges are also a substitute for private colleges, thereby ensuring ease of access.

Table 1

Number of intermediate, degree colleges, and post graduate classes by gender, their enrolment, and teaching staff in Punjab

Year	No. of colleges			Enrolment			Teaching staff		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls
2005-06	672	339	333	619	273	346	19131	10677	8454
2007-08	744	379	365	676	306	370	20255	11448	8807
2009-10	901	461	440	724	339	385	23096	12645	10451
2012-13	994	492	502	837	416	421	26312	14490	11822
2014-15	1095	543	552	937	455	482	26823	14997	11826

Source: Punjab Development Statistics (Various issues).

The instrument was calculated as follows:

$$\begin{aligned} &\textit{Total no. of colleges per } 10{,}000 \textit{ individuals}_{dk} = \\ &= \frac{\textit{Total no. of colleges in a district}_{dk}}{\textit{District population}_{dk}} \times 10{,}000, \end{aligned} \tag{3}$$

where d is any district in Punjab, k is the year in which individual i was in the normal age range for going to college; more about k in the next section.

3.2. First stage

To empirically test if the expansion of tertiary education translates into a greater accumulation of tertiary education, the first stage of this analysis was a regression of years of education attained by individual i on the number of colleges per 10,000 individuals available in a district in year k in which individual i was at the age of going to college. Since the study pooled cross-section spanning over a decade, this also allowed to include in the sample individuals who in the latest year did not fall into the relevant college-going age range, i.e. 16 to 24, which is the standard range for going to college. The final sample included students between the ages 16 to 32, as one typically enters college at an age of 16. Hence, anyone of this age in any of the included rounds of PSLM, i.e. 2006, 2008, 2010, 2012, 2014, was included in the sample. Similarly, the upper bound for the sample was 32 years, as anyone of that age in the latest year of the analysis, i.e. 2014, would be 24 in 2006, and would have just finished their Master's degree, thus the maximum that these data allowed to include regarded individuals of 32 years of age in 2014. A complete description of the sample in tabular form is given in Table 2, which shows if an individual at a particular age in a particular round of PSLM can be included in the sample or not, which depends on whether they were in the college-going age range, namely 16-24, in any of the included rounds. The highlighted cells (blue) are the age ranges from each round included in the final sample.

Table 2

Sample description

Age in	Year when	Age in	Year when	Age in	Year when	Age in	Year when	Age in	Year when
2014	24	2012	24	2010	24	2008	24	2006	24
32	2006	32	2004	32	2002	32	2000	32	1998
31	2007	31	2005	31	2003	31	2001	31	1999
30	2008	30	2006	30	2004	30	2002	30	2000
29	2009	29	2007	29	2005	29	2003	29	2001
28	2010	28	2008	28	2006	28	2004	28	2002
27	2011	27	2009	27	2007	27	2005	27	2003
26	2012	26	2010	26	2008	26	2006	26	2004
25	2013	25	2011	25	2009	25	2007	25	2005
24	2014	24	2012	24	2010	24	2008	24	2006
23	2015	23	2013	23	2011	23	2009	23	2007
22	2016	22	2014	22	2012	22	2010	22	2008
21	2017	21	2015	21	2013	21	2011	21	2009
20	2018	20	2016	20	2014	20	2012	20	2010
19	2019	19	2017	19	2015	19	2013	19	2011
18	2020	18	2018	18	2016	18	2014	18	2012
17	2021	17	2019	17	2017	17	2015	17	2013
16	2022	16	2020	16	2018	16	2016	16	2014

Source: author's own analysis.

The first stage of this analysis is as follows:

$$S_{idt} - \pi_0 + \pi_1 Z_{dk} + \mu_d + \alpha_t + \varepsilon_{idtk}, \tag{4}$$

where S_{idt} stands for years of education, Z_{dk} is number of colleges per 10,000 individuals as calculated in equation (3), available in district d that individual i is from when he/she was at the age k of going to college, μ_d is the district fixed effects which are critical to control for any unobserved time-invariant district attributes that affect college availability and educational attainment in a district. This allowed for a more robust test proving the first identifying assumption; α_t is the year fixed effects to control for time-specific trends that allow to look for any change over time in the choice or preference of individuals, e.g. increase in demand for education or more progressive thinking overtime, etc.

The main hypothesis tested in the first stage was that exposure to a greater number of colleges does not affect educational attainment; $H_0 : \pi_1 = 0$. This allowed for correlation of errors within the district, namely the use of cluster-robust standard errors.

3.2.1. Instrument validity

Instrument relevance. The proposed instrument is relevant given these colleges are widely dispersed across the entire province, ensuring greater access to education for both males and females. It is important to highlight here that the very policy that guides the setting up of these colleges ensures greater access. These colleges are set up by the Higher Education Department (HED, a ministerial department responsible for higher education) to improve access to education. The criteria considered before setting up a college in a locality are: (i) there is enough population in the area and (ii) the number of students who pass, out of SSC and intermediate levels from that area, and (iii) land available for college building (Higher Education Commission, 2007). With a motive of setting up an educational facility in every neighbourhood, these colleges find their way to localities where there is enough population to take advantage of this facility, and where a college was not already present in that location.

Instrument exogeneity. To satisfy the exclusion restriction one needs to prove that other district or community level attributes are uncorrelated with the supply of colleges. To prove exogeneity it is necessary to show that the increased availability of opportunities to acquire education is not reflective of the better overall development of a region. Later, the results showed that for the relevant age range the effect, for the most part, is driven by the variation in the availability of colleges even after controlling for the indicators of development. Additionally, controlling for district fixed effects allowed to control for unobserved time-invariant district attributes that may affect both the college presence and the educational attainment and may bias the estimated coefficient of the instrument in the first stage.

The other conjecture to strengthen the exogeneity condition relies on the assumption that the exact timing of college opening in a given district is not driven by demand for education, therefore the preferences or demand of citizens for greater opportunities to acquire education is not a concern here. Hence, one can exploit the fact that the contemporaneous supply of colleges in a district is not driven by the demand for education, but is reflective of pent-up demand as it takes time and involves incurring financial costs to respond to the demand for educational institutions, and consequently to set up an educational facility. To ensure that any time-varying trends or preferences are controlled for year, fixed effects were included in the study.

To further strengthen the argument that demand is not so much of a concern here, the role that political influence plays is also worth discussing. One can imagine that if in a certain district the member of the parliament from that district belongs to the opposition party, they would find it difficult to get funding/approval for a new college in the district, whereas if they are from the ruling party, they may get funding/approval for a new college in the district even in the absence of demand. Thus, the link between demand and the opening of a new college is weak. Therefore, the role of political connections needs to be also considered when thinking about the second identifying assumption.

Additionally, the suggestion that the supply of colleges is not driven by demand is further confirmed by the fact that private colleges (which one could assume are a product of demand) cater to only one-fourth of the total student body that goes to these degree colleges to attain tertiary education, and this has consistently been the case for all the years included in the analysis (Statistics of Arts and Science, 2015). However, so as not to be completely ignorant of the effect of demand, the study controlled for district and year fixed effects to account for preferences and changing trends.

3.2.2. Identifying assumptions

Therefore, in running the first stage the two identifying assumptions were:

1. The relationship between changes in college availability and changes in educational attainment is not reflective of changes in development otherwise. This was addressed in the analysis later by showing that the first stage results hold only for the relevant age range that could have benefited from the increased availability of college and are robust to the district fixed effects.
2. The exact timing of college opening in a given district is not driven by the demand for education. This was addressed by controlling for district and year-fixed effects.

3.3. Second stage

The second stage made use of the estimated years of education from the first stage to estimate the returns to years of education. To find the gender gap in returns to education the study also included in the second stage an indicator for gender, and also interacted the gender indicator with the years of education. The second stage specification was as follows:

$$Ln\left(Earnings\right)_{idt} = \beta_0 + \beta_1 \hat{S}_{idt} + \beta_2 Male_{idt} + \beta_3 \hat{S}_{idt} Male_{idt} + \mu_d + \alpha_t + \varphi_i + \omega_i + e_{idt}, \quad (5)$$

where $Ln(Earnings)_{idt}$ is the log of yearly earnings of person i, in district d, in year t, $\hat{S}_{idt}$ is the estimated years of education for person i, in district d, in year t, $Male_{idt}$ is an indicator variable for gender. It is one for males and zero for females, $\hat{S}_{idt} Male_{idt}$ is the interaction of $\hat{S}_{idt}$ and $Male_{idt}$, μ_d is the district fixed effects to control for unobserved time-invariant district attributes, α_t is the year fixed effects to control for time trends, φ_i is the sector fixed effects to control for unobserved time-invariant industry attributes that may influence both an individual's educational attainment and their eventual returns. For instance, many women in Pakistan end up joining either the education sector or the health sector. Finally, ω_i is the occupation fixed effects.

The main hypothesis proposed to test in the second stage is that higher educational attainment does not affect earnings; $H_0 : \beta_1 = 0$. The study allowed for correlation of errors within district, namely using cluster-robust standard errors.

4. Data and the descriptive statistics

To carry out this analysis the author used a pooled cross-section of five rounds of PSLM for 2006, 2008, 2010, 2012, and 2014. The data for the supply of education were collected from the *Punjab Development Statistics and Statistics of Arts and Science Intermediate and Degree Colleges* for the above stated years.[3] Table 3 below is a concise snapshot of the data used for estimating the impact of tertiary education on yearly wages and how it differs by gender across the years included in the analysis. The table shows that males in the sample are on average older, have more years of experience, and earn more than females, there is also a higher chance that the men in the sample, compared to the women are married. However, the highest education level attained by women is higher than by men, a confirmation of the statistics presented in Table 1 that female enrolment in tertiary education has risen so much that it has been higher than for males in recent years. All of these differences between genders are significant at a 1% level of significance as shown by the t-values of difference in the means test. These differences between genders have been the same across all the years. The important thing to note from this table is how women are improving in terms of their prime human capital determinants in that the difference between males and females in years of experience has been shrinking over time. Secondly, women, in trying to catch up with men in terms of human capital, have surpassed them as far as the years of education attained is concerned.

As shown in Table 3, men in this sample on average earn significantly more than women, the same phenomenon is evident in the kernel density graph shown in Figure 1. The data suggest that the distribution of yearly wages for males for different levels of education peaks to the right of the series' mean compared to that of women who have a much wider distribution. It is also noteworthy that the shape of the wage distribution for women changes from bimodal to just like that of men from Intermediate to Master's level. This points to the fact that there is a greater inequality of wages between men and women at lower educational levels. At higher educational levels such Master's, women tend to be doing much better in catching up with men. There are three lessons to learn from these figures: (i) it appears that women do not tend to work in jobs that offer very low wages(as the distribution for women starts at a level higher than for men), (ii) for all the three levels of education, women outnumber men at the lower end of the distribution which means that no matter what level of education they achieve, women earn less than men, and (iii) just above the mean, the distribution for women is lower than men – suggesting that at the higher end of wage distribution, women are outnumbered by men.

[3] Data available on the Punjab Bureau of Statistics website. http://www.bos.gop.pk/ developmentstat

Table 3

Summary statistics

Variable	Female	Male	Difference	Count
	(1)	(2)	(3)	(4)
Year: 2006				
Age	25.002	26.397	1.395***	2,488
Experience (Years)	6.897	8.943	2.046***	2,488
Married	0.309	0.43	0.121***	2,488
Years of Education	13.15	12.478	-0.672***	2,488
Yearly Earnings	69,877.63	116157	46,279.520***	2,483
Observations	573	1,915	2,488	
Year: 2008				
Age	24.955	26.48	1.525***	2,524
Experience (Years)	6.594	8.968	2.373***	2,524
Married	0.262	0.438	0.176***	2,524
Years of Education	13.383	12.541	-0.841***	2,524
Yearly Earnings	88,985.24	158628	69,642.836***	2,524
Observations	572	1,952	2,524	
Year: 2010				
Age	25.11	26.589	1.479***	2,488
Experience (Years)	5.685	7.883	2.197*** 2,488	
Married	0.261	0.45	0.189***	2,488
Years of Education	14.536	13.759	-0.776***	2,488
Yearly Earnings	111171.02	204334	93,162.688***	2,485
Observations	655	1,833	2,488	
Year: 2012				
Age	25.126	26.595	1.468***	2,743
Experience (Years)	6.751	8.605	1.854***	2,743
Married	0.293	0.49	0.198***	2,743
Years of Education	13.415	12.999	-0.416***	2,743
Yearly Earnings	143537.5	236195	92,657.758***	2,739
Observations	816	1,927	2,743	
Year: 2014				
Age	25.248	26.158	0.910***	2,298
Experience (Years)	7.023	8.28	1.257***	2,298
Married	0.352	0.479	0.127***	2,298
Years of Education	13.257	12.887	-0.370***	2,298
Yearly Earnings	131008.49	228949	97,940.266***	2,291
Observations	690	1,608	2,298	

Source: author's own calculations.

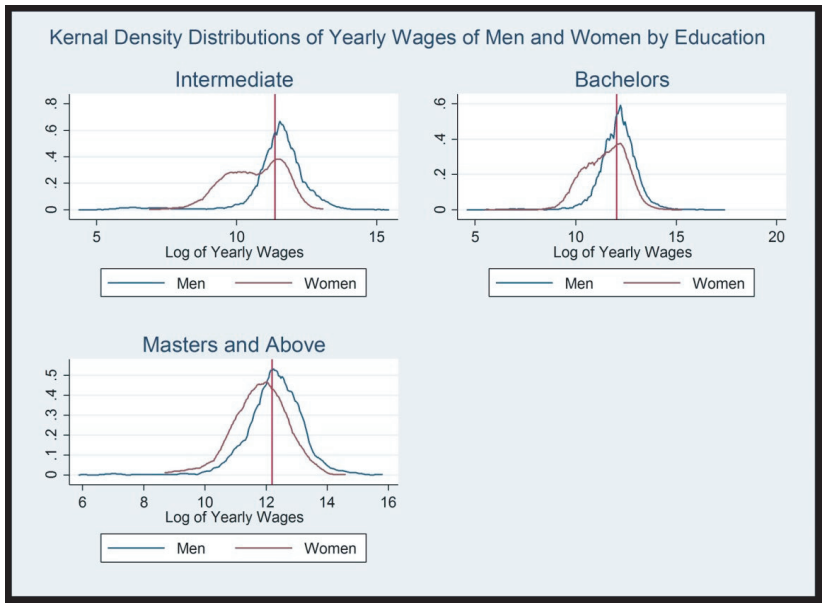


Fig. 1. Wage densities by education for men and women

Source: author's own calculations using various rounds of PSLM data.

As the wage distribution of women is 'bulkier' toward the lower end of the distribution, and begins to decline on the higher end before the men's wage distribution declines, there is a clear indication that women earn less than men. Although the raw data indicates a gender gap in wages earned by men and women who have acquired more than ten years of education, the next section reconfirms this observation using OLS and instrumental variables regression.

5. Results

The results of the first stage are reported in Table 4, where one can see that the chosen instrument predicts the highest education level attained by an individual. The total number of educational institutions per 10,000 individuals in a district in the year when an individual was at the age of going to college significantly affects the years of education. As hypothesised, the greater the number of tertiary education institutions the person is exposed to when they were at the age of going to college, significantly greater the likelihood of going to college and therefore attaining a longer period of education. Moreover, the study also controlled for the unobserved time-invariant factors by checking for district fixed effects and time-varying year fixed effects. The author's argument that increasing the schooling inputs per capita improves access to education subsequently leading to an increase in the highest level of education attained seems to be valid.

To prove that the first stage results were not driven by the overall development in a district, the first stage controlled additionally for community-level development indicators, which are: a source of drinking water, grocery store, public transport, primary school, secondary school, hospital, and a population welfare centre available within thirty minutes' distance of the household. In column 2 in Table 4, one can see that the instrument still continues to hold its significance; although the coefficient's size was reduced, it remains positive and significant.

To prove that the changes in the years of education do not reflect changes in development in general, and that for the sample, i.e. within the given age range additional years of education attained above matriculation are significantly affected by college availability in the district, the first stage was run for individuals who did not fall in the desired sample age range, i.e. aged 16-32, in any of the included years. The logic behind doing this was to see if the relationship between years of education and number of colleges is spurious. If so, then one should see years of education increasing regardless of college presence, even for individuals who are not at the age of going to college.

Table 4

First stage regression

Dependent variables	**Years of Education**	
	(1)	**(2)**
Total number of colleges per 10,000 individuals	4.52**	0.936*
	(1.821)	(0.558)
Year FE	YES	YES
District FE	YES	YES
Development Indicators	NO	YES
Observations	11,677	11,677
Mean of the dependent variable	12.86	12.86

Note: Robust standard errors in parentheses. SE's clustered by districts. Controls in the first stage include experience, experience squared, married. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Source: author's own calculations.

Therefore, the analysis was run on very narrow age bands around the upper (32 yrs) and lower (16 yrs) age cut-off to make the comparison between roughly similar groups. Thus, the comparison at the lower end of age distribution was made between individuals from the age bracket 12-15 and those from16 to 19. At the upper end of the age distribution the comparison was made between individuals from the

age bracket 29-32, and those from 33 to 36. For each of these four regressions other than the basic controls included in Table 4, development indicators were also controlled for. The coefficient plots of these regressions are presented in Figure 2. Therefore, for individuals who are at the age of going to college, it can be argued safely that college presence matters. If this was a spurious result, then one would also see some impact in the similar groups, and human capital increasing regardless of college presence, thus reflecting the general progression of the society. Additionally, year fixed effects were also incorporated in generating Figure 2, to control for trends that change over time, such as a preference for longer periods of education, etc.

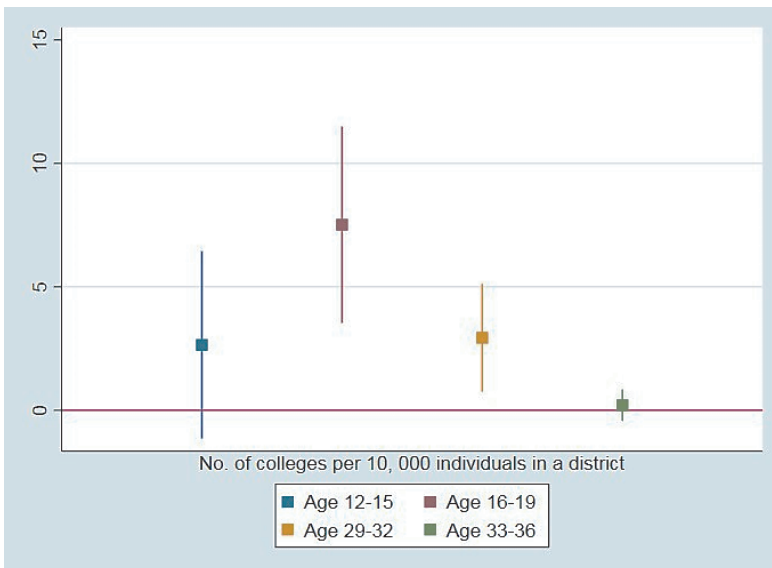


Fig. 2. Estimates of beta coefficients of number of colleges for tighter time windows controlling for community development

Note: Community development indicators controlled for are: access to piped water, grocery store, public transport, primary school, middle school, hospital, and population welfare centre.

Source: author's own analysis.

The results for the second stage show that there are positive returns to attaining tertiary education, and that there is a gender gap in those returns in favour of men. The results for the second stage are presented in Table 5. Column 1 shows that years of education beyond matriculation significantly affect one's wage. Moreover, *Male* – the indicator for gender, shows that men have higher wages than women on average. However, an additional year of education brings a significantly greater increment in wages of women compared to those of men, as the coefficient on the interaction term of *Male* and *Years of education* is significant and negative. This result is in line with earlier findings in the literature, which also suggest that as years of schooling

increase, gender wage gap tends to fall (Blau and Kahn, 2017; Blundell et al., 2000). These higher marginal returns for women could result from the dual impact, i.e a direct effect of human capital accumulation on returns which is also true for men, but in addition, there is an indirect impact for women which is due to the attenuation of the impact of discrimination, tastes and circumstances (DTC) (Dougherty, 2005).

Tirmazee (2021) confirmed this finding by showing that the gender wage gap is highest at the lowest end of the wage distribution, and is contributed in large part by the unexplained gap. The inverse relationship between DTC (hence the wage gap) and the years of schooling could probably be as more educated women have a degree or a formal qualification required in a job that offers a standardised wage, or highly educated women may be able to deal well with discrimination, or may even be able to find better job openings for herself where her characteristics are rewarded fairly (Dougherty, 2005). In column 2 in Table 5, one can see that all of these results are robust to controlling for development indicators.

Table 5

Second stage regression

Dependent variables	Log of Yearly Earnings	
	(1)	(2)
Years of Education	0.827***	1.041**
	(0.080)	(0.434)
Male	1.885***	2.292***
	(0.140)	(0.807)
Male* Yrs of Education	-0.553***	-0.757*
	(0.073)	(0.410)
Observations	11,677	11,677
Mean	11.62	11.62
Year FE	YES	YES
District FE	YES	YES
Industry FE	YES	YES
Occupation FE	YES	YES
Development Indicators	NO	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. *** p<0.01, ** p<0.05, * p<0.1. Controls include: experience, experience squared, marital status, marital status interacted with gender. Column (2) additionally controls for development indicators.

Source: author's own analysis.

An important consideration in estimating the second stage is the possibility of difference in the quality of education imparted in male and female colleges. However, that is not problematic, as all of these colleges are set up by the provincial ministerial

department HED under uniform guidelines and are of similar quality. As mentioned above, these colleges are also cheaper, more accessible options made available for students who cannot afford to study in private colleges. The teaching staff in these colleges are recruited through a central standard procedure and are rotated periodically between colleges. All the colleges must meet the minimum requirements of available legal and physical infrastructure as outlined by the Higher Education Commission in the PU-01 proforma for setting up higher education Institutions.[4]

6. Robustness checks

6.1. Selection's correction

Next, the author aimed to correct for selection bias given the very selected sample of individuals who are in paid employment and also have acquired more than ten years of education. The fact that the decision to join the workforce for women is not random, given their expected gender role, they are required to look after the family and the household. The need to maintain a work-life balance may affect women's decision to participate in the workforce, and depending on that, may also affect their choice of job, profession, or industry. There could also be some self-selection on the men's side, if they deliberately decide to enter paid work rather than be self-employed. This possibility is equally valid for women as well. All of these decisions made by men and women are dependent upon their socio-economic conditions and may ultimately affect the returns they earn in the labour market. The extent to which a certain level of education is financially beneficial for men and women is thus not independent of the choices made, and therefore can be expected to interact with these work-related choices to either attenuate or augment the returns to education.

Another source of the sample selection bias is the very selected sample of individuals who have more than ten years of education, since a very privileged group of people continue to post-secondary education in Pakistan. Although women have caught up with men at tertiary educational level, at lower levels there is still a substantial gender gap which means that there is a very narrow group of women who enter into tertiary education.

To correct for both selection biases, the study used the Heckmann two-step procedure (Heckman, 1976, 1979). Correcting for selection in paid employment involved first estimating the probability of participating in the workforce using the probit model. The exclusion restriction of the participating equation included dependent children and adults aged less than seven years and above sixty years respectively in the household. The probability of salaried employment estimated

[4] https://hec.gov.pk/english/services/universities/Documents/887_HEC2_Criteria_of_university_institutions.pdf

from the participating equation was used to estimate the inverse Mills ratio or the selectivity term (lambda) which was later used as one of the controls in the second stage.

To correct for selection in higher education, the author ran a second selection function where the probability of having acquired higher education was regressed on all the controls in the main wage equation, along with the average education level of the household as an exclusion restriction.[5] The probability of acquiring higher education estimated from this selection function was used to estimate another inverse Mills ratio that was also used as an additional control in the second stage.

To correct for endogeneity between years of education and wages along with correcting for the two selection biases mentioned above, the study incorporated selectivity terms (IMR1 and IMR2) as controls in the IV regression. The results are shown in columns 2 and 3 of Table 6. Here too, the coefficients' sizes, their signs, and significance are similar to those obtained in the simple IV regression in Tables 4 and 5. The results from this section suggest that when corrected for the selection bias, the main results continue to hold and therefore it is safe to assume for Punjab's labour market, that for both men and women who have more than ten years of education additional years of education bring greater returns, but that those additional years are comparatively more beneficial for women than for men, leading to a reduction in the gender wage gap at each successive level of education.

6.2. Are the results driven by affluent districts?

There is also some a-priori evidence from the data to suggest that in districts where the enrolment per capita is higher, the Human Development Index[6] is also higher for the year 2015. The defined concept of Human Development Index, reflects the increase in the capabilities of people by providing them with increased opportunities and the 'freedom of choice' to avail those opportunities. The choropleth maps of the province of Punjab below show that there is indeed a correlation between the HDI (Figure 3) and the enrolment in higher education (Figure 4) across districts.

This is especially true for districts in the north and the centre. Thus, for instance, Rawalpindi (north) and Lahore (centre) with a very high Human Development Index, also show very high levels of enrolment per capita in higher education. Similarly, districts in the south such as Rajanpur, Bahawalnagar, and Bahawalpur, and those in the West, such as Dera Ghazi Khan, and Muzaffargarh, have both lower Human Development Indices and also lower enrolements per capita.

---

[5] Although a better exclusion restriction could have been the parental education (Asadullah and Xiao, 2019), the data allowed this information only for a very small subsample, therefore the study used the average education level of the household excluding one's own.

[6] The HDI for Punjab has been taken from the Pakistan Human Development Index report 2017.

Table 6

Correcting for selection

Dependent variables	Log of Yearly Earnings	Yrs. of Education
	Second Stage	First Stage
	(1)	(2)
Yrs. of Education	1.128** (0.445)	
Male	2.526*** (0.821)	
Male* Yrs. of Education	-0.860** (0.418)	
IMR1	-0.007 (0.135)	
IMR2	-0.320*** (0.060)	
Log Total Number of Colleges		0.951* (0.551)
Year FE	YES	YES
District FE	YES	YES
Industry FE	YES	YES
Occupation FE	YES	YES
Development Indicators	YES	YES
Observations	11,398	11,398

Note: Robust standard errors in parentheses. SE's clustered by districts. ***p<0.01,**p<0.05, *p<0.1. The dependent variable in the participation equation is salaried employment. The exclusion restriction for the participation equation is dependent children and adults aged < 7 yrs and > 60 yrs respectively in the household. IMR1 is the inverse Mills ratio calculated from this equation. The dependent variable in the selection function for selection into higher education is if individuals have acquired more than ten years of education. The exclusion restriction used is the average education level of the household excluding one's own education. IMR2 is the inverse Mills ratio estimated from this second selection function.

Source: author's own analysis.

Therefore, there is an indication from the data that higher human capital accumulation is correlated with better human development. This points to the need of devising ways that can ease the accumulation of human capital and providing physical infrastructure, i.e. schools and colleges is one of the policy options available. The study provides evidence that the responsiveness of human capital accumulation is higher in districts that are relatively worse off as indicated by HDI, making it all the more policy-relevant to invest in physical infrastructure in poorer or worse-off areas.

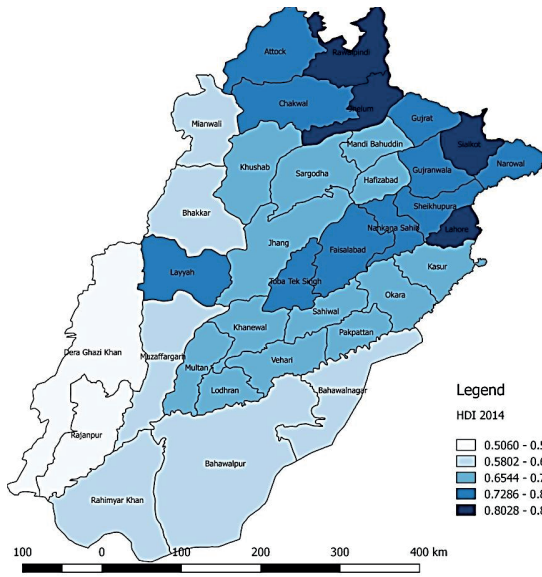


Fig. 3. Human Development Index, 2015

Source: author's own analysis using HDI figures from the UNDP Human Development Index report, 2017.

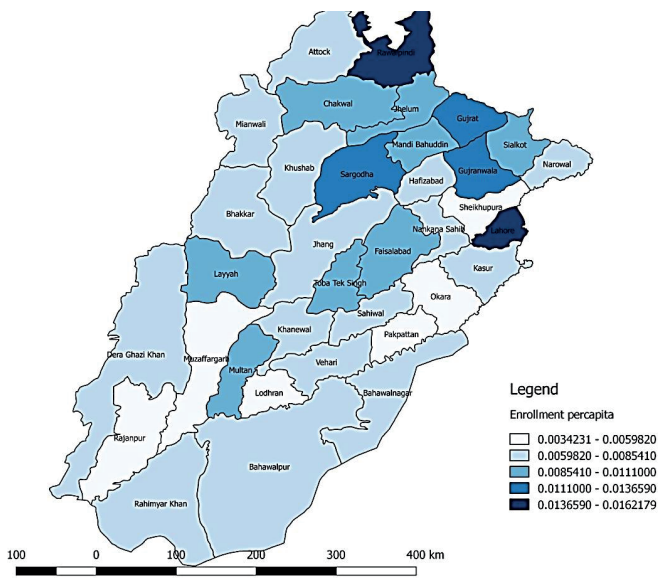


Fig. 4. Enrolment per capita, 2015

Source: author's own analysis using enrolment figures from the Punjab Development Statistics, an annual publication of the Punjab Bureau of Statistics.

Table 7

IV regression: by HDI

Variables	High HDI		Low HDI	
	Second Stage	First Stage	Second Stage	First Stage
	(1)	(2)	(3)	(4)
Yrs. of Education	1.074		1.322***	
	(0.800)		(0.495)	
Male	2.321		2.851***	
	(1.497)		(0.951)	
Male* Yrs. of Education	-0.770		-1.049**	
	(0.760)		(0.476)	
Log Total Number of Colleges		0.447		1.684***
per 10,000 individuals		(0.559)		(0.439)
Observations	7,075	7,075	4,602	4,602
Year FE	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. ***p<0.01, **p<0.05, *p<0.1.

Source: author's own analysis.

To rule out the possibility that the effect observed in the first stage may be driven by the more affluent districts where the human capital accumulation and physical infrastructure is relatively abundant, the author ran the analysis separately for better-off and worse-off districts. The distinction was made based on the HDI of the districts in 2015. The results are shown in Table 7. The results of the first stage are significant for poorer low HDI districts while they are insignificant for the richer high HDI districts. This shows that there is a greater impact of investing in physical infrastructure where the opportunities are already lagging. The results for the second stage also are significant for the low-HDI regions. The earnings for men are higher on average but an extra year of education brings comparatively greater returns for women than for men.

6.3. Are the results driven by men?

To rule out the possibility that the effect observed in the first stage may be driven by the male colleges, and that the availability of colleges may be increasing their enrolment relatively more than those for females, the colleges were split by gender

and by HDI levels to see the impact of each on the educational attainment. To distinguish the responsiveness of female human capital to the presence of physical infrastructure from that of males, the author ran the IV regression by breaking down the total number of colleges into male and female colleges, taking the two as separate instruments in the first stage.

Table 8

IV regression: men and women colleges

Variables	High HDI		Low HDI	
	Second Stage	First Stage	Second Stage	First Stage
	(1)	(2)	(3)	(4)
Yrs. of Education	1.024***		1.002***	
	(0.157)		(0.203)	
Male	2.21***		2.186***	
	(0.285)		(0.375)	
Male* Yrs. of Education	-0.716***		-0.738***	
	(0.148)		(0.190)	
Log Total Number of Girls' Colleges		0.0004		0.004*
per 10,000 girls		(0.001)		(0 .002)
Log Total Number of Boys' Colleges		0.001		0.001
per 10,000 boys		(0.001)		(0.001)
Observations	6,937	6,937	4,764	4,764
Year FE	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. ***p<0.01, **p<0.05, *p<0.1.

Source: author's own analysis.

The results shown in Table 8 demonstrate that when the colleges are split up by gender, only the coefficient for female colleges is significant and positive in the first stage in the poorer districts; the coefficient for male colleges is not significant in any of the regressions. In the richer districts, the significance of female colleges also disappears, which is an even stronger indication of the result obtained in section 6.2. College availability seems to make a difference to the marginalised group in lagging areas which include districts in the south and west of the province. These districts are also much more conservative in their values regarding educating women, let alone

allowing them go to a distant college in a neighbouring district or provincial capital. Therefore, increased college availability in a district may ease the mobility constraint for a lot of female students, making gender-segregated tertiary education institutes available for them (where others from similar backgrounds may be able to acquire higher education degrees). The results of the second stage stay the same.

Conclusion

This paper is an investigation of the gender gap in the returns to tertiary education in Punjab using the Instrumental Variables technique. The study used exogenous variation in the expansion of the supply of higher education institutions to men and women to identify and compare the returns to tertiary education for both genders in Punjab, Pakistan. A large number of colleges in a given district can affect the probability of moving from secondary to tertiary education, since their accessibility improves by alleviating two constraints, namely the high cost of acquiring a higher degree, and mobility if a college is built in one's locality. To carry out this analysis the author used a pooled cross-section constructed from five rounds of the PSLM survey for 2006, 2008, 2010, 2012, and 2014.

The results of this analysis suggest that there is a significant positive relationship between years of education beyond matriculation and the earnings of individuals. Moreover, the marginal returns to acquiring one extra year of education are higher for women than for men, suggesting that gender inequality tends to fall as human capital accumulation improves. The first stage results confirm that it is important to invest in building infrastructure to increase educational attainment. Having controlled for other development indicators and shown that this relationship holds for the appropriate age range (16-32, the sample should cover college-going age during one or more years included in the study), we remove doubts regarding the first stage results as not being causal.

Some important policy lessons to be learned from this analysis are: firstly, the significance of investing in higher education, both because it increases the prospects of graduates in the labour market by increasing labour market returns, and also it decreases gender inequality in the labour market returns. Secondly, the significance of investing in the physical infrastructure such as universities or higher education institutions, as this facilitates the accumulation of human capital by making it less costly for households to invest in it when a higher education institution is built in their locality. Cheema et al. (2019) described the glass walls hindering women to take up training, and showed that once an education centre is housed in their village, it significantly increases their take-up rates.

Thirdly, the responsiveness of human capital investment to investment in physical capital is greatest in the less developed regions of Punjab, hence the greatest returns can be achieved by targeting the expansion of educational institutes to lagging

regions. Using physical capital as an instrument is also important, as an increase in human capital ultimately affects the lives and living conditions of citizens, reducing poverty. There is evidence in the literature to suggest that the availability of schools positively affects school enrolment rates owing to the increased and easier access to opportunities to attain education. In addition, the availability of schools is linked to improving socio-economic conditions (Carneiro et al., 2013; Case and Deaton, 1999; Duflo, 2001; Currie and Moretti, 2003). Moreover, Valero and Van Reenen (2019) showed that human capital accumulation, as well as innovation, is an important mediating factor between universities and regional growth.

Fourthly, a more indirect lesson – a spin-off of the first two lessons – is the spillover effects of investing in physical infrastructure, i.e. the government, by establishing tertiary education institutions in the less developed regions could promote the growth of private low-cost high schools in the area, as graduates from the higher education institutions, enter the labour market to increase the supply of teachers at primary and secondary levels of schooling. This is possible because of an increase in the supply of school teachers graduating from these tertiary education institutions (Andrabi et al., 2008).

There are some important directions that this work could be extended into, that are currently either beyond the scope of this study or because of data being unavailable. The first important direction for the future research is to find out which subject streams or fields tend to reduce the gender gap in the returns to tertiary education the most. Secondly, the distance to a tertiary education institution is a better reflection of improvement in access and could be a better instrument. In this case, the data does not allow to include the distance from a household to a college. Another important research dimension is to examine, while tertiary education is expanding, what is happening to the quality of higher education imparted across different institutions, and what implications could this have for the gender gap in labour market outcomes. Lastly, the fate of graduates largely depends on the balance between supply and demand. An interesting extension of this analysis could be to find out how much of the results are driven because of the expansion of access, or because of the expansion of the demand for graduates in the labour market.

References

Andrabi, T., Das, J. and Khwaja, A. I. (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review*, 52(3), 329–355.

Angrist, J. D. and Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014.

Asadullah, M. N. and Xiao, S. (2019). Labor market returns to education and English language skills in the People's Republic of China: An update. *Asian Development Review*, 36(1), 80–111.

Ashenfelter, O. and Zimmerman, D. J. (1997). Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers. *Review of Economics and Statistics*, 79(1), 1–9.

Aslam, M. (2009). Education gender gaps in Pakistan: Is the labor market to blame? *Economic Development and Cultural Change*, 57(4), 747–784.

Attanasio, O. and Kaufmann, K. (2009). Educational choices, subjective expectations, and credit constraints. Technical report, National Bureau of Economic Research.

Banerjee, A., Glewwe, P., Powers, S. and Wasserman, M. (2013). *Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence*. Citeseer.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Bertoni, M. and Brunello, G. (2016). Later-borns don't give up: The temporary effects of birth order on European earnings. *Demography*, 53(2), 449–470.

Bingley, P., Christensen, K., and Walker, I. (2009). The returns to observed and unobserved skills over time: Evidence from a panel of the population of Danish twins. *Danish National Institute for Social Research*.

Blau, F. D. and Kahn, L. M. (1992). The gender earnings gap: Learning from international comparisons. *The American Economic Review*, 82(2), 533–538.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865.

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000). The returns to higher education in Britain: Evidence from a British cohort. *The Economic Journal*, 110(461), F82–F99.

Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D. and Spector, T. (2002). Estimating returns to education using a new sample of UK twins. *Queen Mary and Westfield College, London*.

Card, D. (1993a). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Card, D. (1993b). Using geographic variation in college proximity to estimate the return to schooling.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics*, volume 3, pages 1801–1863. Elsevier.

Carneiro, P., Meghir, C. and Parey, M. (2013). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association*, 11 (suppl 1), 123–160.

Case, A. and Deaton, A. (1999). School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, 114(3), 1047–1084.

Cheema, A., Khwaja, A. I., Naseer, F. and Shapiro, J. N. (2019). Glass walls: Experimental evidence on access constraints faced by women. Technical report, Mimeo, Harvard University.

Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly journal of economics*, 118(4), 1495–1532.

Dickens, W. T. and Lang, K. (1993). Labor market segmentation theory: Reconsidering the evidence. *Labor economics: Problems in analyzing labor markets*, 141–180.

Dickson, M. and Smith, S. (1995). What determines the return to education: An extra year or a hurdle cleared? *Economics of Education Review*, 30(6), 1167–1176.

Dinkelman, T., Martinez, A. C. (2014). Investing in schooling in Chile: The role of information about financial aid for higher education. *Review of Economics and Statistics*, 96(2), 244–257.

Dougherty, C. (2005). Why are the returns to schooling higher for women than for men? *The Journal of Human Resources*, 40(4), 969–988.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795–813.

Epstein, C. F. (1988). *Deceptive distinctions: Sex, gender, and the social order*. Yale University Press.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American EconomicReview*, 104(4), 1091–1119.

Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, 85(5), 1278–1286.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. In *Annals of economic and social measurement*, 5(4), 475–492, NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.

Higher Education Commission (2007). Guidelines for the establishment of a new university or an institution of higher education.

Isacsson, G. (1999). Estimates of the return to schooling in Sweden from a large sample of twins. *Labour Economics*, 6(4), 471–489.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2), 515–548.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *The Quarterly Journal of Economics*, 127(2), 753–792.

Kantarevic, J. and Mechoulan, S. (2006). Birth order, educational attainment, and earnings an investigation using the PSID. *Journal of Human Resources*, 41(4), 755–777.

Khan, F. (2017). Barriers to pay equality in Pakistan. *International Labor Organization*, Geneva, Switzerland.

Khan, S. (2021). Unmet expectations: The impacts of school construction on female outcomes in rural Punjab, Pakistan.

Krueger, A. and Ashenfelter, O. (1992). Estimates of the economic return to schooling from a new sample of twins. *National Bureau of Economic Research.*

Lavy, V. (1996). School supply constraints and children's educational outcomes in rural Ghana. *Journal of Development Economics*, 51(2), 291–314.

Levanon, A., England, P. and Allison, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950–2000 U.S. census data. *Social Forces*, 88(2), 865–891.

Lillard, L. A. and Willis, R. J. (1994). Intergenerational educational mobility: Effects of family and state in Malaysia. *Journal of Human Resources*, 29(4), 1126–1166.

Maluccio, J. A. (1998). Endogeneity of schooling in the wage function. Technical report, International Food Policy Research Institute (IFPRI).

Mazumder, B., Rosales-Rueda, M. and Triyana, M. (2019). Intergenerational human capital spillovers: Indonesia's school construction and its effects on the next generation. In *AEA Papers and Proceedings*, 109, 243–249.

Miller, P., Mulvey, C. and Martin, N. (1995). What do twin studies reveal about the economic returns to education? A comparison of Australian and U.S. findings. *The American Economic Review*, 85(3), 586–599.

Mincer, J. (1962). Labor force participation of married women: A study of labor supply. In *Aspects of labor economics*, 63–105, Princeton University Press.

Mincer, J. (1974). Schooling, experience and earnings. Series *Human Behavior Social Institutions*, NBER.

Neumark, D. and Korenman, S. (1992). Sources of bias in women's wage equations: Results using sibling data, NBER Working Papers 4019, National Bureau of Economic Research, Inc.

Söderbom, M., Teal, F., Wambugu, A. and Kahyarara, G. (2006). The dynamics of returns to education in Kenyan and Tanzanian manufacturing. *Oxford Bulletin of Economics and Statistics*, 68(3), 261–288.

Tirmazee, Z. (2021). Too much of a good thing? Increasing gender wage disparity in face of rapidly expanding post-secondary education in Punjab, Pakistan. Technical report.

Trostel, P., Walker, I. and Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. *Labour economics*, 9(1), 1–16.

Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68, 53–67.

Received: September 2021, revised: February 2022

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Monika Kwiecińska**, *Katarzyna Grzesik***,
*Anna Siewierska-Chmaj****, *Anna Popielska-Borys*****

Generational differences in values and patterns of thinking in the workplace

The aim of this article is to empirically identify and analyse the differences between generations X, Y and Z in terms of values and patterns of thinking in the workplace. For the purposes of achieving the set objective, quantitative studies were performed. Data were obtained with the use of the MindSonar psychometric test. The results were subjected to a statistical analysis, and 435 economically active respondents took part in the study. The study shows that there are numerous statistically significant differences in preferred values and patterns of thinking among representatives of various generations (the differences occur in 2 out of the 7 measured levels of existence and their corresponding values, and in 9 out of the 32 measured patterns of thinking, or metaprograms). The obtained results may help HR managers and specialists focus on real, diagnosed differences in employees, which may be used for determining and communicating organisational priorities for current and future employees.

Keywords: generations, values, patterns of thinking, workplace, Poland

JEL Classification: J24, M54
DOI: 10.15611/aoe.2023.1.05

Quote as: Kwiecińska, M., Grzesik, K., Siewierska-Chmaj, A., Popielska-Borys, A. (2023). Generational differences in values and patterns of thinking in the workplace. *Argumenta Oeconomica*, 1(50), 95-118.

1. Introduction

The co-existence of various generations of employees in the workplace constitutes one of the present challenges of managing human resources (Singh, 2014; Culpin et al., 2015; Holian, 2015; Guerin-Marion et al., 2018). What is significant in this context

* Business and Management Faculty, Wroclaw University of Economics and Business, Poland. ORCID: 0000-0002-6536-2550.
** Business and Management Faculty, Wroclaw University of Economics and Business, Poland. ORCID: 0000-0002-3998-8445.
*** Institute of Social Sciences and Humanities, Tischner European University, Kraków, Poland. ORCID: 0000-0001-7943-3746.
**** Institute of Social Sciences and Humanities, Tischner European University, Kraków, Poland. ORCID: 0000-0001-7545-0653.

is the complexity of the process of creating a work environment that attracts and satisfies different generations of employees (Hansen and Leuty, 2012).

The increasing interest in generational diversity in the workplace is reflected in numerous scientific studies. Some of those are based on the assumption that employees from different generations differ significantly in terms of their goals, expectations and values in relation to the workplace (Madera et al., 2011; Costanza et al., 2012; Lyons and Kuron, 2014). The said works also include studies which formulate hypotheses based on stereotypes regarding generational differences (Giancola, 2006; Posthuma and Campion, 2009; Costanza and Finkelstein, 2015; Hayes et al., 2018; van Rossem, 2019). The existing research indicates that there are certain differences between generations in terms of their attitudes to work and career; these differences, however, are usually not large (Macky et al., 2008) and sometimes they defy well-established stereotypes.

A large number of studies make the general assumption that each generation has its characteristic needs, values and attitudes which relate both to private as well as professional life (Glass, 2007; Dries et al., 2008; Lyons and Kuron, 2014). Such an approach is justified by the fact that members of a given generation experience together common events, phenomena and trends, which makes them have a similar perception of the world (Kindrick Patterson, 2007). Changes occurring in the nature of work and the fact that each generation entered the labour market at a different period may imply that there are differences in values and conduct in the workplace between representatives of various generations. Lester et al. (2012) indicated that the perception of generational differences in the workplace does exist, even if these differences are not always proven empirically. The perception of differences between given generations may pertain to different aspects of work, such as work values (Parry and Urwin, 2011), work attitudes (Costanza et al., 2012), and the matching of the values of an organisation with the values of an individual (Cennamo and Gardner, 2008). There is practically no research in which dependent variables are values and patterns of thinking preferred by members of given generations, the so-called metaprograms, which translate into specific behaviour in the workplace.

Therefore, the aim of this article was to empirically identify and analyse the differences between generations X, Y and Z in the scope of the preferred values and patterns of thinking in the workplace. This aim was achieved by conducting empirical examinations. The study utilised the CAWI method, in which the MindSonar psychometric test was used to collect data. The obtained results were subjected to a statistical analysis in order to determine differences between the studied representatives of the three generations. This article contributes to explaining intergenerational differences from the perspective of assumed values and patterns of thinking, and the results show how identified values and patterns of thinking can shape the expectations, attitudes and behaviours of employees from different generations. The study results may prove useful in taking actions aimed at limiting conflicts or misunderstandings based on generational stereotypes, and may result in

an increment of knowledge in the general discussion on generational differences in the workplace.

The first part of the article is devoted to a theoretical framework of the issue of intergenerational differences. Later, the article presents the methodology of empirical study, followed by a presentation of the results of the conducted studies, which are then interpreted and related to the research of other authors in the 'discussion' part. The final part of the article contains key conclusions from the study.

2. Literature review

In today's labour market, there are employees representing various generations. Whether a given person belongs to a given generation depends most of all on the age bracket (date of birth). The generation of Baby Boomers comprises persons born in the period 1946-1964, Generation X those born in 1965-1979, whereas the determinant of the next generation has not been precisely established yet. Some researchers and practitioners refer to the last generation as Millennials or Generation Y (1980-2000), while others distinguish Generation Y and Generation Z within this group. Generally, it is assumed that Generation Z are persons born starting from 1995 (Schawbel, 2014; Berkup, 2014).

Generation groups, or cohorts, understood as groups comprising persons based on cut-off values of their dates of birth constitute a simplified theoretical structure for generational studies (Kupperschmidt, 2000; Smola and Sutton, 2002; Foster, 2013). Most studies concerning generational differences in the workplace are carried out from the cohort perspective and assume the occurrence of psychological and behavioural differences between various generations (Lyons and Kuron, 2014).

Generational identity does not only mean a similar age of persons classified within the same generation, but is also the result of historical and social experiences of individual generation cohorts (Egri and Ralston, 2004). According to Smola and Sutton (2002), the social context in which a given cohort develops affects their values and beliefs regarding the organisation, work ethics, cause and mode of action as well as goals and aspirations for their social lives. Factors such as globalisation processes and the quick technological progress affect the changing values and expectations of individual employees in the perspective of consecutive generations. In addition, the complexity of generational problems may be better explained when taking into account variables such as the level of education, race, gender, age, and geographical location.

In spite of the growing number of studies concerning generational changes in the workplace, today scientists and practitioners dealing with management face a disorientating disorder of proof generated in various contexts using various methodological and theoretical perspectives (Lyons and Kuron, 2014). The conclusions reached on the basis of cross-cutting studies do not give grounds for their comparison, which hinders the identification of common, recurring patterns, because these studies

feature samples from different countries or trades, or they compare generations using different scopes of generational identity.

Concentrating on the values of work constitutes one of the basic scopes of studies of generational differences. The assessment of differences in work values is performed based on distinguished features of work, such as: work ethics, free time, external values (money, status), internal values (sense of meaning, talent), altruistic values (charity work, helping others), social values (need to belong). Parry and Urwin (2011) carried out a review of research in this area, and found a lack of clear results confirming the occurrence of generational differences in work values. Studies analysing the meaning of hard work for individual generations point to a decreasing trend for this value with each consecutive generation (Cogin, 2012; Gursoy et al., 2013). This trend is also reflected in the results of research demonstrating an increase of demand for more free time with each consecutive generation (Takase et al., 2009; Cogin, 2012; Twenge et al., 2010). Most studies concerning altruistic values, such as charity work and helping others, did not find any significant differences (Cennamo and Gardner, 2008; Twenge et al., 2010; Hansen and Leuty, 2012); however, the results of research conducted by Chen and Choi (2008) show that Baby Boomers value altruistic work higher than members of the younger generation. Hansen and Leuty (2012) demonstrated that younger generations put greater emphasis on social links at work. The study conducted by Twenge et al. (2010) indicated that the significance of internal values, such as remuneration or prestige, drops slightly over the course of generations. Lester et al. (2012) found that older generations value professionalism more than younger ones. This is in line with Wilsa et al. (2011), finding that, compared to younger generations, Baby Boomers give greater importance to observing rules and codes of conduct, while paying less attention to entertainment and stimulation at work.

The matching of the values of an organisation with the values of an individual is important for all generational groups (Cennamo and Gardner, 2008). The first question that emerges in this respect is: which values are preferred by various generations in the workplace and do they differ? Thus, the following research hypothesis was tested.

Hypothesis 1. *Generations X, Y, Z differ in terms of preferred values in the workplace.*

One of the main subjects of research in the area of generational differences is work attitudes, most notably organisational commitment and work satisfaction (Costanza et al., 2012). A large number of studies in this scope cover only one organisation, trade or profession, which does not allow formulating generalised conclusions regarding entire populations. The results of these studies suggest that organisational commitment differs across generations and it almost always decreases in subsequent, younger generations (Lyons and Kuron, 2014). The results of those studies, however, are not consistent as some of them indicate the occurrence of a declining level of work satisfaction in consecutive generations (Beutell and Wittig-Berman, 2008; Benson and Brown, 2011), Kowske et al. (2010) found decreasing

work satisfaction in the case of Baby Boomers and then a slightly higher level of satisfaction in younger generations, with the general satisfaction of Generation Y being above the average. Similar findings were determined in the case of work satisfaction. Some studies indicate the existence of certain significant generational differences in attitudes to work (Park and Gursoy, 2012), while others in the same scope, point to the existence of similarities (Barron et al., 2014; Cucina et al., 2018). Some studies suggest decreasing work commitment with each consecutive generation (D'Amato and Herzfeldt, 2008; Brunetto et al., 2012; Lub et al., 2012). In turn, Davis et al. (2006) found that younger generations demonstrate greater involvement in their organisation than Baby Boomers.

An analysis of the results of the existing research indicates the occurrence of specific differences in attitudes at work across generations. There have been, however, no results relating to patterns of thinking which would determine the existence of certain attitudes, and, consequently, specific behaviour in the workplace resulting from them. In order to explore generational differences in patterns of thinking in the workplace, the following research hypothesis was tested.

Hypothesis 2. *Generations X, Y, Z differ in terms of preferred patterns of thinking in the workplace.*

3. Research methodology

In order to test the research hypothesis, quantitative studies were carried out using the CAWI method. The selection of the sample was intentional and voluntary. The research sample consisted in graduates of universities where the authors of this article carry out their research. Invitations to take part in the study were sent by e-mail with the use of the universities' databases on graduates. The study was conducted on participants who had responded to a study announcement and filled in a psychometric on-line test. In order to enter the study, respondents had to meet the following conditions: being employed for at least 3 months, having higher education and being part of generation X or Y or Z.

The study assumed that generation X includes persons born in the years 1965-1979; generation Y are persons born in 1980-1994; and generation Z are those born after 1995. The study covered 435 respondents from Poland, of which 26 persons represented generation X, 252 generation Y, and 157 generation Z. The average age of the respondents was: for generation X: 42.3, generation Y: 27.0, and generation Z: 22.8. The study did not include the generation of Baby Boomers (BB) born in the period 1946-1964 because this generation is slowly ending their professional activity, whereas the respondents identified with generation X were treated as a control group. The surveys were conducted in 2019 and 2020.

The survey was carried out with the use of the MindSonar psychometric test, a functional psychological system which examines how people think in certain

situations (contexts) and what their internal values and patterns of thinking and acting are. MindSonar measures values and attributes them to 7 levels of existence in accordance with the Spiral Dynamics model, and to 32 patterns of thinking, or metaprograms (Černý, 2017; Hollander, 2014). The MindSonar test presents the respondent with 76 questions and two tasks (criteria sorting and criteria categorisation). Then, the program asks the respondent to define four things they find important in the chosen context (four criteria). Once the criteria have been defined, the respondent is asked to order them from the most to least important (hierarchy of criteria). He or she is then shown the four criteria and asked to fill in their opposites. There are five types of test items for metaprograms:

- *Identification items* – Photographs showing people thinking different things (in text balloons). The respondent picks the person who thinks most like them in that context.
- *Symbolic items* – The respondent chooses from a set of symbols.
- *Avoidance items* – An avoidance question is asked ("What do you want to prevent?").
- *Keyword items* – The respondent chooses from different keyword combinations.
- *Straightforward items* – The respondent is asked directly about the metaprogram in a question ("Do you think more like this or more like this?").

MindSonar measured the following metaprograms – patterns of thinking. They have a determining effect on the functioning of individuals, because they translate into both verbal and non-verbal behaviour, affecting emotions and feelings. They indicate certain cognitive and perceptual preferences (Bolstad and Hamblett, 2001; Brown, 2006) which filter and sort information (Brown, 2003; Brown, 2004; Hustinx and Durlinger-van der Horst, 2005). MindSonar is a process-oriented instrument. Process theories, on the other hand, attempt to provide a generalised explanation of processes and the behaviours these processes lead to, describing the major conditions necessary for explaining the process. They intend to describe how people think, what the processes are in their minds that induce their behaviour (Kispál-Vuitai, 2016). They contain constructs (Binning, 2016) that are not necessarily linked together in a coherent theory, but explain behaviour and allow to predict future behaviour. Process theories of motivation concern themselves with *how* people arrive at wanting something (Hollander et al., 2020). MindSonar measures 32 metaprograms divided into 13 inter-connected groups of thinking patterns (Table 1).The term 'metaprogram,' as well as the distinctions themselves, come from 'Neuro-Linguistic Programming (NLP)' (Bandler and Grinder, 1975, 1979). In NLP, sequences of inner sensory experience (images, sounds and feelings) were originally referred to as 'strategies' or 'programming'. Sometimes, patterns common to several strategies in the same person were noted during change of work. These distinctions were 'meta' to programming, hence the name 'metaprograms'. This is the qualitative, experiential basis of the metaprogram concepts, with NLP falling clearly in the 'process-oriented' category of theories.

Table 1

Characteristics of metaprograms in MindSonar

No.	Metaprograms	Characteristics of metaprograms
1	Proactive versus Reactive	Proactive = a preference for acting quickly and taking the initiative Reactive = a preference for waiting, considering, and reflecting
2	Towards versus Away From	Towards = a focus on achieving goals Away From = a focus on avoiding problems
3	Internal Reference versus External Reference	Internal Reference = using one's own standards in evaluations External Reference = using other people's standards in evaluations
4	Options versus Procedure	Options = a preference for many different possibilities Procedure = a preference for step-by-step planning
5	General versus Specific	General = a focus on the broad overview Specific = a focus on the small details
6	Matching versus Mismatching	Matching = a focus on what is good and correct Mismatching = a focus on what is bad and incorrect
7	Internal locus of control versus External locus of control	Internal locus of control = a focus on how someone influences their circumstances versus External locus of control = focus on how someone's circumstances influence them
8	Maintenance versus Development versus Change	Maintenance = a preference for things staying the same Development = a preference for gradual change Change = a preference for fast and radical change
9	People versus Activity versus Information	People = a focus on people and what moves them Activities = a focus on activities being done Information = a focus on information; facts and figures
10	Concept versus Structure versus Use	Concept = a focus on essentials and principles Structure = a focus on relationships between elements Use = a focus on practical applications
11	Together versus Proximity versus Solo	Together = a preference for working closely together with shared responsibility Proximity = a preference for mutual support with individual responsibility Solo = a preference for working alone
12	Past versus Present versus Future	Past = a focus on past events Present = a focus on the "here and now" Future = a focus on future events
13	Visual versus Auditory versus Kinaesthetic	Visual = a focus on images and movies Auditory = a focus on sounds and words Kinaesthetic = a focus on feelings and movement

Source: Hollander, J. (2014), MindSonar Certification Training Manual, IEP Institute for Eclectic Psychology, Staringstraat 1, 6511 PC Nijmegen, Netherlands.

The criteria indicate what someone finds important in a given context. In the TOTE-model (Test-Operate-Test-Exit) of goal-directed behaviour, the present situation is compared with a criterion in order to determine whether more operations (actions) are necessary (Miller et al., 1960). In terms of criteria, MindSonar asks the respondent to define:

- four criteria (four things they find important in the context in which their thinking style is measured for);
- a meta-criterion (what happens when the first four criteria are met);
- the opposites of all criteria (for instance, the opposite of 'Inspiration' might be, for a particular person, 'Dullness');
- hierarchy of the criteria (their order of importance).

The top two positions in the hierarchy of criteria were tested. The respondent is asked whether or not he or she would accept a *small loss* of criterion #2 in return for a *large gain* in criterion #1. For example: is he or she willing to accept *a little* loneliness in return for *a lot of* creativity? If the respondent does not accept the offer, they are directed back to their list of criteria and encouraged to make changes. Sometimes, criteria are components of or conditions for other criteria. MindSonar resolves this by encouraging respondents to combine criteria. For instance, if a respondent believes that they can only be creative together with other people, they cannot accept some loneliness to get a lot more creativity, because loneliness will in turn decrease their creativity. The respondent is then advised to combine 'creativity' and 'communication' into one new criterion ('creative communication' for instance).

Originally, MindSonar simply took stock of people's criteria by storing their verbal descriptions. This made it difficult, however, to compare criteria. For instance, three different people who all define 'honesty' as their number one criterion might enter 'lying', 'being unemotional' and 'financial theft' as their respective opposites. This illustrates the different meanings people attach to the same word. To solve this problem, the authors wanted to be able to attach numerical values to criteria, and chose the Graves (Spiral Dynamics) model. The respondent is asked to distribute balls over seven buckets to indicate what categories their criterion belongs to. Each bucket is labelled with two rotating key words representing Graves values. The more important criteria (higher in the hierarchy) carry more weight in the categorisation. This makes it possible to accurately compare and define criteria; also, the scores often offer the respondent's insight into the essence of their values.

The MindSonar test uses the Spiral Dynamics model to examine and classify values (Kompagne, 2008). This model is the effect of the research conducted by Graves, Cowan and Beck, which resulted in the concept of the development of biopsychosocial systems in adults, or the theory of levels of existence. This theory is based on the theory of systems and developmental psychology. It concentrates on the mature personality in action, together with its transformations, positive and negative traits. It integrates the roles of biological/genetic, psychological and sociocultural factors in creating levels of

existence experienced by individuals or groups (Graves, 1965; Graves, 1970; Graves, 2005; Cowan and Todorovic, 2000). Spiral Dynamics allows the understanding of internal mechanisms of decision-making and the shaping of individual behaviour through the identification of hidden values and patterns which correspond to certain levels of existence (Beck and Covan, 2006; Prinsloo, 2013).

Spiral Dynamics distinguishes between eight levels of existence, each with specific values and patterns of behaviour (Table 2), the MindSonar test does not measure 'beige' level values.

MindSonar works with a gradual responding system, meaning that the respondent does not have to make absolute yes-or-no choices. He or she indicates to what extent an alternative applies to him or herself. Cronbach's alpha reliability coefficient for this tool, measured for this research sample, is in the range from 0.7 to 0.85 for the measurements used in the test. Values and metaprograms may change depending on the situation in which a given person finds themselves, that is why MindSonar measures them in defined contexts. In the study in question, the context consisted of preferred values and modes of thinking and conduct in the workplace.

In order to test the hypotheses, statistical analyses were conducted with the use of the PQ Stat 1.6.8. program. The study applied descriptive statistics, normality tests (the Kolmogorov-Smirnov test), and tests of differences between groups (the Kruskal-Wallis ANOVA test) together with post-hoc tests (Dunn-Bonferroni).

Then, in the phase of comparisons of preferred values and metaprograms in individual generations in the workplace, normality tests were conducted first. The Kolmogorov-Smirnov test showed that the distribution of most of the tested variables, in one or in all generations, significantly deviates from normal distribution (asymptotic significance $p<0.05$). In addition, the three compared groups of respondents significantly differ in terms of their numbers. These arguments determined the selection of non-parametric tests for the analysis of differences between groups.

The theoretical concept applied in the MindSonar tool was selected by the authors of the study due to its applicability in management sciences, in particular in the field of examining levels of development, levels of awareness, methods of cognition, worldviews, methods of organising rules or logics of action among managers and leaders (Brown, 2011; Rooke and Torbert, 2005; Torbert, 2004; Aitken and Higgs, 2010; Anderson et al., 2006) or the effect of such levels on decision-making processes (Černý, 2017). The results of such studies show that the development stage of a given person affects what he or she notices or may be aware of, and thus what he or she may describe, articulate, think about, and what he or she may influence and change. Researchers believe that individuals at later stages of development demonstrate a more effective understanding and can influence other individuals from their own level or lower levels of development. This is due to the fact that they can act from earlier levels and assume developmental perspectives of those operating

Table 2

Characteristics of the levels of existence in Spiral Dynamics

Level	Dominant world-view, values (needs), behaviour	Manifestations in the organisation
1	2	3
1. Beige	**World-view: Instinctive.** Values/needs: Biological needs, survival. Behaviour: Few words, focus on actions which guarantee survival. Acquire food and roof over one's head. Instant fight/flee reactions of the reptilian brain. Instinct as the driving force.	Caring for basic needs of the body, intuition, instinct.
2. Purple	**World-view: Magical, spiritualistic, animistic.** Values/needs: Orientated at the community (clan), security and survival. Behaviour: Following orders of spiritual creatures in a mystical space. Showing loyalty to the elders, to customs, to the tribe. Cultivating sacred places, objects, rituals. Creating bonds in order to survive and find security. Living in an enchanted, magical village. Searching for harmony with the rhythms of nature.	People are strongly attached to their communities, groups. Paternalistic organisations, nepotism. Employees give their lives and souls to the organisation – like to a parent. Each one will sacrifice themselves without a question, when the group's survival depends on that. A change requires the approval of the "elders."
3. Red	**World-view: Impulsive**, mythical, the law of the strongest, "life is a jungle." A world of honour vendettas. Values/needs: Power, domination, respect, freedom, instant gratification. Behaviour: Strength and dominance are most important. Escape from the dominance of others. Avoiding shame, lack of the feeling of guilt, earning respect. Instant satisfaction of basic drives and satisfaction of whims. Fight for maintaining control at all cost. Lack of awareness of the consequences of one's actions.	Strong, directive leadership. Clear hierarchy, emphasis on power. People need the domination of a greater force to gratify them and keep their desires under control. Employees will put up with a lot, if their basic needs are regularly satisfied. Salary is an effect, everyone gets what they deserve. No one can be trusted. Rewarding for subordination, punishing for disloyalty. Strict control of information and choices.
4. Blue	**World-view: Absolutist**, **mythical, fundamentalist**. Order based on higher laws. Values/needs: Justice, unambiguous truth, stability, identity, predictability, clear rules. Behaviour: Sacrifice of an individual for group goals. Order and stability, people share the same beliefs, law and ethos. Each human has their own place in accordance with an overriding plan. Hierarchical structures, bureaucracy, commands and control.	Employees are cogs in the machine of the system, they perform roles which they are destined for. Hard work is the only right way to be appreciated and to keep one's job. People work best when they know how they should do something correctly, commanding attitude. Autho-rity is earned by others through proper rules, not through fear. Employees are loyal, if the organisation cares for their well-being. A change must be ordered by authorities, consistent with the rules.
5. Orange	**World-view: Modernistic**, **individualistic, meritocratic, materialistic, rationalistic.** Values/needs: Material success, status, self-fulfilment, competition, consumption, independence. Behaviour: Striving for success, achievements, autonomy and constant changes. Searching for a "good life" and wealth. Development through searching for the best solutions, experiments, advanced technologies. Education through personal experience of trials and errors. The world is rational and full of opportunities. Strive for independence and fulfilment of one's goals. What makes one win is competence, rivalry and strategic thinking.	Constant assessment of results. Competition increases productivity. Humans are driven by: development, success, growth, achievements, promotion and material rewards.

1	2	3
6. Green	**World-view: Pluralistic, post-modernist, ecological, egalitarian. Social liberalism.** Values/needs: Harmony, reciprocity, own truth, commonality, equality. Behaviour: Commonality and cooperation. Sharing the community's resources between everyone. Unanimous decision-making. Involvement in legitimate social projects. Emotionality, empathy, sensitivity to injustice.	Pressure on consensus and compromise. People want to feel accepted by others. Sharing and cooperation lead to better results than rivalry. All members of an organisation should have the opportunity to speak and be included in decision-making processes. The organisation is responsible for the well-being of its community. People are motivated by care for relationships, being a part of a community.
7. Yellow	**World-view: Post-rational, global, network-like. Deep ecology.** Values/needs: Independence, development, sense of meaning, health of the system. Behaviour: Striving for the synthesis of various world-views. Acting in one's own interest without harming others. Experiencing the life to the full, here and now. Striving for growth, evolution, flexibility, spontaneity, usefulness and open systems.	People achieve fulfilment through work which matches who they are by nature, concentration on functionality, talents and competences. Emphasis on access to knowledge, information and experience. People are motivated by learning and understanding, rather than salaries and punishments. People value freedom and valence of work, without instructions on how it should be done.
8. Turquoise	**World-view: Holistic, integral, transpersonal.** Values/needs: planetary community, altruism, transcendence. Behaviour: Using all previous levels, expansion, multi-dimensionality, systemic awareness. Great emphasis on the welfare of all living creatures. Expanding consciousness with all known methods. The sense of "Self" as a part of a larger, conscious whole, a part of global networks. Striving for a minimalistic lifestyle – "small is beautiful."	Holistic view on reality. Ability to combine systems and paradigms. Simultaneous functioning on the rational and on the intuitive level. Lack of egotism. Spiritual bonds push people and the organisation towards each other. Work must have meaning, must serve the welfare of all. The world is a unity of balanced, interacting forces. Organisations are responsible for their effect on human lives and the environment.

Source: Beck, D. E. and Cowan, C. C. (2006), Spiral dynamics: Mastering values, leadership and change, Malden: Blackwell Publishing.

Graves C.W. (2005), The newer- ending quest: Clare W. Graves explores human nature, ECLAT Publishing, Santa Barbara.

within earlier stages. It can be assumed, therefore, that if there are differences between the participants belonging to a given generation, it will be possible to identify them in detail using MindSonar.

4. Research findings and discussion

4.1. Generational differences in values

An analysis of preferred values in the workplace conducted in accordance with the Graves model (Table 1) shows that values from the "yellow," "blue" and "green" level are dominant and equally important for each generation. Differences between generations are negligible. Generations X and Z show equal preference for values

from the "yellow," "green" and "blue" level. In generation Y, values from the "blue" level are chosen more often than the values from the "green" level, while the values from the "yellow" level appear more often than in generations X and Z. Of all the tested generations, generation X chooses values from the "orange" level least often, however preferring values from the "turquoise" level more often than other generations (Table 3).

Table 3

Differences in values across generations

Levels of existence and their corresponding values	Descriptive statistics						H	P	Significant differences
	X		Y		Z				
	M	SD	M	SD	M	SD			
Purple	1.45	1.14	1.42	1.11	1.22	1.04	3.19	0.20	
Red	1.32	0.92	1.27	0.94	1.34	0.89	1.00	0.60	
Blue	1.78	0.81	1.81	1.08	1.76	1.14	0.71	0.70	
Orange	1.01	0.99	1.36	1.02	1.50	1.14	5.89	0.05**	X<Z
Green	1.83	1.32	1.63	0.96	1.78	1.04	1.78	0.41	
Yellow	1.87	1.19	2.19	1.35	2.14	1.12	1.46	0.48	
Turquoise	0.70	0.89	0.32	0.60	0.24	0.55	8.05	0.02*	X>Y, X>Z

Notes: *Significant difference at the level of $p<0.05$; ** Borderline significant.

Source: own elaboration.

A detailed analysis of the results demonstrates a significant difference in the frequency of the occurrence of values from the "turquoise" level among the youngest and the oldest respondents. These values are chosen significantly more often by generation X than generation Z. There are differences between generations also in terms of the frequency of selecting values attributed to the "orange" level. The paired difference test (post-hoc) showed that a significant difference applies to the X-Z generational pair. Generation Z chooses values from this level significantly more often than generation X. The differences in choosing values from other levels are not statistically significant.

4.2. Generational differences in metaprograms

Analysing the choices of generation X (Table 4), it can be noted that they achieve a high result for the "matching," "internal locus of control," "general" and "towards" metaprograms (Table 2). They also stand out regarding the mode of thinking cha-

Table 4

Differences in choosing metaprograms across generations

Metaprograms	Descriptive statistics						H	P	Significant differences
	X		Y		Z				
	M	SD	M	SD	M	SD			
Proactive	5.03	1.87	4.75	1.98	4.63	1.71	0.98	0.61	
Reactive	4.97	1.87	5.25	1.98	5.37	1.71	0.98	0.61	
Towards	6.69	1.46	6.58	1.50	6.40	1.66	0.87	0.65	
Away from	3.32	1.46	3.43	1.50	3.61	1.66	0.89	0.64	
Internal reference	5.50	.08	5.36	1.77	4.89	1.88	7.32	0.03*	Y>Z
External reference	4.50	2.08	4.64	1.77	5.11	1.88	7.32	0.03*	Z>Y
Options	5.86	1.65	5.47	1.74	5.49	1.55	2.01	0.37	
Procedure	4.14	1.65	5.43	1.74	4.51	1.55	2.01	0.37	
Matching	7.19	1.26	6.53	1.63	6.23	1.49	9.92	0.01*	X>Z
Mismatching	2.81	1.26	3.47	1.63	3.78	1.49	10.10	0.01*	Z>X
Internal locus of control	7.08	0.94	6.63	1.16	6.43	1.36	7.18	0.03*	X>Z
External locus of control	2.92	0.94	3.37	1.16	3.57	1.36	7.18	0.03*	Z>X
General	6.92	1.96	4.80	2.14	5.38	1.99	25.72	<0,001***	X>Y, X>Z, Z>Y
Specific	3.08	1.54	5.20	2.14	4.62	1.99	25.73	<0,001***	Y>X, Y>Z, X<Z
Maintenance	2.35	1.24	2.25	1.49	2.34	1.49	0.15	0.93	
Development	5.28	1.09	5.29	1.45	5.34	1.49	0.82	0.66	
Change	2.37	0.80	2.46	1.33	2.32	1.36	0.52	0.77	
People	3.14	0.90	2.95	1.`123436	3.13	1.36	1.67	0.43	
Activities	3.39	0.79	3.48	1.11	3.41	1.99	0.38	0.83	
Information	3.48	1.11	3.58	1.33	3.48	1.99	1.13	0.57	
Concept	2.77	1.01	2.89	1.16	2.97	1.13	1.06	0.59	
Structure	3.18	1.07	3.10	1.07	3.07	1.01	0.41	0.81	
Use	4.04	1.38	4.02	1.19	3.96	1.11	0.03	0.98	
Together	3.21	1.12	2.77	1.46	2.79	1.6	2.75	0.25	
Proximity	4.22	1.13	4.75	1.35	4.52	1.37	5.90	0.05	
Solo	2.57	0.82	2.48	1.49	2.69	2.03	0.39	0.82	
Past	1.37	1.59	1.61	1.03	1.58	1.04	0.89	0.64	
Present	4.17	1.66	4.39	1.49	4.21	1.27	0.96	0.62	
Future	4.46	0.85	4.01	1.45	4.22	1.34	3.38	0.18	
Visual	3.54	1.17	4.02	1.38	4.33	1.31	11.45	0,003**	Z>X, Z>Y
Auditory	2.64	1.58	2.49	1.19	2.35	1.21	2.28	0.32	
Kinaesthetic	3.82	1.14	3.49	1.50	3.33	1.24	3.70	0.16	

Notes: *Significant differences at the level of $p<0.05$; **Significant differences at the level of $p<0.01$; ***Significant differences at the level of $p<0.001$.

Source: own elaboration.

racteristic for the "internal reference" metaprogram. They achieve low results for the following metaprograms: "mismatching," "external locus of control," "specific," "auditory" and "external reference." The respondents from generation X do not tend to take action in line with the "past" metaprogram, instead, they opt for "future" and, to an almost identical degree, for "present."

The results of generation X are similar to the results of generation Y (Table 4). Its representatives prefer the following programs: "matching," "internal locus of control" and "towards," "internal reference." The differences in the results between metaprograms are slightly less distinct than those in generation X. Unlike in generation X, generation Y demonstrates a higher propensity for the "specific" rather than the "general" metaprogram. Furthermore, in action, it is more oriented towards the "present" rather than the "future" metaprogram, and, similarly to generation X, it is not oriented at "past."

Generation Z also shows tendencies similar to generations X and Y (Table 4),still, three metaprograms prevail: "towards," "matching," "internal locus of control." The preferences are slightly less visible than in generations X and Y. Just like generation X, generation Z shows preference for the "future" and "present" metaprograms, and, similarly to generations X and Y, it is not "past"-oriented. Unlike generations X and Y, generation Z prefers an "external reference"-oriented mode of thinking.

A detailed analysis of the results in the scope of metaprograms shows that generation X significantly more often than generation Z chooses the "matching" metaprogram. In turn, generation Z significantly more often than generation X chooses the "mismatching" metaprogram. The option "internal locus of control" is also significantly more often chosen in generation X than in generation Z. A characteristic trait of generation Y is that its orientation towards the "specific" metaprogram is significantly greater than in other groups. Generation Y more often than generation Z chooses the "internal reference" metaprogram. In generation Z, there are significantly more "visual" choices than in generation X.

4.3. Discussion

In the workplace, all generations demonstrate a similar preference for values corresponding to the levels of existence determined by Spiral Dynamics as "yellow," "green" and "blue." An analysis of the metaprogram tests suggests that following metaprograms achieve high results in all generations: "towards," "matching," and "internal locus of control." Other results are more or less counterbalanced, which means that the respondents do not show any clear preferences in relation to these metaprograms in the workplace.

Values from the "yellow" level are most significant in all generations. This means that what counts more for them is learning and understanding rather than salaries and punishments. They want to work in a way that will provide them with self-fulfilment. They value freedom and spontaneity at work, without instructions on how it should

be done, and expect that the organisation will benefit from their diverse competences, skills and unique talents. Their development is driven by the will to explore, gain knowledge, create flexible structures and integrate a multitude of solutions and approaches. The obtained test results differ from the results presented by other researchers (Jurkiewicz and Brown, 1998; D'Amato and Herzfeldt, 2008; Lester et al., 2012; Roongrerngsuke and Liefooghe, 2013), which indicate that younger generations put more emphasis on life-long learning and personal development than older generations.

The next result in all generations was obtained by the values from the "green" level, which means a preference for values resulting from commonality. For persons preferring this level of values, sharing and participation are better than competition. They value the commitment and participation of all employees in decision-making processes, because diversity enriches the results. Contact with other people, sharing their experiences and maintaining harmonious relationships are important to them.

The obtained result is slightly different from the results presented by other researchers. In some of them, it was concluded that older generations are more willing to co-operate with others and prefer team-work more than younger generations, whereas others suggest the increased significance of team-work and social activities in younger generations (Jurkiewicz and Brown, 1998; Yrle et al., 2005; Gursoy et al., 2008; Cogin, 2012; Lub et al., 2012).

The third result in generations was obtained by values from the "blue" level. This means that people need structures and order, they have to be informed in order to be able to do things properly. They work for the right cause and in order to retain their jobs; work provides them with a sense of security.

The obtained results are in line with the results of the research of Brunetto et al. (2012), which demonstrates that all generations put emphasis on strong supervisory relationships. Leschinsky and Michael (2004), Wong et al. (2008), Takase et al. (2009), Lub et al. (2012), Hansen and Leuty (2012) demonstrated that each generation deems job security important.

A detailed analysis of the results reveals a significant difference in the values from the "turquoise" level among the youngest and the oldest respondents (Table 3). These values are chosen significantly more often by generation X than generation Z. Values from this level are connected with: responsibility for human life and Earth as a planet, holism, balance and the integration of various approaches, which means that these persons wish to feel responsible for a greater whole. They value a broad and integrated image of reality, and think highly of everything that is ecological. They prefer minimalism. Work must have meaning and must serve the welfare of all. The significance of having useful work is also confirmed by Hajdu and Sik (2018), who concluded that it increases with age in European and Euro-Atlantic countries.

Differences between generations can also be found in values attributed to the "orange"" level. Generation Z chooses values from this level significantly more often than generation X. This means that they are more likely to prefer values

connected with rivalry. Other important values for this level include: success, achievements, wealth, effectiveness, progress and influence. Similar results were obtained by Cennamo and Gardner (2008), who conclude that younger generations put more emphasis on "status" than older ones. According to Smola and Sutton (2002), Leschinsky and Michael (2004), Wong et al. (2008), younger generations pay more attention to opportunities for promotion and want to be promoted more quickly than older generations, whereas other studies indicate that each generation considers competitive remuneration important in their work (Jurkiewicz and Brown, 1998; Leschinsky and Michael, 2004; Takase et al., 2009; Lub et al., 2012; Roongrerngsuke and Liefooghe, 2013).

The measurement of differences in values between generations in the workplace was conducted with the use of the Spiral Dynamics theoretical model, in which individual values are attributed to their corresponding levels of existence. In turn, the level of existence on which a given individual operates is the effect of the interaction of biological, psychological and social factors. Therefore, the values change over time, and the dynamics of these changes can be different for individual persons or groups. Research shows that values at work are also unstable and are subject to change during an individual's transition from puberty to adulthood (Jin and Rounds, 2012).

Hence the generational differences observed in the results of conducted studies may be more connected with the age of the respondents or the stage of their life (Wong et al., 2008) than with their generational identity. The representatives of generation X are most likely at different career stages, they have a well-established professional standing, they have achieved financial stability, they have invested more in their work, organisation, career than persons who are only starting work (Ng and Feldman, 2008). Thus, it is completely justified to expect the younger employees to be more interested in the development of their careers and being promoted, building economic security, purchasing houses and starting families (i.e. values from the "orange" level), while the older persons will put less emphasis on such values. Those results partly confirm the outcome of the research of Twenge et al. (2012), that younger generations put more emphasis on external values, e.g. money, fame (values from the "orange" level), while civic participation, e.g. through interest in social problems (values from the "turquoise" level), decreases with age.

The high result for the "towards" metaprogram and the low result for the "away from" metaprogram means that the representatives of the studied generations focus on achieving goals. Their strong motivation is the mere completion of an activity and they usually do not pay attention to what may go wrong. The high result for the "matching" metaprogram and the low result for the "mismatching" metaprogram means that they strongly focus on what is good, correct and which they agree with in their actions.

The studied representatives of generations X, Y, Z achieve a high result for the "internal locus of control" metaprogram and a low result for the "external locus of

control" metaprogram. This means that they strongly believe that they have influence on their surroundings and can shape them. They demonstrate high optimism regarding their own agency and effect on the change of their surroundings. The respondents also prefer the "proximity" metaprogram, which means that they prefer remaining in stable though relaxed cooperation with others, allowing mutual help and support, should such need arise. They clearly opt for project cooperation with others, without sharing responsibility. They want to be sure what each member of the group will contribute, and they want each person to be responsible for his or her work.

They quite strongly focus on the here and now, demonstrated by the high result of the "present" metaprogram. They concentrate a lot on a broad picture of a given situation and usually do not pay attention to details (the "general" metaprogram). They focus on the "future" and are interested in what will happen. They show a talent for predicting the future, and assess the present based on its effect on the future. The strength of this metaprogram is the ability to create plans.

The results of the tests indicate that a greater similarity may be observed between generations X and Y, while generation Z differs from both of them (Table 4).

Generation Z significantly more often than generation X chooses the "mismatching" metaprogram. Persons thinking in terms of "differentiating" concentrate on what is not going well, what is not correct, what fails to meet their criteria. The "external locus of control" answer is significantly more often chosen in generation Z than in generation X. Persons with the "external locus of control" metaprogram feel that matters depend on external circumstances. They perceive their own behaviour as an effect rather than a cause.

Generation Z also more often than generation Y chooses the "external reference" metaprogram. This means that they base their judgement on what others think is important. They find motivation outside, beyond themselves. The criteria of others determine their actions. In order to function, they need feedback and other opinions. They easily accept decisions made by others, and cope well in situations which require adhering to instructions, customer support and service. Some studies of other researchers partly confirm the obtained results. For instance, Gursoy et al. (2013) found that younger generations express greater needs in the scope of providing advice, guidelines and direction from their leaders/ mentors at work. Other studies found that generation Y has a greater need than older generations to be acknowledged, appreciated and provided with instant feedback (Yu and Miller, 2003; Busch et al., 2008; Lester et al., 2012; Gursoy et al., 2013; Mencl and Lester, 2014).

Indeed, in generation Z there are significantly more persons with the "visual" metaprogram than in generation X,they prefer the sense of sight, perceive the world in images, photographs and films. Their manner of thinking has a shape, colour, depth, etc. They often speak quickly and use high tones. Their ideas, memories and concepts have a form of mental images. They learn through visual information and first need to see things in order to be able to understand them and start taking action.

The generational differences in the area of metaprograms, similarly to differences in values, as observed in the results of conducted tests, can be explained through factors which are better backed by theories and data than arguments of generational identity. These factors are: individual traits, historical period in which persons are assessed (Costanza et al. 2012), external environmental factors, such as technology (Sackett, 2002), and developmental changes which occur in humans over time (Elder, 1994). The representatives of generation Z are only just entering adulthood and the labour market, they lack professional experience and consequently they demonstrate a greater externalisation of the locus of control (Twenge et al., 2008), putting more emphasis on such aspects of work as mentoring, training (Loughlin and Barling, 2001; Lyons, 2004) and development (Iorgulescu, 2016). These results may explain the increased significance of the "external locus of control" or "external reference" metaprogram in younger generations.

The basic differences between generations as diagnosed by Twenge (2010), which concern, among others, the manner of communication, using modern technologies or work methods, indicate a connection with progressing globalisation, development of the Internet and Internet-based technologies (Lyons, 2004). These studies show that the most adjusted generation in terms of technological skills is generation Z. Since such skills are connected with a specific channel used for obtaining information and for communication, this may explain the higher proportion of the "visual" metaprogram in generation Z in relation to other generations.

As argued by Costanza and Finkelstein (2015), notwithstanding whether generational differences are real or not, people believe they exist. Therefore, the obtained results may prove useful in taking actions directed at reducing conflicts or misunderstandings stemming from generational stereotypes. In organisations with employees of diverse ages, the obtained study results may help managers focus on real, diagnosed differences occurring in the staff. In spite of only a few differences between the examined generations in terms of values and patterns of thinking in the workplace, the obtained results may be used by organisations to determine the communication of priorities for present and future employees with a view to achieving a high level of matching of mutual expectations. The development and communication of organisational priorities in combination with appropriate techniques of recruitment and assessment may help HR specialists reduce staff turnover and costs of recruitment.

One of the limitations of this study was the use of cross-sectional data and the division of respondents into groups by age. This made it difficult to determine whether differences between generations result from age, stage of career, stage of life, or from actual generational differences. In order to separate generational differences from age and other variables, longitudinal studies should be conducted. Real generational differences may be identified only through a study of groups over time. Another limitation was the research sample, which is homogeneous in terms of

education and nationality of the respondents. Further exploration of the tested variables on a diverse sample in terms of education or social status of the respondents could be recommended.

Conclusion

An analysis of the test results leads to a conclusion that neither of the above hypotheses may be unequivocally accepted or rejected. The results presented by the researchers indicate that the generations in question demonstrate numerous similarities, and statistically significant differences between the generations in the scope of tested variables are sparse (for 2 out of the 7 levels of existence and their corresponding values, and in 9 out of the 32 measured metaprograms). The test results also prove that a greater similarity of preferred values and patterns of thinking in the workplace can be observed between generations X and Y, while generation Z differs from them both.

A detailed analysis of the results indicates that the group of representatives of generations X, Y, Z studied by the researchers have numerous common features in preferred values and patterns of thinking in the workplace. The values which are most important to them in the workplace are: freedom, personal development, independence, creativity, being inquisitive, structure and order, duty, discipline, reliability and control. They prefer being independent, favour critical thinking and a will to learn as well as maintaining order and performing duties, compliance with rules and behaving in line with regulations. Equally important to them is social contact as well as consensus and solidarity, warm relationships, being helpful, having empathy, sharing ideals and solidarity with others.

Generation X is orientated towards the future and the future situation more than other generations. Generation X, as the oldest one, tends the most to choose values resulting from responsibility for the fate of future generations, namely values resulting from the "turquoise" level of existence.

Generation Z, socialised with the electronic media of transmitting and receiving information, demonstrates a higher propensity for perceiving reality through the sense of sight more than generations X and Y. As the age of the respondents increases, so does the sense of agency ("internal locus of control"). The youngest generation is the most other-directed and has the least developed own standards of judgement, which may make its representatives seem more reactive in action and may need permanent feedback.

The youngest generation shows the greatest propensity for preferring the following values: results, development of competences, effectiveness, success, effects, progress, competition, being promoted as well as tangible benefits (values characteristic for the "orange" level of existence).

To sum up, the results of this study indicate the existence of both similarities and differences between the generations in terms of the preferred values and patterns of thinking at work. These results partly confirm those found in the literature. Although

the identified differences are statistically significant, they are minimal and occur in the case of only a few of the tested variables. Most of the observed differences can be explained by differences in the stage of life of the respondents rather than by generational differences. Furthermore, a detailed analysis of the collected data indicates the existence of greater differences between individuals in a single generation than between various generations. From this perspective, it appears significant to extend the research by taking into account the personality-related variable.

References

Aitken, P., and Higgs, M. (2010). Developing Change Leaders. The principles and practices of change leadership development. Oxford: Elsevier.

Anderson, B., Stuart, J., and Klein, E. (2006). The Leadership Circle: Bringing Spiritual Intelligence to Work, Founder & CEO, The Leadership Circle, www.theleadershipcircle.com. In Wilber, K., *Integralna teoria wszystkiego. Wizja dla biznesu, polityki nauki i duchowości*. Poznań: Zysk i S-ka.

Barker, E. H., and Torbert, W. R. (2011). Generating and measuring practical differences in leadership performance at post-conventional action-logics. In Pfaffenberger, A., Marko, P. and Combs, A. (eds.), *The post conventional personality: Assessing, researching, and theorizing higher development*. Albany: Suny Press, 39-95.

Bandler, R., and Grinder, J. (1975). The Structure of Magic. A Book about Language and Therapy. Palo Alto: Science & Behavior Books.

Bandler, R., and Grinder, J. (1979). Frogs into princes: Neuro Linguistic Programming. Santa Cruz: Real People Press.

Barron, P., Leask, A., and Fyall, A. (2014). Engaging the multigenerational workforce in tourism and hospitality. *Tourism Review*, 69, 245–263.

Beck, D. E., and Cowan, C. C. (2006). Spiral dynamics: Mastering values, leadership and change. Malden: Blackwell Publishing.

Benson, J., and Brown, M. (2011). Generations at work: are there differences and do they matter? *The International Journal of Human Resource Management*, 22 (9), 1843–1865.

Berkup, S. B. (2014). Working with generations X and Y in generation Z period: Management of different generations in business life. *Mediterranean Journal of Social Sciences*, 5(19), 218-229.

Beutell, N. J., and Wittig-Berman, U. (2008). Work-family conflict and work-family synergy for generation X, baby boomers, and matures: Generational differences, predictors, and satisfaction outcomes. *Journal of Managerial Psychology*, 23(5), 507–523.

Binning J. F., Construct (2016). Encyclopedia Britannica, [online], http://www.britannica.com/science/construct, accessed 16.03.2022.

Bolstad, R. (2002). Resolve: A New Model of Therapy. Carmarthen: Crown House Publishing.

Brown, N. (2003). A comparison of the dominant meta programme patterns in accounting undergraduate students and accounting lecturers at a UK business school. *Accounting Education*, 12(2), 159–175.

Brown, N. (2004). What makes a good educator? The relevance of meta programmes. *Assessment & Evaluation in Higher Education*, 29(5), 515–533.

Brown, N. (2006). The development of a questionnaire assessing metacognitive patterns of students majoring in accounting in higher education. *Accounting Education: An International Journal*, 15(3), 301–323.

Brown, B. C. (2011). Conscious leadership for sustainability: How leaders with late-stage action logic design and engage in sustainability initiatives [online], http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.6432&rep=rep 1& type=pdf, accessed 5.01. 2021.

Brunetto, Y., Farr-Wharton, R., and Shacklock, K. (2012). Communication, training, wellbeing, and commitment across nurse generations. *Nursing Outlook*, 60(1), 7–15.

Busch, P., Venkitachalam, K., and Richards, D. (2008). Generational differences in soft knowledge situations: Status, need for recognition, workplace commitment and idealism. *Knowledge and Process Management*, 15(1), 45–58.

Cennamo, L., and Gardner, D. (2008). Generational differences in work values, outcomes, and person-organisation fit. *Journal of Managerial Psychology*, 29(8), 891–906.

Černý, J. (2017). The development of the importance of internal values of Gen Y employees, The role of human resources development strategy in achieving corporate social responsibility at the SMEs level. *Conference Paper*, PERFnet, Brno, Stávková, J. (ed.), Proceedings.

Chen, P., and Choi, Y. (2008). Generational differences in work values: A study of hospitality management. *International Journal of Contemporary Hospitality Management*, 20, 595–615, 2008.

Cogin, J. (2012). Are generational differences in work values fact or fiction? Multi-country evidence and implications. *The International Journal of Human Resource Management*, 23(11), 2268–2294.

Costanza, D. P., and Finkelstein, L. M. (2015). Generationally based differences in the workplace: Is there a there there? *Industrial and Organizational Psychology*, 8(3), 308–323.

Costanza, D. P., Badger, J. M., Fraser, R. L., Severt, J. B., and Gade, P. A. (2012). Generational differences in work-related attitudes: A meta-analysis. *Journal of Business and Psychology*, 27, 375–394.

Cowan, C. C., and Todorovic, N. (2000). Spiral dynamics: The layers of human values in strategy. *Strategy and Leadership*, 28(1), 4–11.

Cucina, J., Byle, K., Martin, N., Peyton, S., and Gast, I. (2018). Generational differences in workplace attitudes and job satisfaction: Lack of sizable differences across cohorts. *Journal of Managerial Psychology*, 33(3), 246–264.

Culpin, B., Millar, C., and Peters, K. (2015). Multi-generational frames of reference: Managerial challenges of four social generations in the organisation. *Journal of Managerial Psychology*, 30(1), 1–7.

D'Amato, A., and Herzfeldt, R. (2008). Learning orientation, organizational commitment and talent retention across generations: A study of European managers. *Journal of Managerial Psychology*, 23(8), 929–953.

Davis, J., Pawlowski, S., and Houston, A. (2006). Work commitments of baby boomers and gen-xers in the IT profession: Generational differences or myth? *Journal of Computer Information Systems*, 40, 43–50.

Dries, N., Pepermans, R., and de Kerpel, E. (2008). Exploring four generations' beliefs about career – is 'satisfied' the new 'successful'? *Journal of Managerial Psychology*, 23(8), 907–928.

Egri, C. P., and Ralston, D. A. (2004). Generation cohorts and personal values: A comparison of China and the US. *Organization Science*, 15(2), 210–220.

Elder, G. H., Jr. (1994). Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly,* 57, 4–15.

Foster, K. (2013). Generation and discourse in working life stories, *The British Journal of Sociology*, 64, 195–215.

Giancola, F. (2006). The generation gap: More myth than reality. *Human Resource Planning*, 29(4), 32–37.

Glass, A. (2007). Understanding generational differences for competitive success. *Industrial and Commercial Training*, 39(2), 98–103.

Graves, C. W. (1965). Value systems and their relation to managerial controls and organisational viability, Paper presented to the College of Management Philosophy, San Francisco.

Graves, C. W. (1970). Levels of existence: an open systems theory of values. *Journal of Humanistic Psychology*, 10(2), 131–155.

Graves, C. W. (2005). The newer-ending quest: Clare W. Graves explores human nature. Santa Barbara: ECLAT Publishing.

Guerin-Marion, C., Manion, I., and Parsons, H. (2018). Leading an intergenerational workforce: An integrative conceptual framework. *International Journal of Public Leadership*, 14(1), 48–58.

Gursoy, D., Chi, C. G., and Karadag, E. (2013). Generational differences in work values and attitudes among frontline and service contact employees. *International Journal of Hospitality Management*, 32, 40–48.

Gursoy, D., Maier, T., and Chi, C. G. (2008). Generational differences: an examination of work values and generational gaps in the hospitality workforce. *International Journal of Hospitality Management*, 27(3), 448–458.

Hajdu, G., and Sik, E. (2018). Age, Period, and Cohort Differences in Work Centrality and Work Values. *Societies,* 8(11), 6–33.

Hansen, J. C., and Leuty, M. E. (2012). Work Values Across Generations. *Journal of Career Assessment*, 20(1), 34–52.

Hayes, J. B., Parks, C., McNeilly, S., and Johnson, P. (2018). Boomers to millennials: Generational stereotypes at work in academic librarianship. *Journal of Academic Librarianship,* 44(6), 845–853.

Holian, R. (2015). Work, career, age and life-stage: assumptions and preferences of a multigenerational workforce. *Labour & Industry-A Journal of the Social And Economic Relations of Work*, 25(4), 278–292.

Hollander, J. (2014). MindSonar Certification Training Manual, IEP Institute for Eclectic Psychology, Staringstraat 1, 6511 PC Nijmegen, Netherlands.

Hollander, J., Minne, C., Leal, J., and Arguello, C. (2020). Mindsonar: measuring mindsets. A complementary, process oriented type of measurement, unpublished material.

Hustinx, G., and Durlinger-van der Horst, A. (2005). Voorbij je eigen wijze: effectief communiceren met metaprogramma's in professionele relaties, Nelissen, Soest.

Iorgulescu, M.C. (2016). Generation Z and its perception of work. *Cross Cultural Management Journal*, 18(1), 47–54.

Jin, J., and Rounds, J. (2012). Stability and change in work values: A meta-analysis of longitudinal studies. *Journal of Vocational Behaviour*, 80, 326–339.

Jurkiewicz, C. L., and Brown, R. G. (1998). Generational comparisons of public employee motivation. *Review of public personnel administration*, 18(4), 18–37.

Kindrick Patterson, C. (2007). The impact of generational diversity in the workplace. *Diversity Factor*, 15(3), 17–22.

Kispál-Vitai, Z. (2016). Comparative analysis of motivation theories. *International Journal of Engineering and Management Sciences (IJEMS)*, 1(1).

Kompagne, J. (2008). Samenhangtussen MPA MindSonar® en Spiral Dynamics, Artikel in het kader van MPA-certificering, Instituut voor Eclectische Psychologie, Nijmegen.

Kowske, B. J., Rasch, R., and Wiley, J. (2010). Millennials' (lack of) attitude problem: An empirical examination of generation effects on work attitudes. *Journal of Business and Psychology*, 25(2), 265–279.

Kupperschmidt, B. R. (2000). Multigeneration employees: Strategies for effective management. *The Health Care Manager*, 19, 65–76.

Leschinsky, R. M., and Michael, J. H. (2004). Motivators and desired company values of wood products industry employees: investigating generational differences. *Forest Products Journal*, 54(1), 34–39.

Lester, S. W., Standifer, R. L., Schultz, N. J., and Windsor, J. M. (2012). Actual versus perceived generational differences at work: An empirical examination. *Journal of Leadership & Organizational Studies*, 19, 341–354.

Loughlin, C., and Barling, J. (2001). Young workers' values, attitudes, and behaviours. *Journal of Occupational and Organizational Psychology*, 74(4), 543–58.

Lub, X., Bijvank, M. N., Bal, P. M., Blomme, R., and Schalk, R. (2012). Different or alike?: Exploring the psychological contract and commitment of different generations of hospitality workers. *International Journal of Contemporary Hospitality Management*, 24(4), 553–573.

Lyons, S. (2004). An exploration of generational values in life and at work, Dissertation Abstracts International, 3462A (UMI No. AATNQ94206).

Lyons, S., and Kuron, L. (2014). Generational differences in the workplace: A review of the evidence and directions for future research. *Journal of Organizational Behaviour*, 35(1), 139–157.

Macky, K., Gardner, D., and Forsyth, S. (2008). Generational differences at work: introduction and overview. *Journal of Managerial Psychology*, 29(8), 857–861.

Mencl, J., and Lester, S. W. (2014). More Alike Than Different: What Generations Value and How the Values Affect Employee Workplace Perceptions. *Journal of Leadership & Organizational Studies*, 21(3), 257–272.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). Plans and the Structure of Behavior. Henry Holt & Co.

Ng, T. W. H., and Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology*, 93, 392–422.

Park, J., and Gursoy, D. (2012). Generation effects on work engagement among US hotel employees. *International Journal of Hospitality Management*, 31(4), 1195–1202.

Parry, E., and Urwin, P. (2011). Generational differences in work values: A review of theory and evidence. *International Journal of Management Reviews*, 13, 79–96.

Patterson, Ch. (2007). The impact of generational diversity in the workplace, *Diversity Factor,* 15, 17–22.

Posthuma, R. A., and Campion, M. A. (2009). Age stereotypes in the workplace: Common stereotypes, moderators, and future research directions. *Journal of Management*, 35(1), 158–188.

Prinsloo, M. (2013). Consciousness models in action: Comparisons. *Integral Leadership Review*, 12(3), available at: http://integralleadershipreview.com/7166-consciousness-models-in-action-comparisons/ (accessed 17 April 2019).

Rooke, D., and Torbert, W. R. (2005). Seven transformations of leadership. *Harvard Business Review*, 83(4), 66–76.

Roongrerngsuke, S., and Liefooghe, A. (2013). Attracting gold-collar workers: Comparing organizational attractiveness and work-related values across generations in China, India and Thailand. *Asia Pacific Business Review*, 19(3), 337–355.

Schawbel, D. (2014). Generation Z. Understanding the Next Generation of Worker, available at: http://millennialbranding.com/2014/geny-genz-global-workplace-expectations-study (accessed 20 May 2019).

Singh, A. (2014). Challenges and Issues of Generation Z. *IOSR Journal of Business and Management*, 16(7), 59–63.

Smola, K. W., and Sutton, C. D. (2002). Generational differences: Revisiting generational work values for the new millennium. *Journal of Organizational Behaviour*, 23, SP1, 363–382.

Takase, M., Oba, K., and Yamashita, N. (2009). Generational differences in factors influencing job turnover among Japanese nurses: An exploratory comparative design. *International Journal of Nursing Studies*, 46(7), 957–67.

Torbert, W. R., Fisher, D, and Rooke, D. (2004). Action inquiry: The secret of timely and transforming leadership. San Francisco: Berrett-Koehler.

Twenge, J. M., and Campbell, S. T. (2008). Generational differences in psychological traits and their impact on the workplace. *Journal of Managerial Psychology*, 29(8), 862–877.

Twenge, J. M. (2010). A Review of the Empirical Evidence on Generational Differences in Work Attitudes. *Journal of Business and Psychology*, 25(2), 201–210.

Twenge, J. M., Campbell, W. K., and Freeman, E. C. (2012). Generational differences in young adults' life goals, concern for others, and civic orientation, 1966-2009. *Journal of Personality and Social Psychology*, 102(5), 1045–1062.

Van Rossem, A. H. D. (2019). Generations as social categories: An exploratory cognitive study of generational identity and generational stereotypes in a multigenerational workforce. *Journal of Organizational Behavior,* 40(4), 434–455.

Wils, T., Saba, T., Waxin, M.-F., and Labelle, C. (2011). Intergenerational and Intercultural Differences in Work Values in Quebec and the United Arab Emirates. *Relations Industrielles*, 66(3), 445–469.

Wong, M., Gardiner, E., Lang, W., and Leah, C. (2008). Generational differences in personality and motivation. Do they exist and what are the implications of the workplace? *Journal of Management Psychology*, 23(8), 878–890.

Yrle, A. C., Hartman, S. J., and Payne, D. M. (2005). Generation X: Acceptance of others and teamwork implications. *Team Performance Management*, 11(5/6), 188–199.

Yu, H. C., and Miller, P. (2003). The generation gap and cultural influence-a Taiwan empirical investigation. *Cross Cultural Management: An International Journal*, 10(3), 23–41.

Received: August 2021, revised: March 2022

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Marcin Salamaga**

Study on the influence of foreign direct investment on innovations in enterprises in Poland using the ECM panel model

In many theories of economics and empirical research, foreign direct investment (FDI) is perceived as a potential technology transfer channel bringing tangible benefits to FDI exporting countries and host countries. In light of some theories, such as Vernon's product life cycle theory or Dunning's pull factor theory, and the results of empirical research, the impact of inward FDI flows inhibits the development of innovation or has a neutral effect on innovation in the host country. In the era of the growing internationalisation of enterprises, and the search for opportunities for enterprises to compete on the domestic and foreign markets, innovation is becoming one of the most important elements of building a competitive advantage. Innovation and new technologies are also of fundamental importance for Polish enterprises that want to compete effectively. The question is, to what extent FDI in Poland favours this process, and to what extent limits it? In the empirical studies conducted so far for the Polish economy in the field of the relationship between the inflow of FDI and innovation, there are not many models that would allow to describe the cause-effect relationship between the discussed categories. The authors of this article attempt to fill this research gap. The main goal was to examine the directions and intensity of the impact of foreign direct investment on the level of innovation of enterprises in the main sectors of the economy in Poland. The study used, among others, panel data models with error correction mechanism (ECM) and the Granger causality test. The results confirmed the positive impact of foreign direct investment on the innovativeness of enterprises both in the industrial sector and in the service sector. Moreover, the impact of FDI on the innovativeness of enterprises in the industrial sector turned out to be stronger than the impact on the innovativeness of companies from the service sector.

Keywords: innovation, foreign direct investment (FDI), ECM, Granger causality

JEL Classification: C23, F21, O31
DOI: 10.15611/aoe.2023.1.06

Quote as: Salamaga, M. (2023). Study on the influence of foreign direct investment on innovations in enterprises in Poland using the ECM panel model. *Argumenta Oeconomica*, 1(50), 119-136.

* Department of Statistics, Cracow University of Economics, Poland. ORCID: 0000-0003-0225-6651.

1. Introduction

According to various economic theories and empirical studies carried out to date, the inflow of foreign direct investment is one of the key factors in the development of innovation in the host countries. This applies to developing countries but not exclusively. Inward FDI flows have an influence on innovations of enterprises thanks to the mechanism of technology and knowledge transfer between the investor's country and the host country of FDI. Unlike typical portfolio investments, this form of investment is aimed at building a new business entity from scratch or at taking some shares in the existing enterprise which enables its control. To ensure the proper development of an enterprise created with inward FDI, transfers of the relevant production and technological solutions, know-how, management organisation techniques and marketing techniques are necessary to improve the efficiency of an enterprise, its innovations and competitiveness. Thus, foreign direct investment may affect innovations and competitiveness of enterprises both in the material and non-material spheres. The former includes obtaining modern machinery by beneficiaries of FDI, while the latter is about gaining technical, organisational and managerial expertise, improving the quality of human capital by providing training and courses for staff at various levels. FDI also contributes to the general improvement of company management by adapting the most efficient techniques of financial planning, improving work efficiency and better cost control. These elements, combined with greater availability of foreign capital, are conducive to or even decisive for companies' innovations. The potential effects of the development of innovations on the enterprise's environment should also be mentioned here as the implemented product or organisational innovations are usually later followed and imitated by other domestic companies, and this contributes to the development of other industries and the entire economy of the host country of FDI.

It should also be emphasised that the spillover effects linked to the adaptation of new technologies and to qualified staff (who start working for other companies), or to the creation of cooperating logistics chains with vertical links, depend on the capacities of economy sectors of the FDI host country to absorb new production and organisational solutions, and on the nature of FDI. If foreign direct investment projects are focused on access to sales markets, the investor usually tries to limit the spillover effect of technologies because technology transfer is meant to be beneficial only to the company's operations. When FDI is focused on low prices of production factors, then cooperation links with local companies are required and the diffusion of technologies to the sector and beyond becomes faster. Nevertheless, there are theoretical concepts, such as Vernon's product life cycle theory (Vernon, 1966), which assume at some stage of the investment process even inhibiting or at least delaying the impact of inward FDI flows on innovation. However, Dunning's pull factor theory (Dunning, 1995) suggests that a feedback effect is possible, i.e. that it

is innovations in the host country that attract foreign investors. According to this theory, investors are looking for investment opportunities in companies that have more modern technologies than are owned by the investing company. These technologies are then sent by the investor to the parent company. Ultimately, therefore, foreign investors do not incur R&D expenditure in the country hosting the FDI. There are also theories that assume a mixed impact of FDI on innovation, where the benefits and disadvantages of the inflow of FDI may balance each other in some way (Yang et al., 2013). In light of the theoretical concepts mentioned here, as well as empirical research reviewed later in this article, the impact of inward FDI on innovation is not always one-way and does not always brings benefits. The debate in the literature on the subject continues on this topic and concerns also the countries of Central and Eastern Europe, including Poland as inward FDI beneficiaries. The question is to what extent do inward foreign direct investment projects in Poland support this process, and to what extent do they limit it? Poland is a country with great economic potential, attractive for foreign investors, where the share of inward FDI stocks in GDP exceeds 40%[1], so the assessment of the direction and strength of the impact of FDI on innovation seems fully justified. The studies conducted so far show that the inflow of FDI to Poland increases the innovativeness of the economy. However, in general, research on the relationship between FDI and innovation does not take into account the close cause-effect relationship, but only includes independent analyses of the inward FDI structure and innovation indicators. Yet, the results of the research conducted so far lack models that will make it possible to describe the cause-effect relationship between the discussed categories for the Polish economy.

This article is the author's attempt to bridge this gap by analysing the influence of foreign direct investment on innovations of enterprises in Poland in cause and effect terms. For this purpose, panel data models with error correction mechanism and the Granger causality test were applied. Such research methodology allows for determining the actual direction and strength of the impact of FDI on innovations in companies from the major sectors of the economy.

2. Theoretical concepts of the relationship between FDI and innovation

The debate on the relationship between inward FDI flows and innovation in the literature concerns three various theoretical concepts. They assume that the impact of FDI on innovation can be positive, negative or neutral. The positive impact of FDI on innovation is the most motivated one. These theories emphasise the importance of channels, such as transfers of new technologies and knowledge from the investing country to the FDI host country, through which the improvement of innovation occurs (Aghion et al., 2009; Bertschek, 1995; Caves, 1974). According to these

[1] Source: www.nbp.pl (download date: 17.02.2022)

concepts, the entry of foreign investors increases competitiveness and thus stimulates production efficiency and generates additional economies of scale for local companies, enabling them to improve their productivity. Along with the increase in productivity, the company has additional financial resources that can be spent on innovation and R&D (Aghion et al., 2001). According to Blomström and Kokko (1998), the competitiveness of enterprises, in turn, improves the allocative and technical efficiency of companies and stimulates innovation, therefore FDI is expected to improve the innovativeness of companies. The positive impact of FDI on local companies is twofold: it has a positive effect on the companies that support them, and secondly, it has a positive impact on other local companies in the same economy or sector. In addition to excellent knowledge transfer, host firms also benefit enormously from FDI in strengthening their capital base. Thanks to additional capital resources, such companies can afford to employ highly qualified and creative staff, which can be used in the implementation of innovative solutions in the production and management process (Glass, Saggi, 2002). The positive impact of FDI on innovation may also be the result of increased demand for the products of local companies (Rodrigue-Clare, 1996). If their production companies are not able to keep up with the growing demand, they can implement innovative solutions allowing for more efficient production. Knowledge transfer driven by FDI can help here.

Attention should also be paid to the imitation effect, bringing benefits to local companies in the country hosting FDI (Salomon, 2006). The spillover effect in terms of innovation is possible, among others, thanks to the transfer of employees between companies. Some researchers believe that the inflow of FDI inhibits the innovativeness of the economy in the host country. In his theory of product life, Vernon (1966) claims that Multinational Enterprises (MNEs) spend more on innovation in their operations in the early stages of the company's life cycle, and that moving to host countries in the mature stage involves less R&D spending.

There are also concepts according to which MNEs look for such destinations for their FDI, in which local companies will only be responsible for such stages of production that do not require advanced technology or knowledge transfer (e.g. for final processing or distribution), and the stages requiring it will be implemented in parent companies (Yang et al., 2013). Then, the innovation benefits for the FDI-powered company will be limited. On the other hand, according to Dunning's theory (Dunning, 1995), one of the motives behind FDI is the reverse transfer of technology and knowledge – from the host country to the investor's country. Then, investors are interested in taking over the technological solutions used in the beneficiary's country and adopting them in the parent company. Such a situation will not bring benefits in terms of innovation in the economy of the host country.

The third group of theoretical concepts assumes that the impact of FDI on innovation is mixed or neutral. Dunning (1993) considered the relationship between the inflow of FDI and benefits for companies and for the entire economy, and showed

that if local firms benefit from the inflow of FDI, then the whole economy does not benefit, and vice versa. Blind and Jungmittag (2004) proved that the way FDI influences innovation depends on the type of FDI: in the case of greenfield investments, the benefits of FDI are the greatest. Yet, the benefits of mergers and acquisitions depend on whether the investing enterprise or the enterprise that is the target of the investment has more advanced innovations. According to another concept, the degree of benefit from innovation will depend on the size of the technological gap between the investor's country and the FDI host country – the greater the gap, the greater the impact on innovation (Blind, Jungmittag, 2004).

3. Review of empirical literature

Polish and foreign literature has seen research on the relations between FDI and innovation ratios for a long time, but mostly in the context of broader considerations on the influence of FDI on the economy in general. In the case of the Polish economy, the research results usually confirm the beneficial impact of FDI on the innovations of enterprises by industries, regions or countries (Stiebale, Reize, 2010; Wiśniewska, 2001). These conclusions are drawn based on the analysis of dynamics of the inflow of FDI and the analysis of available innovation ratios for economy sectors or countries with the omission of the modelling of cause-effect relations. World literature including empirical research often uses econometric modelling to describe these relations. Public and private expenditure on research and development, the number of patent applications, productivity of production factors, export volume of technologically advanced goods, and the inflow of foreign direct investments are the subjects of modelling. Researchers usually use regression models, panel data models, dynamic econometric models and VAR models. The research refers to whole economies, regions or sectors of the economy. Table 1 provides an overview of the major research findings on the relationship between FDI and innovation.

For example, the positive impact of FDI on innovations in the Chinese economy measured by the number of patent applications was confirmed, e.g. by Cheung and Lin (2004), Hsu and Yu-En (2015). Iacovone et al. (2009) came to similar conclusions when examining the investments of Walmart in the Mexican market. The same kind of influence on the innovations of the West German economy was confirmed by Bertschek (1995) who analysed 1,270 companies in the industry sector. Similar conclusions for the German economy were also obtained by Stiebale and Reize (2010). Temiz and Gökmen (2014) showed that FDIs are the driving force of economic growth and development both in developed and developing countries. Ghazel and Zulkhibri (2015), as well as Khachoo and Sharma (2016) in their separate studies, noted that FDI is an effective catalyst in the innovation capacity of host companies. Moreover, Khachoo and Sharma (2016) showed that greater benefits from the development of innovation caused by the inflow of FDI are visible in companies operating in identical industries.

Table 1

Overview of the study results of the relationship between FDI and innovation

Type of relationship	Study area	Research tools	Relevant studies
Positive	Germany	Probit model	Bertschek (1995)
	Taiwan	Negative binomial model, logit model, Principal Component Analysis (PCA)	Lin, Lin (2009)
	China	GMM estimator	Liu, Zou (2008)
	India	Poisson model, negative binomial model	Khachoo, Sharma (2016)
	Developing countries	Negative binomial model	Ghazel, Zulkhibri (2015)
	China	Fixed Effect (FE) and Random Effect (RE) models	Cheung, Lin (2004)
	Germany	Tobit model	Stiebale, Reize (2010)
	Mexico	Logit model	Iacovone et al. (2009)
	Asian countries	Gravity model, GMM estimator	Hsu, Yu-En (2015)
	Turkey	VAR model	Temiz, Gökmen (2014)
	Poland	Synthesis of the results of various studies	Wiśniewska (2001)
Negative	Spain	Poisson regression model, negative binomial regressions, GMM regression	Garcia et al. (2013)
	Central and Eastern Europe	Tobit model	Maaso et al. (2013)
	Sub-Saharan Africa	Stochastic frontier analysis (SFA)	Barasa et al. (2018)
Mixed, neutral	13 different countries	Dynamic panel data model	Potterie, Lichtenberg (2001)
	Czechoslovakia	Linear regression model	Kinoshita (2000)
	Spain	GMM estimator	Rosell-Martinez, Sanchez-Sellero (2012)
	China	Tobit model	Girma et al. (2005)
	54 developing countries	Panel threshold model	Loukil (2016)
	Southeast Asia	Fixed Effect (FE) and Random Effect (RE) models	Sivalogathasan, Wu (2014)

Source: author's own study.

However, some researchers proved that the benefits of FDI in terms of innovations refer only to the sectors and companies which invest in research and development (R&D). Such results were obtained by e.g. Kinoshita (2000) for the Czech market, whereas Rosell-Martinez and Sanchez-Sellero (2012) demonstrated that in Spain FDI flows to research sectors and to sectors experiencing an intense development. Similar results for the Chinese economy were obtained by Girma et al. (2006), who performed econometric modelling based on the sample of 30,000 companies. The same researchers confirmed that research and development are the major component elements of enterprise innovations, so the inflow of FDI can be endogenous and the spillover effects can be limited. A similar unequivocal influence of FDIs on the innovations of the economy was found by Loukil (2016), who concluded that below a certain threshold of technological development FDI projects can have an adverse effect on innovations in companies accepting innovations whereas above the same threshold, FDI has a positive impact on innovations of companies in developing countries.

The negative effects of the inflow of FDI for the innovativeness of local companies were confirmed by Garcia et al. (2013), and Barasa et al. (2018), who showed that foreign technology has a negative impact on the technological efficiency of companies in host countries. Similar conclusions were reached by Maaso et al. (2013), who found that the inflow of FDI is not conducive to innovation in host countries. Such an unequivocal influence of FDI on innovations in enterprises worldwide and the research gap in the application of advanced quantitative tools in research incline researchers interested in the Polish economy to further develop research on this matter.

4. Research methodology

The modelling of dependencies between foreign direct investment and innovations in businesses used data on the inflow of FDI to individual provinces in Poland, whereas expenditure on research and development by provinces and the number of patent applications in individual provinces were assumed as innovation ratios. These types of variables are most often adopted in studies on the innovativeness of economies, as they reflect fairly accurately the level of innovativeness of the economy (expenditure on research and development represents the side of capital expenditure on innovation, and the number of patents filed – the effects of innovation). All data were cross-sectional and time-based, and panel models of error correction mechanism (ECM) were used in the research. This is a combination of econometric tools known from the analysis of time series with the panel data analysis. On the one hand, the application of this approach was dictated by the nature of the data, but on the other hand, by the need to take into account the impact of the historical values of variables (autoregressive processes) on their current values and finally the search for a long-term relationship between FDI and enterprise innovation. The proposed

approach allows for the determination of the time-independent balance path for cointegrated economic processes along with short-term deviation from balance. This research methodology requires the examination of the existence of a unit root of the relevant time series (Baltagi, Kao, 2000).

The starting point in the testing of the stationarity of time series in panel models with error correction mechanism is the AR process:

$$y_{it} = \delta_i y_{t-1} + \alpha_i X_{it} + \varepsilon_{it}, \tag{1}$$

where: y_{it} – dependent variable, $i = 1, 2, \ldots, N$, n – number of units in the cross-section (of objects) or time series, $t = 1, 2, \ldots, T$ – number of periods during which objects are observed, X_{it} – endogenous variables which take account of trends and fixed effects, ε_{it} – error term.

In this model, the fulfilment of condition $|\delta_i| < 1$ by parameter δ_i suggests stationarity of the time series, whereas condition $|\delta_i| = 1$ means the existence of a unit root, hence non-stationarity of the series. Stationarity tests of variables for dynamic panel models usually are based on the assumption that parameter δ_i is identical in all cross-sections or that it can change its values arbitrarily within cross-sections (Im et al., 2003). In the former case, the Breitung test is used (2000), while in the latter the Fisher tests were based on ADF tests. This study used both groups of tests.

The key element of the modelling of innovation ratios and foreign direct investment is the analysis of the existence of a cointegrating relation between the time series of variables which form equations:

$$\ln P_{it} = \beta_{i0} + \beta_{i1} \ln FDI_{it} + u_{it}, \tag{2}$$

$$\ln RD_{it} = \gamma_{i0} + \gamma_{i1} \ln FDI_{it} + v_{it}, \tag{3}$$

where: P_{it} – number of patent applications in i-th object (here: province) during period t, RD_{it} – expenditure on R&D in i-th object during period t, u_{it}, v_{it} – error terms which are uncorrelated white noise processes.

Engle and Granger (1987) proved that a linear combination of two or more nonstationary time series can be stationary. If there is such a stationary linear combination, then the time series are cointegrated and this combination is the cointegration equation. The equation can be interpreted as long-term dependence between variables. For panel data models, the Johansen procedure (Maddala, Wu, 1999) and the Pedroni (1999) procedure were are used. The Engle-Granger procedure tests the stationarity of residuals in the model with variables with stationary increments (Kao, 1999). The Johansen procedure, in turn, applies empirical probabilities to combine individual, independent results where each panel unit is treated individually and allows for defining individual cointegration. The Johansen test is based on trace statistics of the matrix and maximum eigenvalue. The application of the Johansen procedure allows for determining at least r cointegration vectors.

In this research, both types of tests were used to detect cointegration. For stationary variables, it is possible to build ECM models and thus estimate short-term dependencies. A single-equation model of error correction for stationary increments of variables used in analyses of short-term dependencies for panel data may be written in the following form:

$$\Delta y_{it} = \alpha_0 + \sum_{j=1}^{p} \alpha_{1j} \Delta y_{i,t-j} + \sum_{j=1}^{q} \alpha_{2j} \Delta x_{i,t-j} + \gamma ECT_{i,t-1} + \varepsilon_{it}, \tag{4}$$

where: $ECT_{i,t-1}$ – error-correction term representing a long-term relationship, p, q – orders of lagged differences in variables (selected using the Schwarz information criterion), $y_{it} \in \{\ln P_{it}; \ln BR_{it}\}; x_{it} \in \{\ln BIZ_{it}\}$, ε_{it} – error term of the model.

Granger's analysis of variable causality was conducted using the Dumitrescu and Hurlin (2012) approach, which provides for the heterogeneity of panel data models. In this case, the basis is the model

$$y_{it} = \alpha_0 + \sum_{k=1}^{K} \gamma_i^{(k)} y_{i,t-k} + \sum_{k=1}^{K} \beta_i^{(k)} x_{i,t-t} + \varepsilon_{it}. \tag{5}$$

When testing the causality of variable X in relation to Y, the tested hypothesis was that the autoregressive structure of process X equals zero (H_0: $\beta_i = 0$ for $I = 1, 2, .., N$), as opposed to the hypothesis that β_i parameters differ from zero in each cross-section of panel data. The following test statistics using means from Wald test statistics determined separately for each cross-section is calculated below

$$\hat{Z}_{N,T}^{HNC} = \sqrt{\frac{n}{2K}} \left(\bar{W} - k \right), \tag{6}$$

where: n – number of observations, k – lag order of variables in the model, $\bar{W}$ – mean value of Wald statistics from all panels.

With the null hypothesis being true, statistic (6) has an asymptotic standard normal distribution. The results of the statistics calculated based on the appropriate vector autoregressive model (VAR) enable the examination of causality of FDI variables and innovation ratios. Data for calculations were retrieved from the databases of the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

5. Results of empirical research

Cross-sectional and time-based data considering the variables mentioned below were taken into consideration in the analysis of the influence of foreign direct investments on innovation of enterprises: *FDI* – inflow of foreign direct investment

in PLN million[2], *RD* – research and development expenditure in enterprises in PLN million[3], *P* – number of patent applications filed by businesses with the Patent Office of the Republic of Poland[4]. The values of variables were observed in 16 individual provinces during the period 2010-2019, whereas the calculations were performed individually for businesses from the industry sector and the services sector. The study applied logarithm variables. In accordance with the methodology of the construction of ECM models, the stationarity of the following time series of variables was analysed first: $\ln P_{it}$, $\ln BR_{it}$, $\ln FDI_{it}$ for $i = 1,2,...,16$, and $t = 1,2,\ldots,10$. Table 2 presents the results of panel unit root tests (p-value is given in brackets).

Table 2

The results of panel unit root tests

Sector	Test	$\ln P_{it}$	$\ln RD_{it}$	$\ln FDI_{it}$	$\Delta ln P_{it}$	$\Delta \ln RD_{it}$	$\Delta \ln FDI_{it}$
Industry	Fisher PP	15.145 (0.058)	9.541 (0.125)	12.658 (0.084)	30.145 (0.000)	27.212 (0.001)	42.104 (0.000)
	Breitung	-1.245 (0.214)	0.875 (0.321)	1.0234 (0.624)	-2.784 (0.009)	-2.889 (0.012)	-3.745 (0.000)
Services	Fisher PP	13.876 (0.092)	16.057 (0.065)	10.337 (0.074)	42.108 (0.000)	34.508 (0.000)	37.604 (0.000)
	Breitung	-0.544 (0.107)	1.024 (0.233)	0.983 (0.544)	-3.745 (0.012)	-4.714 (0.000)	-4.204 (0.001)

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

In light of the results of the conducted stationarity tests, it should be concluded that the first differences in the analysed variables are stationary. Each of time series was thus integrated of order I(1). Furthermore, it was verified whether there is cointegration between the relevant variables. For this purpose, the Johansen procedure (Table 3), and the Pedroni and the Kao tests (Table 4) were used.

Based on the results presented in Tables 3 and 4, it should be concluded that all the tests indicate the existence of cointegration between variables, except for the ADF Pedroni panel test which shows the absence of cointegration between expenditure on R&D among enterprises in the services sector and the inflow of FDI to the same sector. However, because the Kao panel test and the panel data test using the Johansen procedure suggest the existence of cointegration, in this case the decision was made that there is cointegration in all the analysed dependencies.

[2] Source: the National Bank of Poland (https://www.nbp.pl/), Orbis database (https://www.bvdinfo.com/en-gb/our-products/data/international/orbis)

[3] Source: Main Statistical Office of Poland (www.gus.gov.pl)

[4] Source: Patent Office of the Republic of Poland (https://uprp.gov.pl/pl)

Table 3

The results of panel cointegration tests using the Johansen procedure

Sector	Cointegration relationships	Statistical hypothesis	Max-Eigen Stat.
Industry	$\ln P_{it}=\beta_{i0}+\beta_{i1}\ln FDI_{it}+u_{it}$	H_0: Absence of cointegration	52.354 (0.000)
		H_1: At most one cointegrating vector	4.214 (0.451)
	$\ln RD_{it}=\gamma_{i0}+\gamma_{i1}\ln FDI_{it}+v_{it}$	H_0: Absence of cointegration	42.055 (0.000)
		H_1: At most one cointegrating vector	6.851 (0.342)
Services	$\ln P_{it}=\beta_{i0}+\beta_{i1}\ln FDI_{it}+u_{it}$	H_0: Absence of cointegration	61.874 (0.000)
		H_1: At most one cointegrating vector	5.983 (0.624)
	$\ln RD_{it}=\gamma_{i0}+\gamma_{i1}\ln FDI_{it}+v_{it}$	H_0: Absence of cointegration	74.365 (0.000)
		H_1: At most one cointegrating vector	2.124 (0.650)

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

Table 4

The results of panel cointegration – the Pedroni and the Kao tests

Sector	Cointegration relationships	Test	
		Pedroni ADF	Kao ADF
Industry	$\ln P_{it}=\beta_{i0}+\beta_{i1}\ln FDI_{it}+u_{it}$	-3.358 (0.002)	-6.251 (0.000)
	$\ln RD_{it}=\gamma_{i0}+\gamma_{i1}\ln FDI_{it}+v_{it}$	-1.441 (0.054)	-3.324 (0.032)
Services	$\ln P_{it}=\beta_{i0}+\beta_{i1}\ln FDI_{it}+u_{it}$	-4.205 (0.000)	-7.521 (0.000)
	$\ln RD_{it}=\gamma_{i0}+\gamma_{i1}\ln FDI_{it}+v_{it}$	-3.657 (0.003)	-5.797 (0.000)

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

In view of this, when the Fully Modified Least Squares (FMOLS) method was used, long-term relations between variables *FDI* and *RD* as well as *FDI* and *P* were estimated in accordance with models (2) and (3). The results given in Table 5 are

statistically significant and the obtained models of cointegration equations imply fairly good adjustment to empirical data (coefficients of determination are given in the last column of the table). When analysing the obtained parameters, it can be concluded that in the long run the impact of foreign direct investments on innovations in enterprises in the industrial sector is stronger than in the services sector. A 1% increase in the inflow of FDI to the industrial sector causes an increase in the number of patent applications by approximately 0.308%, and an increase in R&D expenditure in the same sector by approximately 0.645%. The same increase in the inflow of FDI to enterprises in the services sector has an influence on the increase in the number of patent applications in this group of businesses by approximately 0.208%, and increases R&D expenditure by approximately 0.386% on average. What is notable, is that the fact that long-term elasticity of R&D expenditure in relation to FDI is higher than the elasticity of the number of patent applications both in the industrial sector and the services sector. Therefore, from the long-term perspective, the inflow of FDI supports companies' innovation by higher R&D expenditure to a greater extent than by the number of invention patents.

Table 5

The results of estimation of long-term relations in models (2) and (3) with the use of the FMOLS estimator

Sector	Model	Parameter	Coefficient	R^2
Industry	$\ln P_{it} = \beta_{i0} + \beta_{i1} \ln FDI_{it} + u_{it}$	β_{i0}	1.364 (0.002)	0.856
		β_{i1}	0.308 (0.001)	
	$\ln RD_{it} = \gamma_{i0} + \gamma_{i1} \ln FDI_{it} + v_{it}$	γ_{i0}	-0.961 (0.024)	0.943
		γ_{i1}	0.645 (0.006)	
Services	$\ln P_{it} = \beta_{i0} + \beta_{i1} \ln FDI_{it} + u_{it}$	β_{i0}	2.671 (0.000)	0.971
		β_{i1}	0.208 (0.003)	
	$\ln RD_{it} = \gamma_{i0} + \gamma_{i1} \ln FDI_{it} + v_{it}$	γ_{i0}	1.220 (0.004)	0.826
		γ_{i1}	0.386 (0.007)	

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

In order to analyse the short-term relations between FDI and innovation ratios, the ECM panel model was estimated with the use of the LSDV estimator. Based on the results of the estimation of model (4), the parameters of elasticity for short-term

dependencies and the parameters of the error-correction term are presented in Tables 6 and 7. The parameters of the models were generally statistically significant at the significance level of 0.05, which allows for the generalisation of the conclusions based on the calculations.

A short-term 1% increase in the inflow of FDI in the industrial sector during the previous period caused an increase in R&D expenditure during the current period by approximately 0.079% *ceteris paribus*, whereas a negative parameter next to the

Table 6

The results of estimation of short-term elasticity for R&D expenditure in the ECM panel models

Sector	Explanatory variable	Parameter	Coefficient
Industry	$\Delta \ln FDI_{i,t-1}$	α_{21}	0.079 (0.012)
	$ECT_{i,t-1}$	γ	-0.062 (0.096)
Services	$\Delta \ln FDI_{i,t-1}$	α_{21}	0.041 (0.048)
	$ECT_{i,t-1}$	γ	-0.052 (0.057)

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

Table 7

The results of estimation of short-term elasticities for number of patents in ECM panel models

Sector	Explanatory variable	Parameter	Coefficient
Industry	$\Delta \ln FDI_{i,t-1}$	α_{21}	0.188 (0.032)
	$ECT_{i,t-1}$	γ	-0.113 (0.101)
Services	$\Delta \ln FDI_{i,t-1}$	α_{21}	0.107 (0.029)
	$ECT_{i,t-1}$	γ	-0.073 (0.034)

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

correction term provided for the balance during the previous period; approximately 9.6% of imbalance in comparison to the long-term growth path was corrected by a short-term adjustment process. In the services sector, the short-term relationship between R&D expenditure and the inflow of FDI was weaker than in the industrial sector; an increase in the inflow of FDI during the previous period by 1% caused an increase in R&D expenditure during the current period by approximately 0.041% on average, *ceteris paribus*. The error correction mechanism was slower here: 5.2% of deviations from a long-term growth path were corrected using the error correction mechanism.

Based on the results given in Table 7, it can be concluded that an increase in the inflow of FDI to the industrial sector by 1% in the previous year implied an increase in the number of patent applications during the current year by approximately 0.188% *ceteris paribus,* and approximately 11.3% of imbalance of deviations from the long-run relationship was corrected by a short-term correction process regulated by the error correction mechanism.

In the services sector, a short-term reaction to the inflow of FDI observable in the number of patent applications filed was again weaker in the industrial sector; a 1% increase in the inflow of FDI during the previous year caused an increase in the number of patent applications during the current period by approximately 0.107% on average, *ceteris paribus*, while adaptation to a short-term growth path took longer than in the industrial sector (7.3% of deviation from the long-term relationship was corrected here by the process of short-term adaptations). The Granger analysis of causality was an important element of the analysis of dependencies between FDI and innovations. The results of the panel causality test assuming the heterogeneity of panel model data (Dumitrescu, Hurlin, 2012) are given in Table 8.

Table 8

Results of the panel causality test of variables *RD*, *P* and *FDI*

Sector	Causality direction	Z_{NT}^{HNC}
Industry	*FDI* =>*RD*	4.823***
	RD =>*FDI*	2.012**
	FDI =>*P*	3.982***
	P =>*FDI*	1.058
Services	*FDI* =>*RD*	3.088***
	RD =>*FDI*	2.235**
	FDI =>*P*	4.214***
	P =>*FDI*	2.124**

Note: The significance of results at level 0.01; 0.05 and 0.1 is marked: ***, **, *.

Source: author's own study based on data from the Main Statistical Office of Poland, the National Bank of Poland, Orbis, and the Patent Office of the Republic of Poland.

Based on the results given in Table 8, it can be concluded that the inflow of FDI was a cause in the Granger's sense (at significance level of 0.01) of both R&D expenditure and of the number of patent applications filed. Such causality was observed both in the industrial sector and in the services sector while the causality in the direction from *FDI* to *RD* was stronger in the industrial sector, whereas the causality in the direction from *FDI* to *P* was stronger in the services sector than in the industrial sector. Furthermore, the presence of a statistically significant feedback reaction should be noted; R&D expenditures also provided a cause for the inflow of FDI in the industrial and services sectors, whereas the number of patent applications was a significant cause of the inflow of FDI in the services sector (dependencies significant at the significance level of 0.05). The reverse mechanism showed that not only the inflow of FDI contributes to innovation, but also innovation in enterprises can be an important magnet attracting foreign direct investment.

Conclusions

Direct foreign investment brings new technologies and innovative solutions in the sphere of production and organisation. However, as research shows the benefits FDI brings to innovations in specific industries depend on the capacity of these industries to absorb new technological solutions, the efficiency of the spillover effect, and other factors. For this reason, the examination of the issue for various economies, including the Polish economy, seems justified, especially that scientific analyses relating to the relation between FDI and innovation in Poland have so far lacked the modelling of cause-effect relations between the values discussed here.

The results presented in this paper confirm the hypothesis of the positive influence of foreign direct investment on innovation of enterprises both in the industrial and services sectors. These results are consistent with the results obtained by other researchers of the economies of developed and developing countries (Wiśniewska, 2001; Bertschek, 1995; Ghazel, Zulkhibri, 2015; Hsu, Yu-En, 2015). The influence of FDI on innovation of the industrial sector turned out to be stronger than the influence on innovation of companies from the services sector, which was proven in cointegration models and the Granger causality test. In the long run, the greater influence of FDI on expenditure than on the number of patent applications was visible, whereas in the short term the change in the number of patent applications turned out to be a stronger response to the inflow of FDI than the change in R&D expenditure. The correction mechanism of deviation from long-term relationship appeared to be stronger in the case of the industrial sector than in the services sector. This may result from the fact that in Poland the industrial sector absorbs FDI more easily than the services sector and is characterised by smaller inertia. Yet, a bidirectional causality of FDI-innovation was also observed, which indicates that the innovations introduced by enterprises are also important determinants for incoming FDI, both for the services and the industrial sectors. Therefore, it should be concluded

that in order to further narrow the technological gap in Poland and develop innovations, the intensification of development processes of a knowledge-based economy is important and should attract FDI even more. In addition, internal strengthening of enterprise innovation will, as has been proven, be an important element to support the inflow of FDI. However, it should be emphasised that the effect of strengthening innovation depends not only on the specific industry, but also on the type of FDI (greenfield vs. brownfield, wholly-owned vs. joint-ventures, vertical vs. horizontal). Investigating the detailed dependencies in this area requires further research. This study does not exhaust the whole spectrum of dependencies between FDI and enterprise innovation. Processes in knowledge-based economies are quite dynamic and relations between FDI and innovations will change. Therefore, it will become necessary to continue and expand such research by including other variables which represent new technologies. This will enable the constant monitoring of the direction and strength of the relation between FDI and innovation.

References

Aghion, P., Blundell, R., Griffith, R., Howitt, P., Prantl, S. (2009). The effects of entry on incumbent innovation and productivity. *The Review of Economics and Statistics*, 91(1), 20–32.

Baltagi, B. H., Kao, C. (2000). Nonstationary Panels, Cointegration in Panels and Dynamic Panels, A Survey, in Baltagi, B. (ed.). Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Amsterdam: JAI Press, 15, 7–52.

Barasa, L., Vermeulen, P., Knoben, J., Kinyanjuni, B., Kimuyu, P. (2018). Innovation inputs and efficiency: Manufacturing firms in Sub-Saharan Africa. *European Journal of Innovation Management*, 59–83.

Bertschek, I., (1995). Product and process innovation as a response to increasing imports. *Journal of Industrial Economics*, 43(4), 341–357.

Blind, K., Jungmittag, A. (2004). Foreign direct investment, imports and innovations in service industry. *Review of Industrial Organisation*, 25(2), 205–227.

Blomström, M., Kokko, A. (1998). Multinational corporations and spillovers. *Journal of Economic Surveys*, 12(3), 247–277.

Breitung, J., *The Local Power of Some Unit Root Tests for Panel Data*, [in:] *Advances in Econometrics, Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. B. Baltagi, Amsterdam: JAI Press, 15, p. 161–178.

Caves, R. E. (1974). Multinational firms, competition and productivity in host country markets. *Economica*, 41 (162), 176–193.

Cheung, K-Y., Lin, P. (2004). Spillover effects of innovation in China: Evidence from the provincial data. *China Economic Review*, 15, 25–44.

Choi, I. (2001). Unit Root Tests for Panel Data. *Journal of International Money and Finance*, 20(2), 249–272.

Dumitrescu, E. I., Hurlin, Ch. (2012). Testing for Granger Non-causality in Heterogeneous Panels. *Economic Modeling*, 29(4), 1450–1460.

Dunning J. H. (1995). Reappraising the Eclectic Paradigm in an Age of Alliance Capitalism. *Journal of International Business Studies*, Palgrave Macmillan; Academy of International Business, 26(3), 461–491.

Dunning, J. (1993). Multinational enterprises and the global economy. Addison Wesley: New York.

Engle, R. F., Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251–276.

García, F., Jinb, B., Salomon, R. (2013). Does inward foreign direct investment improve the innovative performance of local firms? *Research Policy*, 42(1), 231–244.

Ghazel, R., Zulkhibri, M. (2015). Determinants of innovation outputs in developing countries: Evidence from panel data negative binomial approach. *Journal of Economic Studies*, 42(2), 237–260.

Girma, S., Gong, Y., Gorg, H. (2006). Can you teach an old dragon new tricks? An innovation activity in Chinese state-owned enterprises. *ZA Discussion Paper*, 2267, 1–26.

Harris, R. D. F., Tzavalis, E. (1999). Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed. *Journal of Econometrics*, 91(2), 201–226.

Hsu, J., Tiao, Y-E. (2015). Patent rights protection and foreign direct investment in Asian countries. *Economic Modelling,* 44, 1–6.

Im, K. S., Pesaran, M. H., Shin, Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, 115 (1), 53–74.

Jacovone, L., Javorcik, B., Keller, W., Tybout, J. (2009). Walmart in Mexico: The impact of on innovation and industry productivity. *World Bank Working Paper*, 1–36.

Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics*, 90(1), 1–44.

Khachoo, Q., Sharma, R. (2016). FDI and innovation: An investigation into intra-and inter-industry effects. *Global Economic Review*, 45(4), p. 311–330.

Kinoshita, Y. (2000). R&D and technology spillovers via: Innovation and absorptive capacity. *William Davidson Institute Working Papers*, 349, 1–24.

Lin, H., Lin, S. E. (2009). FDI, trade and product innovation: Theory and evidence. *Southern Economic Journal*, 77(2), 434 – 464.

Liu, X., Zou, H. (2008). The impact of greenfield FDI and mergers and acquisitions on innovation in Chinese high-tech industries. *Journal of World Business*, 43(3), 352–364.

Loukil, K. (2016). Foreign direct investment and technological innovation in developing countries. *Oradea Journal of Business and Economics*, 1(2), 31–40.

Maaso, J., Roolaht, T., Varblane, U. (2013). Foreign directinvestment and innovation in Estonia. *Baltic Journal of Management*, 8(2), 231–248.

Maddala, G. S., Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, 61, 631-652.

Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics*, 61(4), p. 653–670.

De la Potterie, P. B., Lichtenberg, F. (2001). Does foreign direct investment transfer technology across borders? *The Review of Economics and Statistic*, 83(3).

Rodrigue-Clare, A. (1996). The division of labour and economic development. *Journal of Development Economics*, 49(1), 3–32.

Rosell-Martine, J., Sanchez-Sellero, P. (2012). Foreign direct investment and technical progress in Spanish manufacturing. *Applied Economics*, 44(19), 2473–2489.

Salomon, R. M. (2006). Spillovers to foreign market participants: Assessing the impact of export strategies on innovative productivity. *Strategic Organization*, 4(2), 135–164.

Sivalogathasan, W., Wu, X. (2014). The effect of foreign direct investment on innovation in South Asian emerging markets. *Global Business and Organizational Excellence*, 33(3), 63–76.

Stiebale, J., Reize, F. (2010). The impact of through mergers and acquisitions on innovation in target firms. *International Journal of Industrial Organisation*, 29(2), 155–167.

Temiz, D., Gökmen, A. (2014). Inflow as an international business operation by MNCs and economic growth: An empirical study on Turkey. *International Business Review*, 23, 145–154.

Wiśniewska, J. (2001). Bezpośrednie inwestycje zagraniczne w procesie kreowania innowacyjności i konkurencyjności przedsiębiorstw. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 21.

Vernon, R. (1996). International investment and international trade in the product cycle. *The Quarterly Journal of Economics*, 106(2), 190–207.

Yang, Y., Yang, X., Doyle, B. W. (2013). The location strategy and firm value creation of Chinese multinationals. *Multinational Business Review*, 21(3), 232–256.

Received: August 2021, revised: March 2022

Acknowledgement: *This publication was financed from the funds allocated to Cracow University of Economics.*

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

Gero Szepannek, Karsten Lübke***

How much do we see? On the explainability of partial dependence plots for credit risk scoring

Risk prediction models in credit scoring have to fulfil regulatory requirements, one of which consists in the interpretability of the model. Unfortunately, many popular modern machine learning algorithms result in models that do not satisfy this business need, whereas the research activities in the field of explainable machine learning have strongly increased in recent years. Partial dependence plots denote one of the most popular methods for model-agnostic interpretation of a feature's effect on the model outcome, but in practice they are usually applied without answering the question of how much can actually be seen in such plots.

For this purpose, in this paper a methodology is presented in order to analyse to what extent arbitrary machine learning models are explainable by partial dependence plots. The proposed framework provides both a visualisation, as well as a measure to quantify the explainability of a model on an understandable scale. A corrected version of the German credit data, one of the most popular data sets of this application domain, is used to demonstrate the proposed methodology.

Keywords: credit scoring, interpretable machine learning (IML), partial dependence plot (PDP), explainability

JEL Classification: C52
DOI: 10.15611/aoe.2023.1.07

Quote as: Szepannek, G., Lübke, K. (2023). How much do we see? On the explainability of partial dependence plots for credit risk scoring. *Argumenta Oeconomica*, 1(50), 137-150.

1. Introduction

During the last few years several frameworks for automated machine learning (autoML, Hutter et al., 2018) have been proposed. One such framework is provided by the R package mlr3 (Lang et al., 2019). It allows to define a chain of modelling steps, including data preprocessing operations such as dimensionality reduction and imputation up to the final model evaluation using different strategies such as cross-validation, bootstrap and also holdout sets. All model specification choices can be

** Stralsund University of Applied Sciences, Germany. ORCID: 0000-0001-8456-1283.
** FOM University of Applied Sciences, Dortmund, Germany.

defined as so-called hyperparameters, and algorithms are provided in order to optimise these hyperparameters with regard to a predefined performance measure. As a consequence of the free availability of tools such as mlr3, the use of complex machine learning algorithms has been facilitated also for companies with comparatively low experience in this field. The resulting models are able to detect complex nonlinear multivariate dependencies without the need for the analyst to explicitly specify the kind of the functional relationship of the dependence. For this reason, such models are often called black box models.

In the application context of credit risk scoring, traditionally white box logistic regression models (Crook et al., 2007; Szepannek, 2022) are frequently used in business practice. Nonetheless, numerous benchmark studies have shown that properly parametrised modern machine learning algorithms, such as random forests and gradient boosting, are often of superior predictive accuracy compared to the aforementioned traditional scorecard models (for an overview cf. Louzada et al., 2014). A comprehensive benchmark study which evaluates several algorithms on a set of domain-specific data sets on a meta-level can be found in Baesens et al. (2002) and has been updated by Lessmann et al. (2015). The specific situation of unbalanced classes was addressed by Vincotti and Hand (2002) and Brown and Mues (2012), and investigated together with a systematic hyperparameter tuning for several classes of machine learning algorithms in a comprehensive benchmark study (Bischl et al., 2014). In Crook et al. (2007) and Szepannek (2017), the current challenges are discussed in a broader context, e.g. reject inference (Banasik and Crook, 2007), the Basel 2 accord (Basel Committee on Banking Supervision, BCBS, 2005), and profit scoring (Verbraken et al., 2014).

In order to prevent the concomitant lack of model understanding, the BCBS established a number of requirements on transparency from the perspective of regulation. The ”selection of certain risk drivers and rating criteria should be based not only on statistical analysis, but the relevant business experts should be consulted on the business rationale and risk contribution of the risk drivers under consideration” (European Banking Authority, 2017). This underlines the need for an appropriate methodology to understand what the models have learned, and still for their explanation.

According to Szepannek and Aschenbruck (2020), there can be different requirements to the explanation of a model depending on the context. Several authors recently applied methods of interpretable machine learning to credit scoring (Biecek et al., 2021; Bussmann et al., 2020; Dastile and Celik, 2021; Demajo et al., 2020; Torrent et al., 2020). In Bücker et al. (2021), the different requirements are linked to the corresponding methodology within a unified framework for *Transparency, Auditability and eXplainability for Credit Scoring* (*TAX4CS*). According to this, the methods can be distinguished into either *global* explainability on the model level such as variable importance (Breiman, 2001), partial dependence (PD, Friedman 2001), or accumulated local effects (ALE, Apley, 2016), or *local* explainability on

the level of individual predictions such as Shapley additive explanations (SHAP, Strumbelj and Kononenko, 2014), breakdown plots (Staniak and Biecek, 2018), or local interpretable model explanations (LIME, Ribeiro et al., 2016). Many of them can be accessed via the DALEX framework (Biecek, 2018).

This paper concentrates on partial dependence, denoting a popular and well-known method for a model-agnostic assessment of a feature's effect on the model outcome. Despite its popularity and its frequent use in practice, partial dependence analysis is usually applied without addressing the corresponding question how much can actually be seen in the resulting plots. For this purpose, the methodology is presented in order to analyse to what extent arbitrary machine learning models are explainable by partial dependence plots. The proposed framework provides both a visualisation as well as a measure to quantify the explainability of a model on an understandable scale.

Molnar et al. (2020a) pointed out that the superior performance of complex machine learning models results from their ability to detect high order dependencies and nonlinearities. Such dependencies are difficult to understand for analysts, while it has to be noted that the trade-off between predictive accuracy and interpretability is not necessarily given for any data (Rudin, 2019). As a potential solution, criteria are proposed that help to quantify the interpretability of a model. Model selection can thus consist in multi-objective optimisation of both predictive performance and interpretability. The approach followed in this paper differs from this in the sense that it assumes an existing model (which may be the one with the largest predictive accuracy). Afterwards, the question addressed is "How much can we see in the interpretation given by the partial dependence plots for a given model?"

In Section 2, partial dependence is reviewed. Based on this, a measure is presented that allows to quantify how far it explains a given model. In the case study, the methodology is applied to the real-world context of credit scoring using the South German credit data (Groemping, 2017; Szepannek and Luebke, 2021). In Section 3, an algorithm is presented that can be further used to identify a subset of variables which best serve to explain model. Finally, the research results are summarised in Section 4.

2. Quantifying explainability

2.1. Partial dependence

Referring back to Friedman (2001), partial dependence plots (PDP) are a popular tool to understand the effect of one or several features w.r.t. the output of a predictive model. One of their advantages is that they can be used for different kinds of predictive models $\hat{f}(x)$: the set of predictor variables $x = (x_s, x_c)$ is split into disjoint subsets and the partial dependence function for a subset x_s is given by:

$$PD_s(X) = \int \hat{f}(X_s, X_c)\, dP(X_c) \tag{1}$$

This means that a partial dependence function computes the expected prediction given X_s takes the values x_s. For a data set with n observations, it is estimated by:

$$\widehat{PD}_s(x_s) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_s, x_{ic}), \tag{2}$$

where x_{is} are the values that observation i takes in X_s. Note that for $X_s = X$ it is $PD(X) = \hat{f}(X)$ corresponds to the model itself and for $s = \varnothing$ or in other words $X_c = X$, the partial dependence function ends up in:

$$PD_{\varnothing} = \int \hat{f}(X)\, dP(X), \tag{3}$$

which is a constant that can be estimated by $\frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_i)$.

2.2. Application to the South German credit data

The South German credit data is publicly available at the UCI ML benchmark repository (Dua and Graff, 2017) and has been made available by Groemping (2019; see also Szepannek and Luebke, 2021). It has 1000 observations and 21 variables where 7 predictors are numeric and 13 are categorical plus a binary target variable. The predictable event describes the default status of a loan. The overall prior default rate on the data is 0.3. For the purpose of this paper a random forest model was trained on the South German credit data using default parameters according to Liaw and Wiener (2002), which turned out to be a good choice for this purpose (Szepannek,

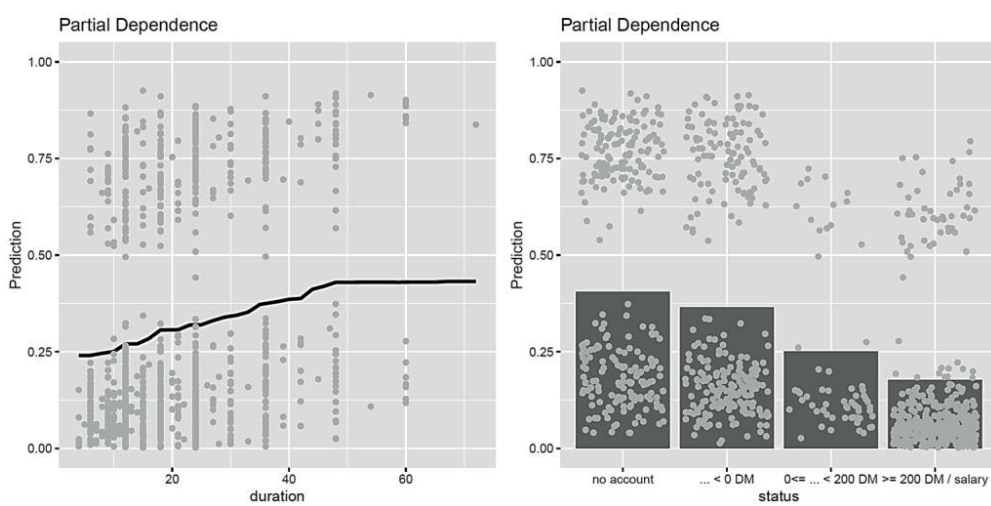


Fig. 1. Partial dependence plot for the variables duration (left, black line) and status account (right, bars), as well as the predictions on the training data (grey dots).

Source: authors' own.

2017). Usually, the data are split into training, validation and test sets in order to ensure a proper model selection and validation. As these aspects are beyond the scope of this paper, but rather the interpretability of the resulting model is of interest, no additional splits of the data were undertaken and the forest was trained on the entire data.

Figure 1 illustrates the partial dependence curves for the numeric variables duration and status account. This allows for a visual analysis of the effect of the variable on the predicted default probability and it can be easily seen that the risk (i.e. the default probability) increases for longer maturity time, whereas from roughly four years (45 months) onwards the risk stays constantly high. Analogously, it can be seen from the right plot that the risk decreases with a larger amount of money in the account. Nonetheless, when adding the predicted training data points to the graph it has to be noted that the PDP only partly explains the predictions by the models which cover a much broader range than the PDP indicates. This is obvious, as partial dependence is obtained by averaging. In turn, relying on the PD can be misleading.

2.3. Explainability

In the following step, a measure is derived to quantify the degree of explanation given by a partial dependence function for a model. A perfect explanation will have the same values for the partial dependence function and the predictions of the data. In this case, all points in a scatterplot of predictions vs. explanation (PX-plot) will lie on the diagonal. Such a plot is shown in Figure 2, where compared to Figure 1, the x-axis changed.

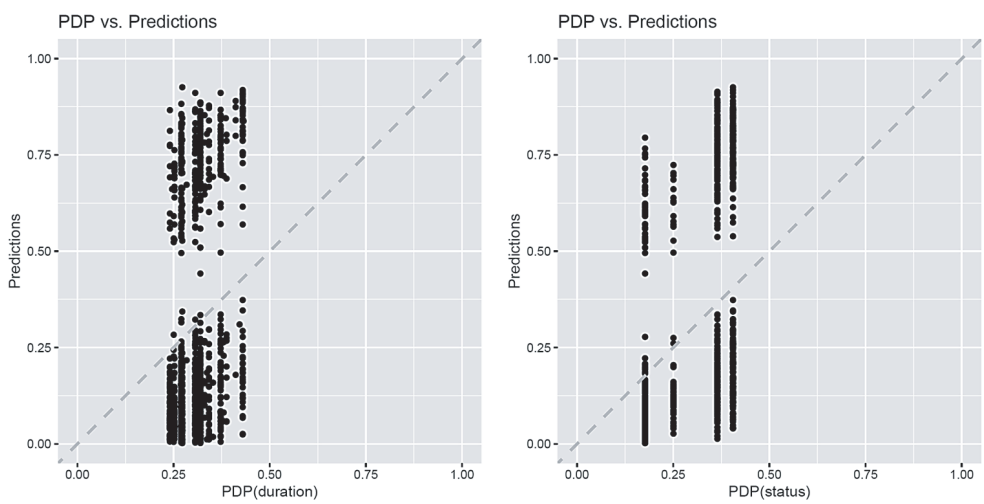


Fig. 2. Partial dependence (abscissa) vs. true predictions (ordinate) of training data for the variables duration (left) and status account (right).

Source: authors' own.

The above allows for a graphical analysis of the explainability. The more representative a PDP for a model, the closer the points to the diagonal. From this plot, it can be seen that the PDP covers a much smaller range of predicted values compared to the true model's predictions. Note that the x-axis of the right plot for the categorical variable status account takes only for distinct values – one for each category of the variable. In addition, the range of the partial dependence values is broader compared to those for the status account variable, and in particular for this variable there are only few observations with low values of the PDP ≤ 0.25 and large predictions > 0.75.

In order to quantify the confidence in an explanation given by a partial dependence plot, one can measure the differences between the partial dependence function $PD(X_s)$ and the model's predictions. A natural way of doing this is obtained by computing the expected squared difference (ESD):

$$ESD(PD_s) = \int \left(\hat{f}(X) - PD_s(X) \right)^2 dP(X_s). \tag{4}$$

Note that in contrast to common error functions, the ESD does not measure the difference between predictions and observations, but instead between the partial dependence function $PD_s(X)$ and the model's predictions $\hat{f}(X)$.

For an easier interpretation $ESD(PD_s)$ can be benchmarked against $ESD(PD_\varnothing)$:

$$ESD(PD_\varnothing) = \int \left(\hat{f}(X) - PD_\varnothing \right)^2 dP(X). \tag{5}$$

The comparison of both $ESD(PD_s)$ and $ESD(PD_\varnothing)$ can be used to quantify the **explainability** Υ of model $\hat{f}(X)$ by a partial dependence function PD_s via the ratio:

$$\Upsilon(PD_S) = 1 - \frac{ESD(PD_S)}{ESD(PD_\emptyset)}. \tag{6}$$

Note that Υ in (6) is somehow similar to the common R^2 as used in linear regression: Υ close to 1 states that a model is well represented by a PDP and the smaller it is, the fewer of the model's predictions are explained in the PDP. Real data plug-in estimates for $ESD(PD_s)$ and $ESD(PD_\varnothing)$ are obtained using $\widehat{PD}_s(x_s)$ and $\widehat{PD}_\varnothing$ as described above.

2.4. Application to the South German credit data

Table 1 (column $\widehat{\Upsilon}$) shows the explainability of the random forest model on the South German credit data for all variables. Among all the numeric variables, duration has the highest explainability of only $\widehat{\Upsilon} = 0.077$, which is nonetheless pretty far from 1 and thus reflects the visual impression as gained by considering Figures 1 and 2.

Columns $\widehat{\Upsilon}_k$ of the table describe the explainability for increasing number of variables k in the subset X_s (cf. Section 3). It can be seen that for two subsets $X_s \subset X_{s^*}$, it is $\Upsilon(PD_s) \leq \Upsilon(PD_{s^*})$ with $\Upsilon(PD_s) = 1$ for $X_s = X$. The PX-plot in Figure 3 illustrates the fit of PD_s from Table 1 with $\dim(X_s) = 9$ and $\widehat{\Upsilon} = 0.8$ (which obviously cannot be visualised anymore). Compared to Figure 2, the PDP covers a broader range and the points are closer to the diagonal.

Table1

Explainability for all variables of the South German credit data

Variable	$\hat{\Upsilon}$	k	$\hat{\Upsilon}_k$	Variable	$\hat{\Upsilon}$	k	$\hat{\Upsilon}_k$
status.account	0.221	1	0.221	rate.to.income	0.004	11	0.878
duration	0.077	2	0.304	personal.status	0.001	12	0.910
credit.history	0.074	3	0.366	job	0.000	13	0.937
credit.amount	0.054	4	0.434	resident.since	0.000	14	0.960
purpose	0.039	5	0.521	housing	0.013	15	0.977
savings	0.044	6	0.595	other.debtors	0.004	16	0.989
age	0.023	7	0.671	num.credits	0.001	17	0.995
employment.since	0.018	8	0.742	telephone	0.001	18	0.998
property	0.017	9	0.805	numb.people.liable	0.000	19	1.000
other.installments	0.017	10	0.843	foreign.worker	0.001	20	1.000

Source: authors' own.

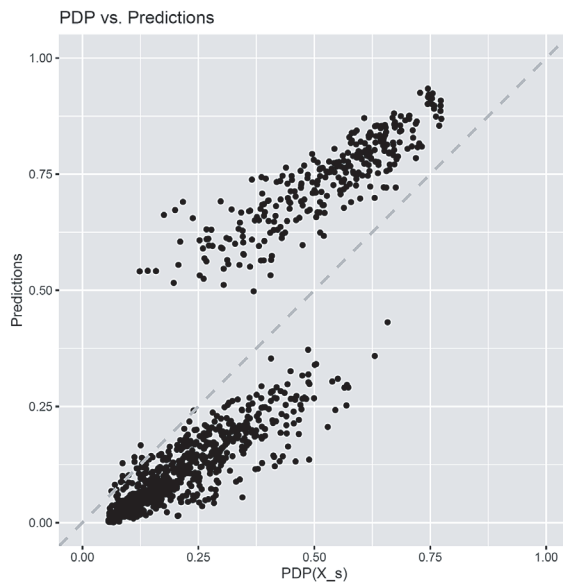


Fig. 3. Partial dependence vs. true predictions of training data for dim(X_s) = 9 and = 0.8.

Source: authors' own.

2.5. Connection to the existing methodology

Note that the proposed measure of explainability Υ reflects the difference between the PDP and the model's prediction, which is an important but not the only aspect of interest with regard to explainability. A well-known limitation of partial dependence curves is that they might be misleading in the case of correlated predictor variables

(Hooker and Mentch, 2019). In Friedman and Popescu (2008), an H^2 statistic is proposed that can be used to identify the existence of interactions between predictor variables. For correlated predictor variables, accumulated local effect plots (ALE, Apley, 2016) have been shown to be more appropriate than partial dependence plots. ALE plots are beyond the scope of this paper, but the extension of Υ for ALE plots may be a subject for future research.

A popular visual tool to analyse hidden variability behind a partial dependence curve are individual conditional expectation (ICE) curves (Goldstein et al., 2015), where instead of averaging over all the observations, a separate PD curve is drawn for each observation x_i:

$$\widehat{ICE}_s(x_i) = \hat{f}(x_s, x_{ic}). \tag{7}$$

The resulting plot of the ICE curves enables to understand the heterogeneity of the PD as a function of x_s (cf. Figure 4 (left) for the variable duration). In particular, ICE plots can be used for a visual analysis of whether the individual curves show the same trend. Yet, to the best of the author‘s knowledge, this can only be analysed visually, but no objective measure has been proposed in order to quantify this. In contrast, explainability Υ quantifies the observed variation hidden behind a partial dependence function into one single and interpretable value that is close to 1 (for small variation) and close to 0 (for strong variation) by integrating over the distribution $P(X_s)$.

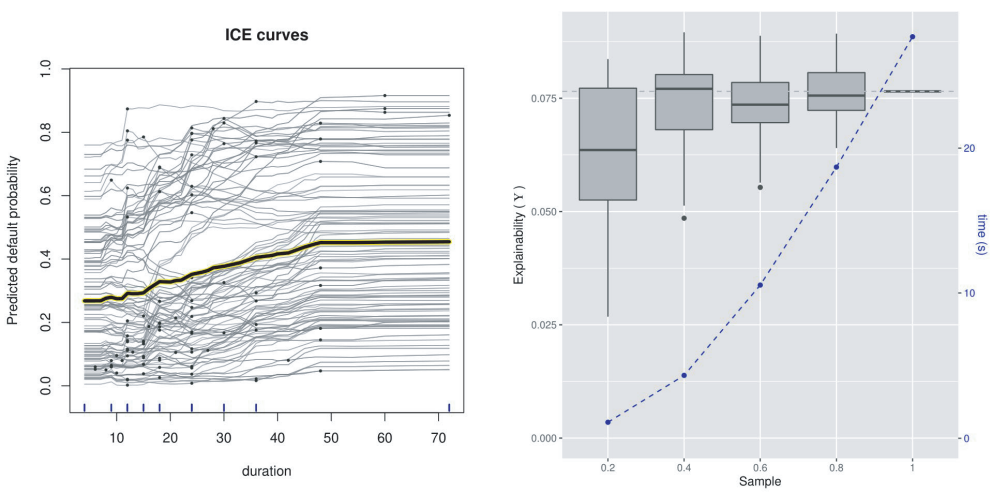


Fig. 4. ICE curves for the variable duration (left) and simulation results for the computation time based on subsamples of different size, as well as the resulting distribution of $\hat{\Upsilon}$ for the variable duration (right).

Source: authors' own.

Another issue of PDPs is their extrapolation to areas where little or no training data is available (Hooker and Mentch, 2019, Molnar et al., 2020b). Note that explainability as a global measure reflects the distribution of the training data w.r.t. the predictor variables, i.e. a large value of Υ does not prevent from misinterpreting extrapolations of the model outside the range of the training data.

Note that for the individual curves in an ICE plot, the values of x_s are varied regardless of how likely they are to occur, conditional on x_{ic}, which might be misleading. Υ, in addition, takes into account the joint distribution of variables from X_s and X_c, as from each curve only the observed points x_{is} (dots in the graph) are used.

2.6. Computational considerations

For the common implementations of partial dependence plots, e.g. those in Greenwell (2017), Biecek (2018) and Molnar et al.,(2018), the scope consists in visuali-sation of the PD curve and it is sufficient to restrict on computing $\widehat{PD}_s$ for a subset grid of the data. In contrast, $\widehat{\Upsilon}$ accounts for distribution $P(X_s)$, and thus requires computation of the partial dependence $\widehat{PD}_s(x_i)$ for all observations.

Computation of $\widehat{PD}_s(x_i)$ requires the Cartesian product $x_s \otimes x_c$ of the two variable subsets x_s and x_c of the data, therefore the calculation of for given data is $O(n^2)$ in the number of observations n with regard to both computation time and memory usage. In order to circumvent this issue arising with large sample sizes, an alternative consists in its computation on a random subsample of the x_{is}, $i = 1, ..., n$. Note that a similar approach was proposed to reduce the computation cost for Shapley additive explanations (Strumbelj and Kononenko, 2014), where random subsets of variables are used in order to avoid enumerating all possible permutations of variable subsets. Naturally, this trades off with the variance of the estimate. Figure 4 (right) illustrates both the reduction in (average) computation time (dashed line) as well as the increasing variability of the estimates (box plots) for 50 random samples of the x_s using an INTEL Xeon CPU E3-1505M v5 2.8Ghz 8 core with 32GB RAM.

3. Maximising explainability

3.1. Based variable selection

According to Table 1 (column $\widehat{\Upsilon}$), Υ can be used to compare different variables with regard to their ability to explain a model (using a PDP). Consequently, a forward variable selection can be carried out to maximise the explainability of a model with as few variables as possible (cf. Algorithm 1). Note that, as opposed to traditional variable selection or variable importance, the variables here are not selected with regard to the model's performance but rather with regard to the degree of explanation that they provide for an existing model.

Algorithm 1 $\hat{\Upsilon}$ based forward variable selection in order to maximize explainability.

Initialize $X_s = \emptyset$ and $X_c = X$.
repeat
  for all variables $X_j \in X_c$ **do**
    $X_s^{candidate} = X_s \cup X_j$
    Compute $\hat{\Upsilon}(X_s^{candidate})$
  Determine X_{j^*} that maximizes $\hat{\Upsilon}(X_s^{candidate})$.
  Set new $X_s = X_s \cup X_{j^*}$ and $X_c = X_c \setminus X_{j^*}$
until $X_c \neq \emptyset$

Source: authors' own.

3.2. Application to the South German credit data

Table 1 (column $\widehat{\Upsilon}_k$) provides an example of variable selection based on Υ to maximise explainability (the step number is indicated in column k): a PDP of only two variables already provides an explainability of 0.304 and for dim (X_s) = 5 (/9/12) an explainability $\widehat{\Upsilon}_{\dim(X_s)}$ = 0.5 (/0.8 /0.9) is obtained. Figure 5 shows a trellis visualisation (Cleveland, 1993) of a two-dimensional PDP (as implemented in e.g. Greenwell, 2017) for the two variables: status account and duration, with the highest explainability. It reveals the same trend of increasing risk with longer maturity times for all status levels of the account, but an observable interaction exists for existing accounts with a low or negative balance where the increase in risk is stronger.

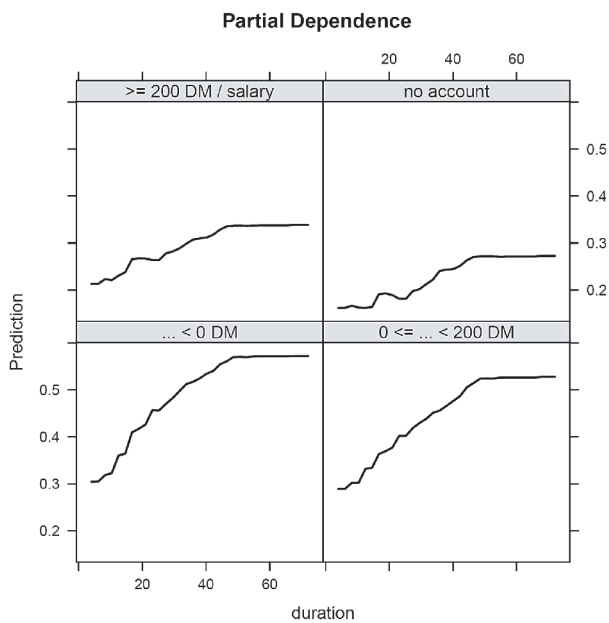


Fig. 5. 2D PDP for the variables status account and duration.

Source: authors' own.

Although in general, partial dependence functions are not restricted with regard to the dimension of X_s, their visualisation is limited to one or two dimensions. For more than two variables one can create scatterplot matrices (Cleveland, 1993), but this still does not allow to visualise higher order interactions between variables. This should be kept in mind when partial dependence plots are used to explain black box machine learning models. For the random forest model on the South German credit data, the most explainable two-dimensional PDP from Figure 5 only explains 30% of the variation of the model's predictions.

Conclusion

In recent years, several failures of AI applications have occurred. As a result, regulatory requirements for business applications of machine learning and the ongoing hype around the methodology for explainable AI (*XAI*) have emerged, but a unified framework on up to what extent the explanations by *XAI* can be misleading is still missing. Hence, a methodology was presented that allows to analyse to what degree predictive black box machine learning models can be explained by partial dependence plots. The framework provides both a graphical analysis of the mismatch between the PD curve and the predictions by the model in terms of PX-plots, as well as a measure (Υ) to quantify explainability of a model by a PDP on an interpretable scale. An algorithm was presented to maximise explainability with a low-dimensional PDP.

The proposed methodology was applied in this study to the publicly available South German credit data using a random forest model. It appears that a reasonable and well-interpretable partial dependence curve as it is observed for the variable duration, can still deviate noticeably from the predictions of the model – which has to be taken into account when explaining it. The proposed measure of explainability Υ can help to support business decisions by validating the model's interpretability. A PDP of the two most explainable variables, i.e. status account and duration, is more appropriate in order to understand how the model behaves.

In general, the explainability of the model becomes better when an increasing number of variables are taken into account, but for >2D PDPs can no longer be visualised and thus an analyst will not be able to understand any high-order dependencies that impact on the model's predictions. An R package with implementations of the described methodology is available on Github under https://github.com/g-rho/xPDPy.

Note that the proposed measure of explainability Υ only reflects the difference between the partial dependence curve and the predictions by the model under investigation, which is an important but not the only aspect of interest with regard to explainability. For example, individual conditional expectation (ICE) plots allow for a visual analysis whether the individual curves for all the observations show the same trend. Yet, currently there is no objective measure to quantify this – which could be a scope of future research.

There is also a need for ongoing research to develop methodology to understand high order interactions, e.g. based on the ideas presented in Britton (2019), and Gosiewska and Biecek (2019). Psychology provides a reasonable number of dimensions that can be simultaneously assessed by humans, there seems to be somewhere around seven (Miller, 1956), while naturally there might be differences depending on the experience of the analyst. However, it is questionable to what degree humans will ever be able to understand nonlinear high-order interactions. For this reason, the proposed measure can be considered as a tool to quantify the degree of explainability of a black box machine learning model.

Molnar et al. (2020a) suggested an approach to simultaneously optimise a trade-off between predictive accuracy and interpretability. In contrast, other authors claim to rather use interpretable models (Rudin, 2019) which may trade off with predictive power (but not always, cf. e.g. Buecker et al., 2021). To conclude, the benefits of more complex but uninterpretable models over interpretable ones should be carefully analysed during model selection.

An important challenge consists in the development of fair scoring models (Kusner and Loftus, 2020; Szepannek and Luebke, 2021) and future research on this topic will be based on causal inference (cf. e.g. Luebke et al., 2020 for some examples). According to the results of Zhao and Hastie (2019), partial dependence curves can be used for this purpose. This makes the suggested measure of explainability also an important concept on the road towards developing fair scores.

References

Apley, D. (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. arXiv:1612.08468.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen J. (2002). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *Journal of the Operational Research Society*, 54(6), 627-635.

Banasik, J. and Crook, J. (2007). Reject inference, augmentation and sample selection. *European Journal of Operational Research*, 183, 1582-1594.

Basel Committee on Banking Supervision. International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Bank for International Settlements, 2005.

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models. *R. Journal of Machine Learning Research*, 19(84), 1-5.

Biecek, P., Chlebus, M., Gajda, J., Gosiewska, A., Kozak, A., Ogonowski, D., Sztachelski, J. and Wojewnik, P. (2021). Enabling Machine Learning Algorithms for Credit Scoring – Explainable Artificial Intelligence (XAI) methods for clear understanding complex predictive models. arXiv:2104.06735.

Bischl, B., Kühn, T. and Szepannek, G. (2014). On Class Imbalance Correction for Classification Algorithms in Credit Scoring. In Operations Research Proceedings 2014. *Selected Papers of the Annual International Conference of the German Operations Research Society (GOR)*, 37-43, 2016.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

Britton, M. (2019). VINE: Visualizing Statistical Interactions in Black Box Models. arXiv:1904.00561.

Brown, I. and Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453.

Bücker, M., Szepannek, G., Gosiewska, A. and Biecek, P. (2021). Transparency, auditability and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, 73(1), 70-90.

Bussmann, N., Giudici, P., Marinelli, D. and Papenbrock, J. (2020). Explainable AI in fintech risk management. *Frontiers in Artificial Intelligence*, 3.

Cleveland, W. (1993). Visualizing Data. Hobart Press.

Crook, J., Edelman, D., and Thomas, L. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183, 1447-1465.

Dastile, X., and Celik, T. (2021). Making deep learning-based predictions for credit scoring explainable. *IEEE Access*, 9.

Demajo, L., Vella, V. and Dingli, A. (2020). Explainable AI for interpretable credit scoring. arXiv:2012.03749.

Dua, D., and Graff, C. (2017). UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/index.php, 2017.

European Banking Authority. Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189-1232.

Friedman, J. and Popescu, B. (2008). Predictive learning via rule ensembles. *Annals of Applied Statistics*, 2(3), 916-954.

Goldstein, A., Kapelner, A., Bleich, J. and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44-65.

Gosiewska, A. and Biecek, P. (2019). iBreakDown: Uncertainty of model explanations for non-additive predictive models. arXiv:1903.11420.

Greenwell, B. (2017). An R Package for constructing partial dependence plots. *The R Journal*, 9(1), 421-436.

Groemping, U. (2019). South German credit data: Correcting a widely used data set. Department II, Beuth University of Applied Sciences Berlin.

Hooker, G. and Mentch, L. (2019). Please stop permuting features: An explanation and alternatives. arXiv:1905.03151.

Hutter, F., Kotthoff, L. and Vanschoren, J. (2018). Automated Machine Learning: Methods, Systems, Challenges. Springer.

Kusner, K. and Loftus, J. (2020). The long road to fairer algorithms. *Nature*, 534, 34-36.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L. and Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework. *R. Journal of Open Source Software*.

Lessmann, S., Baesens, B., Seow, H.V., and Thomas, L. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136.

Liaw, A. and Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2(3), 18-22.

Louzada, F., Ara, A. and Fernandes, G. (2016). Classification methods applied to credit scoring: A systematic review and overall comparison. *Surveys in OR and Management Science*, 21(2), 117-134.

Luebke, K., Gehrke, M. Horst, J. and Szepannek, G. (2020). Why we should teach causal inference: Examples in linear regression with simulated data. *Journal of Statistics Education*, 28(2), 133-139.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2).

Molnar, C., Bischl, B and Casalicchio, G. (2018). iml: An R package for interpretable machine learning. *Journal of Open Source Software*, 3(26).

Molnar, C., Casalicchio, G. and Bischl, B. (2020a). Quantifying model complexity via functional decomposition for better post-hoc interpretability. In Machine Learning and Knowledge Discovery in Databases (pp. 193-204). Springer International Publishing.

Molnar, C, König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C., Casalicchio, G., Grosse-Wentrup, M. and Bischl, B. (2020b). Pitfalls to avoid when interpreting machine learning models. arXiv:2007.04131.

Ribeiro, M., Singh, S. and Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215.

Staniak, M. and Biecek, P. (2018). Explanations of model predictions with live and break Down Packages. *The R Journal*, 10(2), 395-409.

Strumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3), 647-665.

Szepannek, G. (2017). On the practical relevance of modern machine learning algorithms for credit scoring applications. *WIAS Report Series*, 29, 88-96.

Szepannek, G. (2022). An Overview on the landscape of R packages for open source scorecard modelling. *Risks*, 10(3), 1-33.

Szepannek G. and Luebke, K. (2021). Facing the challenges of developing fair risk scoring models. *Frontiers in Artificial Intelligence*, 4.

Szepannek G. and Aschenbruck, R. (2020). Predicting eBay prices: selecting and interpreting machine learning models – Results of the AG DANK 2018 Data Science Competition. *Archives of Data Science* A, 7(1), 1-17.

Torrent, N., Visani, G. and Enrico Bagli, E. (2020). PSD2 Explainable AI model for credit scoring. arXiv:2011.10367.

Verbraken, T., Bravo, C., Weber Richard, and Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, 238(2), 505-513.

Vincotti, V. and Hand, D. (2002). Scorecard construction with unbalanced class sizes. *Journal of the Iranian Statistical Society*, 2, 189-205.

Zhao, Q. and Hastie, T. (2019). Causal interpretations of black-box models. *Journal of Business and Economic Statistics*, 2019.

Received: June 2022, revised: October 2022

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Mária Vojtková**, *Patrik Mihalech*****

Intraday liquidity modelling using statistical methods

The correct approach to liquidity risk management in banks is essential for securing their financial stability. The position of liquidity risk among the rest of bank risks is specific because the negative outcome is not just a loss, but directly the bankruptcy of the institution. Such an occurrence might start a chain reaction and bring uncertainty into the entire financial system. This paper focused on one source of liquidity risk, i.e. management of liquidity throughout the day. The management of intraday liquidity is related to cash inflows and outflows occurring during the business day, their timing and settlement. In 2013, the BCBS published the document *Monitoring tools for intraday liquidity management*, often referred to by the regulatory authorities. It offers basic concepts of intraday liquidity monitoring and sketchily defines stress scenarios. The author suggests possibilities of how to perform intraday liquidity stress testing in a bank, which is often required by supervisors, even though no detailed approach or methodology as to how to proceed was introduced by the regulators. The research was carried out on anonymised data of cash inflows and outflows recorded on a central bank reserves account of one of the Slovak commercial banks. Both a base and four stress scenarios were developed and suggested for the better understanding of expected cashflows in standard conditions and during stress. The author's aim was to develop scenarios in a non-traditional way by means of a basic and EWMA historical bootstrap simulations, respectively. Stress scenarios are supposed to simulate reputation crisis, disruption in RTGS payment system, increased deposit outflows and bank run. The purpose of the proposed intraday liquidity monitoring scenarios was to strengthen resilience not only for a concrete bank, but also the entire financial system. Intraday liquidity monitoring is a key factor in securing stability of the financial sector.

Keywords: liquidity risk in bank, intraday liquidity, bootstrap simulation, stress testing

JEL Classification: G21, G32, C15, C83
DOI: 10.15611/aoe.2023.1.08

Quote as: Vojtková, M., Mihalech, P. (2023). Intraday liquidity modelling using statistical methods. *Argumenta Oeconomica*, 1(50), 151-178.

* Department of Statistics, Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia. ORCID: 0000-0001-6257-901X.
** Department of Statistics, Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia. ORCID: 0000-0002-1161-5874.

1. Introduction and literature

Due to their business active ties, banks are exposed to a wide range of risk. Given the fact that banks have a substantial impact on the financial sector and the national economy as a whole, they are under banking supervision in order to avoid their bankruptcy. The main goal of banking regulation is to ensure that banks hold sufficient amounts of capital necessary to cover their risk exposure (Hull, 2018). Banking regulation cannot eliminate all sources of bankruptcy, since it is not possible, but its aim is to ensure that risk exposure is reasonable, and the probability of bankruptcy is sufficiently low (Skoglund and Chen, 2015). By regulating this sector, governments seek to create a stable economic environment, where households and enterprises have confidence in the banking sector.

In commercial banks, the need for liquidity results from fact that their cashflow profile is uncertain. Banks have to make sure that they can cope with increased outflows and potentially decreased inflows in any given time, for these changes might be fully unexpected (Smolík, 1995). From the terminology point of view, one can come across the concepts of liquidity and liquidity risk. Some authors tend to use these concepts interchangeably, however, it is beneficial to distinguish between the two. Farahvash (2020), suggested that liquidity can be defined as ability to repay obligations in time of their maturity and capability to transform any asset to cash by market price. From this point of view, measuring liquidity stands for the estimation of the expected development, while the measurement of liquidity risk represents the estimation of the negative deviation from the expected development with a given probability. Liquidity risk monitoring was further elaborated by, for example Cucinelli (2013), Drehman and Nikolaou (2013), Hong et al. (2014), Ippolito (2016) and Khan (2017).

It is also necessary to distinguish between liquidity and the solvency of a financial institution (Scannella, 2016). The theoretical concept of both risks is similar but not the same. Liquidity represents the ability of a bank to manage cash outflows promptly and economically relevantly, while solvency is related to a bank's ability to repay its obligations in the long term, and is linked mostly to a sufficient amount of own funds of a bank.

The first serious attempt to unify liquidity risk management across different institutions and countries was *A framework for measuring and managing liquidity* created by the Basel Committee for Banking Supervision (1992). However, it did not succeed in the methodology definition, nor in the motivation of bank institutions to increase consistency and improve their processes in the field of liquidity risk monitoring. The BCBS introduced several definitions of liquidity risk and the development of its management and regulation, but progress in this topic was remarkably slow and inadequate to the speed of the banking industry's development. In 2006, this approach to liquidity risk management was still very notable among banks, and regulators insisted on the development of different heterogeneous models for liquidity profile evaluation (Castagna and Fede, 2013).

The crisis that began in 2007 showed that the banking sector was completely unprepared for the management of strong liquidity shocks, and the models used by banks for liquidity crisis forecasting turned out to be ineffective. In the same manner, the models applied by the regulatory institutions were also overly optimistic. The measurement and management of liquidity risk were not considered a priority among bankers, and the literature dedicated to this topic also failed to cover this as a whole, resulting in the non-existence of any integrated management process of liquidity risk (Giordana and Schumacher, 2013).

Scannella (2016) distinguished between two types of liquidity risk: funding and market liquidity risk. Market liquidity risk can be caused either by external factors (such as the condition of the financial markets) or internal factors (such as the size and structure of the financial institutions' bond portfolio). Funding liquidity risk is identified by the fact that the bank is not able to manage the expected or unexpected cash outflows effectively. In other words, this occurs when a bank cannot satisfy its obligations in time of maturity. Among the sources of funding liquidity risk one can include:

- liquidity mismatching risk – a mismatch between the size and maturity of cash inflows and outflows,
- liquidity contingency risk – future events may cause an increased need for liquidity,
- intraday liquidity risk – the inability to settle payments throughout the day and fulfill collateral requirements.

The reviewed studies and publications focus on intraday liquidity monitoring by means of parametric methods. The aim of this research was to outline the possibility of using a non-parametric simulation method. Inflows and outflows are not generated using a known or assumed distribution. This study was based on the use of an empirical distribution, which allows avoiding erroneous assumptions.

1.1. Intraday liquidity

Apart from strategic liquidity risk management over longer time horizons, banks must also deal with the availability of liquidity throughout each business day. They ought to have a sufficient amount of resources to settle all cash operations which are due on a given day and time of their maturity. The sources of intraday liquidity are highly liquid assets which are available throughout the banking day for the settlement of payments. The management of intraday liquidity needs represents a set of metrics and procedures carried out in order to secure the timely settlement of obligations (Ball et al., 2011, Farahvash 2020). Banks as intermediary institutions execute a large number of payments, which can be either cash inflows – the bank is the recipient of the cashflow/s – the bank sends money to another bank. The volume and multitude of cash inflows and outflows may significantly vary throughout the day.

An extensive number of cash outflows may lead to problems related to insufficient resources available for settlement. Managing intraday liquidity must play an important role in the risk management of a bank (Soprano, 2015).

Settlement of payments entered by clients into the banking system is realised in minimum reserve requirements (MRR) accounts, which banks must have in the central bank of a given jurisdiction; for Slovakia, this is the National Bank of Slovakia (NBS). The minimum reserve requirement is a given amount of money that banks must hold in the NBS on MRR accounts. The minimum requirements are determined for 6-week periods, based on the bank balance sheet. The required pre-set amounts of money have to be available in the central bank, on average over the maintenance period. Due to the settlement of intraday payments, banks tend to have much higher resources allocated in MRR accounts as prescribed by the regulations. Payments between banks are usually realised by means of Real-Time Gross Settlement payment systems (RTGS). In terms of the eurozone, the most important is TARGET2. One can simply say that all payments in TARGET2, which is the bank – either a sender or a recipient – are deducted or added to the MRR account. For this reason, banks tend to maintain sufficient amounts of cash in MRR accounts.

Intraday liquidity management came into the domain of the BCBS and in 2013 *Monitoring tools for intraday liquidity management* was published. This document defined the basic concepts and approaches to the measurement and management of liquidity cash flows during the day, according to which banks should perform the following activities:

- measurement of expected daily gross liquidity inflows and outflows and anticipate their timing where possible,
- monitoring of intraday liquidity positions in terms of expected activities of a bank,
- securing of sufficient funding sources to cover intraday liquidity needs,
- management of timing of cash outflows in line with intraday objectives,
- development of plan how to proceed in cases of unexpected intraday cash outflows.

Sources of intraday liquidity are used throughout the day mostly on payments realised by payment systems for correspondent banks, dedicated lines offered to clients for intraday usage and unexpected expenses related to the failure of payment transactions. Thus the BCBS introduced seven monitoring tools whose aim is to identify intraday liquidity needs (BCBS, 2013). BCBS here suggested monthly reporting of these tools to regulators, however, their monitoring by banks is expected to be carried out on a daily basis. Given that not all of these tools are applicable to all banks, they were divided into three categories. The monitoring tools and their respective categories are shown in Table 1, and the briefly described tools are discussed in this paper.

Table 1

Intraday liquidity monitoring tools

Category A: Tools applicable to all reporting banks	
A(i)	Daily maximum intraday liquidity usage
A(ii)	Available intraday liquidity at the start of the business day
A(iii)	Total payments
A(iv)	Time-specific obligations
Category B: Tools applicable to reporting banks that provide correspondent services	
B(i)	Value of payments made on behalf of correspondent banking customers
B(ii)	Intraday credit lines expended to customers
Category C: Tool applicable to reporting banks which are direct participants	
C(i)	Intraday throughput

Source: based on BCBS (2013).

Category A: Monitoring tools applicable to all reporting banks

- **A(i) – Daily maximum intraday liquidity usage**

This tool allows supervisors to monitor bank's intraday liquidity usage under standard operating conditions. The measurement is based on the net balance of all payments – sent and received during the day in all central bank accounts or with a correspondent bank. The largest net negative position occurred during the day stands for maximum intraday liquidity usage. All payments are recorded in order of settlement and maxima of liquidity usage is calculated at the end of the day. Cash flow CF_t in time t stands for inflow CF_t^{In} when cash flow is positive and outflow CF_t^{Out}, when cash flow is negative. Net cumulative liquidity position in time t can be expressed as the sum of cash flows up to time t:

$$CF_t^{Net} = \sum_{i=1}^{t} CF_i = \sum_{i=1}^{t} CF_i^{In} - \sum_{i=1}^{t} CF_i^{Out}, \tag{1}$$

$$CF_{(i)}^{In} = \begin{cases} CF_i & if \quad CF_i > 0 \\ 0 & if \quad CF_i \le 0 \end{cases}, \qquad CF_{(i)}^{Out} = \begin{cases} 0 & if \quad CF_i \ge 0 \\ -CF_i & if \quad CF_i < 0 \end{cases}.$$

Maximal usage of intraday liquidity is expressed by the largest negative net cumulative position and can be written as follows, where n stands for number of payments occurred on a given day:

$$Maximal\ liquidity\ usage = \min_{0<t\le n} \left\{ CF_t^{Net} \right\}. \tag{2}$$

- **A(iii) – Total payments**

This tool is aimed at monitoring of the bank's payment activity – sent and received payments during the day through the central bank account, or with correspondent banks. Total inflows and outflows can be expressed in the following way:

$$total\ inflow = \sum_{i=1}^{T} CF_i^{In}, \qquad total\ outflow = \sum_{i=1}^{T} CF_i^{Out}. \tag{3}$$

- **B(i) – Value of payments made on behalf of correspondent banking customers**

This tool is applicable only to banks which provide correspondent banking services to other financial institutions. It is calculated as the total amount of payments realised in the name of all correspondent banks' customers.

- **C(i) – Intraday throughput**

The last tool is relevant only for banks which are direct participants of payment systems. The goal is to measure the share of outflows in a given time period (hourly for example) to total daily outflows. Outflows Out_t^{Tot} profile up to time t can be expressed as stated below, where T stands for total number of time intervals into which daily cash flows were divided (so denominator stands for total 1 day outflow):

$$Out_t^{Tot} = \frac{\sum_{i=1}^{t} CF_i^{Out}}{\sum_{i=1}^{T} CF_i^{Out}}. \tag{4}$$

1.2. Intraday liquidity stress testing

The monitoring tools described previously, provide information about intraday liquidity profile of a bank under standard conditions. However, they say nothing about how liquidity profiles change in cases of stress situations in the financial market, respectively, in the case of stress which occurred because of unavailability of funding sources, or due to reputational risk concerning a specific bank. The BCBS states that banks and supervisors should also consider intraday liquidity requirements in stress conditions. Intraday liquidity stress testing was further elaborated for example by León (2012), Pagratis (2017), Roncalli (2021) and Liermann et al. (2021). The BCBS proposes general examples of stress scenarios, and banks are encouraged to choose which of them are most relevant for their business model. The following scenarios are proposed:

- Own financial stress – the bank encounters stress situations due to counterparties declining their payments or denying intraday credit lines. The bank faces a stress situation because it will be forced to use more sources of liquidity to prevent delaying its own payments.
- Counterparty stress – one of the major counterparties in the intraday banking payment system faces a stress situation that makes it unable to realise payments.

This results in a situation where the bank will not receive any payments from this counterparty.

- A customer bank's stress – which may result in deferring payments to customers, generating further loss of intraday liquidity.
- Market-wide credit or liquidity stress – in the case of crisis in the financial market, it might happen that the market value of high liquid assets held for intraday liquidity purposes will significantly degrade. A severe decrease in the market value or credit ranking of unencumbered liquid assets may result in the inaccessibility of intraday liquidity from the central bank because these assets might not then meet the criteria for intraday loan anymore.

However, the Basel Committee states that these scenarios serve only as an example, and banks are encouraged to define their own stress scenarios. Stress testing of intraday liquidity is often required by supervisors during the *Supervisory Review and Evaluation Process* (SREP) as well. Despite the fact that stress testing of intraday liquidity is frequently required, no direct methodology of how to proceed has been elaborated. In this paper the author introduces the possibilities of using a historical bootstrap simulation for the estimation of inflows and outflows in the standard and specific under-stress conditions.

2. Methodology

2.1. Historical bootstrap simulation

Bootstrapping is a computationally demanding technique used for the estimation of a variety of statistical metrics. Aside from standard statistical approaches where inference about population is made from a single sample, bootstrap estimate is based on random sampling with a replacement. Bootstrap belongs among a broader class of resampling methods and allows to estimate sample distributions of almost any statistic. The term 'bootstrap' was first used by B. Efron in his paper *Bootstrap methods: Another look at the jackknife* (1979). The importance of this approach started to increase with the development of information technology and the creation of specialised packages devoted to bootstrap techniques in statistical software. Among the most used packages in programming language R (used to perform computations in this paper) are the *bootstrap* package created by Efron and Tibshirani in 1993 (Efron and Tibshirani, 1993), and the package *boot*, programmed by A. J. Canty. The popularity of bootstrapping techniques has increased thanks to its high flexibility and relative simplicity (Hesterberg, 2011).

Suppose one wants to make inference about parameter θ of random variable X based on sample data $(x_1, x_2, \ldots, x_n)$ with distribution function $F(x;\theta)$. When the probability distribution of random variable X is not known, one replaces the observed sample $(x_1, x_2, \ldots, x_n)$ with a new sample obtained from the given sample by random

sampling. By following this approach, one obtains one bootstrap sample. To obtain a bootstrap estimate of parameter θ of random variable X, one proceeds in a following way (Fox and Weissberg, 2018):

- From observed values $(x_1, x_2, \ldots, x_n)$ of random sample $(X_1, X_2, \ldots, X_n)$, one calculates $\hat{\theta}$ as an estimate of parameter θ.
- Next, B random bootstrap samples are created by a replacement with sample size n from observed values $(x_1, x_2, \ldots, x_n)$. The accuracy of the estimate increases with the increasing number of bootstrap samples. A disadvantage of a large number of samples is a higher computational complexity.
- For each of the bootstrap samples one can calculate an estimate of parameter θ denoted $\hat{\theta}_i$, where $i = 1, 2, \ldots, B$.

The concept based on repeated random sampling can also be applied to an analysis of intraday cash flows. In this case, the author does not estimate a single parameter, but a process – the development of cash flows during the day. The usage of simulatin methods in the modelling of non-maturing liabilities was described by Kalkbrenner and Willing (2005) and Castagna and Fede (2013). Next, a way to apply this process to intraday cash flows is proposed. The main difference is that, while in non-maturing liabilities modelling only outflows are modelled, for intraday liquidity modelling purposes one must simulate both outflows and inflows and then create net cash flows. For modelling the study used following approach:

- First, denote time horizon T and period [0, T]. In the given case, the time horizon is one business day, and inflows and outflows are divided into hourly intervals in order to obtain estimates of cashflows on an hourly basis. The start of the business day in this dataset is at 7.00 and ends at 18.00. Thus, the time horizon is divided into 11 parts ($T = 11$).
- Next, simulate B trajectories of inflows and outflows cumulated by hours. Inflows and outflows are simulated separately, and each trajectory can be understood as one bootstrap sample.
- Calculate the expected level of cumulated inflows $In(0, T_i)$ and cumulative outflows $Out(0, T_i)$ for each step in projection $i \in \{0, 1, \ldots, T\}$ by averaging the B scenarios.
- Next, calculate the stressed level of inflows volumes on confidence level p – $In^p(0, T_i)$ and outflows $Out^p(0, T_i)$ for each projection step $i \in \{0, 1, \ldots, T\}$ by averaging the B scenarios.
- Finally, denote stress scenarios (by means of choosing confidence level p for inflows and outflows) and calculate net liquidity flows for each scenario as the difference between cumulated inflows and outflows. For liquidity risk management purposes, it is relevant to analyse increase net liquidity outflows, i.e. cash inflows will be lower as in standard conditions and outflows will be higher.

2.2. EWMA historical bootstrap simulation

In the standard bootstrap technique, each element has in every moment the same probability of being chosen into the bootstrap sample. Therefore, for each element of the sample, the probability of being selected is 1/*n* (where *n* is the total number of elements in the sample). However, this might not always be the desired state. In this case, the subject of analysis are the previously recorded inflows and outflows, and it may happen that newer observations reflect the actual situation more accurately than older ones, so it could be beneficial to choose newer cash flows into bootstrap samples more often than older ones. In this case one can assign a vector of probabilities to the selected sample, assigning each element the probability of being chosen by using the exponentially weighted moving average – EWMA (Barbe and Bertail, 1995; Hall and Maesono, 2002). Suppose one has time series of outflows and inflows recorded during a given time period (e.g. several days). One calculates cumulative inflows and inflows recorded each day divided into hourly intervals, and to each of the cash flows a weight is assigned, which denotes the probability of being chosen. In the case of EWMA historical bootstrap simulation, this weight will be exponentially decreasing as the records are older. The weights assigned to the records can be expressed as follows:

$$W_\lambda(r) = \lambda^{r-1}\frac{1-\lambda}{1-\lambda^n}, \tag{5}$$

where $W_\lambda(r)$ stands for weight assigned to *r*-th record, *r* is number of elements since most actual record to given record, λ is decay factor and *n* is number of records in bootstrap sample (sample size). Weights *W* represent the probability distribution of a record being chosen into a bootstrap sample in any step of the simulation. The sum of all weights has to be equal to 1:

$$\sum_{i=1}^{n} W_\lambda(i) = 1. \tag{6}$$

During each record selection, a stochastic process is implemented and a random number from 0 to 1 interval is picked. This number (respectively quantile) is approximated by cumulative distribution $Q(r)$.

$$Q(r) = \frac{1-\lambda^r}{1-\lambda^n}. \tag{7}$$

By inverting the function above for each random quantile, one obtains the position number *r* of the bootstrap record to be used in the simulation step. The weight function is highly dependent on the choice of parameter λ. Parameter λ must be a number higher than 0 and smaller than 1. The smaller the parameter, the more the weight function decreases, which means the increasing probability of choosing newer records instead of older ones. In the case when the parameter λ is closer to 1,

each record has the same chance of being picked and thus the basic bootstrap is obtained.

By the correct choice of λ one can determine the effective sample size. The correct sample size can be checked by *Kish's effective sample size* (Masuku and Singh, 2014), which calculates how many elements have the real probability of being chosen, and is denoted as a proportion of 1 to the sum of squared weights w_i:

$$n_{Kish} = \frac{1}{\sum_{i=1}^{n} w_i^2}. \tag{8}$$

3. Data

The analysis was performed on anonymised data from a Slovak commercial bank. Data consist in balances in one of the central bank accounts used for settlement of transactions, and was modified and cleared of some specific cash flows which occurred in a given period that might have violated the results. Additionally, some data quality issues and extreme values identified to be of non-random character were fixed and excluded from the analysis. The data were recorded from 23/2/2021 to 30/9/2021, with a gap (from 4/6 to 18/6) due to insufficient data quality. Outflows

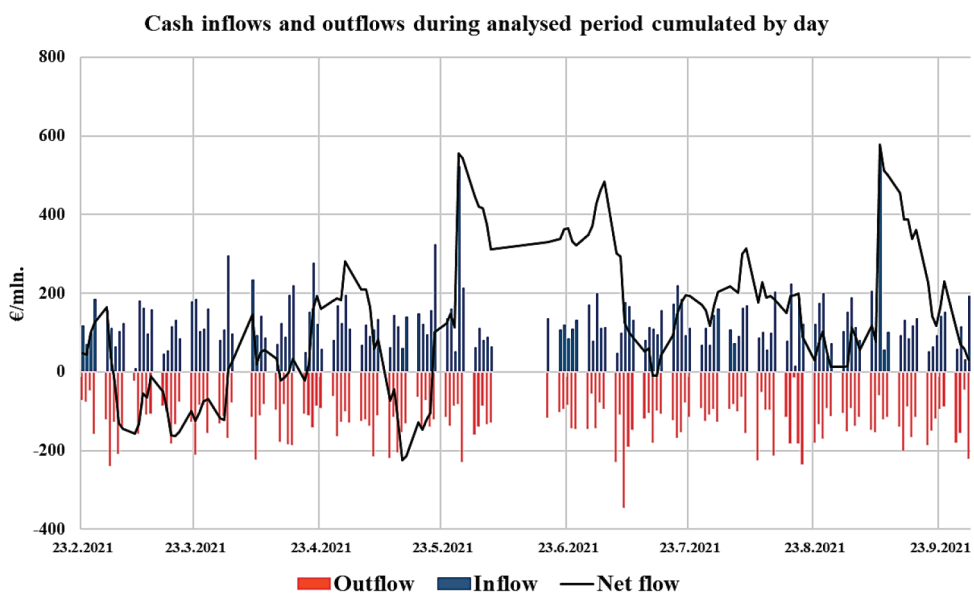


Fig. 1. Cash flows used for the analysis. Cumulative daily amounts of inflows and outflows are shown along with net cash flows.

Source: own study.

and inflows cumulated by days are shown in Figure 1. Net cash flow in time t is calculated as follows:

$$CF_t^{Net} = CF_{t-1}^{Net} + \left(CF_t^{In} - CF_t^{Out} \right), \tag{9}$$

where CF_t^{Net} stands for net cash flow in time t, CF_t^{In} is inflow in time t and CF_t^{Out} – outflow in time t.

The goal of the analysis was to elaborate a liquidity profile of cash flows by means of a historical bootstrap simulation. For this purpose, the author used inflows and outflows cumulated by an hour in each recorded day. The main objective was to create an intraday liquidity profile under normal conditions and a stressed profile which might indicate a level of outflows and inflows in the case of specific stress situations described below. The purpose of these scenarios is the better understanding of intraday cash flows and identifying the possible need to increase intraday funding. It is also a proposal for banks on how to approach intraday liquidity stress testing, which is often required by supervisor.

4. Results and discussion

4.1. Historical bootstrap simulation

Outflows CF^{Out} and inflows CF^{In} were both simulated separately, and the process can be presented in the following steps:

1. Sufficiently large numbers of simulations have to be chosen in order to obtain stable results. The number of simulations was set to 100 000, i.e. the number when the results were sufficiently stable and calculation time in R was not insufficiently long.
2. Random sampling with replacement was performed for inflows and outflows. Sampling was on hourly basis (e.g. flows for the interval 7:00-8:00 were simulated from cumulated cash flows that occurred only during this hour). Cash flows were then added up from 7:00 to 18{00 for all simulations to obtain total cash inflow and total cash outflow for the entire day. Inflow for the day can be expressed in a following way (the same stands for outflow):

$$In\left(0, T_k\right) = \sum_{i=1}^{k} CF_{In}, \quad i \in 1,2,...,k. \tag{10}$$

Given that business day starts at 7:00 and ends at 18:00, in total 11 hourly cash flows were added up. The visualised trajectories of 100 simulations are shown in Figure 2 (outflows are shown with negative operator):

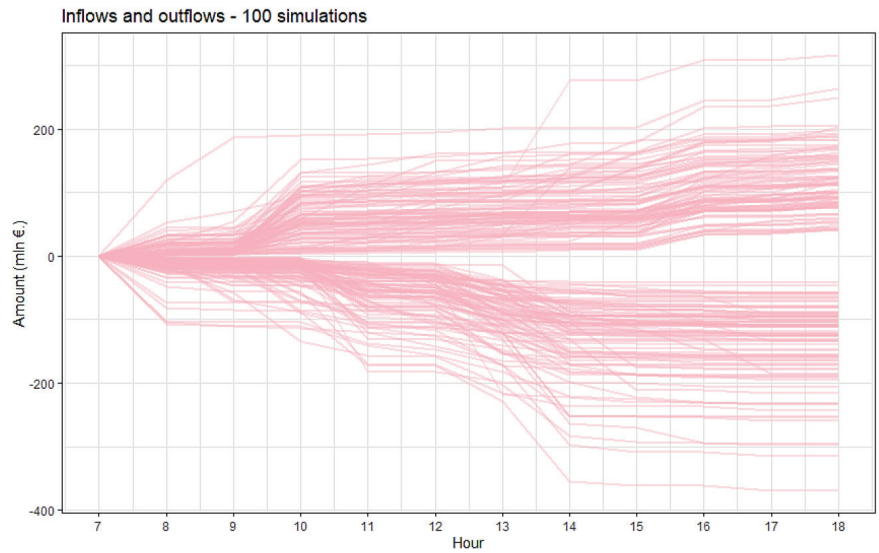


Fig. 2. 100 simulated inflows and outflows during the day grouped by hour.

Source: own elaboration.

3. Cash flows recorded in all bootstrap simulations distribution function of inflows (outflows) in a given time interval (hour). Based on this function, from all the simulations one can determine confidence level p – $In^p\left(0,T_i\right)$ / $Out^p\left(0,T_i\right)$. Thus, it can be said that with p% probability, inflows (outflows) are smaller than the given inflow (outflow). In these terms, $In^{0,5}\left(0,T_{11}\right)$ stands for the cumulative inflow recorded throughout the entire day in the middle of all the simulations lined up in ascending order.
4. After the computation of all the simulations, net cash flows were calculated as the difference between the inflows and outflows recorded on given confidence level p:

$$CF_i^{Net} = In^p\left(0,T_i\right) - Out^p\left(0,T_i\right), \quad i \in 1,2,\ldots,11\,. \tag{11}$$

Inflows and outflows computed for individual hours can be calculated as the difference between the actual and previous cash flow on given confidence level:

$$\Delta In^p\left(0,T_i\right) = In^p\left(0,T_i\right) - In^p\left(0,T_{i-1}\right), \quad i \in 2,\ldots,11. \tag{12}$$

5. Stress scenarios supposed to simulate intraday liquidity stress were then set up on a qualitative basis. For each scenario, different confidence levels were considered.

Base scenario

As a representation of the base scenario, the author chose median cash flows. Both cash inflows and outflows amount to the middle simulation in terms of volume, therefore this scenario can be expressed in the introduced terminology as $In^{0,5}\left(0,T_k\right)$

and $Out^{0,5}(0,T_k)$. This scenario represents cash flows under standard conditions. The resulting gross cash flows are shown in Figure 3 with dashed lines, and hourly inflows and outflows with bars.

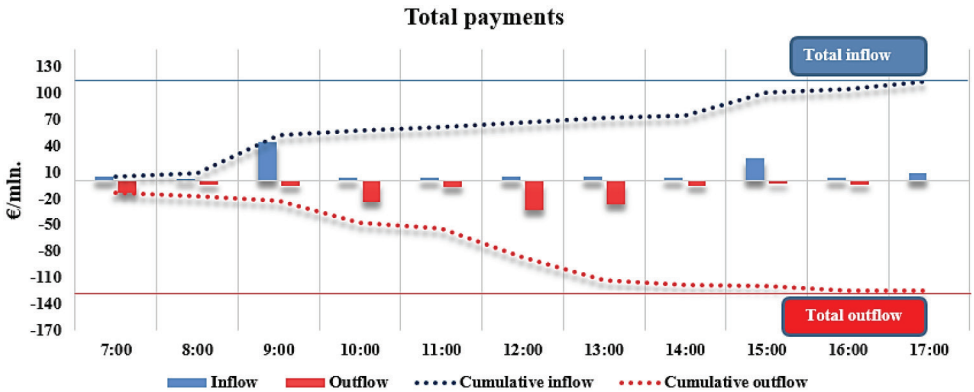


Fig. 3. Gross cash flows simulated for base scenario.

Source: own elaboration.

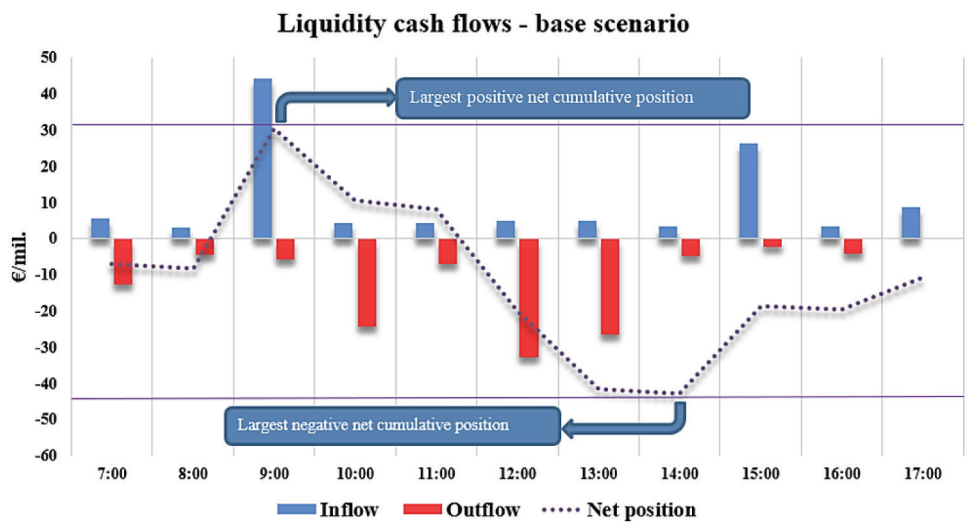


Fig. 4. Net liquidity flow – standard conditions.

Source: own elaboration.

Total cumulative inflows in the base scenario reached EUR 114 million and total cumulative outflows EUR 125 million. Total payments are one of the previously mentioned monitoring tools introduced by the BCBS, specifically A(iii) – Total payments, which allows to evaluate the expected amount of total payments. Another useful indicator is cumulative net liquidity position. The largest negative net cumulative

position stands for the amount of liquidity sources that banks use in standard conditions and which therefore must be always available. Net cash flows are shown in Figure 4. standard

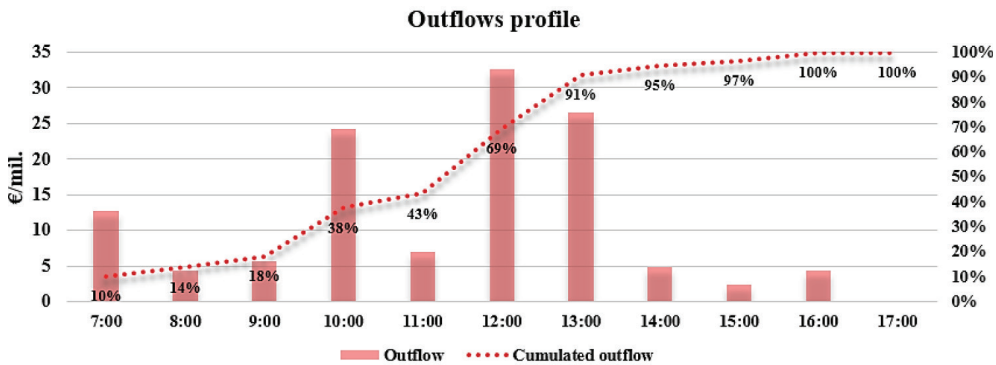


Fig. 5. Cash outflows profile – base scenario.

Source: own elaboration.

The net cumulative position is represented by the dashed line. The maximum stands for the largest positive net cumulative position and basically represents the largest amount of additional sources of liquidity available resulting from intraday cash operations. The minimum is the largest negative net cumulative position representing required sources of funding that banks use during the day. The largest positive net liquidity position in the amount of EUR 30 million was reached during the 9:00-10:00 time interval, while the largest negative liquidity position in the amount of EUR –43 million occurred in the period 14:00-15:00. The net cumulative position is related to tool A(i) – Daily maximum intraday liquidity usage. Another BCBS tool that can be calculated from simulated cash flows is C(i) – Intraday throughput, showing a profile of cash outflows recorded on an hourly basis as a proportion to total outflows during the day. This shows which hour requires the highest liquidity for settlement of payments. The outflows profile is shown in Figure 5; on the left y-axis is the amount of outflows in million EUR, while on the right y-axis there are relative cumulative outflows up to a given hour. The largest outflow in standard conditions occurs during the period 12:00-13:00 in the amount of EUR 33 million. It is worth noting that the majority of standard outflows happen up to 14:00, i.e. 91%. This means that outflows in the morning and early afternoon hours are highest, and liquidity required in later hours is not that high.

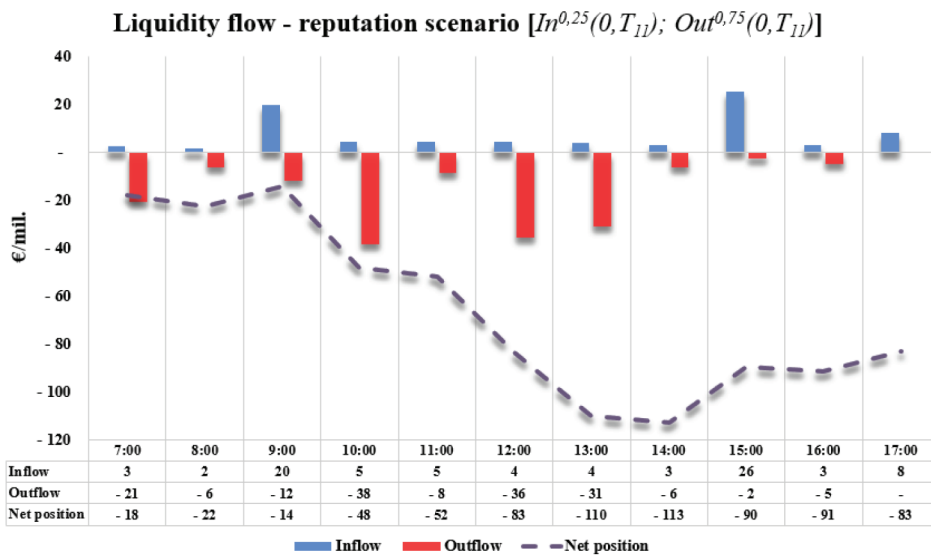
Stress scenarios

The benefit of the base scenario is that it helps with understanding cash flows' behaviour in standard conditions. This section defines several stress scenarios, where cash flows should reflect the occurrence of a non-standard event. All these scenarios

are specified on a qualitative basis by determining cash flow quantiles from a historical bootstrap simulation. It is necessary to point out that in the case of a real stress situation, historical data might fail to forecast the correct outcomes, therefore the results are just a quantification of the estimate. Four stress scenarios were developed:

1. Reputation crisis,
2. Disruption in RTGS payment system,
3. Increased deposit outflows,
4. Black scenario – bank run.

The first scenario is supposed to reflect a reputation crisis. In cases when the bank is exposed to reputation risk, e.g. some negative information about the bank's ability to repay its obligations is shown in the media, even if this information is false it tends to influence clients' behaviour and they might withdraw their money from the bank. A decreased amount of inflows might also be expected, because clients will avoid sending money to this bank and redirect their cash flows elsewhere. For this scenario, simulations $In^{0,25}(0,T_{11})$ and $Out^{0,75}(0,T_{11})$ were carried out. Cash flows for each hour will be at the level of 25% simulations with the lowest inflows, and outflows at the level of 75% simulations with the lowest outflows (this designation is used for all upcoming scenarios).



	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
Inflow	3	2	20	5	5	4	4	3	26	3	8
Outflow	- 21	- 6	- 12	- 38	- 8	- 36	- 31	- 6	- 2	- 5	-
Net position	- 18	- 22	- 14	- 48	- 52	- 83	- 110	- 113	- 90	- 91	- 83

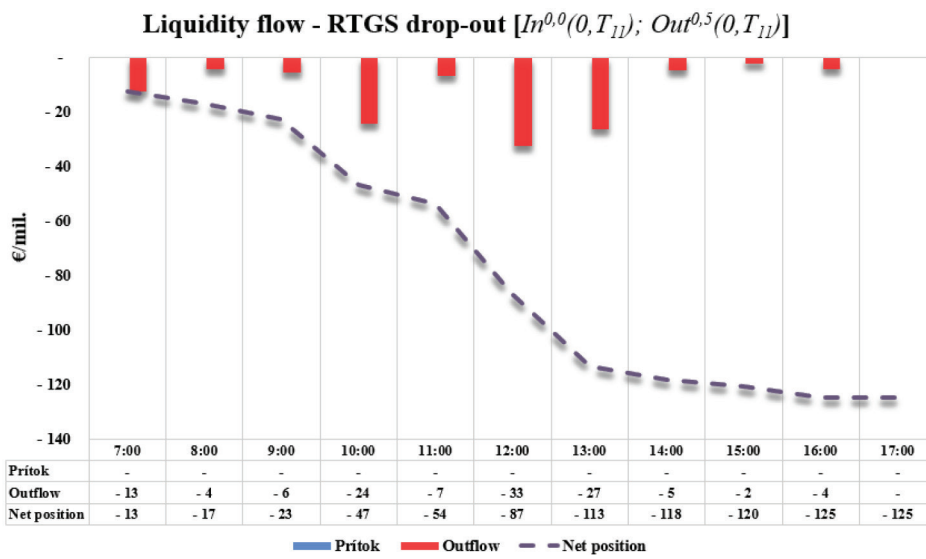
Fig. 6. Intraday liquidity flows in reputation stress scenario.

Source: own elaboration.

As expected, this change has quite a significant impact on net liquidity position and total payments; net position and cash flows are shown in Figure 8. Unlike in the base scenario, where net position is close to zero (outflows and inflows are close to

equal at the end of the day), in a reputation crisis scenario outflows significantly exceed inflows. This impacts on net liquidity flows, which at the end of the day amount to EUR –90 million. The largest net negative cumulative position is EUR –113 million during the 14th hour, and reflects the amount of liquidity sources that must be available in the bank to cover all the obligations due during the day. The cumulated outflow reached EUR 165 million, and the cumulated inflow EUR 83 million.

The second scenario was labelled as a disruption of the RTGS payment system. This scenario simulates a situation of an unexpected failure in the payment system in the case of a hacker attack, or due to other external impacts that cause situation when a bank is unable to accept payments from other banks. Outgoing payments will be working without restrictions and this scenario can be defined in line with the former designation as $In^{0}(0,T_{11})$ and $Out^{0,5}(0,T_{11})$. In other words, the scenario simulates a situation when inflows are zero and outflows are standard. In this case, net liquidity outflow will be equal to cumulated outflow from the base scenario and reach the highest negative net cumulative position at the end of the day amounting to EUR 125 million.



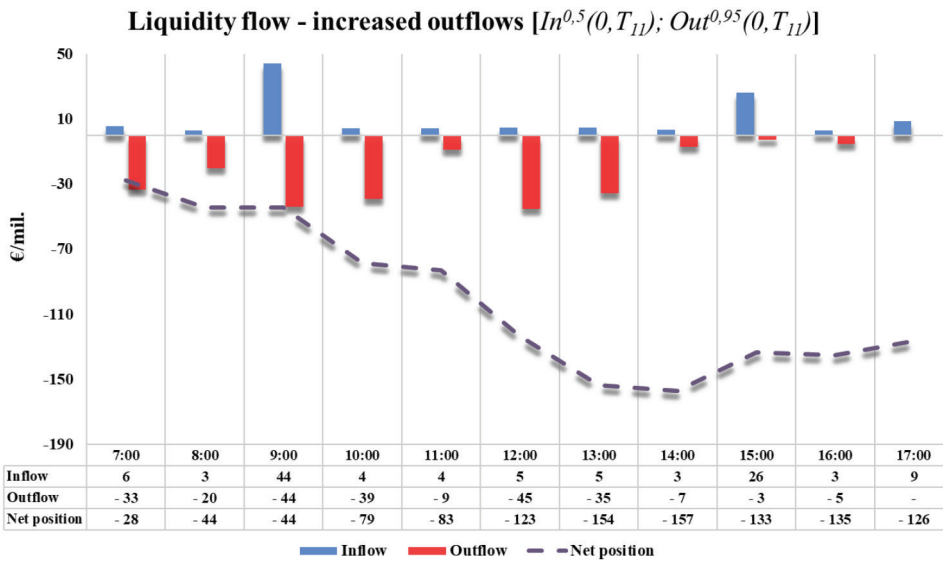
	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
Prítok	-	-	-	-	-	-	-	-	-	-	-
Outflow	- 13	- 4	- 6	- 24	- 7	- 33	- 27	- 5	- 2	- 4	-
Net position	- 13	- 17	- 23	- 47	- 54	- 87	- 113	- 118	- 120	- 125	- 125

Fig. 7. Intraday liquidity flows in RTGS drop-out scenario.

Source: own elaboration.

The third scenario simulates increased outflows due to a higher withdrawal rate from deposit accounts, not necessarily because the situation of mass withdrawals is due to some bank specific crisis, but just a higher level of standard outflows caused by the tax due date or a similar event. Nowadays, corporate clients withdraw their

funds at a higher rate to pay off taxes and this might impact liquidity position. In addition, the business day before a holiday might record a higher outflow rate from retail customers. This scenario should reflect such occurrences, and scenario is designed as $In^{0,5}\left(0,T_{11}\right)$ and $Out^{0,95}\left(0,T_{11}\right)$ meaning standard inflows and higher outflows. In this case, the net liquidity position at the end of the day reached EUR −126 million, and the largest negative net cumulative position EUR −157 million over the period 14:00-15:00. Total outflows were EUR 240 million and inflows remained unchanged to the base scenario in the amount of EUR 114 million.



	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
Inflow	6	3	44	4	4	5	5	3	26	3	9
Outflow	- 33	- 20	- 44	- 39	- 9	- 45	- 35	- 7	- 3	- 5	-
Net position	- 28	- 44	- 44	- 79	- 83	- 123	- 154	- 157	- 133	- 135	- 126

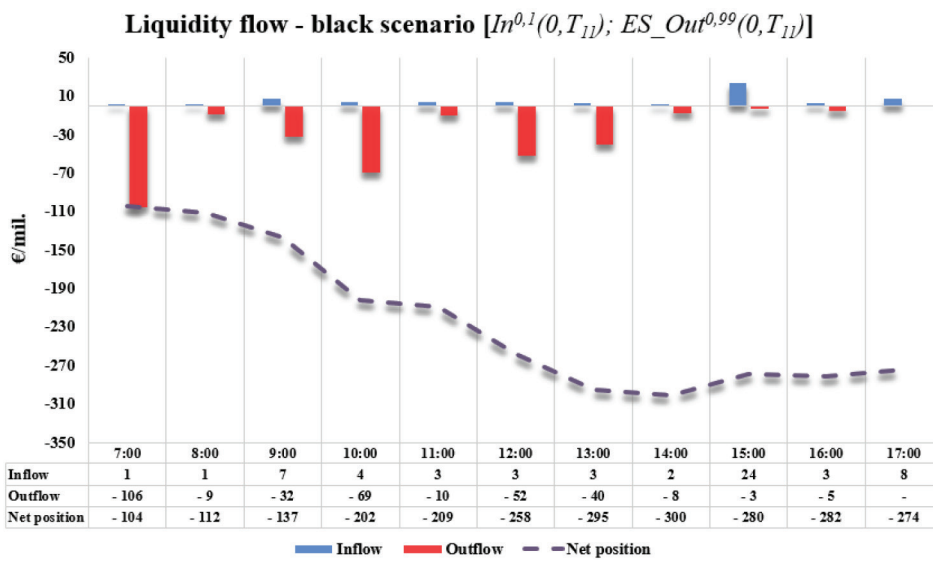
Fig. 8. Intraday liquidity flows in increased deposit outflows scenario.

Source: own elaboration.

The fourth and the final defined scenario is labelled the black scenario. All the aforementioned scenarios are based on the quantile value of a historical bootstrap simulation. Given that cash inflows and outflows are strongly positively skewed, outflows recorded at tails of distributions might be significantly higher than related quantile value. For this reason, the author introduced conditional value-at-risk (also referred to as expected shortfall – ES), which can be interpreted as expected loss from the values exceeding a given quantile. The study applied this metric on cash outflows so that conditional value at risk stands for average outflow from outflows that exceed a given percentage of simulations. To project the study's designation into this methodology, $Out^{0,95}\left(0,T_{11}\right)$ could be labelled as value at risk – $VaR\_Out\left(0.95\right)$. In order to calculate average outflow from simulations with higher outflows than 95% of other simulations, there would be conditional value at risk, in other words

$ES\_Out(0.95)$. Expected shortfall is also a coherent risk measure as it satisfies the sub-additivity property, unlike standard *VaR* (Horáková, 2015).

All the previously mentioned scenarios were intended to deal with a specific field of increased liquidity needs during the day. However, the need for liquidity might be significantly higher in the case of a bank run, i.e. a situation, when clients, due to some reason – most often a reputational problem, start to withdraw all their deposits from the bank. It should be noted that for the rest of the risks a negative outcome is loss, in the case of inadequate liquidity needs it is bankruptcy. Therefore, a crisis related to a bank run is one of the most severe situations banks can face, at the same time it is also stressed that this might have no clear cause. Sometimes even an incorrect interpretation of some steps carried out by banks can cause that clients commence mass withdrawals of their funds. When other clients recognize this behaviour, they tend to panic as well, and also withdraw their deposits. This triggers off a withdrawal spiral when everyone removes their funds from the bank, and even when the bank is otherwise financially healthy it might face severe difficulties to withstand the crisis.



	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
Inflow	1	1	7	4	3	3	3	2	24	3	8
Outflow	- 106	- 9	- 32	- 69	- 10	- 52	- 40	- 8	- 3	- 5	-
Net position	- 104	- 112	- 137	- 202	- 209	- 258	- 295	- 300	- 280	- 282	- 274

Fig. 9. Liquidity flow – black scenario (bank run)

Source: own elaboration.

In the case of a bank run, inflows are expected to decrease and outflows strongly increase. For this scenario the author proposed $In^{0,1}(0,T_{11})$ and $ES\_Out^{0,99}(0,T_{11})$ cash flows, reflected by 10% of lowest inflows and outflows being the average from 1% simulations with the highest outflows. With these assumptions, simulated outflows reach EUR 334 million, and inflows EUR 60 million. The net cumulative

liquidity position at the end of the day is EUR –274 million, and the lowest recorded during period 14:00-15:00 amounted to EUR –300 million. In comparison to other scenarios, this time the liquidity needs are much higher. Here it is necessary to note that the study did not have historical data with a bank run included (as it is for most of the banks), and therefore predicting client behaviour is difficult. The usage of historical simulation as shown here may serve as the basis for some expert judgement adjustments to the model.

4.2. EWMA historical bootstrap simulation

As already noted, in a standard bootstrap each element has the same probability of being chosen for simulation at any step of the simulation. Bearing in mind the fact that the author chose records of different age, it might not be the best approach. Older data have the same probability of being selected as the newer, even though the latter might predict the current situation more credibly. This is valid specifically in cases of long time series that were used as entry data for bootstrap simulation. With only roughly 7 months of observations being available, this should not be the case, however the author carried out the EWMA simulation for the purposes of comparison purposes on these dataset as well.

First, one must choose the value for parameter lambda and calculate weights $W_\lambda(n)$ for all the records in the sample. For the purposes of the EWMA bootstrap simulation, it was decided to apply an effective sample size of 3 months, bearing in mind that the highest probability of being chosen is linked to the newest record, and that the probability is exponentially decreasing when going further back into the past. The author chose a parameter lambda equal to 0.9677, and the correctness of the chosen value can be verified by Kish's effective sample size. In relation to the sample (observations from 147 days), an effective sample size with the use of $\lambda = 0.9677$ is equal to 60, what can be considered the average amount of business days for three months, and thus parameter λ was chosen correctly.

$$n_{Kish} = \frac{1}{\sum_{i=1}^{n} w_i^2} \rightarrow \frac{1}{0.0167} = 59.95076\,.$$

By calculating of the weight's cumulative distribution using Kish's effective sample size of 60 and lambda 0.9677, one obtains the value of 0.8675. This value can be interpreted in a way that a randomly chosen cash flow in any step of simulation has approximately 87% probability of not being older than 3 months.

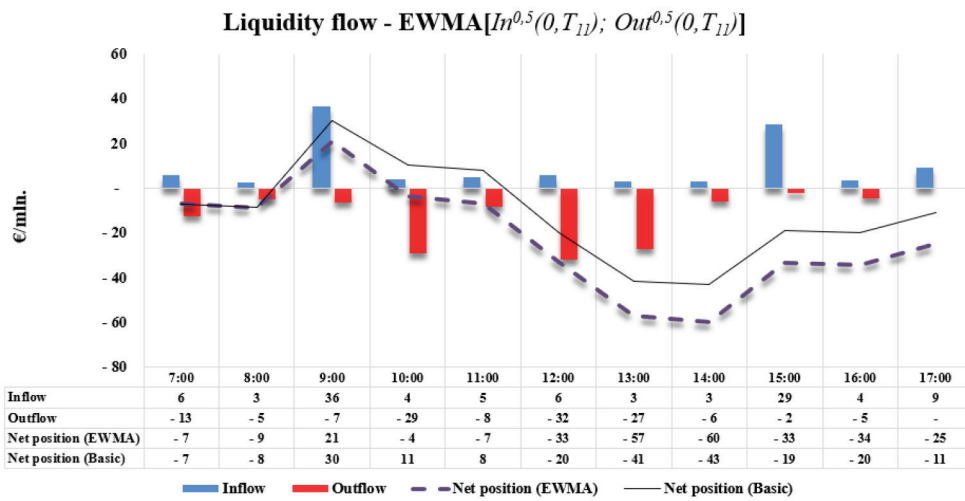
$$Q(r) = \frac{1-\lambda^r}{1-\lambda^n} \rightarrow \frac{1-0.9677^{60}}{1-0.9677^{147}} = 0.8675.$$

Due to the high number of simulations, one can also say that cash flows from the last three months make up 88% of all cash flows selected into the simulation. This results in newest cash flows having a higher impact on the calculated net flows than

older ones; however, older records are not completely excluded – just their impact is smaller. The meaning of EWMA bootstrap simulation increases in situations when cash flows behaviour changes in time, e.g. when the bank faces a recent crisis. This crisis is then reflected in an EWMA simulation with a higher impact than in the basic one.

Basic scenario EWMA

First, simulate the EWMA basic scenario. All the inputs remain the same, only the probability for each record to be chosen is different and determined by the weights function. The cash flows simulated by the EWMA methodology are shown in Figure 10 (net flows from the basic simulation are also shown for comparison). The net liquidity flow is in this case EUR –25 million, and the largest negative net cumulative position is EUR 60 million during the period 14:00-15:00. The net liquidity position is worse by EUR 14 million, in comparison to the basic bootstrap methodology. One can conclude that there was a slight worsening of net liquidity positions in the EWMA simulation, due to higher outflows in the last months shown in the sample. Total inflows reached EUR 108 million (+6 m EUR) and total outflows were EUR 133 million (+8 m EUR). Changes in the outflows profile were minimal.



	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00
Inflow	6	3	36	4	5	6	3	3	29	4	9
Outflow	- 13	- 5	- 7	- 29	- 8	- 32	- 27	- 6	- 2	- 5	-
Net position (EWMA)	- 7	- 9	21	- 4	- 7	- 33	- 57	- 60	- 33	- 34	- 25
Net position (Basic)	- 7	- 8	30	11	8	- 20	- 41	- 43	- 19	- 20	- 11

Fig. 10. Cash flows – EWMA basic scenario.

Source: own elaboration.

Stress scenarios EWMA and comparison to basic bootstrap

In stress scenarios, the differences between basic and EWMA simulation are negligible, which confirms that no period of recent stress was visible in the data. Due to slight differences between the basic and EWMA stress scenarios, the author introduced only a net liquidity flows comparison. The biggest difference was found

during the first few hours in the third scenario related to increased outflows. In the first hour of a working day this difference reached EUR 43 million, and in the second EUR 24 million. The difference in net flows at the end of the day, however, was almost zero. The first and the second scenarios (reputation and RTGS failure, respectively) yielded better results for the basic simulation and a slight worsening was observed in the case of EWMA. In general, it can be concluded that there was not a very significant difference between both approaches, which was expected – given that the dataset was relatively small, and no period of stress occurred in the underlying data. A comparison of all the scenarios is shown in Figure 11. The numbering of the scenarios follows the order defined above.

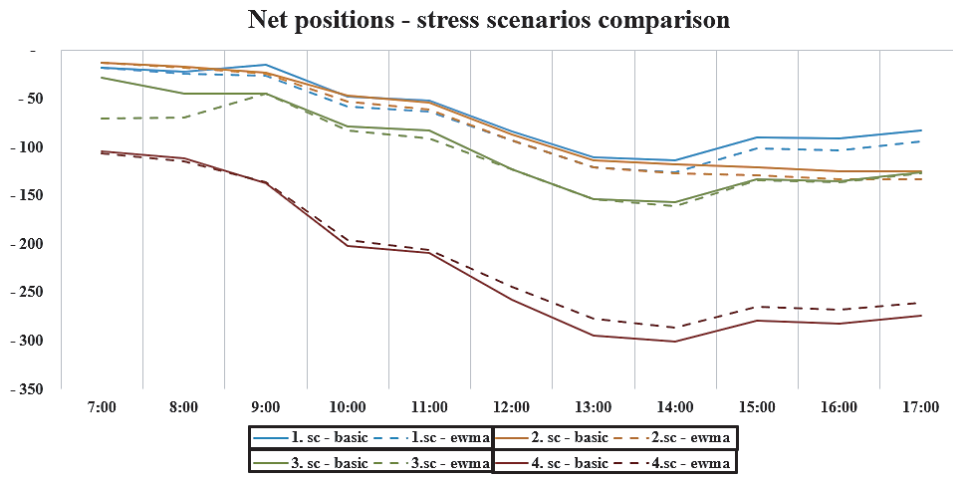


Fig. 11. Comparison of basic and EWMA simulated net positions for stress scenarios.

Source: own elaboration.

For the black scenario, the results of the EWMA simulation are better than for basic one (net outflows at the end of the day amounted to EUR –287 million for the EWMA simulation, and EUR –287 million for the basic one).

As previously mentioned, a comparison of the basic and EWMA bootstrap serves mostly as an example. EWMA would be most appropriate when possessing data recorded in the past several years; in this case one might want to increase the effective sample size to e.g. one year. From the long-term point of view, it is also necessary to stress that cash flows have a notional value. If data were taken from several years, the real value of older cash flows might be different from the notional one in that time. This is important with respect to the actual trend of increased inflation. If inflation persisted, cash flows recorded before this period might look small in comparison to actual amounts, however, in time of their realisation they might appear higher. This might lead to an underestimation of stress outflows in current conditions. For this reason, in the case of the biggest dataset it might be beneficial not only to use

Table 2

Comparison of scenarios. Base and EWMA simulation (million EUR)

Scenario	Cumulated inflow	Cumulated outflow	Net position end of day	Net position Max	Net position Min
Basic	113.78	-124.62	-10.84	30.43 / 9:00	-42.77 / 14:00
Basic EWMA	108.43	-133.16	-24.73	21.29 / 9:00	-59.57 / 14:00
1 sc. In_0,25/Out_0,75	82.17	-165.16	-82.99	0 / 7:00	-113.04 / 14:00
1 sc. In_0,25/Out_0,75 EWMA	76.26	-170.45	-94.20	0 / 7:00	-125.76 / 14:00
2 sc. In_0/Out_0,5	0.00	-124.62	-124.62	0 / 7:00	-124.62 / 17:00
2 sc. In_0/Out_0,5 EWMA	0.00	-133.16	-133.16	0 / 7:00	-133.16 / 17:00
3 sc. In_0,5/Out_0,95	113.78	-240.10	-126.32	0 / 7:00	-157.03 / 14:00
3 sc. In_0,5/Out_0,95 EWMA	108.18	-234.88	-126.71	0 / 7:00	-160.39 / 14:00
4 sc. In_0,1/Out_ES_0,99	60.22	-334.30	-274.07	0 / 7:00	-300.49 / 14:00
4 sc. In_0,1/Out_ES_0,99 EWMA	56.97	-317.61	-260.64	0 / 7:00	-286.83 / 14:00

Source: own elaboration.

the EWMA bootstrap but also to recalculate the cash flows to their actual fair value. The final comparison of all the scenarios and their cumulative cash flows and net positions is presented in Table 2.

Conclusion

Liquidity risk is one of the major banking risks, especially bearing in mind that the results of the liquidity crisis might be severe not only for the bank itself, but could also spread through the entire financial system. This paper focused on one particular part of liquidity risk management in commercial banks – intraday liquidity cash flows management. The research was based on *Monitoring tools for intraday liquidity management* framework (BCBS, 2013). The Basel Committee also encouraged banks to perform stress testing of intraday liquidity, however, no detailed approach on how to do that was suggested. This research focused on providing a relatively straightforward and easily repeatable solution of doing that by means of a historical bootstrap simulation.

The approach was introduced on the anonymised data of bank inflows and outflows occurring during the day, grouped by hour as recorded from February to October 2021 in a commercial bank operating in the Slovak Republic. Four stressed scenarios were suggested, however they served only as an example, and other scenarios can be developed in the same manner. The biggest limitation of this solution is, naturally, relying on historical data. In the case of a real stress situation, there is no guarantee that cash flows would behave in the same way.

Cash flows were cumulated by hour, and the entire simulation was carried out on an hourly basis, however different time intervals can also be considered as well. With the hourly approach, stress testing could be performed at any given hour of the day

(not just from the start of business hours at 7:00, but e.g. at 12:00). Through this approach one could stress cash flows for only the remaining part of the day. For example, by starting the prediction at 14:00, most of the outflows would have occurred up to this hour (91%), and therefore the resulting net flows might not be as severe as in a full-day simulation.

This paper also deals with cash flows in their notional amount. This information itself does not directly reveal if the amount is high or not. All inflows and outflows occur in the bank's current account of obligatory reserves in the central bank. For a determination of the severity of outflows in stress scenarios it might be beneficial to compare their amount to the average level of the bank's reserves in the central bank. If this ratio is high, it means that banks rely on a bigger portion of their reserves for intraday payments purposes and an increase of these outflows might cause problems for the banks to cover them promptly. Vice-versa, lower percentage ratio signals that banks use only a small portion of their reserves on intraday transactions and their liquidity reserves are at a sufficient level.

A limitation of the research was also the fact that the author had access only to information about cash flows but no information as to where the outflows go or from which bank the inflows come from. Another suggestion for stress scenarios could be a counterparty or a country-specific stress focus on countries from which the most inflows usually come, and which carry out bigger transactions. Additionally, information whether cash flow is related to retail, corporate or treasury operation might be of interest and provide additional insights into cash flows structure. To obtain this information, further research of this issue is planned in order to understand the intraday liquidity position as well as possible.

References

Ball, A., Denbee, E., Manning, M., and Wetherilt, A. (2011). Intraday Liquidity: Risk and Regulation. *Bank of England Financial Stability Paper*, 11.

Barbe, P., and Bertail, P. (1995). The Weighted Bootstrap. Springer.

BCBS (2013). Monitoring tools for intraday liquidity management. Basel Committee on Banking Supervision. Retrieved from: https://www.bis.org/publ/bcbs248.pdf.

Castagna, A., and Fede, F. (2013). Measuring and Managing Liquidity Risk. Wiley 1st Edition.

Cucinelli, D. (2013). The Determinants of Bank Liquidity Risk within the Context of Euro Area. *Interdisciplinary Journal of Research in Business*, 2(10), 51–64.

Drehman, M., and Nikolaou, K. (2013). Funding liquidity risk: Definition and measurement. *Journal of Banking & Finance*, 37(7), 2173–2182.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. Springer.

Efron, B., and Tibshirani, R. (1993). An Introduction to the Bootstrap. CRC Press 1st Edition.

Farahvash, P. (2020). Asset-Liability and Liquidity Management. Wiley 1st Edition.

Fox, J., and Weisberg, S. (2018). R Companion to Applied Regression: Bootstrapping Regression Models. R. SAGE publications, 3rd Edition.

Giordana, G., and Schumacher, I. (2013). Bank liquidity risk and monetary policy. Empirical evidence on the impact of Basel III liquidity standards. *International Review of Applied Economics*, 27(5).

Hall, P., and Maesono, Y. (2002). A weighted bootstrap approach to bootstrap iteration. *Series B: Statistical Methodology*, 62(1), 137–144.

Hesterberg, T. (2011). Bootstrap. *Interdisciplinary review: Computational statistics*, 3(6), 497–526.

Hong, H., Huang, J.-Zh., and Wu, D. (2014). The information content of Basel III liquidity risk measures. *Journal of Financial Stability*, 15, 91–111.

Horáková, G., Páleš, M., and Slaninka, F. (2015). Teoria rizika v poistení. Wolters Kluwer, prvé vydanie.

Hull, J. (2018). Risk Management and Financial Institutions. Wiley 5th Edition.

Ippolito, F., Peydró, J.-L., Polo, A., and Sette, E. (2016). Double bank runs and liquidity risk management. *Journal of Financial Economics*, 122(1), 135–154.

Kalkbrenner, M., and Willing, J. (2004). Risk management of non-maturing liabilities. *Journal of Banking & Finance*, 28(7), 1547–1568.

Khan, M., Scheule, H., and Wu, E. (2017). Funding liquidity and bank risk taking. *Journal of Banking & Finance*, 82, 203–216.

León, C. (2012). Estimating Financial Institutions' Intraday Liquidity Risk: A Monte Carlo Simulation Approach. *Borradores de Economia, SSRN Electronic Journal*.

Liermann, V., Li, S., and Steuer, R. (2021). Kafka – A Practical Implementation of Intraday Liquidity Risk Management. *The Digital Journey of Banking and Insurance*, 3(1), 105–116.

Masuku, M., and Singh, A. (2014). Sampling Techniques and Determination of Sample Size in Applied Statistics Research: An Overview. *International Journal of Economics, Commerce and Management*, 2(11),1–22.

Pagratis, S., Topaloglou, N., and Tsionas, M. (2017). System stress testing of bank liquidity risk. *Journal of International Money and Finance*, 73, Part A, 22–40.

Roncalli, T. (2021). Liquidity Stress Testing in Asset Management – Part 3. Managing the Asset-Liability Risk.

Scannella, E. (2016). Theory and Regulation of Liquidity Risk Management in Banking. *International Journal of Risk Assessment and Management*, 19(1/2), 4–21.

Skoglund, J. and Chen, W. (2015). Financial Risk Management: Application in Market, Credit, Asset and Liability Management and Firmwide Risk. Wiley.

Smolík, M. (1995). Manažment bankových aktív a pasív. ELITA.

Soprano, A. (2015). Liquidity Management: A Funding Risk Handbook. Wiley.

Received: June 2022, revised: August 2022

Acknowledgements: *This research was supported by the Slovak VEGA Grant No. 1/0561/21 "The impact of the COVID-19 crisis on business demography and employment in the Slovak Republic and the EU".*

APPENDIX

Results of simulations (million EUR)

Base scenario (basic bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	5,614	-12,662	-7,047	5,614	-12,662	10%
8:00	3,144	-4,330	-8,233	8,758	-16,992	14%
9:00	44,339	-5,678	30,428	53,098	-22,669	18%
10:00	4,326	-24,197	10,557	57,423	-46,866	38%
11:00	4,456	-7,045	7,968	61,879	-53,911	43%
12:00	4,977	-32,686	-19,741	66,857	-86,598	69%
13:00	4,854	-26,575	-41,461	71,711	-113,172	91%
14:00	3,498	-4,807	-42,771	75,209	-117,980	95%
15:00	26,486	-2,354	-18,639	101,695	-120,333	97%
16:00	3,307	-4,283	-19,615	105,002	-124,617	100%
17:00	8,778	0,000	-10,837	113,780	-124,617	100%

1. stress scenario – reputational crisis (basic bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	2,670	-20,515	-17,845	2,670	-20,515	12%
8:00	1,746	-6,258	-22,357	4,416	-26,773	16%
9:00	19,886	-11,842	-14,313	24,302	-38,615	23%
10:00	4,694	-38,447	-48,066	28,996	-77,062	47%
11:00	4,535	-8,463	-51,994	33,530	-85,525	52%
12:00	4,410	-35,694	-83,278	37,941	-121,218	73%
13:00	4,215	-30,886	-109,949	42,156	-152,105	92%
14:00	2,934	-6,024	-113,039	45,090	-158,129	96%
15:00	25,704	-2,257	-89,593	70,793	-160,386	97%
16:00	3,112	-4,777	-91,258	73,905	-165,163	100%
17:00	8,269	0,000	-82,989	82,174	-165,163	100%

2. stress scenario – RTGS drop-out (basic bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	0,000	-12,662	-12,662	0,000	-12,662	10%
8:00	0,000	-4,330	-16,992	0,000	-16,992	14%
9:00	0,000	-5,678	-22,669	0,000	-22,669	18%
10:00	0,000	-24,197	-46,866	0,000	-46,866	38%
11:00	0,000	-7,045	-53,911	0,000	-53,911	43%
12:00	0,000	-32,686	-86,598	0,000	-86,598	69%
13:00	0,000	-26,575	-113,172	0,000	-113,172	91%
14:00	0,000	-4,807	-117,980	0,000	-117,980	95%
15:00	0,000	-2,354	-120,333	0,000	-120,333	97%
16:00	0,000	-4,283	-124,617	0,000	-124,617	100%
17:00	0,000	0,000	-124,617	0,000	-124,617	100%

3. stress scenario – increased outflows (basic bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	5,614	-33,157	-27,543	5,614	-33,157	14%
8:00	3,144	-20,037	-44,435	8,758	-53,194	22%
9:00	44,339	-44,017	-44,113	53,098	-97,211	40%
10:00	4,326	-38,897	-78,684	57,423	-136,107	57%
11:00	4,456	-8,795	-83,023	61,879	-144,903	60%
12:00	4,977	-45,036	-123,082	66,857	-189,939	79%
13:00	4,854	-35,356	-153,584	71,711	-225,295	94%
14:00	3,498	-6,945	-157,031	75,209	-232,240	97%
15:00	26,486	-2,597	-133,143	101,695	-234,837	98%
16:00	3,307	-5,266	-135,101	105,002	-240,103	100%
17:00	8,778	0,000	-126,323	113,780	-240,103	100%

4. stress scenario – bank run (basic bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	1,124	-105,512	-104,388	1,124	-105,512	32%
8:00	1,431	-8,898	-111,855	2,555	-114,410	34%
9:00	7,416	-32,482	-136,921	9,971	-146,892	44%
10:00	3,815	-69,070	-202,176	13,785	-215,962	65%
11:00	3,470	-9,985	-208,692	17,255	-225,947	68%
12:00	3,382	-52,329	-257,639	20,637	-278,276	83%
13:00	3,256	-40,472	-294,855	23,893	-318,749	95%
14:00	1,895	-7,534	-300,494	25,788	-326,282	98%
15:00	23,989	-3,035	-279,541	49,777	-329,318	99%
16:00	2,675	-4,980	-281,845	52,452	-334,297	100%
17:00	7,771	0,000	-274,074	60,223	-334,297	100%

Base scenario (EWMA bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	5,865	-12,662	-6,796	5,865	-12,662	10%
8:00	2,827	-4,679	-8,648	8,692	-17,340	13%
9:00	36,469	-6,535	21,286	45,162	-23,875	18%
10:00	4,168	-29,021	-3,568	49,329	-52,897	40%
11:00	4,955	-8,355	-6,968	54,284	-61,252	46%
12:00	6,002	-31,956	-32,921	60,286	-93,208	70%
13:00	3,258	-27,184	-56,848	63,544	-120,392	90%
14:00	3,223	-5,950	-59,575	66,767	-126,342	95%
15:00	28,562	-2,220	-33,233	95,329	-128,562	97%
16:00	3,789	-4,593	-34,038	99,118	-133,155	100%
17:00	9,309	0,000	-24,729	108,426	-133,155	100%

1. stress scenario – reputational crisis (EWMA bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	3,059	-20,515	-17,456	3,059	-20,515	12%
8:00	1,281	-7,969	-24,144	4,340	-28,484	17%
9:00	13,881	-16,240	-26,503	18,222	-44,724	26%
10:00	4,100	-35,412	-57,814	22,322	-80,136	47%
11:00	4,758	-10,263	-63,319	27,080	-90,399	53%
12:00	4,794	-34,527	-93,052	31,874	-124,926	73%
13:00	2,916	-30,961	-121,097	34,791	-155,887	91%
14:00	2,645	-7,307	-125,758	37,436	-163,194	96%
15:00	26,749	-2,421	-101,430	64,185	-165,615	97%
16:00	3,468	-4,837	-102,799	67,653	-170,452	100%
17:00	8,604	0,000	-94,195	76,257	-170,452	100%

2. stress scenario – RTGS drop-out (EWMA bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	0,000	-12,662	-12,662	0,000	-12,662	10%
8:00	0,000	-4,679	-17,340	0,000	-17,340	13%
9:00	0,000	-6,535	-23,875	0,000	-23,875	18%
10:00	0,000	-29,021	-52,897	0,000	-52,897	40%
11:00	0,000	-8,355	-61,252	0,000	-61,252	46%
12:00	0,000	-31,956	-93,208	0,000	-93,208	70%
13:00	0,000	-27,184	-120,392	0,000	-120,392	90%
14:00	0,000	-5,950	-126,342	0,000	-126,342	95%
15:00	0,000	-2,220	-128,562	0,000	-128,562	97%
16:00	0,000	-4,593	-133,155	0,000	-133,155	100%
17:00	0,000	0,000	-133,155	0,000	-133,155	100%

3. stress scenario – increased outflows (EWMA bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	5,614	-75,707	-70,093	5,614	-75,707	32%
8:00	2,827	-1,474	-68,740	8,441	-77,181	33%
9:00	36,469	-12,492	-44,763	44,910	-89,673	38%
10:00	4,168	-42,483	-83,079	49,078	-132,156	56%
11:00	4,955	-12,316	-90,439	54,033	-144,472	62%
12:00	6,002	-38,465	-122,901	60,035	-182,937	78%
13:00	3,258	-34,042	-153,685	63,293	-216,978	92%
14:00	3,223	-9,926	-160,388	66,516	-226,904	97%
15:00	28,562	-2,596	-134,423	95,078	-229,500	98%
16:00	3,789	-5,383	-136,017	98,866	-234,883	100%
17:00	9,309	0,000	-126,708	108,175	-234,883	100%

4. stress scenario – bank run (EWMA bootstrap)

Hour	Inflow	Outflow	Net position	Cumulative inflow	Cumulative outflow	Cumulated outflow (%)
7:00	1,615	-107,576	-105,960	1,615	-107,576	34%
8:00	1,175	-10,109	-114,895	2,790	-117,685	37%
9:00	5,536	-26,959	-136,317	8,326	-144,643	46%
10:00	3,107	-62,236	-195,447	11,433	-206,879	65%
11:00	3,209	-14,270	-206,508	14,642	-221,150	70%
12:00	3,739	-41,415	-244,184	18,381	-262,565	83%
13:00	2,248	-35,488	-277,424	20,629	-298,053	94%
14:00	1,735	-11,145	-286,834	22,364	-309,198	97%
15:00	24,204	-2,424	-265,054	46,568	-311,623	98%
16:00	2,740	-5,984	-268,299	49,308	-317,607	100%
17:00	7,661	0,000	-260,638	56,969	-317,607	100%

Source: own elaboration.

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Marcin Spychała**

The variation in the level of the socio-economic development of the NUTS-3 subregions in the European Union

A very important research problem is the specification of the level of the socio-economic development of the EU regions. Within the cohesion policy, one may differentiate between regions more and less advanced in terms of general development, including the worth of GDP per capita. Following that measure, one can establish the areas eligible for getting help from EU budget support programmes. The purpose of the article was to present the variation in the level of the socio-economic development of 28 EU countries within the 1,347 NUTS-3 units. The level of the development was established on a multilevel basis, subcategorising three factors of regional development: human capital, the natural environment and the economy. The article specifies the extent of the NUTS-3 unit general progression following the analysis of 31 indicative measures structured around publicly available Eurostat statistical data (as of 2019). The extent of the progression was demonstrated based on a synthetic measure revealing the taxonomic distance of a particular region from the assumed arrangement measuring the development. The research procedure consisted of five development levels, namely: the establishment of factors of the progression of regions, a choice of variable factors, a decrease in the multi-factor space, an identification in the extent of the social and economic progression of the researched units and a subcategorisation of the units on the scale of the social and economic progression structured around a ranking prepared using the analysis of a lowering synthetic measure. The research resulted in spatial variation of 1,347 NUTS-3 unit subregions in 28 EU states presented in terms of the level of the social and economic progression and the three subparts of the progression. The highest synthetic measure was noted in the areas including the capitals of the states included in the research and in a belt comprising the Irish subregions, Central England, the Benelux states, Western and Southern Germany, up until the Alpine subregion. Moreover, significant differences between the levels of the development within the respective states were presented. The results of the research may be a source of inspiration for EU institutions within the scope of the manner of specifying the richest and the poorest EU regions, whose purpose is to ensure the efficient introduction and conclusion of the cohesion policy in the programming periods specified.

Keywords: regional development, NUTS-3 units, developmental discrepancies, distance from the pattern, Hellwig's data reduction

JEL Classification: O11, O20, O47
DOI: 10.15611/aoe.2023.1.09

Quote as: Spychała, M., (2023). The variation in the level of the socio-economic development of the NUTS-3 subregions in the European Union. *Argumenta Oeconomica*, 1(50), s. 179-204.

* Department of Public Finance, Poznań University of Economics and Business, Poland.
ORCID: 0000-0002-3860-303X.

1. Introduction

Identifying the level of the socio-economic development as well as its changes is a very important research issue both in the theory of economics and in economic practice. The level of the development heavily influences the EU regional classification in terms of the better and worse developed regions, and the latter in turn impacts on the scope of the EU fund allocation in the regions, as well as the intensity of state aid made available in the respective territorial units (Matsuura 2015; Albulescu, Goyeau 2014; Nistor, Glodeanu 2014). Within the EU cohesion policy, the regions are classified as richer and poorer and based on that, territorial units eligible for EU aid funds are subcategorised. The abovementioned classification was, however, conducted solely based on GDP per capita of a particular region as well as relating its value to the EU average. The purpose of the report was to specify the level of the development of the EU regions, however taking into account of the larger number of indicators, not just the GDP per capita.

Researching the core, the cause and the consequences of the process of the level of the socio-economic development constitutes the subject of many scientific compilations (Dreyer et al. 2006; Mukhametzhan et al. 2020; Jašková, Havierniková 2020; Vučković et al. 2018; Orlova et al. 2018; Shikverdiev et al. 2019). The regions under scrutiny progressed in spatially diverse ways. The increasing discrepancies of regional development are one of the core issues of the modern economy, and the main goal of the EU cohesion policy is convergence, i.e. activities aimed at decreasing the discrepancies in the level of EU regional progression (Beugelsdijk et al. 2018; Martin, Sunley 1998; Charron et al. 2014).

The article examines the level of the socio-economic development of all 1,347 EU NUTS-3 unit subregions (as of 2019), analysed based on 31 indicative factors included within the three factors (subcomponents) of regional progression: human capital, the environment, as well as the economy. The results were obtained using widely accessible Eurostat data. The article was written in order to present the extent of the variation of 28 EU regions within 1,347 territorial units of the NUTS-3 level, i.e. the third level of Eurostat categorisation of units of territory in a statistical manner. The article presents the socio-economic situation of the EU subregions at a more detailed level than the one used by the European Commission within the cohesion policy, both in terms of the number of subjects of the research and the scope of the indicators taken into account. The variation between the level of development aimed for, and the actual one demonstrates the level of progression.

A review of the literature within the scope of the factors of regional development is presented first, followed by the methodology of the research, together with a description of the respective stages of the research procedure. Next, the obtained results of the research are examined extensively. In the final section, the conclusions as well as the recommendations within the scope of exhibiting the extent of the progression between regions are indicated.

2. Factors of regional development in the subject literature

The regional development may be considered both in the quantitative and in the qualitative dimension. Striving towards an increase in productivity may be considered to be quantitative activities, and stimulating the development of certain spheres of activity with the simultaneous lack of incentives for other spheres may pass as qualitative activities (see: Vučković et al. 2018). It is well worth noting that the quantitative dimension of development is economic development, hence it is a notion wider than growth, as apart from qualitative changes, it is accompanied by structural changes (Bystrova et al. 2015).

In the subject literature one may find numerous reviews of the concepts and theories of regional development, some attempting to systematise them, taking account of different criteria (Illeris 1993; Martin 2015). The article includes a review of concepts of regional development conducted in terms of the factors of the development based on two main trends in economics:

– neoclassical, according to the liberal doctrine, minimising the extent of the intervention of the state in business and treating the free market as a regulating force;
– neo-Keynesian, considering state interventionism as indispensable and the most important regulatory mechanism of the regional development.

A concept of regional development relating to Smith's and Ricardo's theories representing the classical school of economics, is the theory of comparative costs. Ricardo proved not only the falsehood of the Mercantilists' beliefs concerning international trade, who thought only one party of the transaction obtains benefits, but also supported Smith's arguments on the advantages of free international trade. Smith saw the benefits of both partners, if each, thanks to specialisation has a cheaper product at their disposal, created using a relatively lower labour input compared to their partner (Friedmann 1983). Therefore, the main assumption of the theory is that the comparison of the efficiency and the labour cost between two regions shapes the level of the development, which is co-operation and international exchange.

Another neoclassical concept is the theory of convergence, formulated by the Nobel prize winner, J. Tinbergen. The theory has an immediate relation with the comparative cost theorem, and its supporters claim trade exchange between the highly and the low developed countries may with time lead to a decrease in the variations in their level of progression, going as far as to make even the level of generated revenue minus tax in both samples. The exchange between countries, the capital increase and the technology advancement are also significant (see: Henrekson, Jakobsson 2003).

In the first part of the twentieth century, as opposed to the neoclassical concepts, theories emerged connected with Keynes' doctrine. The previous concepts focused on the supply aspect, while Keynes and his followers centred around the demand side of the equation. The demand concepts postulate different administrative

and management activities directed at creating or reinforcing the factors of the progress, such as: boosting export, improving the labour force competencies, supporting the development of entrepreneurship, investments in infrastructure and creating innovativeness (Li et al. 2019; Diebolt, Hippe 2019; Zemtsov, Smelov 2018; Florida 2002).

The importance of innovation in the process of the development between regions was emphasised by J. Schumpeter, leader of the Austrian School. In his opinion, progress is similar to the process of creative destruction (Emami-Langroodi 2018; De Castro et al. 2018). Innovativeness provides a structural transformation and growth on the one hand, as well as the destruction of previous socioeconomic structures that turned out to have been inefficient and maladapted to the changing reality on the other. Some such units are unprepared for such changes. Occasionally, some maladapted units or structures have to destroy themselves in order to make room for technological transformation.

A novel way of perceiving development, proposed by P. Romer, assumes the possibility of collecting the factors of development, meaning the option to achieve sustainable development, as well as maintaining and increasing the economic differences between the regions. A stable and long-lasting development is termed within the concept as an endogenous development. The main stimulating factors are: human capital, the financial capital and also the technological advancement of the economy. The theory states that the lesser developed regions may not close the gap in developmental differences in a manner different than by increasing their technological advancement and investing in human competencies (Baklanov 2020; Arranz et al. 2019; Rodionov et al. 2018; Benner 2003).

In line with the neoclassical trend, a contemporary model was also created of the so-called new economic geography, which combines three factors: transport cost, the benefits of an agglomeration and the cost of production flow resources (see: Krugman 1998). The model established that there is a possibility of the occurrence, and even of the intensifying, the differences between regions, which according to Grosse (2018) is an effect of the trend of accumulating the factors of development in the most developed capital regions of the metropolitan area. New economic geography along with the new theory of development have found a widespread use within the concept of the policy of endogenous regional development (see: Eshugova 2018). According to this theory, the economic development of a particular region depends on factors such as: the endogenous material capital, the human and social capital as well as the intraregional policy of the development. Regional development should be an internal effect, as well as managed by a region "from the bottom" (Gallego et al. 2010). The conditions for sustainable regional development should be created by an efficiently managed regional policy. Thus the concept rejects the neoclassical assumption of the "invisible hand of the market" to the benefit of neo-Keynesian state interventionism.

Regional development may take place in a spontaneous manner or in a directed manner. The latter is connected with shaping development through stimulating the factors of the regional development within the regional policy. The problem of regional policy is a subject of lively debate among the exponents of different fields of knowledge: economics, geography, law and political science. Defining the regional policy leads to some difficulty, especially in the context of the activities taken within the structural policy and the cohesion policy of the European Union (Milenković et al. 2021). The notions are very often interchangeable, as both the regional policy, the cohesion policy and the structural policy have the same purpose which is to decrease the economic and social differences between the EU regions (Moroshkina 2020). The European Commission (Smętkowski, Dąbrowski 2019) concluded that regional policy is a conscious and purposeful activity of the central bodies of the public authority directed towards regulating the interregional proportions of development. Structural policy is a notion used in the European Union virtually always within the policy of cohesion (Downes 1996).

According to Smętkowski (2015), regional policy in the economic dimension comprises all forms of state intervention which are directed towards a change in the spatial location of the business venture activity. Within such a perspective, it aims to correct the effects of the functioning of free market forces directed at ensuring economic development, along with a change in income redistribution (Pfirrmann 1995). From the general perspective, regional policy may be defined as activities striving towards maximising of the usability function, i.e. improving the economic situation of one or a few regions (Démurger 2001).

To summarise, regional policy is linked with the occurrence of variations in the level of the development of the respective regions. Uneven progress made by the regions resulting from the decisions of the investors, the workings of the mechanism regulating the market, and the factors connected with the geography, lead to different amounts of the registered income and the quality of life. The basic task of regional policy is thus the necessity to decrease the differences in scale of those phenomena (Liu et al. 2018).

Among the factors most frequently mentioned in the literature on regional development are: economic aspects and aspects related to human capital. Following the literature review, however, a research gap was identified in relation to including environmental aspects among regional development factors. This may be caused by the difficulty in identifying and insufficient indexing of environmental elements, therefore the author of this study decided to address the research gap and include environmental aspects in his considerations as one of the most important regional development factors. However, due to the limited availability of empirical data directly referring to the natural environment, the author is aware of certain imperfections related to the construction of a synthetic indicator.

3. Methodology

The article provides a confirmed hypothesis according to which the development of society and the economy, including the NUTS-3 unit subregion variation in the European Union, and its highest level of the said development is noted in the territorial units comprising the capitals of countries, and the lowest in the regions furthest from the capital units exhibited. The research comprises all NUTS-3 subregions subcategorised in 28 EU states as at the end of 2019 – 1,347 units in total.

A cross-sectional measure of distance from the assumed pattern was used. The research procedure was carried out simultaneously – both statically (based on the values of measures as of 2019) and dynamically (based on the change of the value of indicators in 2010-2019, i.e. a 10-year period with the most up-to-date data). The research procedure consisted of five stages:

(1) an identification of subparts – i.e. the factors contributing to the progression of the regions (the so-called subcomponents of the development),
(2) a choice of contributing factors – a construction of a matrix of geographical information,
(3) a decrease in the multi-factor space,
(4) a specification of the extent of the social and economic development of the researched units,
(5) a classification of the subregions on the scale of the socio-economic progress resulting from the data included in a ranking created according to the decreasing value of the synthetic measure.

In the first stage, a conducted review of literature comprising the concept of regional development demonstrates that the most significant subparts were established, i.e. the subparts constituting the progression of the respective areas. A factor influencing the progression might be a subpart, a property of the region or an event which puts the most pressure on the socio-economic development of a particular territorial unit (Khasanova et al. 2020; Naydenov 2019; Feldman 1999; Yun et al. 2017). In this article, the socio-economic development was characterised based on its three subcomponents: human capital, the natural environment, and the economy. The ‘human capital’ subcomponent was specified through the indicators presenting the population potential of a particular subregion, relating to the migration balance and the natural development, the level of education, the birth rate, and the age structure. Within the ‘natural environment’ subcomponent, the indicators concerning the use of land, the structure of farms, urban waste, the burden connected with road transport of goods, and the use of energy for heating and cooling the living quarters were also included. Within the ‘economy’ subcomponent, indicators concerning entrepreneurship were subcategorized, as well as the GDP, the structure of the employment and the scale of innovativeness measured as the number of trademarks and consumer goods obtained. All the factors of regional development mentioned above are interrelated. In striving towards the competitive development of a particular region, these factors should be included in its long-term strategy (cf. Cristina et al.

2021). According to Churski (2008), one may thus assume that regional development comprises both the dynamic processes occurring under the influence of specific factors which determine the character, the direction as well as the speed of the socio-economic changes, and also the changes purposefully directed at realising the tasks within the regional policy. In the research, publicly available Eurostat data were used. The author understands the inadequacies and incompleteness of the structure of the respective indicators characterising the factors of the development. Specific characteristics could be built more precisely, however all the available public data taken into account may determine the level of the development of the respective subcomponents of the development. It is well worth considering a wider catalogue of indicators, e.g. from other sources.

Table 1

Indicative factors analysed in the establishment of the respective area development

The indicative factor of development	**Indicative factors**
Human capital (11 variables)	the growth rate counted per 1,000 inhabitants (s); the ratio of migration per 1,000 inhabitants (s); the share of people with higher education in the total population (s); the percentage of individuals at working age in the general population (s); the percentage of individuals at pre-production age in the general population (s); *the percentage of people at non-working age in the general population (d);* the ratio of people at non-working age against one hundred individuals at working age (d); *the number of people at non-working age per 100 people at pre-working age (d);* the birth rate in total (s); the population median age (d); the average age of women when giving birth (d)
The natural environment (10 variables)	the area of farming land and the green areas in the area in total (s); the percentage of farmers owning farms under the age of 35 in the number of farm owners in total (s); the percentage of farms under 5 hectares in the number of farms in total (d); road transport of products in tonnes measured against 1,000 inhabitants (d); the noting of minor offences and crimes against the natural environment measured against 1,000 inhabitants (d); the electrical energy exploitation for the purpose of lowering the temperature of the living areas (as the average in the EU) (d); *the electricity use for heating the living areas (as the average in the EU) (d);* municipal waste in relation to1 inhabitant (d); municipal waste disposed of measured against 1 inhabitant (s); accommodation per 1,000 inhabitants (s)
The economy (10 variables)	microenterprises in the economic entity number in total (s); the creation of enterprise coefficient (s); the number of information sector employees in the general workforce (s); the percentage of farming employees in the total number of the employed (d); the percentage of professional service sector employees in the general workforce (s); *the share of the employed in the financial sector in the total number of the employed (s);* the percentage of the employed in services in the general workforce (s); *consumables in relation to 1m inhabitants (s);* the number of trademarks per 1m inhabitants (s); Gross Domestic Product per capita (as the EU average) (s)

Notes: (s) – stimulant, (d) – destimulant, *italics* – rejected indicators.

Source: own research.

The subsequent part of the study involved geographical data obtained from 31 indicating factors (Table 1), highlighting the level of development of the NUTS-3 parts in 2019 and changes related to its three subparts: human capital, the environment and the economy. Next, Pearson's correlation coefficients were included among all the final indicators researched separately for 2019, and separately for their change in 2010-2019. It is important for the indicators selected for a synthetic measure of distance from the assumed model of the development to be achieved, not to be correlated with each other. Thus, the information capacity of the indicators is varied.

The obtained Pearson's correlation coefficient matrices were the foundation for the reduction of a departure variable decrease by means of Hellwig's method, whose purpose was to seclude the indicative features, i.e. those indicators to be included in the further research procedure (Balcerzak 2016). In Hellwig's indicative measure decrease method, the correlation coefficients between different variables were manipulated for the purpose of the final calculation. The main indicative feature had the highest correlation of the value sum of the absolute correlation coefficients with the other features (then called the central feature). Next, the variables for which the value of the correlation coefficient was bigger than the central value indicated formed against the pattern below specified, were eliminated (Nowak 2018):

$$r^* = \sqrt{\frac{\left(t^*\right)^2}{n-2+\left(t^*\right)^2}} \tag{1}$$

where: r^* – the central value of Pearson's coefficient measuring linear correlation [**= 0.35**], t^* – the t-Student statistic (at $p = 0.05$), n – departing indicative measures (variables) [= **31**]

Next, the varying measures were cancelled which were statistically significant for the diagnostic feature (termed as the satellite factors). At each subsequent stage, there was a reduction of the correlation matrix by a significant feature as well as the lesser significant features. Hellwig's method was used again by the achievement of new decreased value correlation matrices until the set of features is exhausted or isolated features were separated (Hauke, Kossowski 2011). The procedure of variable reduction was conducted eight times: with respect to the level of the general socio--economic development, and for the level of the development of each of the three subparts of the development both in the static dimension (for the data for 2019), as well as the dynamic dimension (for the data for 2010-2019).

At the following stage of the research procedure, a model and an anti-model of the social and economic progression were specified. The model considered the highest standardised values of the respective diagnostic features, and an anti-model – their minimum values (Spychała 2020). At the subsequent stage, the diversion of every thus examined subregion of the NUTS-3 level from the desired model of development formed against the data comprising the pattern specified below was established (Reiff et al. 2016):

$$d_{i0} = \sqrt{\sum_{j=1}^{m}(z_{ij} - z_{0j})^2}, \quad (2)$$

where: d_{i0} – the diversion of the i^{th} subregion from the desired model of development, z_{ij} – the standard value of the j^{th} indicative measure (feature) for the i^{th} subregion, z_{0j} – the standard value of the j indicative measure (feature) for the pattern of the development.

At the last stage of the research procedure, a synthetic measure being an indicator of the level of the development in a given subregion was devised for each territorial unit of the NUTS-3 level. The value of the synthetic measure was calculated for the total level of socio-economic development and separately for each of the three factors of development. The synthetic measure was calculated based on the following pattern:

$$v_i = 1 - \frac{d_{i0}}{d_0}, \quad (3)$$

where v_i – a synthetic measure of the level of the i^{th} subregion development, d_{i0} – the diversion of the i^{th} subregion from the established model of development, d_0 – the diversion of the model from an anti-model of development.

A synthetic measure of the extent of the researched development took on figures from 0 to 1, noting that the higher its value, the higher the level of the development of the researched area. Against the values of the synthetic measures counted, a set of the ranked 1,347 subregions of the NUTS-3 stage in the EU countries was established, and subsequently divided into five groups:

1. group – extremely high (20% of the areas exhibiting the biggest value of the synthetic measure values within the ranking 1-269);
2. group – high (the next 20% of the areas within the ranking 270-539);
3. group – medium (territorial units ranked within 540-808 including their lowered position resulting from a particular synthetic measure);
4. group – low (areas within the ranking 809-1,078);
5. group – very low (20% of the subregions with the lowest value of the synthetic measure within the ranking 1,079-1,347).

Regarding the research conducted dynamically, areas of the NUTS-3 level for which the measure registered the highest value figures (20% of the studied areas) were subcategorised into a set of a very big change in the level of the development of a specific phenomenon, and the units for which the measure took the lowest values (20% of the researched subregions), classified into the group at a relatively low change in the level of development of a specific area of research.

In Figures 1 and 2 and in Tables 2 and 3, the research outcomes are presented. Table 2 shows the NUTS-3 level units of the highest and lowest values of the synthetic measure within the respective factors of the social and economic development counted separately for 2019, while Table 3 shows the NUTS-3 subregions of extreme figures

Table 2

The highest and the lowest values of the synthetic measure within the specific factors of regional development in 2019

The highest figures of the synthetic measure (2019)			The lowest figures of the synthetic measure (2019)		
Item	**The NUTS-3 area**	**Value**	**Item**	**The NUTS-3 area**	**Value**
Human capital					
1	Seine-Saint-Denis (FR)	0.581	1347	Stuttgart, Stadtkreis (DE)	0.139
2	Barking and Dagenham and Havering (UK)	0.569	1346	Esslingen (DE)	0.142
3	Val-d'Oise (FR)	0.568	1345	Böblingen (DE)	0.151
4	Thurrock (UK)	0.566	1344	Evrytania (EL)	0.325
5	Mid-East (IE)	0.566	1343	Zamora (ES)	0.360
The natural environment					
1	Hautes-Alpes (FR)	0.596	1347	Florina (EL)	0.147
2	Savoie (FR)	0.582	1346	Seine-Saint-Denis (FR)	0.334
3	Primorsko-notranjska (SL)	0.551	1345	Darlington (UK)	0.341
4	Koroska (SL)	0.547	1344	Byen København (DK)	0.348
5	Kerkyra (EL)	0.542	1343	Arr. de Bruxelles-Capitale (BE)	0.356
The economy					
1	Camden and City of London (UK)	0.642	1347	Vaslui (RO)	0.027
2	Westminster (UK)	0.554	1346	Kardzhali (BG)	0.029
3	Luxembourg (LU)	0.329	1345	Razgrad (BG)	0.031
4	Tower Hamlets (UK)	0.291	1344	Neamt (RO)	0.033
5	Groot-Amsterdam (NL)	0.244	1343	Silistra (BG)	0.035
The level of the social and economic progression in general					
1	Camden and City of London (UK)	0.518	1347	Esslingen (DE)	0.194
2	Westminster (UK)	0.489	1346	Böblingen (DE)	0.202
3	Luxembourg (LU)	0.423	1345	Stuttgart, Stadtkreis (DE)	0.214
4	Tower Hamlets (UK)	0.406	1344	Florina (EL)	0.217
5	Dublin (IE)	0.379	1343	Evrytania (EL)	0.237
6	Groot-Amsterdam (NL)	0.378	1342	Vidin (BG)	0.256
7	München, Landkreis (DE)	0.372	1341	Fokida (EL)	0.258
8	Kensington and Chelsea and Hammersmith and Fulham (UK)	0.366	1340	Alto Tâmega (PT)	0.262
9	Miasto Warszawa (PL)	0.364	1339	Terras de Trás-os-Montes (PT)	0.262
10	München, Kreisfreie Stadt (DE)	0.363	1338	Arta, Preveza (EL)	0.264

Source: own research.

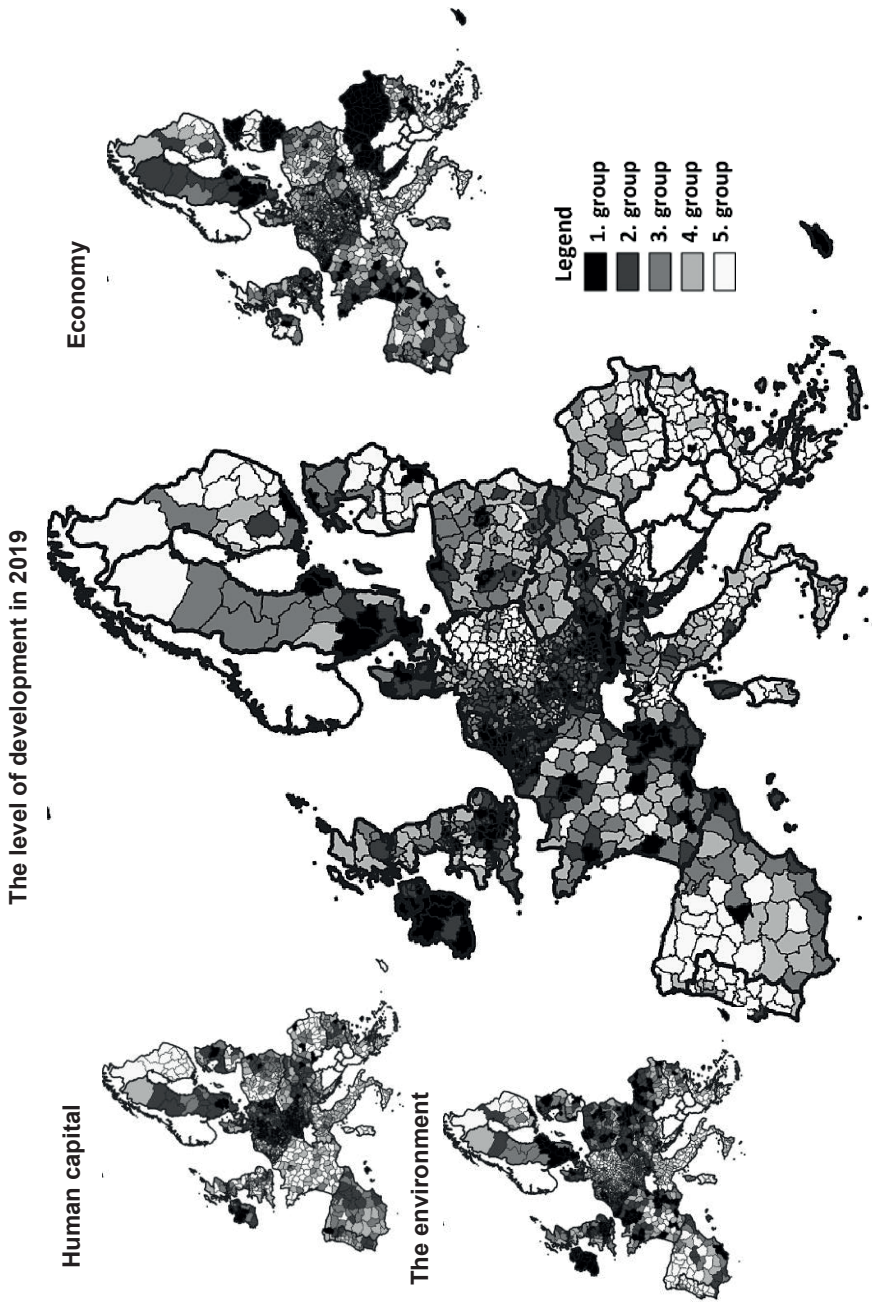
Table 3

Extreme figures of the synthetic measure within the specific factors of regional development in 2010-2019

The highest values of the synthetic measure (the period of 2010-2019)			The lowest values of the synthetic measure (the period of 2010-2019)		
Item	**The NUTS-3 area**	**Value**	**Item**	**The NUTS-3 area**	**Value**
Human capital					
1	Hagen, Kreisfreie Stadt (DE)	0.606	1347	Oost-Zuid-Holland (NL)	0.270
2	Gelsenkirchen, Kreisfreie Stadt (DE)	0.584	1346	Etelä-Savo (FI)	0.368
3	Salzgitter, Kreisfreie Stadt (DE)	0.581	1345	Kymenlaakso (FI)	0.373
4	Schweinfurt, Kreisfreie Stadt (DE)	0.581	1344	Kainuu (FI)	0.378
5	Ingolstadt, Kreisfreie Stadt (DE)	0.580	1343	Kanta-Häme (FI)	0.384
The natural environment					
1	Norrbottens län (SE)	0.422	1347	Florina (EL)	0.160
2	Kerkyra (EL)	0.395	1346	Menorca (ES)	0.212
3	Außerfern (AT)	0.393	1345	Ithaki, Kefallinia (EL)	0.230
4	Grevena, Kozani (EL)	0.389	1344	Fokida (EL)	0.233
5	Västerbottens län (SE)	0.381	1343	Byen København (DK)	0.237
The economy					
1	Camden and City of London (UK)	0.388	1347	Pieriga (LV)	0.155
2	Malta (MT)	0.357	1346	Bratislavský kraj (SK)	0.201
3	Westminster (UK)	0.356	1345	Zemgale (LV)	0.207
4	Ilfov (RO)	0.355	1344	Vidzeme (LV)	0.209
5	Wolfsburg, Kreisfreie Stadt (DE)	0.355	1343	Banskobystrický kraj (SK)	0.214
The level of the social and economic progression in general					
1	Norrbottens län (SE)	0.371	1347	Florina (EL)	0.255
2	Außerfern (AT)	0.369	1346	Pieriga (LV)	0.264
3	Cluj (RO)	0.361	1345	Fokida (EL)	0.264
4	Västerbottens län (SE)	0.361	1344	Thesprotia (EL)	0.279
5	Vas (HU)	0.359	1343	Oost-Zuid-Holland (NL)	0.281
6	Bucuresti (RO)	0.358	1342	Bratislavský kraj (SK)	0.283
7	Ingolstadt, Kreisfreie Stadt (DE)	0.357	1341	Banskobystrický kraj (SK)	0.284
8	Camden and City of London (UK)	0.356	1340	Ithaki, Kefallinia (EL)	0.285
9	Vilniaus apskritis (LT)	0.356	1339	Etelä-Savo (FI)	0.287
10	Gyor-Moson-Sopron (HU)	0.354	1338	Nitriansky kraj (SK)	0.287

Source: own research.

of the synthetic measure counted for the changes in the period 2010-2019. Figure 1 contains choropleth maps representing the spatial variation of the level of the socio-economic development of the NUTS-3 subregions located in the continental part of the European Union in 2019 (the UK excluded), as well as the changes of the level of development in 2010-2019. Figure 2 shows the division of the NUTS-3 subregions


The level of development in 2019
Economy
Human capital
The environment
Legend
1. group
2. group
3. group
4. group
5. group

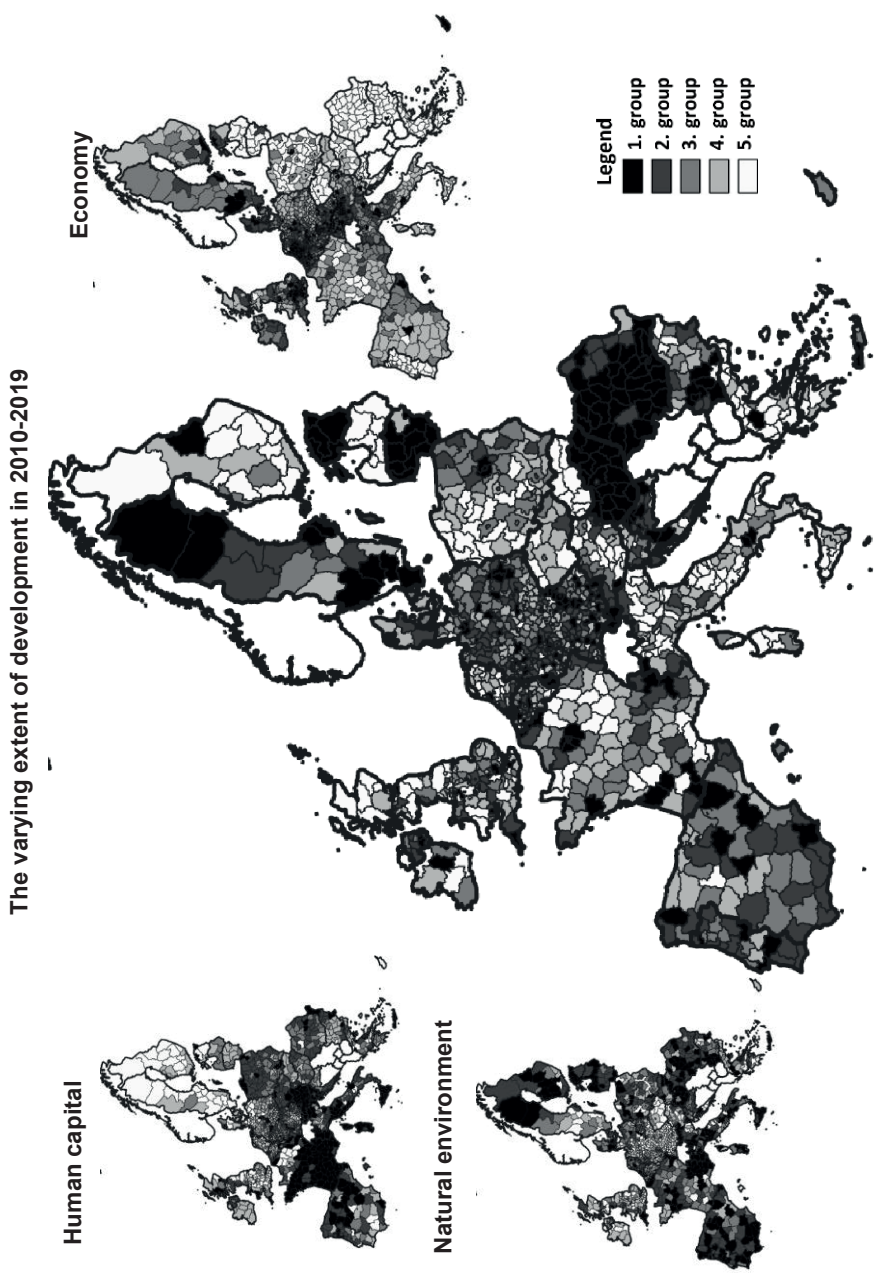


Fig. 1. Variation of the extent of regional development of the subregions in the EU countries

Source: own research.

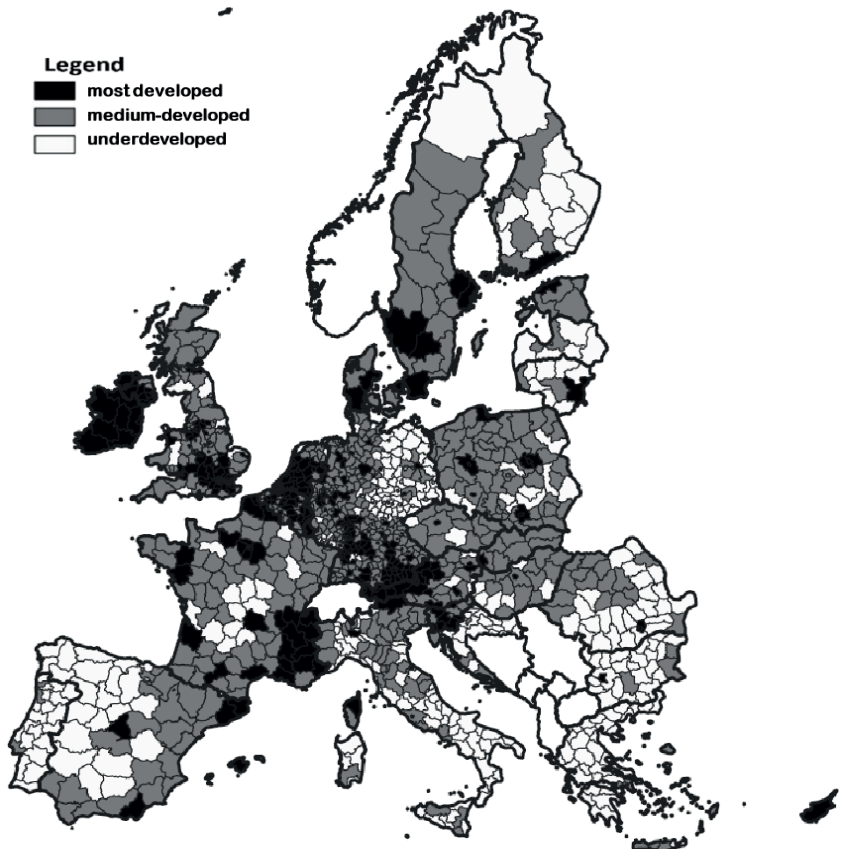


Fig. 2. The most developed, medium-developed and socio-economically underdeveloped subregions

Source: own research.

into the most developed (darkest colour), medium-developed and underdeveloped (lightest colour). The above classification was made on the basis of the average value of the total synthetic measure and its standard deviation.

4. Conclusions from the scientific examination of the selected respective factors of development

The research carried out for the purpose of this study resulted in the spatial variation of 1,347 NUTS-3 level subregions in 28 EU countries demonstrated with regard to the extent of socio-economic development, and the three subcomponents of the development (Figure 1). In the set of units examined, the value of the synthetic measure reflecting the level of the social and economic development in 2019 varied from 0.194 to 0.518 (Table 2). The value of the measure reflecting the variation in the social and economic rate of development of the researched subregions in the period

2010-2019 ranged from 0.255 to 0.371 (Table 3). A parallel variation was noted in the case of human capital (0.139–0.581 for 2019, and 0.270–0.606 for the change in 2010-2019), the natural environment (0.147–0.596 and 0.160–0.422, respectively), and in the economy (0.027–0.642 and 0.155–0.388, respectively).

It is thus worth noting that the highest variation in the specific areas was observed for the economy in 2019, and the biggest similarity of the researched units occurred in the case of the changes of the general level of the socio-economic development in 2010-2019. In the researched subregions, the biggest discrepancies within the scope of their economic development, while for human capital and the natural environment such big fluctuations were not registered.

Regarding the rate of development of the human capital subcomponent, the most significant values of the synthetic measure in 2019 were noted in the following subregions: Seine-Saint-Denis (FR – comprising the northern outskirts of Paris), Barking and Dagenham and Havering (the UK – forming part of the Greater London) and Val-d'Oise (FR – also located in the Île-de-France region), and the lowest – in the German subregions: Stuttgart, Esslingen and Böblingen. Their significant place within the ranking of the indicated NUTS-3 units was decided by: a high birth rate and a positive migration balance, a favourable age structure of the population, as well as a significant share of people with higher education. A low position of the respective units was decided by: the negative birth rate, and a very high share of people at the post-production age in the total number of people. In the group of the subregions of the lowest level of the development of human capital, there were the NUTS-3 units in which the ageing of the population may be observed. Apart from the German subregions (including specifically the ones located in the area of previous East Germany), one may thus subcategorise, among others, a high number of Portuguese, Spanish, Italian and Greek subregions. Considering the analysis conducted in the dynamic dimension, the biggest change in the level of the development of the human capital subcomponent in 2010-2019 was observed in the German subregions of Hagen, Gelsenkirchen and Salzgitter. Generally, the biggest improvement in the human capital in the researched years was noted in the German subregions, related to, among others, an increase in the level of the education of the population and also a relatively big improvement in the demographic structure (resulting from a very unfavourable starting position in 2010). The least significant variation in the rate of the development of the human capital impacting factor was noted in the areas: Oost-Zuid-Holland (NL), Etelä-Savo (FI) and Kymenlaakso (FI). The weaker position of the indicated NUTS-3 level units in the research on the change in the level of the development of human capital was influenced by: the falling birth rate, the increasing indicator of the demographic burden, and the significant deepening of the negative migration balance.

Regarding the factors impacting the condition of the environment, the highest value of the synthetic measure in 2019 was observed in the areas: Hautes-Alpes (FR), Savoie (FR) and Primorsko-notranjska (SL), and the lowest in Florina (EL), Seine-Saint-Denis (FR) and Darlington (UK). The position within the ranking of the NUTS-3

units was decided by: the number of accommodation places per 1,000 inhabitants, a significant share of land used for farming and the green spaces in the area in total, as well as the low share of road transport of goods. A high position within the ranking within the natural environment, involved tourist areas, located among others in South and Central France, in the Alpine region, in Austria, in Northern and Central Italy, on the Black Sea coast, by the Polish coast of the Baltic Sea and on the Mediterranean islands. A low ranking position was decided by: a low share of the area of farming land and the green spare in the area in total, a high number of offences and crimes connected with the natural environment per 1,000 inhabitants, and a heavy burden connected with road transport of products. When considering the dynamic development, the biggest improvement in the natural environment in 2010-2019 was observed in the following subregions: Norrbottens län (SE), Kerkyra (EL) and Außerfern (AT), and the lowest in Florina (EL), Menorca (ES) and Ithaki, and Kefallinia (EL). The lowest position of the studied subregions was decided by: an increase in urban waste generated per 1 inhabitant, a decrease in accommodation places, and an increase in electrical energy for cooling the living quarters. A high position within the ranking of the respective units was decided by: a decrease in the nuisance of road transport of products, a high increase in the share of the natural green areas in the total area (the biggest improvement was observed in the Austrian subregion of Außerfern), as well as a significant increase in the share of the treated waste.

Regarding the rate of development of the economy, the highest value of the synthetic measure in 2019 was observed in the British subregion forming part of the Inner London: Camden and the City of London and Westminster, as well as in Luxembourg (LU) – these NUTS-3 units showed the biggest number of patents registered per 1 mln inhabitants, the biggest share of the employed in the financial sector in the total number of the employed, as well as the biggest GDP per 1 inhabitant. The lowest value of the synthetic measure was observed in the following units: Vaslui (RO), Kardzhali (BG) and Razgrad (BG) (the biggest share of the employed in farming, the lowest number of patents per 1 mln inhabitants, as well as a very low GDP per capita). Thus the highest value of the synthetic measure within the scope of the economy was observed in the subregions located within the belt extending from the British capital region , through Benelux, South-West Germany, up to the Alpine region. The value of the measure decreased together with the increasing distance from the abovementioned area of the development of the economic level. For the dynamic set, the biggest improvement of the economy subcomponent in 2010-2019 was registered in: Camden and City of London (UK), Malta (MT) and Westminster (UK), and the lowest in the subregions of Pieriga (LV), Bratislavský kraj (SK) and Zemgale (LV). The position within the ranking of the NUTS-3 units in the dynamic version of the research was decided by: the percentage of the professional service sector employees in the general workforce, the share of the employed in the financial sector in the total number of the employed, an increase in the number of registered patents per 1 mln inhabitants, as well as the share of the employed in the information sector and in the communication sector in the total

number of the employed. Interestingly, in the group of the subregions in which the biggest increase in the economy subcomponent was registered, there was a majority of Romanian, Hungarian, Lithuanian, Estonian and Croatian units. Moreover, it is worth noting that the value of the synthetic measure within the economy reflects to a large extent the level of the general social and economic development of the EU areas under study.

5. Discussion – the overall level of the socio-economic development of the NUTS-3 regions in the EU member states

Summarising the results of the research conducted on the rate of the social and economic development of all 1,347 NUTS-3 areas in the EU countries, the following general remarks can be made. The rate of the overall development in 2019 was characterised according to 31 indicators subcategorised as the three indicative factors of the development: human capital, the natural environment, and the economy. The most significant indicative factor of the synthetic measure was registered in the subregions comprising the capitals of the respective countries: Camden and the City of London (UK), Westminster (UK) and Tower Hamlets (UK) as parts of London, Luxembourg (LU), Dublin (IE) and Groot-Amsterdam (NL). The statement to be verified, given at the start of the article, was confirmed. What is more, in the first thirty of the best developed NUTS-3 subregions in the European Union, 16 capital units were found – comprising the capitals of 11 countries (six units were parts of London), and the group of 20% of the best developed subregions numbered 22 out of 28 units, including the capitals of the respective EU countries (constituting 79% of their population). Including the dynamic compilation, the biggest improvement of the socio-economic development in 2010-2019 was observed in the following subregions: Norrbottens län (SE), Außerfern (AT) and Cluj (RO). Among the capital subregions, the top thirty places in the ranking were: Bucuresti (6th), Camden and the City of London (8th), Vilniaus apskritis (9th), Põhja-Eesti (17th position) and Luxembourg (30th). The group of 20% of the subregions demonstrating the biggest change in the level of the socio-economic development included 15 out of 28 subregions including the capitals of the respective EU countries (i.e. 54% of all the capital regions). Apart from those previously mentioned, the group also included: Warsaw (42th position in the ranking), Berlin (45th), Malta (60th), Sofia (64th), Dublin (83th), Stockholms län (92th), Grad Zagreb (93th), Groot-Amsterdam (97th), Área Metropolitana de Lisboa (162th) and Madrid (244th). The first half of the compilation also included Budapest (450the), Osrednjeslovenska (470th) and Hlavní mesto Praha (609th). Lowest positions among the capitals were taken by: Rome, Byen København, Arr. de Bruxelles-Capitale, Athens and Bratislavský kraj (1167th, 1196th, 1262th, 1304th and 1342th, respectively). Remarkably, the areas exhibiting a very high level of social and economic development were usually those units which experienced the biggest change in the rate of their development in 2010-2019 (and conversely). Apart from the capitals indicated, a group of the subregions of the

highest level of the development in 2019 included among others, units located in a belt extending from Ireland, through Central England, Benelux, Western and Southern Germany up to the Alpine regions (i.e. an area similar to the units characterised by the highest level of development of the economy), as well as the Danish, Southern Swedish and French-Italian border subregions. The group of units of the most significant alteration of the level of socio-economic development – outside of the capitals – included above all the subregions from the areas with a weaker level of socio-economic development in 2010, and which within the last 10 years made up for the developmental discrepancies and effectively caught up with the richer regions in terms of the level of the development (mainly the Romanian, Hungarian, Croatian and the Baltic states' regions). Conversly, the regions with the lowest rate of socio-economic development were the NUTS-3 units located on the outskirts and also away from the most advanced areas, including, most notably, the NUTS-3 units in Greece, Southern Italy, and the furthest overseas departments and dependent units, such as: Martinique, Guadeloupe, French Guyana, Ciudad de Melilla, Ciudad de Ceuta, the Azores and Madeira. One may thus conclude that – on the one hand – the current level of development of the respective subregions of the EU member states to a large extent depends on the means of activities undertaken within the last ten years, while on the other hand – the increasingly higher developmental discrepancies of the NUTS-3 subregions are noted. To the greatest extent, this applies to the level of socio-economic development in the economically strongest regions, being mostly capital subregions of the respective countries, and to the least extent – to the relatively weaker developed subregions, usually located away from the socio-economic centre of Europe, within the belt extending from London, through Benelux, up to the Alpine regions. Large developmental discrepancies may also be observed in the respective countries. In the examined areas of each subregion, there were regions both exhibiting a very high level of socio-economic development, and the units classified into the group of the 20% of the least developed NUTS-3 units in the respective states.

The conducted research analysis is specific to the related studies, and it was not possible to find another publication in which the level of the regional development of the EU was first of all conducted in a synthetic manner for all the 28 EU member states on the level of all 1,347 NUTS-3 subregions, and secondly, where a synthetic measure of development was used for that purpose prepared on the basis of Hellwig's data reduction procedure. The statements of other authors researching regional development in the EU, who use other procedures, are however similar. They also indicate the largest intensity of development of the capital regions, while the level of development usually decreases together with the distance from the most important regions. Similarly, analyses concerning the NUTS-2 unit regions by Eurostat, based on GDP per capita, classify the capital units among the richest regions. The abovementioned compilation thus validates the appropriateness of the achieved final criteria – the applied procedure notwithstanding – classifying a specific area into a more or less developed group in a correct manner.

Conclusions

To summarise, it is worth looking into the results relating to the respective countries (Table 4). The rate of the social and economic development varied greatly also in the respective countries, as shown in Table 4 (apart from the countries in

Table 4

The most and least advanced NUTS-3 subregions in the EU member states in 2019

Country	The most developed NUTS-3 region			The least developed NUTS-3 region		
	The name of the region	Value	Place	The name of the region	Value	Place
Belgium	Arr. de Bruxelles-Capitale	0.359	15	Arr. Veurne	0.309	737
Bulgaria	Sofia (stolitsa)	0.326	254	Vidin	0.256	1342
Czech Rep.	Hlavní mesto Praha	0.335	119	Karlovarský kraj	0.299	1021
Denmark	Byen København	0.352	26	Bornholm	0.292	1148
Germany	München, Landkreis	0.372	8	Esslingen	0.194	1347
Estonia	*Põhja-Eesti*	*0.353*	*23*	*Kirde-Eesti*	*0.309*	*738*
Ireland	Dublin	0.379	5	Mid-West	0.324	294
Greece	Andros, Thira, Kea, Milos…	0.320	391	Florina	0.217	1344
Spain	Girona	0.330	177	Ourense	0.267	1332
France	Hauts-de-Seine	0.352	24	Creuse	0.283	1261
Croatia	Grad Zagreb	0.321	370	Vukovarsko-srijemska zupanija	0.276	1307
Italy	Bolzano-Bozen	0.335	118	Medio Campidano	0.270	1324
Cyprus	*Kypros*	*0.338*	*90*	*Kypros*	*0.338*	*90*
Latvia	*Pieriga*	*0.315*	*552*	*Latgale*	*0.283*	*1268*
Lithuania	Vilniaus apskritis	0.328	217	Utenos apskritis	0.277	1305
Luxembourg	*Luxembourg*	*0.423*	*3*	*Luxembourg*	*0.423*	*3*
Hungary	Budapest	0.325	286	Békés	0.292	1142
Malta	*Malta*	*0.354*	*21*	*Gozo and Comino*	*0.313*	*592*
Netherlands	Groot-Amsterdam	0.378	6	Oost-Groningen	0.293	1124
Austria	Salzburg und Umgebung	0.349	34	Südburgenland	0.295	1079
Poland	Miasto Warszawa	0.364	9	Sandomiersko-jedrzejowski	0.291	1169
Portugal	Área Metropolitana de Lisboa	0.319	426	Alto Tâmega	0.262	1340
Romania	Ilfov	0.329	195	Teleorman	0.270	1325
Slovenia	Osrednjeslovenska	0.345	47	Pomurska	0.302	930
Slovakia	Bratislavský kraj	0.341	64	Trenciansky kraj	0.301	972
Finland	Helsinki-Uusimaa	0.343	53	Etelä-Savo	0.266	1333
Sweden	Stockholms län	0.359	12	Norrbottens län	0.294	1109
the UK	Camden and City of London	0.518	1	Eilean Siar (Western Isles)	0.289	1195

Notes: *italics* – countries in which only one or two NUTS-3 subregions are separate.

Source: own research.

Table 5

Average values of the synthetic measure in the subregions of the researched countries

Item	Belgium	Bulgaria	Czech Rep.	Denmark	Germany	Estonia	Ireland	Greece	Spain	France	Croatia	Italy	Cyprus	Latvia
Number of NUTS-3 subregions	44	28	14	11	401	5	8	52	59	101	21	110	1	6
The level of development in 2019														
General - the average value of the synthetic measure	0.323	0.283	0.307	0.323	0.313	0.320	0.338	0.287	0.304	0.316	0.292	0.298	0.338	0.297
Human capital - the average value of the synthetic measure	0.500	0.460	0.501	0.494	0.468	0.495	**0.549**	0.455	0.462	0.495	0.461	0.439	0.518	0.486
The natural environment - the average value of the measure	0.416	0.418	0.427	0.405	0.418	0.422	0.408	0.422	0.432	0.441	0.419	0.434	0.385	0.413
The economy - the average value of the synthetic measure	0.111	0.049	0.070	0.119	0.109	0.103	0.114	0.060	0.087	0.087	0.066	0.089	0.145	0.065
Transforming rate of development in 2010-2019														
General - the average value of the synthetic measure	0.320	0.321	0.316	0.318	0.324	0.340	0.320	0.310	0.322	0.320	0.326	0.315	0.321	0.292
Human capital - the average value of the synthetic measure	0.485	0.496	0.482	0.487	**0.531**	0.485	0.522	0.477	0.494	0.464	0.469	0.467	0.450	0.521
The natural environment - the average value of the measure	0.296	0.302	0.290	0.295	0.292	**0.316**	0.288	0.299	0.301	0.303	0.295	0.307	0.291	0.307
The economy - the average value of the synthetic measure	0.290	0.283	0.287	0.286	0.290	0.314	0.286	0.268	0.286	0.290	0.307	0.273	0.305	0.214

Item	Lithuania	Luxembourg	Hungary	Malta	Netherlands	Austria	Poland	Portugal	Romania	Slovenia	Slovakia	Finland	Sweden	the UK	EU
Number of NUTS-3 subregions	10	1	20	2	39	35	73	25	42	12	8	19	21	179	1347
The level of development in 2019															
General - the average value of the synthetic measure	0.293	**0.423**	0.305	0.334	0.330	0.321	0.312	0.287	0.298	0.322	0.313	0.298	0.316	0.324	0.311
Human capital - the average value of the synthetic measure	0.476	0.520	0.488	0.499	0.490	0.483	0.498	0.444	0.493	0.489	0.509	0.455	0.499	0.510	0.479
The natural environment - the average value of the measure	0.414	0.415	0.420	0.419	0.407	0.451	0.424	0.423	0.423	**0.479**	0.411	0.402	0.402	0.402	0.421
The economy - the average value of the synthetic measure	0.062	**0.329**	0.077	0.134	0.138	0.102	0.082	0.066	0.056	0.088	0.084	0.093	0.101	0.113	0.096
Transforming rate of development in 2010-2019															
General - the average value of the synthetic measure	0.337	**0.345**	0.343	0.326	0.315	0.321	0.319	0.325	0.338	0.319	0.292	0.310	0.328	0.320	0.321
Human capital - the average value of the synthetic measure	0.455	0.507	0.508	0.468	0.465	0.509	0.485	0.487	0.462	0.488	0.499	0.416	0.510	0.492	0.496
The natural environment - the average value of the measure	0.307	0.292	0.308	0.267	0.305	0.304	0.300	0.307	0.304	0.308	0.301	0.315	0.299	0.293	0.298
The economy - the average value of the synthetic measure	0.324	**0.342**	0.323	0.332	0.277	0.277	0.283	0.289	0.327	0.277	0.223	0.272	0.297	0.289	0.287

Source: own research.

which only one or two NUTS-3 subregions are separate, i.e. Cyprus, Luxembourg and Malta). The biggest developmental discrepancies measured by the value of the synthetic measure were registered in Great Britain. The largest discrepancies in socio-economic development measured by the position within the ranking of the best developed subregions were observed in Germany. The countries with a strong variation in the level of socio-economic development were: Greece, Italy, the Netherlands, France, Poland, Spain, Finland, Bulgaria and Romania, whose capitals mostly counted as 20% of the best developed subregions, and numerous NUTS-3 units were found in the last positions of the ranking. Taking into account the value of the synthetic measure, as well as the position of the respective NUTS-3 subregions within the ranking, the countries showing the lowest developmental discrepancies were: Ireland, Estonia, Latvia and Belgium.

In the results discussed here, the most commonly mentioned were the German, British and French subregions, therefore one may have an illusion that those features are the best (the least) developed. However, a higher frequency of the occurrence of the subregions from those countries results from the fact that in Germany there were 401 NUTS-3 units (30% among all those under study), in the United Kingdom – 179 subregions (13% of the researched group), and in France – 101 units (7%). This is related to the assumption that a NUTS-3 subregion must number – apart from certain extraordinary circumstances – at least 150,000 inhabitants, and at most 800,000 inhabitants, thus the majority of such units were created in the most densely populated EU states, and vice-versa.

Table 5 shows a different attitude to the specification of the level of development from the perspective of the respective countries. The level of socio-economic development was presented, as well as the changes in the NUTS-3 unit subregions of the EU member states, taking account of the average values of the synthetic measure of all the subregions in a particular country. The highest average value of the synthetic measure of the general level of socio-economic development and its changes, as well as the 'economy' subcomponent was registered in Luxembourg. This stems from the fact, however, that in Luxembourg only one NUTS-3 subregion was subcategorized, which at the same time is the third most developed EU subregion. Among the countries in which at least three NUTS-3 units were subcategorised, the highest average value of the synthetic measure of the general level of socio-economic development was registered were Ireland, the Netherlands, the United Kingdom, Denmark and Belgium, while the highest average of the synthetic measure of the changes of the rate of development was observed in Hungary, Romania and Estonia. The highest level of development of the 'human capital' component in 2019 was registered in Ireland, and the biggest change of the subcomponent in the period 2010-2019 was observed in the German subregions. The most significant value of the synthetic measure within the 'economy' subcomponent, among the member states of at least three NUTS-3 subregions in the static dimension, was observed in the Netherlands, and in the dynamic dimension – in Romania. Within the natural

environment, the highest value of the synthetic measure in the static dimension was registered in Slovenia, and the biggest improvement of the state of the environment in 2010-2019 was observed in Estonia. To sum up, it could be thus stated that in the analysis of the respective NUTS-3 subregions, dominated German, British and French units due to their largest number. Taking account of the average value of the respective synthetic measures, the best results were registered for Irish, Dutch and Romanian regions. The abovementioned considerations, the conducted research and the final figures arrived at, may therefore be a stimulus for further studies, including those of future development and shaping the cohesion policy at the subsequent stages.

References

Albulescu, C., and Goyeau, D. (2014). EU Funds Absorption Rate and Economic Growth. *Timisoara Journal of Economics and Business*, 6(20), 25–26.

Arranz, N., F. Arroyabe, C., and Fernandez de Arroyabe, J. C. (2019). The effect of regional factors in the development of eco-innovations in the firm. *Business Strategy and the Environment*, 28(7), 1406–1415.

Baklanov, P. Y. (2020). Geopolitical Factors in Regional Development. *Customs Policy of Russia in the Far East*, *93*(4), 75–83.

Balcerzak, A. P. (2016). Multiple-criteria evaluation of quality of human capital in the European union countries. *Economics and Sociology*, 9(2), 11–26.

Benner, C. (2003). Labour flexibility and regional development: The role of labour market intermediaries. *Regional Studies*, *37*(6–7), 621–633.

Beugelsdijk, S., Klasing, M. J., and Milionis, P. (2018). Regional economic development in Europe: the role of total factor productivity. *Regional Studies*, 52(4), 461–476.

Bystrova, T. Y., Larionova, V. A., Osborne, M., and Platonov, A. M. (2015). Introduction of open E-learning system as a factor of regional development. *Economy of Region*, 4, 226–237.

Charron, N., Dijkstra, L., and Lapuente, V. (2014). Regional Governance Matters: Quality of Government within European Union Member States. *Regional Studies*, 48(1), 68–90.

Churski, P. (2008). Czynniki rozwoju regionalnego i polityka regionalna w Polsce w okresie integracji z Unią Europejską. Poznań: Wydawnictwo Naukowe UAM.

Cristina, I. O. M., Nicoleta, C., Cătălin, D. R., and Margareta, F. (2021). Regional Development in Romania: Empirical Evidence Regarding the Factors for Measuring a Prosperous and Sustainable Economy. *Sustainability*, 13(7).

De Castro, E. A., Marques, J., and Viegas, M. (2018). The Schumpeter creative destruction hypothesis: A spatial assessment on Portuguese regions. *Global Business and Economics Review*, 20(2), 198–212.

Démurger, S. (2001). Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China? *Journal of Comparative Economics*, 29(1), 95–117.

Diebolt, C., and Hippe, R. (2019). The long-run impact of human capital on innovation and economic development in the regions of Europe. *Applied Economics*, 51(5), 542–563.

Downes, R. (1996). Economic transformation in central and eastern Europe: The role of regional development. *European Planning Studies*, 4(2), 217–224.

Dreyer, L. C., Hauschild, M. Z., and Schierbeck, J. (2006). A framework for social life cycle impact assessment. *International Journal of Life Cycle Assessment*, 11(2), 88–97.

Emami-Langroodi, F. (2018). Schumpeter's Theory of Economic Development: A Study of the Creative Destruction and Entrepreneurship Effects on Economic Growth. *SSRN Electronic Journal*.

Eshugova, S. K. (2018). A factor analysis of models for regional development. *International Journal of Engineering and Technology(UAE)*, 7(3.15 Special Issue 15), 340–342.

Eurostat. (2021). Regional statistics by NUTS classifications. Retrieved June 15, 2021, from https://ec.europa.eu/eurostat/web/main/data/database

Feldman, M. P. (1999). The new economics of innovation, spillovers and agglomeration: a review of empirical studies. *Economics of Innovation and New Technology*, 8(1–2), 5–25.

Florida, R. (2002). The economic geography of talent. *Annals of the Association of American Geographers,* 92(4), 743–755.

Friedmann, J. (1983). Life space and economic space: contradictions in regional development. *The Crises of the European Regions*.

Gallego, A., Rodríguez, L., Hospido, A., Moreira, M. T., and Feijoo, G. (2010). Development of regional characterization factors for aquatic eutrophication. *International Journal of Life Cycle Assessment*, 15(1), 32–43.

Grosse, T. G. (2018). EU Cohesion policy and the peripheries of the new Member States. In *Regional Development in Central and Eastern Europe: Development processes and policy challenges*.

Hauke, J., and Kossowski, T. (2011). Comparison of the values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87–93.

Henrekson, M., and Jakobsson, U. (2003). The transformation of ownership policy and structure in Sweden: Convergence towards the Anglo-Saxon model? *New Political Economy*, *8*(1), 73–102.

Illeris, S. (1993). An inductive theory of regional development. *Papers in Regional Science*, 72(2), 113–134.

Jašková, D., and Havierniková, K. (2020). Human resources as an important factor of regional development. *International Journal of Business and Society*, 21(3), 1464–1478.

Khasanova, A., Amirova, N., and Sargina, L. (2020). Natural resource potential as a factor of regional development. *CITISE*, *24*(2).

Krugman, P. (1998). What's new about the new economic geography? *Oxford Review of Economic Policy*, 14(2), 7–17.

Li, W., Wang, J., Chen, R., Xi, Y., Liu, S. Q., Wu, F., Masoud, M., and Wu, X. (2019). Innovation-driven industrial green development: The moderating role of regional factors. *Journal of Cleaner Production*, 222, 344–354.

Liu, G., Liu, Y., and Zhang, C. (2018). Factor allocation, economic growth and unbalanced regional development in China. *World Economy*, 41(9), 2439–2463.

Martin, R. (2015). Rebalancing the spatial economy: The challenge for regional theory. *Territory, Politics, Governance*, 3(3), 235–272.

Martin, R., and Sunley, P. (1998). Slow convergence? The new endogenous growth theory and regional development. *Economic Geography*, 74(3), 201–227.

Matsuura, M. (2015). Contribution of EU Funds to Economic Growth in Poland. *Russian and East European Studies*, 2015(44), 87–98.

Milenković, M., Vaseashta, A., and Vasović, D. (2021). Strategic planning of regional sustainable development using the factor analysis method. *Polish Journal of Environmental Studies*, 30(2), 1317–1323.

Moroshkina, M. (2020). Geographical Factor in Regional Development. *Vestnik Volgogradskogo Gosudarstvennogo Universiteta. Ekonomika*, 4*,* 90–98.

Mukhametzhan, S. O., Junusbekova, G. A., and Daueshov, M. Y. (2020). An econometric model for assessing the asymmetry of urban development as a factor of regional economic growth: The case of Kazakhstan. *Industrial Engineering and Management Systems*, 19(2), 460–475.

Naydenov, K. (2019). Human resources development as a factor for regional development. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 19(5.4), 475–481.

Nistor, R. L., and Glodeanu, A.-C. (2014). Regional economic development influenced by the EU funds absorption rate. *Managerial Challenges of the Contemporary Society. Proceedings* 7(1), 115–118.

Nowak, P. (2018). Regional variety in the quality of life in Poland. *Oeconomia Copernicana*, *9*(3), 381–401.

Orlova, L., Gagarinskaya, G., Gorbunova, Y., and Kalmykova, O. (2018). Start-ups in the field of social and economic development of the region: a cognitive model. *Entrepreneurship and Sustainability Issues*, 5(4), 795–811.

Pfirrmann, O. (1995). Path Analysis and Regional Development: Factors Affecting RandD in West German Small and Medium Sized Firms. *Regional Studies*, 29(7), 605–618.

Reiff, M., Surmanová, K., Balcerzak, A. P., and Pietrzak, M. B. (2016). Multiple criteria analysis of European Union agriculture. *Journal of International Studies*, *9*(3), 62–74.

Rodionov, D. G., Kudryavtseva, T. J., and Skhvediani, A. E. (2018). Human development and income inequality as factors of regional economic growth. *European Research Studies Journal*, 21(Special Issue 2), 323–337.

Shikverdiev, A. P., Oganezova, N. A., Mazur, V. V., Obrezkov, N. I., and Ichetkina, M. A. (2019). Development of regional competitiveness as a factor in creating a favorable business environment. *Espacios*, 40(28).

Smętkowski, M. (2015). Spatial patterns of regional economic development in central and eastern European countries. *Geographia Polonica*, 88(4), 539–555.

Smętkowski, M., and Dąbrowski, M. (2019). Economic crisis, Cohesion Policy and the eroding image of the European Union at the regional level. *Regional Science Policy and Practice*, 11(4), 713–732.

Spychała, M. (2020). The absorption of EU funds and the socio-economic development in the subregional dimension in Poland. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 64(3), 78–91.

Vučković, S. Đ., Đorđević, J., Jovanov, J. M., Bibić, L. I., and Đorđević, T. (2018). Socio-economic characteristics as limiting factors of regional development. The case of Kolubara District, Republic of Serbia. *Romanian Journal of Geography*, 62(2).

Yun, J. H. J., Won, D. K., Park, K. B., Yang, J. H., and Zhao, X. (2017). Growth of a platform business model as an entrepreneurial ecosystem and its effects on regional development. *European Planning Studies*, 25(5), 805–826.

Zemtsov, S. P., and Smelov, Y. A. (2018). Factors of regional development in Russia: Geography, human capital and regional policies. *Zhournal Novoi Ekonomicheskoi Associacii*, 40(4), 84–108.

Received: September 2021, revised: November 2021

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Cristi Frenţ**

Regionalizing a Tourism Satellite Account: A top-down approach based on existing data sources

Purpose: Developing Tourism Satellite Account (TSA) at regional level is a challenging task for any country that wishes to have data for its component regions or its sub-national entities. There are numerous conceptual and methodological issues to be faced in such an endeavour. Considering data availability, particularly the lack of demand-side data at regional level, the purpose of this paper was to employ a top-down method for regionalizing TSA data in Romania at the level of its eight development regions.

Design/methodology/approach: This method is based on using multiple regional indicators and existing data sources coming from the supply side: tourism statistics, passenger transport statistics, culture statistics, administrative data, structural business survey, labour cost survey, as well as national and regional accounts.

Findings: Regionalizing TSA was obtained but restricted only to calculating the TSA aggregates at regional level.

Originality: This paper will contribute to enhancing the TSA development at sub-national (regional) level by proposing a quick top-down method based only on the existing data sources.

Research limitations/implications: The major limitation is the lack of a direct reconciliation of data between regional supply and regional demand which is actually the philosophy of any TSA.

Practical implications: The regionalization of TSA data proposed in this paper can be illustrative for countries having national TSAs and wishing to make advancements at sub-national (regional) level.

Keywords: tourism satellite account (TSA); regional tourism satellite account (RTSA); top-down approach; regionalization; regional statistics; Romania

JEL Classification: R12, Z30
DOI: 10.15611/aoe.2023.1.10

Quote as: Frenţ C. (2023). Regionalizing a Tourism Satellite Account: A top-down approach based on existing data sources. *Argumenta Oeconomica*, 1(50), 205-226.

* National Institute of Research Development in Tourism, Bucharest, Romania.
ORCID: 0000-0001-5589-1240.

1. Introduction

Since tourism does not exist as a separate sector in economic statistics, particularly in the national accounts, the statistical instrument of Tourism Satellite Account (TSA) was officially proposed more than 20 years ago in order to measure in a standardised manner the tourism's contribution to the economy. In 2010 there were a total of 60 countries that embraced the TSA project at national level according to an assessment made by the World Tourism Organization (2010). Nevertheless, recently at regional (sub-national) level only 14 countries were identified in 2019 to have a regional TSA (Frenţ and Frechtling, 2020). At the same time, it has to be considered that developing TSA at sub-national (regional) level is a much more challenging issue compared with national TSA due to "differences in statistical resources and systems, in policy priorities and in technical capabilities between regions" (Dwyer et al., 2020).

In a certain perspective, one can say that, especially the literature on regional tourism satellite accounts (RTSAs) is as old as TSAs in general. The first academic paper on this issue seems to be that by Rütter and Berwert (1999). This was even before the initial official adaptation of the internationally agreed document on TSA by the World Tourism Organization (WTO) in 2001. In 1999, Rütter and Berwert proposed a pragmatic way of applying TSAs in some regions of Switzerland which was considered to be an input for establishing TSA at national level.

At institutional level, discussions on RTSAs were illustrated by Quevedo (2002), and later officially by the WTO (2005). Actually the paper of Quevedo (2002) was considered to be "the first time UNWTO referred to a regional TSA" (INRouTe, 2016). In the same year of 2002, important contributions to the topic of regionalization of TSA from the Spanish perspective (seen as a medium-term project at that time) were made by Cañada-Martinez (2002). After some years of international consultations one should also mention the 2005 WTO conference held in Iguazu Falls where a special session was devoted to "Tourism Satellite Accounts: The Regional Perspective". The central paper of this conference prepared by Jones (2005) concluded that regions should envisage a regional TSA only where there are policy needs as well as statistical systems and proper human resources.

Meanwhile, some pioneering initiatives on regional TSA were made by countries such as Norway (Braendvang et al., 2001), Canada (Barber-Dueck and Kotsovos, 2002), the United Kingdom – Wales and Scotland (Jones et al., 2003; University of Strathclyde, 2003), Spain – Andalusia (Working Group, 2004), Denmark (Zhang, 2005), Finland (Konttinen, 2006) and Australia (van Ho et al., 2008; Pham et al., 2009).

An important contribution to the topic of RTSAs was also made by Frechtling (2009) in the main paper of the 2008 UNWTO Malaga conference on "Measurement and analysis of tourism economic contribution for sub-national regions through Tourism Satellite Account". The author proposed some principles for TSA validation

and discussed three scenarios for TSA development: under the current TSA condition, under the outdated TSA condition and under the non-existent TSA condition; more precisely, when a country has a TSA, two options are considered for developing a regional TSA: (A) elaborating TSA based on input-output table of a region and (B) distributing the TSA macroeconomic aggregates among the regions by a set of indicators (p. 189). Option (B) is envisaged in this paper. However, Frechtling (2009) warned that "this approach does not produce proper TSAs at regional level, as it is inherently a modelling exercise" and used the term "Experimental regional TSA – ER-TSA" (p. 196). In this paper one wants to challenge that statement by proposing a method where national TSA data are disaggregated at the level of its component regions (sub-national entities). While the author of this paper agrees with the term ER-TSA, he suggests that the top-down approach (proposed in this paper) does produce certain TSA results at regional level that cannot be considered strictly a modelled exercise.

Meanwhile, the topic on regional TSAs has also emerged in other countries, such as Austria (Smeral, 2010), France – Reunion Island (Perrainn and Jean-Pierre, 2011), India (Pandey and Singh, 2013), Belgium – Flanders (Weekers and Maesschalck, 2014), Poland (Skalska and Dziedzic, 2014), Italy (Maresca, 2014) and Portugal – Madeira and Azores (Direção Regional de Estatística da Madeira DREM, 2019a, 2019b; Serviço Regional de Estatistica dos Açores SREA, 2018). Some RTSAs initiatives of other regions within a country should also be added, for instance in Spain – Community of Madrid, Basque country, Canary Islands (Cañada, 2013).

At the same time, it is important to include Cañada (2013) as a specific paper for regional TSA published by UNWTO which provides some general guidelines for developing a regional TSA admitting to be "experimental in character but sufficiently complete to cover the essential objectives of a TSA" (p. 30). The author stresses the importance of developing interregional origin/destination matrices for tourism consumption while recognizing the origin/destination matrices to be "one of the most difficult fields for estimating RTSA". Meanwhile, the author considers two fundamental pillars for a RTSA: regional Supply and Use Table (or a partial set of regional accounts) and the system of tourism statistics at regional level "used to move up from accounting data to the estimation of a TSA" with examples for the community of Madrid (p. 31).

Cañada (2013) named two approaches for developing a regional TSA: regionalization versus regional estimation. When referring to the first approach it is said that it "attempts to apportion territorially certain parts or variables of an available national TSA, using different indicators and methods" (p. 1); however Cañada's paper failed to provide any details on how this apportionment should be made admitting that his paper is "in line with the second approach". This second approach envisages a separate TSA for a region developed in the same manner as a national TSA; obviously, the second approach is more suitable but it is a costly exercise as it should be applied only by the regions where tourism is well developed, and moreover

where there is a strong regional administration as stated by the WTO (2005). A combination of these two approaches has also emerged, and this was well illustrated in the case of Australia (Dwyer et al., 2016), and described as a "hybrid approach (part bottom-up part top-down)".

This paper wants to additionally report about another country in the regional TSA developments around the world, namely Romania as a case study for a centralised country in terms of administrative regional affairs, and thus in the official statistical system. A new perspective is highlighted showing a method of regionalization of its TSA aggregates by using a pure top-down approach which is based only on the existing data sources. This is an evidence-based case that can serve as an example for other countries wishing to produce their own rough estimates of TSA at regional level. Needless to say, the existence of a national TSA is a precondition to talk about regionalization of a TSA, hence the author considered the "under the current TSA condition" as proposed by Frechtling (2009).

2. Methodology

Starting from 2013, TSA has been presented annually in Romania at national level by the National Institute of Statistics (INS). However, on ad hoc basis, previous experimental TSAs in Romania for the reference years 2001 and 2007 that were prepared by National Institute of Research Development in Tourism (INCDT) within some research projects carried out in the period 2003-2010 also have to be mentioned. INS publishes TSA data every year in mid-December in a specific publication also presenting detailed data (in eight tables of TSA results). The following TSA aggregates are calculated at national level in accordance with the TSA:RMF (2008) methodology: Internal Tourism Expenditure, Internal Tourism Consumption, Gross Value Added of Tourism Industries, Direct Tourism Gross Value Added, Direct Tourism Gross Domestic Product and Employment in the tourism industries.

2.1. A step-by-step approach

In order to regionalize these TSA aggregates, the following eight steps were employed:

I. Perform a general assessment of the data sources on which national TSA is based (see Appendix A in the case of Romania). The purpose is to identify the regional breakdown of the data used in the TSA compilation. In principle and if feasible, all data sources used at national level should be also employed at regional level. However, this is not a recipe for success, as many of these data sources might provide data only at national level.
II. Take the national TSA data as reference values in the process of regionalization of TSA aggregates. In essence, the top-down approach means the regional allocation of TSA data having as a starting point the national TSA.

III. Since both national and regional accounts are involved in TSA regionalization (see Figure 1), the regional accounts as a point of reference are also taken as (Cañada-Martinez, 2002).

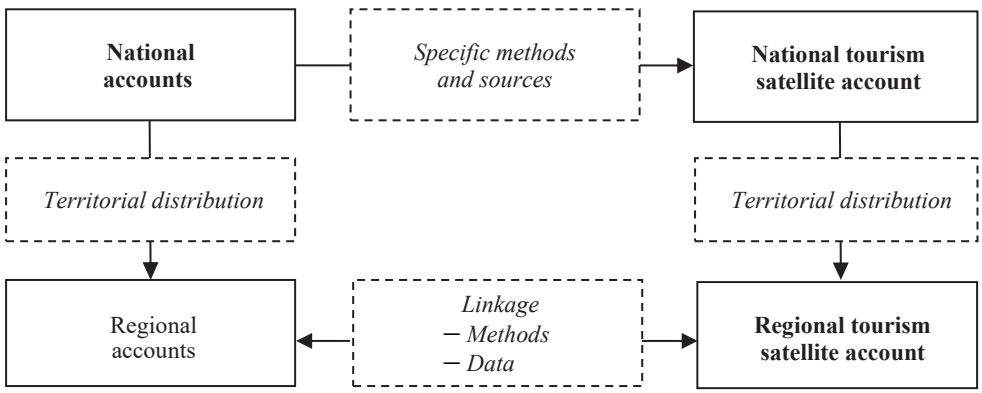


Fig. 1. Regionalization of TSA based on National and Regional Accounts.

Source: Cañada-Martinez (2002), p. 91.

It is important to mention that in Romania, developing regional accounts are mainly based on a top-down approach, while mixed methods are applied for some industries like electricity production and distribution, mining, fishery, postal services, chemical industry (INS, 2019, p. 15) – but none of these are related to tourism industries. To be more precise, the rules from regional accounts regarding the distribution of gross value added by industries and regions were also applied in the case of tourism industries (i.e. different indicators in the case of regionalizing air and rail transportation industries).

IV. Setting the general approach – regionalize each of the TSA aggregates through a specific procedure. Details on the regionalization procedure of each of these aggregates for Romania are presented in Table 1. The data sources pertaining to each regionalized TSA aggregated in Romania are presented in Appendix B.

V. Identifying regional indicators to be used to regionalize national TSA data. In this endeavour, in the regionalization process the indicators used are "as close as possible to the variable to be estimated" (Eurostat, 2013, p. 34). This method is facilitated by the existence in each region of a uniform system of data collection. WTO (2005) considers that for the top-down approach to be feasible in a country having national TSA "it is essential to have access to a set of homogeneous tourism-related regional indicators so that the national aggregates may be regionalized" (p. 23). These regional indicators were identified in Romania based on available data sources (see Table 2).

Table 1

Regionalization procedure of the TSA aggregates in Romania

TSA aggregate	Regionalization procedure
Internal Tourism Expenditure	Different procedures for Inbound tourism expenditure and Domestic tourism expenditure. Inbound tourism expenditure is regionalized at the level of total expenditure (no products breakdown) while domestic tourism expenditure is regionalized at the level of each constituent product. In the case of domestic tourism expenditure, no separation is made between overnight tourists and same-day visitors while for inbound tourism expenditure, separate estimation procedures are employed for these two segments.
Internal Tourism Consumption	Regionalization at the level of total expenditure (no products breakdown) resulted from summing the components of Internal Tourism Consumption
Gross Value Added of Tourism Industries (GVATI)	Regionalization at the level of each tourism industry based on National and Regional Accounts practice on disaggregating gross value added by each industry
Direct Tourism Gross Value Added (TDGVA)	Regionalization at the level of each industry based on domestic tourism consumption breakdowns by products
Direct Tourism Gross Domestic Product (TDGDP)	For the constituent part (i.e. Taxes less subsidies on products) the same procedure as for TDGVA will be employed.
Employment in the tourism industries	Regionalization at the level of each tourism industry

Source: own elaboration.

VI. Calculating regional distribution keys based on indicators identified in the previous step. Except for data from Family Budget Survey (for Other products category and Country-specific products category – in the case of domestic tourism expenditure), all indicators can be used to provide each year some distribution keys to be further applied to the national data. However, when employing Family Budget Survey data, in addition some disparity indexes were calculated in relation with the national level that were further used and to derive indirectly the related regionalized products (i.e. for Other products and Country-specific products). It should be noted that Romania does not have a tourism regionalized demand-side survey.

VII. Regionalizing TSA data based on the distribution keys and deriving regionalized TSA aggregates. Distribution keys are the input to regionalize TSA data. After TSA data is regionalized, the calculation of TSA aggregates at regional level is performed as a sum of its regionalized components. For instance, regionalized Internal tourism consumption is the sum of regionalized inbound tourism expenditure, regionalized domestic tourism expenditure and regionalized Other components of tourism consumption.

Table 2

List of indicators with regional breakdown used for regionalizing TSA aggregates in Romania

TSA aggregate / Component of the aggregate	**Indicators**
Internal tourism consumption / Inbound tourism expenditure	Foreign arrivals at border points (excluding airports) from citizens of the neighbouring countries (only in the case of same-day visitors) Number of arrivals of non-residents in accommodation establishments Number of overnight stays of non-residents
Internal tourism consumption / Domestic tourism expenditure	Turnover of local units
	Number of overnight stays of Romanian residents
	Number of restaurant services provided by accommodation establishments
	Number of embarked passengers in domestic traffic (by airport of origin)
	Estimated number of passengers at regional level calculated/derived from: Average revenue per train-kilometre in 2016 Number of train-kilometres operated Length of transport routes operated by trains (km)
	Number of passengers-kilometres in interurban and international road transport (only regular and occasional services taken into consideration)
	Number of passengers crossings with ferry (inland) Departures of Romanian citizens abroad by border points with ferries
	Number of visitors at museums
	Data from Family Budget Survey – category Expenditure for service payments – Payments for tourist services and COICOP category 'Other products and services'
Internal tourism consumption / Other components of tourism consumption	Number of holiday homes Number of dwellings Data from Family Budget Survey – category Expenditure for service payments – Payments for rent Number of treatment tickets supported by social insurance by spa localities Ticket price for each accommodation unit in spa localities Number of bed – places for particular accommodation establishments located in spa localities
Gross Value Added of the Tourism Industries	Expenditure for remuneration of employees (wages and salaries) Turnover of local active units Gross salaries Average number of employees Average gross monthly salary
Employment in the tourism industries	Personnel employed in active local units Personnel employed in museums at the 31th of December

Source: own elaboration.

VIII. Data validation through various checks with some related indicators (e.g. by calculating location quotients based on the tourism intensity indicator defined as the ratio between number of overnight stays and total resident population). This final step is to assure the reliability of obtained data.

2.2. Specificities of the proposed method

Overall, even if the regionalization of TSA in Romania is based exclusively on a top-down method, and what is specific is that it is based only on the existing data sources (and moreover without either a regional demand-side survey or a regional Input-Output table). However, a particularity emerges in the regionalization of Tourism Direct Gross Value Added (TDGVA). Due to the lack of regional input-output tables, there was no reconciliation between production at regional level and regionalized internal tourism consumption at the level of each tourism product/industry. Instead, each portion of tourism gross value added related to an industry was regionalized based on the distribution keys derived from the domestic tourism

Table 3

Simplified form for deriving Tourism Direct Gross Value Added

	Tourism industries				**Other industries (residual)**		**Output of domestic producers (at basic prices)**	
	Accommodation for visitors		**...**					
	Output	**Tourism share (1)**	**Output**	**Tourism share (...)**	**Output**	**Tourism share (13)**	**Output**	**Tourism share (total)**
Production (A)								∑
Intermediate consumption (B)								
Gross Value Added (C) = (A) – (B)								∑

Notes:

Internal Tourism Consumption (∑ = Internal Tourism Consumption in the country/region)

Tourism Gross Value Added (∑ = Tourism Gross Value Added in the country/region)

Source: adapted upon TSA:RMF (2008).

consumption of the corresponding product. This is due to the fact that over 80% of internal tourism consumption at national level is based on domestic tourism consumption (86.3% in 2016).

It is important to mention that in the TSA, the aggregate of TDGVA is derived from internal tourism consumption, as a result of a reconciliation between demand and supply – in fact the essence of the TSA framework (see Table 3). Moreover, it is believed that under the circumstances where there is a direct proportionality between gross value added on the one hand, and intermediate consumption and production on the other, regionalizing TDGVA in a similar manner with internal tourism consumption is fully justified (see Table 3 for a simplified representation of deriving TDGVA). It should be stated that internal tourism consumption is equal to tourism output at market prices for tourism services if imports are not considered. A somewhat similar simplified procedure for estimating TDGVA, but limited only to national level, is also found at UNWTO (2018). This confirms that the method proposed in this paper has in a way the same validity from the methodological point of view as that proposed by UNWTO (2018).

In addition, the representation from Table 3 is only illustrative and does not consider taxes on products. In fact, TDGVA (at basic prices) does not include taxes on products while Internal Tourism Consumption includes those taxes which refer to value added tax, excise duties on tobacco and alcohol, etc. Yet, in the regional accounts practice, the regionalization of taxes is carried out in a similar manner as with Gross Value Added (GVA), since by convention these taxes "are allocated on the basis of relative size of GVA of all industries in the region, valued at basic prices" (Eurostat, 2013 p. 52). Therefore, in this simplified example from Table 3, it is reasonable to assume that the issue of exclusion of taxes will not greatly influence the regionalization of the main TSA aggregates.

3. Results

3.1. The TSA aggregates at regional level

The following aggregates were calculated for each of the eight development regions of Romania (defined at NUTS 2 level), annually in the period 2011-2017:

- Internal Tourism Consumption (ITC),
- Gross Value Added at Tourism Industries (GVATI),
- Direct Tourism Gross Value Added (DTGVA),
- Direct Tourism Gross Domestic Product (DTGDP),
- Employment in the Tourism Industries (ETI).

All these are presented in Table 4 for the year 2017.

Table 4

TSA aggregates by regions in Romania in 2017

Regions	ITC	GVATI	DTGVA	DTGDP	ETI
	Millions RON				*Persons*
North-West	5,707.1	5,072.3	2,605.6	2,675.2	50,533
Centre	6,875.9	5,182.6	3,558.2	3,566.9	48,126
North-East	4,606.7	3,856.9	2,106.7	2,154.2	39,500
South-East	5,798.5	4,639.9	3,414.2	3,400.1	42,895
South-Muntenia	3,818.3	3,531.5	1,962.6	1,975.3	29,778
Bucharest-Ilfov	17,116.1	13,557.7	6,942.1	7,473.8	104,656
South-West Oltenia	1,939.7	2,394.0	962.6	979.7	25,367
West	3,693.3	3,416.5	1,644.6	1,683.8	32,219
Total Romania	**49,555.5**	**41,651.3**	**23,196.5**	**23,909.0**	**373,074**

Source: INCDT (2020).

One should note that these aggregates are in line with TSA:RMF (2008) provisions. The only minor difference is given by the lack of Internal tourism expenditure as aggregate, in this case very close to Internal tourism consumption, so it would have been somehow redundant to use both aggregates.

3.2. Components of internal tourism consumption at regional level

For each region, internal tourism consumption at regional level was calculated as the sum of three components at regional level:

- Inbound tourism expenditure,
- Domestic tourism expenditure,
- Other components of tourism consumption.

It should be kept in mind that the concepts of inbound and domestic tourism consumption are similar to those used at national level, hence there is no distinct terminology employed at the level of regions.

From an analytical perspective, one can calculate an indicator representing the level of internationalization of each region's internal tourism consumption, in other words the share of inbound tourism expenditure in total internal tourism consumption in a region (see Table 5). At national level, this share is around 15%. In the period 2011-2017, two regions (Bucharest-Ilfov and Centre) had a level of internationalization higher than the national average. At the bottom end there was South-East region where the level of internationalization was the lowest (it was only 5.9% in 2017).

Table 5

Level of internationalization of internal tourism consumption in each region, 2011-2017

Regions	2011	2012	2013	2014	2015	2016	2017
North-West	15.0%	14.1%	12.6%	12.1%	15.3%	16.4%	13.2%
Centre	24.6%	25.8%	21.5%	18.0%	19.1%	20.4%	16.9%
North-East	13.2%	11.9%	12.2%	11.6%	13.3%	14.8%	12.7%
South-East	8.2%	8.0%	7.9%	6.4%	6.0%	6.7%	5.6%
South-Muntenia	16.1%	17.0%	13.9%	9.1%	11.7%	10.5%	8.7%
Bucharest-Ilfov	24.0%	20.4%	23.1%	22.9%	21.9%	20.5%	17.6%
South-West Oltenia	8.7%	13.4%	11.3%	8.7%	11.2%	11.7%	9.8%
West	17.9%	17.9%	17.5%	15.5%	15.8%	16.8%	13.3%
Romania (national level)	**18.6%**	**17.4%**	**17.5%**	**15.9%**	**16.4%**	**16.5%**	**13.7%**

Source: INCDT (2020).

3.3. Regional tourism as a share of national tourism

Some shares from national TSA aggregates can be illustrated as regional distributions (see Table 6). These reflect the importance of regional tourism within national tourism.

Table 6

Distribution of Tourism Direct GDP by regions, 2011-2017

Regions	2011	2012	2013	2014	2015	2016	2017
North-West	11.4%	11.0%	11.4%	10.6%	11.6%	11.3%	11.4%
Centre	10.9%	11.5%	15.1%	14.8%	13.9%	13.5%	15.3%
North-East	8.8%	8.5%	8.0%	8.3%	9.4%	10.2%	9.2%
South-East	13.1%	18.3%	15.2%	15.2%	14.9%	14.4%	14.9%
South-Muntenia	6.9%	6.2%	7.4%	9.7%	8.3%	8.8%	8.8%
Bucharest-Ilfov	36.2%	32.9%	31.9%	30.3%	30.4%	30.9%	28.9%
South-West Oltenia	4.0%	3.8%	4.1%	4.6%	4.1%	4.1%	4.3%
West	8.7%	7.8%	7.1%	6.5%	7.4%	6.8%	7.3%
Total Romania	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Source: INCDT (2020).

For instance, one can calculate Regional Tourism Gross Domestic Product as a share of the country's Tourism Gross Domestic Product. One can see that the Bucharest-Ilfov capital region ranks first, followed by the Centre and South-East regions, while at the bottom end there are the South-West Oltenia and West regions.

Besides those, it is important to note the decreasing share of the capital region, which is in contrast with the slow increase of the share for the Centre and North-East regions.

3.4. Tourism's direct contribution to the regional economy

For each region, one can compute the tourism's contribution to the regional GDP and regional GVA since the latter aggregates are available at NUTS 2 level (in the European Union there is the Nomenclature of Territorial Units for Statistics (NUTS) system used to classify the sub-national territories at three levels which are defined according to population threshold: NUTS 1 from 3 million to 7 million residents, NUTS 2 from 800,000 to 3 million, and NUTS 3 from 150,000 to 800,000 (Eurostat, 2020b)). This shows the economic significance of tourism in the regional economy, and can also be compared with the national level (tourism's contribution to the GDP in the country). Only three regions posted levels of contributions to their regional GDP superior to the national level: South-East (4.0%), Centre (3.8%) and Bucharest-Ilfov (3.1%). The same patterns are seen in the case of DTGVA (see Table 7).

Table 7

Tourism's direct contribution to regional and national economy in Romanian 2017 (%)

Regions	GVATI	DTGVA	DTGDP	ETI*
North-West	5.3	2.7	2.6	7.4
Centre	5.9	4.1	3.8	7.5
North-East	4.8	2.6	2.4	7.3
South-East	5.8	4.3	4.0	8.4
South Muntenia	3.9	2.1	2.0	5.0
Bucharest-Ilfov	6.4	3.3	3.1	8.1
South-West Oltenia	4.1	1.7	1.5	6.7
West	4.7	2.2	2.1	5.8
Total Romania	**5.4**	**3.0**	**2.8**	**7.2**

* Based on number of employed persons taken from the Structural Business Survey

Source: INCDT (2020).

A particular case is the aggregate of Employment in the tourism industries, calculated in this table independently from national TSA data, using exclusively the indicator of number of employed persons (in local active units) from Structural Business Statistics, where activities from agriculture were excluded. One can consider this indicator as a proxy of estimating the regional importance of tourism from the labour force perspective. This situation occurred due to the fact that national TSA data does not include total employment figures for all economic activities, which would have allowed the calculation of a share of employment in the tourism

industries in total employment at national level. In addition, from a methodological point of view it would not have been correct to use total employment figures from other data sources (i.e. national accounts) since there is not the same methodology employed in the compilation of national TSA employment data. In any case, the values presented in Table 7 confirm to some extent the same hierarchy of regions in terms of their regional importance of tourism.

4. Discussion and conclusions

This paper proposes a pure top-down approach for regionalizing TSA using a set of specific indicators available at regional level. While recognizing the need to have proper tourism-related regional indicators, WTO (2005) admits that due to the specific nature of tourism and its territorial characteristic, the existence of homogenous indicators cannot always be guaranteed in advance (p. 23). Indeed, the fact that there are some homogenous indicators available for regions is not a guarantee that the regionalization process is made properly. It is always questionable to what extent these indicators are relevant, and reflect completely the tourism characteristics of a region. The latter issue was not solved in this paper. It is believed that only a different approach of regional TSA compilation (i.e. a bottom-up or a hybrid approach) is proper in such endeavour.

It is supposed that if there are no data sources specific for tourism at regional level (i.e. tourist surveys for a region) and the methods used in the regional accounts compilation in a country are developed in a top-down approach, the choice of a pure top-down approach for TSA regionalization is feasible in the lack of any other alternative.

The level of territorial disaggregation is an important aspect to be discussed. The question is to what territorial level TSA should be developed. This paper applied a pragmatic approach and used only the NUTS 2 level (as territorial level), not going deeper to NUTS 3 level. The reason for this was twofold: some supply-side data sources (i.e. structural business statistics which provides the core supply-side data in the regional disaggregation) provide data only at NUTS 2 level; the lack of any demand-side data for tourism at regional level is a serious constraint and choosing NUTS 3 level would have posed much more difficulties than NUTS 2 level. Moreover, it is considered that the proper territorial level used for a TSA cannot be established universally since a tourism destination cannot always be defined by its administrative borders (OECD, 2010).

4.1. Limitations and delimitations

An important limitation is given by the lack of any data to construct an origin-destination matrix for tourism consumption. Hence tourism expenditure between regions was not quantified properly (there was no survey capturing expenditure

neither at the place of origin nor at the place of destination). Instead, the top-down regionalization of TSA data based on properly-chosen indicators tried to eliminate this major lack. Being aware that this is far from being an accurate approach, in a pragmatic way, one must admit that it is the only solution to tackle the lack of data. Additionally, Jones at al. (2009) admitted that "national surveys are not always stratified to ensure an adequate return for each industry activity for each region" which is an important difficulty of a top-down approach (p. 305).

At the same time, the author is aware that the pure top-down approach applied in this paper complies only partially with the two fundamental pillars proposed by Cañada (2013) due to the lack of a system of tourism statistics at regional level in Romania, in particular the lack of regional demand-side data. The lack of regional SUT is another shortcoming, but this is compensated by the existence of the alternative proposed by Cañada (2013), namely "partial set of regional accounts" (p. 31) which are "with reference to the example of the European Union, the ESA regional accounts and regional system (A/N which) are confined to a limited set of accounting elements" (p. 23). As an EU member state since 2007, Romania has had such regional accounts as the country has to comply with European regulations in terms of National Accounts statistics. In other words, gross value added and gross domestic products are produced at NUTS 2 level, so the regions have these major economic aggregates already calculated.

Another issue refers to the impossibility of compiling any TSA results disaggregated by regional forms of tourism (i.e. regional inbound tourism expenditure and regional domestic tourism expenditure). It should be recalled that when speaking strictly at regional level, inbound tourism includes not only foreign tourists (non--residents for the country where the region is located), but also residents from different regions of the country of reference. Moreover, at regional level domestic tourism expenditure would refer only to the expenditure of residents of the region of reference made inside this region. However, in all the data compilation the paper used only the concepts from national level (having the same meaning as at national level) that were further disaggregated by regions.

In other words, at regional level there was no specific terminology adopted (and used) since there was no demand-side data available at regional level. Consequently, this is maybe the major limitation of the top-down approach since there was no reconciliation between domestic supply and demand at regional level which is the core *modus operandi* of any TSA. However, this reconciliation is practically impossible in the absence of regional input-output tables. Therefore, it should be acknowledged that in this case, TDGVA and TDGDP at regional level are not derived directly from this reconciliation. TSA:RMF (2008) clearly states that (at national level) these are "indicators emanating from a reconciliation of tourism consumption and supply, and their values will depend on the scope of measurement of tourism consumption that a country adopts" (p. 48). Nevertheless, in order to respect these principles to some extent, regionalization of TDGVA was made taking internal

tourism consumption at regional level (more precisely, domestic tourism consumption) as a proxy indicator used in this regionalization process. It is believed that using distribution keys coming from the regionalization of a demand aggregate (i.e. part of internal tourism consumption) is much more relevant that using the distribution keys coming from the regionalization of gross output, namely gross value added of tourism industries (this also comprises the non-tourism production). The author has to admit that the whole process is in fact simply a redistribution by regions of the national TSA aggregates using different indicators and data sources.

Nevertheless, the experimental character of the results obtained has to be always kept in mind. It was the first exercise on the TSA regionalization in Romania through a method based on assumptions applied to various data sources. Apart from tourism indicators, data coming from other related fields were integrated, most notably transportation and culture. However, these fields provide statistics that are not always designed to respond to the specificity of tourism. Put more simply, not all museum visitors can be considered tourists and not all passengers travelling have tourism related purposes. Yet, there are always some assumptions that have to be made without having the perfect statistics (i.e. assuming that long-distance travel is always part of tourism or museum visitors are mostly tourists).

INRouTe (2016) admits that in practice "it is not so easy to measure what one region produces and which part of it is consumed by visitors in another region" (p. 123). Only a specific demand-side survey carried out in each region would provide some indication about these measurements. Unfortunately, this was not the case of Romania, which lacks any tourism surveys carried out from demand side at regional level. This was a major constraint of the applied method.

OECD (2016) saw RTSA as "an irregular project and a one-off exercise" (p. 11). However, the author of this paper wanted to prove that this irregularity of regional TSA can be tackled with a method of regionalization of TSA using a top-down approach. This might have the potential to encourage other countries with national TSAs to develop a similar exercise. Finally, more TSA applications at regional level may demonstrate that there is always room for improvements and future developments in this field.

Thus, one can sum up some important delimitations of this paper:

- The lack of any direct relationship between regional tourism demand and regional tourism supply at the level of regional tourism products as this is a core part of any TSA.
- Regionalization of TDGVA aggregate at the level of each tourism industry (which is based only on domestic tourism expenditure breakdowns by products) does not in fact consider the inbound tourism expenditure. This was the case even if one is aware that there are different levels of internationalization of internal tourism consumption in each region, which can influence somehow the results. It is believed that regions that have a lower level of internationalization of tourism consumption (i.e. the South East, South Muntenia and South West Oltenia

regions) will be more accurately reflected in the regionalization process. Instead, for the regions with a higher level of internationalization of tourism consumption (e.g. Bucharest-Ilfov, Centre and West), there will clearly be a loss of accuracy. However, the magnitude of this influence is rather unknown at this moment due to the lack of any data in this field. Yet, it is assumed that since for most of the regions the level of internationalization is below 20%, thus the approach taken will not greatly influence the results.

- The issue of neglecting intra-regional tourism, since both domestic tourism expenditure and inbound tourism expenditure are defined in the same way as at national level.
- The estimation of tourism expenditure of non-residents staying in non-commercial establishments (VFR and second/holiday homes) was not envisaged since this type of expenditure is not yet estimated in the national TSA.
- As regards the accuracy of results, one should assume that our approach is far from being very accurate but the magnitude of this potential inaccuracy is indeed unknown until new data sources are developed.

4.2. Conclusion

The regionalization of TSA data proposed in this paper can be illustrative for any country that wishes to have some TSA data at sub-national level, especially for countries that are EU members or those countries that are complying with the ESA regulations on regional accounts and have already developed TSA at national level. According to Eurostat (2019) data collection, 25 EU countries (out of 28 in 2019) provide TSA data. These countries compile regional accounts data that are transmitted annually by Eurostat, so they provide a set of regional accounts data seen as statistical reference in a macroeconomic framework. This is one of the pillars mentioned by Cañada (2013), while the other pillar refers to the system of tourism statistics at regional level. Regarding the latter, in the EU there are accommodation statistics figures available at regional level (NUTS 2), as well as some statistics for transport and culture available at NUTS 2 level. Thus, a synergy between tourism statistics and other related fields (i.e. transportation, culture) should be created at regional level. This will help tourism statistics to take advantage of data coming from other related fields. This paper proved that this synergy with statistics from other related fields (i.e. transportation, culture) could be an option in the absence of demand-side data at regional level.

There will always be a trade-off between data availability and the theoretical and methodological provisions that have to be adopted. The belief is that, above all, it is important that all available data sources are examined and utilized to the maximum extent possible. This was also the approach adopted in this paper by using regionalized existing data sources not all of them coming from tourism but also from other related fields (i.e. transportation, culture). Naturally, the principles and rules of TSA and

regional accounts were envisaged throughout this process, albeit with no direct reconciliation of data between regional supply and demand. This was in fact the major compromise in the absence of specific tourism demand-side data at regional level.

Nevertheless, the paper proved that disaggregated TSA data for regions can be obtained at least at the level of the main TSA aggregates that international standards (i.e. TSA:RMF 2008) have proposed. These aggregates should be interpreted both in relation to the national levels as a share from total country's tourism economy, and in a comparative manner as contributions to the regional economy. Finally, the TSA aggregates provide very useful data in characterizing the economic importance and the size of tourism in each region. Obviously these are in fact very useful key figures used by policy-makers and any other stakeholders in a region.

References

Barber-Dueck, C. and Kotsovos, D. (2002). *The Provincial and Territorial Tourism Satellite Accounts 1996* (Report). Statistics Canada. https://www150.statcan.gc.ca/n1/en/pub/13f0063x/4193714-eng.pdf?st=KMMDaq3s (accessed 6 December 2018).

Braendvang, A., Dybedal, P., Johansen, S. and Sorensen, K. (2001). *Regional Satellite Accounts for Tourism. Data, Concepts, Methods and Applications*. the 41st Congress of the European Regional Science Association, Zagreb, Croatia. http://www-sre.wu.ac.at/ersa/ersaconfs/ersa01/papers/full/195.pdf (accessed 23 November 2018).

Cañada–Martinez, A. (2002). Spain's Tourism Satellite Account (TSA): Methodological considerations regarding the regionalization of transactions and other accounting elements. In World Tourism Organization, *Enzo Paci Papers on Measuring the Economic Significance of Tourism (Volume 2)* (pp. 81–102). World Tourism Organization.

Cañada, A. (2013). *Regional Tourism Satellite Account*. UNWTO Statistics and TSA Issue Paper Series STSA/IP/2013/02 (Online). http://statistics.unwto.org/en/content/papers (accessed 19 November 2018).

Dwyer, L., Forsyth, P., Spurr, R. and Van Ho, T. (2016). *Developing an integrated suite of regional tourism satellite accounts (TSAs): A case study from Australia*. Travel and Tourism Research Association: Advancing Tourism Research Globally. https://scholarworks.umass.edu/ttra/2007/Presented_Papers/48 (accessed 17 January 2019).

Dwyer, L., Forsyth, P. and Dwyer, W. (2020). *Tourism Economics and Policy*. Second edition, Channel View Publications, Bristol, UK.

Direção Regional de Estatística da Madeira DREM (2019a, January 29). *Tourism Satellite Account – 2015*. https://estatistica.madeira.gov.pt/en/download-now-3/economic/contaseconomicas-gb/contaseconomicas-cst-gb/contaseconomicas-cst-quadros-gb/viewcategory/662-tourism-satellite-account-tables.html.

Direção Regional de Estatística da Madeira DREM (2019b, January 29). *ContaSatélite do Turismo da RegiãoAutónoma da Madeira 2015*.https://estatistica.madeira.gov.pt/download-now-3/economic/contaseconomicas-gb/contaseconomicas-cst-gb/contaseconomicas-cst-emfoco-gb/finish/300-em-foco/9711-em-foco-2015.html.

Eurostat (2020a, June 25). *Gross domestic product (GDP) at current market prices by NUTS 2 regions*. https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do.

Eurostat (2020b, April 21). NUTS – Nomenclature of territorial units for statistics – Principles and Characteristics. https://ec.europa.eu/eurostat/web/nuts/principles-and-characteristics.

Eurostat (2019, May 4). *Tourism Satellite Accounts (TSAs) in Europe. Edition 2019*. https://ec.europa.eu/eurostat/web/products-statistical-reports/-/KS-FT-19-007.

Eurostat (2013). *Manual on regional accounts methods*. 2013 Edition, Luxembourg: Publication Office of the European Union.

Frechtling, D.C. (2009). Measurement and analysis of tourism economic contributions for subnational regions through the Tourism Satellite Account. In: World Tourism Organization, *Enzo Paci Papers on Measuring the Economic Significance of Tourism (Volume 6)* (pp. 189-206). World Tourism Organization.

Frenț, C. and Frechtling, D. C. (2020). Regional tourism satellite accounts: Towards international comparability. *Tourism Review*. .

International Network on Regional Economics, Mobility and Tourism, InRouTe (2016). *Tourism, territory and sustainability: a statistical insight at subnational levels. Toward a set of UNWTO Guidelines*. http://www.inroutenetwork.org/wp-content/uploads/2016/12/INRouTe2016.pdf (accessed 23 November 2018).

Institutul Național de Cercetare Dezvoltare în Turism, INCDT [National Institute of Research Development in Tourism] (2020). *Compilarea pilot a unui CST la nivel regional în România [Pilot compilation of a TSA at regional level in Romania]*, Bucharest, October 2020.

Institutul Național de Statistică [National Institute of Statistics], INS (2019). *Regional National Accounts 2012-2016.* https://insse.ro/cms/sites/default/files/field/publicatii/regional_national_accounts_2015-2016.pdf.

Jones, C., Munday, M. and Roberts, A. (2009) Top down or bottom up? Issues in the development of sub-national tourism satellite accounts. *Current Issues in Tourism*, 12(4), 301-313.

Jones, C., Munday, M. and Roberts, A. (2003). Regional Tourism Satellite Accounts: A Useful Policy Tool. *Urban Studies*, 40(13), 2777-2794.

Jones, C. (2005, October 3-6). *Tourism Satellite Accounts: The Regional Perspective* [Paper presentation]. UNWTO conference The Tourism Satellite Account (TSA): Understanding Tourism and Designing Strategies, Iguazu Falls, Argentina/Brazil/Paraguay.

Konttinen, J. P. (2006, June 1-3). Regional Tourism Satellite Account and Economic Effects in Finnish regions (paper presentation). The International Conference on Regional and Urban Modelling, Brussels, Belgium.

Laimer, P. (2012, November 14-16). *Regional Tourism Satellite Accounts in Austria: sufficient information for regional tourism policy?* (paper presentation). 11th Global Forum on Tourism Statistics, Reykjavik, Iceland.

Maresca, S. (2014). Open problems and new possibilities for the development of regional TSA in Italy. In: INRouTe, *Proceedings: INRouTe II international Seminar on Regional Tourism: Moving towards a regional TSA approach,* 8-23.

OECD (2010). *OECD Tourism Trends and Policies 2010*.

OECD (2016). *An OECD Review of Statistical Initiatives Measuring Tourism at SubnationalLevel* (OECDTourism Papers 2016/01), OECD Publishing.

Quevedo, J. (2002). The TourismSatellite Account (TSA) from the regional point of view: Reflections for Debate. In World Tourism Organization, *Enzo Paci Papers on Measuring the Economic Significance of Tourism (Volume 2)*, 159-180.

Pandey, R.N. and Singh, S. (2013, November 6-7). *Status of Regional Tourism Satellite Accounts (TSA) in India – Problems and Challenges*. (paper presentation). MOVE 2013 3rd International Conference on the Measurement and Economic Analysis of Regional Tourism – American Chapter, Medellin, Colombia.

Perrain, D. and Jean-Pierre, P. (2011, April 22). *A new approach to Tourism Satellite Account compilation. An application in Reunion Island*. https://ideas.repec.org/p/hal/journl/hal-01485104.html.

Pham, T. D., Dwyer, L. and Spurr, R. (2009). Constructing a regional tourism satellite account: The case of Queensland. *Tourism Analysis*, 13, 445-460. https://10.3727/108354208788160397.

Rütter, H. and Berwert, A. (1999). A regional approach for tourism satellite accounts and links to the National Account. *Tourism Economics*, 5(4), 353-381.

Serviço Regional de Estatistica dos Açores, SREA (2008). *Contas Satélite do Turismo da Macaronésia – 2001-2002*. https://srea.azores.gov.pt/Conteudos/Media/file.aspx?ida=1968 (accessed 29 January 2019).

Skalska, T. And Dziedzic E. (2014). Methodology of the Regional Tourism Satellite Account for Poland – Concept and Feasibility Study. *Proceedings: INRouTe II International Seminar on Regional Tourism: Moving towards a Regional TSA approach*, 34-57.

Smeral, E. (2010). Application of the TSA in a regional context: the case of Vienna. *Tourism Review*, 65(1), pp. 57-65.

University of Strathclyde (2003). Towards a Tourism Satellite Account or Scotland. Augmented final report, Professor Iain McNicoll, August 2003.

United Nations, Commission for the European Communities Eurostat, Organization for Economic Cooperation and Development and World Tourism Organization (2010). Tourism Satellite Account: Recommended Methodological Framework 2008 (herein called TSA:RMF 2008). Luxembourg, Madrid, New York, Paris: United Nations publication.

United Nations World Tourism Organization, UNWTO (2018). Measuring Sustainable Tourism – Proposals for estimating Tourism Direct GDP with limited data, paper prepared by Carl Obst, UNWTO consultant, Institute for Development of Environmental-Economic Accounting (IDEEA Group).

Van Ho, T., Pambudi, D., Forsyth, R., Spurr R., Dwyer, L. and Hoque, S. (2008). Development of Regional Tourism Satellite Account: A case study from Australia, paper presented at the International Conference on Measuring Tourism Economic Contribution at Sub-National Levels, *Knowledge as Value Advantage for Tourism Destinations*, 29-31 October 2008, Málaga, Spain.

Weekers, K. and De Maesschalck, P. (2014). *TSA (Tourism Satellite Account) Flemish Region and Brussels-Capital Region 2012* [Executive Summary] https://www.toerismevlaanderen.be/sites/toerismevlaanderen.be/files/assets/documents_KENNIS/onderzoeken/TSA%202012-%20Executive%20summary%20.pdf (accessed 21 November 2018).

Working Group for the development of TSA-Andalusia (2004). The singularity of the territory in tourism accounting: Andalusia as a case study. In World Tourism Organization, *Enzo Paci Papers on Measuring the Economic Significance of Tourism (Volume 3)*, 191-209.

World Tourism Organization (2005). *Adapting the National Tourism Satellite Account (TSA) Project to Subnational Levels – a discussion paper*, WTO Department of Statistics and Economic Measurement of Tourism, Madrid.

World Tourism Organization (2010). *TSA data around the world – worldwide summary*, http://statistics.unwto.org/sites/all/files/pdf/tsa_data.pdf (accessed 6 March 2019).

Zhang J. (2005). *Documentation of Regional Tourism Satellite Accounts in Denmark*. https://www.kora.dk/media/272138/udgivelser_2005_pdf_rtsa.pdf (accessed 5 December 2018).

Received: October 2021, revised: January 2022

Acknowledgement: *This work was supported by a grant of Romania's Ministry of Research and Innovation. CNCS - UEFISCDI project number PN-III-P1-1.1-TE2016-0238, within PNCDI III.*

Appendix A

Data Sources for TSA Compilation at national level in Romania and Assessing the Possibilities for Regional Breakdowns

Data source	TSA tables of results	Regional breakdown of data (NUTS 2 level)	Comments
1	2	3	4
ACNER – Survey on tourism expenditure of non-residents staying in collective accommodation establishments	Table 1	No	Data is representative only at national level.
TOUR_PA – Survey on tourism expenditure of non-residents staying in private accommodation establishments	Table 1	No	There is no regional breakdown of data. However, the total expenditure by tourists (at the place of accommodation) does not reflect accurately the region where this expenditure was actually made.
ACTR – Survey on Tourism Demand of residents in Romania	Tables 2, 3, 10	No	Data is representative only at national level.
Family Budget Survey	Table 2	Partially	Although data on the structure of consumption expenditure is available at regional level, this structure is rather aggregated which is not very useful for TSA purposes. However, there is a major limitation since the destination (region) where expenditure is not required in the survey's questionnaire. All expenditure is allocated implicitly where the household has the residence.
Survey on tourist actions organized by travel agencies	Table 3	No	Since there is a cut-off sample, only national data representativeness can be guaranteed.
CHDEP – Survey on travel of employees	Tables 3, 4	No	The coverage of the survey is limited only to some institutions whose activity is difficult to be regionalized from theoretical point of view (i.e. governmental agencies, ministries).
Input-Output Table (IOT)	Tables 5, 6	No	There is no IOT for regions (only IOT for the national level).

1	2	3	4
ASI – Structural Business Survey	Tables 2, 5, 10	Yes	ASI provides data for regions (NUTS 2 level). However, only four indicators are collected at regional level: turnover, expenditure for the compensation of employees, average number of employees and investment expenditure.
AMIGO – Labour Force Survey	Tables 7, 10	Yes	Even if the sample is representative at regional level, the level of detail for industries in the survey does not allow a clear identification of tourism industries.
Balance of Payments	Table 3	No	Not applicable – by its nature Balance of Payments is only designed for national level.
The statistics of neighbouring countries Regarding day trips to Romania (mirror statistics)	Table 1	No	Some reasonable assumptions can be made depending on the location of the border crossing-points of the regions with borders.
Existent tourism accommodation establishments on 31st July	Table 10	Yes	Non-monetary data derived from the location of accommodation establishments.
Occupancy of accommodation establishments	Table 10	Yes	Non-monetary data derived from the location of accommodation establishments.
REGIS statistical register	Table 10	Yes	Indicators derived from the location of reporting units. However, REGIS does not provide detailed data for subsidiaries/branches of enterprises – only aggregated data for the headquarters of enterprises.

Source: own elaboration based on INS publications.

Appendix B

Data sources used in the regionalization of the TSA aggregates in Romania

TSA aggregate	Details of the aggregate	Data sources
Internal Tourism Expenditure	Internal Tourism Expenditure= Inbound tourism expenditure + Domestic tourism expenditure	Administrative data (border crossings counts) ACNER tourism expenditure survey for non-residents Survey on occupancy of tourism establishments Survey on the existent tourism accommodation establishments at 31st of July Structural Business Statistics Survey on museums Survey on passenger transportation at airports Survey on road passenger transportation Family Budget Survey Reports and data from Transportation authorities (e.g. Romanian Naval Authority, Railway Reform Authority)
Internal Tourism Consumption	Internal Tourism Consumption = Internal Tourism Expenditure + Other components of tourism consumption; the latter is restricted to the value of imputed rent of accommodation services in own vacation homes and social transfers in kind (subsidies for trips for medical treatment)	Estimates for Other components of tourism consumption based on: • Population and Dwellings Census • Family Budget Survey • Social Insurance Statistics • Survey on the existent tourism accommodation establishments on 31st of July
Gross Value Added of Tourism Industries (GVATI)	Only 10 tourism industries for international comparability considered	Structural Business Survey Labour Cost Survey
Direct Tourism Gross Value Added (TDGVA)		Similar with domestic tourism consumption
Direct Tourism Gross Domestic Product (TDGDP)	TDGDP = TDGVA + *Taxes less subsidies on products*	Similar with TDGVA
Employment in the tourism industries	Only 10 tourism industries for international comparability considered	Labour Cost Survey Structural Business Survey

Source: own elaboration.

ARGUMENTA OECONOMICA
No 1 (50) 2023
ISSN 1233-5835; e-ISSN 2720-5088

*Jakub Karnowski**, *Andrzej Rzońca****

Should Poland join the euro area? The challenge of the boom-bust cycle

The article makes references to the existing cost-benefit analyses on adopting the euro in Poland and other new member states of the EU. In general, they considered a decrease in the cost of capital to be among the crucial benefits of adopting the common currency. As a result, the costs of a credit boom-bust cycle, to which adoption of the euro may lead, were at best underestimated. The authors discussed the possible solutions to mitigate the risk of boom-bust cycles and concluded that although they are conceptually simple, their effective implementation may be difficult, mostly due to political pressures. To overcome the feasibility problem, the power over undertaking macro prudential measures should be delegated to the central bank. Its past performance in stabilising the economy (at least until recently) suggests that it could make a good use of such a power. Furthermore, adopting the euro would resolve the problem of the 'difficult cohabitation' of monetary and macro prudential policy that such a delegation would otherwise create.

Keywords: euro adoption, boom-bust cycle, macro prudential measures

JEL Classification: E32, E58, E61, E62, G0
DOI: 10.15611/aoe.2023.1.11

Quote as: Karnowski, J., Rzońca, A. (2023). Should Poland join the euro area? The challenge of the boom-bust cycle. *Argumenta Oeconomica*, 1(50), 227-262.

1. Introduction

Since the beginning of its political transition, Poland has become highly integrated with the euro area in terms of trade, FDI inflows and other financial linkages. The integration notably gained in pace after accession to the European Union in 2004. It has proven to be beneficial to income convergence and technological advance of the Polish economy, however the cornerstone of integration, i.e. adoption of the euro, still lies ahead of Poland.

Through joining the EU, Poland became "a Member State with a derogation". Such a status means that there is no room for the question of *whether Poland will join the Euro area*, and only the question of *at what point it will happen* remains open.

* Warsaw School of Economics, Poland. ORCID: 0000-0003-4447-4753.
** Warsaw School of Economics, Poland. ORCID: 0000-0002-8631-2595.

As of now, Poland is not eligible to join the euro area, which would require meeting the Convergence Criteria (in other words: the *Maastricht Criteria*). As the latest Convergence Report[1] of June 2020 indicated, Poland had hardly fulfilled the price stability criterion even then. While HICP stood within the 1.5 p.p. of acceptable deviation from the reference value[2], the European Commission pointed out that "there are concerns regarding the sustainability of inflation convergence in Poland over the longer term" (European Commission 2020a, p. 95). These concerns proved to be justified (see Figures 1-2). Poland has also ceased to comply with the second monetary convergence criterion, that is the long-term interest rates criterion (see Figures 3 and 4). Interest rates are being pushed upwards not only by the currency risk (the central bank's ability to keep the currency stable), but the country risk as well. Due to pandemic, the general government deficit increased well above the Maastricht limit, while sovereign debt has dangerously come close to that limit (see Figure 5). In 2021, both the deficit and the sovereign debt declined (although the exact figures are not available yet). However, that improvement of fiscal stance was largely driven by an inflation surprise, which in the years to come will result in strengthened pressure on the government to increase expenditure, which will further intensify due to the Russian invasion of Ukraine, as Poland will have to cover the cost of aid to refugees and to increase military spending. Lastly, Polish zloty does not participate in ERM II.

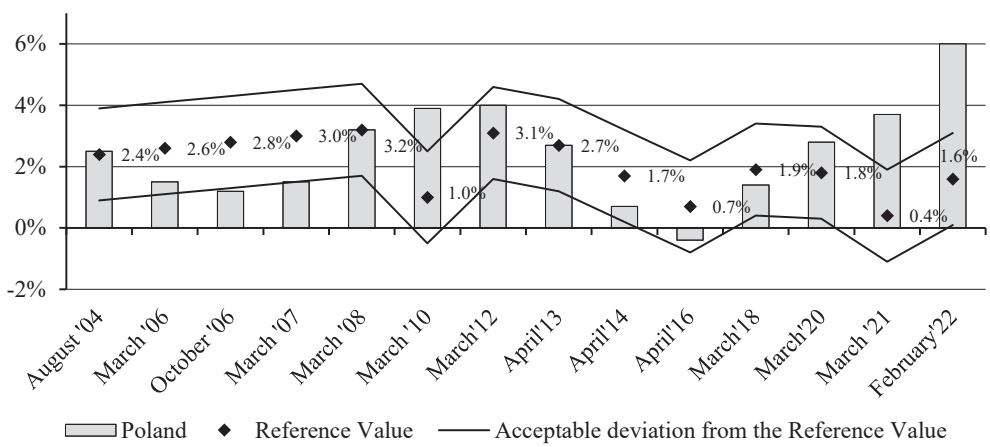


Fig. 1. Inflation in Poland vs. Maastricht criterion (2004-2021)

Source: authors' elaboration based on the European Commission (2020) and Eurostat, Series: HICP monthly data (12-month average rate of change).

[1] Convergence report is published every two years.

[2] Calculated for March 2020, as the average HICP rate of the three best performers in the EU: Portugal, Cyprus, Italy.

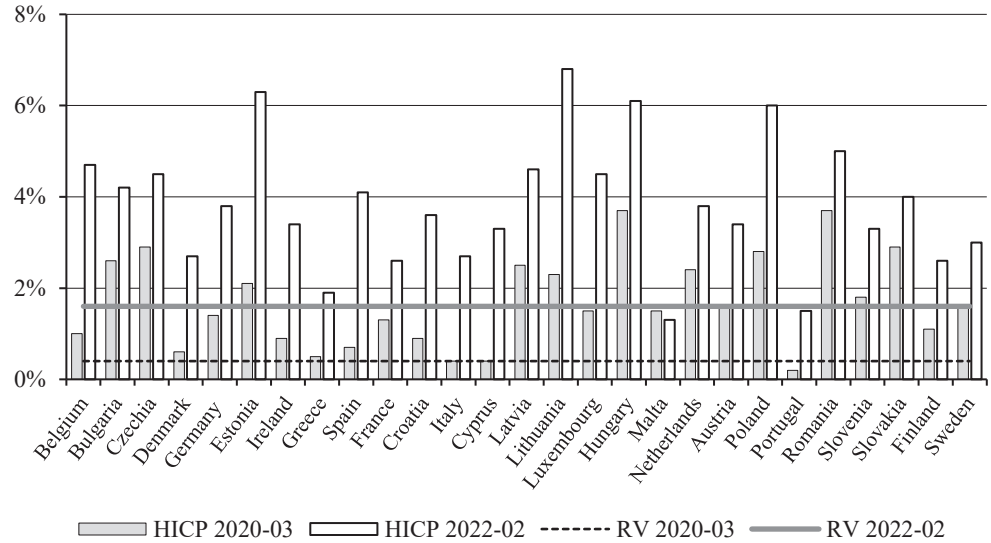


Fig. 2. Inflation vs. Maastricht criterion across EU countries in 2020 and 2021

Source: authors' elaboration based on the Eurostat, Series: HICP monthly data (12-month average rate of change).

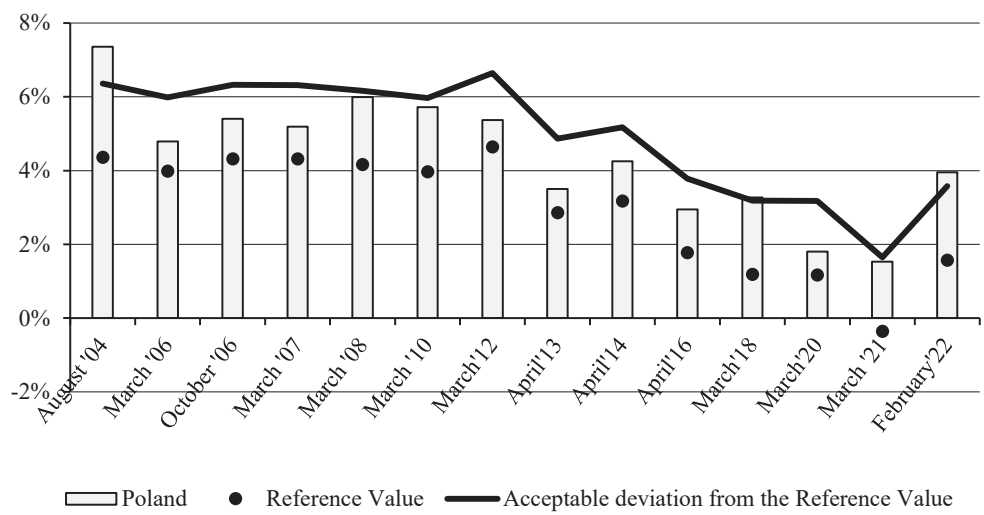


Fig. 3 Long-term interest rates in Poland vs. Maastricht criterion (2004-2021)

Source: authors' elaboration based on the European Central Bank Statistical Data Warehouse. Series: Harmonised long-term interest rates for convergence assessment purposes, (percentages per annum; period averages; secondary market yields of government bonds with maturities of close to ten years).

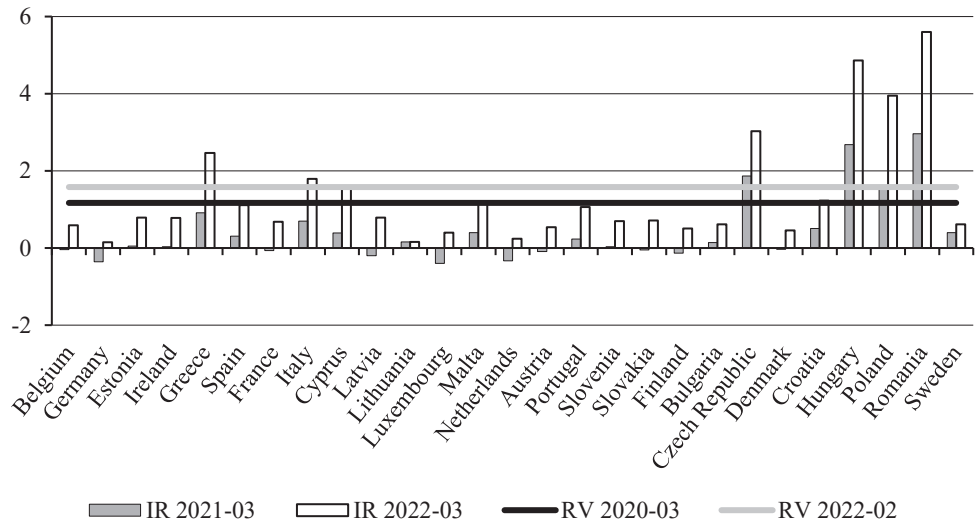


Fig. 4. Long-term interest rates vs. Maastricht criterion across EU countries (2020 and 2021)

Source: authors' elaboration based on the European Central Bank Statistical Data Warehouse. Series: Harmonised long-term interest rates for convergence assessment purposes, (percentages per annum; period averages; secondary market yields of government bonds with maturities of close to ten years).

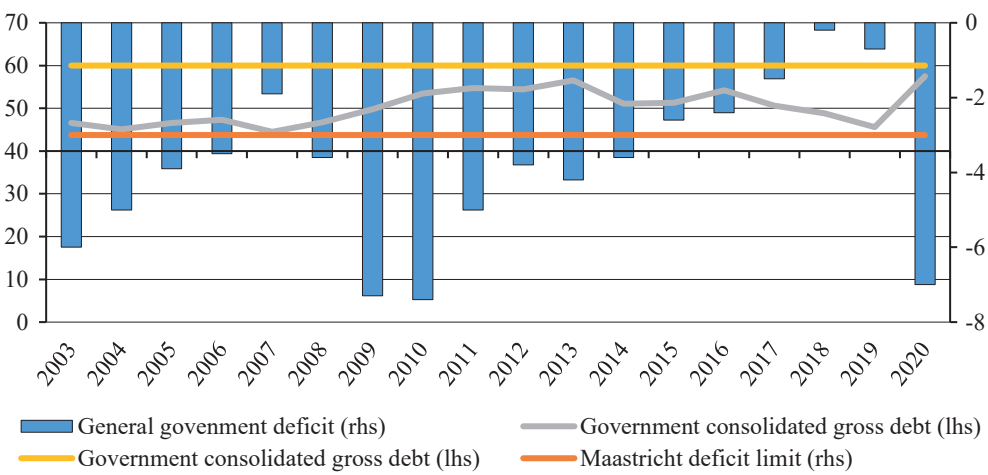


Fig. 5. Fiscal stance of Poland vs. Maastricht Criteria

Source: AMECO, Chapter 16 (General Government).

Even if Poland met the Convergence Criteria, political polarisation would render the adoption of the euro unlikely in the foreseeable future as this would require amending the Polish constitutional law, which is unfeasible with the current

distribution of votes in Parliament.[3] This would involve the votes of most MPs representing the ruling majority – the United Right (ZP, Zjednoczona Prawica), in power since 2016, which definitely opposes joining the euro zone. Its leader, Jarosław Kaczyński, stated that Poland will not join the euro area "until the level of income will catch up to the standards of the 'Old' EU" (Business Insider, 2019).

The results of political calculus reflect public sentiment. The Global Financial Crisis (GFC) and its consequences considerably weakened public support for the euro. At the time of the most intense tensions in the peripheral countries of the euro area dealing with the debt crisis, the share of Poles opting for the common currency dropped to a quarter, the proportion reversed compared to the time before EU accession. In 2002 almost two-thirds of Poles supported joining the euro area, even outnumbering supporters of EU accession (CBOS, 2013). That being said, the public support for the common currency has rebounded recently. The percentage of opponents to the introduction of the Euro (49%) still outweighs the share of its proponents (48%), but only marginally (European Commission, 2020b). The Russian invasion of Ukraine is likely to further increase the share of its proponents.

Most of the existing cost-benefit analyses suggest that adopting the euro should be beneficial for Poland (Borowski (ed.) 2004 and NBP 2009). However, these analyses do not account for the costs of credit boom-bust cycle which may come as a result of euro adoption – as was, for instance, the case of Slovenia. The NBP (2014) has addressed this gap and shows that countries with weak macroeconomic fundamentals may suffer from the surge of macroeconomic imbalances. Therefore, joining the euro area shall be perceived only as a potential opportunity to accelerate economic growth, which is by no means certain.

The article focuses on the issue of adopting the euro from the credit boom-bust cycle perspective. It approaches the issue in a general theoretical way. Nevertheless, if political conditions allowed Poland to join the euro area, then the paper could become a source of guidance for institutional changes necessary for the euro to have a positive effect on economic growth.

The article consists of five further sections, numbered from 2 to 6:

Section 2 briefly describes the extent to which Poland has benefited so far from integration with the European Union. The focus is on these mechanisms (trade, FDI, financial integration), which could be strengthened if Poland joined the euro area.

Section 3 refers to the existing cost-benefit analyses on adopting the euro in Poland and the other new member states of the EU, and concentrates on whether they take into account the risk of a credit boom-bust cycle which the adoption of the euro may result in.

Section 4 presents the sources of a credit boom-bust cycle and its costs. It compares the pre-crisis and post-crisis experience of Poland and other new member states of the

[3] Amending constitutional law requires collecting the supermajority – i.e. support of two thirds of voting MP's – at the quorum of 50%.

EU, with the similar experience of the euro area countries, placing emphasis on the factors behind the resilience of the Polish economy during the GFC. Based on this comparison the section outlines conditions under which adopting the euro may increase both the risk and costs of a credit boom-bust cycle.

Section 5 discusses the possible methods for mitigating the risk of boom-bust cycles after euro adoption and assesses their feasibility. The authors refer to Poland's experience with fiscal rules and macro prudential policies, and also discuss the ability of the Polish economy to smoothly reallocate capital and labour. Then the section deals with certain other challenges of adopting the euro, related to the legacy of the sovereign debt crisis and the completion of institutional reforms in the euro area. The section concludes with a discussion of some political consequences of Poland staying outside the euro area.

Section 6 summarises the main findings of the article. These are as follows:

1. Poland has become deeply integrated with the euro area, and certain links between Poland and the euro area are even stronger than among euro area members.
2. Potential benefits for Poland from euro adoption are related to a deepening of the integration.
3. The main risk is associated with a credit boom-bust cycle or, more generally, with a halt in catching-up (which may result not only from a boom-bust cycle, but also from undermined cost competitiveness or slow restructuring).
4. Theoretical solutions to mitigate the above risk are broadly available. Their feasibility, however, is questionable.
5. The feasibility problem might be overcome through delegating the power of undertaking macro prudential measures to the national central banks.
6. Euro adoption would resolve the problem of the 'difficult cohabitation' of monetary and macro prudential policies.

The article draws from various strands of the literature, starting from cost-benefit analyses of euro adoption, through the literature on determinants and costs of financial crises, and ending with research on finance and growth, and macro-prudential measures. The authors see the paper's main contribution in linking these strands of the literature in order to indicate the main risk related to adoption of the euro and a feasible solution to mitigate that risk.

2. Integration of Poland with the euro area

The past 32 years count among the most prosperous in Polish history, characterised by stable economic growth and convergence with Western Europe. In fact, the gap in income per capita between Poland and Western Europe is thought to be at its lowest levels since the 15th century (Piątkowski 2011). One of the crucial factors behind this success is ongoing integration with the European Union, and its core, i.e. the euro area. Joining the EU in 2004 facilitated the consolidation of broad institutional changes initiated in 1989. However, taking into account the magnitude of the

transformation of the Polish economy since the end of socialism, separating the benefits of European integration from ongoing transition is not an easy task. Furthermore, integration is a process, not a well-defined single event. On the one hand, certain benefits of EU membership were obtained even before accession, following the gradual harmonisation of the legal system and the inflow of foreign direct investment (FDI), based on the assumption that Poland would ultimately become an EU member. On the other hand, building economic links takes time, and thus certain benefits of integration have not fully materialised yet.

That being said, these links between Poland and other EU member states, in particular members of the euro area, have already become very strong. Trade with neighbouring Germany alone accounts for approximately 29% of Polish foreign trade[45], which in 2019 concentrated more on the euro area countries than the trade in the average euro area member, e.g. Spain (see Figure 6). Due to its size Spain can be regarded as a suitable benchmark for Poland.

Trade links have been enhanced by the massive inflow of FDI. The bulk of FDI in Poland originates from the euro-area countries. As of 2019, investment in Poland by euro area members accounted for 33% of Polish GDP. The analogous figure for Spain stood at 36% (see Figure 7).

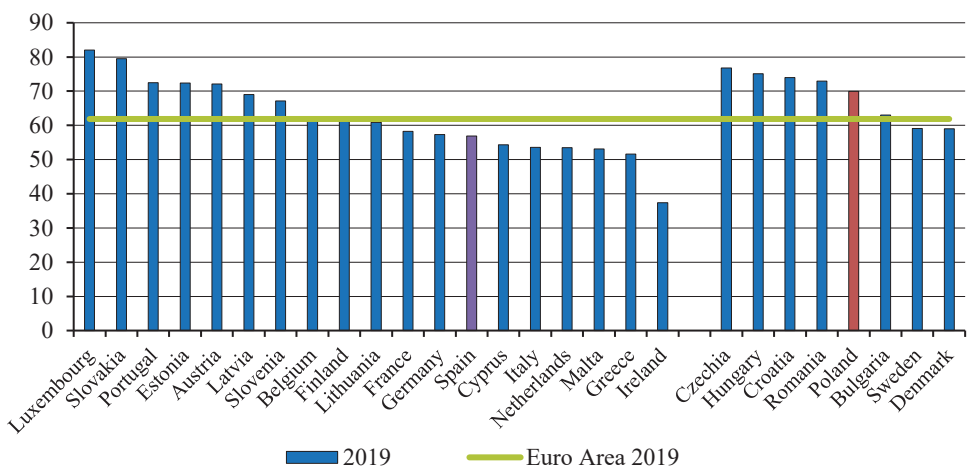


Fig. 6. Intra-EU trade in goods as a share of total trade in goods (2019)

Source: Eurostat. Series: Intra and Extra-EU trade by Member State and by product group.

[4] Eurostat data, as of 2020.

[5] Although it is not surprising that Germany is the main trading partner of Poland, it is also worth noting that Poland is a significant trading partner for Germany. As of 2019 (the last year not distorted by Covid-19) Poland was the 8th largest partner of Germany in terms of exports – ahead of countries such Belgium and Switzerland; in terms of imports Poland was Germany's 5th largest partner, ahead of countries like Italy, Czechia and Switzerland (data from Destatis).

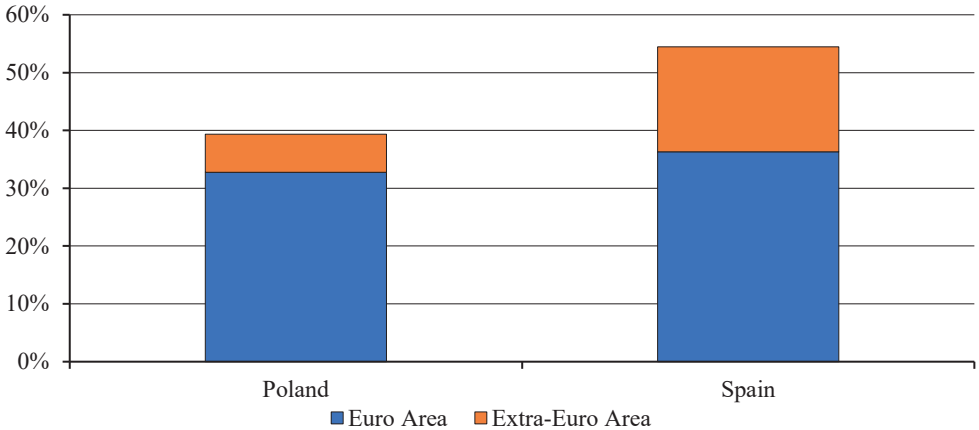


Fig. 7. FDI stock in reporting economy as % of GDP in 2019

Source: Eurostat, Series: EU direct investment positions, breakdown by country and economic activity.

Both the growing foreign trade and inflow of FDI, have allowed Poland to avoid the deindustrialisation that most euro area economies have suffered from. Polish companies often work as subcontractors for European companies, in particular German companies. The share of manufacturing in value added in Poland is quite similar to that of Germany. The two aforementioned factors have also helped Poland to move its export up the value chain. In 1995, labour- and resource-intensive products comprised 61% of Polish manufactured goods exports (see Figure 8).

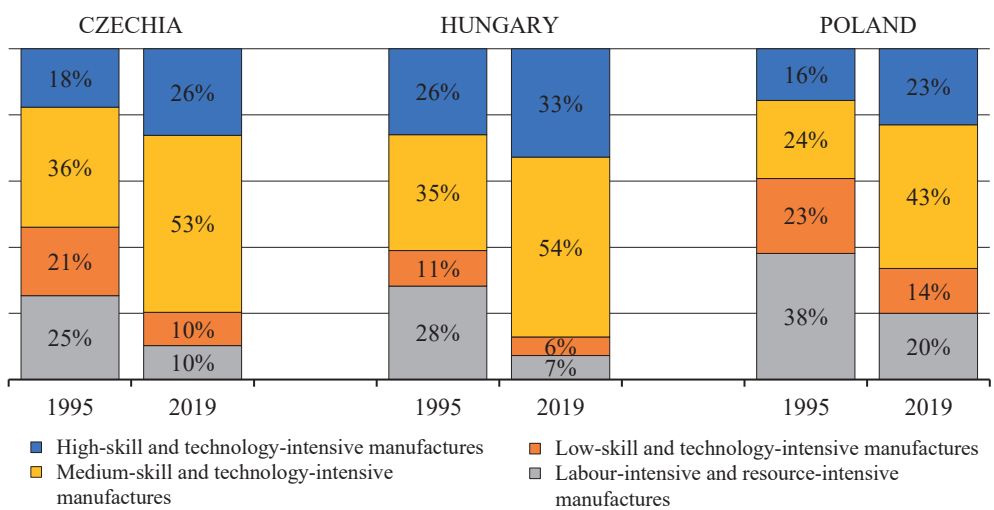


Fig. 8. Structure of Merchandise Exports in 1995 and 2019

Source: authors' elaboration, based on UNCTAD. Series: Merchandise trade matrix – exports of individual economies in thousands US dollars, annual Table summary).

By 2019, the share declined to 34%, as the result of rapidly growing exports of products that require medium to high technology to be manufactured.

The links are also strong in the financial sector. The Polish banking sector is largely foreign-owned, with approximately one-third owned by investors from the uro -area countries (see Figure 9). All this suggests that in the case of Poland, the risk of insufficient business cycle synchronisation is likely to be lower than for many Euro area members (cf. Fidrmuc and Korhonen, 2006; Rinaldi-Larribe, 2013).

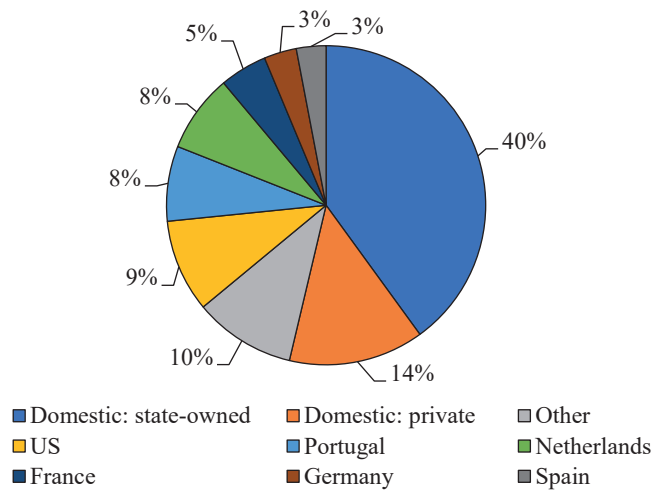


Fig. 9. Ownership structure of the Polish banking sector in 2019 (percent of assets)

Source: The Polish Financial Supervision Authority (KNF), 2019, p. 15.

The rapid development of tradable sectors that have gone through significant technological change has contributed to the fast catching-up of Poland with the euro area. The gap in terms of GDP per capita between Poland and euro area has narrowed by almost 22% since EU accession. Interestingly, it has narrowed even in the case of the worst performers among the New Member States (NMS) (see Figure 10). The available studies confirm that Poland (and other NMS) significantly benefited from joining the EU. For example Campo et al (2014), using the Synthetic Counterfactual Method, estimated that EU accession increased Polish GDP by approximately 6% (the average gain for all the NMS amounted to 14%[6]). Due to the way the method was implemented, the results should be deemed conservative or even the lower bound of the true benefits of integration (that is why the authors refer to that study instead of numerous others). By comparison, studies ordered by the Polish Ministry

[6] 14% is the average for 2004 and 2007 enlargements. The average for all the enlargements since the 1970s was 12%. According to the authors, the only country that would be better off without joining the EU is Greece, which requires further studies.

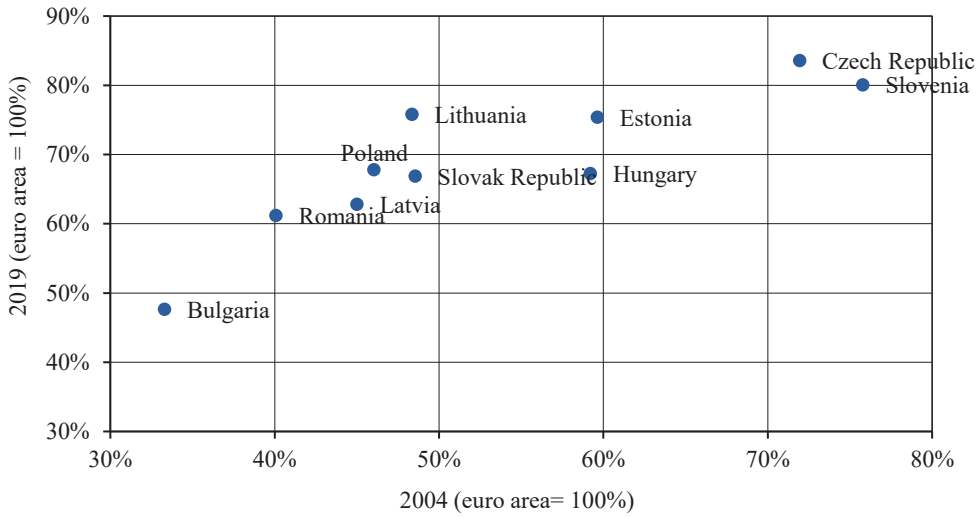


Fig. 10. GDP per capita of NMS vs. the euro area in 2004 and 2019

Source: IMF's World Economic Outlook 2021.

of Regional Development estimated that the cohesion policy only boosted the GDP of Poland by between 4.2% and 7.2% (as of 2011)[7].

Poland has benefited from the EU membership mainly through institutional, trade and financial integration, and euro adoption should further deepen this integration. In particular, trade ties will be strengthened by the elimination of exchange rate risk, increased price (and cost) transparency and consequently easier price comparisons, and – albeit to a much lesser extent – a decrease in transaction costs.

The importance of falling transaction costs as a mechanism to deepen international trade is often overstated (see eg. Borowski [ed.], 2004). First, it is merely a revenue shift between the non-financial sector and the financial sector. Second, the development of online FX platforms has led to a significant reduction in exchange costs (Sławiński, 2014).

For specialised exporters or importers even the exchange rate risk or the lack of full price transparency are not barriers to develop international trade. They are able to hedge against exchange rate risk at low cost, and can even benefit from the lack of full price (and cost) transparency because it gives them an advantage over other economic agents in making relevant comparisons. Yet, the elimination of exchange rate risk and full price (and cost) transparency should make it easier for new players that previously operated only in the local market to engage in international trade.

[7] The impact of the policy was estimated with the use of the CGE model developed by IBnGR (2012), the DSGE model by IBS and HERMIN model updated also by IBS (2012a, 2012b.).

International trade is an important source of technology transfer, which in small open economies explains, according to some estimates, up to 90% of technical progress. This transfer is facilitated not so much by exports (although they do force the learning of foreign standards), but primarily by imports. Imports give domestic companies access to more advanced technology, and components of lower price or higher quality (Keller, 2004).

In addition, the deepening of international trade as a result of membership in the euro area will be another impulse – after trade liberalisation at the beginning of the transition and accession to the European Union – strengthening the competitive pressure from abroad. The importance of this pressure for productivity growth in Poland is evidenced by much faster productivity growth in manufacturing than in other sectors less exposed to international trade than manufacturing. Manufacturing is also distinguished by an allocation efficiency ratio close to the highest in the euro area, while in other sectors it is clearly lower than in euro area countries (NBP, 2014).

The benefits for the new EU member states of the deepening of international trade as a result of euro adoption was the focus of the analysis by Schadler et al. (2005), which builds in particular on Frankel and Rose (2002) and Rose (2002). It shows that adoption of the euro could raise GDP per capita in Poland over 20 years in the range from 1% to 8%.

Nevertheless, a caveat is required that according to some studies, this effect should be clearly weaker (Podpiera et al., 2015). Havranek (2010) estimated it at 0% and Baldwin (2006) at 2% to 3%. These studies argue that it is not so much membership in the euro area, as the EU membership, that leads to a deepening of international trade. However, even if this were indeed the case, participation in the euro area could still be desirable as a way to consolidate the membership in the European Union.

3. Available estimates of the economic benefits and costs of euro adoption

Analyses of the potential economic benefits and costs of adopting the euro often focus on assessing the extent to which a country forms an optimal currency area with the euro area, whose main features were defined, in particular, by Mundell (1961), McKinnon (1963), and Kenen (1969).

In the literature on optimal currency areas, the benefits of a common currency are mainly seen in the reduction of the exchange rate (as well as terms of trade) volatility, which is expected to promote the deepening of international trade, and also in the integration of financial markets. It is further indicated that for countries without a long history of price stability, a common currency can be a way to import monetary policy credibility (see McKinnon, 2004; Tavlas, 1993).

The costs of the common currency are instead linked to the loss of autonomy in monetary policy and the abandonment of the floating exchange rate, which can no longer absorb shocks, but are emphasized as limited by the synchronisation of business cycles, facilitated by large international intra-industry trade (Frankel and Rose, 1998). However, since cycles within a currency area are never perfectly synchronised, it is argued that the construction of an optimal currency area requires prices and wages to be as flexible as possible, so that shocks are absorbed as much as possible by price adjustments not forced by quantity changes (in output and employment/unemployment). An additional mechanism for absorbing asymmetric shocks is freedom of labour flows between the countries forming a common currency area. It was also pointed out that the costs of such shocks can be dissipated by the geographical diversification of the income sources of the economic agents, enabled by the integration of financial markets, which facilitates investment in the assets from different countries (McKinnon, 2004).

Even a superficial analysis of the conclusions from the literature on the optimal currency areas suffices to conclude that joining the euro area should be beneficial for Poland, especially if it increases flexibility of the product and labour market. Poland is strongly economically integrated with the euro area, and companies that are part of the European value chain play an important role in Polish foreign trade (see the previous section), and Poles have full freedom to work in the euro area countries. Thus, the assessment of the balance of benefits and costs of euro adoption is much more complex. The literature on optimal currency areas and the reports based on it, which analyse potential benefits and costs of adopting the euro, largely ignore the risk of a boom-bust cycle.

The two most comprehensive analyses of benefits and costs of euro adoption for Poland were prepared by the National Bank of Poland (NBP). The first one, edited by Borowski, was published in 2004, followed by the second one in 2009. On the benefits side, both highlighted lower transaction costs and reduced risk premiums, which were to lead to higher investment and trade deepening. Moreover, better financial integration and a more competitive environment were indicated as possible benefits. On the cost side, both reports focused on the loss of sovereign monetary and exchange rate policies in the context of asymmetric shocks absorption, and the degree of synchronisation of business cycles. The first report was prepared before EU accession, and concluded that ongoing integration with the euro area would foster synchronisation of business cycles. The authors emphasised the insufficient flexibility of the Polish labour market, but concluded that opening the labour markets of the EU countries for employees from Poland might attenuate domestic rigidities. Five years later, when the second report was prepared, the Polish economy was more interlinked with the euro area (Germany in particular), which limited the problem of business cycles' synchronisation. Furthermore, the Polish labour market became more flexible. Unfortunately, between 2004 and 2009 there was no significant

improvement in the fiscal policy. Both reports pointed to the problem of persistent structural deficit limiting fiscal space.

Borowski et al. (2004) estimated that due to the lower cost of capital, joining the euro area would increase the GDP growth rate by 0.2 p.p. annually, boosting Poland's GDP level by approximately 6% by 2030. Similar gains were expected due to increased FDI inflows. According to the report, joining the euro area in 2007 would boost GDP by approximately 12% by 2030, while the report from 2009 also concluded that joining the euro area would have a net positive impact on the Polish economy. As a result of euro adoption, Polish GDP was projected to grow by additional 7.5%, with the majority of results realised during the first ten years after accession. According to the authors, the ongoing GFC (the report was published in 2009) does not affect the long term benefits of euro area membership, but makes the short-term effects difficult to estimate and raises questions about the optimal euro adoption date.

Unlike the first NBP report, the second report recognised the risk of a boom-bust cycle due to adoption of the euro but concluded that the risk is limited. The report examined the experiences of Greece, Ireland and Portugal where such a cycle occurred, and assessed that these countries avoided a deterioration of the health of the banking sector. With the benefit of hindsight, it is clear that such an assessment was wrong, as all these countries later experienced significant problems in their banking sectors. In the case of Poland, the NBP simulations concluded that adopting the euro should not lead to a significant increase in private credit to GDP ratio, in spite of its low starting level.

Another NBP report was released in 2014. This publication, unlike the two previous, focused on the matter of optimal conditions for joining the euro area instead of carrying out the cost-benefit analysis. Special emphasis was put on the issue of the credit boom bust cycle.

According to the report, ECB interest rates could turn out to be too low for Poland. As long as there will be real convergence in Poland, the appreciation of the real exchange rate will continue. Without own currency, this appreciation would take the form of higher inflation and accelerated wage growth. Therefore, the ECB interest rates, nominally the same in each country, would be lower in real terms in Poland than in the euro area as a whole. On the one hand, too low interest rates may cause excessive credit growth and speculative bubble, while on the other, it may weaken the cost competitiveness of domestic enterprises, mainly due to wage growth spreading from the construction sector (where it would be fuelled by the growth of real estate prices) to the tradables sector.

The authors considered that the accumulation of the above-mentioned imbalances would be the main peril for the stability of the Polish economy after joining the euro area. Moreover, the NBP experts admitted that it would be much more severe than possible asymmetric shocks, i.e. the focus of the 2004 and 2009 reports.

It was stressed that dedicated preventive mechanisms would have to be created before accession to the Economic and Monetary Union (EMU). Otherwise, if the imbalances were allowed to cumulate, their subsequent reduction would require much time and cause significant costs. A macro-prudential policy was recommended as the core preventive mechanism, however it was stipulated that macro prudential measures cannot perfectly substitute autonomous monetary policy. In the case of those measures, both the length of the decision process and their leverage on the economy are much longer than in relation to the monetary policy.

The main conclusion from the 2014 report was that faster growth of Poland's economy after euro adoption cannot be taken for granted. Reaching those benefits will depend on how Poland will be prepared to operate within constraints set up by the ECB monetary policy. Accordingly, before euro adoption, a series of reforms should be implemented.

Reports on the effects of euro adoption conducted in other NMS also focused on the degree of business cycles synchronisation, responses to asymmetric shocks and fulfilment of the criteria of optimal currency areas[8]. Only two of them explicitly mentioned the risk of an asset prices boom due to a decrease in interest rates. They were published by the National Bank of Hungary (Cssajbok and Csermely, 2002) and the National Bank of Slovakia (2006). The Hungarian report concluded that proper microeconomic prudential regulations can limit such risk. The Slovakian report had broader recommendations including not only forward looking financial supervision but also a proper insolvency regime, a slightly restrictive fiscal and monetary policy (before adopting the euro), a prudent wage policy and a flexible labour market.

It should be remembered that the majority of cost-benefit analyses of joining the euro area were conducted before the GFC. Since then, the euro area has undergone significant changes, and the process of reforms is not yet finished. Hence, further changes to the cost-benefit balance may be expected – the authors return to this issue in Section 5.

---

[8] These include: Adopting the Euro in Hungary: Expected Costs, Benefits and Timing, by researchers from the National Bank of Hungary from 2002 (Cssajbok A., Csermely, 2002), Programme for ERM II entry and adoption of the euro. Joint programme of the Slovenian Government and the Bank of Slovenia, by the Bank of Slovenia and the Slovenian Government from 2003, Impact of Euro Adoption on the Economy of Latvia, published by the Bank of Latvia from 2004 (Bitans M., Kauzens 2004), The Effects of Euro Adoption on the Slovak Economy published by the National Bank of Slovakia from 2006 (Suster [ed.] 2006), Impact of Euro Adoption of the National Economy: an Overview of the Quantitative Assessment by the Central Bank of the Republic of Lithuania from 2013. Note that some of these documents were published as working papers, reflecting only the views of the authors, while others were official documents.

4. The credit boom-bust cycle and its costs

After 2007, 13 EU members fulfilled the quantitative criteria proposed by Laeven and Valencia (2008) to be classified as cases of banking crisis. This number seems to be underestimated, for an episode to qualify as a banking crisis according to Laeven and Valencia either (i) significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations) occur, or (ii) significant banking policy intervention measures in response to major losses in the banking system are taken. In the case of countries such as Estonia, Lithuania and Bulgaria, the crisis defined in such a way, did not occur only due to support granted to local banks by their parent companies, which in turn was possible only due to the significant public support in their home countries. Thus, the crisis that started in the EU in 2008 can be classified as a banking crisis even to a larger extent than the strict use of definition by Laeven and Valencia would suggest.

It is generally agreed that recessions associated with banking crises are particularly deep and last longer than others (cf. Cerra et al., 2009, and Claessens et al., 2009). The costs of the crisis that hit the European Union in 2008 have been significant. In 2013, i.e. five years after the GFC outbreak (which is a time span usually long enough for an economy after shocks to return towards trend), GDP per capita in the member states was 18% below its pre-crisis trend. The gap in the most affected countries, such as Ireland, Greece and Latvia exceeded the costs of the previous banking crisis in developed countries, such as Japan and Finland, in the 1990s. Obviously, it can be argued that in some cases the pre-crisis growth rates were impossible to sustain, as the premium from catching up in converging countries would naturally decrease, so the cost of the crisis is overstated. However, taking as a benchmark the IMF forecasts from April 2008, yields results were of a similar magnitude – in 2013 GDP per capita of EU member states was on average 17% below the forecasted level (see Figure 11) – despite changes to the relative positions of the selected countries.

The size of the gap can be explained by three main factors: initial conditions, policies after the crisis and structural features of the economies (Laszek 2014). The larger the size of imbalances accumulated in economies as of 2007, the worse their performance after the outbreak of the crisis. The imbalances were reflected in low saving rates, high investment rates, structural deficits in public finances, high growth rate of credit before the crisis and worsening net investment position. Taken together, these factors account for up to 30% of variation in post-crisis performance. Apart from the initial imbalances, policy reaction also mattered. The notably larger deterioration in structural fiscal deficit between 2007 and 2009 resulted in worse performance in 2013. The initial imbalances can explain to some extent the decrease in general government revenue, even after the adjustment for cyclical factors. The increase in spending, however, was much more discretionary than the decrease in revenue. Euro area members (with easier access to financing than most EU countries from outside of the euro area) and EU countries with already unsustainable public

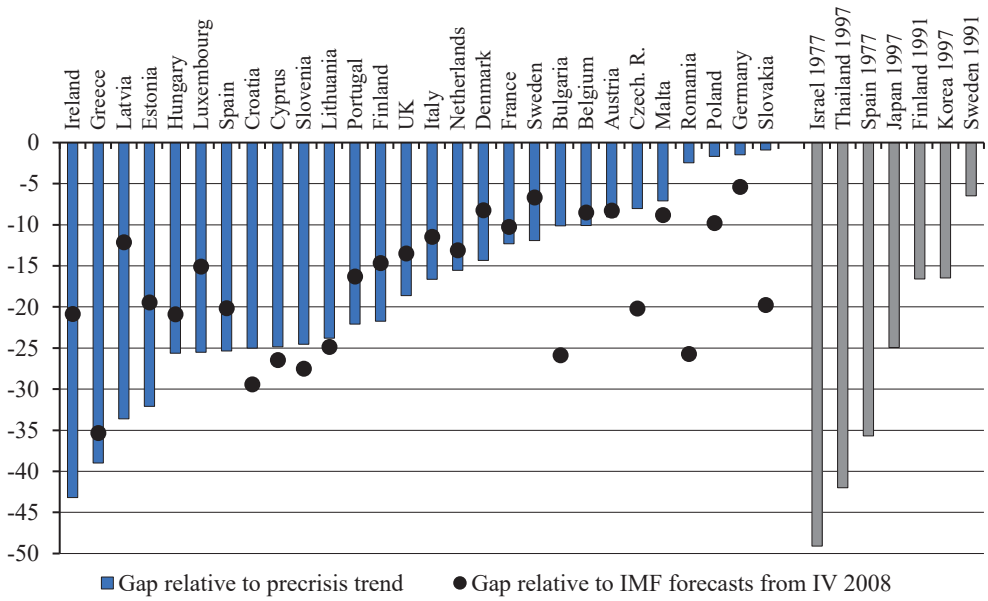


Fig. 11. Gap between actual GDP per capita in 2013 and pre-crisis trend/pre-crisis forecasts

Note: Trend fitted to 12 pre-crisis years 1995-2006, smoothed with HP filter (λ=6.25)

Source: IMF World Economic Outlook IV 2008, IMF World Economic Outlook IX 2013, AMECO.

finances[9], were more eager to increase spending in the midst of the crisis. As far as the structural features of economies are concerned, countries with more flexible product market regulations fared better.

The importance of the initial imbalances in explaining the post-crisis performance of EU countries highlights the costs of the credit boom-bust cycle. The most extreme case were the Baltic states which, after growing at a rate approaching 10% in 2006 and 2007, recorded a decrease in GDP of approximately 15% in 2009 (-17.7% for Latvia). The story behind the boom and bust is documented by Bakker and Gulde (2010), who showed how the countries, lagging behind at the turn of the century, accelerated the pace of reforms, which resulted in higher growth rates. Robust economic growth and optimism, associated with EU accession, led to increased capital inflows further encouraged by the fixed exchange rate regime. Capital inflows, channelled through the banking sector, fuelled a lending boom and an even more rapid GDP growth. The lending was concentrated in the non-tradables sector. The growing wage pressure resulting from the boom eroded the competitiveness of countries in question. To some extent, a similar story regarding Ireland was documented by Whealan (2010). Fixing chronically ill public finances in the second half of the

[9] These are countries for which the Ageing Report (European Commission 2009) was forecasting the largest increase in old age related spending.

1990s unleashed growth potential and led to fast convergence, however in the late 1990s, the initially healthy growth started to turn into an unsustainable bubble. Fuelled by over-optimism and low interest rates, it paved the way for the subsequent crash. It should be noted though, that the magnitude of the boom-bust cycles in NMS was much larger than in Ireland or other 'old' EU members (see Figure 12).

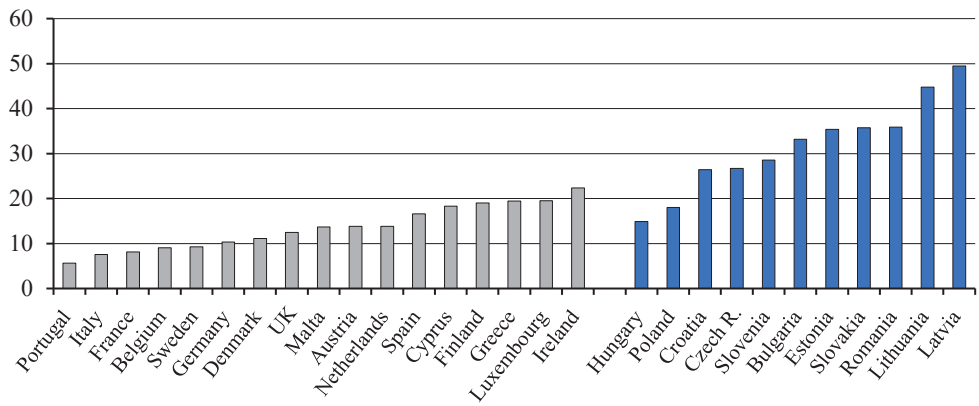


Fig. 12. Boom-bust magnitude. Cumulated GDP growth 2004-2008 minus cumulated GDP growth 2008-2010 (percentage points)

Source: Ameco, Chapter 6 (domestic product).

One of the reasons behind this might be uncertainty about the potential growth of the rapidly changing transition economies, which is visible in the magnitude of the revisions in the estimates of the output gap for 2007 (Figure 13).

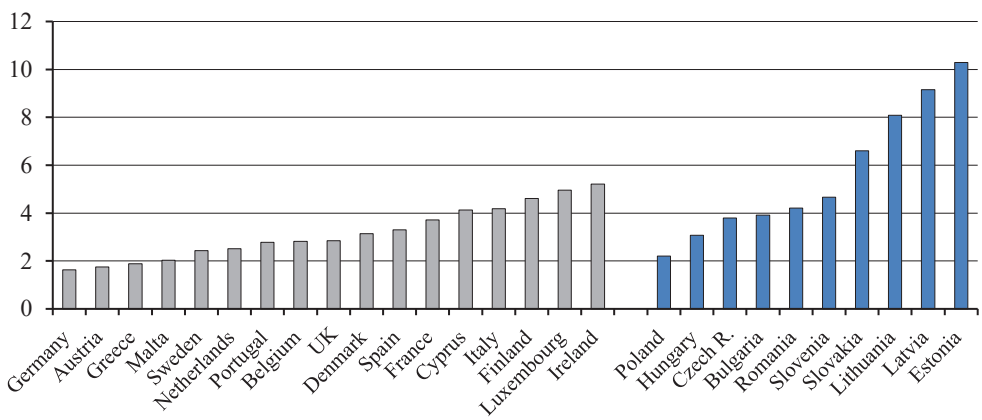


Fig. 13. Ex post revision of the output gap in 2007 (the difference between estimates from EC Spring Forecast 2014 and EC Autumn Forecast in 2007, %)

Source: EC Spring Forecast 2014 and EC Autumn Forecast 2007.

Although the magnitude of the boom-bust cycle in NMS was larger than in the Old Member States, Poland was among the few outliers and did not experience a bust, becoming the fastest growing economy during the crisis (see Figure 14). There are several factors behind the resilience of the Polish economy to the crisis. First, Poland, being the largest economy among the NMS, depends less on foreign demand. Second, 2009 was the year when the effective tax cuts enacted before the crisis were implemented, thus providing fiscal stimulus right after the outbreak of the crisis. Third, the stimulus was amplified by a large increase in public investment financed mostly by an inflow of EU funds. Thus, luck clearly matters for economic performance, however cautious policies are needed to take advantage of good luck. Poland was generally pursuing cautious policies before the crisis. Due to the tightening of monetary policy directly after EU accession, as well as the actions of financial supervision that limited the growth of loans denominated in foreign currencies (FX loans), the credit boom in Poland before the crisis was very short-lived. It started only in late 2006 and was over in 2008 after the Lehman Brothers collapse. These policies also helped the floating exchange rate to cushion a large part of the external shock in 2009. If the stock of FX loans in Poland had been larger, the floating exchange rate could have acted not as a damper, but as an amplifier of the shock. Note that the large volume of FX liabilities accumulated in NMS with fixed exchange rates during the boom phase made, abandoning the peg by those countries prohibitively costly.

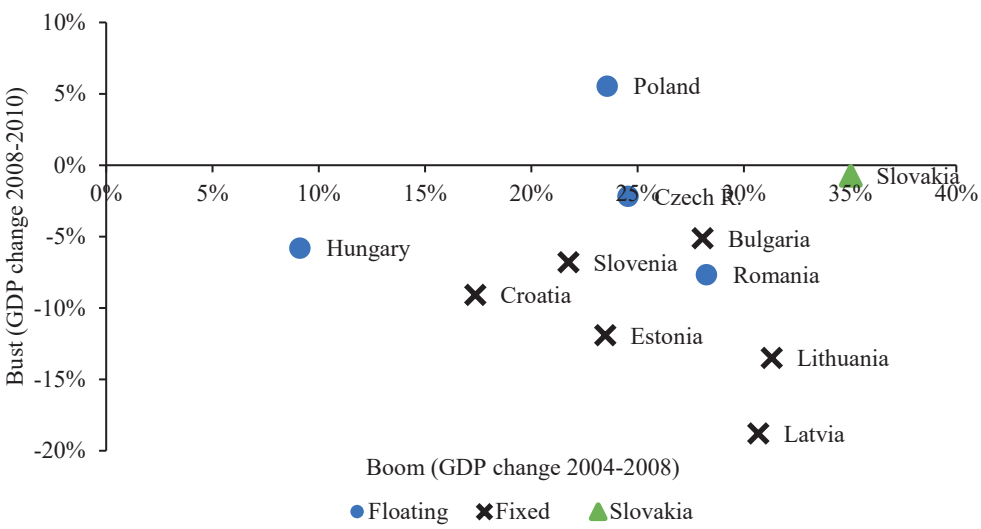


Fig. 14. Boom-bust in New Member States

Note: Although Slovakia adopted the euro in January 2009 and the Slovak koruna was in the ERM II from November 2005, it would be improper to classify Slovakia as a country with a fixed exchange rate, as the koruna exchange rate was adjusted twice: in March 2007 and in May 2008.

Source: Ameco, Chapter 6 (domestic product).

The above comparisons prove that adopting the euro may increase both the risk and costs of the credit boom-bust cycle under the conditions outlined below.

In a dynamically efficient economy interest rates exceed the growth rate of per capita income, otherwise it would be possible to take a loan to increase consumption in the current piod, next pay the interest with another loan, and then wait until the debt to income ratio decreases to zero. In other words, larger consumption today would not require lower consumption in the future (cf. Fischer and Easterly, 1990). Hence, in the standard DSGE model, natural interest rate is given by the following equation:

$$r_t = \sigma\left(y^P_{t+1} - y^P_t\right) + (1-\beta)/\beta, \tag{1}$$

where r_t is natural interest rate, σ is the parameter of relative risk aversion of households (usually assumed to be larger than one), β is the household's subjective discount factor and $y^P_{t+1} - y^P_t$ is the growth rate of potential output.

It follows that for an economy that grows more rapidly than the euro area, adopting the common currency means a fall of interest rates below their natural level, unless measures introduced at the national level increase the price of credit or constrain its volume; these measures are referred to in the next section. With a given pace of economic growth, the decrease in interest rates is deeper, and the less monetary conditions before adopting the euro depend on the monetary conditions in the euro area. Therefore, the decline in interest rates is deepened by the greater previous freedom of the exchange rate to float, and reduced by increasing financial integration with the euro area.

Interest rates below their natural level contribute to the boom-bust cycle through various channels: the search for yield and underpricing of risk (Rajan, 2005), easing of credit standards by banks (e.g. Maddaloni and Peydró, 2010; Jiménez et al., 2009) and the increased likelihood of housing booms and bubbles (Sa, Towbin and Wieladek, 2011).

It is worth noting that even if such a cycle does not result in a financial crisis, it may still decelerate economic growth. During the boom phase, a country's cost competitiveness is undermined, which contributes to a distortion of sectoral composition of output. Tradables sectors lose their share in output for the benefit of non-tradables sectors (in particular the construction sector). Such a structural change impedes technology transfer from abroad and inhibits the diffusion of knowledge on how to produce more efficiently. In brief, when adjustment of the interest rate is constrained, then restoring economic equilibrium may require an adjustment in the natural interest rate, which may take the form of economic growth deceleration. Note however that contrary to the concerns of some economists (see in particular Kawalec and Pytlarczyk, 2016), a member of the euro area has the potential to regain competitiveness. This is allowed by fiscal devaluation, which is not just a theoretical peculiarity, known at least since the time of Keynes. The Baltic States and Ireland in particular, have used it frequently and with success (see Ciżkowicz et al., 2020).

As already mentioned, the costs of a credit boom-bust cycle depend on three main groups of factors: initial imbalances, policies after the crisis, and structural features of economies. Adopting the euro by an economy growing faster than the euro area may increase initial imbalances, by increasing the supply of and demand for credit and lengthening the period of unsustainable credit growth. Note, however, that imbalances' accumulation can be blocked (or amplified) by measures introduced at national level. Furthermore, these measures shape the two remaining factors determining the costs of the credit boom-bust cycle. If, during the bust period, policies force a quick restructuring of overdeveloped sectors with low productivity, the recession is short. The economy rapidly returns to its pre-crisis growth or even its growth accelerates, as was in the case of Sweden after the crisis at the beginning of the nineties (cf Borio et al., 2010). In turn, the more flexible an economy, the less costly the reallocation of production factors, which is necessary after the boom.

A caveat is required here. Adopting the euro softens governmental budgetary constraints. First, this softening facilitates hiding the costs of delaying reforms aimed at enhancing the flexibility of the economy, as demonstrated by the euro area periphery (Fernandez-Villaverde 2013). Second, it expands the possibilities to postpone necessary adjustments during the bust period (see, e.g. Gross and Alcidi, 2013). The comparison of post-crisis performance of the euro area periphery with the performance of the NMS that experienced a particularly strong pre-crisis boom (i.e. Bulgaria, Estonia, Latvia and Lithuania), is instructive on that score. The former group achieved worse economic results than the United States after the outbreak of GFC, whereas all the latter countries outperformed the United States (in terms of growth of per capita income) (Balcerowicz et al., 2013). In the former group, the adjustments, although ultimately impressive, were delayed and gradual. The latter countries managed or were forced to introduce sharp and deep adjustments already at the onset of the crisis.

5. The possible methods for mitigating the risk of boom-bust cycles after adopting the euro

In theory, in order to mitigate the risk of a credit boom-bust cycle after euro adoption, it is sufficient to:

(i) increase government savings adequately,
(ii) drive a wedge between the interest rate set by the ECB and credit costs (notably for mortgage loans[10]), or
(iii) constrain the volume of credit appropriately.

Due to those solutions, a halt in convergence, which the boom-bust cycle may result in, should not be automatically treated as an argument against adopting the

[10] Empirical studies suggest that mortgage loans can lead to housing bubbles (Sa, Towbin and Wieladek 2011), and in themselves are not necessarily supportive for growth (Beck et al., 2012).

common currency. Furthermore, sovereign monetary policy does not necessarily protect against credit boom-bust cycles. The pre-crisis experience of the United States demonstrates that even the most sovereign central bank managed by the most distinguished governor and hiring the most competent staff, may overlook accumulation of dangerous imbalances. Indeed, the transcripts of FOMC meetings in 2006, the last year under Alan Greenspan term of office, show an overwhelming impression of the perception of success and control.[11]

The above solutions indicate areas that need enhancements before euro adoption. Solution (i) calls for continuous fiscal restraint. Solution (ii) may be achieved through either fiscal instruments (taxation) or macro prudential measures. Solution (iii) is a natural domain of macro prudential regulations and policy.

Note that the solutions themselves may have an impact on the pace of convergence.

Solution (i) appears to be potentially most advantageous in these terms, although it also creates certain risks to the convergence process. If it includes cuts to social transfers for the working age population and to wages and salaries, it encourages households to increase labour supply, alleviates wage pressure, and thereby improves the cost competitiveness of domestic enterprises, as well as their ability and propensity to invest. In turn, larger corporate investment supports technological progress, as the latter is largely embodied by the new capital (cf. Geenwood, Hercowitz and Krusell, 1997). If it includes VAT increases, it lowers the profitability of non-tradables relative to tradables. Government savings also increase room for internal devaluation or, more generally, for reductions in most distortionary taxes. However, such a fiscal restraint may have an inappropriate composition for economic growth, i.e. it may be based on distortionary taxation or on cuts to potentially growth-enhancing public expenditure (e.g. expenditure on infrastructure). Moreover, government purchases of assets or, more generally, its management over growing stock of savings, may significantly distort capital allocation.

The remaining two solutions may distort capital allocation and accumulation to an even higher extent. They are associated with the risk of over-taxation and over-regulation of the financial sector, respectively. The materialisation of that risk would undermine the capacity of the domestic financial sector to encourage private saving, and to allocate it to productive projects.[12] Furthermore, in an economy which is open to capital flows, domestic credit could be replaced by loans taken abroad. In this case a credit boom could still occur. Thus, the main positive effect of solutions (i) and (ii)

---

[11] The overall mood was well captured by Janet Yellen at the last meeting chaired by Alan Greenspan in January 2006, when she said "the situation you're handing off to your successor is a lot like a tennis racquet with a gigantic sweet spot."

[12] Note that these solutions can prevent growth of the financial sector, and in particular of banking sector, to a size when it ceases to support economic growth, and may even inhibit it. Vast literature on finance and growth is surveyed e.g. by Levine (2005). For more on nonlinearity in the link between growth and financial sector's size see, e.g. Arcand, Berkes and Panizza (2012), Cecchetti and Kharroubi (2012), Cecchetti, Mohanty and Zampolli (2011), and Pagano (2012).

would be less frequent instances of forbearance lending during the bust (and thereby the faster restructuring of imbalances accumulated over the boom[13]). However, the negative aspect of less frequent forbearance during a bust is a sharper recession.

To address the risks of halting the development of the domestic financial sector and of substitution between domestic credit and foreign loans, regulation or taxation should be levied directly on borrowers rather than on domestic financial institutions. Moreover, to be effective, it should be very broadly based, even at the cost of being "too blunt". During the boom, the substitutability of various liabilities (assets) increases and in such an environment the narrow, targeted measures are likely to fail.

Unfortunately, all the aforementioned solutions are difficult to implement and, except for solution (ii)[14], their adequate calibration is an even more complex issue.

The feasibility of the solutions is questionable primarily due to political reasons, however not exclusively. Political pressure is more likely the more successful the measures are (cf. Elliott, Feldberg, and Lehnert, 2013).

It is not by coincidence that very few countries managed so far to run surplus in public finances for a longer period of time (and the few that succeeded are mostly undemocratic). The effective implementation of solution (i) would require continuous increases in government surplus over the convergence period.[15] The longer that period, the larger the necessary surplus. If the surplus ceased to increase in spite of economic growth in a given country exceeding that of the euro area, it would no longer protect the domestic economy against the credit boom-bust cycle. It is difficult for governments to convince their electorate that the benefits of adopting the euro should be saved instead of being spent immediately.

Even the less ambitious goal of building a fiscal space sufficient to absorb potential costs of credit boom-bust cycle appears to be infeasible for at least two reasons. First, the costs can be very large. The fiscal burden of rescuing financial institutions can exceed 30% of GDP, sometimes by a very large margin (as in the case of Ireland and Iceland after the GFC; see Laeven and Valencia's database on financial crisis). Second, the credit boom blurs the actual fiscal stance. On the one hand, it produces artificially large tax revenues. On the other hand, it encourages governments to introduce still new or more generous social schemes, whose full costs are not revealed until households have significant alternatives to social benefits. The stronger the boom, the seemingly healthier the fiscal stance. However, once the recession hits, even a large surplus in public finances may transform into a huge deficit (as shown by the experience of e.g. Ireland and Spain). Thus, it is not surprising that a financial crisis is often followed by a fiscal crisis (cf. Reinhart and Rogoff, 2009 and 2011) and that this pattern was also seen in the euro area periphery.

---

[13] For more on costs of forbearance lending, see, e.g. Caballero, Hoshi and Kashyap (2008).

[14] At least, if it takes the form of a tax on loan principal.

[15] Moreover, tightening of fiscal policy should not only be continuous but also occur on time to effectively curb risk of a credit-boom bust cycle. If it lagged behind credit growth, it could hardly prevent a boom. The less economic agents remain credit-constrained, the more the fall in private saving offsets the increase in government saving (cf. Elmendorf and Mankiw, 1998).

As much as taxing loan principal is unpopular, there are many countries that offer generous tax breaks or direct subsidies for mortgage loans. The majority of countries offer tax incentives for financial leverage in the corporate sector. The outbreak of the GFC has not convinced governments to withdraw tax benefits for debt financing, they have been expanded rather than limited (see the case of the UK in particular).

From the perspective of the political economy, taxing the non-core liabilities of banks appears to be a more feasible way of implementing solution (ii), given the banks are still charged with the responsibility for the outbreak of the GFC. However, an increase in domestic credit costs, spurred on by the tax on those liabilities, may encourage domestic agents to pursue financing options abroad.

The GFC has greatly revived the debate on macro-prudential measures which could be part of the solution (ii) or (iii), or both. Most of the discussed and implemented measures, such as higher capital requirements, liquidity coverage, net stable funding ratio, and leverage ratio have a regulatory flavour. Yet, most countries are afraid of regulation arbitrage, which renders them reluctant to introduce/calibrate those measures unilaterally. Thus, economic knowledge on the effectiveness and the workings of macro-prudential measures is likely to remain limited in the foreseeable future (cf. Repullo and Saurina, 2011).

The political economy suggests that the feasibility of macro-prudential measures is not independent of fiscal policy. If there was no fiscal policy reaction to excessive credit growth, then the macro-prudential response would have to be more restrictive, however a decisive macro-prudential response from a hesitant government is unlikely, unless the decision belongs to an institution sufficiently independent of government.

That independence should be very deeply rooted, given that macro-prudential decisions would impact on the distribution of income (which would not be expected to expire over the business cycle, unlike those of e.g. monetary policy). Such effects inspire political pressures. Apart from independence, macro-prudential policy also requires systemic macroeconomic perspective, and both these requirements indicate the central bank as the appropriate institution for pursuing that policy.

It is worth noting that euro adoption would alleviate the problem of the difficult 'cohabitation' of monetary and macro-prudential policies (provided that macro-prudential measures would be decided at the national level). When the national central bank is primarily responsible for price stability, it is naturally tempted to subordinate macro-prudential actions to monetary policies. That temptation stems from the fact that inflation is observable and systemic risk is not. At the same time, the business cycle is much shorter than the credit cycle and, as a result, the risk of inflation's deviation from the target increases much more frequently than the risk of financial crisis. Furthermore, when credit begins to fuel a speculative bubble, inflation usually subsides (cf. Christiano et al., 2010).

For Poland, the importance of introducing a solution to effectively mitigate the risk of a credit boom-bust cycle is increased by the underdeveloped private rental

market (see Figure 15). This underdevelopment might exacerbate the responsiveness of house prices and residential investment to financing conditions, and thereby the magnitude of the boom-bust cycle.

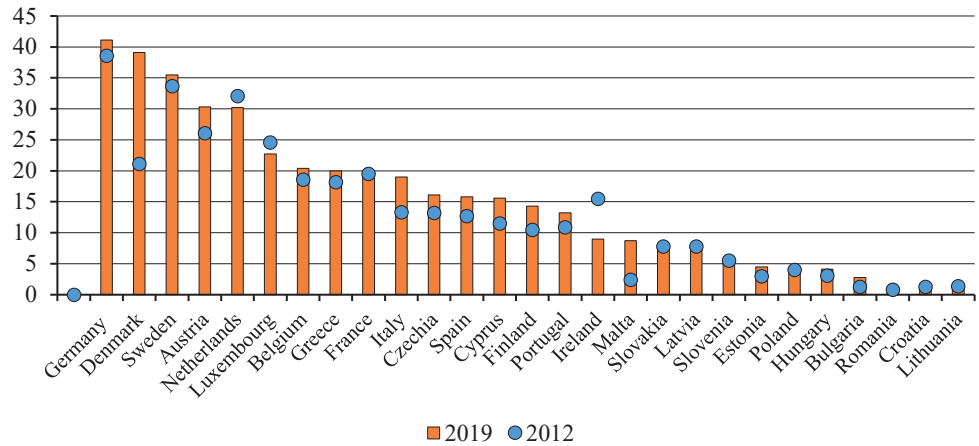


Fig. 15. The share of households living in houses rented in the private market across EU countries in 2012 and 2019

Source: Eurostat Series: distribution of population by tenure status, type of household and income group.

Poland's experience with fiscal discipline in general and fiscal rules in particular is mixed (at best). On the one hand, debt ceilings have helped to keep public debt below 60% of GDP before 2020, while on the other, the rules have encouraged fiscal discipline only during economic slowdowns. During booms, the deficit remained high, and except for 2007 and 2015-2019 it always exceeded 3% of GDP[16]. Fiscal restraint, when introduced, was not always conducive to economic growth in the long term (e.g. government outlays on infrastructure were cut and the capital pillar of the pension system has been marginalised). It was also often supplemented with changes in definition of public debt, which rendered public finances less transparent. Lastly, when in 2013 the debt ceiling became a hindrance for increasing the deficit, it was suspended. This precedent encouraged the next government to modify the most recent fiscal rule in Poland, i.e. the stabilising expenditure rule, whenever it interfered with projected expenditure. The major softening of the rule was carried out already in the autumn of 2015. The new government modified then the formula setting out the public spending ceiling by replacing inflation (then actually deflation) with the inflation target (2.5%). By 2019 just this one change raised the public

[16] Before the reduction of the capital pillar albeit until recently, the deficit was largely offset by savings from mandatory pension funds.

spending ceiling by a total of about 45 billion PLN (2% of GDP). Despite such a dilution, the stabilising expenditure rule has been circumvented through inflating revenue allegedly from curbing tax evasion and tax avoidance. During the pandemic, numerous new off-budget funds not covered by the rule were created; meanwhile the rule has been suspended altogether.

The pressure to increase government spending is strengthened in Poland (and other NMS) by comparisons of the quality of public services in Poland and Western Europe. Adopting the euro would facilitate such comparisons, in particular with regard to various social benefits and wages. The average Pole would not accept the statement that "despite the booming economy and a budget surplus, slowing the pace of increases in pension and wages is necessary due to necessary restraint of the credit boom bust cycle". Therefore, even though enhancing fiscal discipline in general and fiscal rules in particular is necessary (aside from the matter of euro), it cannot be the only or even the main shelter from the credit boom bust cycle.

The Polish experience with macro-prudential measures seems more encouraging. Already before the outbreak of the GFC, Poland introduced some measures (in particular recommendation "S") aimed at curbing the growth of FX loans, which were very popular at the time. In August 2008, just before the collapse of Lehman Brothers and the most severe phase of the GFC, the share of FX mortgages in all newly issued mortgages exceeded 70%. This tendency has reversed since 2008, and over the subsequent few years the share of newly issued FX loans in Poland became negligible.

However, the more limited credit boom prior to the GFC than in most other New Member States appears to be related more to the sharp monetary policy tightening in 2004, which contributed[17] to the interruption of the post-EU-accession boom. This tightening and, more generally, the success of the central bank in stabilising the economy, at least until recently (see Figures 16-19), suggests it can make a good use of its independence (cf. Ciżkowicz and Rzońca, 2015).

The NBP should be given wide competencies in the field of macro-prudential policy as soon as possible. The sooner it becomes responsible for this kind of measures, the more time it will have to learn how to use them. By the time of joining the euro area, the NBP will know how to adjust the effect of the ECB interest rates on Poland's economy if they are considered inappropriate; prevention against credit booms would then become its primary task. However, performing this task could be hindered by excessive harmonisation of the macro-prudential policy at EU level, signs of which can be seen in the CRDIV directive and CRR regulation (cf. NBP, 2014).

[17] Jointly with the positive outcomes of the Hausner Plan.

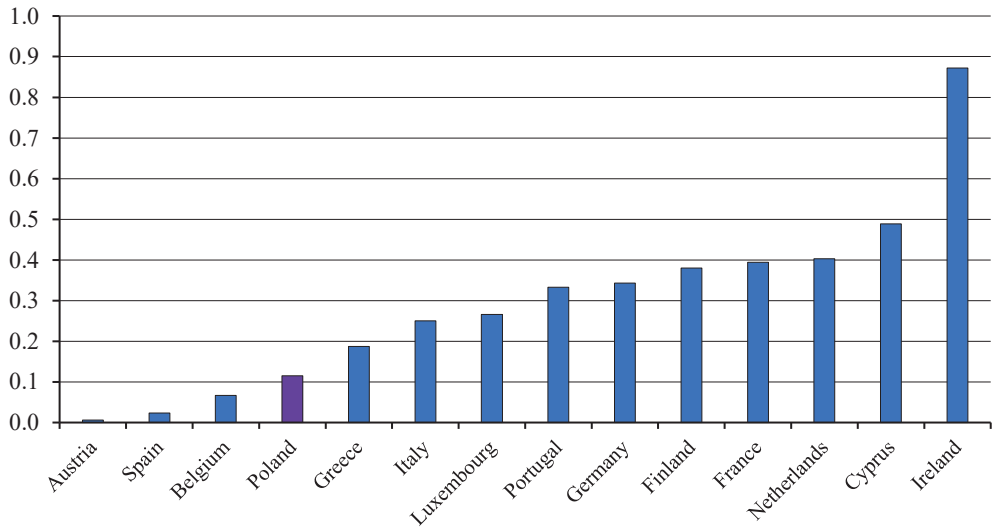


Fig. 16. Average deviation from the inflation target (01.2004-12.2019) in Poland and the euro area – 12 countries. Absolute values, percentage points.

Source: authors' elaboration, based upon Eurostat Series HICP monthly data (12-month average rate of change).

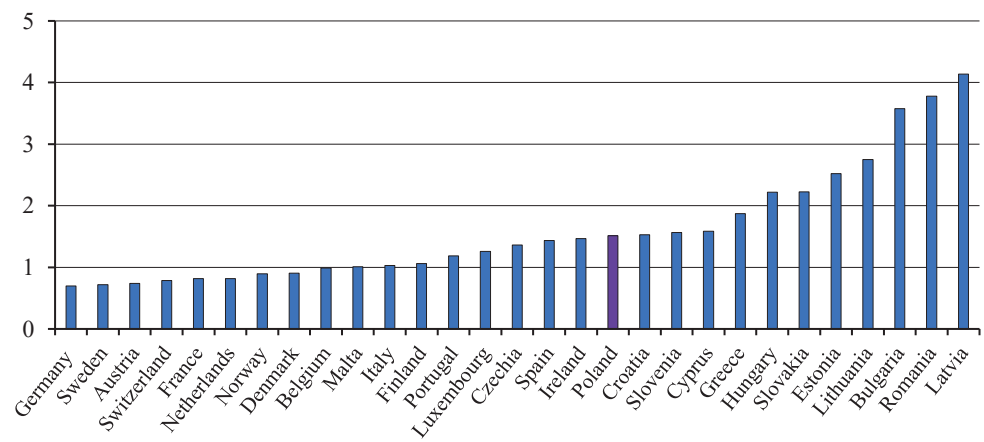


Fig. 17. HICP volatility in the EU (standard deviation, 01.2004-12.2019)

Source: authors' elaboration, based upon Eurostat. Series: HICP monthly data (12-month average rate of change).

Economic turbulence cannot be avoided entirely, and Poland should better prepare for its occurrence before joining the euro area. Improving resilience will require carrying out two kinds of actions.

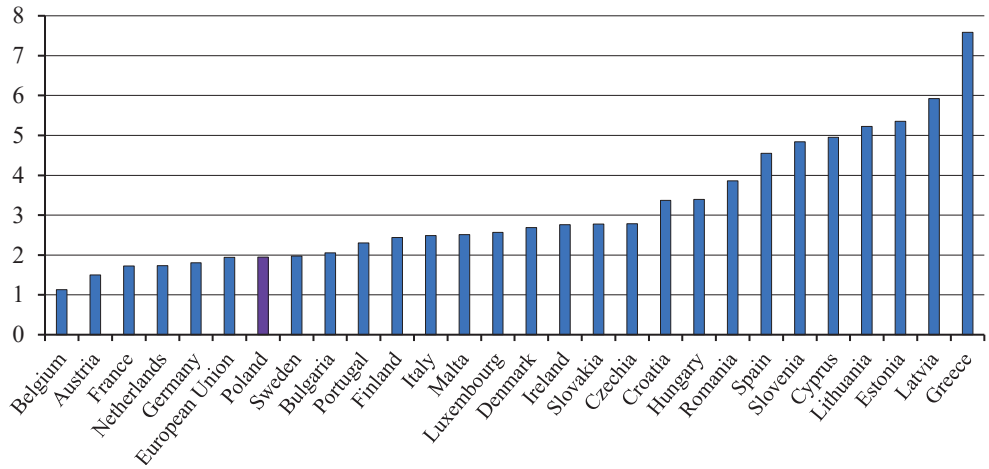


Fig. 18. Output gap volatility in the EU (standard deviation, 2004-2019)

Source: Ameco.

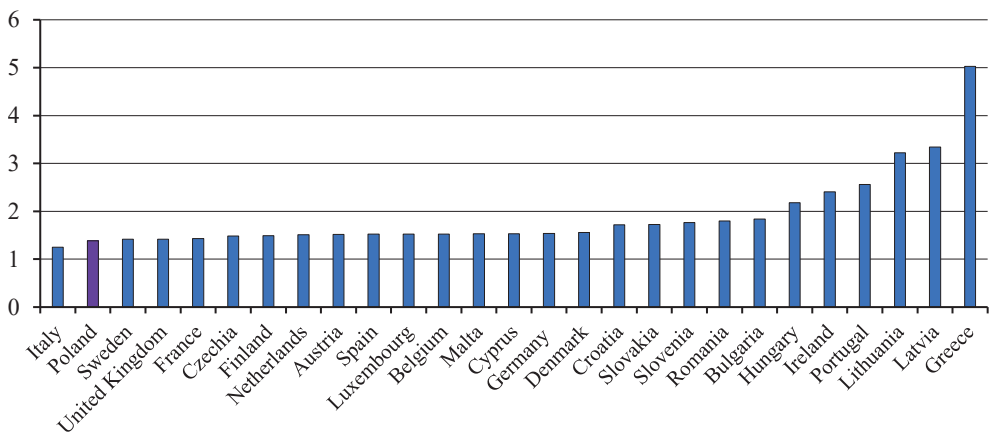


Fig. 19. Long-term interest rate volatility (2004-2019)

Source: Eurostat Series: EMU convergence criterion.

First of all, fiscal space to mitigate moderate-negative shocks should be created. To make this possible, the structural deficit has to decrease by at least 3% of GDP, otherwise it will remain far above the medium-term budgetary objective (MTO) adopted by Poland. The stabilising expenditure rule was designed to ensure the achievement of the MTO. However, as outlined before, this rule cannot play its role as long as it is circumvented.

Secondly, the economy's ability to restructure should be strengthened, and the costs of such adjustment to the shocks should be reduced. To accomplish this, labour

and capital reallocation must be enhanced. It has decelerated in Poland in recent years, e.g. flows into and out of the unemployment have become quite limited by European standards (see Figure 20)[18]. This may reflect the relatively flexible labour marker, facilitating wage adjustments, as well as the shortage of labour, encouraging labour hoarding (see Figure 21).

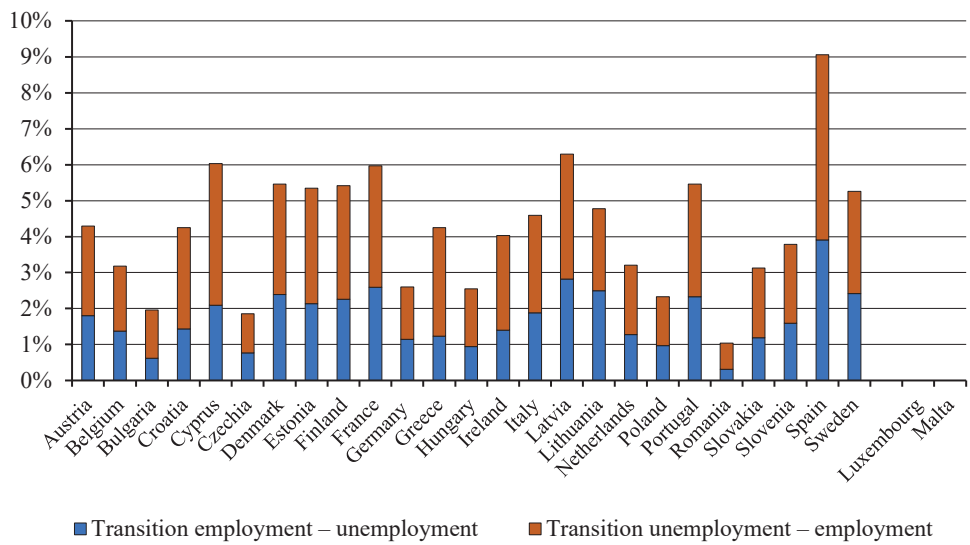


Fig. 20. Flows from employment to unemployment and opposite. Data from 2019.

Source: Eurostat Series: Labour market transitions.

On the one hand, Poland's economy flexibility is increased by an overdeveloped sector of micro-firms (hiring less than 10 employees). Across the EU, this was larger than in Poland only in Slovakia and in the Mediterranean countries: Cyprus, Spain, Portugal and Italy. Micro-firms may reduce the number of their poorly protected staff overnight or cut their wages, however they are also a cause of Poland's backwardness in relation to Western Europe. Labour productivity in micro-firms does not exceed a third of the productivity of their peers in the 'old' EU (apart from the peripheral states). In contrast, in small (10-49 employees) and medium (50-249 employees) enterprises, this amounts to three-thirds of the Western level, while in large firms it reaches 90% of that level (cf. Ciżkowicz, Łaszek and Rzońca, 2020).

[18] That stays in stark contrast with the pace of labour reallocation directly after the GFC. Then the probability of a person changing employment status during a quarter was 12% in Poland, whereas in Italy 5%, in Germany less than 6%, while in France less than 8% (Bukowski, Kowal and Lewandowski, 2011).

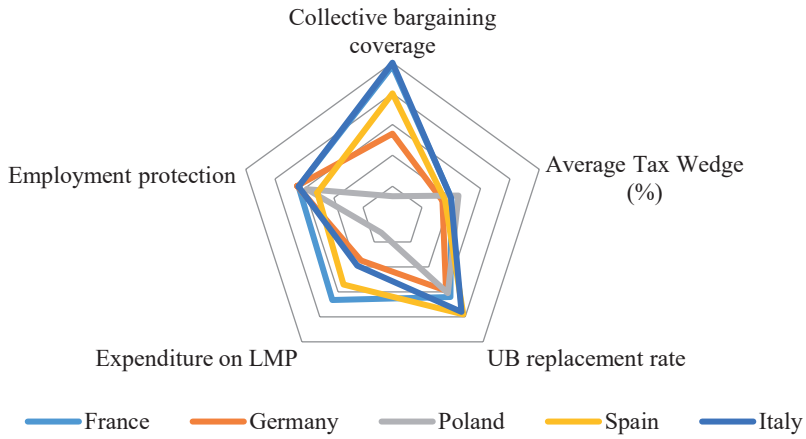


Fig. 21. Employment protection

Source: authors' elaboration based upon OECD Stat. Series: Collective bargaining coverage, Average Tax Wedge, Net replacement rate in unemployment, Public expenditure on LMP and Strictness of employment protection.

On the other hand, the flexibility of Poland's economy is much lower than for the OECD top performers, in particular as far as product market is concerned.[19] This is due to the large sector of state-owned enterprises. State control is particularly strong in upstream sectors, impacting basically on the operational costs of each company. These sectors in Poland are more monopolised than on average in the EU or the OECD. Barriers to competition are especially acute in the gas and aviation industry. Moreover, the retail sector in Poland operates under tighter regulations than in most other EU countries; it accounts for about 20% of GDP and has a strong impact on the efficiency of suppliers (cf. McKinsey, 2006).

Since this flexibility is also fundamental for the pace of the convergence process, its enhancement is of crucial importance irrespective of when Poland joins the euro area. Indeed, Poland would benefit most in terms of growth if it deregulated its product market in line with OECD best practice (cf. Bourlès et al., 2010).

The authors conclude this section with a very brief discussion of some other challenges of euro adoption.

The euro area deals with the large public debt exacerbated by the outbreak of the pandemic (see Figure 22). The high level of public debt is detrimental to economic growth (cf. Cecchetti, Mohanty and Zampolli, 2011; Kumar and Woo, 2010; Reinhart, Reinhart and Rogoff, 2012; Reinhart and Rogoff, 2010), and requires fiscal adjustment

[19] The already mentioned underdevelopment of the private rental market is another important sphere which requires far-reaching improvement. On top of the increasing amplitude of a potential boom-bust cycle, it decreases spatial labour mobility, and thereby hinders reallocation.

even if it is to be accompanied by output losses in the short run. Fortunately, evidence exists that when public debt is large (namely exceeds 60% of GDP), fiscal adjustment is not necessarily costly in terms of output in the short term (and is beneficial in the long run – cf. Ilzetzki, Mendoza and Vegh, 2013). At the same time, the public debt crisis in peripheral economies has confirmed that delaying fiscal adjustment, if feasible at all, could be very costly – both for countries which need this adjustment and for the euro area as a whole.

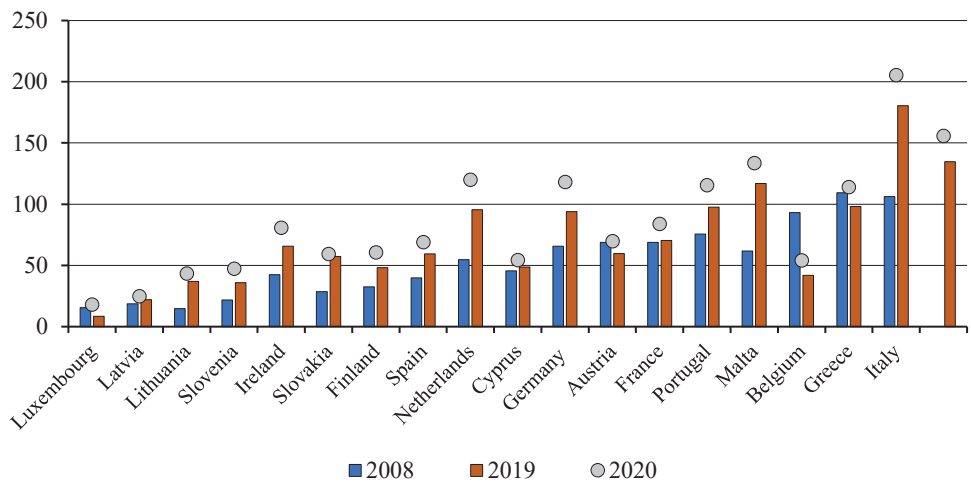


Fig. 22. General government debt across the euro area members (in % of GDP)

Source: Eurostat, National accounts indicator (ESA 2010) Government consolidated gross debt.

High public debt levels are not the only barrier to growth and a source of potential vulnerability which the euro area must face. The response to the GFC and public debt crisis in the euro area, and then to the pandemic, has too often consisted in policies that resulted in postponing deep restructuring. In particular, accommodative monetary policy pursued by the European Central Bank could discourage banks and enterprises from restructuring and weaken incentives of Euro area members to reform (cf. Balcerowicz et al., 2013).

Completing the institutional reforms is still another challenge for the euro area, which could considerably affect the cost-benefit analysis of adopting the euro. Questions of critical importance will remain unanswered for a long time: how effective will the Single Resolution Mechanism (SRM) be in preventing the build-up of risks in the banking system and reducing the costs of potential resolutions? How will the cooperation of different institutions involved in the resolutions work in practice? etc. However, the full balance of benefits and costs of euro adoption cannot ignore the political consequences of staying outside the euro area.

The first political argument in favour of Poland's accession to the euro area is associated with the radical reduction of the risk that Poland will leave the EU or that European integration will continue without Poland – only inside the euro area. A sign that further European integration is very likely to take place within that area is the European Commission's proposal that the new EU Multiannual Financial Framework for 2021-2027 should secure funds to be used exclusively by euro area members, or to accelerate preparations for euro area membership.

The second political argument for Poland's membership in the euro area concerns external security. Even if security is not breached in a given period, fears of economic agents of a significant external threat alone, can negatively affect economic growth. In contrast, the resolution of these fears promotes economic growth (see, for example, Landau, 1996; Baffes and Shah, 1998; Aizeman and Glick, 2003). The economic potential of Russia, Poland's aggressive and unpredictable neighbour, is small compared to the euro area. The largest economy, Germany, is two and a half times larger than Russia's. France, being the second largest economy in the euro area, is almost twice as large, while the entire euro area is about ten times as large. Naturally, Russia spends a much larger percentage of GDP on its army than any other euro area country. Even so, the difference in economic potential is so large to Russia's disadvantage that the combined national defence spending of France and Germany, just two euro area members, exceeds that of Russia by half. If the euro area countries combined their defence capabilities, they would be able to build a credible defence system against Russia without increasing their national defence spending (De Grauwe, 2018).

Concluding comments

Poland has become deeply integrated with the euro area, and some links (e.g. foreign trade) between Poland and the euro area are even stronger than among this area's members. The potential benefits for Poland from euro adoption would stem from deepening this integration.

By contrast, a decrease in cost of capital, considered to be a benefit of adopting the euro in the existing cost benefit analyses, appears to represent a source of the most significant threat. As long as Poland keeps growing faster than the euro area, the interest rates set by the ECB will be too low for the Polish economy. This might easily result in a credit boom-bust cycle, the risk and costs of which would be exacerbated by the underdeveloped rental market. Such a cycle might stop the process of convergence of the Polish economy. A halt in this process might result not only from the boom-bust cycle, but also from undermined cost competitiveness or slow restructuring. However, Ireland and the Baltic States have demonstrated that a member of the euro area can regain competitiveness through fiscal devaluation.

Theoretical solutions to mitigate the risk of a credit boom-bust cycle are broadly available. It suffices to increase the price of credit or constrain its volume adequately

through fiscal or macro prudential measures. However, the feasibility of these solutions is questionable, primarily due to political pressures.

The feasibility problem might be overcome through delegating the power of undertaking macro prudential measures to the National Bank of Poland. This institution is independent of political pressures and its track record with stabilising the economy (at least until recently) suggests that it could make good use of such power. The NBP should be assigned with macro-prudential competencies as soon as possible. Thereby, by the time of joining the euro area, it would already have the necessary experience and could correct the effects of ECB interest rates on the Polish economy. After adopting the euro, the NBP could focus on macro prudential measures. Hence, the problem of the 'difficult cohabitation' of monetary and macro prudential policy would be resolved.

Before joining the euro area, Poland should also improve its flexibility to reallocate labour and capital across companies and sectors as it lags far behind OECD top performers on that score.

References

Aizeman, J., and Glick, R. (2003). Military expenditure, threats, and growth, *FRBSF Working Paper*, 2003-08.

Arcand, J.-L., Berkes, E., and Panizza, U. (2012). Too Much Finance?. *IMF Working Paper*, 12/161.

Baffes, J., and Shah, A. (1998). Productivity of Public Spending, Sectoral Allocation Choices, and Economic Growth. *Economic Development and Cultural Change*, 46(2), 291-303.

Bakker, B. B., and Gulde, A.M. (2010). The Credit Boom in the EU New Member States: Bad Luck or Bad Policies? *IMF Working Papers,* 10/130.

Balcerowicz, L., Rzońca, A., Kalina, L., and Łaszek, A. (2013). *Economic Growth in the European Union*. Brussels: Lisbon Council.

Baldwin, R. (2006). The Euro's Trade Effects. *ECB Working Paper,* 594.

Bank of Slovenia and Slovenian Government (2003). *Programme for ERM II entry and adoption of the euro. Joint programme of the Slovenian Government and the Bank of Slovenia*. Ljubljana: Bank of Slovenia and Slovenian Government.

Beck, T., Büyükkarabacak, B., Rioya, F. K., and Valev, N. T. (2012). Who Gets the Credit? And Does It Matter? Household vs. Firm Lending Across Countries. *B.E. Journal of Macroeconomics*. 12(1), Art. 2.

Bitans, M., and Kauzens, E. (2004). Impact of the Euro Adoption on the Economy of Latvia. *Bank of Latvia Working Paper*, 2/2004.

Borio, C., and Drehmann, M. (2009). Assessing the risk of banking crises – revisited. *BIS Quarterly Review*, Bank for International Settlements, March.

Borio, C., Vale, B., and von Peter, G. (2010). Resolving the financial crisis: are we heeding the lessons from the Nordics? *BIS Working Papers* 311.

Borowski, J. (ed.) (2004). Raport na temat korzyści i kosztów przystąpienia Polski do strefy euro. Warszawa: Narodowy Bank Polski.

Bourlès, R., Cette, G., Lopez, J., Mairesse, J., and Nicoletti, G. (2010). The Impact on Growth of Easing Regulations in Upstream Sectors. *CESifo DICE Report,* 3/2010, 8-12.

Bukowski, M., Kowal, P., and Lewandowski, P. (2011). *Modelowa analiza rynków pracy o różnej strukturze instytucjonalnej*. Warszawa: Instytut Badań Strukturalnych.

Business Insider (2019). Kaczynski mówimy „nie" euro, mówimy „nie" europejskim cenom", 13.04.

Caballero, R. J., Hoshi, T., and Kashyap, A. K. (2008). Zombie Lending and Depressed Restructuring in Japan. *American Economic Review*. 98(5), 1943-1977.

Campos, N.F., Coricelli, F., and Moretti, L., (2014). Economic Growth and Political Integration: Estimating the Benefits from Membership in the European Union Using the Synthetic Counterfactuals Method. *IZA Discussion Paper*, 8162.

CBOS (2013). *Obawy i nadzieje związane z wprowadzeniem euro w Polsce*. Warszawa: Fundacja Centrum Badania Opinii Społecznej.

Cecchetti, S. G., and Kharroubi, E. (2012). Reassessing the impact of finance on growth. *BIS Working Paper,* 381.

Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011). The real effects of debt. *BIS Working Paper,* 352.

Central Bank of the Republic of Lithuania (2013) *Impact of the Euro Adoption of the National Economy: an Overview of the Quantitative Assessment*. Vilnius: Central Bank of the Republic of Lithuania.

Cerra, V., Saxena, S. C., and Panizza, U. (2009) International Evidence on Recovery from Recessions. *IMF Working Papers*. 09/183

Christiano, L., Ilut, C., Motto, R., and Rostagno, M. (2010). Monetary Policy and Stock Market Booms. *Paper prepared for the Federal Reserve Bank of Kansas City Annual Conference Jackson Hole, Wyoming*. August 28.

Ciżkowicz, P., Łaszek, A., and Rzońca, R. (2020). How big are labour productivity disparities in Poland? *Financial Law Review*, 20 (4), 1-21.

Ciżkowicz, P., Radzikowski, B., Rzońca, A., and Wojciechowski, W. (2020). Fiscal devaluation and economic activity in the EU. *Economic Modelling*, 88(C), 59-81.

Ciżkowicz, P., and Rzońca, A. (2015). Inflation Targeting and its Discontents: The Case of Poland. *Acta Oeconomica*, 65(supplement), 107-122.

Claessens, S., M., Kose, A., and Terrones, M.E. (2009). What Happens During Recessions, Crunches and Busts. *Economic Policy,* 24(60).

Cssajbok, A., and Csermely, A. (eds.) (2002). Adopting the Euro in Hungary: Expected Costs, Benefits and Timing. *NBH Occasional Papers*, 24.

De Grauwe, P. (2018). Why Russia is politically and militarily strong while being an economic dwarf. *VOX*, 16 June.

Elliott, D. J., Feldberg, G., and Lehnert, A. (2013). The History of Cyclical Macroprudential Policy in the United States. *Federal Reserve Board Divisions of Research & Statistics and Monetary Affairs. Finance and Economics Discussion Paper*, 2013-29.

Elmendorf, D. W., and Mankiw, G. N. (1998). Government debt. *Paper prepared for Handbook of Macroeconomics.*

European Commission (2007). *Autumn Forecast 2007*. Brussels: European Commission.

European Commission (2009). *The 2009 Ageing Report: Economic and Budgetary Projections for the 27 EU Member States (2008-2060)*. Brussels: European Commission.

European Commission (2014). *Spring Forecast 2014*. Brussels: European Commission.

European Commission (2020a). *The Convergence Report.* Brussels: European Commission.

European Commission (2020b). *Eurobarometer 487 – Introducing the euro is seen as positive for the euro area countries*. Brussels: European Commission.

Fernandez-Villaverde, J., Garicano, L., and Santos, T. (2013). Political Credit Cycles: The Case of the Euro Zone. *NBER Working Paper*, 18899.

Fidrmuc, K., and Korhonen, I. (2006). Meta-analysis of the business cycle correlation between the euro area and the CEECs. *Journal of Comparative Economics*, 34(3), 518-537.

Fischer, S., and Easterly, W. (1990). The economics of the government budget constraint. *World Bank Research Observer,* 5(2), 127-142.

Frankel, J. A., and Rose, A. K. (1998). The Endogeneity of the Optimum Currency Area Criteria. *Economic Journal*, 108(449), 1009-1025.

Frankel, J. A., and Rose, A. K. (2002). An Estimate of the Effect of Common Currencies on Trade and Income. *Quarterly Journal of Economics*, 117(2), 437-466.

Geenwood, J., Hercowitz, Z., and Krusell, P. (1997). Long-run implications of investment-specific technological change. *American Economic Review.* 87(3), 342-362.

Gelain, P., and Lansing, K. J. (2014). House Prices, Expectations, and Time-Varying Fundamentals. *Federal Reserve Bank of San Francisco Working Paper*, 2013-03.

Gros, D., and Alcidi, C. (2013). Country adjustment to a 'sudden stop': Does the euro make a difference? *European Commission Economic Paper*, 492.

Havránek, T. (2010). Rose Effect and the Euro: Is the Magic Gone? *Review of World Economics*, 146(2), 241-261.

IBnGR (2012). Wpływ realizacji polityki spójności na kształtowanie się głównych wskaźników dokumentów strategicznych NPR 2004-2006 i NSS 2007-2013. *Raport półroczny 2012*. Gdańsk, Warszawa: Prevision.

IBS (2012a). *Wpływ realizacji polityki spójności na kształtowanie się głównych wskaźników dokumentów strategicznych – Narodowego Planu Rozwoju 2004-2006 i Narodowej Strategii Spójności 2007-2013 oraz innych wybranych wskaźników na poziomie krajowym i regionalnym, Raport półroczny 2012*, Warszawa: Instytut Badań Strukturalnych.

IBS (2012b). *Wpływ realizacji polityki spójności na kształtowanie się głównych wskaźników dokumentów strategicznych – Narodowego Planu Rozwoju 2004-2006 i Narodowej Strategii Spójności 2007-2013 oraz innych wybranych wskaźników makroekonomicznych na poziomie krajowym i regionalnym za pomocą modeli HERMIN (krajowego i regionalnych)*. Warszawa: Instytut Badań Strukturalnych.

Ilzetzki, E., Mendoza, E. G., and Végh, G. C. A. (2013). How big (small?) are fiscal multipliers? *Journal of Monetary Economics* 60(2), 239-254.

IMF (2014). World Economic Outlook April 2014, Washington D.C.: IMF.

Jiménez, G., Ongena, S., Peydró, J. L., and Saurina, J. (2009). Hazardous Times for Monetary Policy: What do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking? *Banco de España Working Paper*, 0833.

Kawalec, S., and Pytlarczyk, E. (2016). *Paradoks euro. Jak wyjść z pułapki wspólnej waluty?*. Warszawa: Wydawnictwo Poltext.

Keller, W. (2004). International Technology Diffusion. *Journal of Economic Literature*, 42(3), 752-782.

Kenen, P. B. (1969). The Theory of Optimum Currency Areas: An Eclectic View. In Mundell R. A., Swoboda A. (eds.) Monetary Problems of the International Economy. The University of Chicago Press, Chicago.

KNF (2020). *Informacja na temat sytuacji sektora bankowego. Opracowanie: Departament Bankowości Komercyjnej i Specjalistycznej. Zespół Analiz Sektora Bankowego*, Warszawa: Urząd Komisji Nadzoru Finansowego.

Kumar, M. S., and Woo, J. (2010). Public debt and growth. *IMF Working Paper,* 10/174.

Laeven, L., and Valencia, F. (2008). Systemic Banking Crises: A New Database. *IMF Working Paper*, 08/224.

Landau, D.(1996). Is one of the 'peace dividends' negative? Military expenditure and economic growth in the wealthy OECD countries. *Quarterly Review of Economics and Finance*, 36(2), 183-196.

Laszek, A. (2014). Wpływ kryzysów bankowych na długookresowy wzrost gospodarczy. *Mimeo.*

Levine, R. (2005). Finance and Growth: Theory and Evidence, in: Aghion P., Durlauf S. N. (eds.) *Handbook of Economic Growth. Volume 1A*. Amsterdam: Elsevier North-Holland.

Lünnemann, P., and Wintr, L. (2010). Downward wage rigidity and automatic wage indexation: Evidence from monthly micro wage data. *ECB Working Paper*, 1269.

Maddaloni, A., and Peydró, J.-L. (2010). Bank Risk-Taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro Area and U.S. Lending Standards. *Review of Financial Studies*, 24, 2121-2165.

McKinnon, R. I. (1963). Optimum Currency Areas. *American Economic Review*, 53(4), 717-724.

McKinnon, R. I. (2004). Optimum Currency Areas and Key Currencies: Mundell I versus Mundell II. *Journal of Common Market Studies*, 42(4), 689-715.

Mundell, R. A. (1961). A Theory of Optimum Currency Area. *American Economic Review*, 51, 657–665.

Mundell, R. A., and Swoboda, A. (eds) (1969), Monetary Problems of the International Economy. Chicago: University of Chicago Press.

NBP (2009). *Raport na temat pełnego uczestnictwa Rzeczypospolitej Polskiej w trzecim etapie Unii Gospodarczej i Walutowej*. Warszawa: Narodowy Bank Polski.

NBP (2014). *Ekonomiczne wyzwania integracji Polski ze strefą euro*. Warszawa: Narodowy Bank Polski.

Pagano, M. (2012). Finance: Economic Lifeblood or Toxin? *Paper prepared for the CEPR – Chicago Fed Annual International Banking Conference on The Social Value of the Financial Sector: Too Big to Fail or Just too Big?* November 15.

Piątkowski, M. (2013). Poland's New Golden Age Shifting from Europe's Periphery to Its Center. *World Bank Policy Research Working Paper,* 6639.

Podpiera, J., Wiegand, J., and Yoo, J. (2015). Euro Adoption in NMS: Macroeconomic Benefits and Challenges, in: Selected Issues Paper on Central and Eastern Europe: New Member States (NMS) Policy Forum, 2014. *IMF Country Report*, 15/98.

Rajan, R. G. (2005). Has financial development made the world riskier? *NBER Working Paper*, 11728.

Reinhart, C. M., Reinhart ,V. R., and Rogoff, K. S. (2012). Debt Overhangs: Past and Present*, NBER Working Paper,* 18015.

Reinhart, C. M., and Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press.

Reinhart, C. M., and Rogoff, K. S. (2010). Growth in a Time of Debt. *American Economic Review*, 100(2), 573-578.

Reinhart, C. M., and Rogoff, K. S. (2011). From Financial Crash to Debt Crisis. *American Economic Review*, 101(5), 1676–1706.

Rinaldi-Larribe, M.-J. (2013). Is economic convergence in New Member States sufficient for an adoption of the euro? *European Journal of Comparative Economics*, 5(2), 133-154.

Rose, A. K. (2002). The Effect of Common Currencies on International Trade: Where Do We Stand? *University of California at Berkeley Occasional Paper*, 22.

Sá, F., Towbin, P., and Wieladek, T. (2011). Low interest rates and housing booms: the role of capital inflows, monetary policy and financial innovation. *Bank of England Working Paper*, 411.

Schadler, S., Drummond, P., Kuijs, L., Murgasova, Z., and van Elkan, R. (2005). Adopting the Euro in Central Europe: Challenges of the Next Step in European Integration. *IMF Occasional Paper*, 234.

Sławiński, A. (2014). Abenomics to nie dla nas. *Forbes*, 2.

Suster, M. (ed.) (2006). The Effects of Euro Adoption on the Slovak Economy. National Bank of Slovakia.

Tavlas, G. S. (1993). The 'New' Theory of Optimum Currency Areas. *World Economy*, 16(6), 663-685.

Whelan, K. (2010). Policy Lessons from Ireland's Latest Depression. *The Economic and Social Review, Economic and Social Studies* 41(2).

Received: January 2022, revised: April 2022

Reviewers 2022/2023

Prof. dr hab. Katarzyna Cheba
Prof. dr hab. Grażyna Dehnel
Prof. dr hab. Ewa Dziedzic
Prof. dr hab. Ewa Dziwok
Prof. dr hab. Eugeniusz Gatnar
Prof. dr hab. Aldona Glińska-Neweś
Prof. dr hab. Witold Jurek
Prof. dr hab. Magdalena Kalińska-Kula
Prof. dr hab. Marcin Kuzel
Prof. dr hab. Dagmara Lewińska
Prof. dr hab. Paweł Lula
Prof. dr hab. Michał Majsterek
Prof. dr hab. Iwona Markowicz
Prof. dr hab. Anna Matuszyk
Prof. dr hab. Barbara Pawełek
Prof. dr hab. Marek Ratajczak
Prof. dr hab. Marcin Salamaga
Prof. dr hab. Teresa Skalska
Prof. dr hab. Małgorzata Tyrańska
Prof. dr hab. Piotr Trąpczyński
Prof. dr hab. Paweł Ulman
Prof. dr hab. Bogdan Włodarczyk
Prof. dr hab. Robert Zajkowski
Prof. dr hab. Bogusława Ziółkowska

ARGUMENTA OECONOMICA

Scope
Papers published in the ARGUMENTA OECONOMICA will encompass, but not be limited, to the following areas of Economics and Management:
– theoretical foundations
– practical evaluation of methodologies
– case studies and justification of research projects
– surveys of the state-of-the-art

Language
English

Note to contributors
– No manuscripts submitted shall have been or will be published elsewhere.
– Papers for publication should be submitted via https://review.ue.wroc.pl/
– Editors will initially screen submissions, prior to reviewing, and inform authors quickly if the paper appears unsuitable for this journal.
– Tables and Figures should be numbered.
– The manuscript should begin with an abstract and keywords. The abstract should not contain any formulae or references. Its purpose is to summarize clearly the principal conclusion of the paper.
– The paper should be subdivided into sections and subsections if necessary. Main headings should designate major sections, side headings should designate minor sections.
– References must be listed alphabetically by the surname of the first author followed by the title, journal name, year. Volume, pages, year for journal references; and author(s), title, publisher, city and year for the book references.
– All references should be indicated in the manuscript by the author's surname followed by the year of publication, e.g. (Hellwig 1968). Where the references include three or more authors, the form (Hellwig et al. 1968) should be used.
– Footnotes should be avoided. Parentheses should be used in their place.
– No maximum length for the paper is prescribed; however authors should write concisely.

Wroclaw University of Economics and Business
Komandorska 118/120

Argumenta Oeconomica – Editorial Office
e-mail: aoe@ue.wroc.pl
website: www.argumentaoeconomica.ue.wroc.pl