MANAGERIAL ECONOMICS

SEMIANNUAL

vol. 27 no. 1 ■ 2026

Wydział Zarządzania Akademii Górniczo-Hutniczej

EKONOMIA MENEDŻERSKA

PÓŁROCZNIK

tom 27 nr 1 ■ 2026

WYDAWNICTWA AGH

KRAKÓW 2026

Faculty of Management
AGH University of Krakow

MANAGERIAL ECONOMICS

SEMIANNUAL
vol. 27 no. 1 ■ 2026

AGH UNIVERSITY PRESS
KRAKOW 2026

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

Language Editor: *Raymond Clarke of Skrivanek sp. z o.o.*

Statistical Editor: *Anna Barańska*

Editorial support: *Magdalena Grzech*

Typesetting and Desktop Publishing: *Wydawnictwo JAK*

Cover and title page design: *Zofia Łucka*

DOI: https://doi.org/10.7494/manage

ISSN 1898-1143 (paper version), ISSN 2353-3617 (on-line version)

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

Wydawnictwa AGH (AGH University Press)
al. A. Mickiewicza 30, 30-059 Kraków
tel. 12 617 32 28
e-mail: redakcja@wydawnictwoagh.pl; www.wydawnictwo.agh.edu.pl

CONTENTS

Managerial Economics
2026, Vol. 27, No. 1, pp. 7–36
https://doi.org/10.7494/manage.2026.27.1.7

Agnieszka Bieńkowska*, Katarzyna Tworek**

The role of Employees' Dynamic Capabilities in shaping job performance through fake and authentic leadership

1. Introduction

Employees' Dynamic Capabilities (EDCs) (Bieńkowska, Tworek 2020) are proving to be an important factor shaping contemporary organizational performance. In principle, this seems natural for at least two reasons. Firstly, in strategic management theory, the resource approach is no longer sufficient (Teece et al. 1997). Nowadays, gaining a competitive advantage requires not only the ability to have the resources but also the ability to allocate them appropriately: preferably anticipating or also reacting to changes in the environment. The concept of dynamic capabilities of organizations by Teece, Pisano and Shuen (Teece et al. 1997) responds to the need for a dynamic view of organizations operating in turbulent environments. Secondly, employees are nowadays considered the most important resource of an organization. Their competences, in the form of knowledge, skills and attitudes, influence the effectiveness of achieving the goals set for the organization. According to the concept of human resource management, an organization's goals are achieved through the people it employs (cf. Syed, Jamal 2012; Saridakis et al. 2017). Combining both of those reasons, the idea of developing EDCs (Bieńkowska, Tworek 2020) should be assessed as legitimate, considering both the importance of human resources in an organization, as well as the necessity of their dynamic adaptation to the changes taking place in the environment and in the organization in the context of achieving the goals set for

* Wroclaw University of Science and Technology, Wrocław, Poland, e-mail: agnieszka.bienkowska@pwr.edu.pl, ORCID: 0000-0002-7498-6322
** Wroclaw University of Science and Technology, Wrocław, Poland, e-mail: katarzyna.tworek@pwr.edu.pl, ORCID: 0000-0002-6276-2436

that organization. Such a view is confirmed by numerous studies in this area (cf. Al Wali et al. 2020; Bocoya-Maline et al. 2024; Wang et al. 2024).

The impact of EDCs on organizational performance has been confirmed and presented primarily in an EDC-based model of job performance (Bieńkowska, Tworek 2020). The results of the study made it possible to confirm that EDCs influence job performance primarily through person-job fit, followed by job-related attitudes such as work motivation, job satisfaction, and work engagement (Bieńkowska, Tworek 2020). Further research confirmed that EDCs have a significant influence not only on the organization as a whole (Tworek et al. 2023), but also on human resources and organizational reliability (Bieńkowska et al. 2020), as well as, for example, on innovative work behavior (Al Wali et al. 2020). It has also been confirmed that the impact of EDCs on organizational performance, as well as on job performance, occurs during the crisis caused by a Black Swan event, which was particularly relevant during the COVID-19 pandemic crisis (Bieńkowska et al., 2021; Tworek et al., 2023). The above studies show that the concept of EDCs has the potential to play an important role in the management of contemporary organizations, not only in human resource management.

However, when considering the functioning of the employees in the organization, it is also important to discuss the role of the EDCs in terms of the context in which an employee with more or fewer EDCs operates. A fundamental element of an employees' broader work environment is their leader (who shapes this work environment) (Amabile et al. 2004; Omilion-Hodges, Ptacek, 2021), who influences employees in a specific, direct way in order to achieve an intended goal in the context of achieving the vision, mission and strategy of the organization as a whole. The way in which a leader influences employees can vary and is referred to as the concept of leadership. Leadership itself can be understood either as a trait of the leader's person or as a process of influence on subordinates aimed at achieving goals set by the leader (Bieńkowska, Tworek 2024). "Leadership has been described as a process, but most theories and research on leadership look at a person to gain understanding" (Horner 1997, p. 270). In this context, "Leadership is typically defined by the traits, qualities, and behaviours of a leader" (Horner 1997, p. 270).

There are many theories of leadership today. In the context of this study, three classical ones seem to be particularly important: contingency-based theory by Fiedler (Fiedler 1967) situational leadership theory by Hersey and Blanchard (Hersey, Blanchard, 1977), and goal-path theory by House and Mitchell (House, Mitchell 1974). Situational leadership theory allows for the consideration of different leadership styles appropriate to the specific situations an organization faces (internal and external environment) (cf. also Northouse 2021), while path-goal theory

treats leaders as those who are primarily responsible for helping their employees (followers) develop behaviors enabling them to achieve goals or desired outcomes (Horner 1997). The above makes it reasonable to conclude after Horner (Horner 1997, p. 272) that "there are many appropriate ways to lead or styles of leadership."

There is a distinction to be made between positive (constructive) and negative (destructive) leadership styles in contemporary literature. Positive ones enhance the results achieved by employees and by the organization as a whole, especially in terms of the goals it achieves, and negative ones generally undermine the aforementioned results (cf. e.g. Einarsen et al. 2007; Burns 2017). The literature discusses many aspects of the influence of positive and negative leadership styles on employees and the organization (cf. e.g. Alimo-Metcalfe 2013; Northouse 2021). Also, attempts are made to explain the mechanisms that accompany (build and explain) this influence (e.g. Starratt, Grandy, 2010; Boddy 2017; Palmen et al. 2021; Bieńkowska, Tworek 2024; Khalid et al., 2024; Liu et al., 2024). However, the mechanism of EDCs' influence on the relationship between a particular type of leadership and the job performance of employees has not yet been addressed, neither in theoretical nor in empirical terms, which constitutes a research gap. However, the existing research studies suggest that it must be assumed that such an influence exists, as the characteristics of the employees either mitigate or boost the influence of the leader on employees and on the organization as a whole (cf. Carmeli et al. 2013; Audenaert, Decramer 2018; Freiherr von Fircks 2024; Jia et al. 2024). In this context, the research questions arise: do (and in what way) EDCs have the potential to strengthen the positive influence of positive leadership on employees' job performance, and, conversely, do EDCs have the potential to mitigate the negative impact of negative leadership on employees' job performance? Moreover, whether EDCs' ability to mitigate the influence of negative leadership on employees' job performance should be seen as an asset or a threat to both employees and the organization should be examined.

In this context, the study will aim to examine the role of EDCs in the mechanism of shaping job performance by negative leadership (using the example of fake leadership), as well as in the mechanism of shaping job performance by positive leadership (using the example of authentic leadership). Both styles have been chosen as being clearly positioned at the two opposite ends of leadership styles – regarding their influence on employees and the organization (Bieńkowska, Tworek 2024). Authentic leadership is characterized by a pro-organization and pro-employee orientation, while fake leadership is characterized by an anti-organization and anti-employee orientation (Bieńkowska, Tworek 2024). Achieving the described aim will help to fill the identified research gap in the area of organizational behavior within the management sciences.

2. Fake and authentic leadership – models of influence on job performance

Fiedler's contingency-based theory (Fiedler 1967), situational leadership theory by Hersey and Blanchard (Hersey, Blanchard, 1977), and goal-path theory by House and Mitchell (House, Mitchell 1974), as mentioned in the introduction, allows for the consideration of different leadership styles. In general terms, they can be divided into negative and positive ones, taking into account the positive or negative impacts of different leadership styles on employees and the organization as a whole. Einarsen, Aasland and Skogstad (Einarsen et al. 2007) developed a two-dimensional model of destructive and constructive leadership behavior. Basing on that, Bieńkowska and Tworek (Bieńkowska, Tworek 2024) developed a three-dimensional approach that distinguishes between the orientation of leadership traits (self-orientation vs. collective orientation) in addition to the leader's behavior toward the employees and the organization. In both cases, positive (constructive) and negative (destructive) leadership styles can be distinguished. Constructive leadership styles combine a focus on both mission achievement and the well-being of the team and employees. "Constructive leaders combine human qualities such as honesty, respect, sincerity, fairness, and honour with organisational/team strengths such as confidence, focus, achievement and a drive for the greater good. Team members respect and place trust in constructive leaders, in contrast to what happens when they are led by a disruptive leader" (Burns 2017, p. 34). Destructive leadership styles are the opposite of that. Pelletier (Pelletier 2010, p. 375) defined destructive leadership as "systematic and repeated behavior by a leader, supervisor or manager that undermines the legitimate interests of the organization by undermining and/or sabotaging the goals, objectives, resources and effectiveness of the organization and/or the motivation, well-being or job satisfaction of subordinates."

Leadership has long been recognized as a key determinant of employee attitudes and performance (Avolio et al. 2004; Northouse 2021). Positive leadership has been associated with increased trust, engagement, and performance outcomes (Walumbwa et al. 2008; Wang, Hsieh 2013). Conversely, destructive or toxic leadership undermines employee well-being and organizational effectiveness (Einarsen et al. 2007; Pelletier 2010; Boddy 2017). This is why it seems relevant to analyze both mechanisms through which leadership can influence employees, their attitudes and performance.

In the present study, the concepts of authentic and fake leadership were chosen as extreme examples in the three-dimensional approach of Bieńkowska and Tworek (Bieńkowska, Tworek 2024). Authentic leadership has the pro-organization and pro-employee orientation and takes into account the collective

orientation of leader traits, whereas fake leadership has the anti-organization and anti-employee orientation and takes into account the self-orientation of leader traits (Bieńkowska, Tworek, 2024).

The concept of authentic leadership was developed by Cameron and colleagues (Cameron et al. 2003) and Walumbwa and colleagues (Walumbwa et al. 2008). Authentic leadership will be understood following Luthans and Avolio "as a process that draws on both positive psychological capacities and a highly developed organizational context, resulting in both greater self-awareness and self-regulated positive behavior on the part of leaders and colleagues, fostering positive self-development"(Luthans, Avolio 2003, p. 243). In contrast, Walumbwa and colleagues (Walumbwa et al. 2008, p. 94) defined "authentic leadership as a pattern of leader behavior that draws on and promotes both positive psychological capacities and a positive ethical climate to foster greater self-awareness, internalized moral perspective, balanced information processing and relational transparency on the part of leaders working with subordinates, fostering positive self-development." In this context, the authentic leadership can be understood as – first of all - comprising of full spectrum of positive traits and behaviors of a leader, who is focused on the collective goals and exhibits pro-employee and pro-organizational behaviors. Moreover as connected to sincere and authentic intent of a leader exhibited in their behaviors toward themself, the employees and the organization as well as showing leader's mindfulness to be transparent in relationship with employees, without any hidden intentions toward them or the organization, which is a basis for mutual trust (Bieńkowska, Tworek 2024).

The concept of fake leadership was developed by Bieńkowska and Tworek (Bieńkowska, Tworek 2024). It will be understood as a pattern of behavior aimed at fulfilling the individual goals of a leader, regardless of means necessary to do so, coupled with the need to hide the real intent of the leader, mitigating intra-organizational trust and employee self-assessment and self-esteem. In this context, fake leadership is understood as – above all – exhibiting full spectrum (not isolated instances) of traits and behaviors, with leader's focus on the individual goals, regardless of their value for the employees or organization, and anti-employee and anti-organizational behaviors. Moreover as comprising of leader's intentional engagement in negative behaviors toward employees and the organization as a whole, as well as showing leader's mindfulness in hiding the intent from employees and manifesting traits and behaviors aligned with authentic leader, engaging in positive behaviors towards them and the organization (Bieńkowska, Tworek 2024).

The two distinctive leadership styles have very different ways of influencing an employee's job performance. In each case, intra-organizational trust and classic job-related attitudes are important. Authentic leadership has a positive

influence on job performance through trust, work motivation, work engagement, and organizational commitment) (cf. George 2003; Avolio et al. 2004; George et al. 2007; Wang, Hsieh 2013; Bieńkowska, Tworek 2024). Fake leadership has a negative influence on job performance through trust, work motivation, work engagement, and organizational commitment (Bieńkowska, Tworek 2024). Mechanisms of the influence of both styles on employees' job performance are described by Bieńkowska and Tworek (Bieńkowska, Tworek 2024). Diagrams of both impacts are presented in Figure 1.

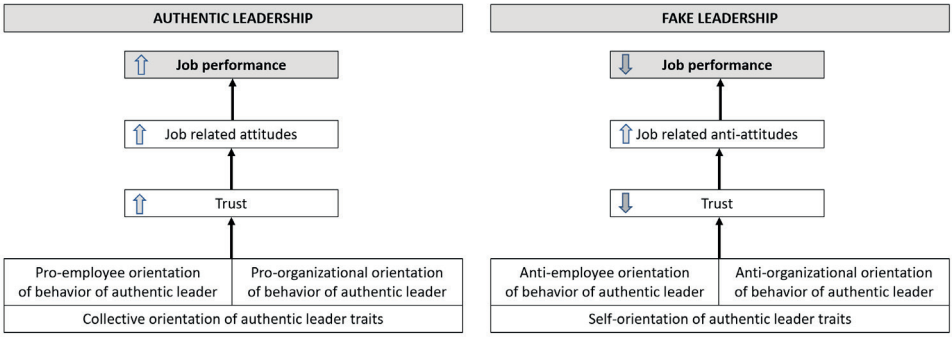


Figure 1. Models of influence on authentic and fake leadership on job performance. Source: own elaboration based on (Bieńkowska, Tworek 2024)

3. Employees' Dynamic Capabilities and their role in shaping job performance by fake and authentic leadership

Dynamic capabilities theory emphasizes the ability to sense, seize, and reconfigure resources in response to environmental change (Teece et al. 1997). While originally conceptualized at the organizational level, recent studies highlight the micro-foundations of dynamic capabilities at the employee level (Bocoya-Maline et al., 2024; Wang et al. 2024), linking them to adaptability, proactive behavior, and performance outcomes.

In essence, the EDCs are responsible for employees' ability to adapt to the environment (organizational or external), which is by its very nature in constant change today. The components of EDCs explicitly refer to abilities to be sensitive to changes in the environment, to adapt to changes in the environment, to proactively solve problems arising in the workplace (if they occur), and include innovations in the workplace, as well as the abilities for continuous personal

development and learning (Bieńkowska, Tworek 2020). They are therefore related to the job performance of the employees and, consequently, the organization's response to changes in the environment. They translate into both employee job performance (see: EDC-based model of job performance (Bieńkowska, Tworek 2020)) and organizational performance (see: Tworek et al. 2023).

At the same time, as indicated earlier, the leader is an element of the environment significantly affecting employees and the organization. The leader's impact on employees in fact has a similar mechanism, which is also used by the EDCs. Both influence job-related attitudes. In the case of EDCs, such impact occurs through a person-job fit (Bieńkowska, Tworek 2020), and in the case of leadership, such impact occurs through intraorganizational trust (Bieńkowska, Tworek 2024). Moreover, it should be noted that leadership is a factor external to the employees, while EDCs are internal characteristics of the employees. Therefore, it seems natural to assume that EDCs and leadership enter into a mutual relationship.

Despite extensive research on leadership and dynamic capabilities, limited attention has been paid to their interaction effects at the employee level, particularly in explaining how individual capabilities condition leadership influence mechanisms. Existing studies suggest that employee characteristics may moderate leadership effectiveness (Carmeli et al. 2013; Audenaert, Decramer 2018), but the specific role of Employees' Dynamic Capabilities (EDCs) remains underexplored.

3.1. Authentic leadership and EDCs

An authentic leader is intrinsically interested in harnessing the potential of employees and in achieving the goals of the organization as a whole through this potential. At the same time, the exploitation of the employee's potential implies, among other things, the delegation of autonomy to employees (Avolio et al. 2004; Walumbwa et al. 2008), which should be accompanied by commitment and responsibility for the entrusted goals within the limits set by the organization (Wang, Hsieh 2013; Frederrick et al. 2016; Corriveau 2020;). Authentic leadership, understood in this way, naturally influences the job performance of employees (cf. George 2003; Avolio et al. 2004; George et al. 2007; Bieńkowska, Tworek 2024).

EDCs are a response not only to the changing and dynamic environment, resulting in the fact that the nature of today's work and workplaces are also dynamic, but also to the need to redefine expectations of today's employees, who are the main organizational resource determining the possibility of sustainable development (Bieńkowska, Tworek, 2020). The literature rarely discusses the link between authentic leadership and dynamic capabilities. It is usually discussed in the context of the organization as a whole. Kleynhans and colleagues

(Kleynhans et al. 2021) stated that authentic leadership has the potential to positively influence the dynamic capabilities of the organization as a whole, and such a relationship is mediated by trust within the organization. Their research concerned the dynamic capabilities of the organization as a whole, not those of the employees. However, it confirms that authentic leadership shapes dynamic capabilities. The question concerning the dynamic capabilities of the employees, not those of the organization as a whole, is addressed by Semedo and colleagues (Semedo et al. 2016), who connected authentic leadership to the performance of the employees, showing that job resourcefulness (an element of EDCs) plays a crucial role in this relationship. It is also known that authentic leadership supports elements that are part of EDCs, such as proactive attitudes of employees (cf. e.g. Sumanth et al. 2023).

At the same time, it is important to state that employees with low EDCs will inherently be less able to proactively adapt to changes in the organization and its environment than employees with high EDCs. They will be less able to solve emerging problems, even if they perceive that they are occurring. Finally, they will be less sensitive to the need for continuous learning. Thus, it can be argued that, despite the capabilities, efforts, and endeavors of an authentic leader, they will not be able to respond to their behavior and be influenced in the way that those with high EDCs will: contributing to the increase of both intra-organizational trust, job-related attitudes and job performance. This situation is similar to that described by Hu and colleagues (Hu et al. 2018), where they showed that it is possible to identify some personal determinants that influence the effectiveness of authentic leadership on employees. They proved that there are "positive predictive effects of authentic leadership on proactive behaviour but also further revealed a mediating effect of psychological capital in the relationship between authentic leadership and proactive employee behaviour. It has also revealed moderating effects of compassion at work in the relationships between authentic leadership and psychological capital, and between psychological capital and proactive behaviour" (Hu et al. 2018, p. 6). In the case discussed in the study, such a factor is directly connected to EDCs. Hence, it may be concluded that employees with low EDCs are less responsive than employees with high EDCs, especially to factors that would have to cause a change in their attitudes and behavior toward work, even though, as noted earlier, authentic leadership naturally supports proactive behavior of employees (Yamak, Eyupoglu 2021; Sumanth et al. 2023).

Therefore, the following hypothesis can be formulated:

H1: The higher the EDCs, the stronger the authentic leadership and positive influence on job performance, mediated by intraorganizational trust, work engagement, organizational commitment, and work motivation.

The job performance model of authentic leadership based on trust and job-related attitudes moderated by EDCs is shown in Figure 2.

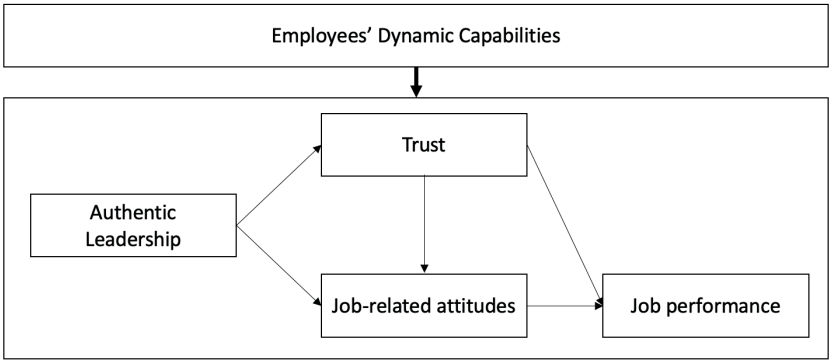


Figure 2. Job performance model of authentic leadership based on trust and job-related attitudes moderated by Employees' Dynamic Capabilities

Source: own work based on (Bieńkowska, Tworek 2024)

3.2. Fake leadership and EDCs

A fake leader effectively expects obedience from their employees and passive performance of tasks according to their expectations. In doing so, this leader restricts freedom and abstracts from responsibility in the sense of the autonomous action of people in organizations.

In this context, the attitudes of employees toward the negative behavior of the fake leader seem interesting. It seems to be important to consider the issue of employees' EDCs in analyses of these attitudes. It seems that employees with low EDCs, even if they disagree with the views and actions of the fake leader, will not be able to demonstrate proactive attitudes toward such behavior and attitudes, even if they perceive the necessity for such attitudes, which in itself is unlikely. In contrast, when employees with high EDCs do not agree with the views and behaviors of the fake leader, they lose trust in the leader (which seems to be key in the described relationship). Hence, they may inherently be able to adopt a proactive stance aimed at mitigating the negative impact of fake leadership on job performance, but it is not clear whether they will do so (this is influenced by other factors). However, it seems that it is not entirely clear which path will lead to such an effect. Indeed, two situations can be imagined.

The first situation is when a proactive stance may aim to proactively counteract the fake leader's actions, which are judged to be bad and detrimental

to employees and the organization, and subsequently to counteract the effects of the negative impact of the fake leadership on employees and the organization. Often, this means in practice that it is necessary to unequivocally oppose these actions (Bebbington 2014; Grabarczyk-Ponimasz 2021) and – even if only partially – take control of the situation. Knowing that the fake leader is using forbidden techniques, employees can start to defend themselves by restoring healthy (or healthier) relationships within the organization, and thus weakening the negative impact of the fake leader on both employees and the organization. Grabarczyk-Ponimasz (Grabarczyk-Ponimasz 2021) states that employees who are willing and able to do so are characterized by higher independence and social responsibility, supported by the ability to take risks and act proactively. Hence, it seems that employees with high EDCs will be more prone to such behaviors. Moreover, it should be noticed that EDCs equip employees with the ability to not resort to conformity (out of fear), leading to a vast decrease in job performance, but to retain some level of organizational trust and through teamwork maintain a less negatively impacted level of job performance. EDCs may enable employees to establish façades of conformity, allowing them to not be exposed to the entire set of fake leader negative behaviors, but simultaneously retaining some of their job performance (Liang 2022).

The second situation is when a proactive stance taken may imply bribery- or fear-driven conformism as a reaction to the perceived leader's behavior in the organization in the form of anti-employee and anti-organizational actions. This is because the fake leader teaches fear management (based on bribery and fear), which is consistent with obedience by authoritarianism emphasizing respect for authority and "relations of domination and subordination as the main relations existing in a hierarchical world" (Grabarczyk-Ponimasz 2021, p. 82). "This is accompanied by a tendency to use stereotypes, i.e. simplified ideas about those who are placed outside of one's own group" (Grabarczyk-Ponimasz 2021, p. 83). It seems that it is possible here to adapt, or even to adopt an ally attitude, in the face of such extreme conditions, perhaps preceded by a distorted rationalization of the situation. Such an attitude is also more achievable for employees with high EDCs. Perhaps in the short term this will mitigate the negative impact of the fake leadership on the job performance of these employees. However, the positive impact of such an option on the organization is questionable. Especially employees outside the so-called allied group will still decrease the performance parameters of the organization as a whole.

Therefore, the following hypothesis should be formulated:

H2: The higher the EDCs, the weaker the fake leadership's negative influence on job performance, mediated by intraorganizational trust, work engagement, organizational commitment, and work motivation.

The job performance model of fake leadership, based on trust and job-related attitudes, moderated by Employees' Dynamic Capabilities, is presented in Figure 3.

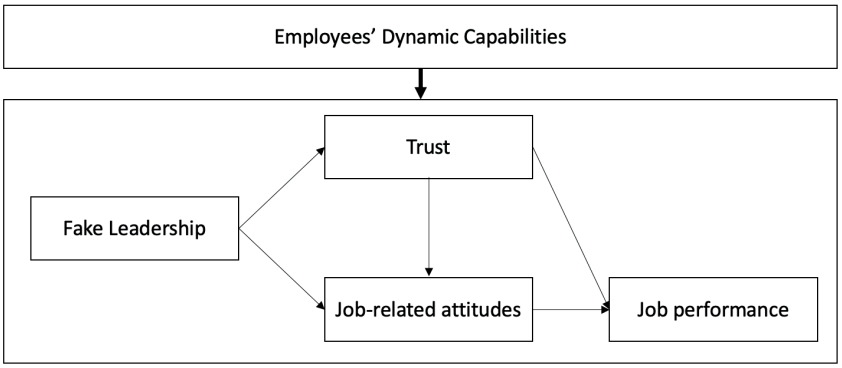


Figure 3. Job performance model of fake leadership based on trust and job-related attitudes moderated by Employees' Dynamic Capabilities

Source: own work based on (Bieńkowska, Tworek 2024)

4. Research methodology

The proposed theoretical model was verified based on empirical research. Two stages of empirical research were used for the analysis and verification of the hypotheses concerning fake leadership and authentic leadership (and its role in shaping job performance among organizations with high and low levels of EDCs among employees).

The first stage was conducted as a pilot study to validate the measurement tool, followed by the second stage, which comprised the main research. The empirical study employed a questionnaire as the primary research instrument. This questionnaire included a developed measurement scale (questions) for authentic and fake leadership variables, along with previously validated, less extensive measurement scales (questions) for each variable included in the models, as well as various control variables. Top and mid-level managers from organizations operating in Poland, who possess a comprehensive understanding of their organizations, completed the questionnaire, with only one questionnaire per organization being submitted.

The first stage focused on verifying the quality of the questionnaire as a research tool, encompassing all proposed measurement scales (questions). It was

completed by 25 carefully selected, competent judges who were top-level managers. Their feedback allowed for revisions of several questions to ensure they would be clearly understood by respondents. The second stage, which constituted the main research, aimed to analyze and verify the proposed hypotheses and was conducted in the first quarter of 2023.

4.1. Sample overview

The empirical research was carried out using the CAWI (Computer-Assisted Web Interview) method. The sample was constructed using a purposive sampling approach targeting organizations operating in Poland across multiple sectors. Respondents were recruited via professional research panels to ensure access to individuals holding managerial roles with sufficient organizational insight. Each organization was represented by a single respondent (top or mid-level manager), which ensured independence of observations and avoided clustering effects. No organization contributed more than one completed questionnaire. A total of 289 organizations participated in the study. Table 1 provides a detailed description of the sample, indicating that geographical location was the only limiting factor. The study concerned organizations, not individual respondents; therefore, the sample description is given accordingly. Although the sample selection was not representative, the diversity of the organizations and their varied industries allowed meaningful conclusions to be drawn. Additionally, the KMO (Kaiser-Meyer-Olkin) index was calculated for each variable and model to confirm that the sample size and diversity were sufficient for the intended statistical analyses.

Table 1

Sample overview

		Size of employment			
		fewer than 10	10–50 people	50–250 people	over 250 people
Primary source of revenue	production	11	16	26	17
	trade	8	28	23	13
	services	9	28	51	35
	education	0	5	12	3
Total		28	77	112	68

Source: (Bieńkowska, 2024)

For the purpose of the study, the sample was divided into two groups: organizations characterized by low levels of EDCs (74 organizations) and high levels of EDCs (214 organizations).

4.2. Description of variables

To verify the proposed theoretical hypotheses, the following variables were employed: Fake Leadership, Authentic Leadership, Intraorganizational Trust, Organizational Commitment, Work Engagement, Work Motivation, Job Performance, and Employees' Dynamic Capabilities. The variables were assessed using a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree," with a midpoint of "Neither Agree nor Disagree." The questionnaire was developed by the authors based on established theoretical frameworks and previously validated scales. Measurement items for key constructs were adapted from existing literature where available (e.g., Walumbwa et al., 2008, for authentic leadership), while the scale for fake leadership was developed and validated in prior research (Bieńkowska, Tworek 2024). All items were reviewed and refined during the pilot study.

Fake Leadership (FL) encompassed three latent dimensions – self-orientation, employee orientation, and organizational orientation – measured with 14, 22, and 12 items, respectively.

Authentic Leadership (AL) was evaluated using a similar 5-point Likert scale, with 13 items focused on the traits and behaviors characteristic of a true leader.

Intraorganizational Trust (Trust) measured employees' trust toward management and the organization, including perceptions of fair treatment, using three items on a 5-point Likert scale.

Organizational Commitment (OrgCom) assessed employees' happiness and sense of belonging within the organization through three items on a 5-point Likert scale.

Work Engagement (WrkEng) measured employees' attitudes toward their job, including enthusiasm and job immersion, using three items on a 5-point Likert scale.

Work Motivation (Motiv) evaluated employees' willingness and readiness to perform tasks, including the effort put into them, using three items on a 5-point Likert scale.

Job Performance (JobPer) assessed task proficiency, meticulousness, and work discipline with four items on a 5-point Likert scale.

Employees' Dynamic Capabilities (EDC) measured the approach to the need for adaptability among employees and proactive behaviors on a 5-point Likert scale.

Each of these variables was carefully measured to ensure the reliability and validity of the data collected, facilitating a robust analysis of the proposed hypotheses.

4.3. Measurement scales verification

To conduct the statistical analysis for verifying the proposed hypotheses, all measurement scales used in the study were analyzed and validated (the detailed process was published in book (Bieńkowska, Tworek 2024)). The initial step involved confirming that the collected data exhibited a normal distribution. The subsequent step focused on analyzing the measurement scales through three statistical tests. A Cronbach's Alpha analysis was used to verify the coherence and reliability of the measurement scale. A Cronbach's α value above 0.7 indicates average reliability, while values above 0.8 are considered highly satisfactory (Drost 2011), which was the case for all variables within this study. This analysis was performed using IBM SPSS. A Confirmatory Factor Analysis (CFA) assessed the coherence and internal consistency of the measurement scale. The Average Variance Extracted (AVE) should exceed 0.5, indicating that the latent variable explains a sufficient percentage of the variance. Model fit statistics should fall within these limits: $c2/df < 5.000$; $TLI > 0.800$; $CFI > 0.800$; $GFI > 0.800$; $RMSEA < 0.2$ (Hopwood and Donnellan, 2010), which was the case in all models obtained for measurement scales in this study. The CFA was also conducted using IBM SPSS. The Kaiser-Meyer-Olkin (KMO) test evaluated the adequacy of the sample size for the analysis. KMO values above 0.5 are deemed sufficient for further analysis (Kaiser 1974, 2000), which was the case for all variables within this study. Additionally, as a third step, discriminant validity was tested to ensure that all latent variables representing different theoretical concepts are statistically distinct. The results of the HTMT (Heterotrait-Monotrait) ratio tests yielded values below 0.65 (Ab Hamid et al. 2017), indicating that the chosen variables are appropriate for further analyses, including correlation, regression, and path analysis. The results of these tests are presented in Table 2 for all variables included in the study. The findings confirm that the selected measurement scales are internally consistent, reliable, and coherent, making them suitable for further analysis.

Table 2

Scales reliability and internal consistency analysis

Variable name	Variable name	Number of items	AVE	KMO	Cron-bach's Alpha	N
Fake Leadership	FL	3 (14, 22, 12)	0.951	0.757	0.974	280
Authentic Leadership	AL	13	0.716	0.968	0.967	281
Intraorganizational Trust	Trust	3	0.742	0.721	0.823	287
Organizational Commitment	Org-Com	3	0.570	0.524	0.706	289
Work Engagement	WrkEng	3	0.663	0.689	0.744	289
Work Motivation	Motiv	3	0.753	0.697	0.836	289
Job Performance	JobPer	4	0.638	0.672	0.806	289
Employees' Dynamic Capabilities	EDC	5	0.734	0.698	0.888	289

Source: based on own work and (Bieńkowska, Tworek 2024)

4.4. Multigroup path analysis results

After the first step concerning scales analysis, the main statistical reasoning based on a multigroup path analysis was performed using the IBM SPSS Amos software. Two sets of multigroup path analysis models were developed: one for authentic leadership and one for fake leadership.

The statistically significant and well-fitted models were obtained through multigroup path analysis, and for both of them, baseline comparison showed that the unconstrained models' characteristics were within the margins for them to undergo statistical reasoning. Based on them, the full assessment of the model was performed. The fit of the model was assessed with a Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) (sufficient values above 0.8) and RMSEA (sufficient values below 0.2). The model obtained for authentic leadership was statistically significant and well-fitted: Chi^2 (6) = 36.303; p = 0.001; CFI = 0.981; TLI = 0.886; RMSEA = 0.091 (PCLOSE = 0.001). The model obtained for fake leadership was also statistically significant and well-fitted: Chi^2 (6) = 20.833; p = 0.001; CFI = 0.989; TLI = 0.925; RMSEA = 0.093 (PCLOSE = 0.001). In both models, the sufficient RMSEA values were obtained based on structural weights, intercepts,

and residuals, showing a very good fit in the model (calculated when CMIN has a chi-square distribution under the assumption that the fitted model is correct). Moreover, sufficient values of CFI and TLI were also obtained for both models, establishing that goodness of fit is sufficient to form conclusions based on the obtained model.

4.4.1. Model for authentic leadership

The overview of the authentic leadership model obtained for organizations with high values of EDCs is presented in Table 3, and for organizations with low values of EDCs is presented in Table 4. Total and indirect effects occurring in the model are presented in tables 5–8.

Table 3

Tegression estimates for authentic leadership model – high levels of EDCs

			Estimate	S.E.	C.R.	P
Trust	←	AL	0.731	0.041	17.825	<0.001
WrkEng	←	Trust	0.502	0.054	9.332	<0.001
WrkEng	←	AL	0.343	0.051	6.754	<0.001
OrgCom	←	AL	0.280	0.047	5.914	<0.001
OrgCom	←	Trust	0.344	0.054	6.380	<0.001
OrgCom	←	WrkEng	0.147	0.058	2.524	0.012
Motiv	←	WrkEng	0.342	0.062	5.497	<0.001
Motiv	←	OrgCom	0.228	0.073	3.138	0.002
Motiv	←	AL	0.316	0.054	5.861	<0.001
Motiv	←	Trust	0.174	0.062	2.809	0.005
JobPer	←	Motiv	0.416	0.058	7.146	<0.001
JobPer	←	OrgCom	0.406	0.073	5.599	<0.001

Table 4

Regression estimates for authentic leadership model – low levels of EDCs

			Estimate	S.E.	C.R.	P
Trust	←	AL	0.430	0.108	3.997	<0.001
WrkEng	←	Trust	0.513	0.096	5.367	<0.001

Table 4 cont.

WrkEng	←	AL	0.214	0.097	2.203	0.028
OrgCom	←	AL	0.074	0.092	0.808	0.419
OrgCom	←	Trust	0.238	0.103	2.307	0.021
OrgCom	←	WrkEng	0.347	0.107	3.245	0.001
Motiv	←	WrkEng	0.276	0.102	2.700	0.007
Motiv	←	OrgCom	0.014	0.104	0.132	0.895
Motiv	←	AL	0.075	0.083	0.912	0.002
Motiv	←	Trust	0.794	0.095	8.335	<0.001
JobPer	←	Motiv	0.262	0.094	2.794	0.005
JobPer	←	OrgCom	0.372	0.127	2.921	0.003

Table 5

Total Effects for authentic leadership model – high levels of EDCs

	AL	Trust	WrkEng	OrgCom	Motiv
Trust	0.731	0.000	0.000	0.000	0.000
WrkEng	0.709	0.502	0.000	0.000	0.000
OrgCom	0.635	0.418	0.147	0.000	0.000
Motiv	0.830	0.441	0.376	0.228	0.000
JobPer	0.604	0.353	0.216	0.501	0.416

Table 6

Indirect effects for authentic leadership model – high levels of EDCs

	AL	Trust	WrkEng	OrgCom	Motiv
Trust	0.000	0.000	0.000	0.000	0.000
WrkEng	0.367	0.000	0.000	0.000	0.000
OrgCom	0.356	0.074	0.000	0.000	0.000
Motiv	0.515	0.267	0.034	0.000	0.000
JobPer	0.604	0.353	0.216	0.095	0.000

Table 7

Total Effects for authentic leadership model – low levels of EDCs

	AL	Trust	WrkEng	OrgCom	Motiv
Trust	0.430	0.000	0.000	0.000	0.000
WrkEng	0.435	0.513	0.000	0.000	0.000
OrgCom	0.328	0.416	0.347	0.000	0.000
Motiv	0.541	0.941	0.281	0.014	0.000
JobPer	0.263	0.401	0.202	0.375	0.262

Table 8

Indirect effects for authentic leadership model – low levels of EDCs

	AL	Trust	WrkEng	OrgCom	Motiv
Trust	0.000	0.000	0.000	0.000	0.000
WrkEng	0.221	0.000	0.000	0.000	0.000
OrgCom	0.253	0.178	0.000	0.000	0.000
Motiv	0.466	0.147	0.005	0.000	0.000
JobPer	0.263	0.401	0.202	0.004	0.000

Tables 3 and 4 show the relations within the models obtained for high and low levels of EDCs in an organization. It is important to underline two main differences. The first concerns the relations with statistical significance. In the case of low levels of EDCs, some relations within the model are not statistically significant. The influence of authentic leadership on organizational commitment is not statistically significant. However, the influence of authentic leadership on trust, work engagement, and work motivation remains statistically significant. In the case of high levels of EDCs, all relations within the model are statistically significant, showing that various paths of authentic leadership have a positive influence on job performance. The second concerns the strength of the relations. In the case of high levels of EDCs, the relations between authentic leadership and trust, work engagement, and work motivation is much stronger than in the case of low levels of EDCs. Also, as shown in tables 6 and 8, the indirect and total effects occurring in the model obtained for high levels of EDCs are much stronger than those obtained for low levels of EDCs. Taken together, this makes it reasonable to accept hypothesis H1, stating that the higher the EDCs, to stronger the positive influence of authentic leadership on job performance.

4.4.2. Model for fake leadership

The overview of the fake leadership model obtained for organizations with high values of EDCs is presented in Table 9, and for organizations with high values of EDCs is presented in Table 10. Total and indirect effects occurring in the model are presented in tables 11–14.

Table 9

Regression estimates for fake leadership model – high levels of EDCs

			Estimate	S.E.	C.R.	P
Trust	←	FL	−0.300	0.091	−3.291	<0.001
WrkEng	←	Trust	0.548	0.095	5.796	<0.001
WrkEng	←	FL	−0.128	0.079	−1.618	0.106
OrgCom	←	FL	−0.033	0.073	−0.457	0.648
OrgCom	←	Trust	0.246	0.103	2.388	0.017
OrgCom	←	WrkEng	0.360	0.106	3.397	<0.001
Motiv	←	WrkEng	0.284	0.102	2.786	0.005
Motiv	←	OrgCom	0.021	0.104	0.199	0.842
Motiv	←	FL	−0.040	0.065	−0.605	0.545
Motiv	←	Trust	0.800	0.096	8.369	<0.001
JobPer	←	Motiv	0.262	0.094	2.794	0.005
JobPer	←	OrgCom	0.372	0.127	2.921	0.003

Table 10

Regression estimates for fake leadership model – low levels of EDCs

			Estimate	S.E.	C.R.	P
Trust	←	FL	−0.567	0.058	−9.695	<0.001
WrkEng	←	Trust	0.746	0.045	16.435	<0.001
WrkEng	←	FL	−0.067	0.047	−1.421	<0.001
OrgCom	←	FL	−0.079	0.039	−2.031	0.042
OrgCom	←	Trust	0.425	0.056	7.546	<0.001
OrgCom	←	WrkEng	0.282	0.057	4.970	<0.001
Motiv	←	WrkEng	0.446	0.060	7.411	<0.001
Motiv	←	OrgCom	0.338	0.069	4.887	<0.001
Motiv	←	FL	−0.196	0.039	−5.011	<0.001

Table 10 cont.

			Estimate	S.E.	C.R.	P
Motiv	←	Trust	0.167	0.064	2.620	0.009
JobPer	←	Motiv	0.417	0.058	7.165	<0.001
JobPer	←	OrgCom	0.405	0.072	5.593	<0.001

Table 11

Total effects for fake leadership model – high levels of EDCs

	FL	Trust	WrkEng	OrgCom	Motiv
Trust	−0.300	0.000	0.000	0.000	0.000
WrkEng	−0.292	0.548	0.000	0.000	0.000
OrgCom	−0.212	0.444	0.360	0.000	0.000
Motiv	−0.367	0.965	0.291	0.021	0.000
JobPer	−0.175	0.417	0.210	0.377	0.262

Table 12

Indirect effects for fake leadership model – high levels of EDCs

	FL	Trust	WrkEng	OrgCom	Motiv
Trust	0.000	0.000	0.000	0.000	0.000
WrkEng	−0.164	0.000	0.000	0.000	0.000
OrgCom	−0.179	0.197	0.000	0.000	0.000
Motiv	−0.327	0.165	0.007	0.000	0.000
JobPer	−0.175	0.417	0.210	0.005	0.000

Table 13

Total effects for authentic fake model – low levels of EDCs

	FL	Trust	WrkEng	OrgCom	Motiv
Trust	−0.567	0.000	0.000	0.000	0.000
WrkEng	−0.490	0.746	0.000	0.000	0.000
OrgCom	−0.457	0.635	0.282	0.000	0.000
Motiv	−0.664	0.715	0.541	0.338	0.000
JobPer	−0.462	0.556	0.340	0.547	0.417

Table 14

Indirect effects for fake leadership model – low levels of EDCs

	FL	Trust	WrkEng	OrgCom	Motiv
Trust	0.000	0.000	0.000	0.000	0.000
WrkEng	−0.423	0.000	0.000	0.000	0.000
OrgCom	−0.379	0.210	0.000	0.000	0.000
Motiv	−0.468	0.548	0.095	0.000	0.000
JobPer	−0.462	0.556	0.340	0.141	0.000

Tables 9 and 10 show the relations within the models obtained for high and low levels of EDCs in an organization. It is again important to underline two main differences. The first concerns the relations with statistical significance. In case of high levels of EDCs, not all relations within the model are statistically significant. The influence of fake leadership on work engagement, organizational commitment, and work motivation is not statistically significant. Only the influence of fake leadership on trust remains statistically significant. In the case of low levels of EDCs, all relations within the model are statistically significant, showing that various paths of fake leadership have a negative influence on job performance. The second concerns the strength of the relations. In the case of high levels of EDCs, the only remaining relation – between fake leadership and trust – is much weaker (−0.300) than in the case of low levels of EDCs (−0.576). Also, as shown in tables 12 and 14, the indirect and total effects occurring in the model obtained for high levels of EDCs are much weaker than those obtained for low levels of EDCs. Hence, it also becomes reasonable to accept hypothesis H2, stating that the higher the EDCs, to weaker the negative influence of fake leadership on job performance.

5. Discussion

The empirical research results on the role of Employees' Dynamic Capabilities (EDC) in shaping job performance under different leadership styles, specifically fake and authentic leadership, have significant implications for organizational behavior and human resource management. These findings align closely with the existing literature, particularly the theories on dynamic capabilities and leadership, contributing to their development, especially in the context of three theories: situational leadership by Hersey and Blanchard (Hersey, Blanchard, 1977), contingency theory by Fiedler (Fiedler 1967) and path-goal theory by House and Mitchell (House, Mitchell 1974).

Dynamic capabilities, as discussed by Teece, Pisano, and Shuen (Teece et al. 1997), emphasize the importance of not only possessing resources but also the ability to adapt and reconfigure them in response to environmental changes. This concept is echoed in the findings of Bieńkowska and Tworek (Bieńkowska, Tworek 2020), which confirm that EDCs significantly influence job performance through mechanisms such as Person-Job fit (P-J fit), work motivation, job satisfaction, and work engagement. The results obtained from this study further develop the theory concerning the significance of dynamic capabilities in the context of leadership, highlighting the adaptive nature of their role in shaping job performance through leadership.

Authentic leadership, characterized by a focus on genuine leader-follower relationships and ethical behavior, has been shown to positively influence job performance, and the obtained results confirm this, particularly for organizations with high levels of EDCs. The empirical research indicates that authentic leadership enhances intraorganizational trust, work engagement, organizational commitment, and work motivation, all of which contribute to improved job performance. This contributes to the theoretical framework (e.g. Bass, Steidlmeier 1999; Gardner et al. 2011), confirming that authentic leaders, by fostering trust and engagement, can effectively leverage the dynamic capabilities of their employees to achieve higher performance levels. Moreover, it seems that even with lower levels of EDCs, the positive influence of authentic leadership on job performance remains in place, showing that this particular leadership style is aiming at true support for employees, regardless of their level of capabilities.

Conversely, fake leadership, which is detrimental and characterized by self-serving behaviors and manipulation, negatively impacts job performance (Bieńkowska, Tworek, 2024). However, the empirical research vastly contributes to that theoretical framework, showing the importance of EDCs in that process. The findings as a whole are consistent with the literature that suggests negative leadership styles undermine employee morale and productivity (Northouse 2021; Alimo-Metcalfe 2013), and even some groups of employees who support a fake leader are unable to significantly improve the overall job performance, which is more meaningfully negatively influenced by this leadership style. However, a critical insight and contribution from the study concerns the moderating role of EDCs in the relation between fake leadership and job performance. It shows that high levels of EDCs can mitigate the adverse effects of fake leadership by enabling employees to adapt and respond proactively to negative behaviors, as employees are better equipped to not resort to conformity. EDCs also mitigate the decrease in intra-organizational trust, making it possible to maintain some level of job performance. The study also reveals that the negative effects of fake

leadership are more pronounced in organizations with low levels of EDCs. In these environments, employees are less equipped to counteract the negative influences of a fake leader, resulting in reduced trust, engagement, and overall performance.

The results align with the hypothesis that dynamic capabilities provide a buffer against unfavorable leadership, as employees with high EDCs can employ strategies to protect their performance even under poor leadership, which constitutes a topic for further studies.

An interesting aspect of future research seems to relate to the mechanism of the high EDCs mitigating effect on the negative impact of fake leadership on job performance. It seems necessary to conduct further research and, for example, to include the person-supervisor fit (P-S fit) as a component of the employee's fit with the work environment (whereby "P-E fit is broadly defined as the compatibility between an individual and a work environment that occurs when their characteristics are well matched" (Kristof-Brown 2005, p. 281)). P-S fit refers to the match between employees' and supervisors' characteristics. It is assumed that it is determined by the match of values, personality, lifestyle, and work style (Chuang, Shen 2007; Vianen van et al. 2011, p. 914). Thus, if the P-S fit is high, i.e. employees share the supervisor's values, personality, lifestyle, and work style, then they will naturally accept and even imitate the supervisor's attitudes and behaviors, which will be facilitated by high EDCs, but also not hindered by low EDCs. However, with a low P-S fit and high EDCs, employees will actively oppose the leader's actions both toward the employees and toward the organization or, despite their own different views, in a conformist manner and/or due to fear of the supervisor's actions, they will conform to the supervisor's activities in an attempt to improve work outcome parameters, including job performance. Identifying a valid option requires in-depth research into the level of conformism among employees in the organization. As a whole, it provides directions for possible future research work.

6. Conclusions

This study addressed an identified research gap in the organizational behavior field of study by exploring the role of EDCs in shaping the relation between leadership styles and job performance. Previous research has extensively discussed the individual impacts of dynamic capabilities on leadership and job performance. However, this study uniquely integrates these concepts, including the specific role of EDCs, providing empirical evidence on how EDCs can influence the mechanism behind authentic and fake leadership's influence on job

performance. Hence, the aim of the study was to examine the role of EDCs in shaping job performance by the negative leadership concept using the example of fake leadership, as well as in shaping job performance by the positive leadership concept using the example of authentic leadership. The aim was successfully fulfilled, contributing various conclusions to the current body of knowledge in the examined field of research.

The obtained results showed that authentic leadership has a universally positive impact on job performance; however, still being significantly enhanced by high levels of EDCs. This leadership style fosters trust, engagement, organizational commitment, and motivation among employees, leading to improved job performance, and EDCs strengthen such positive impact, showing that authentic leadership leverages Employees' Dynamic Capabilities.

The results also showed that fake leadership negatively affects job performance, regardless of the level of EDCs. However, high EDC levels mitigate these adverse effects by enabling employees to adapt and respond proactively to negative leadership behaviors, limiting the negative impact of fake leadership on work motivation, work engagement, and organizational commitment. It shows that employees are better equipped to not only mitigate the decrease of intra-organizational trust but also to limit their resort to conformity and maintain some level of job performance. This shows the protective role of EDCs in buffering against unfavorable leadership styles, which makes a significant contribution to the contingency theory and path-goal theory of leadership.

This study contributes to the literature in three ways. First, it extends dynamic capabilities theory to the micro (employee) level by demonstrating its moderating role. Second, it integrates positive and negative leadership frameworks within a single analytical model. Third, it provides empirical evidence that employee capabilities can act as both amplifiers and buffers in leadership-performance mechanisms.

The study performed has some limitations. The research was conducted among organizations in Poland, which may limit the generalizability of the findings to other cultural and organizational contexts. Future studies should consider a more diverse geographical scope to validate these results globally. The study's cross-sectional design captures a snapshot in time, which may not fully account for the dynamic nature of EDCs and leadership interactions over periods of time. Longitudinal studies should be considered in the future to observe these relations over time. The reliance on self-reported data from managers may introduce bias, as responses could be influenced by personal perceptions or organizational pressures. This results in the need to incorporate multi-source data, including subordinate feedback and objective performance metrics, in future research, which would enhance the robustness of the findings.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT 4o (version June 2024) to improve the flow of the article (corrections in English language). After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

1. Ab Hamid, M.R., Sami, W. and Sidek, M.M. (2017). 'Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion', *Journal of Physics: Conference Series*, vol. 890, no. 1, 012163, 10.1088/1742-6596/890/1/012163.
2. Al Wali, J., Muthuveloo, R., Ping, T.A. and Bataineh, M. (2020) 'An overview of employees' dynamic capabilities, job performance with innovative work behavior: A conceptual paper', *Global Journal of Management and Business Research: A Administration and Management*, 14 Jan 2020, [Online], Available: https://journalofbusiness.org/index.php/GJMBR/article/view/102645/5-An-Overview-of-Employees_JATS_NLM_xml [13 Apr 2026].
3. Al Wali, J., Muthuveloo, R., Teoh, A.P. and Al Wali, W. (2023) 'Disentangling the relationship between employees' dynamic capabilities, innovative work behavior and job performance in public hospitals', *International Journal of Innovation Science*, vol. 15(2), pp. 368–384, https://doi.org/10.1108/IJIS-01-2022-0012.
4. Alimo-Metcalfe, B. (2013) 'A critical review of leadership theory', in Skipton Leonard, H., Lewis, R., Freedman, A.M., Passmore, J. (ed.) The Wiley-Blackwell Handbook of the Psychology of Leadership, Change, and Organizational Development, pp. 13–47. Chichester: Wiley-Blackwell.
5. Amabile, T.M., Schatzel, E.A., Moneta, G.B. and Kramer, S.J. (2004) 'Leader behaviors and the work environment for creativity: Perceived leader support', *The Leadership Quarterly*, vol. 15(1), pp. 5–32, https://doi.org/10.1016/j.leaqua.2003.12.003.
6. Audenaert, M. and Decramer, A. (2018) 'When empowering leadership fosters creative performance: The role of problem-solving demands and creative personality', *Journal of Management & Organization*, vol. 24(1), pp. 4–18, https://psycnet.apa.org/doi/10.1017/jmo.2016.20.
7. Avolio, B.J., Gardner, W.L., Walumbwa, F.O., Luthans, F. and May, D.R. (2004) ' Unlocking the mask: A look at the process by which authentic leaders impact follower attitudes and behaviors', *The Leadership Quarterly*, vol. 15(6), pp. 801–823, https://doi.org/10.1016/j.leaqua.2004.09.003.

8. Bass, B.M., and Steidlmeier, P. (1999). 'Ethics, character, and authentic transformational leadership behavior'. *The leadership quarterly*, vol. 10(2), pp. 181–217, https://doi.org/10.1016/S1048-9843(99)00016-8.
9. Bebbington, D.W. (2014) The nonconformist conscience, London: Routledge.
10. Bieńkowska, A. and Tworek, K. (2020) 'Job performance model based on Employees' Dynamic Capabilities (EDC)', *Sustainability*, vol. 12(6), 2250, https://doi.org/10.3390/su12062250.
11. Bieńkowska, A. and Tworek, K. (2024) Leadership styles and job performance: The impact of fake leadership on organizational reliability, New York: Routledge
12. Bieńkowska, A., Koszela, A. and Tworek, K. (2021) 'Verification of the job performance model based on Employees' Dynamic Capabilities in organisations under the COVID-19 pandemic crisis', *Engineering Management in Production and Services*, vol. 13(3), pp. 66–85, https://doi.org/10.2478/emj-2021-0022.
13. Bieńkowska, A., Tworek, K. and Zabłocka-Kluczka, A. (2020) Organizational reliability: Human resources, information technology and management, New York: Routledge.
14. Bocoya-Maline, J., Calvo-Mora, A. and Rey Moreno, M. (2024) 'Predictive and mediation model for decision-making in the context of dynamic capabilities and knowledge management', *Management Decision*, vol. 62(7), pp. 2164–2188, https://doi.org/10.1108/MD-06-2023-0956.
15. Boddy, C.R. (2017) 'Psychopathic leadership a case study of a corporate psychopath CEO', *Journal of Business Ethics*, vol. 145(1), pp. 141–156, https://doi.org/10.1007/s10551-015-2908-6.
16. Burns Jr, W.A. (2017) 'A descriptive literature review of harmful leadership styles: Definitions, commonalities, measurements, negative impacts, and ways to improve these harmful leadership styles', *Creighton Journal of Interdisciplinary Leadership*, vol. 3(1), pp. 33–52.
17. Cameron, K.S., Dutton, J.E., and Quinn, R.E. (2003) 'Foundations of positive organizational scholarship', in Cameron, K.S., Dutton, J.E., and Quinn, R.E. (ed.) Positive organizational scholarship: Foundations of a new discipline, San Francisco: Berrett-Koehler Publishers Inc.
18. Carmeli, A., Gelbard, R. and Reiter-Palmon, R. (2013) 'Leadership, creative problem-solving capacity, and creative performance: The importance of knowledge sharing', *Human Resource Management*, vol. 52(1), pp. 95–121, https://doi.org/10.1002/hrm.21514.
19. Chuang, A., Shen, C.T. and Judge, T.A. (2016) 'Development of a multidimensional instrument of person–environment fit: The Perceived Person–Environment Fit Scale (PPEFS)', *Applied psychology*, vol. 65(1), pp. 66-98. https://doi.org/10.1111/apps.12036.

20. Corriveau, A.M. (2020) 'Developing authentic leadership as a starting point to responsible management: A Canadian university case study', *The International Journal of Management Education*, vol. 18(1), 100364, https://doi.org/10.1016/j.ijme.2020.100364.
21. Drost, E.A. (2011) 'Validity and reliability in social science research', *Education Research and Perspectives*, vol. 38(1), pp. 105–123, [Online], Available: https://www.brown.uk.com/teaching/HEST5001/drost.pdf [13 Apr 2026].
22. Einarsen, S., Aasland, M.S. and Skogstad, A. (2007) 'Destructive leadership behaviour: A definition and conceptual model', *The Leadership Quarterly*, vol. 18(3), pp. 207–216, https://doi.org/10.1016/j.leaqua.2007.03.002.
23. Fiedler, F.E. (1967) A theory of leadership effectiveness, New York: McGraw-Hill.
24. Frederick, H.R., Wood Jr, J.A.A., West, G.R.B. and Winston, B.E. (2016) 'The effect of the accountability variables of responsibility, openness, and answerability on authentic leadership', *Journal of Research on Christian Education*, vol. 25(3), pp. 302–316, https://doi.org/10.1080/10656219.2016.1237907.
25. Freiherr von Fircks, E. (2024) 'Culture and leadership: A Lewinian perspective of organizational problem solving', *Human Arenas*, vol. 7(1), pp. 1–14, https://doi.org/10.1007/s42087-021-00256-5.
26. Gardner, W.L., Cogliser, C.C., Davis, K.M. and Dickens, M.P. (2011), 'Authentic leadership: A review of literature and research agenda', *The Leadership Quarterly*, vol. 22, pp. 1120-1145, https://doi.org/10.1016/j.leaqua.2011.09.007.
27. George, B. (2003) Authentic leadership: Rediscovering the secrets to creating lasting value, San Francisco: JosseyBass.
28. George, B., Sims, P., McLean, A.N. and Mayer, D. (2007) 'Discovering your authentic leadership', *Harvard Business Review*, vol. 85(2), pp. 129–138, [Online], Available: https://hbr.org/2007/02/discovering-your-authentic-leadership [13 Apr 2026].
29. Grabarczyk-Ponimasz, E. (2021) Psychologiczne mechanizmy bierności i posłuszeństwa, [Online], Available: http://www.oipc.pl/pliki/Opowiedzialnosc biernych_Ponimasz.pdf [13 Apr 2026].
30. Hersey, P. and Blanchard, K.H. (1977) Management of organizational behavior: Utilizing human resources, 3rd edition, New Jersey: Prentice-Hall.
31. Horner, M. (1997) 'Leadership theory: past, present and future', *Team Performance Management: An International Journal*, vol. 3(4), pp. 270–287.
32. House, R.J. and Mitchell, R.R. (1974) 'Path-goal theory of leadership', *Journal of Contemporary Business*, vol. 3(4), pp. 81–97.
33. Hu, Y., Wu, X., Zong, Z., Xiao, Y., Maguire, P., Qu, F., Wei, J. and Wang, D. (2018) 'Authentic leadership and proactive behavior: the role of psychological capital and compassion at work', *Frontiers in Psychology*, vol. 9, 2470, https://psycnet.apa.org/doi/10.3389/fpsyg.2018.02470

34. Jia, S., Khassawneh, O., Mohammad, T. and Cao, Y. (2024) 'Knowledge-oriented leadership and project employee performance: the roles of organisational learning capabilities and absorptive capacity', *Current Psychology*, vol. 43(10), pp. 8825–8838, https://doi.org/10.1007/s12144-023-05024-y.
35. Kaiser, H.F. (1974), 'An index of factorial simplicity', *Psychometrika*, vol. 39, pp. 31–36, [Online], Available: https://link.springer.com/article/10.1007/BF02291575 [13 Apr 2026] .
36. Khalid, S., Malik, N. and Atta, M. (2024) 'Employee silence predicted by abusive leadership and workplace ostracism: Role of employee power distance', *International Journal of Educational Leadership and Management*, vol. 12(1), pp. 13–35, http://dx.doi.org/10.17583/ijelm.10373.
37. Kleynhans, D.J., Heyns, M.M. and Stander, M.W. (2021) 'Dynamic organisational capabilities: The role of authentic leadership and trust', *SA Journal of Industrial Psychology*, vol. 47(1), pp. 1–12, https://doi.org/10.4102/sajip.v47i0.1877.
38. Kristof-Brown, A.L., Zimmerman, R.D. and Johnson, E.C. (2005) 'Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit', *Personnel Psychology*, vol. 58(2), pp. 281–342, https://psycnet.apa.org/doi/10.1111/j.1744-6570.2005.00672.x.
39. Liang, H.L. (2022) 'Compulsory citizenship behavior and facades of conformity: a moderated mediation model of neuroticism and citizenship pressure', *Psychological Reports*, vol. 125(6), pp. 3141–3161, https://doi.org/10.1177/00332941211031794.
40. Liu, L., Liu, P., Yang, H., Yao, H. and Thien, L.M. (2024) 'The relationship between distributed leadership and teacher well-being: The mediating roles of organisational trust', *Educational Management Administration & Leadership*, vol. 52(4), pp. 837–853, https://doi.org/10.1177/17411432221113683.
41. Norman, S.M., Avolio, B.J. and Luthans, F. (2010) 'The impact of positivity and transparency on trust in leaders and their perceived effectiveness', *The Leadership Quarterly*, vol. 21(3), pp. 350–364, https://doi.org/10.1016/j.leaqua.2010.03.002.
42. Northouse, P.G. (2021) Leadership: Theory and Practice, 9th edition, London: Sage.
43. Omilion-Hodges, L.M. and Ptacek, J.K. (2021) Leader-member exchange and organizational communication: Facilitating a healthy work environment, Springer International Publishing.
44. Palmen, D.G., Kolthoff, E.W. and Derksen, J.J. (2021) 'The need for domination in psychopathic leadership: A clarification for the estimated high prevalence of psychopathic leaders', *Aggression and Violent Behavior*, vol. 61, 101650, https://doi.org/10.1016/j.avb.2021.101650.

45. Pelletier, K.L. (2010) 'Leader toxicity: An empirical investigation of toxic behavior and rhetoric', *Leadership*, vol. 6(4), pp. 373–389, https://psycnet.apa.org/doi/10.1177/1742715010379308.
46. Saridakis, G., Lai, Y. and Cooper, C.L. (2017) 'Exploring the relationship between HRM and firm performance: A meta-analysis of longitudinal studies', *Human Resource Management Review*, vol. 27(1), pp. 87–96, https://doi.org/10.1016/j.hrmr.2016.09.005.
47. Semedo, A.S.D., Coelho, A.F.M. and Ribeiro, N.M.P. (2016) 'Effects of authentic leadership, affective commitment and job resourcefulness on employees' creativity and individual performance', *Leadership & Organization Development Journal*, vol. 37(8), pp. 1038–1055, https://doi.org/10.1108/LODJ-02-2015-0029.
48. Starratt, A. and Grandy, G. (2010) 'Young workers' experiences of abusive leadership', *Leadership & Organization Development Journal*, vol. 31(2), pp. 136–158, https://psycnet.apa.org/doi/10.1108/01437731011024394.
49. Sumanth, J.J., Černe, M., Hannah, S.T. and Škerlavaj, M. (2023) 'Fueling the creative spark: How authentic leadership and LMX foster employees' proactive orientation and creativity', *Journal of Leadership & Organizational Studies*, vol. 30(3), pp. 356–374, https://doi.org/10.1177/15480518231180064.
50. Syed, Z.J.W. and Jamal, W. (2012) 'Universalistic perspective of HRM and organizational performance: meta-analytical study', *International Bulletin of Business Administration*, vol. 13(4), pp. 47–57.
51. Teece, D.J., Pisano, G. and Shuen, A. (1997) 'Dynamic capabilities and strategic management', *Strategic Management Journal*, vol. 18(7), pp. 509–533.
52. Tworek, K., Bieńkowska, A., Hawrysz, L. and Maj, J. (2023) 'The model of organizational performance based on employees' dynamic capabilities–verification during crisis caused by Black Swan event', *IEEE Access*, vol. 11, pp. 45039–45055, https://doi.org/10.1109/ACCESS.2023.3273608.
53. Vianen van, A.E.M., Shen, C.T. and Chuang, A. (2011) 'Person–organization and person–supervisor fits: Employee commitments in a Chinese context', *Journal of Organizational Behavior*, vol. 32(6), pp. 906–926, https://psycnet.apa.org/doi/10.1002/job.726.
54. Walumbwa, F.O., Avolio, B.J., Gardner, W.L., Wernsing, T.S. and Peterson, S.J. (2008) Authentic leadership: Development and validation of a theory-based measure, *Journal of Management*, vol. 34(1), pp. 89–126, https://doi.org/10.1177/0149206307308913.
55. Wang, D.S. and Hsieh, C.C. (2013) 'The effect of authentic leadership on employee trust and employee engagement', *Social Behavior and Personality: An International Journal*, vol. 41(4), pp. 613–624, https://psycnet.apa.org/doi/10.2224/sbp.2013.41.4.613.

56. Wang, G., Mansor, Z.D. and Leong, Y.C. (2024) 'Unlocking digital performance: exploring the mediating role of employee competitive attitudes, behaviors, and dynamic capabilities in Chinese SMEs under high-involvement human resource management practice', *Journal of Innovation and Entrepreneurship*, vol. 13(1), 37, https://doi.org/10.1186/s13731-024-00395-3.
57. Yamak, O.U. and Eyupoglu, S.Z. (2021) 'Authentic leadership and service innovative behavior: mediating role of proactive personality', *SAGE Open*, vol. 11(1), https://doi.org/10.1177/2158244021989629.

Summary

This article examines the role of Employees' Dynamic Capabilities (EDC) in shaping job performance through the influence of authentic and fake leadership styles. It addresses a critical gap by exploring how EDCs moderate the relations between leadership styles and job performance, mediated by intraorganizational trust, work engagement, work motivation, and organizational commitment. Data obtained from questionnaires from 289 organizations in Poland were analyzed using multigroup path analysis. The findings show that authentic leadership positively influences job performance by enhancing trust, engagement, commitment, and motivation, and the strength of this influence is higher among organizations with higher levels of EDCs. Conversely, fake leadership negatively impacts job performance, with more pronounced effects in low EDC environments. Most importantly, high EDC levels can buffer against the negative effects of fake leadership. The study highlights the importance of fostering EDCs to leverage authentic leadership benefits and mitigate fake leadership harms.

JEL codes: M12, O15

Keywords: *fake leadership, authentic leadership, Employees' Dynamic Capabilities, job performance, management*

Managerial Economics
2026, Vol. 27, No. 1, pp. 37–52
https://doi.org/10.7494/manage.2026.27.1.37

Dubravko Sabolić*

On the structural design of the EU Inter-TSO Compensation mechanism

1. Introduction

Cross-border electricity flows in the European power system generate network externalities that are not fully internalized by national transmission tariffs or market prices. Power flows follow physical laws rather than commercial schedules, and in meshed transmission networks, they frequently traverse control areas that are neither the origin nor the destination of the underlying transaction. In the European Union, where system operation remains institutionally fragmented across several dozen transmission system operators (TSOs), these externalities have been addressed through the Inter-TSO Compensation (ITC) mechanism.

The ITC system provides a multilateral settlement framework intended to remunerate TSOs that host cross-border flows and incur associated losses, while charging those deemed responsible for generating them. Its legal basis was initially established through ETSO agreements and later codified in Commission Regulation (EU) No. 838/2010, with the current framework anchored in Regulation (EU) 2019/943 (ETSO 2007; European Commission 2010; EU 2019). Implementation and transparency are supported by ENTSO-E's methodological reporting and ACER's (The EU Agency for Cooperation of Energy Regulators) annual monitoring (ENTSO-E 2023; ENTSO-E 2024; ACER 2024; ACER 2025).

From an economic perspective, the design problem faced by the ITC mechanism is nontrivial. The system seeks to allocate costs that arise from complex network interactions using a limited set of observable quantities, primarily boundary power flows sampled at standardized snapshots. In meshed networks with internal congestion and loop flows, such observables provide only indirect and potentially ambiguous signals of underlying physical causation.

* University of Zagreb, Croatia, e-mail: dubravko.sabolic@gmail.com, ORCID: 0000-0002-7892-8701

Early analytical work by Daxhelet and Smeers (Daxhelet, Smeers 2005) demonstrated that, under realistic conditions, the ITC rules can generate paradoxical or indeterminate outcomes, while subsequent equilibrium-based analyses embedded the mechanism within a broader market-design context and highlighted interactions with congestion management and trade (Daxhelet, Smeers 2007).

A parallel strand of the literature examined alternative compensation approaches and their properties. Comparative analyses assessed different Inter-TSO Compensation schemes against criteria such as cost causation, neutrality, and implementability, concluding that no method simultaneously satisfies all desirable properties in a meshed transmission network (Olmos, Pérez-Arriaga 2007b; Olmos and Pérez-Arriaga 2007a).

Further work explored refinements and hybrid designs, emphasizing the inherent trade-off between theoretical consistency and practical feasibility (Stoilov et al. 2011; Stoilov, Stoilov 2013). Collectively, these contributions underscore that boundary-based allocation rules necessarily rely on proxies that may only imperfectly reflect the physical drivers of network usage and losses.

A number of additional contributions have approached the ITC problem from complementary perspectives. Early work by Glavitsch and colleagues (Glavitsch et al. 2004) proposed a flow-based methodology for allocating cross-border network usage, highlighting the inherent difficulty of linking physical flows to economic responsibility. From a regulatory standpoint, CEER (the Council of European Energy Regulators; CEER 2004) emphasized the long-term design challenges of the ITC mechanism, including the trade-off between cost causation, neutrality, and practical implementability.

Somewhat more recent modeling efforts have explored the optimization of cost and benefit allocation within the ITC framework under alternative assumptions and system conditions (Andročec et al. 2011). These contributions reinforce the view that the allocation of cross-border network costs remains a structurally complex problem for which no fully satisfactory solution exists.

A more conceptual critique was advanced in Sabolić's work (Sabolić 2017), which showed that even in a simple two-zone setting, internal congestion within one area can induce physical power detours through neighboring networks in the absence of commercial exports. In such cases, the ITC mechanism records symmetric inflows and outflows and assigns transit to both zones, despite the fact that only one zone causally generates the detour. The argument is structural rather than statistical: increasing measurement precision or temporal granularity refines the proxy but does not alter its informational content. As a result, the mechanism may fail to distinguish between zones that induce network stress and zones that merely host it.

More recent work has revisited the ITC mechanism from an institutional and descriptive empirical perspective. Using party-level data published by ACER and

ENTSO-E, a legal–economic reappraisal documented that aggregate ITC outcomes often display a high degree of internal consistency by construction, while material deviations persist at the level of individual parties (Sabolić 2025). That analysis also highlighted the continued reliance on the with/without-transit (WWT) approach for loss attribution and on standardized snapshots, despite ongoing discussions about higher temporal granularity (ACER 2023; ENTSO-E 2024). These findings suggest that procedural evolution and improved transparency have not fundamentally altered the informational basis of the mechanism.

At the operational level, both the literature and public documentation indicate that TSOs have actively responded to unscheduled and loop flows by investing in controllability measures, most notably phase-shifting transformers (PSTs). Case studies involving the German–Polish border and other highly interconnected interfaces illustrate how such investments can materially alter physical flow patterns (PSE 2014; ENTSO-E 2016; Opala, Ogryczak 2017). While these measures are primarily justified on operational grounds, they also affect the incidence of measured transit and losses and, indirectly, ITC settlements, reinforcing the importance of understanding what the mechanism does and does not capture.

Against this background, the present paper offers an additional analytical contribution by developing a new stylized three-zone counterexample that builds on, but goes beyond, earlier constructions. The example considers a configuration in which internal congestion in one control area induces a physical flow detour through another, non-causal area, leading to compensation outcomes that are difficult to reconcile with the physical origin of the burden.

Unlike the two-zone case, the presence of a third control area allows a clearer separation between the zone that induces network stress and the zone that passively accommodates it. The purpose of the exercise is not to claim empirical prevalence or to derive welfare rankings, but to show how, under admissible and transparent network conditions, the ITC settlement rules can map identical observable inputs into paradoxical compensation outcomes that are inconsistent with the intended redistribution goals of the mechanism.

2. Methodology

The analysis in this paper is analytical and relies on a stylized thought experiment. The approach follows the tradition of earlier paradox-based analyses of the Inter-TSO Compensation (ITC) mechanism, such as those developed by Daxhelet and Smeers (Daxhelet, Smeers 2005) and by Sabolić (Sabolić 2017). As in those articles, the purpose is to examine how the ITC settlement rules operate under transparent and economically meaningful network configurations, rather than to assess their empirical frequency or statistical relevance.

The thought experiment considers a simplified three-zone transmission system in which physical power flows differ from commercial schedules due to internal network physics. The construction focuses on how such configurations are recorded by the boundary-flow observables used in the ITC mechanism and how these observables translate into compensation outcomes. The emphasis is on the logical mapping from physical flows to ITC-relevant quantities, not on equilibrium behavior, welfare comparisons, or parameter calibration.

The present work extends this line of analysis by introducing a new configuration that complements previously identified paradoxes and sheds additional light on the interpretation and incentive properties of the mechanism.

An analytical approach is particularly appropriate in this context, as the objective is to identify structural properties of the ITC mechanism that arise from its very design and are not contingent on specific empirical configurations.

3. EU Inter-TSO Compensation mechanism

A brief description of the ITC mechanism and its underlying data architecture (European Commission 2010) is provided in Sabolić's works (Sabolić 2017, 2025). Nevertheless, we shall reproduce it here for the reader's convenience.

ITC settlement is carried out on a monthly basis and is constructed from a predefined set of standardized system states, commonly referred to as *snapshots*. Let $t = 1, \ldots, N$ index these snapshots, where N is determined by the prevailing methodology. Each snapshot is intended to represent a class of hourly operating conditions, and monthly quantities are obtained by aggregating across the N snapshots.

For a given control area i and snapshot t, let $U_i(t)$ denote the total physical outflow to neighboring areas, and $I_i(t)$ the total physical inflow. Let $G_i(t)$ and $K_i(t)$ denote total internal generation and consumption, respectively. A boundary-based proxy for transit through area i is defined as:

$$T_i(t) = \min\{U_i(t), I_i(t)\} \tag{1}$$

Aggregation over the accounting month is performed across snapshots using the weights specified by the ITC methodology, reflecting the representativeness of the snapshots (ENTSO-E 2024). For expositional convenience, without affecting the analytical argument and in line with earlier analytical formulations, the scaling ratio is written as an unweighted average:

$$\rho_i = \frac{1}{N} \sum_{t=1}^{N} \frac{T_i(t)}{T_i(t) + \max\{G_i(t), K_i(t)\}} \tag{2}$$

which relates measured transit to total activity in the area.

Infrastructure component. The entitlement of area *i* to infrastructure compensation is computed as:

$$R_i = \rho_i \cdot LRAIC_i \tag{3}$$

where $LRAIC_i$ denotes the calibrated long-run average incremental cost of the extra-high voltage network. Summing across all participating TSOs yields the total infrastructure fund:

$$F = \sum_i R_i \tag{4}$$

Financing of this fund is based on a measure of the so-called absolute accumulated cross-border flow:

$$A_i = \sum_{t=1}^{N} \left| U_i(t) - I_i(t) \right| \tag{5}$$

so that the contribution of area *i* is given by:

$$D_i = F \cdot \frac{A}{\sum_j A_j} \tag{6}$$

where the index *j* runs over all TSOs participating in the ITC mechanism. The resulting net position in the infrastructure component is:

$$Net_i^{infra} = R_i - D_i \tag{7}$$

Following the entry into force of Commission Regulation (EU) No. 838/2010, the infrastructure component was partially modified. While 75% of the infrastructure fund continues to be allocated according to the above rules, the remaining 25% is redistributed using a consumption-related adjustment factor defined as:

$$f_{K,i}(t) = \frac{T_i(t)}{\sum_j T_j(t)} \cdot \frac{T_i(t)}{T_i(t) + K_i(t)} \tag{8}$$

This factor moderates pure transit-based allocation by incorporating the ratio of transit to internal consumption. At the system level, its quantitative impact on redistribution has been reported to be limited (ACER 2014).

Losses component. Compensation for transmission losses is based on a with/without transit (WWT) approach, which estimates the reduction in network losses under a counterfactual state without transit relative to the observed state

(ACER 2023). Let δP_i denote this loss difference in energy units. Valuing losses at the prior-year average procurement cost C_i yields:

$$L_i = \delta P_i \cdot C_i \tag{9}$$

and the corresponding losses fund is:

$$F_L = \sum_i L_i \tag{10}$$

Contributions to the losses fund are allocated using the snapshot-level factor:

$$f_{L,i} = \frac{T_i(t)}{T_i(t) + K_i(t)} \tag{11}$$

which is aggregated across the month using snapshot weights. In current practice, ITC settlement relies on $N = 6$ standardized snapshots per month (3:30 a.m., 11:30 a.m., and 7:30 p.m. CET/CEST on the third Wednesday and the preceding Sunday). Each hour of the month is mapped to one of these snapshot types, and the corresponding hour shares are used as weights w_t (ENTSO-E 2024; ACER 2025).

Formally, letting $\sum_{t=1}^{N} w_t = 1$, we can define the monthly transit-loss weighted average as:

$$\bar{f}_{L,i} = \sum_{t=1}^{N} w_t \frac{T_i(t)}{T_i(t) + K_i(t)} \tag{12}$$

so that the contribution of area i to the losses fund is:

$$D_{L,i} = F_L \cdot \frac{\bar{f}_{L,i}}{\sum_j \bar{f}_{L,j}} \tag{13}$$

The construction is intentionally based on quantities that can be measured consistently across all participating control areas using boundary metering and a limited number of standardized system states. As a result, the mechanism abstracts from internal network topology and does not attempt to identify the physical origin of individual power flows. This reliance on perimeter-based observables ensures implementability but also constrains the informational content available for settlement. The implications of this feature are central to the analytical discussion that follows.

Firm-level implications for TSOs

From the perspective of an individual transmission system operator, the ITC mechanism operates as a largely ex post and non-strategic clearing arrangement.

Monthly settlement outcomes are effectively price-taking: the methodology, the mapping of hours to standardized snapshots, and the valuation of losses under the with/without-transit framework are determined outside the control of individual firms. Day-to-day operational decisions – such as unit commitment, remedial actions, and redispatch – affect ITC outcomes only indirectly, through their influence on boundary-measured physical flows.

In a coupled zonal market design without nodal (LMP) signals, TSOs face limited scope to manage network externalities through prices. This institutional setting is consistent with broader assessments of European pricing-rule alternatives, which emphasize the weak signaling properties of zonal pricing for internal congestion management (Weißensteiner 2023). As a result, managerial discretion shifts away from short-run operational adjustments toward slower-moving and capital-intensive instruments. These include investments in network controllability (most notably phase-shifting transformers), engagement in cross-border capacity governance and coordinated operational arrangements, and internal policies for procuring and valuing technical losses.

The experience on the Poland–Germany border illustrates this pattern. Persistent unscheduled flows associated with northern German wind generation led to coordinated investments in phase-shifting transformers, formalized in the 2014 agreement between PSE and 50Hertz and followed by concrete steps to regulate flows on the Poland–Germany interface (PSE 2014; ENTSO-E 2016). Engineering studies document how automatic PST adjustment can be used to steer active power flows on international lines (Opala, Ogryczak 2017). These interventions are structural rather than transactional: they reshape physical flow patterns and thereby modify exposure to ITC allocations indirectly.

Related evidence from the Germany–Netherlands–Belgium–France interface highlights a complementary issue. Boundary-based proxies may allocate compensation in ways that diverge from cost causation, so that a system inducing large internal transfers can appear as a beneficiary, while neighboring systems that physically host the resulting loop flows bear a disproportionate burden (Daxhelet, Smeers 2005). Such outcomes reinforce the incentives for TSOs to invest in controllability and to advocate institutional changes that reduce unwanted transits.

The managerial implications are therefore twofold. In the short run, ITC outcomes should be regarded as largely exogenous realizations of a proxy-based settlement rule. In the medium run, however, exposure management is closely tied to structural decisions – network investments, topology choices, and governance arrangements – rather than to tactical operational behavior. While firm-level challenges in this environment are substantial, a detailed managerial analysis lies beyond the scope of this paper. For the purposes of the present study, the key observation is that the ITC mechanism leaves relatively little room for active short-term management within its existing institutional framework.

4. A three-zone thought experiment

This section develops a stylized three-zone thought experiment that illustrates how the ITC infrastructure settlement responds to changes in commercial exchange patterns when the underlying physical configuration of the transmission system remains unchanged. In this example, the loss component of the ITC compensation will not be analyzed.

The intuition behind the example is as follows. Internal congestion in one control area may induce physical loop flows through neighboring systems without any corresponding commercial transaction. If compensation is based solely on boundary-measured flows, the mechanism may attribute responsibility symmetrically across affected areas, even when the underlying physical causation is asymmetric. The construction below isolates this effect and examines how such situations are reflected in ITC settlement outcomes.

The construction is deliberately simple and should be read jointly with Figure 1, Table 1, and the defining equations in Section 3. Its purpose is to make the mapping from physical flows to ITC compensation transparent and to examine the incentive properties implied by that mapping.

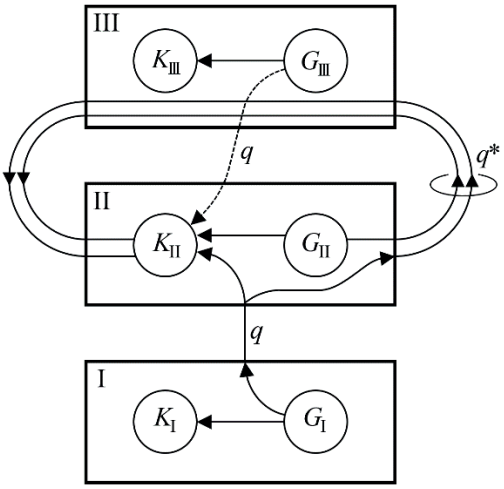


Figure 1. Stylized three-zone transmission system used in the thought experiment

Figure 1 illustrates the structure of the example. Area I exports power to Area II under a commercial schedule that corresponds exactly to the cross-border flow q. However, due to internal constraints within Area II, part of the power

destined for its consumers is physically rerouted through Area III, giving rise to an unscheduled loop flow q^*. Importantly, this transit through Area III is not associated with any commercial transaction involving Area III, but is instead caused by internal network conditions in Area II. In Model 2, an additional commercial export of magnitude q from Area III to Area II (dashed arrow) is added relative to Model 1, without altering the underlying physical configuration of the system.

In Table 1, variables are indexed by Roman numerals corresponding to the control areas I, II, and III from Figure 1. It applies: $F = \rho_{II} + \rho_{III}$, where $\rho_{II} = q^*/(q^* + K_{II})$ and $\rho_{III} = q^*/(q^* + G_{III})$. Δ denotes the change in net financial position between Model 2 and Model 1. All variables T, ρ, R, F, A, D, and *Net* are defined as in the corresponding equations in Section 3. For simplicity, the *LRAIC* parameter from (3) is normalized to unity for all three areas, and all snapshot observations are assumed to be identical.

Table 1

Infrastructure component of ITC compensation in the three-zone thought experiment

Model	Parameter	I	II	III
Model 1	T	0	q^*	q^*
	ρ	0	$q^*/(q^* + K_{II})$	$q^*/(q^* + G_{III})$
	R	0	$q^*/(q^* + K_{II})$	$q^*/(q^* + G_{III})$
	A	q	q	0
	D	$F/2$	$F/2$	0
	Net	$-(\rho_{II} + \rho_{III})/2$	$(\rho_{II} - \rho_{III})/2$	ρ_{III}
Model 2	T	0	q^*	q^*
	ρ	0	$q^*/(q^* + K_{II})$	$q^*/(q^* + G_{III})$
	R	0	$q^*/(q^* + K_{II})$	$q^*/(q^* + G_{III})$
	A	q	$2q$	q
	D	$F/4$	$F/2$	$F/4$
	Net	$-(\rho_{II} + \rho_{III})/4$	$(\rho_{II} - \rho_{III})/2$	$(3\rho_{III} - \rho_{II})/4$
	Δ	$+(\rho_{II} + \rho_{III})/4$	0	$-(\rho_{II} + \rho_{III})/4$

Consider three control areas, denoted I, II, and III. Each area contains internal generation G and consumption K, connected by an internal network. Area II is internally constrained: its transmission system cannot accommodate the full transfer required to supply internal demand using internal generation and imports alone. As a consequence, part of the power destined for consumers in Area II

flows physically through Area III, giving rise to a loop flow of magnitude q^* that is caused by internal physical conditions in the grid of Area II; Area III merely hosts the resulting unintentional and commercially unscheduled transit.

We now consider two configurations, referred to as Model 1 and Model 2. In both cases, the internal topology, the magnitude of the loop flow q^*, and the physical burden borne by Area III are identical. The only difference lies in the pattern of commercial exchanges.

In Model 1, Area I exports a quantity q to Area II. Area II imports q and, due to its internal bottleneck, induces a loop flow q^* through Area III. Area III is commercially neutral. The resulting ITC-relevant quantities are reported in the upper panel of Table 1. Area III is entitled to infrastructure compensation through ρ_{III}, while Areas I and II finance the infrastructure fund equally.

By construction, the variables ρ take values in the range [0, 1]. With this in mind, the net financial position of Area II depends on the relative magnitudes of ρ_{II} and ρ_{III}. When Area III is electrically significantly larger than Area II (as, for example, Germany compared to Austria), ρ_{II} may exceed ρ_{III}, and Area II receives a positive net payment, even though it is the only one of the three zones that actually causes the loop flows.

The resulting outcome can be interpreted as a situational form of free-riding: Area II benefits not from its behavior within the system, but from its relative electrical size compared to the hosting area. This advantage is merely a coincidence – a fact of life that has nothing to do with the behavior of Area II in the system composed of the three areas. Note, however, that in the present model setting, where transit is caused solely by loop flows originating in Area II, the coefficient ρ_{II} cannot in principle be much larger than 1/2. A higher value would imply that Area II has virtually no internal transmission grid, forcing almost the entire flow $K_{II} + q$ to be routed through Area III. Therefore, Area II can be net-positive only if the coefficient ρ_{III} is smaller than approximately 1/2.

Model 2 introduces an additional commercial exchange without altering the physical structure of the system. Area III now exports a quantity q to Area II, so that Area II imports $2q$ in total. Internal congestion and the resulting loop flow q^* remain unchanged.

The lower panel of Table 1 reports the corresponding ITC outcomes. Although Area II increases its reliance on external supply and remains the exclusive cause of the loop flow, its net infrastructure position is unchanged relative to Model 1, as indicated by $\Delta_{II} = 0$. The comparison across models highlights a structural feature of the ITC design.

Because settlement is driven by boundary-based transit proxies rather than by physical causation, changes in commercial behavior that increase dependence on external networks need not translate into higher compensation payments by the importing area.

In the present construction, the area responsible for the loop flow is not penalized for worsening the behavior of the system, even though it increases its import by q, while the additional financial burden is shifted to another party that increases its export by exactly the same amount q. Under the logic of the ITC mechanism – specifically, the notion of absolute accumulated flow – both actions are meant to be treated equally, yet only one of them – the "less guilty" party in this example – bears the entire additional cost. This outcome follows mechanically from the settlement rules summarized in Section 3. It does not rely on special parameter choices, nor does it require extreme assumptions about system configuration or size.

5. Conclusions and further research

This paper has examined the structural properties of the EU Inter-TSO Compensation (ITC) mechanism using a stylized three-zone thought experiment. Building on earlier analytical critiques, the construction isolates a configuration in which internal congestion in one control area induces physical loop flows through another area that is not causally responsible for the network stress. By varying commercial exchange patterns while holding the physical configuration constant, the analysis highlights how boundary-based transit proxies map identical physical outcomes into compensation results that differ in their economic interpretation.

The central finding is not merely that the ITC mechanism can misattribute costs in the presence of loop flows – an issue already documented in the literature – but that it may fail to penalize worsening behavior by the area that causes those flows. In the example developed here, an importing area can increase its reliance on external networks without facing a corresponding deterioration in its net ITC position, while another area undertaking an exactly symmetric action bears the entire additional financial burden. This outcome follows directly from the design of the infrastructure component and, in particular, from the use of absolute accumulated flow and boundary-based transit ratios as proxies for responsibility.

The analysis does not rely on extreme assumptions or parameter choices. All variables take admissible values, and the configurations considered are consistent with transparent and economically meaningful network conditions. The results therefore point to a structural feature of the ITC design rather than to a pathological special case. While the mechanism ensures implementability and administrative simplicity, it does so by abstracting from internal network topology and physical causation in a way that constrains the economic interpretation of its outputs.

These observations have implications for how ITC outcomes should be understood in practice. Party-level net positions should not be interpreted mechanically as indicators of responsibility or burden, especially in systems characterized by

internal congestion and loop flows. More broadly, the findings suggest that the ITC mechanism, in its current form, provides weak incentives to internalize the network externalities associated with increased cross-border dependence.

A final point concerns the interpretation of ITC outcomes in relation to the stated economic objective of the mechanism. The ITC framework is commonly understood as an attempt to approximate cost causation in a system where direct attribution is infeasible. The present analysis shows that, even under transparent and internally consistent conditions, the mapping from physical flows to financial outcomes may systematically diverge from this objective. In particular, identical physical configurations can give rise to different allocations solely due to changes in commercial exchange patterns, without any corresponding change in the underlying network burden. This suggests that the mechanism should be interpreted primarily as a pragmatic settlement rule based on observable proxies, rather than as a reliable indicator of economic responsibility within the system.

Several directions for further research follow naturally. One avenue is to explore whether alternative proxy designs – still implementable with limited and standardized data – could better align compensation with physical causation without sacrificing administrative feasibility. Another is to examine how ITC settlements interact with investment incentives in network controllability and grid reinforcement, particularly in systems where loop flows are persistent. Finally, integrating the present analytical insights with descriptive empirical evidence may help clarify under which conditions the structural features identified here are likely to be most relevant in real-world operation.

References

1. ACER (Agency for the Cooperation of Energy Regulators) (2014) Report to the European Commission on the implementation of the ITC mechanism in 2013, [Online], Available: https://www.acer.europa.eu/sites/default/files/documents/Publications/ITC%20Monitoring%20Report%202014.pdf [10 Jan 2026].
2. ACER (Agency for the Cooperation of Energy Regulators) (2023) Recommendation No 01/2023 of the European Union Agency for the cooperation of energy regulators of 13 April 2023 on the treatment of losses for the purpose of the ITC mechanism, [Online], Available: https://www.acer.europa.eu/sites/default/files/documents/Recommendations/ACER_Recommendation_01_2023_on_the_Treatment_of_Losses_for_the_Purpose_of_the_ITC_Mechanism.pdf [10 Jan 2026].
3. ACER (Agency for the Cooperation of Energy Regulators) (2024) Report on the implementation of the ITC mechanism in 2022, [Online], Available: https://www.acer.europa.eu/sites/default/files/documents/Publications/ACER_2024_Report_ITC_mechanism_2022.pdf [10 Jan 2026].

4. ACER (Agency for the Cooperation of Energy Regulators) (2025) Report on the implementation of the ITC mechanism in 2023, [Online], Available: https://www.acer.europa.eu/sites/default/files/documents/Publications/ACER-2025-ITC-mechanism.pdf [10 Jan 2026].
5. Androcec, I. Krajcar, S. and Wangensteen, I. (2011) 'Optimization of costs and benefits in Inter-TSO Compensation mechanism', 2011 8th International Conference on the European Energy Market (EEM), https://doi.org/10.1109/EEM.2011.5953091.
6. CEER (2004) CEER position paper. Long term requirements of the Inter-TSO Compensation mechanism (ITC). 2004-03-02 (Electricity Working Group, Cross Border Trade Task Force), [Online], Available: https://www.ceer.eu/wp-content/uploads/2004/03/Long-term-requirements-of-the-the-Inter-TSO-Compensation-Mechanism-ITC-%E2%80%93-a-CEER-Position-Paper-_02032004.pdf [7 Apr 2026].
7. Daxhelet, O. and Smeers, Y. (2005) 'Inter-TSO Compensation mechanism', Working Paper. Version: 7 November 2005, Cambridge, MA: Harvard Electricity Policy Group, John F. Kennedy School of Government, Harvard University, [Online], Available: https://hepg.hks.harvard.edu/sites/g/files/omnuum10586/files/hepg/files/daxlet_smeers_11.07.05.pdf [10 Jan 2026].
8. Daxhelet, O. and Smeers, Y. (2007) 'The EU regulation on cross-border trade of electricity: A two-stage equilibrium model', *European Journal of Operational Research*, vol. 181(3), pp. 1396–1412, https://doi.org/10.1016/j.ejor.2005.12.040.
9. ENTSO-E (2016) Step towards more power flows regulation between Germany and Poland, [Online], Available: https://www.entsoe.eu/news/2016/04/13/step-towards-more-power-flows-regulation-between-germany-and-poland/ [10 Jan 2026].
10. ENTSO-E (2023) ENTSO-E ITC transit losses data report 2022. Published following the requirements of Articles 4.2 and 4.3 of the Annex of Regulation (EU) No 838/2010, Part A, [Online], Available: https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/mc-documents/ITC_Transit_Losses_Data/entso-e_ITC_Transit_Losses_Data_report_2022.pdf [10 Jan 2026].
11. ENTSO-E (2024) ENTSO-E ITC transit losses data report 2023. Published following the requirements of Articles 4.2 and 4.3 of the Annex of Regulation (EU) No 838/2010, Part A, https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/mc-documents/ITC_Transit_Losses_Data/entso-e_ITC_transit_losses_data_report_2023_241108.pdf [10 Jan 2026].
12. ETSO (2007) ITC Agreement 2008–2009, [Online], Available: https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/pre2015/publications/etso/Inter_TSO_Compensation/Explanatory%20Notes%20on%20ITC%202008-9.pdf [10 Jan 2026].

13. EU (2019) Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (Recast), *Official Journal of the European Union*, L 158/54, 14.6.2019, pp. 54–124, [Online], Available: https://eur-lex.europa.eu/eli/reg/2019/943/oj/eng [10 Jan 2026].
14. European Commission (2010) Commission Regulation (EU) No 838/2010 of 23 September 2010 on laying down guidelines relating to the inter-transmission system operator compensation mechanism and a common regulatory approach to transmission charging, *Official Journal of the European Union*, L 25024.9.2010, pp. 5–11, [Online], Available: https://eur-lex.europa.eu/eli/reg/2010/838/oj/eng [10 Jan 2026].
15. Glavitsch, H., Andersson, G., Lekane, Th., Mariën, A., Mees, E. and Naef U. (2004) 'A flow-based methodology for the calculation of TSO to TSO compensations for cross-border flows', *International Journal of Electrical Power & Energy Systems*, vol. 26(1), pp. 49–56, https://doi.org/10.1016/S0142-0615(03)00072-3.
16. Olmos Camacho, L. and Pérez-Arriaga I.J. (2007a) 'An assessment of inter-TSO compensation algorithms in the internal electricity market of the European Union', *International Journal of Electrical Power & Energy Systems*, vol. 29(10), pp. 699–712, https://doi.org/10.1016/j.ijepes.2007.05.004.
17. Olmos Camacho, L. and Pérez-Arriaga I.J. (2007b) 'Comparison of several inter-TSO compensation methods in the context of the internal electricity market of the European Union', *Energy Policy*, vol. 35(4), pp. 2379–2389, https://doi.org/10.1016/j.enpol.2006.09.004.
18. Opala, K. and Ogryczak, T. (2017) 'Automatic adjustment of phase shifting transformers – the ability to control the active power flow in international exchange lines', *Acta Energetica*, vol. 2(31), pp. 149–157, https://doi.org/10.52710/ae.180.
19. PSE (Polskie Sieci Elektroenergetyczne) (2014) Agreement between Polish PSE and German 50Hertz transmission system operators on phase shifting transformers marks important step towards completion of the European energy market, [Online], Available: https://www.pse.pl/documents/20182/51490/20140312_Press_Release_PST_Agreement.pdf/96f51e25-a8ad-45c3-bcc9-8a5ee2c340bd [10 Jan 2026].
20. Sabolić, D. (2017) 'On economic inefficiency of the European Inter-TSO compensation mechanism', *Energy Policy*, vol. 110, pp. 548–558, https://doi.org/10.1016/j.enpol.2017.08.047.
21. Sabolić, D. (2025) The EU Inter-TSO compensation mechanism: A legal-economic reappraisal, SSRN, Elsevier BV, https://doi.org/10.2139/ssrn.5624808.
22. Stoilov, D., Dimitrov, Y. and François, B. (2011) 'Challenges facing the European power transmission tariffs: The case of Inter-TSO compensation', *Energy Policy*, vol. 39(9), pp. 5203–5210, https://doi.org/10.1016/j.enpol.2011.05.044.

23. Stoilov, D. and Stoilov, L. (2013) 'Improving inter-transmission compensation in EU', *Energy Policy*, vol. 62, pp. 282–291, https://doi.org/10.1016/j.enpol.2013.07.031.
24. Weißensteiner, F. (2023) 'Literature review. Pricing rule alternatives for the European day-ahead market', *Managerial Economics*, vol. 24(2), pp. 187–217, https://doi.org/10.7494/manage.2023.24.2.187.

Summary

This paper examines the structural design of the EU Inter-TSO Compensation (ITC) mechanism, which aims to remunerate transmission system operators for the costs associated with cross-border electricity flows. Building on the established legal and institutional framework and earlier analytical critiques, the paper develops a new stylized three-zone counterexample to assess how the mechanism allocates costs under realistic network interactions. The analysis shows that settlement rules based on boundary-flow proxies can assign compensation in ways that diverge from the physical origin of network burdens and may fail to penalize behavior that increases reliance on external networks. This finding complements existing paradoxes in the literature and highlights how proxy-based clearing can mute or distort operational and investment incentives, particularly in meshed networks with internal congestion. The contribution is analytical rather than empirical and is intended to clarify the scope and limitations of what the ITC mechanism can reasonably be expected to achieve, given its current design and informational basis.

JEL codes: L94, D47, Q48, L51

Keywords: *Inter-TSO Compensation, cross-border electricity flows, transmission networks, zonal market design, network externalities*

Managerial Economics
2026, Vol. 27, No. 1, pp. 53–84
https://doi.org/10.7494/manage.2026.27.1.53

Marek Straszak*

Accuracy as one of the dimensions of the quality of stock market recommendations published by Polish brokerage houses

1. Introduction

Thanks to the analyses of various aspects (internal and external) of the functioning of listed companies, investment recommendations increase the amount of available and professionally processed information, thereby improving the efficiency of capital markets.

In addition to their undoubtedly important function in improving the efficiency of capital markets, recommendations are primarily a source of knowledge for investors (Jegadeesh, Kim 2006; Hashim, Strong 2016). Equity investment requires market participants to devote more and more attention and time, and to have the appropriate analytical skills (Lusardi, Mitchell 2014). This is not only due to the growing number of listed financial instruments but also because of the increasing amount of information available to market participants. This information covers both the general macroeconomic situation and individual segments of the economy, as well as data from companies.

Due to their function, recommendations are an important tool for investors seeking support in the investment process and for the capital market in its development. It is crucial that they are prepared with due care and objectivity by persons with the appropriate knowledge and qualifications. Brokerage houses employ stock market analysts who provide their clients with professional assistance in decisions regarding capital allocation. In addition to daily commentary and minor ad hoc analyses, the essence of analysts' work is to publish recommendations that

* Krakow University of Economics, Krakow, Poland, e-mail: straszak.marek.pawel@gmail.com, ORCID: 0009-0001-0366-2128

indicate the expected direction of changes in the prices of financial instruments. The qualifications of equity analysts, combined with the standards for making recommendations set by brokerage houses, should ensure the high quality of the published materials.

This is also particularly important in the context of technological changes taking place in the financial sector. An increasingly common service offered by some financial institutions is **robo-advisory**. Currently, there is no legal definition of this type of service, but in market practice it is assumed that robo-advisory is a process in which recommendations are made and communicated using algorithms and automatic or semi-automatic systems (UKNF, 2024). Research shows that recommendations generated using artificial intelligence are less optimistic and are more evenly distributed between positive, negative, and neutral. In addition, they are published or updated more frequently and are more effective in the long term than those developed by stock market analysts (Coleman et al. 2022). As a result, they may become serious competition for traditionally produced analytical reports, especially if the latter are not sufficiently useful for investors.

Stock market analysts are usually qualified professionals whose skills in financial modeling and company valuation are confirmed by reputable certificates (e.g. Chartered Financial Analyst or Investment Advisor license). They are also subject to market standards and guidelines imposed by the institutions that employ them, as well as customer expectations, which may influence their work and the valuations of listed companies consequently (Brown et al. 2015). The impact of psychological (behavioral) factors, which may affect the content of the recommendations issued, is also significant (Mokoaleli-Mokoteli et al. 2009; Machado 2021). The influence of these factors means that the reports they produce may not always contain information processed in a way that ensures the highest value for investors. Therefore, there is a need to examine the quality of recommendations, which determines their macroeconomic (impact on financial market efficiency) and microeconomic (impact on the financial performance of investors using them) functions.

This research makes three distinct contributions to the literature. Firstly, the study extends the literature on investment recommendation by defining the concept of investment recommendation quality and classifying the dimensions through which it can be assessed. While prior studies have often described the results on different kinds of recommendation parameters, this article provides a systematic review of the dimensions of their quality and contributes to the theoretical understanding of the difference between the effectiveness and accuracy of investment recommendations.

Secondly, previous publications on the accuracy of recommendations have mostly focused on individual aspects. This study is intended to be a comprehensive analysis of the recommendations published for companies listed on the Warsaw Stock Exchange and proposes four indicators to examine the accuracy of target prices in recommendations in order to show the full picture of their quality in this dimension. Moreover, this article unveils a novel approach to examining accuracy by positive and negative recommendations and by sector.

Thirdly, what makes this research stand out among other works on the Polish stock exchange is the number of examined recommendations and the long range of data. The inclusion of 10,469 recommendations issued between 2005 and 2019 in the scope of the study made it possible to cover more than four economic and stock market cycles, and therefore should well illustrate the changing conditions on the capital market.

This article aims to answer the following research questions:

- What percentage of recommendations accurately indicate the direction of investment?
- What percentage of recommendations are implemented at the end of their validity period?
- What percentage of recommendations are implemented during their validity period?
- Are stock market analysts wrong when determining the investment potential of companies, and if so, to what extent?

Chapter 2 is devoted to defining the concept of recommendation quality and describing the dimensions by which it can be assessed. Chapter 3 reviews the existing literature on the accuracy of recommendations. Chapter 4 presents the scope of the data used and the research methodology applied. Chapters 5 and 6 are devoted to presenting the empirical results and conclusions of the study.

2. Quality, relevance, and effectiveness of recommendations

There are many definitions of quality in the literature that attempt to capture its meaning from different angles (Rura-Polley, Clegg 1999; Seawright, Young 1996). Garvin (Garvin 1984) divided definitions of quality into seven categories: general (transcendent), production-related, product-related, user-related, value-related, multidimensional, and strategic. Another division of definitions was proposed by Reeves and Bednar (Reeves, Bednar 1994), who believe that quality can be

perceived as: excellence, value, compliance with specifications, meeting or exceeding customer requirements, and as a dynamic process. Quality can be defined as the degree to which an object (e.g. a product, service, process, organization, or phenomenon) meets the requirements of a specific entity (Badura 2022). It can also be considered from the point of view of meeting the manufacturer's expectations (profitability and competitiveness of the product) and from the point of view of satisfying the user's needs (Frąś 2000). According to ISO 8402-1986, quality can be defined as the totality of features and characteristics of a product or service that determine its ability to satisfy stated or implied needs (Szutkowska, 2016). Although this definition could well fit the description of recommendation quality, it seems too general.

Since there is no single consistent definition of recommendation quality in the literature, it is reasonable to fill this gap based on a framework drawn from management and quality science. With regard to the shares of listed companies, the quality of a recommendation encompasses all the features and characteristics that determine its ability to systematically provide reliable and professional-standard information on the company's financial situation and the expected value of its securities, thereby effectively supporting investors' decision-making, improving efficiency and promoting the capital market.

There is no consensus among researchers as to whether the quality of recommendations (manifested primarily in their accuracy) has a significant impact on their effectiveness. There are studies indicating the existence of such a relationship (Loh, Mian 2006; Ertimur et al. 2007), but there are also those that deny it (Mikhail et al. 1999; Hall, Tacon 2010). However, the quality of a recommendation should somehow translate into its effectiveness. However, the relationship between the quality and effectiveness of a recommendation is sometimes ambiguous. For example, it may turn out that even though a recommendation was made carefully, in accordance with regulations and applicable standards, and even proved to be accurate, the investment made on its basis generates a lower rate of return than the market benchmark. The opposite situation may also occur, where a poor-quality recommendation leads to a successful investment. It follows that the effectiveness of recommendations may largely depend on their skillful use in the investment process applied by the market participant. Based on recommendations, appropriate investment strategies can be developed that exploit their certain properties while mitigating any shortcomings. The relationship between effectiveness, quality and its dimensions, as well as the factors that shape it, is shown in Figure 1.

The quality of recommendations can be assessed in many dimensions. Based on the literature, an attempt was made to classify them, as shown in Figure 2.

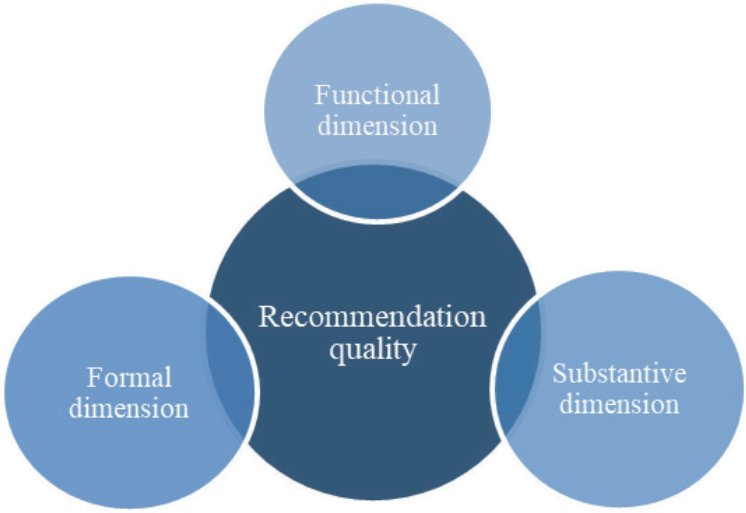


Figure 1. Effectiveness and quality of recommendations

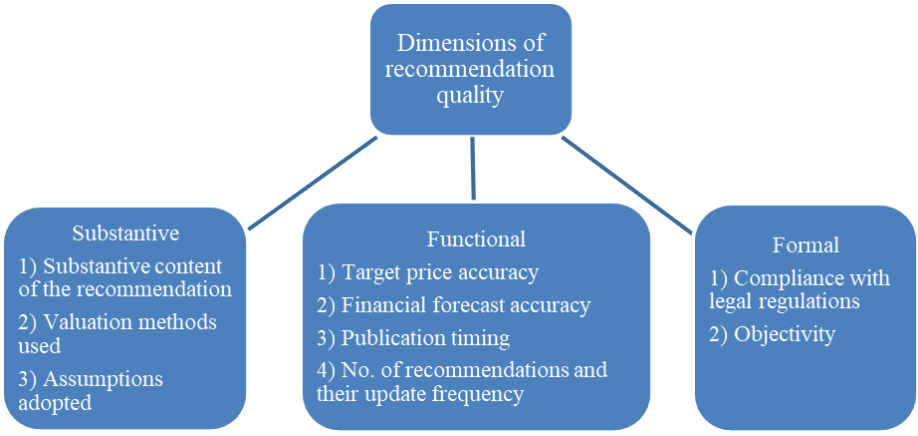


Figure 2. Dimensions of recommendation quality

Firstly, the substantive dimension of analytical reports can be assessed. To this end, it is necessary to check whether the recommendation contains an appropriate macroeconomic and sectoral analysis, whether historical data from financial statements are included and presented in an appropriate manner, and what macroeconomic assumptions have been made for the purpose of forecasting

results. Włodarczyk shows that while almost every recommendation includes historical financial data generated by the company, more than half of the reports do not contain a macroeconomic analysis and a description of the sector in which it operates (Włodarczyk 2020). The research also shows that the majority of recommendations do not include assumptions about the future macroeconomic environment. These were described in only one-third of the cases examined.

Secondly, quality can be determined based on the number and adequacy of the valuation models used. The choice of valuation methods may prove crucial for determining the value of a company, as different approaches usually indicate different investment potentials. A well-executed valuation requires the adoption of an appropriate approach to determining the intrinsic value of a company, so the use of inappropriate methods may result in an incorrectly calculated target price. However, there is no single market practice indicating how many and which methods should be used. Asquith and colleagues, after analyzing 1,126 analytical reports, revealed that almost all of them used a comparative valuation method based primarily on P/E and EV/EBITDA ratios, and only 12.8% of the reports contained a valuation based on discounted cash flow models (Asquith et al. 2005). Demirakos and colleagues, examining recommendations issued for companies in the United Kingdom, noted that multiplier analysis was used in 52.86% of cases, which is only slightly more often than the DCF method, which analysts used in 47.14% of recommendations (Demirakos et al. 2010). They also noted that analysts more often use the DCF method for smaller companies whose cash flows are characterized by greater uncertainty and for those for which the peer group in a given industry is limited. In addition, they pointed out that DCF methods are used more often in a declining market, while comparative methods are used more often in a growing market. Huang and colleagues noted that the DCF model is more readily used by analysts in relation to companies with poorer performance and higher risk (Huang et al. 2023). In addition, they observed that investors react more strongly to recommendations supported by the DCF model. A similar relationship is also confirmed by the research of Sayed, who points out that although analysts are least willing to use the DCF method, this approach provides the most accurate target price forecast (Sayed 2015). On the other hand, Prusak examined the accuracy of valuation models used in stock market recommendations on the Polish market (Prusak 2015). It turns out that the highest accuracy can be achieved by using a combined valuation, i.e. a weighted average of two other valuations, e.g. DCF and comparative valuation. Next in line in terms of effectiveness was comparative valuation, followed by DCF valuation.

It is not only the choice of the method itself that can affect the calculated value of a company. The assumptions made, and errors in their application, are also important. This is particularly important in the DCF method, which depends on

many parameters, which are decided on a discretionary basis by the analyst or imposed from above by the brokerage house's policy. Research indicates numerous shortcomings committed in this regard by analysts (Kowalski 2018; Zarzecki 2024). It has been shown that the recommendations lack information on how to calculate the cost of capital used for discounting, especially in relation to the risk-free rate and the risk premium. Another shortcoming of the valuations prepared is that when determining the beta coefficient, leverage is often not applied (i.e. adjusting its value depending on the changing level of debt in the company), but is arbitrarily assumed to be 1 throughout the forecast period. Analysts also often use the nominal rather than the effective tax rate. Furthermore, detailed calculations of residual value are very rarely included in analytical reports. Researchers also point to simple calculation errors made by analysts. Using various valuation methods, primarily comparative methods, analysts often apply various types of subjective adjustments to make their valuations more consistent with the characteristics of a given company. These adjustments relate, among other things, to the structure of the company, its relative market position, or past share price performance. Research shows that recommendations in which analysts apply subjective valuation adjustments are more accurate (Bonini et al. 2022).

Investment recommendations issued by analysts may also be assessed from a formal perspective. Analytical reports should be prepared with due diligence, in accordance with the law and objectively when determining the target price for a given security. However, it should be borne in mind that achieving this goal is difficult in practice. Analysts' access to information, especially from companies, is not always equal. Furthermore, close cooperation between an investment firm and an issuer on many levels may lead to conflicts of interest, often resulting in overvaluation. On the other hand, thanks to its contacts with the issuer on various levels, a brokerage house may be able to develop more accurate financial forecasts (Mehran, Stulz 2007).

Unfortunately, there are many violations of the principle of fairness toward clients on the capital market in order to achieve benefits for a financial institution or stock market analyst. This problem is analysed by Buczek (Buczek 2005). He points to several examples of practices by brokerage houses and analysts themselves that may affect the objectivity of the analysis presented and the valuation of the company. These include the following relationships:

- securities issuers are clients of brokerage houses,
- analysts prefer to issue positive recommendations in order to maintain good relations with company management,
- positive recommendations increase interest in transactions involving financial instruments.

The first two groups of recommendation quality dimensions mentioned above (substantive and formal) relate more to the way reports are produced. The last group, in turn, refers to their usefulness (functionality) for the final recipients, i.e. investors. An important parameter for assessing the quality of investment recommendations is the timing of their publication and the number and frequency of their updates. In this context, the uniqueness of the information provided in published recommendations is also important. Yezegel (Yezegel 2015) points out that almost 25% of all recommendations are issued within three days of the publication of quarterly results. On the other hand, Ivković and Jegadeesh (Ivković, Jegadeesh 2004) noted that recommendations have the greatest informational value when they are issued just before the publication of financial results by public companies, which may be due to the leakage of confidential information. Jacob and colleagues (Jacob et al. 1999) indicate that analysts who respond more frequently to new information by issuing new recommendations increase their accuracy. In turn, research by Chen and Cheng suggests that more frequent updates of reports lead to a reduction in their informational value and, consequently, to a smaller market reaction after their publication (Chen, Cheng 2003). Hobbs and colleagues indicate that the recommendations of those analysts who update their reports most frequently provide investors with the most value in both the short and long term (Hobbs et al. 2012). They note that the most profitable recommendations are those issued as soon as possible after the publication of important information for the company, but this is not true in the case of quarterly results announcements. The authors also point out that this may be at least partly related to the ability of analysts to discover unique information that other market participants are not yet aware of. Shroff and colleagues also point out that the strongest market reaction occurs in the case of recommendations that were published earliest after the release of specific information, but such reports are characterized by less accurate forecasts (Shroff et al. 2004). Agapova and Filatova noted that analysts tend to issue more positive investment recommendations in the last month of the quarter and more negative recommendations in the first month of the quarter (Agapova, Filatova 2023). This phenomenon is particularly evident in companies with a large share of institutional investors and is explained by analysts working on behalf of large clients. The presented research indicates that the capital market needs both recommendations that disclose new information as quickly as possible and those that are issued later but contain much more accurate forecasts, so quality assessment based on this dimension may depend on the function they perform.

However, the most frequently studied qualitative feature of recommendations in the literature is their accuracy, both in terms of the accuracy of analysts' forecasts of listed companies' results (Das et al. 1998; Brown, Huang 2013; Wróblewski 2016), as well as the accuracy of the target price contained therein (Bonini et al. 2010 ;

Bradshaw et al. 2013). It would seem that the accuracy of company earnings forecasts should translate into the accuracy of the calculated target price. The validity of this hypothesis is supported by research by Gleason and colleagues, who revealed that analysts who were better at forecasting earnings per share were able to estimate the target price more accurately (Gleason et al. 2013). On the other hand, Eames and colleagues pointed to the absence of such a relationship, indicating that analysts who are positive/negative about a given company forecast excessively high/low earnings, respectively (Eames et al. 2006). Similarly, Hwang and Lou point to the absence of this relationship (Hwang, Lou 2011). However, they showed that analysts with positive or negative recommendations forecast excessively low or high profits, respectively, so that on the day of their announcement, the market reaction will be stronger and allow them to reach their target price.

All the above-mentioned features may affect the predictive value of the recommendations issued, but one of the most important pieces of information resulting from an analytical report is the calculated target price, as it ultimately influences the message conveyed by the report. For this reason, the assessment of the quality of recommendations in this paper will be conducted from the perspective of the accuracy of the forecast prices contained therein. This aspect is particularly important from the point of view of the process of constructing an investment portfolio based on the recommendations of stock market analysts.

3. Literature review

Academics address the issue of recommendation accuracy in two aspects, i.e. the accuracy of target prices and the accuracy of forecasts of companies' financial results. The results of research on the accuracy of target prices will be discussed in more detail, as from the point of view of the empirical research conducted in this study, it seems to be the most important dimension of the quality of recommendations. The methodology used by various authors to study this issue is extensive. Asquith and colleagues examined accuracy using a dynamic accuracy coefficient for reports issued between 1997 and 1999 (Asquith et al. 2005).[1] Their tests showed that for 54.28% of recommendations, the price exceeded their target price and was on average 37.27% higher than it, while for the remaining positive (negative) recommendations that did not reach their target price, the market price approached on average 84.38% of the projected potential at its maximum (minimum) point. Bradshaw and Brown measured the accuracy of recommendations using a static

[1] Dynamic/static accuracy coefficients refer to the percentage of recommendations that exceeded their target price during their validity period/on the last day of their validity period.

and dynamic accuracy coefficient for recommendations issued between 1997 and 2002 (Bradshaw, Brown 2005). The results of their tests showed that only in 24% and 45% of cases, respectively, did the stock market price exceed the target price. Using the same methodology, similar results for recommendations issued between 2000 and 2006 on the Italian market were obtained by Bonini and colleagues, who calculated that the above-mentioned coefficients were 20% and 33%, respectively (Bonini et al. 2010). Extensive analyses of this problem were carried out in a similar manner by Bradshaw and colleagues (Bradshaw et al. 2012). They showed that 64% of recommendations issued between 2000 and 2009 reached their target prices within 12 months of publication, while at the end of this period only 38% exceeded the prices assumed by analysts. The study also showed that target prices were on average 15% higher than the actual prices achieved, and the absolute forecast error averaged 45%. The authors noted that investors follow changes in target prices regardless of the analyst's reputation and ability. The hypothesis that analysts have the ability to predict future stock prices was positively confirmed as statistically significant, but its economic value was described by the authors as weak. Kerl calculated the dynamic accuracy coefficient for analytical reports issued in 2002–2004 on the German market (Kerl 2011). His results showed that this accuracy averages 73.64%, but varies for positive recommendations (75.69%) and negative recommendations (59.43%). The accuracy of analysts' forecasts was also tested by Bilinski and colleagues (Bilinski et al. 2013). They examined recommendations for 16 countries between 2002 and 2009. The analysis shows that the target prices in the published recommendations were exceeded in 59.1% of cases, with an absolute forecast error of 44.7%. According to the authors, these results allowed them to conclude that analysts demonstrate a consistent ability to forecast target prices. Tiberius and Lisiecki conducted accuracy studies for the German market (Tiberius, Lisiecki 2019). Their analysis shows that the average forecast error, measured by the average ex post error indicator, was +0.0095, which may indicate the very good accuracy of the recommendations issued. However, the standard deviation of this error was 0.3287, which led to the conclusion that the forecast for a single recommendation was poor. It also turned out that the distribution of the forecast error is very similar to the normal distribution. Kadam and Sethi conducted a study on the accuracy of target prices of recommendations issued on the Indian market in 2016–2020 in dynamic terms (Kadam, Sethi 2024). The recommendation accuracy coefficient was 63%.

Compared to developed markets, the topic of recommendation accuracy on the Polish capital market is much less frequently discussed in the literature. Extensive research on this topic was conducted by Prusak, who tested not only the accuracy but also the effectiveness of analytical reports issued by Polish brokerage houses (Prusak 2015). The study covered 470 recommendations issued between

2009 and 2012, using several approaches to assess the accuracy of analytical reports. Static accuracy tests showed that the actual price 6, 9 and 12 months after the recommendation was issued and at the end of its validity period differed from the target price by less than 10% in only 30.00%, 24.68%, 21.91% and 27.45% of cases, respectively. These results led to the conclusion that accuracy is very low and decreases over time. In the dynamic approach, the results showed that the analytical reports were at least good (defined by the author as a deviation from the target price of no more than 10%) in the corresponding periods in 63.37%, 67.82%, 71.78% and 67.82% of the analyzed reports. The second method proposed was to verify whether future prices fall within the range assumed by the brokerage house. It turned out that in all the above-mentioned time periods, analysts correctly predicted prices for only about one-third of all recommendations in static terms. In the dynamic approach, the accuracy rates increased to about 50%.

The assessment of the accuracy of recommendations on the Polish capital market was also a subject investigated by Biedrzycki, who examined the deviations of the actual price from the target price after 9 months for 516 recommendations (Biedrzycki 2008). The range of these deviations was from −77.66% to 922.86%, while the mean and median were 44.77% and 29.46%, respectively. In addition, it turned out that the number of underestimated companies was 271, while the number of overestimated companies was 120. The author noted that only in 57% of recommendations did the price follow the same direction as the investment recommendation. Significant deviations from target prices indicate that the recommendations are not very accurate. However, it is worth noting that the research was conducted in 2006–2007, which naturally resulted in unnatural valuations related to the bull market that was ongoing at the time. The second stage of the study was to assess the recommendations as accurate or inaccurate, depending on whether they fell within the ranges proposed by the author. The results revealed that only 43.2% of the recommendations were accurate. Czyżycki and Kłóska also conducted a study on the accuracy of analytical reports (Czyżycki, Kłóska 2010). They calculated Pearson's correlation coefficient between the investment potential resulting from the recommendation and the actual change in the share price after a period of three months. This coefficient was 0.73 and was statistically significant. The authors demonstrated a fairly high degree of convergence between analysts' forecasts and actual price changes. A study covering only the five largest companies in the WIG20 index was conducted by Dąbrowski (Dąbrowski 2013). The results indicate that only 44.63% of the recommendations issued in 2007–2011 reached their target price during their validity period. The author also ranked brokerage houses, which showed that the best houses in terms of recommendation accuracy at that time were Millennium Brokerage House, Goldman Sachs and DI BRE (now mBanku Brokerage House). A study of the competitiveness of

brokerage houses for recommendations issued between 2006 and 2012 was also conducted by Czyżycki (Czyżycki 2013). In this study, he analyzed the accuracy of the proposed target prices for periods from 1 to 100 days. It turned out that in the short term, Wood&Company was the best institution, while over time, Dom Maklerski BDM achieved better results. Brycz and Włodarczyk examined the accuracy of recommendations for Polish companies issued in 2012 using an accuracy coefficient in a dynamic approach (Brycz, Włodarczyk 2017). The index they calculated was 66.7%. Another assessment of the accuracy of recommendations as the deviation of the market price from the target price was carried out by Włodarczyk (Włodarczyk 2020). Her research covered the years 2005–2012 and proved that recommendations have predictive value only in the short term, as the accuracy of target price forecasts is highest for short periods (6–8 days from the publication of the report). However, over time, this accuracy deteriorates and is lowest for long-term periods (one and two years). The accuracy of recommendations issued in 2004–2016 was also assessed by Wnuczak (Wnuczak 2021). Using a static accuracy coefficient, he showed that in the case of buy recommendations, the target price was exceeded in only 30% of cases after one year. The results of the target price accuracy study are summarized in Table 1.

Table 1

Accepted ranges for assessing the accuracy of recommendations for individual indicators

Author(s)	Year of publication	Accuracy measures	Result of a given measure	Conclusion adopted by author
P. Asquith	2005	Dynamic accuracy coefficient	54.2%	No conclusion
		MAX/MIN average deviation of the target price from the actual price for realized recommendations	37.3	No conclusion
		MAX/MIN average deviation of the target price from the actual price for unrealized recommendations	−15.6	No conclusion
M. Bradshaw, L. Brown	2006	Static accuracy coefficient	24.0	Analysts' inability to predict prices
		Dynamic accuracy rate	45.0	Analysts' inability to predict exchange rates

Table 1 cont.

S. Bonini et al.	2010	Static accuracy coefficient	20.0	No accuracy
		Dynamic accuracy coefficient	33.1	No accuracy
A.G. Kerl	2011	Dynamic accuracy coefficient	56.5	No conclusion
		MAX/MIN average deviation of the target price from the actual price for realized recommendations	42.0	No conclusion
		MAX/MIN average deviation of the target price from the actual price for unrealized recommendations	−13.8	No conclusion
M.T. Bradshaw et al.	2012	Deviation of the target price from the actual price	15.0	Economically weak
		Absolute forecast error	45.0	Economically weak
		Static accuracy coefficient	38.0	Economically weak
		Dynamic accuracy coefficient	64.0	Economically weak
P. Bilinski et al.	2013	Dynamic accuracy coefficient	59.1	Sustainable ability to forecast target prices
		Absolute forecast error	44.7%	Sustainable ability to forecast target prices
V. Tiberius, L. Lisiecki	2019	Average forecast error	0.0095	Very high accuracy
		Deviation of the mean forecast error	0.3287	Poor accuracy for a single company
S. Kadam, M. Sethi	2024	Dynamic accuracy coefficient	63	Good accuracy
R. Biedrzyński	2008	Percentage of recommendations for which the price followed the investment recommendation	57	Poor accuracy
		Average absolute deviation between the market price and the target price	44.8	Poor accuracy
		Proprietary accuracy coefficient	43.2	Poor accuracy
R. Czyżycki, R. Kłóska	2010	Pearson correlation coefficient	73.0	High accuracy

Table 1 cont.

Author(s)	Year of publication	Accuracy measures	Result of a given measure	Conclusion adopted by author
P. Dąbrowski	2013	Dynamic accuracy coefficient	44.63	Low quality
B. Prusak	2015	Static accuracy coefficient	21.9	Very low accuracy
		Dynamic accuracy coefficient	71.8	Not very high added value
		Static accuracy coefficient for brokerage house intervals	~33	Very low accuracy level
		Dynamic accuracy coefficient for brokerage house ranges	~50%	Low accuracy level
B. Brycz, K. Włodarczyk	2017	Dynamic accuracy coefficient	66.7	No conclusion
K. Włodarczyk	2020	Average absolute deviation between market price and target price	From a few to several dozen per cent	Good accuracy only in the short term
P. Wnuczak	2021	Static accuracy coefficient	30% for buy recommendations	Recommendations should not be the only guide for investors

Based on the literature cited, the accuracy of recommendations should be assessed as average. Studies conducted to date do not provide clear confirmation of whether analysts accurately forecast the target prices of listed companies. Often, the authors of these studies do not provide a clear assessment of the results obtained. Table 2 was created based on the results of previous tests (summarized in Table 1) and the author's subjective assessment. It will be used to evaluate the results of research on the accuracy of recommendations.

Table 2

Adopted ranges for assessing the accuracy of recommendations for individual indicators

Indicators used to assess the accuracy of recommendations	Low accuracy	Average relevance	High accuracy
Percentage of cases where the price of a given asset followed the direction indicated by the recommendation	<67%	≥67%; ≤80%	>80%
Static accuracy coefficient	<33%	≥33%; ≤50%	>50%
Dynamic accuracy coefficient	<60%	≥60%; ≤75%	>75%
Average absolute deviation between market price and target price	>20%	≥10%; ≤20%	<10%

4. Data and methodology

The study used recommendations for companies listed on the Warsaw Stock Exchange published on www.bankier.pl and historical share prices downloaded from Bloomberg. The number of analytical reports was an important factor in determining the correctness of the proposed research. Therefore, it was assumed that only companies for which at least 99 recommendations were issued in 2005–2019 by all 69 investment firms operating on the Polish capital market at that time would qualify for the research sample. This limited the sample to companies that are most frequently analyzed by analysts and investors and have an adequate history of recommendations. It also excludes from the study companies that are valued sporadically, which may distort the impact of reports on their share prices and raise doubts about their quality.

Finally, 45 companies were selected for the study, for which a total of 10,469 recommendations were issued during this period. During the period under review, these companies were mainly included in the WIG20 and mWIG40 indices, although it should be added that the classification of some of them in a given index changed during this time. Although the names of individual companies and investment firms also changed during the period under review, the most current versions were used in this study. To avoid the "survivorship bias" effect (Liu et al. 2022), which consists in excluding from consideration entities that have ceased operations or been delisted from the stock exchange, neither companies nor brokerage houses that went bankrupt, were taken over or ceased operations during that time were eliminated from the research sample.

In order to ensure a uniform approach to all calculations, the closing prices for a given day were used. These prices are most commonly used in market practice and form the basis for the valuation of, among other things, investment funds.

Table 3 presents the companies selected for the study by sector, together with the total number of recommendations issued for all securities in a given sector.

Table 3

Companies participating in the research sample by sector, with the number of recommendations published for them

Sector	Companies	No. of companies	No. of recommendations
Financial	PKO BP, PEKAO SA, PZU, Alior, Millennium, Getin Holding, mBank, Santander, Bank Handlowy, ING	10	3135
Fuel and energy Energy	PKN, Lotos, MOL, PGNIG, PGE, ENEA, Tauron, CEZ	8	1820
Technology, Media, and Telecommunications	Cyfrowy Polsat, Orange, Netia, Agora, Asseco Poland, Comarch, Sygnity	7	1734
Construction and Real estate	Budimex, Trakcja, Polimex-Mostostal, PBG, Elektrobudowa, GTC, ECHO, Dom Development	8	1205
Retail	LPP, CCC, Amrest, Eurocash	4	874
Mining	KGHM, JSW, Bogdanka	3	857
Industrial	Kęty, Grupa Azoty, Police, Puławy, Ciech	5	844
Total		**45**	**10,469**

The accuracy of the investment recommendations was assessed using the following measures:

- percentage of cases where the price of a given asset followed the direction indicated by the recommendation,
- static accuracy coefficients in different periods,
- dynamic accuracy coefficient,
- deviation of the actual price at the end of the recommendation's validity period from the target price forecast by the analyst.

First, in how many cases the rate of return at the end of their validity period had the same sign (positive or negative) corresponding to the investment recommendation (positive or negative) was checked for positive and negative recommendations.

Second, the static accuracy coefficients of the recommendation were calculated for different periods using the following formula:

$$\text{Static accuracy coefficient}_t = \frac{\text{No. of recommendations for which the target prices were reached in the period } t}{\text{No. of issued recommendations}} \quad (1)$$

Various periods were assumed, where t (subscript) = 5, 30, 90, 182, 273 days and the end of the recommendation validity period (273 or 365 days, depending on the brokerage house). Third, the dynamic accuracy coefficient of the recommendation was calculated using the following formula:

$$\text{Dynamic accuracy coefficient} = \frac{\text{No. of recommendations for which the target prices were reached at the end of their validity period}}{\text{No. of issued recommendations}} \quad (2)$$

The dynamic accuracy coefficient differs from static accuracy in that it includes all cases where a company's share price exceeded the target price on any day during the entire validity period of the recommendation (273 or 365 days). Naturally, it must be higher than or equal to the static coefficient. Both studies also took into account only analytical reports with positive or negative recommendations. Those with neutral recommendations were omitted, as they usually include a forecast of price changes within a narrow range of ±10% set by the brokerage house. Including neutral reports in the study would distort the correct conclusions about their accuracy, as a large proportion of them would not result from trends affecting the price, but could be realized through natural market volatility, which would lead to an overestimation of the accuracy coefficient.

The fourth method of verifying the accuracy of the recommendations was to analyze the relative and absolute deviations of the actual share price at the end of the recommendation period from the target price. Firstly, the relative deviations were taken into the account, which were calculated using the following formula:

$$\text{Deviation of market price from target price} = \frac{\text{Market price at the end of the recommendation validity period} - \text{Target price}}{\text{Target price}} \cdot 100\% \quad (3)$$

This formula was applied to positive and neutral recommendations, while for negative recommendations, it was multiplied by a factor of −1 in order to

determine whether the deviation was positive or negative in relation to the target price. Based on the calculated deviations, a histogram showing their distribution was created. The intervals of this histogram, as well as all subsequent ones, were designed according to Scott's rule (Grzegorzewski et al. 2014). The basic parameters of this distribution, such as the mean, median, 1st and 3rd quartiles, as well as the standard deviation, skewness, and kurtosis, were also calculated and presented. Three statistical tests were used to test the conformity of the distribution of deviations with the normal distribution:

- Kolmogorov–Smirnov,
- Lilliefors,
- Anderson–Darling.

The calculated relative deviations are highly informative in the context of examining whether analysts are too conservative or optimistic about target prices. However, they do not show how far the actual price deviated from the target price on average, as positive and negative deviations cancel each other out. For this reason, absolute deviations of actual prices at the end of the recommendation period from target prices were also calculated using Formula 3 (absolute deviations).

Next, histograms were developed for relative deviations, broken down into positive and negative recommendations. Basic descriptive parameters were also calculated for them. As these distributions did not show the characteristics of a normal distribution, the non-parametric Mann–Whitney U test was used to test the equality of their medians. The distributions of the deviations and their parameters were also shown, broken down by sector. All recommendation accuracy tests were calculated using the closing prices on a given day.

5. Empirical results

The study of the accuracy of recommendations began by checking in how many cases for positive and negative reports the rate of return at the end of their validity period had the same sign (positive or negative) as the investment recommendation (positive or negative). The results show that analysts' investment recommendations were directionally accurate in only 47.0% of cases. This means that they were more often wrong about the future direction of price changes than they were right. This test was a prelude to the study of the static accuracy coefficient, and the calculated value was the upper limit that this coefficient could reach.

Next, how often analysts are able to accurately determine the appropriate target price was checked. To this end, we calculated the static accuracy coefficient,

i.e. the percentage of recommendations that reached a specific target price at the end of the assumed period. Only positive or negative reports were included in the study. The values of the static accuracy indicators for individual periods are presented in Table 4.

Table 4

Static recommendation accuracy coefficients in different periods

Number of days	5	30	90	182	273	End of recommendation validity period
Average accuracy [%]	9.1	14.9	22.1	28.4	30.8	29.0

The above data shows that in almost every eleventh recommendation, the target price was reached on average after just one week. In turn, in approximately 22.1% of analytical reports, the price exceeded the target price after one quarter. The accuracy coefficient increases over time, but reaches its highest value (30.8%) nine months after the recommendation is issued. At the end of the validity period, the accuracy rate is lower, at 29.0%. In turn, the calculated dynamic accuracy rate was higher, reaching 55.0%.

Accuracy coefficients for individual companies were also estimated to verify whether there are any specific correlations in this area, e.g. related to membership of a given sector. Deviations from the average static accuracy coefficient for individual securities range from −16.8 p. p. to +19.1 p. p. (Appendix 1). Analysts were least successful in predicting the correct price in the case of companies from the energy sector, i.e. ENEA (where the accuracy coefficient was 12.1%), PGE (16.2%), CEZ (17.3%) and Tauron (22.1%). On the other hand, some of the highest coefficients were achieved for companies in the retail sector: CCC (43.4%), Eurocash (43.0%) and Amrest (41.3%). In the case of the dynamic indicator, these deviations are very similar and range from −16.2 p. p. to +19.3 p. p.

The fourth method used to analyze the accuracy of the analytical reports was the deviation of the actual price at the end of the recommendation period from the forecast target price. The histogram of relative deviations is presented in Figure 3.

The deviations of the actual price at the end of the recommendation period from the target price forecast by analysts have a distribution similar to the normal distribution, but are not consistent with it, mainly due to the high kurtosis. This observation is also confirmed by the results of three independent statistical tests, i.e. Kolmogorov–Smirnov, Lilliefors, and Anderson–Darling, at a significance level of 0.05 (Tab. 5).

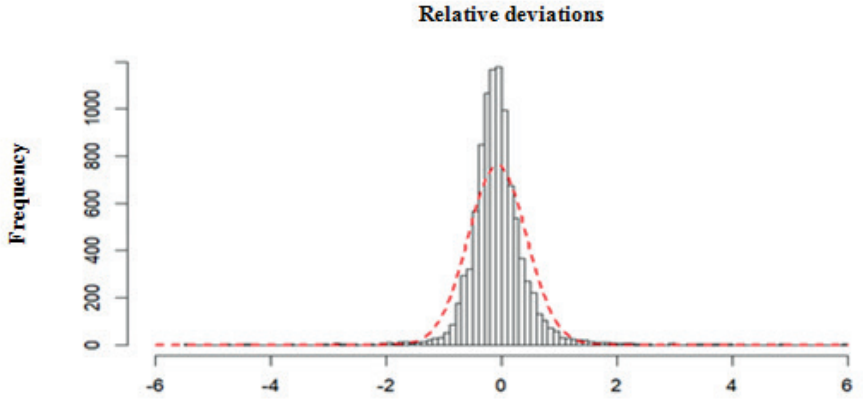


Figure 3. Histogram of relative deviations of the actual price at the end of the recommendation period from the target price forecast by analysts, with a normal distribution superimposed (red dashed line)

Table 5

Tests verifying the normal distribution of relative deviations obtained

Test	p-value	Compliance with a normal distribution at the 0.05 level
Kolmogorov–Smirnov	<0.01	None
Lilliefors	<0.01	None
Anderson–Darling	<0.01	None

The mean of this distribution is below zero and amounts to −7.08%, while the median is equal to −9.77%. This means that the distribution is right-skewed with a longer tail on the right side of the chart (skewness >0) (Tab. 6). Its analysis shows that as many as 62% of recommendations have target prices that are too high or too low (in the case of positive and negative analytical reports, respectively) compared to the actual prices of financial instruments. It can therefore be concluded that, on average, analysts forecast too-high positive investment potential for positive recommendations and too-high negative potential for negative recommendations (upside/downside). There may be at least two mutually non-exclusive reasons for this. Firstly, analysts, wanting to increase the attractiveness of buying or selling a company (and thus also their report), tend to overestimate the investment potential. Investors approach a recommendation that assumes a 10% increase differently than one that predicts, for example, a 40% increase. The latter is more

interesting to them because the range of potential price movement is much greater. On the other hand, when the price of an asset approaches its target price, market participants usually close their positions. This is often because as the price approaches the target price, the ratio of further potential profit to risk becomes less favorable, which encourages investors to close their positions on a given asset.

Table 6

Descriptive statistics of the distribution of relative deviations of the actual price at the end of the recommendation period from the target price forecast by analysts

Parameter	Value
Average	−7.08%
1st quartile	−31.30%
2nd quartile (median)	−9.77%
3rd quartile	13.30%
Standard deviation	49.75%
Kurtosis	15.12
Skewness	0.56

In order to examine the average difference between the actual price at the end of the recommendation period and the target price, the median and mean values of the absolute deviations of the actual price from the target price were also calculated. The average absolute deviation was 33.56%, with a median and standard deviation of 24.36% and 37.41%, respectively. Figure 4 shows a histogram of the absolute deviations of the actual price from the target price.

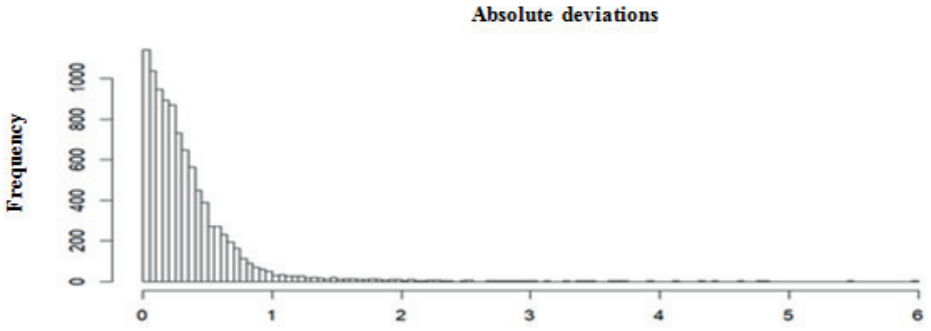


Figure 4. Histogram of absolute deviations of the actual price at the end of the recommendation period from the target price

The study of the accuracy of recommendations was supplemented by a comparison of the distributions of deviations for positive and negative recommendations. Histograms and statistics of these deviations are presented in Figure 5 and Table 7, respectively.

a) b)

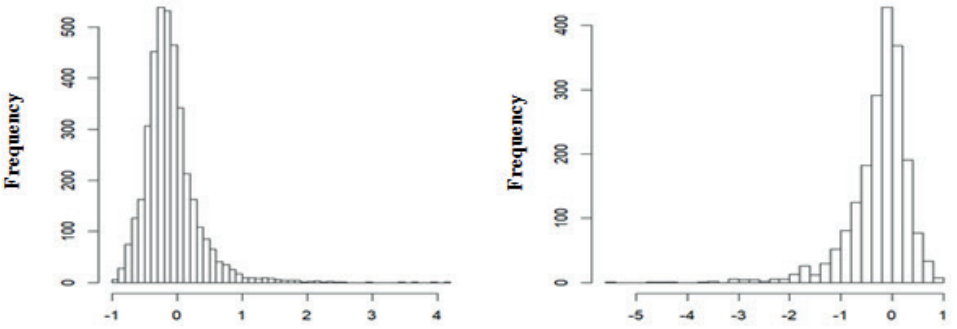


Figure 5. Histogram of relative deviations of the actual price at the end of the recommendation period from the target price forecast by analysts for positive (a) and negative (b) recommendations

Table 7

Descriptive statistics of the distribution of deviations of the actual price at the end of the recommendation period from the target price forecast by analysts for positive and negative recommendations

Parameter	Positive	Negative
Average	−9.50%	−25.47%
1st quartile	−34.19%	−45.93%
2nd quartile (median)	−15.53%	−13.16%
3rd quartile	5.75%	9.16%
Standard deviation	41.82%	59.71%
Kurtosis	12.51	10.43
Skewness	2.31	−2.31

Deviations for positive recommendations are characterized by a right-skewed distribution with a long tail on the right. Observations to the left of the mean take

values not lower than −1. This is a natural consequence of the fact that share prices cannot fall below 0. Both the mean and the median are negative, which indicates that analysts are too optimistic in their estimates of target prices for shares when they recommend buying them. The situation is similar for negative recommendations. The distribution of these deviations is similar to the distribution of positive recommendations. It is characterized by negative skewness and a long tail on the left. In this case, the mean and median take negative values, confirming that analysts also overestimate the negative potential for negative recommendations. However, it is worth noting that the mean in this case is almost twice as small as the median. This indicates that the authors of reports recommending the sale of shares make significantly larger errors (compared to recommendations suggesting a purchase), which means that these deviations more often take on extreme values, e.g. below −100% or −200%. This is also confirmed by the relatively high standard deviation for this group of recommendations, which is 59.71%. Actual prices deviate from the target price by ±10% in only 19.8% of cases for positive recommendations and 21.3% of cases for negative recommendations. The non-parametric Mann–Whitney U test (also known as the Wilcoxon rank test) returned a statistic value of W = 3,952,510 and a p-value <0.01, which allows us to conclude that there is a statistically significant difference between the medians of both distributions.[2]

In the next step, an analysis of the distributions of deviations for individual sectors was performed and their descriptive statistics are presented in Table 8.

Analysis of these results allows us to draw important conclusions that can be used to make more effective use of analytical reports. It is worth noting that only in the case of recommendations for the retail sector are both the mean and the median positive, which indicates that analysts are too conservative in their target price forecasts for companies in this industry. In the case of the other sectors, they are too optimistic, most notably in the case of the mining sector. The standard deviation for this sector (84.39%) is significantly higher than for the other industries. Also noteworthy are the relative deviations of the industrial segment, which is the only one with a platykurtic distribution. This shows that in this case, there are relatively fewer extreme values, which indicates a smaller number of extreme errors by analysts. In statistical terms, the Wilcoxon signed-rank test clearly showed that for all sectors, the medians of relative deviations differ significantly from zero. The results of the test are presented in Table 9.

[2] The t-test was not used due to distributions that deviated significantly from the normal distribution.

Table 8

Descriptive statistics of the distribution of deviations between the actual price at the end of the recommendation period and the target price forecast by analysts for recommendations broken down by sector

Parameter	Financial	Fuel and Energy	TMT	Construction and real estate	Retail	Mining	Industrial
Average	−4.86%	−11.14%	−9.13%	−10.22%	5.83%	−11.24%	−6.40%
1st quartile	−30.20%	−30.15%	−25.44%	−39.78%	−23.62%	−42.06%	−35.13%
2nd quartile (median)	−8.07%	−12.92%	−9.24%	−11.97%	0.80%	−15.79%	−8.36%
3rd quartile	16.67%	6.48%	4.95%	13.96%	30.92%	15.00%	18.33%
Standard deviation	50.29%	32.72%	34.16%	52.07%	53.27%	84.39%	45.12%
Kurtosis	10.08	10.49	10.88	4.76	5.95	12.58	1.32
Skewness	0.81	−0.93	0.80	0.37	0.51	0.49	0.19

Table 9

Results of the Wilcoxon signed-rank test for different sectors

Sector	Test statistic V	p-value
Financial	1,537,843	<0.01
Fuel and Energy	384,353	<0.01
TMT	385,131	<0.01
Construction and real estate	214,176	<0.01
Retail	177,629	0.0274
Mining	111,974	<0.01
Industrial	124,728	<0.01

6. Conclusions

The percentage of recommendations for which the generated rate of return was consistent with the expectations resulting from the investment recommendation

was only 47.0%, which means that analysts were more often wrong than right about the direction in which the company's share price would move in the future. Next, the static and dynamic accuracy coefficients achieved only 29.0% and 55%, respectively. The results obtained are similar to those of previous studies. Static accuracy coefficients in previous papers ranged from 20% to 38%, and the accuracy of recommendations was generally assessed by the authors as low or very low. In turn, dynamic accuracy coefficients in other studies ranged from 33% to 71.5%, and the authors' conclusions were generally consistent with those concerning static accuracy coefficients, i.e. that the accuracy of recommendations is poor. Based on the ranges established in Table 2, the results obtained for both accuracy coefficients justify a negative assessment of analysts' ability to correctly forecast share prices and an assessment of the accuracy of recommendations as low.

The low accuracy may be partly due to the fact that neutral recommendations were not taken into account in this study, and their inclusion would certainly increase the levels of the indicators in question. However, including them could lead to incorrect conclusions about the accuracy of recommendations (especially in the case of dynamic accuracy) due to the low investment potential that could be realized by market volatility rather than by trends shaping the price of a given company.

In the case of the dynamic approach, accuracy improved by 26 percentage points compared to the static approach. This is a natural consequence of the fact that for this approach, all cases were taken into account where the share price exceeded the target price on any day during the recommendation's validity period. It follows that investments based on recommendations should be monitored by market participants, as statically waiting for the end of the recommendation's validity period is less likely to bring the desired results. A dynamic approach to investing, in line with recommendations, therefore provides greater benefits for the investor.

However, the distribution of deviations indicates that only 38% of recommendations exceeded their target price at the end of the recommendation period (the discrepancy with the static accuracy coefficient is due to the fact that neutral recommendations were not included in that study). Moreover, one year after the report was published, the actual price deviated from the target price by more than 10% or 20% in 77.3% and 58.0% of recommendations, respectively. To emphasize the scale of the discrepancy between the forecasts and actual prices, it can also be pointed out that almost one-third of all target price forecasts differed from the actual values by more than ± 40%. Such a wide range of deviations is also confirmed by the standard deviation, which is close to 50%. Such frequent and significant errors can seriously limit the use of recommendations as an effective

tool in the investment process, especially when target prices are the only criterion on which market participants base their decisions.

The average absolute deviation was 33.6%. The presented results of deviations of the actual price from the target price are lower than the deviations in previous studies, but their high levels still justify a negative assessment of the ability of domestic analysts to correctly forecast share prices on the Warsaw Stock Exchange and to determine the accuracy of recommendations to be low.

The results of all the methods used to verify the accuracy of investment recommendations justify a negative assessment. However, this does not mean that analytical reports are not useful for the investment process, as their effectiveness is key in this case.

References

1. Agapova, A. and Filatova, U. (2023) 'Analysts' institutional client catering and reputation tradeoff: Strategic timing of recommendations', *Journal of Accounting, Auditing & Finance*, vol. 40, https://doi.org/10.1177/0148558X231198895.
2. Asquith, P., Mikhail, M.B. and Au, A.S. (2005) 'Information content of equity analyst reports', *Journal of Financial Economics*, vol. 75, pp. 245–282, https://doi.org/10.1016/j.jfineco.2004.01.002.
3. Badura, M. (2022) 'Analiza pojęcia jakości na podstawie literatury', *Przedsiębiorczość i Zarządzanie*, vol. 23(2), pp. 9–22, [Online], Available: https://www.ceeol.com/search/article-detail?id=1166903 [14 Apr 2026].
4. Biedrzycki, R. (2008) Wycena w rekomendacjach sporządzanych przez biura maklerskie a cena rynkowa spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie, praca magisterska, Uniwersytet Ekonomiczny w Szczecinie.
5. Bilinski, P., Lyssimachou, D. and Walker M. (2013) 'Target price accuracy: international evidence', *Accounting Review*, vol. 88(3), pp. 825–851, https://doi.org/10.2308/accr-50378.
6. Bonini, S., Capizzi, V., and Kerl, A. (2022) 'Subjective valuation and target price accuracy', *Journal of Financial Management, Markets and Institutions*, vol. 10(01), 2250005, https://doi.org/10.1142/S2282717X22500050.
7. Bonini, S., Zanetti, L., Bianchini, R. and Salvi, A. (2010) 'Target price accuracy in equity research', *Journal of Business Finance & Accounting*, vol. 37(9–10), pp. 1177–1217, https://doi.org/10.1111/j.1468-5957.2010.02209.x.
8. Bradshaw, M.T. and Brown, L.D. (2005) Do sell-side analysts exhibit differential target price forecasting ability?, [Online], Available: https://assets.csom.umn.edu/assets/37727.pdf [14 Apr 2026].

9. Bradshaw, M.T., Brown, L.D. and Huang, K. (2013) 'Do sell-side analysts exhibit differential target price forecasting ability, *Review of Accounting Studies,* vol. 18, pp. 930955, https://doi.org/10.1007/s11142-012-9216-5.
10. Brown, L.D. and Huang, K. (2013) 'Recommendation-forecast consistency and earnings forecast quality', *Accounting Horizons*, vol. 27(3), pp. 451–467, https://doi.org/10.2308/acch-50482.
11. Brown, L.D., Call, A.C., Clement, M.B. and Sharp, N.Y. (2015) 'Inside the "black box" of sell-side financial analysts', *Journal of Accounting Research*, vol. 53(1), https://doi.org/10.1111/1475-679X.12067.
12. Brycz B. and Włodarczyk K. (2017) 'Target price accuracy and the information content of stock recommendations on the Warsaw Stock Exchange. Trafność ceny docelowej a zawartość informacyjna rekomendacji na Giełdzie Papierów Wartościowych w Warszawie', *Nauki o Finansach. Financial Sciences*, vol. 4(33), pp. 35–50, https://doi.org/10.15611/nof.2017.4.03.
13. Buczek, PP. (2005) Efektywność informacyjna rynków akcji. Teoria a rzeczywistość, Warszawa: Oficyna Wydawnicza Szkoły Głównej Handlowej.
14. Chen, X. and Cheng, Q. (2003) 'What determines the market impact of stock recommendations?', *Sauder School of Business Working Paper*, https://dx.doi.org/10.2139/ssrn.385524.
15. Coleman, B., Merkley, K.J. and Pacelli, J. (2022) 'Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations', *The Accounting Review*, vol. 97(5), pp. 221–244, https://dx.doi.org/10.2139/ssrn.3514879.
16. Czyżycki, R. (2013) 'Marka instytucji finansowych a jakość udzielanych rekomendacji giełdowych', *Marketing i Zarządzanie*, vol. 31, pp. 155–166, https://www.ceeol.com/search/article-detail?id=292200.
17. Czyżycki, R., Kłóska, R. (2010) 'Rekomendacje giełdowe jako źródło wiedzy dla inwestora indywidualnego. Stock market recommendations as a source of knowledge for an individual investor', MAGNANIMITAS, pp. 2371–2378, [Online], Available: https://open.icm.edu.pl/handle/123456789/2409 [14 Apr 2026].
18. Das, S., Levine, C.B. and Sivaramakrishnan, K. (1998) 'Earnings predictability and bias in analysts? Earnings forecasts', *The Accounting Review*, vol. 73, pp. 277–294, [Online], Available: https://www.jstor.org/stable/248469 [14 Apr 2026].
19. Dąbrowski P. (2013) 'Hurraoptymizm rekomendacji maklerskich w okresie giełdowej dekoniunktury', *Studia Ekonomiczne*, vol. 174, pp. 88–104, [Online], Available: https://bazekon.uek.krakow.pl/rekord/171280783 [14 Apr 2026].
20. Demirakos, E., Strong, N.C. and Walker, M. (2010) 'Does valuation model choice affect target price accuracy?', *European Accounting Review*, vol. 19(1), pp. 35–72, https://doi.org/10.1080/09638180902990630.

21. Eames, M.J., Glover, S.M. and Kennedy, J.J. (2006) 'Stock recommendations as a source of bias in earnings forecasts', *Behavioral Research in Accounting*, vol. 18(1), pp. 37–51, https://doi.org/10.2308/bria.2006.18.1.37.
22. Ertimur, Y., Sunder, J. and Sunder, S.V. (2007) 'Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts', *Journal of Accounting Research*, vol. 45(3), pp. 567–606, https://doi.org/10.1111/j.1475-679X.2007.00244.x.
23. Frąś, J. (2000) Zarządzanie jakością w instytucjach gospodarczych, Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
24. Garvin, G.A. (1984) 'What does "product quality" really mean?', *MIT Sloan Management Review*, 1, pp. 25–41, [Online], Available: https://sloanreview.mit.edu/article/what-does-product-quality-really-mean/ [14 Apr 2026].
25. Gleason, C.A., Bruce Johnson, W. and Li, H. (2013) 'Valuation model use and the price target performance of sell-side equity analysts', *Contemporary Accounting Research*, vol. 30(1), pp. 80–115, https://doi.org/10.1111/j.1911-3846.2011.01142.x
26. Grzegorzewski P., Gągolewski M. and Bobecka-Wesołowska K. (2014) Wnioskowanie statystyczne z wykorzystaniem środowiska R, Warszawa: Politechnika Warszawska.
27. Hall, J.L. and Tacon, P.B. (2010) 'Forecast accuracy and stock recommendations', *Journal of Contemporary Accounting and Economics*, vol. 6(1), pp. 18–33, https://doi.org/10.1016/j.jcae.2010.04.003.
28. Hashim, N.A. and Strong, N.C. (2018) 'Do analysts' cash flow forecasts improve their target price accuracy?', *Contemporary Accounting Research*, vol. 35(4), pp. 1816–1842, https://doi.org/10.1111/1911-3846.12369.
29. Hobbs, J., Kovacs, T. and Sharma, V. (2012) 'The investment value of the frequency of analyst recommendation changes for the ordinary investor', *Journal of Empirical Finance*, vol. 19(1), pp. 94–108, https://doi.org/10.1016/j.jempfin.2011.09.006.
30. Huang, S., Tan, H., Wang, X. and Yu, C. (2023) 'Valuation uncertainty and analysts' use of DCF models', *Review of Accounting Studies*, vol. 28(2), pp. 827–861, https://doi.org/10.1007/s11142-021-09658-w.
31. Hwang, B.H. and Lou, D. (2011) "Self-Fulfilling" stock recommendations, Retrieved January, 16, [Online], Available: http://efa2011.efa-online.org/fisher.osu.edu/blogs/efa2011/files/BEH_3_3.pdf [14 Apr 2026].
32. Ivković, Z. and Jegadeesh, N. (2004) 'The timing and value of forecast and recommendation revisions', *Journal of Financial Economics*, vol. 73(3), pp. 433–463, https://doi.org/10.1016/j.jfineco.2004.03.002.
33. Jegadeesh, N. and Kim, W. (2006) 'Value of analyst recommendations: international evidence', *Journal of Financial Markets*, vol. 9(3), pp. 274–309, https://doi.org/10.1016/j.finmar.2006.05.001.

34. Jacob, J., Lys, T. Z. and Neale, M.A. (1999) 'Expertise in forecasting performance of security analysts', *Journal of Accounting and Economics*, vol. 28, pp. 51–82, https://doi.org/10.1016/S0165-4101(99)00016-6.
35. Kadam, S. and Sethi, M. (2024) 'Target price accuracy of sell-side analysts: evidence from India', *Cogent Economics & Finance*, vol. 12(1), 2423261, https://doi.org/10.1080/23322039.2024.2423261.
36. Kerl, A. (2011) 'Target price accuracy, *BuR Business Research Journal*, vol. 4(1), pp. 74–96, [Online], Available: https://ssrn.com/abstract=1806927.
37. Kowalski, M.J. (2018) 'Quality of investment recommendation – evidence from Polish capital market, multiples approach', in Wilimowska, Z., Borzemski, L. and Świątek, J. (ed.) Income Approach, Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT 2018. Part III, pp. 60–70, [Online], Available: https://link.springer.com/book/10.1007/978-3-319-99993-7 [14 Apr 2026].
38. Liu, X.Y., Xia, Z., Rui, J., Gao, J., Yang, H., Zhu, M., Wang C.D. et al. (2022) 'FinRL-Meta: Market environments and benchmarks for data-driven financial reinforcement learning', Advances in Neural Information Processing Systems 35 (NeurIPS 2022), https://doi.org/10.48550/arXiv.2211.03107.
39. Loh, R.and Mian, G.M. (2006) 'Do accurate earnings forecasts facilitate superior investment recommendations?', *Journal of Financial Economics*, vol. 80(2), pp. 455–483, https://doi.org/10.1016/j.jfineco.2005.03.009.
40. Lusardi, A. and Mitchell, O.S. (2014) 'The economic importance of financial literacy: Theory and evidence', *Journal of Economic Literature*, vol. 52(1), pp. 5–44, [Online], Available: https://www.jstor.org/stable/24433857 [14 Apr 2026].
41. Machado, A. and Lima, F.G. (2021) 'Sell-side analyst reports and decision-maker reactions: Role of heuristics', *Journal of Behavioral and Experimental Finance*, vol. 32, 100560, https://doi.org/10.1016/j.jbef.2021.100560.
42. Mehran, H. and Stulz, R.M. (2007) 'The economics of conflicts of interest in financial institutions', *Journal of Financial Economics*, vol. 85(2), pp. 267–296, https://doi.org/10.1016/j.jfineco.2006.11.001.
43. Mikhail, M.B., Walther, B.R. and Willis, R.H. (1999) 'Does forecast accuracy matter to security analysts?', *The Accounting Review*, vol. 74(2), pp. 185–200, [Online], Available: https://ssrn.com/abstract=153828 [14 Apr 2026].
44. Mokoaleli-Mokoteli, T., Taffler, R.J. and Agarwal, V. (2009) 'Behavioural bias and conflicts of interest in analyst stock recommendations', *Journal of Business Finance & Accounting*, vol. 36(3–4), pp. 384–418, https://doi.org/10.1111/j.1468-5957.2009.02125.x.
45. Prusak, B. (2015) Wielokryterialna analiza rekomendacji giełdowych, Warszawa: CeDeWu.
46. Reeves C.A., Bednar D. A., *Defining quality: alternatives and implications*, Academy of management Review, 19(3), 1994, https://doi.org/10.2307/258934.

47. Rura-Polley, T. and Clegg, S. (1999) 'Managing collaborative quality: a challenging innovation', *Creativity and Innovation Management*, vol. 8(1), pp. 37–47, https://doi.org/10.1111/1467-8691.00117.
48. Sayed, S.A. (2015) 'Should analysts go by the book? Valuation models and target price accuracy in an emerging market', *Global Business Review*, vol. 16(5), pp. 832–844, https://doi.org/10.1177/0972150915591626.
49. Seawright, K.W. and Young, S.T. (1996) 'A quality definition continuum', *Interfaces*, vol. 26(3), pp. 107–113, [Online], Available: https://www.jstor.org/stable/25062136 [14 Apr 2026].
50. Shroff, P.K., Venkataraman, R. and Xin, B. (2004) Leaders and followers among security analysts: Analysis of impact and accuracy, [Online], Available: https://www.academia.edu/download/80881442/download.pdf [14 Apr 2026].
51. Szutkowska J. (2016), Definicja jakości w statystyce publicznej, [Online], Available: www.bip.stat.gov.pl/dzialalnosc-statystyki-publicznej/jakosc-w-statystyce/definicja-jakosci-w-statystyce-publicznej [14 Apr 2026].
52. Tiberius V. and Lisiecki L. (2019) 'Stock price forecast accurancy and recommendation profitability of financial magazines', *International Journal of Financial Studies*, vol. 7(4), 58, https://doi.org/10.3390/ijfs7040058.
53. Urząd Komisji Nadzoru Finansowego, Stanowisko Urzędu Komisji Nadzoru Finansowego w sprawie świadczenia usługi robo-doradztwa, [Online], Available: www.knf.gov.pl/knf/pl/komponenty/img/Stanowisko_UKNF_ws_swiadczenia_uslugi_robo_doradztwa_71303.pdf [14 Apr 2026].
54. Włodarczyk, K. (2020) Analiza wpływu rekomendacji maklerskich na podejmowanie decyzji inwestycyjnych, praca doktorska, Politechnika Warszawska.
55. Wnuczak, P. (2021) 'Profitability of investment strategies developed on the basis of buy and sell recommendations', *Journal of Economics and Management*, vol. 43(1), pp. 317–338, https://sbc.org.pl/Content/506616/15.pdf.
56. Wróblewski, M. (2016) 'Problemy trafności prognozowania przepływów pieniężnych w wycenie spółek giełdowych metodą DCF', *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, nr 307, *Współczesne Finanse*, 8, pp. 128–142, [Online], Available: https://www.ue.katowice.pl/fileadmin/user_upload/wydawnictwo/SE_Artyku%C5%82y_291_320/SE_307/10.pdf [14 Apr 2026].
57. Yezegel, A. (2015) 'Why do analysts revise their stock recommendations after earnings announcements?', *Journal of Accounting and Economics*, vol. 59(2–3), pp. 163–181, https://doi.org/10.1016/j.jacceco.2015.01.001.
58. Zarzecki, D. (2024) 'Evaluation of key parameters used by financial analysts to value companies in Poland', *European Research Studies Journal*, vol. 27(Special Issue A), pp. 839–856, https://ersj.eu/journal/3751.

Appendix 1

Dynamic and static accuracy coefficients of recommendations for individual companies at the end of the recommendation validity period

Company	Static accuracy coefficient [%]	Dynamic accuracy coefficient [%]
Enea	12.1	40.7
Sygnity	15.2	48.1
PGE	16.2	44.8
CEZ	17.3	38.8
Alior Bank	17.7	46.8
ECHO	18.3	40.0
Bogdanka	20.5	39.8
PKO	20.9	40.3
Tauron	22.1	44.2
Millennium	22.5	53.5
DOMDEV	23.0	56.3
PGNIG	23.7	57.2
Orange	24.6	62.9
PZU	25.0	54.5
Trakcja	25.4	49.3
mBank	25.5	58.8
Cyfrowy Polsat	25.6	59.7
Asseco Poland	25.7	48.6
Elektrobudowa	26.2	41.0
Police	26.4	29.7
KGHM	26.7	54.7
Ciech	27.4	49.2
Agora	28.6	54.4
Netia	28.6	55.6
Handlowy Bank	28.7	53.2
Pekao	28.8	56.4
Lotos	30.0	56.7
PKN Orlen	31.3	59.7
GTC	31.3	51.1

Appendix 1 cont.

Company	Static accuracy coefficient [%]	Dynamic accuracy coefficient [%]
PBG	32.9	55.3
Comarch	33.0	58.3
LPP	33.9	64.4
Polimex Mostostal	34.0	63.3
Azoty Group	34.2	60.5
Getin Holding	34.3	74.3
Puławy	34.3	57.8
ING Bank	35.1	53.5
MOL	35.8	53.1
Budimex	39.1	60.1
Santander	39.2	66.0
Amrest	41.3	71.3
Kęty	42.1	69.0
Eurocash	43.0	63.9
CCC	43.4	69.8
JSW	48.0	59.1
Average	29.0	55.0

Summary

The aim of this article is to provide an empirical analysis of investment recommendation accuracy on Polish Stock Exchange. The study also intended to defining the concept of investment recommendation quality and classifying the dimensions through which it can be assessed. Based on data from 2005 to 2019 the accuracy of investment recommendation was calculated using four approaches. All four indicators indicate low level of accuracy of investment recommendation in Poland.

JEL codes: G11, G14, G24

Keywords: *investment recommendation, accuracy, quality, effectiveness*

Managerial Economics
2026, Vol. 27, No. 1, pp. 85–116
https://doi.org/10.7494/manage.2026.27.1.85

Milena Suliga*

Linear vs. threshold cointegration approaches to price discovery: The case of the Warsaw Stock Exchange

1. Introduction

Financial markets are continuously exposed to new information, which exerts a significant impact on the price formation process. According to the informational efficiency hypothesis proposed by Fama (Fama 1970), in a fully efficient market, asset prices should instantaneously reflect all available information. The complete absorption of new information into the price of a given instrument is referred to as *price discovery*. In practice, however, the vast majority of markets fail to meet the assumptions of strong-form informational efficiency, and the adjustment of prices to new information arriving in the market is a gradual process that unfolds over time.

Due to the specific nature of derivative instruments, their prices are closely linked to the prices of the corresponding underlying assets traded in the spot market. Incoming market information usually triggers similar investor reactions in both markets (spot and futures). The strong connection between these two markets ensures that any temporary mispricing is exploited by arbitrageurs, whose activity restores equilibrium and guarantees price consistency.

Nevertheless, price changes in the two markets are not perfectly synchronized. In practice, new information may be absorbed more quickly by either the spot or the futures market. When the price of the underlying asset or derivative adjusts first, it is assumed that the other price will follow in a short period of time. Such relationships are associated with the existence of Granger causality between spot and futures prices – past price changes in one market may improve

* AGH University of Krakow, Krakow, Poland, e-mail: msuliga@agh.edu.pl, ORCID: 0000-0001-5719-5679

the forecasting accuracy of price changes in the other, compared to forecasts made without that information.

In the literature, it is generally assumed that the market that first incorporates new price-relevant information plays the leading role in the price discovery process. It should be noted that fulfilling the price discovery function is regarded as one of the key functions of the futures market. A mature and liquid futures market is expected to actively participate in price discovery and contribute to the efficiency of the spot market. For this reason, the analysis of the price discovery process constitutes an important element in assessing the maturity of a derivatives market as well as its impact on the efficiency and functioning of the spot market.

One of the most commonly employed methods of analyzing the price discovery process is the use of a bivariate vector error correction model (VECM). This model enables the identification of both long-run and short-run causal relationships between the prices of futures contracts and their corresponding spot market instruments. This approach has been applied, among others, by Bohl and colleagues (Bohl et al. 2011), Mutlu and Arik (Mutlu, Arik 2015), Marcinkiewicz and Kompa (Marcinkiewicz, Kompa 2013), Chen and colleagues (Chen et al. 2019, 2021), and Suliga (Suliga 2025). However, the classical VECM relies on assumptions that may limit its usefulness in studies of price discovery, particularly in the context of the complex and dynamic nature of financial markets.

First, the model assumes linear and time-invariant relationships between price changes in the two markets. This implies that the response of one variable to the other remains unchanged throughout the entire sample period. In practice, especially when the estimation horizon spans several years, this assumption may be unrealistic. Markets are subject to structural shocks such as financial crises, regulatory changes, or macroeconomic turbulence, which may permanently alter the relationships between derivatives and their underlying assets. The VECM framework does not account for such changes.

Second, the model presumes that the error correction mechanism is independent of both the direction and the magnitude of deviations from the long-run equilibrium. In reality, however, adjustment may occur only after a certain threshold is exceeded; for instance, when the deviation becomes economically significant from the perspective of market participants. Moreover, asymmetric responses are possible; for example, the futures market may react more rapidly to overpricing than to underpricing of the derivative. Such nonlinear and state-dependent reactions cannot be captured by the classical VECM.

Given these limitations, it appears reasonable to apply a nonlinear threshold vector error correction model (TVECM) in the study of price discovery. This model allows the adjustment mechanism to vary depending on the magnitude and direction of deviations from equilibrium, making it possible to identify situations in which the error correction mechanism actually becomes active. As a result, the

TVECM framework may more accurately reflect real market conditions such as transaction costs, limited liquidity, or delayed arbitrage responses. Thus, it offers a more flexible and precise tool for analyzing the price discovery process.

The objective of this study is to compare the usefulness of the classical bivariate VECM and its nonlinear threshold version (TVECM) in analyzing the price discovery process on the Warsaw Stock Exchange. The analysis is conducted for the most liquid derivative instruments traded on the exchange, namely WIG20 index futures and their corresponding underlying index.

The study employs daily data from January 2018 to December 2024. During this period, two major events severely disrupted the domestic financial market and led to a sharp increase in investment risk: the outbreak of the COVID-19 pandemic and the onset of the war in Ukraine. Both during the initial phase of the pandemic and in the first weeks of the war, the WIG20 index and its associated futures contracts experienced steep declines. It can be assumed that the structural shocks triggered by these events also influenced the relationship between the spot and the futures markets. In such a context, the results of a classical linear VECM estimated on data from 2018 to 2024 would likely be distorted and would fail to reliably capture the causal relationships between the analyzed markets.

The comparison of VECM and TVECM outcomes aims to determine whether a nonlinear framework – accounting for market reactions that vary with the magnitude and direction of deviations from equilibrium – can more accurately capture the price discovery mechanism under conditions of heightened volatility and structural disturbances. In particular, the questions are whether the threshold model better reflects actual market behavior during crisis periods, and whether it provides a more effective analytical tool than the classical linear specification.

The remainder of this paper is structured as follows. Section 2 presents a review of the literature on the price discovery process, with particular emphasis on the application of the VECM and TVECM models as tools for investigating causal linkages between derivative markets and their underlying instruments. Section 3 describes the data and outlines the methodological framework. Section 4 presents and discusses the empirical results. Section 5 concludes with a summary, final remarks and a discussion of the study's limitations.

2. Literature review

2.1. Previous research findings from various markets worldwide

A mature futures market, characterized by high liquidity, should play an important role in the process of price discovery. The dominant role of the derivatives market in this process has been confirmed by studies conducted with

respect to the United States market (e.g., Hasbrouck 2003; Chou, Chung 2006), the United Kingdom (Brooks et al. 2001; Gwilym, Buckle 2001), Germany (Booth et al. 1999; Gaul, Theissen 2008), and France (Alphonse 2000; Buckle et al. 2019).

The features that enable a mature futures market to reflect new incoming information faster than the spot market include, among others, lower transaction costs, the leverage inherent in derivative instruments, the absence of short-sale restrictions, and the significant share of institutional investors (cf. Theissen 2002; Bohl et al. 2011; Mutlu, Arık 2015; Fassas, Siriopoulos 2019).

In the case of emerging markets, previous research on the contribution of futures contracts to the price discovery process differs considerably across markets. For example, with regard to the Korean market, Min and Najand found that index futures began to dominate the price discovery process almost immediately after their introduction (Min, Najand 1999). By contrast, as demonstrated by Guo and colleagues, on the Chinese market, along with the development of the futures segment, futures contracts gradually took over the leading role in price discovery from the spot market (Guo et al. 2013). Similarly, in the case of the Greek market, early research on futures contracts indicated inefficiency, with futures prices appearing to be biased as forecasts of spot prices (cf. Kenourgios 2005; Andreou, Pierides 2008). However, studies conducted in later periods demonstrated the dominant role of futures in the price discovery process on that market (cf. Kavussanos et al. 2008; Kavussanos, Visvikis 2011).

Research on the price discovery process on the Warsaw Stock Exchange (WSE) is relatively scarce. An analysis concerning the early years of the WSE's derivatives market (1998–2009) was conducted by Bohl and colleagues (Bohl et al. 2011). Using daily closing prices of the WIG20 index and the corresponding futures contracts, and applying a bivariate VECM, they showed that the spot market played the leading role in price discovery during the analyzed period. By splitting the sample into two subperiods, they also demonstrated that after investment funds were allowed to participate in the futures market in 2004 – which significantly increased the share of institutional investors in derivatives trading – the role of WIG20 futures in price discovery rose. While the average contribution of the futures market to price discovery was estimated at around 16% during 1998–2004, it was assessed at around 35% in 2005–2009.

Causal relationships between the WIG20 index and its futures contracts were also investigated by Marcinkiewicz and Kompa using VAR and VECM models (Marcinkiewicz, Kompa 2013). The authors employed both daily data and high-frequency data (5-, 15-, and 30-minute intervals) from 2008 to 2011. In a summary table, they presented the direction and significance of the identified causal linkages. Their finding that the error correction coefficient was significant in each of the estimated VECM models only in the equation describing futures

price changes was, however, followed by an incorrect conclusion: "this means that in the long run causality runs from futures to spot." The conclusion should be the opposite, since the significance they reported indicates that the WIG20 index is the Granger cause of futures prices. In light of this interpretation and the absence of full estimation results, the evidence provided by Marcinkiewicz and Kompa may not be sufficient to draw firm conclusions about the contribution of index futures to price discovery on the Warsaw Stock Exchange in 2008–2011 (Marcinkiewicz, Kompa 2013). Writing in the conclusion of their work that the futures market reacted faster than the spot market to incoming information and that the influence of the index market on futures was weaker than the reverse causal relationship, they likely overstated the role of the futures market in price discovery during the analyzed period.

Studies on the price discovery process on the Warsaw Stock Exchange were also conducted by Mutlu and Arik (Mutlu, Arik 2015) and Suliga (Suliga 2025). However, these studies focused on causal relationships between the prices of single-stock futures (SSFs) and the prices of their underlying shares. Both studies employed a bivariate VECM, and both indicated that it was the stock market, rather than the futures market, that played the leading role in price discovery.

Mutlu and Arik, in addition to the Polish market, examined the Russian, Korean, and Indian markets (Mutlu, Arik 2015). With respect to Poland, for twenty pairs of instruments (SSF contracts and their underlying stocks), they used daily data covering the period from the introduction of each series of contracts up to August 15, 2014, as well as 60-minute data from April 1 to August 15, 2015. The authors did not include the full estimation results of the VECM models in their paper. They provided only the average contributions of the spot and futures markets to price discovery, estimated on the basis of those models. For the Polish SSF market, this contribution was reported to be around 40% when daily data were used and around 38% when intraday data were applied.

Suliga carried out an analysis of the price discovery process for single-stock futures with the highest contract multiplier (1000) and their underlying shares (Suliga 2025). The study covered the years 2020–2023 and was conducted using daily data. The results demonstrated that in most cases, stock prices were the Granger cause of futures prices, both in the context of short-run and long-run relationships. Causality running from stock prices to futures prices was found in the long run for only three companies and in the short run for only one out of the eight analyzed firms. In particular, Suliga's study showed that a high SSF contract multiplier (and thus higher leverage) does not guarantee the dominance of these contracts in the price discovery process, especially when futures trading volumes are relatively low (Suliga 2025).

2.2. VECM and TVECM models as tools for examining the price discovery process

The classical bivariate vector error correction model (VECM) has been employed in numerous studies conducted so far relating to the price discovery process. This model has been applied to the analysis of causal relationships between the spot and futures markets in relation to the Chinese (Hou, Li 2013; Liu, Qiao 2017; Xu 2018), Greek (Fassas, Siriopoulos 2019), Indian (Kumar, Tse 2009; Mutlu, Arik 2015; Curran et al. 2020), Korean (Mutlu, Arık 2015; Kang et al. 2016), Polish (Bohl et al. 2011; Marcinkiewicz, Kompa 2013; Mutlu, Arik 2015; Suliga 2025), Russian (Mutlu, Arik 2015), and Taiwanese (Chen et al. 2019; Chen et al. 2021) markets.

In the classical VECM, it is assumed that the structure of the cointegrating relationship remains stable throughout the adopted research period and that the causal relationships between the analyzed time series are linear. In practice, these assumptions are not always met, and the economic realism of the model is simplified. For instance, during a long research period, the nature of the cointegrating relationship may change as a result of a crisis or significant regulatory shifts that alter the structure of participants in the futures market. In some studies on price discovery, the research period is therefore divided into subperiods, and the VECM is estimated separately for each (e.g., Bohl et al. 2011; Marcinkiewicz, Kompa 2013; Xu 2018; Chen et al., 2021). This division usually serves the purpose of testing whether an important event significantly influenced the strength and direction of causal relationships between the spot and futures markets.

The linear VECM also assumes that the process of returning to equilibrium is symmetric and linear, regardless of whether the deviation from the long-run equilibrium is positive or negative. In reality, however, it is possible that the response to disequilibria between markets is asymmetric; for example, prices may adjust more rapidly in cases of undervaluation than in cases of overvaluation (or vice versa). Applying the VECM with a division of the research period into subperiods does not solve this problem and does not allow for the detection of potential asymmetries in market mechanisms. Modeling such asymmetries is possible through the application of the threshold vector error correction model (TVECM), originally introduced by Balke and Fomby (Balke, Fomby 1997) and further developed by Hansen and Seo (Hansen, Seo 2002), which extends the classical VECM. In this model, the concept of a threshold (or thresholds) is introduced, assuming that once the threshold is exceeded, the error correction mechanism changes.

In research on price discovery and causal relationships between derivative prices and the prices of their underlying assets, the TVECM has been employed by, among others, Martens and colleagues (Martens et al. 1998), Kim and colleagues (Kim et al. 2010), Mamatzakis and Remoundos (Mamatzakis, Remoundos 2010), Ters and Urban (Ters, Urban 2020), and Xi and colleagues (Xi et al. 2023).

Mamatzakis and Remoundos (Mamatzakis, Remoundos 2010) applied a threshold VECM with a single threshold to analyze cointegration between spot and futures prices of Brent crude oil, using data from 1990 to 2009. Their results confirmed the existence of cointegration, with two distinct regimes emerging. Applying a threshold model instead of the classical linear VECM allowed them to more precisely capture the relationships between the examined markets.

Xi and colleagues (Xi et al. 2023) employed both the classical VECM and a threshold VECM with a single threshold to study the efficiency of the CO_2 futures market in the second phase of the European Union Emissions Trading System (2008–2012). Their findings suggest that the analyzed derivatives played an important role in the price discovery process during the research period and that the use of past futures prices could improve spot price forecasts. By employing the threshold model, Xi and colleagues (Xi et al. 2023) found that, in the short run, the futures market under study was not fully efficient. The error correction mechanism was activated only when the basis fell below −0.07. Futures prices exhibited very slow adjustments to long-run equilibrium in the presence of deviations. The authors attributed this phenomenon to the global financial crisis of 2008 and its aftermath.

Ters and Urban presented a methodology for estimating a three-regime TVECM with an unknown cointegrating vector, based on a novel dynamic grid evaluation (Ters, Urban 2020). The authors highlighted that applying a three-regime threshold model made it possible to estimate the range in which arbitrageurs have no incentive to trade, meaning that the error correction mechanism is not activated. It becomes active only when the basis exceeds a critical threshold, making potential profits from arbitrage greater than the necessary transaction costs. Only then can one expect arbitrageurs to enter the market and execute transactions. Such a mechanism supports the presence of nonlinear adjustment dynamics and the existence of three regimes, with the middle regime corresponding to the no-arbitrage band. These regimes may differ significantly in terms of the characteristics of the relationships between cointegrated spot and futures markets. In the final part of their study, Ters and Urban also presented two short applications of the TVECM with two thresholds, relating to the palladium market and to S&P 500 index futures (Ters, Urban 2020).

A three-regime TVECM was also applied by Martens and colleagues and Kim and colleagues in their studies of causal relationships between the S&P 500 index and its futures contracts (Martens et al. 1998; Kim et al. 2010). Their choice of model was justified similarly to that of Ters and Urban (Ters, Urban 2020), arguing that arbitrageurs enter the market only when the mispricing between the spot and futures markets is large enough to compensate for the transaction costs they incur and the associated risks of interest rate and dividend changes. If the error correction term ec_t deviates only slightly from zero, arbitrage is not profitable. For this reason, the authors considered the use of a three-regime threshold model to be appropriate for studying the price discovery process, with the middle regime corresponding to $|ec_t| \approx 0$.

The results obtained by Martens and colleagues (Martens et al. 1998) showed that in their research period (S&P 500 futures contracts expiring in June and December 1993), the impact of the futures market on the spot market was greater when the error correction term was negative, and that the influence of deviations from long-run equilibrium on current returns increased with the degree to which futures prices diverged from their theoretical values.

An analysis of the nonlinear dynamics of causal relationships between the S&P 500 index and its futures contracts was also conducted by Kim and colleagues using a three-regime TVECM (Kim et al. 2010). The authors identified two thresholds defining the no-arbitrage band (small deviations of prices from the long-run equilibrium relationship), consistent with the cost-of-carry model Kim et al. 2010).

Kim and colleagues found that when prices move outside this band, i.e., when the error correction term is sufficiently large in absolute value, spot and futures price series become stationary mean-reverting processes (Kim et al. 2010). Prices within the no-arbitrage band, however, were found to be non-stationary. Furthermore, Kim and colleagues (Kim et al. 2010) confirmed the earlier findings of Martens and colleagues (Martens et al. 1998), namely that futures prices dominate in the price discovery process (the index adjusts to futures prices, not the other way around).

Taking into account the arguments of Ters and Urban (Ters, Urban 2020) as well as the results obtained by Martens and colleagues (Martens et al. 1998) and Kim and colleagues (Kim et al. 2010), the threshold VECM with two thresholds (three regimes) appears to be a more suitable tool for analyzing the price discovery process in the spot and futures markets than the model with two regimes (a single threshold). In this study, therefore, the classical VECM will be compared in terms of its usefulness for analyzing price discovery on the Warsaw Stock Exchange with the three-regime threshold model. The information criterion (AIC) further confirmed that the TVECM with two thresholds provides a better fit to the analyzed data than the model with a single threshold.

3. Data and research methodology

This section presents the data and methodology used in the empirical analysis. It begins with a description of the dataset, including the WIG20 index and its futures contracts, followed by the specification of the econometric models applied to examine the price discovery process.

3.1. Data used

The study employed daily closing prices of the WIG20 index and its futures contracts for the period from January 2, 2018, to December 30, 2024. Futures contracts on the WIG20 index are the most liquid instruments on the Warsaw Stock Exchange's derivatives market. In the years 2018–2024, the average annual turnover value of these derivatives amounted to PLN 262,695 million, which represented on average 90.92% of the annual turnover of all futures contracts traded on the Warsaw Stock Exchange. The multiplier of WIG20 futures contracts has a value of 20, which means that a change of one index point in the underlying index results in a change of PLN 20 in the contract value. Each quarter, a new series of WIG20 futures contracts is introduced into trading, which remains listed on the exchange for one year. The expiration days of the derivatives are the third Fridays of March, June, September, and December. Thus, at any given time, four consecutive contract series are simultaneously traded. The series of WIG20 contracts with the nearest expiration date usually exhibits the highest liquidity, while on the expiration day of that series, the subsequent series takes over the dominant position in terms of liquidity.

In order to conduct the analysis of the price discovery process using the VECM and TVECM models, it was necessary to construct a uniform time series covering daily closing prices of futures contracts from January 2018 to December 2024. This series was built on the basis of closing prices of the contracts with the nearest delivery date, whereby the transition to the subsequent series was set at the expiration day of the current one. This method of constructing the series makes it possible to obtain the most reliable representation of prices, taking into account trading volume and the number of open positions in the futures market (see Fassas, Siriopoulos 2018).

Figure 1 presents the time series used in the study. The blue line (WIG20) represents the series of WIG20 index closing prices, while the red line (FW20) represents the series of daily closing prices of the futures contracts on this index, constructed according to the principles described above. Figure 1 illustrates the close relationship between spot and futures price levels, which is reflected in the nearly parallel trajectories of their time series. In many instances, these series overlap, and the differences visible at certain points in time remain relatively small.

Figure 1. Daily closing prices of the WIG20 index and WIG20 futures (January 2018 – December 2024).

The conducted study used logarithms of closing prices of the WIG20 index (S_t) and its futures contracts (Ft), as well as daily log returns. The following notations of log prices were adopted:

$$s_t = ln(S_t), f_t = ln(F_t) \tag{1}$$

and of log returns:

$$\Delta s_t = s_t - s_{t-1}, \ \Delta f_t = f_t - f_{t-1} \tag{2}$$

Table A1 in the Appendix reports the basic descriptive statistics for the log returns of the WIG20 index and its futures contracts over the sample period. The interpretation of these results is provided directly below the table. As indicated there, the values of the individual descriptive statistics are very similar across both markets.

3.2. Research methodology

According to the cost-of-carry model, which is the classical economic model used for the valuation of derivative instruments and which assumes the absence

of arbitrage opportunities, the fair price F_t of a futures contract at time t is expressed by the formula:

$$F_t = S_t e^{(r-d)(T-t)} \tag{3}$$

In this equation, S_t denotes the price of the underlying instrument, r is the risk-free interest rate, d is the dividend yield, while $T - t$ represents the time remaining until the expiration of the derivative.

Logarithmizing equation (3) on both sides allows us to observe that the cost-of-carry model assumes the existence of a linear relationship between the log prices of spot and futures:

$$f_t = s_t + (r-d)(T-t) \tag{4}$$

In econometric models, the variability of returns, dividends, and the time remaining until contract expiration is often omitted. As a result, relation (3) is simplified by assuming that the time series s_t and f_t are cointegrated of order (1,1). This assumption means that both series are integrated of the first order (that is, they are non-stationary with a stochastic trend, while their first differences, i.e., the log returns Δs_t, Δf_t are stationary), and that there exists a linear combination of them, ec_t, which is stationary. Without loss of generality, one may assume that in this combination, one of the coefficients equals 1, and therefore, for example, that the futures price is a linear function of the spot price:

$$f_t = \beta_1 s_t + ec_t \tag{5}$$

From equation (5), it follows that the stationary linear combination ec_t, called the error correction term, can be expressed as:

$$ec_t = f_t - \beta_1 s_t \tag{6}$$

The vector of coefficients of this combination, $[1, -\beta_1]$, is called the cointegrating vector.

Classical VECM Model

The classical two-dimensional Vector Error Correction Model (VECM), used to describe the relationships between cointegrated spot and futures prices, takes the following form:

$$\Delta f_t = \mu_f + \alpha_f ec_{t-1} + \sum_{j=1}^{p} \gamma_{fs,j} \Delta s_{t-j} + \sum_{j=1}^{p} \gamma_{ff,j} \Delta f_{t-j} + \varepsilon_{f,t} \tag{7}$$

$$\Delta f_t = \mu_f + \alpha_f ec_{t-1} + \sum_{j=1}^{p} \gamma_{fs,j} \Delta s_{t-j} + \sum_{j=1}^{p} \gamma_{ff,j} \Delta f_{t-j} + \varepsilon_{f,t} \tag{8}$$

In the general case, the *n*-dimensional VECM model takes the following matrix form:

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta Y_{t-j} + \varepsilon_t \tag{9}$$

where $Y_t = \left[y_{t1}, y_{t2}, \ldots, y_{tn} \right]$ is a vector of non-stationary time series.

In order to examine the existence of cointegration between the non-stationary time series s_t and f_t, and thus confirm the validity of employing the VECM model expressed by equations (7) and (8), one may apply the Johansen test (Johansen 1991, 1992), which is based on the estimation results of the VECM model and requires the computation of two test statistics:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} \ln(1 - \lambda_i) \tag{10}$$

$$\lambda_{max}(r) = -T \ln(1 - \lambda_{r+1}) \tag{11}$$

The λ_i values appearing in equations (10) and (11) are the eigenvalues of the matrix Π from the general form (9) of the *n*-dimensional VECM model, ordered in ascending sequence. By means of the trace statistic $\lambda_{trace}(r)$, one tests the null hypothesis that the number of distinct cointegrating vectors is not greater than *r* against the alternative that there are more. The maximum eigenvalue statistic $\lambda_{max}(r)$, on the other hand, is used to verify the null hypothesis of the existence of exactly *r* cointegrating vectors, against the alternative that there are $r + 1$. Naturally, in the case of the two-dimensional model described by equations (7) and (8), at most one cointegrating vector may exist. Hence, performing the Johansen test will serve to confirm that such a vector exists, and thus that spot and futures price series are cointegrated of order (1,1).

In the VECM model defined by equations (7) and (8), the coefficients α_s and α_f accompanying the error correction term ec_{t-1} measure the speed of price adjustment in each market following a disturbance in the long-run equilibrium relation defined by equation (5). For example, a value of α_s significantly different from zero would indicate that, in the event of a disruption of this relation, a significant price adjustment takes place in the spot market. This would signal the existence of a significant long-run causal relationship running from the futures market to the spot market, pointing to a substantial contribution of the futures market to the price discovery process. Conversely, a significant value of the coefficient α_f would indicate an analogous causality running in the opposite direction, and thus a significant contribution of the spot market to price discovery, considered in the context of long-run causal relationships.

The examination, based on the VECM model, of the existence of short-run causal relationships between the prices of the derivative and the prices of its corresponding underlying instrument will be based on determining the significance of the parameters $\gamma_{sf,j}$ ($\gamma_{fs,j}$) appearing in equations (7) and (8) with the lagged values of returns. Confirmation of the joint significance of the parameters $\gamma_{sf,j}$ in equation (7) will mean that futures prices are the Granger cause of spot prices. Conversely, the existence of short-run causality running in the opposite direction will be indicated by the significance of the parameters $\gamma_{fs,j}$ in equation (8).

TVECM with three regimes

As mentioned earlier, the TVECM model was introduced by Balke and Fomby (Balke, Fomby 1997) and later extended by Hansen and Seo (Hansen, Seo 2002). The model employs a regime-based approach that allows the adjustment process to vary depending on the size of the deviations from the long-run equilibrium. Estimation of the three-regime TVECM proceeds as follows. In the first step, the cointegrating relation (5) is estimated and, based on the coefficients of the cointegrating vector obtained, the values of the error-correction term ec_t are computed. Subsequently, depending on the value of ec_{t-1}, the values of the time series s_t and f_t are partitioned into three regimes (R_{-1}, R_0, R_1) according to the rule:

$$\begin{cases} ec_{t-1} < \tau_1 \Rightarrow s_t, f_t \in R_{-1} \\ \tau_1 \le ec_{t-1} \le \tau_2 \Rightarrow s_t, f_t \in R_0 \\ ec_{t-1} > \tau_2 \Rightarrow s_t, f_t \in R_1 \end{cases} \tag{12}$$

where $\tau_1 < 0, \tau_2 > 0$ are threshold values defined in such a way that:

$$0.05 \le P\left(ec_{t-1} \in R_i\right) \le 0.95 \text{ for } i \in \{-1,0,1\}, \tag{13}$$

where P denotes probability. The selection of the optimal thresholds τ_1, τ_2 is based on searching a preliminary grid of values. For each combination of potential thresholds, the parameters of the TVECM model are estimated, and the most optimal of those are chosen, which minimize the sum of squared residuals of the model.

The lower regime R_{-1} will include those observations s_t and f_t, for which, on the previous day ($t - 1$), the futures price was sufficiently undervalued relative to the spot price, whereas the upper regime R_1 will consist of those values s_t, f_t, for which, on the preceding day, the derivative price was correspondingly too high relative to the price of the underlying instrument. The middle regime, in turn, comprises those values of the time series s_t and f_t, for which the error correction term ec_{t-1} falls between the thresholds defining the lower and upper regimes,

meaning that deviations from the long-run equilibrium described by equation (5) are moderate.

After partitioning the data into three regimes, the VECM model is estimated in each of them separately with the same lag order p. The TVECM model can therefore be described by the following equations:

$$\Delta s_t = \mu_{s,i} + \alpha_{s,i} ec_{t-1} + \sum_{j=1}^{p} \gamma_{ss,j,i}\, \Delta s_{t-j} + \sum_{j=1}^{p} \gamma_{sf,j,i}\, \Delta f_{t-j} + \varepsilon_{s,t,i} \tag{14}$$

$$\Delta f_t = \mu_{f,i} + \alpha_{f,i} ec_{t-1} + \sum_{j=1}^{p} \gamma_{fs,j,i}\, \Delta s_{t-j} + \sum_{j=1}^{p} \gamma_{ff,j,i}\, \Delta f_{t-j} + \varepsilon_{f,t,i} \tag{15}$$

where $i \in \{-1,0,1\}$,

To prevent overparameterization, the optimal lag length p in both the VECM and TVECM models is selected based on the Bayesian Information Criterion (BIC), with adjustments made if autocorrelation of the residuals is observed. According to the results obtained by Martens and colleagues (Martens et al. 1998), it may be expected that in the TVECM model, the impact of the error correction term on the values of index and futures returns will be considerably smaller in the middle regime, corresponding to the no-arbitrage band, than in the lower and upper regimes. One may therefore expect the following inequalities: $|\alpha_{s,-1}| > |\alpha_{s,0}|$ and $|\alpha_{s,1}| > |\alpha_{s,0}|$ in the equation describing the changes in spot prices as well as $|\alpha_{f,-1}| > |\alpha_{f,0}|$ and $|\alpha_{f,1}| > |\alpha_{f,0}|$ in the equation relating to futures price movements.

In the final step of the analysis, after estimating the VECM and TVECM, the measures of the average contribution of the spot and futures markets to the price discovery process, the common factor weights, were calculated. These measures were first defined by Schwarz and Szakmary (Schwarz, Szakmary 1994) and have subsequently been employed in other studies addressing the issue under consideration (see, e.g., Bohl et al. 2011; Mutlu, Arik 2015; Fassas, Siriopoulos 2019; Suliga 2025):

$$\theta_s = \frac{|\alpha_f|}{|\alpha_s| + |\alpha_f|} \tag{16}$$

$$\theta_f = 1 - \theta_s = \frac{|\alpha_s|}{|\alpha_s| + |\alpha_f|}. \tag{17}$$

As follows from equations (16) and (17), the common factor weights are based solely on the adjustment coefficients α_s and α_f, and therefore reflect the average

contribution of a given market to the price discovery process only in the context of long-run (error-correction) causal linkages. How should these values be interpreted? The larger the ratio of α_f to α_s in absolute terms, the stronger the response of the futures market (relative to the spot market) to a disturbance from the long-run equilibrium relationship between prices in the two markets. The market that reacts more strongly does not dominate price discovery; rather, it tends to follow the other market, adjusting to changes that have occurred there earlier. A larger absolute ratio α_f / α_s thus translates into a higher value of θ_s, which indicates the spot market's contribution to price discovery. In the abovementioned research, the measures θ_s and θ_f have been computed based on estimates from the classical bivariate VECM. In the present study, they will also be computed within each of the regimes of the threshold VECM (TVECM).

4. Empirical results

The empirical investigation, the results of which are presented in this chapter, followed the structure outlined below. At the outset, whether the analyzed time series satisfy the fundamental assumptions regarding stationarity and cointegration was verified, which are necessary for the application of error-correction models. The main part of the study, conducted after validating these assumptions, was divided into two stages. In the first stage, a classical bivariate vector error correction model was fitted to the data, on the basis of which preliminary conclusions were drawn concerning the contribution of the spot and futures markets to the price discovery process over the period under study. In the second stage, a threshold VECM with three regimes was estimated using the same dataset. The conclusions regarding the roles of the spot and futures markets, derived from the threshold model, were then compared with the results of the first stage of the study in order to assess their consistency.

The results of the Augmented Dickey–Fuller (ADF) test, reported in Table A2 in the Appendix, indicate that the log price series s_t and f_t are nonstationary in levels but stationary in first differences, confirming that they are integrated of order one, I(1).

Table 1 presents the results of the cointegration tests for the series s_t and f_t using Johansen's procedure. The values of both test statistics, λ_{trace} and λ_{max}, clearly indicate that the null hypothesis of no cointegrating vector should be rejected in favor of the alternative hypothesis that the series s_t and f_t are cointegrated. The final column of the table reports the estimated coordinates of the cointegrating vector.

Table 1

Cointegration test results

Johansen test statistics values				Cointegrating vector
λ_{trace}		λ_{max}		
$H_0 : r = 0$	$H_0 : r \leq 1$	$H_0 : r = 0$	$H_0 : r = 1$	
40.90***	5.64	35.27***	5.64	[1; −0.987]

Note: The symbol *** denotes statistical significance at the 0.01 level

The results presented in Table A2 and Table 1 confirm that the data satisfy the fundamental assumptions of the VECM framework. In the next step of the analysis, the parameters of the classical VECM specified by equations (7) and (8) were therefore estimated. The Bayesian Information Criterion (BIC) indicated an optimal lag order of 2; however, due to the presence of autocorrelation in the residuals, the lag order was increased to $p = 3$, which resolved this issue. Residuals were examined for heteroskedasticity, which was detected; therefore, parameter significance was assessed using robust Newey–West standard errors. The estimation results of the model are reported in Table 2.

With respect to the coefficients on the error-correction term, only the parameter α_f differs significantly from zero (at the 0.05 significance level). In the context of long-run causal relationships between spot and futures prices, this result indicates the existence of significant unidirectional causality running from the WIG20 index to its futures contracts. This suggests that during the analyzed sample period, in situations where the long-run equilibrium relationship between spot and futures prices was disrupted, significant price adjustment occurred exclusively in the futures market. Hence, it was the spot market that initiated the disequilibrium, while the derivatives market followed the price changes of the index.

Table 2

Estimation results of the VECM model

Equation	Coefficient values								C.f.w.
Δs_t	μ_s	α_s	$\gamma_{ss,1}$	$\gamma_{ss,2}$	$\gamma_{ss,3}$	$\gamma_{sf,1}$	$\gamma_{sf,2}$	$\gamma_{sf,3}$	θ_s [%]
	0.007	−0.065	0.190	0.038	0.031	−0.202	−0.044	0.022	68.33
Δf_t	μ_f	α_f	$\gamma_{ff,1}$	$\gamma_{ff,2}$	$\gamma_{ff,3}$	$\gamma_{fs,1}$	$\gamma_{fs,2}$	$\gamma_{fs,3}$	θ_f (%)
	0.014**	−0.127**	−0.634***	−0.248	−0.034	0.621***	0.256	0.085	31.67

Note: The symbols *** and ** indicate statistical significance at significance levels of 0.01 and 0.05, respectively, based on Newey–West robust standard errors. The last column (C.f.w.) contains the values of the common factor weights

The coefficient $\gamma_{fs,1,}$ which appears in the equation describing changes in futures prices, is statistically significant (at the 0.01 significance level). This indicates that during the analyzed period, the daily return on WIG20 futures contracts depended significantly on the previous day's return of the WIG20 index, thereby confirming the existence of short-run causality running from the spot market to the futures market. The daily futures returns also depended on their own lagged value, as indicated by the significant (at the 0.01 level) coefficient $\gamma_{ff,1}$. It should be noted that in the equation describing Δs_t, none of the coefficients are significantly different from zero. In particular, none of the coefficients indicating potential causality from the futures market to the spot market (α_s, $\gamma_{sf,1}$, $\gamma_{sf,2}$, $\gamma_{sf,3}$) are significant.

The final column of Table 2 presents the values of the common factor weights, computed according to equations (16) and (17). These results suggest that over the period 2018–2024, the average contribution of the spot market to the price discovery process amounted to approximately 68.33%, compared with a contribution from the futures market that was less than half of this.

In summary, the results presented in Table 2 provide no basis for concluding that WIG20 index futures played a significant role in the price discovery process on the Warsaw Stock Exchange during the period under investigation. Instead, they suggest that this process occurred primarily in the spot market, with the derivatives market following it and performing only a secondary function. Thus, the findings are consistent with earlier results reported by Bohl and colleagues (Bohl et al. 2011) for the WIG20 index and its futures contracts. Taking into account prior studies on the role of index futures in price discovery in other emerging markets (Min, Najand 1999; Kenourgios 2005; Andreou, Pierides 2008; Kavussanos et al. 2008; Kavussanos, Visvikis 2011; Guo et al. 2013), the results reported in Table 2 suggest that the Polish market behaves in an atypical way. While in the cases of the Korean, Chinese, and Greek markets, it has been shown that the derivatives market – if not from the outset, then at least with further development – gradually took over the price discovery function from the spot market and expanded its contribution over time, the findings discussed above for the Polish futures market indicate that its role in the price discovery process remains limited despite more than two decades of existence.

However, as noted in the previous section, the classical VECM suffers from important analytical limitations. Above all, as a linear model, it assumes stability of the cointegrating relationship and homogeneity of adjustment mechanisms throughout the entire sample period. In reality, this assumption may be overly restrictive, particularly in the context of dynamically changing market conditions. The sample period – 2018 to 2024 – includes episodes of exceptionally strong economic disturbances, most notably the global COVID-19 pandemic and the outbreak of the war in Ukraine. These events may have significantly influenced

market participants' behavior and the structure of the linkages between the spot and futures markets. In particular, the years 2020–2022 were characterized by heightened volatility, uncertainty, and structural instability in financial markets.

Under such conditions, the assumption of a homogeneous, linear adjustment process may lead to distorted inferences. The return to equilibrium between spot and futures markets may have taken different forms during stable and unstable periods, and it may also have been asymmetric – i.e., stronger in response to negative shocks than to positive ones. The classical VECM is incapable of capturing such nonlinearities and regime shifts.

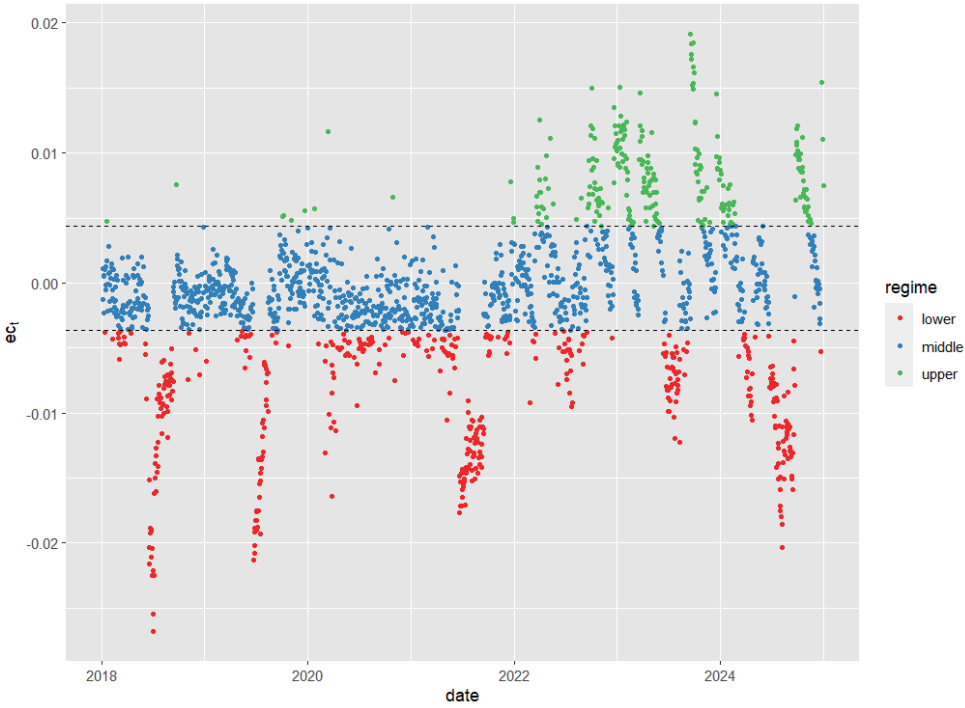
For this reason, the TVECM model was employed in the subsequent stage of the analysis. This framework makes it possible to account for asymmetries and varying adjustment dynamics depending on the extent of the deviation from equilibrium. By allowing for different error-correction mechanisms across regimes, the TVECM may provide a more adequate tool for analyzing periods characterized by instability and structural shocks.

The appropriate specification of the three-regime TVECM was selected on the basis of the Bayesian Information Criterion (BIC). This criterion indicated a TVECM with a lag order of $p = 1$. However, the BIC for the TVECM ($BIC_{TVECM} = -35{,}234.70$) is higher than that for the classical VECM ($BIC_{VECM} = -35{,}339.94$) employed in the previous stage, suggesting that the VECM provides a better fit according to this criterion. The higher BIC value for the TVECM may reflect its greater complexity, as BIC imposes a stronger penalty for additional parameters than AIC.

In contrast, the AIC for the TVECM is lower than for the VECM ($AIC_{TVECM} = -35{,}376.80$, while $AIC_{VECM} = -35{,}247.04$), indicating some support for the TVECM specification. Overall, the mixed evidence from the information criteria suggests that both models provide a comparable fit to the empirical data.

The estimated threshold values appearing in equation (12) amounted to (rounded to three decimal places): $\tau_1 = -0.004$ and $\tau_2 = 0.004$. A total of 26.1% of the observations fell into the lower regime. This means that on approximately one quarter of the trading days under consideration, the error-correction term given by equation (6) was negative and lower than τ_1, implying that the daily closing price of WIG20 futures contracts was undervalued relative to the underlying index. The middle regime comprised 59.5% of the observations, indicating that on most of the trading days, the spot and futures prices did not deviate largely from their long-run equilibrium relationship. Finally, 14.4% of the observations fell into the upper regime, in which the error-correction term ec_t was positive and greater than τ_2, signifying that the WIG20 futures contracts were overpriced relative to the index.

The division of observations across regimes is illustrated in Figure 2. The vast majority of observations are located in the middle regime (blue), corresponding to minor deviations from equilibrium, where adjustment mechanisms are likely to remain inactive. The lower regime (red) is more numerous than the upper regime (green), suggesting that instances of futures undervaluation occurred more frequently than cases in which futures were overvalued relative to the spot market during the sample period. This may point to asymmetric arbitrage costs or the existence of other market frictions. Horizontal dashed lines have been added to the figure to clearly separate the three regimes.



Note: The lower regime corresponds to the values $ec_t < -0.004$, while the upper regime includes the values $ec_t > 0.004$. The values $ec_t \in [-0.004; 0.004]$ determine the middle regime

Figure 2. Division of correction error term values ec_t across the three regimes

The results of the TVECM estimation are presented in Table 3. The residuals were found to be free of autocorrelation but exhibited heteroskedasticity; consequently, the statistical significance of the estimated parameters was evaluated using

robust Newey–West standard errors. In the upper part of Table 3 (Panel A), the estimated parameter values of the model corresponding to the lower regime are reported. Values of both coefficients related to the error correction term (α_s and α_f are not significantly different from zero. This suggests that when futures prices are undervalued relative to the WIG20 index, neither the spot nor the futures market responds in a systematic way to restore equilibrium. The error correction mechanism appears to be locally inactive.

This finding may be explained by the presence of some market frictions that limit arbitrage activity. As a result, observed mispricing may not be perceived as sufficiently profitable relative to the associated risk. Furthermore, the absence of a significant error correction mechanism may be linked to the turbulent market conditions observed during the sample period, including the COVID-19 pandemic and the outbreak of the war in Ukraine. Periods of heightened uncertainty and volatility are typically associated with reduced liquidity and increased risk aversion, which can limit arbitrage activity and lead to a temporary breakdown of the equilibrium adjustment process.

Moreover, Figure 2 reveals clustering of observations within the lower regime. This indicates that deviations from equilibrium are persistent rather than transitory, as the system tends to remain in this regime for several consecutive periods once the threshold is crossed. Such regime persistence is consistent with the insignificance of the error correction coefficients and suggests that the adjustment process is not only weak but also delayed.

The statistically significant $\gamma_{fs,1}$ and $\gamma_{ff,1}$ (at the 0.1 and 0.01 levels, respectively), which appear in the lower regime in the equation related to changes in futures prices, indicate that the daily return on WIG20 index futures depended on both its own lagged value and on the previous day's return of the index. The value of the coefficient $\gamma_{fs,1}$ in particular indicates the potential presence of unidirectional short-run causality: if futures prices are too low relative to the closing value of the WIG20 index on a given day, only the futures market reacts to the information contained in spot prices the following day. Such a unilateral transmission of information between the markets was also detected in the first part of the study, as the classical VECM model indicated short-run causality running only from the WIG20 index prices to the futures prices.

Panel B, constituting the middle part of Table 3, presents the results of the TVECM parameter estimation obtained for the middle regime. As in the lower regime, the values of both coefficients related to the error correction term (α_s and α_f are statistically insignificant. This result is consistent with expectations, since the middle regime corresponds to observations for which the value of the error correction term was close to zero, and thus no significant deviation from the

long-run equilibrium between the spot and futures markets was present. Under such conditions, the adjustment mechanism toward equilibrium was not activated. According to the suggestions of Ters and Urban (Ters, Urban 2020), this regime defines the no-arbitrage band, within which the basis of the futures contract is too small (in absolute terms) for arbitrage transactions to be financially viable.

Table 3

Estimation results of the TVECM model

Equation	Coefficient values				C.f.w.
Panel A: lower regime					
Δs_t	μ_s	α_s	$\gamma_{ss,1}$	$\gamma_{sf,1}$	θ_s [%]
	0.002	0.106	0.153	−0.350	58.09
Δf_t	μ_f	α_f	$\gamma_{ff,1}$	$\gamma_{fs,1}$	θ_f [%]
	0.003**	0.147	−0.714***	0.521*	41.91
Panel B: middle regime					
Δs_t	μ_s	α_s	$\gamma_{ss,1}$	$\gamma_{sf,1}$	θ_s [%]
	−0.001*	−0.033	0.254	−0.149	88.86
Δf_t	μ_f	α_f	$\gamma_{ff,1}$	$\gamma_{fs,1}$	θ_f [%]
	−0.001**	−0.266	−0.451**	0.549***	11.14
Panel C: upper regime					
Δs_t	μ_s	α_s	$\gamma_{ss,1}$	$\gamma_{sf,1}$	θ_s [%]
	0.006**	−0.501**	−0.199	0.056	51.28
Δf_t	μ_f	α_f	$\gamma_{ff,1}$	$\gamma_{fs,1}$	θ_f [%]
	0.005**	−0.528**	−0.278	0.151	48.72

Note: The symbols ***, ** and * indicate statistical significance at significance levels of 0.01, 0.05 and 0.1, respectively, based on Newey-West robust standard errors. The last column (c.f.w.= common factor weights) reports the values of the common factor weights defined in equations (16) and (17)

It is worth noting that the middle regime accounts for about 60% of all observations. This means that on the majority of days covered by the analysis, no significant disturbance of the equilibrium relation between the markets occurred, and the correction mechanism remained inactive. This fact casts doubt on the results obtained in the first stage of the study using the classical VECM model, in which it was assumed that the adjustment mechanism operates regardless of the scale of the deviation from the long-run equilibrium between spot and futures prices.

With respect to the coefficients $\gamma_{ss,1}$ and $\gamma_{ff,1}$, it is noteworthy that in the middle regime – similarly to the previous regime – only $\gamma_{ff,1}$ is statistically significant (at the 0.01 level). This suggests that, during periods of relative stability between the markets, futures returns depend on their own lagged values, whereas this is not the case for WIG20 index returns.

For the short-term relationships between the markets, only the coefficient $\gamma_{fs,1}$ is statistically significant (at the 0.01 level). This indicates, once again, the presence of unidirectional short-run causality running from the spot market to the futures market: futures prices respond to past values of the underlying index, while no reverse relationship is observed. These findings are also consistent with the results obtained from the classical VECM model, which also indicated short-run causality running from the spot market to the futures market.

The lower part of Table 3 (Panel C) presents the estimation results of the TVECM parameters corresponding to the upper regime. In this case, both error correction coefficients, α_s and α_f, are significantly different from zero at the 0.05 level. This regime corresponds to a positive term ec_t, meaning that $f_t > \beta_1 s_t$. The negative estimate of ($\alpha_f = -0.528$) is justified. It means that futures prices decrease to get back to the equilibrium state. However, the estimate α_s was also obtained as negative ($\alpha_s = -0.501$). It means that in the situation of a large positive error correction term, the spot prices also decrease, which is not conducive to restoring an equilibrium state. Bohl and colleagues (Bohl et al. 2011) suggest that it is possible that the error correction coefficients of the spot and futures markets have the same sign. However, to restore the equilibrium in the long run, the inequality $|\alpha_s| < |\alpha_f|$ should hold. This holds in our case in absolute values, but the test of the statistical difference between α_s and α_f does not provide grounds to reject the null hypothesis of equality of the coefficients. The findings thus are contrary to the standard expectations of the error correction mechanism and suggest that both the spot and futures markets respond in a similar direction to deviations from equilibrium. Instead of correcting the deviation, both markets appear to move in parallel, which may prevent the spread from narrowing in a systematic way. This suggests a breakdown of the classical arbitrage-based error correction mechanism in the upper regime.

One possible explanation for this behavior is the presence of common informational shocks affecting both markets simultaneously. In periods of heightened volatility, such as during financial or geopolitical crises, arbitrage activity may become constrained due to increased risk or market frictions. As a result, prices on both markets may adjust concurrently to new information rather than through a mechanism aimed at restoring equilibrium. As can be seen in Figure 2, the upper regime consists predominantly of observations from the period 2022–2024,

corresponding to the time of the war in Ukraine. This confirms that the regime may be associated with increased uncertainty. Moreover, similarly to the lower regime, a clear clustering of observations can be observed. This indicates that once the system enters this regime, it tends to remain there for several consecutive periods, which suggests that the adjustment process in the regime is delayed or weakened rather than immediate. This is consistent with the obtained α_s and α_f values. The results therefore point to the dominance of common factors driving both markets, rather than the operation of an effective arbitrage mechanism linking them.

With respect to short-run dynamics, none of the estimated coefficients are statistically significant. This indicates that daily returns of both the WIG20 index and its futures contracts do not depend on either their own lagged values or the lagged returns observed in the other market.

This lack of short-run interdependence further supports the interpretation that price dynamics are driven primarily by contemporaneous informational shocks in the upper regime rather than by systematic interactions between the two markets. In other words, standard short-run transmission mechanisms appear to be dominated by exogenous factors affecting both markets simultaneously.

The last column of Table 3 contains the values of the common factor weights, calculated according to formulas (16) and (17), separately for each regime. In both the upper and lower regimes, the two coefficients – θ_s and θ_f – are close to 50%, although a slightly higher contribution to price discovery is attributed to the spot market in both cases. The values of θ_s and θ_f computed in the middle regime amounted to 80.19% for the spot market and only 19.81% for the futures market. These results differ substantially from those based on the classical VECM specification (see Table 2), which suggested a dominant role of the spot market in the price discovery process throughout the entire sample period. This discrepancy highlights an important limitation of the linear VECM framework, which assumes a constant adjustment mechanism and stable relationships between markets over time.

In contrast, the TVECM model reveals that the error correction mechanism is either weak, statistically insignificant, or operates in a non-standard way in different regimes. As a result, the process of restoring long-run equilibrium cannot be regarded as uniform across market conditions. Common factor weights rely on the existence of a well-functioning error correction mechanism. However, in the regimes where such a mechanism is absent or ineffective, the estimated weights cannot be considered reliable indicators of the relative contributions of the spot and futures markets to price discovery. The results suggest that the linear VECM model, although it provides a convenient summary measure, oversimplifies the

underlying dynamics between the stock and futures markets by averaging across fundamentally different market states.

When comparing the values of the coefficients related to the error correction term across regimes, the following inequalities can be observed: $|\alpha_s^1| > |\alpha_s^{-1}| > |\alpha_s^0|$ and: $|\alpha_f^1| > |\alpha_f^0| > |\alpha_f^{-1}|$. These inequalities between α_f values are not consistent with earlier expectations that the absolute values of the error correction coefficients in the lower and upper regimes should be higher than in the middle regime. Once again, these results highlight the non-standard behavior of the error correction coefficients across regimes, deviating from the conventional expectations.

When comparing the TVECM estimation results with the findings of similar studies conducted for the U.S. market by Martens and colleagues (Martens et al. 1998) and Kim and colleagues (Kim et al. 2010), partial consistency can be observed with respect to the existence of a no-arbitrage band. The results of the present study suggest that such a band may also be present in the Polish market; however, the adjustment mechanism is not uniformly activated once the threshold is exceeded as deviations from equilibrium are not always effectively eliminated in either the upper or the lower regime.

Furthermore, the results obtained for the WIG20 index and its futures differ from those reported for the S&P 500 index and its derivatives with regard to the price discovery process. While studies for the U.S. market point to a dominant role of the futures market, the evidence for the Warsaw Stock Exchange suggests that price discovery is not stable over time and varies across regimes. In particular, the spot market appears to play a leading role in some regimes, while the relationship between the markets weakens or is dominated by common informational shocks in others. Overall, the findings indicate that, unlike in more mature markets, the price discovery process in the Polish market is less efficient and more sensitive to changing market conditions.

5. Summary and conclusions

The aim of the study was to compare the usefulness of the classical Vector Error Correction Model (VECM) and its nonlinear threshold version (TVECM) in analyzing the price discovery process on the Warsaw Stock Exchange. The study covered daily data on the closing prices of the WIG20 index and its futures contracts from 2018 to 2024, thus encompassing both periods of relative market stability and periods of pronounced market turbulence associated with the outbreak of the COVID-19 pandemic and the war in Ukraine. Such a broad time horizon made it possible to assess to what extent the two models are capable of capturing the

mechanisms of price discovery under conditions of volatility, asymmetry, and potential structural changes.

The estimation results of the classical VECM model indicated unequivocally the dominant role of the spot market in the price discovery process. Both the analysis of the error correction coefficients and the investigation of short-run causality suggested that in the years 2018–2024, it was the WIG20 index that initiated changes, while the futures contracts followed. The measures of the contribution of spot and futures markets to the price discovery process (common factor weights) suggested that the average contribution of the spot market to price discovery was more than twice as high as that of the futures market. These findings confirmed earlier results in this respect (Bohl et al. 2011) and implied a secondary role for WIG20 futures contracts in the price discovery process.

However, when analyzing the estimation results of the classical VECM model, attention was drawn to its limitations stemming from its linear nature and the assumption of constant adjustment mechanisms over the entire sample period. It was therefore hypothesized that the conclusions derived from the VECM model might be oversimplified, particularly in light of the crises that occurred in the market during the analyzed period. To verify this hypothesis, the threshold TVECM model was employed in the subsequent stage of the study, which made it possible to incorporate nonlinearity and differentiated market reactions depending on the magnitude and direction of deviations from equilibrium.

The estimation results of the TVECM model revealed that the error correction mechanism behaves differently and, at least in part, unexpectedly. In line with theoretical expectations, no significant adjustment toward equilibrium was observed in the middle regime, which can be interpreted as a no-arbitrage band. However, no significant adjustment toward the long-run equilibrium was observed in the lower regime, while both markets reacted in a similar direction to deviations in the upper regime, which did not lead to an effective reduction of disequilibrium. These findings indicate that deviations from equilibrium may persist over time and are not necessarily eliminated through the classical arbitrage mechanism.

Moreover, the regime-based analysis showed that short-run relationships between the spot and futures markets are unstable. In the lower and middle regimes, short-run causality from the spot market to the futures market was observed, while no significant interactions were detected in the upper regime, indicating that price transmission mechanisms cannot be adequately described by a single, stable structure.

A comparison of the two models highlights that the VECM tends to oversimplify dynamics by imposing a constant adjustment mechanism, whereas the TVECM captures regime-dependent variations in both the error correction process and short-run interactions. Nevertheless, even the TVECM may not fully account

for price discovery during highly turbulent periods, reflecting the inherent limitations of both models.

Importantly, the results indicate that the contribution of the spot and futures markets to price discovery is not stable over time. While the VECM suggests a dominant role of the spot market, the TVECM indicates that this dominance is conditional and may vary across regimes, with extreme market turbulence potentially driven by common informational shocks rather than classical arbitrage mechanisms.

Finally, it should be noted that the study's results are subject to important limitations stemming from the use of daily data, which were employed due to the lack of access to high-frequency intraday data. While daily data are suitable for analyzing long-term relationships, in modern markets, the informational leadership of a given market in the price discovery process can manifest within minutes, seconds, or even milliseconds. For example, Hasbrouck showed for the U.S. market that informational contributions to price discovery, which remain unclear at a one-second frequency, may only be revealed at the sub-millisecond level (Hasbrouck 2021).

In recent years, most studies have relied on high-frequency intraday data to analyze price discovery and the relative informational contributions of spot and futures markets (e.g., Aggarwal, Thomas 2019; Baur, Dimpfl 2019; Zhou et al. 2021; Han et al. 2025). Daily data cannot capture the information flows with all the details or the true speed of price adjustments across markets. This limitation is particularly relevant during periods of heightened market volatility, such as the COVID-19 pandemic and the outbreak of the war in Ukraine, when intraday information transmission was rapid and markets were highly reactive. Therefore, the findings of this study should be interpreted with caution with respect to the high-frequency dynamics of the price discovery process. Future research will aim to extend this analysis using high-frequency intraday data, which would allow for a more precise assessment of the speed and microstructural dynamics of price discovery between the spot and futures markets.

References

1. Aggarwal, N. and Thomas, S. (2019) 'When stock futures dominate price discovery', *Journal of Futures Markets*, vol. 39(3), pp. 263–278, https://doi.org/10.1002/fut.21973.
2. Alphonse, P. (2000) 'Efficient price discovery in stock index cash and futures markets', *Annales d'économie et de statistique*, vol. 60, pp. 177–188, https://doi.org/10.2307/20076259.

3. Andreou, P.C. and Pierides, Y.A. (2008) 'Empirical investigation of stock index futures market efficiency: The case of the Athens derivatives exchange', *The European Journal of Finance*, vol. 14(3), pp. 211–223, https://doi.org/10.1080/13518470801890768.
4. Balke, N.S. and Fomby, T.B. (1997) 'Threshold cointegration', *International Economic Review*, vol. 38(3), pp. 627–645, https://doi.org/10.2307/2527284.
5. Baur, D.G. and Dimpfl, T. (2019) 'Price discovery in bitcoin spot or futures?', *Journal of Futures Markets*, vol. 39, pp. 803–817, https://doi.org/10.1002/fut.22004.
6. Bohl, M.T., Salm, C.A. and Schuppli, M. (2011) 'Price discovery and investor structure in stock index futures', *Journal of Futures Markets*, vol. 31(3), pp. 282–306, https://doi.org/10.1002/fut.20469.
7. Booth, G.G., So, R.W. and Tse, Y. (1999) 'Price discovery in the German equity index derivatives market', *Journal of Futures Markets*, vol. 19, pp. 619–643, https://doi.org/10.1002/(SICI)1096-9934(199909)19:6<619::AID-FUT1>3.0.CO;2-M.
8. Brooks, C., Rew, A.G. and Ritson, S. (2001) 'A trading strategy based on the lead-lag relationship between the spot index and futures contract for the FTSE 100', *International Journal of Forecasting*, vol. 17, pp. 31–44, https://doi.org/10.1016/S0169-2070(00)00062-5
9. Buckle, M., Chen, J., Guo, Q. and Li, X. (2019) 'The impact of multilateral trading facilities on price discovery: Further evidence from the European markets', *Financial Markets, Institutions & Instruments,* vol. 28(4,) pp. 321–343, https://doi.org/10.1111/fmii.12121.
10. Chen, W.K., Lin, C.T. and Shiu, C.Y. (2019) 'Price discovery and price leadership of various investor types: evidence from Taiwan futures markets', *Review of Quantitative Finance and Accounting*, vol. 53, pp. 601–631, https://doi.org/10.1007/s11156-018-0760-3.
11. Chen, Y.-L., Lee, Y.-H., Chou, R.K. and Chang, Y.-K. (2021) 'Arbitrage trading and price discovery of the regular and mini Taiwan stock index futures', *The Journal of Futures Markets*, vol. 41(6), pp. 926–948, https://doi.org/10.1002/fut.22192.
12. Chou, R.K. and Chung, H. (2006) 'Decimalization, trading costs, and information transmission between ETFs and index futures', *Journal of Futures Markets*, vol. 26(2), pp. 131–151, https://doi.org/10.1002/fut.20189.
13. Curran, E., Hunt, J. and Mollica, V. (2020) 'Trading protocols and price discovery: implicit transaction costs in Indian single stock futures', *Journal of Futures Markets*, vol. 40(11), pp. 1793–1806, https://doi.org/10.1002/fut.22123.
14. Fama, E.F. (1970) 'Efficient capital markets: A review of theory and empirical work', *The Journal of Finance*, vol. 25(2), pp. 383–417, https://doi.org/10.2307/2325486.

15. Fassas, A.P. and Siriopoulos, C. (2019) 'Intraday price discovery and volatility spillovers in an emerging market', *International Review of Economics and Finance*, vol. 59, pp. 333–346, https://doi.org/10.1016/j.iref.2018.09.008.
16. Gaul, J. and Theissen, E. (2008) A partially linear approach to modelling the dynamics of spot and futures prices, CFS Working Paper Series, No. 2008/12, Center for Financial Studies, [Online], Available: https://gfk-cfs.de/media//08_12.pdf [14 Apr 2026].
17. Gwilym, O.A. and Buckle, M. (2001) 'The lead-lag relationship between the FTSE100 stock index and its derivative contracts', *Applied Financial Economics,* vol. 11(4), pp. 385–393, https://doi.org/10.1080/096031001300313947.
18. Guo, B., Han, Q., Liu, M. and Ryu, D. (2013) 'A tale of two index futures: The intraday price discovery and volatility transmission processes between the China Financial Futures Exchange and the Singapore Exchange', *Emerging Markets Finance and Trade*, vol. 49, pp. 197–212, https://doi.org/10.2753/REE1540-496X4905S414.
19. Han, Q., Zhao, C., Chen, J. and Guo, Q. (2025) 'Does asynchronous market update matter? Re-examining the price discovery of stock index and futures in China', *Emerging Markets Review*, vol. 67, 101307, https://doi.org/10.1016/j.ememar.2025.101307.
20. Hansen, B.E. and Seo, B. (2002) 'Testing for two-regime threshold cointegration in vector error-correction models', *Journal of Econometrics*, vol. 110(2), pp. 293–318, https://doi.org/10.1016/S0304-4076(02)00097-0.
21. Hasbrouck, J. (2003) 'Intraday price formation in U.S. equity index markets', *Journal of Finance,* vol. 58(6), pp. 2375–2399, https://doi.org/10.1046/j.1540-6261.2003.00609.x.
22. Hasbrouck, J. (2021) 'Price Discovery in High Resolution', *Journal of Financial Econometrics*, vol. 19(3), pp. 395–430, https://doi.org/10.1093/jjfinec/nbz027.
23. Hou, Y. and Li, S. (2013) 'Price discovery in Chinese stock index futures market: New evidence based on intraday data', *Asia-Pacific Financial Markets,* vol. 20, pp. 49–70, https://doi.org/10.1007/s10690-012-9158-8.
24. Johansen, S. (1991) 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, vol. 59(6), pp. 1551–1580, https://doi.org/10.2307/2938278.
25. Johansen, S. (1992) 'Determination of cointegration rank in the presence of a linear trend', *Oxford Bulletin of Economics and Statistics,* vol. 54(3), pp. 383–397, https://doi.org/10.1111/j.1468-0084.1992.tb00008.x.
26. Kang, H., Kang, J. and Lee, S. (2016) 'Which traders contribute most to price discovery? Evidence from the KOSPI 200 options market', *Emerging Markets Finance and Trade*, vol. 52(10), pp. 2335–2347, https://doi.org/10.1080/1540496X.2016.1196927.

27. Kavussanos, M.G. and Visvikis, I.D. (2011) 'The predictability of non-overlapping forecasts: Evidence from a new market', *Multinational Finance Journal*, vol. 15, pp. 125–156, [Online], Available: https://ssrn.com/abstract=1586822 [14 Apr 2026].
28. Kavussanos, M.G., Visvikis, I.D. and Alexakis, P.D. (2008) 'The lead-lag relationship between cash and stock index futures in a new market', *European Financial Management,* vol. 14(5), pp. 1007–1025, https://doi.org/10.1111/j.1468-036X.2007.00412.x.
29. Kenourgios, D.F. (2005) 'Testing efficiency and the unbiasedness hypothesis of the emerging Greek futures market', *European Review of Economics and Finance*, vol. 4(2), pp. 3–20, [Online], Available: https://ssrn.com/abstract=871364 [14 Apr 2026].
30. Kim, B.-H., Chun, S.-E. and Min, H.-G. (2010) 'Nonlinear dynamics in arbitrage of the S&P 500 index and futures: A threshold error-correction model', *Economic Modelling*, vol. 27(2), pp. 566–573, https://doi.org/10.1016/j.econmod.2009.11.011.
31. Kumar, U. and Tse, Y. (2009) 'Single stock futures: Evidence from the Indian securities market', *Global Finance Journal,* vol. 20(3), pp. 220–234, https://doi.org/10.1016/j.gfj.2009.06.004.
32. Liu, Q. and Qiao, G. (2017) 'The evolving nature of intraday price discovery in the Chinese CSI 300 index futures market', *Empirical Economics*, vol. 52(4), pp. 1569-1585, https://doi.org/10.1007/s00181-016-1115-3.
33. Mamatzakis, E. and Remoundos, P. (2010) Threshold cointegration in BRENT crude futures market, Munich Personal RePEc Archive Paper, No. 19978, University Library of Munich, Germany, [Online], Available: https://mpra.ub.uni-muenchen.de/19978/ [14 Apr 2026].
34. Marcinkiewicz, E. and Kompa, K. (2013) 'Badanie przyczynowości między cenami spot i futures na przykładzie kontraktów terminowych na indeks WIG20', *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, rynki finansowe, ubezpieczenia*, vol. 768(63), pp. 321–331, [Online], Available: http://wneiz.pl/nauka_wneiz/frfu/63-2013/FRFU-63-321.pdf [14 Apr 2026].
35. Martens, M., Kofman, P. and Vorst, T.C.F. (1998) 'A threshold error-correction model for intraday futures and index returns', *Journal of Applied Econometrics*, vol. 13(3), pp. 245–263, https://doi.org/10.1002/(SICI)1099-1255(199805/06)13:3%3C245::AID-JAE480%3E3.0.CO;2-E.
36. Min, J.H. and Najand, M.(1999) 'A further investigation of the lead–lag relationship between the spot market and stock index futures: Early evidence from Korea', *Journal of Futures Markets,* vol. 19(2), pp. 217–232, https://doi.org/10.1002/(SICI)1096-9934(199904)19:2%3C217::AID-FUT5%3E3.0.CO;2-8.
37. Mutlu, E. and Arık, E. (2015) 'Interaction between single-stock futures and the underlying securities: A cross-country analysis', *Emerging Markets Finance and Trade*, vol. 51(3), pp. 647–657, https://doi.org/10.1080/1540496X.2014.998568.

38. Schwarz, T.V. and Szakmary, A.C. (1994) 'Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis', *Journal of Futures Markets,* vol. 14(2), pp. 147–167, https://doi.org/10.1002/fut.3990140204.
39. Suliga, M. (2025) 'Price discovery in single-stock futures: Evidence from the Warsaw Stock Exchange', *Scientific Papers of Silesian University of Technology. Organization and Management Series*, No. 222, pp. 551–570, http://dx.doi.org/10.29119/1641-3466.2025.222.31. 3466.2025.222.31
40. Ters, K. and Urban, J. (2020) 'Estimating unknown arbitrage costs: Evidence from a 3-regime threshold vector error correction model', *Journal of Financial Markets*, vol. 47, 100503, https://doi.org/10.1016/j.finmar.2019.07.002.
41. Theissen, E. (2002) 'Price discovery in floor and screen trading systems', *Journal of Empirical Finance*, vol. 9, pp. 455–474, https://doi.org/10.1016/S0927-5398(02)00005-1.
42. Xi, Z., Pan, H. and Qin, T. (2023) 'Re-examining the efficiency of the EU carbon futures market in phase II: price discovery and intertemporal arbitrage', *Frontiers in Energy Research*, vol. 11, https://doi.org/10.3389/fenrg.2023.1236488.
43. Xu, X. (2018) 'Intraday price information flows between the CSI300 and futures market: an application of wavelet analysis', *Empirical Economics,* vol. 54, pp. 1267–1295, https://doi.org/10.1007/s00181-017-1245-2.
44. Zhou, X., Zhang, J. and Zhang, Z. (2021) 'How does news flow affect cross-market volatility spillovers? Evidence from China's stock index futures and spot markets', *International Review of Economics & Finance*, vol. 73, pp. 196–213, https://doi.org/10.1016/j.iref.2021.01.003.

Appendix

Table A1 presents the basic descriptive statistics for the log returns of the WIG20 index and its futures contracts. The results indicate that the individual descriptive statistics are very similar across both markets. Both return series exhibit negative skewness, while their high kurtosis values suggest a strong concentration of observations around a mean close to zero. Moreover, the significantly positive Jarque–Bera test statistics at the 0.01 significance level indicate that neither distribution can be considered normal.

Table A1

Descriptive statistics for the log returns of the WIG20 index and its futures contracts in the years 2018–2024

Statistics	Time series	
	Δs_t	Δf_t
mean	−0.0001	−0.0001
median	−0.0001	−0.0004

Table A1 cont.

minimum	−0.1425	−0.1275
maximum	0.0810	0.0900
standard deviation	0.0148	0.0152
skewness	−0.7512	−0.5862
kurtosis	11.6772	10.3662
J-B test statistic	5648.3***	4052.1***
N	1748	1748

Note: The symbol *** denotes statistical significance at the 0.01 level

Table A2 reports the results of the stationarity tests for the log prices of the WIG20 index and its corresponding futures contracts, as well as for the log returns. The Augmented Dickey–Fuller (ADF) test statistic obtained for the price series s_t and f_t is not statistically significant for all variants of the ADF test. This implies that there is no basis for rejecting the null hypothesis that the series of log spot and futures prices are nonstationary and exhibit a stochastic trend. By contrast, the ADF test applied to the first differences confirms the stationarity of the log returns Δs_t and Δf_t, as test statistics reported on the right-hand side of the table are statistically significant (at the 0.01 or 0.05 significance levels, respectively). The results presented in Table A2 therefore confirm that the series s_t and f_t are integrated of order one, I(1).

Table A2

Results of stationarity tests for the time series of log prices and log returns of the WIG20 index and its futures contracts

Alternative hypothesis H_1	ADF test statistic			
	s_t	f_t	Δs_t	Δf_t
no drift no trend	−0.349	−0.331	−42.9***	−44***
drift	−2.57	−2.66	−42.9**	−44***
drift and trend	−2.51	−2.61	−42.9***	−44***

Note: The symbols *** and ** indicate statistical significance at significance levels of 0.01 and 0.05, respectively

Summary

This study evaluates and compares the usefulness of the classical Vector Error Correction Model (VECM) and the Threshold Vector Error Correction Model (TVECM) in analyzing the price discovery process on the Warsaw Stock Exchange. The empirical analysis uses daily data on the WIG20 index and its futures contracts from 2018 to 2024. The VECM results indicate unidirectional long-run and short-run causality from the spot market to the futures market, with the latter primarily adjusting to deviations from equilibrium. The estimated common factor weights suggest that the spot market accounts for about two-thirds of the overall price discovery. Based on these findings alone, one might conclude that the dominance of the spot market is stable and persistent. However, the TVECM reveals substantial nonlinearities and regime-dependent dynamics that challenge this conclusion. It identifies three regimes, corresponding to undervaluation, near-equilibrium, and overvaluation of futures, within which the adjustment mechanisms differ notably. In both the lower and middle regimes, the error correction mechanism is weak or statistically insignificant, indicating that deviations from equilibrium are not systematically eliminated. In particular, the middle regime, which accounts for the majority of observations, can be interpreted as a no-arbitrage band in which mispricing is too small to trigger arbitrage activity. In the upper regime, although both markets respond to deviations, their adjustments occur in the same direction, preventing the restoration of equilibrium and suggesting a breakdown of the classical arbitrage mechanism. This behavior may reflect the presence of common informational shocks and heightened market uncertainty rather than a stable lead–lag relationship between the markets. The comparison demonstrates that while the VECM provides a convenient summary of average relationships, it oversimplifies the underlying dynamics by assuming a constant adjustment process. The TVECM offers a more informative framework by capturing regime-specific behavior and revealing that the price discovery process is unstable, asymmetric, and sensitive to market conditions. These findings highlight the importance of nonlinear approaches in analyzing financial market dynamics, particularly in periods of increased volatility.

JEL codes: C32, C58

Keywords: *price discovery, Granger causality, futures on WIG20 index, VECM model, TVECM model*

Instruction for authors

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

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

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

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

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

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

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

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

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

Reference style

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

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

- JOURNAL ARTICLE

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

- BOOKS

Book with one author

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

Book with two authors

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

Book with three or more authors

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

Book – second or later edition

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

Book by same author in the same year

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

Book with an editor

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

A chapter in a book

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

- WORLD WIDE WEB PAGE

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

- CONFERENCE PAPERS

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

- NEWSPAPER ARTICLES

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

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

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

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

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

Double blind peer review procedure

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

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

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

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

– assess:

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

– verity whether:

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

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

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

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