Poznań University of Economics and Business

Michał Hebdzyński

Housing Rental Market in Selected Polish Cities in the Face of the COVID-19 Pandemic – Hedonic Perspective

(Collection of thematically cohesive articles and studies)

Supervisor: dr hab. Radosław Trojanek, prof. UEP

Dissertation received on: …………………………………..

…………………………………………………………………………………………..
Supervisor's signature

Poznań 2024

To all those to whom I devoted too little time while preparing this dissertation.

Table of contents

Part I. Description of the original solution to the scientific problem constituting the subject of the doctoral dissertation in the form of a collection of thematically cohesive articles and studies

I.1. Introduction

The role of the long-term housing rental market (LTR) is to satisfy basic human housing needs. Additionally, it has been proven that a developed LTR market may be considered effective in reducing the volatility of the real estate market and may contribute to the overall economic stability (Rubaszek et al., 2024; Rubaszek & Rubio, 2020). On the other hand, short-term rental contracts (STR) are used mainly for tourism (Guttentag et al., 2017) and last a few days. The STR market is considered a disruptor of the housing market as it decreases housing affordability for citizens and inflates housing prices and rents (Barron et al., 2020).

The COVID-19 pandemic spreading globally since early 2020 induced an unprecedented, multidimensional shock to worldwide economies (Iyer & Simkins, 2022; Kholodilin & Rieth, 2023) and can be considered a super-shock, initiating structural changes (Dolnicar & Zare, 2020). In Poland, it hit not only the economically desired LTR market but also affected the disruptor – the STR market, causing shocks to the sub-markets' supply and demand. Yet, to date, the empirical evidence concerning the impact of the pandemic on the rental market has been limited on both micro and macro levels, across the sub-markets and within their segments. As a primary reason, one may indicate the information gaps. They are reflected in the incomplete comprehension of information signals sent by market participants, imperfection of data sources available for conducting market analysis, and, finally, in problems with interpreting the results of studies based on them.

This dissertation aims to provide comprehensive knowledge about the rental market's adjustment processes to the pandemic-induced shock. However, it intends to go beyond the historical analysis. It strives to show the areas that have gained little attention in the worldwide scientific literature to date, particularly concerning countries with a little-developed private LTR market. They should be regarded in any study of the past, current and future shocks. On the other hand, the dissertation provides novel analytical methods and approaches to hedonic analysis of the market (Lancaster, 1966; Rosen, 1974) that contribute to bridging the detected information gaps. As a result, in the study, the pandemic is considered not only a source of shock leading to market change but also a stimulus to improve the scope of analytical tools needed for market monitoring. It is crucial, as international institutions have highlighted the particular role of housing market supervision since the outbreak of the global financial crisis in 2007-2008, in which one of the leading roles has been attributed to the bursting of the bubble on the real estate market (FSB & IMF, 2009; European Commission, Eurostat, Organisation for Economic Co-operation and Development & World Bank, 2013).

To study the phenomena mentioned above, two research questions have been asked:

1. To what extent does including information on housing quality in hedonic models affect their performance and robustness in the face of the economic shock?
2. Has the COVID-19 pandemic influenced the housing rental market regarding the rent-setting factors and rent changes?

For the research problem and questions thus outlined, a thesis was formulated, which reads: The COVID-19 pandemic affected the housing rental market in selected Polish cities, and to deepen the understanding of the impact, one needs to take account of the drawbacks of the available data sources, particularly with regard to information on housing quality.

The following research hypotheses are an extension of the thesis:

H1: The LTR market listing data may act as a proxy of transactional data in terms of their usefulness for constructing hedonic models.

H2: Hedonic rent indices of the LTR market are robust to minor changes in the selection of the explanatory variables in hedonic models.

H3: Agents and landlords who list apartments for rent on the LTR market tend to overstate their declarations regarding the quality of the apartments.

H4: The Wordscores algorithm may be used to extract quality signals from textual descriptions included in the LTR market listings.

H5: The underspecification of hedonic models of the LTR market in terms of housing quality leads to obtaining inferior statistical properties of the models and imprecise estimates.

H6: During the pandemic, the rent-setting factors in the LTR market changed.

H7: During the pandemic, rent levels in the STR and LTR markets declined, but the reactions varied spatially.

H8: During the pandemic, rent levels in the LTR market declined unevenly across quality-related market segments.

H9: Changes in the STR market rents cause changes in the LTR market rents.

H10: During the pandemic, the decrease in the supply of apartments in the STR market contributed to the decline in rents in the LTR market.

To solve the outlined research problem, prove the formulated thesis and test the hypotheses, it was necessary:

- to examine the already presented evidence on the impact of the pandemic on the housing markets and to identify the crucial yet unexplored market phenomena,
- to critically analyse the latest scientific literature in terms of data and methods used to study the housing rental market,
- to construct the micro-level, pooled cross-sectional database of the LTR market listings, which is unique in terms of the scope of the included information and their precision and has allowed to target the stated hypotheses precisely,
- to get access to the transactional and listing data of the STR and LTR market that have allowed to extend further the analytical scope of the dissertation,
- to detect the gaps concerning the information on the quality of apartments in the listing process and later in the listings-originated datasets,
- to propose novel analytical approaches to bridge the detected information gaps,
- to construct micro- and macro-level models of the housing rental market using primarily, but not exclusively, hedonic methods, as well as their variants,
- to provide multiple robustness checks to ensure the stability of the obtained models and their results.

The implementation of research tasks, subordinated to achieving the purpose of the work, required the definition of the scopes of work:

Objective scope: the LTR market of apartments located in multi-family buildings and the corresponding segment of the STR market.

Spatial scope: Warsaw and Poznan (their locations have been presented in Figure 1). The selection of Warsaw has been motivated by the fact that it is the city with the largest LTR and STR markets in Poland. On the other hand, Poznan is the fifth most populated Polish city, considered a business centre, attracting students and internal and external migrants. Thus, it constitutes a well-suited example to study the reaction of the LTR market to the pandemic-induced shock.

Time scope: the LTR market of Warsaw – 2015-2020, and of Poznan – 2019-2023; the STR market of Warsaw – 2015-2020.

Methodological scope: non-spatial (Ordinary Least Squares, Quantile Regression) and spatial (Spatial Error Model, Multiscale Geographically Weighted Regression) hedonic methods, BMN (Bailey-Muth-Nourse) repeat sales method to construct price indices, supervised machine-learning Wordscores algorithm, Granger causality test.

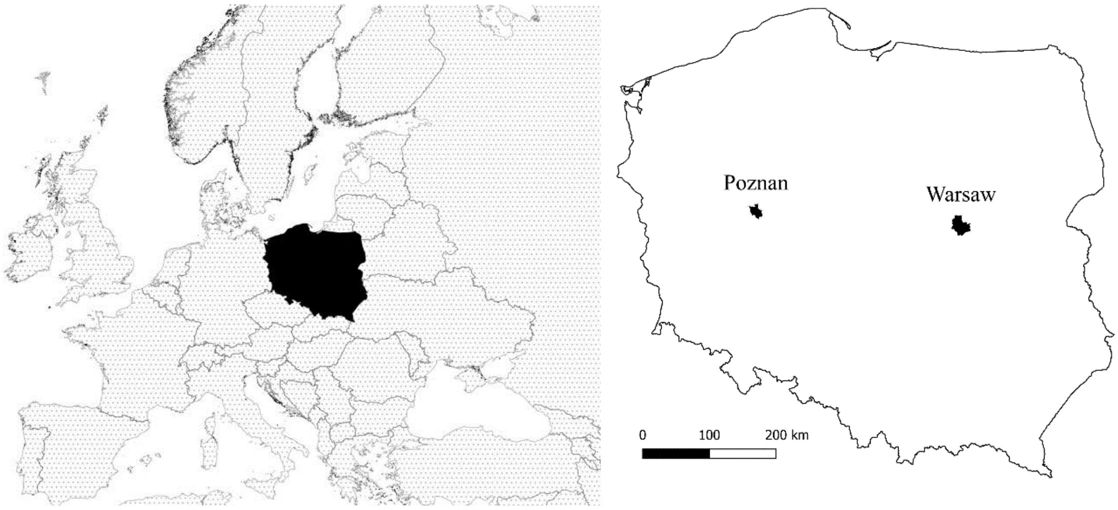


Figure 1. Location of Poland on the map of Europe (left panel), location of Poznan and Warsaw on the map of Poland (right panel)

Source: own elaboration

The research questions, thesis and hypotheses have been targeted in a series of thematically cohesive studies – 5 articles and 1 working paper (presented in logical order):

Article 1 (A1) – Hebdzyński, M. (2024c). Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data?, *Bank i Kredyt, 55*(4). https://bankikredyt.nbp.pl/content/2024/04/bik_04_2024_05m.pdf

Article 2 (A2) – Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures?, *Journal of Housing and the Built Environment, 38*, 2399–2425, https://doi.org/10.1007/s10901-023-10043-z

Article 3 (A3) – Hebdzyński, M. (2024a). Are hedonic models really quality-adjusted? The role of apartment quality in hedonic models of housing rental market. *Real Estate Management and Valuation*, *32*(2), 46-57. https://doi.org/10.2478/remav-2024-0014

Article 4 (A4) – Hebdzyński, M. (2024b). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies, 17*(2), 298-323. https://doi.org/10.14254/2071-8330.2024/17-2/16

Article 5 (A5) – Trojanek, R., Gluszak, M., Hebdzyński, M. & Tanas, J. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

- I am a co-author of the article, having a 10% share in the preparation of the article – I participated in constructing the price and rent indices and in writing the manuscript.

Working paper 1 (WP1) – Hebdzyński, M., Trojanek, R. (2024). *Is there a relationship between rents in the short- and long-term housing rental markets? A case study of Warsaw*. Working paper.

- I am a co-author of the working paper, having a 50% share in the study conceptualisation, formal analysis, calculation, visualisation, and manuscript writing.

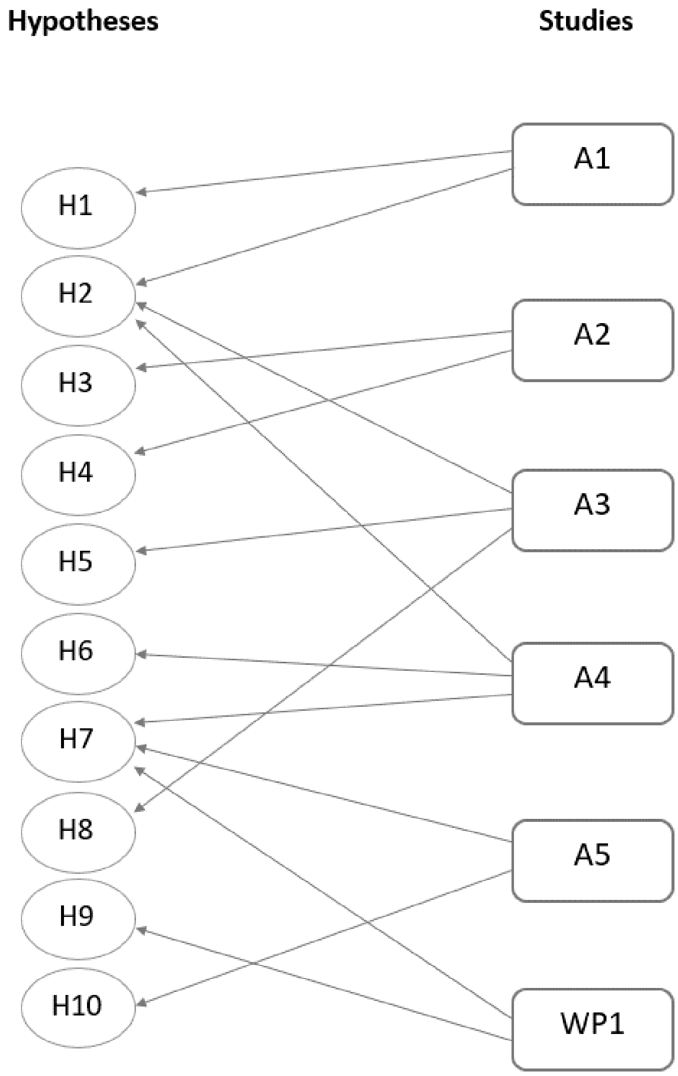


Figure 2. Logical structure of the dissertation – the linkages between the conducted studies and hypotheses

Source: own elaboration

I.2. Description of the research subject and problem

As a long-term rental (LTR) contract one should understand a rental agreement aimed at satisfying housing needs that is typically concluded for 12 months (as it is most often in Poland), but it may also be an agreement for an indefinite period. A developed LTR market has been considered a factor that not only stabilises the real estate market fluctuations but also may contribute to the overall macroeconomic stability (Czerniak & Rubaszek, 2018; Rubaszek et al., 2024; Rubaszek & Rubio, 2020). Additionally, it may be considered an important factor for mobility in the labour market (Łaszek et al., 2021). However, in 2023, only 4.2% of Polish households rented apartments at market prices, much below the European Union average of 20.2% (Eurostat, 2024a). In Poland, the strong preference for owning instead of renting was documented by Rubaszek & Czerniak (2017) and Bryx et al. (2021). The low share of renters may be attributed, among others, to the sale of state-owned apartments following the systemic changes in the last decades of the 20th century (Ronald, 2008). It has always been regarded as an optimal choice to pursue real assets, as housing was not subject to inflation. Still, the increased state involvement in the construction sector and promotion of access to ownership through various subsidies also played a role in this phenomenon. As a result, housing rental in Poland is rarely considered a target tenure, being most often used by young couples, students (Polityka Insight, 2022), and migrants (National Bank of Poland, 2023), for whom renting is a transitional stage on the way to purchasing an apartment. Although over the last decade, the share of Poles living in overcrowded apartments has decreased by 10 percentage points, to 33.9% in 2023, it is still almost twice as high as in the other EU countries (Eurostat, 2024b). Because of the progress in the Polish economy, it is probable that in the near future Poles will strive to improve their living conditions further. Combined with the ongoing transition from the “ownership society” (Ronald, 2008) to the “generation rent” (Byrne, 2020; Ronald & Kadi, 2018), mainly because of the decreasing price affordability of home ownership, it may trigger the development of the housing rental market. Finally, there is a space for the rental market to grow. Rubaszek & Czerniak (2017) found that apartment rental may be treated as the more favourable tenure by 1/3 of Poles after satisfying certain conditions, most of which may be linked with the higher professionalisation of the market.

On the other hand, the affordability of housing to buy or rent has been recently hindered by the growth of the STR market, resulting in an outflow of apartments from the residential housing stock towards the supply of short-term tourist accommodation. Based on the dataset presented later in this study and GUS (2023), at least 3% of all apartments in the central district of Warsaw – Srodmiescie were actively used for the STR market purposes. In Barcelona or Paris, the share equalled over 2% of all housing units (Garcia-López et al., 2020). Yet, a similar situation might be encountered in any other European city that experiences a high tourist demand (Eurostat, 2024c). In its principle, the STR market enhances resource efficiency by renting unused or vacant housing space. Moreover, it makes tourism more affordable and directs the increased money flow from tourists into tourism-related neighbourhoods and businesses (Barron et al., 2020), thus decreasing local unemployment. However, there are also multiple disadvantages of the uncontrolled growth of the STR market. Firstly, tourists are less prone to adhere to social rules, which disturbs local communities. Secondly, the deregulation of the STR rentals attracts tax-evading-oriented property owners. Finally, one should name gentrification (Wachsmuth & Weisler, 2018) or hotelisation of whole city districts, which decreases the housing supply for citizens. As it cannot be supplemented quickly, it has proven to exert a positive effect on local house prices and rents (Barron et al., 2020; Benítez-Aurioles & Tussyadiah, 2020; Chaves Fonseca, 2024; Garcia-López et al., 2020; Horn & Merante, 2017; Lee & Kim, 2023; Mozo Carollo et al., 2024; Reichle et al., 2023).

However, the COVID-19 pandemic exerted a drastic, multidimensional shock, influencing multiple spheres of social and economic reality (Nicola et al., 2020). Especially at the beginning of the health crisis, the reactions of national governments to the rapid transmission of the virus were similar and resulted in the introduction of numerous restrictions on social and economic life (Hale et al., 2021; ZPP, 2021). From the first day of the pandemic, individuals and economic entities started the process of adjusting to the new reality. Firstly, universities switched to distance learning. Based on the report by Centrum AMRON (2020), 27% of surveyed Polish students decided to terminate the rental agreements in the cities where they studied and return to their family homes. Additionally, 30% either renegotiated the rental price or changed the place of residence to a cheaper one. Secondly, the imposed cross-border traffic restrictions limited the influx of external migrants and suspended international tourism. Thus, many apartments rented on the STR market were moved to the LTR market, increasing the LTR market supply (Boros et al., 2020; Marona & Tomal, 2020). Moreover, introducing lockdowns aimed at preventing COVID-19 transmission (starting in March 2020) resulted in Poland's GDP decline, estimated at -2% in 2020 (Eurostat, 2024d). To stimulate the demand, the European central banks decreased interest rates (in Poland – from 1.5% in 2020-02 to 0.1% in 2020-05). The greater availability of mortgages encouraged people to buy apartments instead of renting and to invest in rental apartments. As a result, it should be concluded that the pandemic-related shocks severely affected both the supply and demand side of the market. It may be assumed that the pandemic lasted in Poland until the first quarter of 2022. At that time, the pandemic restrictions reached the lowest level since the beginning of the turmoil (Hale et al., 2021), and Russia invaded Ukraine, starting a war and exerting a new shock to the Polish housing market.

The impact of the pandemic on the housing rental market may be considered a super-shock as it had the potential to trigger structural changes in the market and establish its new rules (Dolnicar & Zare, 2020). First, it was argued that the microeconomic factors shaping rents have changed, driven by lifestyles adjusted to the pandemic reality (Gallent & Madeddu, 2021; Mouratidis, 2021; Nanda et al., 2021). Then, the adjusted lifestyles transformed the supply and demand structure, especially in those market segments where most consumers were from the most affected social groups. This, in turn, resulted in macro-level decreases in the average rent levels (Kuk et al., 2021; Tomal & Marona, 2021). Yet, based on the state of knowledge documented in the studies already conducted, it may be argued that multiple housing rental market phenomena require further examination in the context of the pandemic, on both micro and macro levels of the market, across its sub-markets and within their segments.

To study changes in the real estate rental market, one can use the theoretical model of DiPasquale & Wheaton (1992). In the four-quadrant analysis, in which the real estate market is divided into the market for space and assets, the rent represents the former. It is situated at the core of the long-run analysis of the economic adjustment processes to economic shocks. On the other hand, as a baseline quantitative method to empirically study the changes in the housing rental market, one should consider the microeconomic hedonic methods. They are a numerical representation of the hedonic price theory developed by Rosen (1974) based on Lancaster's (1966) theory of consumer demand. Its main assumption is that the price of a heterogeneous good (as is an apartment) may be presented as a function of its non-separable attributes. Then, using the econometric hedonic methods, one may decompose prices into the marginal prices paid by tenants for particular housing features, which constitute explanatory variables of the model. Housing characteristics, which prove to be statistically significant in the formation of prices in hedonic models, may be referred to as price-setting or rent-setting factors. Moreover, assuming consumer purchases are utility maximising, and the market prices reflect the equilibrium prices from both tenants' and landlords' perspectives, the obtained marginal prices may be regarded as revealed preferences (Samuelson, 1948). Finally, based on the micro-level hedonic models, one may construct hedonic price indices (HPI) or hedonic rent indices (HRI) (Hill, 2004; Hill & Trojanek, 2022; Tomczyk & Widłak, 2010; Trojanek, 2018), which international institutions recommend for the supervision and monitoring of the housing market (European Commission, Eurostat, Organisation for Economic Co-operation and Development & World Bank, 2013). The goals that guide the preparation of a reliable model focused on investigating the price-setting factors comply with the requirements of a model aimed at tracking market price movements.

To construct hedonic models that may be used to deepen the knowledge of the impact of the pandemic on the rental market, one should use cross-sectional data. For this purpose, the information on micro-level transactions and the broadest possible range of information on individual apartments' characteristics should be considered the most reliable. Among the structural and locational housing features, one should pay particular attention to those that allow the differentiation of market segments to verify the pandemic's impact on specific social groups. Yet, transactional micro-level data concerning the housing rental market of the quality required for hedonic modelling are hardly available. This situation may be encountered in Poland, where a high share of rental transactions is concluded without the intermediary of real estate brokers, with no obligation to report them to public institutions for purposes other than tax. Moreover, access to transactional data is severely restricted because they are a source of competitive advantage (in the case of private entities) or are covered by statistical confidentiality (in the case of public entities). Thus, their usefulness for studying the rental market phenomena in the rapidly changing market situation may be limited.

The problematic access to transactional data forces the usage of alternative sources, out of which housing listings are the most popular and constitute the base for most hedonic studies of the rental market. This kind of data is easily accessible with the use of web-scraping algorithms, and their timeliness is one of their most important advantages over transactional data that often incorporate a time lag between the moment of transaction and its reporting.

Additionally, listings are rich in information on housing attributes required for hedonic modelling. Although the mentioned characteristics position them as a viable source of information on the housing market (Ahlfeldt et al., 2023; Anenberg & Laufer, 2017), also in times of economic turmoil (Lyons, 2019), one should remember that they represent only the market's supply side prices, which may divert from the market equilibrium prices. Thus, using them as a proxy of transactional data incorporates uncertainty (Nasreen & Ruming, 2022) and may require an additional quality adjustment as they may differ in the distribution of prices and housing attributes (Shimizu et al., 2016).

Furthermore, in the housing rental market, there is information asymmetry (Akerlof, 1970), as sellers have an information advantage over buyers concerning the knowledge of the true quality of the offered product. To reduce information asymmetry before concluding the transaction, sellers[1] would send signals revealing the quality to buyers (Spence, 1973, 2002). Recently, the first phase of the signalling process in the housing market has been conducted via listings posted on online platforms. The way the information is conveyed influences consumers' search process for apartments to rent and is reflected in the gathered listing data. The outbreak of the pandemic only elevated its importance, as most real estate agencies admitted increasing the scope of new technologies in the listing process (Marona & Tomal, 2023). Therefore, knowledge of the nature of quality signalling is crucial to understanding rental market phenomena and conducting their reliable analysis.

One may distinguish two types of quality signals sellers send via housing listings – direct and indirect. The direct signals – sellers' declarations of apartment quality may be easily interpreted, at least in their literal dimension. The bigger problem arises regarding indirect signals, most often sent via apartment pictures or their textual descriptions. Because of their structure, these signals are much more detailed but harder to interpret. However, in listing data, they are often the only sources of information on the quality of apartments. Grossman (1981) and Milgrom (1981) showed that sellers must provide buyers with information on the product's quality as its non-disclosure would signal low quality. Still, researchers suggest the existence of some configurations of consumer preferences that may incentivise sellers to signal quality not fully (Hotz & Xiao, 2013). Finally, the quality of apartments itself may be perceived differently. It would not be problematic in the case of hard information (Liberti &

[1] For the needs of this dissertation, the word "sellers" would be used in the context of apartment owners or real estate brokers who list apartments for sale or rent.

Petersen, 2019), which can be expressed objectively regarding housing quality. It covers the characteristics of buildings in which the apartments are located (e.g. age, construction technology or geographical location) or the features of apartments (e.g. number of rooms or presence of a balcony). Yet, the information that describes the apartment's finishing, condition, design or adaptability that are more challenging to quantify may be considered soft information (Liberti & Petersen, 2019). In this case, the information is more subjective and depends on the knowledge of the information provider, their motivations, and the context of the process. For further qualitative considerations, hard/soft information on apartment quality will be referred to as hard/soft quality.

I.3. Justification of the addressed research problem

Even though by 2024 the pandemic-related restrictions have been lifted globally, infectious diseases continue to be listed among the top global risks in both short- and long-term perspectives (World Economic Forum, 2024). In this regard, the role of economic research should be seen in drawing conclusions from the pandemic to provide knowledge on the processes of the rental market adjustment to the structural changes in the economic environment. It will allow us to understand the post-pandemic reality of the market better and optimise the response to similar shocks in the future. Moreover, to stimulate the economically desired growth of the Polish LTR market and to reveal its potential, there is a need to take action to reduce risks associated with insufficient information. In this process, comprehending different sources of information and their analytical advantages and drawbacks should be considered a key factor. Then, by observing the needs highlighted during the study of the pandemic shock, one will be able to adjust the methodological toolkit and make it more robust to problems resulting from missing or misspecified information. This will improve the understanding of hedonic models' micro- and macro-level results and allow for more precise market monitoring and supervision.

Although researchers have already targeted the issue of the impact of the pandemic on particular areas within the housing market, the effect on the rental market has been barely analysed, constituting a research gap. It applies to the quantitative analysis of micro-level changes in rent-setting factors and of macro-level adjustment of rent levels in market segments. Furthermore, the cause-effect relationships between the LTR and STR sub-markets have been scarcely tested, particularly in the context of the pandemic. However, taking account of the volatility of the STR market, which collapsed during the pandemic but has recovered quickly afterwards to an even larger scale, may constitute a crucial input for analyses of the past and future shocks to the housing market. Then, an insufficient understanding of the sources of information on the housing rental market, each of which should be considered imperfect, may be regarded as a significant obstacle to conducting market research. As a result, in the first place, this study attempted to improve comprehension of quality signals and propose analytical techniques adjusted to the specificity of the rental market data.

The following subsections present the specific research gaps in the areas related to the study's hypotheses.

Area 1 – Insufficient knowledge on how to interpret the results of hedonic models of the housing market based on imperfect micro-level data – hypothesis H1.

In the listings-based hedonic models of the housing rental market, researchers have often refrained from calling the coefficients estimated for particular housing characteristics "revealed preferences" in Samuelson's (1948) spirit (Brunauer et al., 2010; Cajias et al., 2019; Efthymiou & Antoniou, 2013; Li et al., 2019; Tomal & Helbich, 2022; Zhan et al., 2023; Zhang et al., 2019). Supposedly because of problems resulting from the imperfections of the data, authors have focused on technical, statistical descriptions of the models' outcomes. It hindered the interpretability of the results and made it challenging to provide policy implications.

Little attention has been to date devoted to understanding the meaning of results obtained based on different data sources, each of which incorporates risks that may lead to biases. This was indicated by Shimizu et al. (2016), who considered the choice of dataset to be critically important in constructing HPI. Still, it should be noted that the scarcity of research on the similarity of results of listings- and transactions-based hedonic models of the rental market is also rooted in the lack of available transactional data. The problem has been described on the example of Poland in Section I.2., yet, it may be encountered also in other European countries, e.g. in Greece, as described by Efthymiou & Antoniou (2013), or in Germany (Ahlfeldt et al., 2023).

To date, researchers who compared the results of the listings- and transactions-based hedonic models have focused rather on finding statistical similarities of the obtained price indices than on exploring the reasons of their discrepancy (Ahlfeldt et al., 2023; Anenberg & Laufer, 2017; Ardila et al., 2021; Knight et al., 1994; Lyons, 2019; Micallef, 2022). On the other hand, studies comparing micro-level coefficients obtained for particular housing characteristics in the housing sales market models based on different data sources have been extremely scarce (Shimizu et al., 2016). Finally, a critical approach to the rental market transactional data has not been found in any previous study. Yet, obtaining knowledge in all the above areas is crucial to understanding the advantages and disadvantages of both listings and transactional data. Only then will it be possible to propose methods that mitigate the drawbacks of listing data and elevate the benefits resulting from using information included in listings that are not present in the case of transactional data.

Area 2 – Lack of understanding of signalling soft quality via housing listings, the relations between different types of soft quality signals and their implications for hedonic modelling – hypotheses H2, H3, H4 and H5.

To date, researchers constructing hedonic models of the housing market based on listing data have often not taken account of soft quality in any way, supposedly because of unavailability or incompleteness of direct textual quality signals and the problematic processing of indirect signals (e.g. Efthymiou & Antoniou, 2013; Löchl & Axhausen, 2010; Sieger & Weber, 2023; Tomal, 2020; Tomal & Helbich, 2022). However, soft quality is among the most important attributes contributing to housing attractiveness (Renigier-Biłozor et al., 2022). Moreover, when included in hedonic models, soft-quality-reflecting variables have always proven to be significant determinants of house prices or rents and elevated the models' statistical properties (Brunauer et al., 2010; de Wit & Van der Klaauw, 2013; Gluszak & Trojanek, 2024; Hahn et al., 2018; Liu et al., 2020; Nowak & Smith, 2017; Olszewski et al., 2017; Seo et al., 2020; Shen & Ross, 2021; Trojanek & Gluszak, 2022). Hence, disregarding sellers' quality signals in hedonic studies may lead to the underuse of existing information (and therefore to less accurate results) while excluding the observations without specific quality signals – to the unrepresentativeness of the studied samples. Yet, to the best of my knowledge, no study has documented the size and specificity of the information gap for certain types of quality signals in housing listings. In particular, no study has inspected the quality-related information disclosure strategies of sellers who list apartments for rent, which is later reflected in the accuracy of rental market data.

Although Luchtenberg et al. (2019) showed that in the housing sales market words might mean more than pictures in terms of attracting customers, no study compared different types of soft-quality signals to decide whether they should be used as complementary or substitute sources of information. Additionally, I have found no study that confronted the outcomes of hedonic models utilising different types of quality signals or discussed the possibility of using the quality signals beyond the scope of hedonic models, e.g. for market segmentation, which might be essential in the times of economic shocks. Then, there has been a further gap in understanding the sensitivity of the hedonic model's numerical outcomes to the changes in the models' composition. It applies particularly to taking account of the shock-induced variability of rent-setting factors, the inclusion of explanatory variables reflecting different types of quality signals and information on apartments' geolocation on different levels of precision. Without knowledge about all of the above issues, it is impossible to decide, which types of information should be targeted in the first place for the needs of hedonic modelling.

Area 3 – Scarcity of quantitative studies that evaluate the rent-setting factors in the LTR market and their changes during the pandemic – hypothesis H6.

To study changes in rent-setting factors during the pandemic, one needs first to establish their current state. Apart from the apartment characteristics conventionally included in hedonic models of the housing market, some features have rarely been analysed, especially in the context of the LTR market. It applies particularly to the location of apartments in renovated or revitalised tenement houses that constitute a high share of the housing stock in the studied local markets (GUS, 2023). The lack of knowledge on how the location in the buildings of a particular type and condition influences rental prices should be considered a gap in understanding the market pricing processes that, based on hedonic price theory, mirror consumer choices. Additionally, including a broader scope of explanatory variables reflecting the location of apartments in buildings should enhance hedonic models' statistical properties, hence the results' precision.

Concerning the pandemic, its consequences at the micro level have yet to be completely understood and tested, constituting a research gap. So far, researchers have discussed its several implications for lifestyle and housing preferences (Gallent & Madeddu, 2021; Mouratidis, 2021; Nanda et al., 2021), which was supported by survey studies (Marona & Tomal, 2020; Noszczyk et al., 2022). However, the quantitative evidence has been limited (Gamal et al., 2023; Guglielminetti et al., 2021; Tomal & Helbich, 2022), and the conducted studies have not targeted certain apartment characteristics, the valuation of which could have changed during the pandemic. Among them, one may distinguish the following: availability of an additional room for remote working or studying, availability of a small private garden, availability of a balcony, proximity to green areas, and proximity to university buildings.

Area 4 – Insufficient knowledge about rental price changes during the pandemic in rental market segments, taking into account spatial, qualitative and sub-market type differentiation – hypotheses H7 and H8.

The literature on HRIs has been for long scarce, supposedly because of the hindered access to data. It has recently started to emerge as researchers have increasingly utilised the potential of online listing data. It led to obtaining HRIs of the LTR markets of the biggest Polish cities by Trojanek & Gluszak (2022) and Gluszak & Trojanek (2024). However, the issue of different dynamics of rents across various market segments has not been studied to date, either in the case of spatial or quality-related differentiation within the LTR market. It may be argued that those distinctions may convey essential, previously unknown information.

Concerning rent changes in the STR market, the literature has been underdeveloped and has not provided evidence of the spatially differentiated impact of the pandemic on STR rents. Moreover, the only studies that targeted the issue of rent changes in the STR market (Boto-García, 2022; Cheung, 2023; Hill et al., 2023) took a tourism industry perspective and analysed the whole market supply. Yet, to understand the linkages between the LTR and STR markets, there is a need to examine particularly this part of the STR market supply, which may be successfully used for both submarkets' purposes. In the context of this dissertation, this group has been defined as apartments located in multi-family buildings.

Area 5 – Insufficient understanding of the relations between the STR and LTR markets concerning rent levels and supply on both sub-markets – hypotheses H9 and H10.

The disruptive influence of the STR market on the entire residential market has been confirmed in the studies of European and American markets by Barron et al. (2020), Benítez-Aurioles & Tussyadiah (2020), Chaves Fonseca (2024), Garcia-López et al. (2020), Horn & Merante (2017), Mozo Carollo et al. (2024), Reichle et al. (2023) and Lee & Kim (2023). Yet, the literature on the topic may still be considered scarce, particularly for the Central and Eastern European (CEE) countries. More research is needed to verify the interrelation of the STR and LTR markets in the still unexamined reality of the relatively little-developed LTR markets, such as the Polish one. First, it applies to the relation between the volume of the STR market and LTR rents. In the above-presented studies, the relationship was studied primarily to find the impact of the growth of supply in the STR market on LTR rents. However, the pandemic reversed the trend, and because of the frozen tourist industry, multiple STR apartments have been moved to the LTR market (Boros et al., 2020; Marona & Tomal, 2020). Thus, to understand the pandemic-related behaviour of rental prices in the LTR market, there has been a need to verify whether the relation between changes in the STR market size and the LTR market rents also holds in the opposite direction. It has been studied by Mozo Carollo et al. (2024) and Batalha et al. (2022). Still, to the best of my knowledge, the issue has been targeted for the first time in the study by Trojanek et al. (2021), which is a part of this dissertation.

Finally, because of the scarcity of research measuring rent levels in the STR market, it has not been possible to formally study the price dependency of the LTR and STR markets. Yet, besides establishing knowledge of the relationship between the STR market supply and the LTR market rent level, inspecting relations of the submarkets' rent levels may be crucial to understanding the processes of transmission of apartments between the markets. This, in turn, lies at the root of the problematic decline in housing affordability caused by the development of the STR market. Thus, insufficient knowledge about the submarkets' dependency should be considered among the major obstacles to providing policy implications to mitigate the STR market's negative externalities.

I.4. Research goals

I.4.1. Main objectives

The dissertation aims to provide comprehensive evidence on the multi-dimensional impact of the pandemic on the housing rental market. Moreover, its key goal and the intrinsic part of the exploratory process is to improve the understanding of different sources of information and the results of hedonic analyses based on them. Since the rental market is a place of asymmetric information, all the sources should be considered imperfect. The lack of data of the quality required for hedonic modelling has often hampered the analytical processes of the rental market. Yet, this dissertation aims to show that rather than resigning from conducting the analyses, one should improve the understanding of the available data to exploit their previously untapped potential.

Additionally, there has been a need to adjust analytical methods and widen their scope to provide information on the market response to the pandemic shock. It applies to the adjustment resulting from the changing economic environment and the improving knowledge of the market phenomena. This dissertation identifies inefficiencies of the widely used methodological approaches and proposes solutions to the detected problems. Therefore, the study aims to go beyond understanding one particular economic shock. In the first place, the pandemic should be seen as a period highlighting the importance of previously disregarded research areas concerning market phenomena. To understand better any upcoming economic shocks, they should be included in future research.

I.4.2. Specific objectives

Objective 1 - "Information gaps" – verification of the usefulness of different data sources and novel analytical techniques for studying phenomena in the rental market; proposing the method that bridges the gap in information on the quality of apartments.

First, Hebdzyński (2024c) deepens the understanding of the results of hedonic models of the housing rental market based on different data sources. The study adds to the discussion on whether it is possible to use the listings-based indices as a sufficient proxy of transactional indices or whether listings should be treated only as a source of supplementary information. It aims to assess what part of the difference in the models' outcomes should be attributed to the difference between the listed and transacted prices and what part stems from the structural differences between the different types of datasets used. Furthermore, it addresses the issue of how to understand the coefficients obtained for particular housing characteristics in hedonic models. In the above contexts, it critically assesses not only the listings-based but also the transactions-based datasets, opening the scientific discussion on the topic.

Secondly, Hebdzyński (2023) provides information on sellers' listing strategies and documents the size of the information gap for the sellers' direct textual signals of apartment quality. To bridge the gap, the study proposes a novel adaptation of the Wordscores algorithm (Laver et al., 2003). It shows a way to calculate an independent measure of apartment quality that aims to reflect indirect textual quality signals. Then, Hebdzyński (2024a) empirically verifies the linkages of various types of quality signals with the value of individual apartments for rent and discusses their usefulness in analysing the quality-related market segments. Finally, Hebdzyński (2024a, 2024b, 2024c) widens the understanding of the robustness of the results of hedonic modelling to the choice of analytical approach, to the set of explanatory variables and to their changes over time.

Objective 2 - "Rent-setting factors" - identification of rent-setting factors in the LTR market and their changes during the pandemic.

Hebdzyński (2024b) provides information on the relationship between rents and certain housing characteristics that have not been tested in previous hedonic studies of the housing market. Then, the study focuses on changes in the valuation of particular housing features during the pandemic. As a result, it broadens knowledge of the micro-level adjustment processes to the shock resulting from the pandemic.

Objective 3 - "Rent changes" – identifying rental price fluctuations during the pandemic and assessing price-related interdependency between the LTR and STR markets.

Trojanek et al. (2021) and Hebdzyński (2024b) provide evidence on the LTR market rent changes amidst the pandemic. Then, Trojanek et al. (2021) and Hebdzyński & Trojanek (2024) analyse changes in HRI within different spatial segments of the market to document the unequal influence of the pandemic shock. On the other hand, Hebdzyński (2024a) focuses on rent changes across the market's quality segments. Similar evidence, but restricted to the STR market, has been provided by Hebdzyński & Trojanek (2024). Moreover, Hebdzyński & Trojanek (2024) and Trojanek et al. (2021) investigate the interdependence of the STR and LTR markets. Trojanek et al. (2021) discuss the pandemic-related decrease in the STR market volume and present it as a source of shock to the LTR market. Then, Hebdzyński & Trojanek (2024) assess the dependence of the sub-markets' rent levels.

I.5. Theoretical and conceptual foundations of the study and research hypotheses

H1: The LTR market listing data may act as a proxy of transactional data in terms of their usefulness for constructing hedonic models.

To date, in the listings-based hedonic studies of the LTR market, researchers have often refrained from calling the obtained models' coefficients "revealed preferences", as suggested by the revealed preference theory (Samuelson, 1948), presumably because of imperfections of the used data. Instead, they have often used phrases that have disconnected the market valuations of individual housing characteristics from consumers' choices in the LTR market that reveal indirectly their preferences. Li et al. (2019), Zhang et al. (2019), Zhan et al. (2023) and Efthymiou & Antoniou (2013) focused on the econometric, technical interpretation – "determinant of housing rent" or described characteristics' "association with rents". Similarly, Brunauer et al. (2010) and Cajias et al. (2019) considered the "effect of" housing characteristics on rents, and Tomal & Helbich (2022) referred to "marginal prices of characteristics". Among studies that connect the issues of rent-setting factors and consumer preferences are Crespo & Grêt-Regamey (2013) and Tomal (2020), who wrote about "tenants' willingness-to-pay". Finally, Tomal & Helbich (2023) and Sieger & Weber (2023) used the word "preferences", while the latter study explained that the hedonic attributes based on listings might be considered an accurate representation of revealed preferences only when the rental values would be not negotiated, hence reflecting market equilibrium. Otherwise, when studying the revealed preferences based on listings, one would achieve a supply side, landlord's valuations of particular housing characteristics, rather than tenants' preferences. Yet, if we assume that the final rental price reflects the market equilibrium price, then the landlords' valuations and tenants' preferences will be identical. As a result, in the process of revealed preferences' modelling using listing data, it is the scale of rent negotiation that would determine how much the proxied revealed preferences (obtained based on listed rents) deviate from the actually revealed preferences (obtained based on transacted rents) for the studied sample.

Then, to decide whether it is appropriate to call the listings-based hedonic coefficients "revealed preferences", the results of the hedonic decomposition of the listings- and transactional data should be compared. Shimizu et al. (2016) conducted a micro-level hedonic study of Tokyo house prices at four stages of the sales process, starting from the initial asking price and ending with the final transaction price reported by the buyer. Yet, for each stage, a separate dataset was used. The models that regressed the prices were obtained using quantile regression, and, as a result, no large differences between the models' coefficients were found. The discrepancies in estimates for floor space, age of the building and two commuting-related neighbourhood variables were no bigger than 20%. Secondly, Kolbe et al. (2021) analysed the willingness-to-pay functions (constructed separately for each housing feature but dependent on all estimated coefficients). Contrary to Shimizu et al. (2016), the authors argued that based on the model for house prices in Berlin, the listings- and transactions-based functions for the variable reflecting the age of the building in which the apartments were located differed substantially. At the same time, functions for apartment area were considered similar but not identical. To my best knowledge, equivalent research has not yet been provided for the rental market, which has been done in this dissertation to verify the H1 hypothesis.

On the other hand, to decide on the possibility of using listings as a proxy of transactional data for the needs of constructing the HRI of the LTR market, one should confront two HRIs that represent:

I. The same geographical location

II. The same time range

III. The targeted part of the housing market – LTR market

IV. Listings and transactions of exactly the same apartments

Yet, no study to date has fulfilled all four conditions, hence testing the H1 hypothesis aimed at bridging the gap. While satisfying conditions I and II has been a minimum requirement for the study to be regarded as a reference point for testing the H1 hypothesis, particular attention should be paid to the research that satisfies also conditions III or IV. Micallef (2022) satisfied condition III to prove that the HRIs of listings and transactions are highly correlated in Malta's market. However, he argued that it is necessary to use a longer time series to understand the relation between indices, their timing differences and co-movements in the business cycle. Then, Knight et al. (1994), who satisfied condition IV in the study of Baton Rouge (Louisiana, the USA) found that even though list prices prove to Granger-cause sales prices, they are least informative at peaks and troughs of the business cycle, i.e. in times when timely, precise indices are most needed. Nevertheless, they have proven that listings-based HPIs may successfully predict future housing prices.

Other researchers satisfied only conditions I and II. Ahlfeldt et al. (2023) argued that because of the low accessibility of transactional data in Germany, the listings-based indices may be used as a decent indicator of market trends. In the study, the listings- and transactions-based indices exhibited a positive correlation, and the indicated price trends were very similar in 2007-2016. Anenberg & Laufer (2017) (based on data from the biggest cities in the USA) argued that listings-based indices are accurate and are leading indicators of the level of transaction prices. Similarly, Lyons (2019) applied the Granger causality test to the Irish market and found that even during market turmoil, list price indices can be used as leading indicators of the state of the sales market. On the contrary, Ardila et al. (2021) found that the listed and transacted prices in Switzerland do not Granger-cause each other. However, they have proven to be co-integrated in various market segments. Thus, given the limited availability of transactional data in Switzerland, listings may be considered suitable substitutes for transactional data.

Lastly, one should mention the study by Shimizu et al. (2016). Although it satisfied conditions I, II and IV, used hedonic methodology, and focused on the needs of constructing price indices, the study did not present the index itself. Instead, in the research on Tokyo house prices, the authors focused on the similarity of distributions of prices and house attributes in datasets collected from independent sources, reflecting the various stages of the sales process. They found the price distributions unequal and attributed the differences to the quality structure of datasets. To focus on pure differences between price distributions, they proposed two approaches. The first was to use only observations referring to the same apartments, for which information from the compared stages was available. Secondly, they conducted the quality adjustment based on quantile hedonic regression as proposed by Machado & Mata (2005). After utilising both approaches, they detected only minor differences between the distributions of listed and transacted prices. Although they concluded that list prices could be used to construct HPIs, they argued that to do so, a quality adjustment is necessary.

H2: Hedonic rent indices of the LTR market are robust to minor changes in the selection of the explanatory variables in hedonic models.

Diewert & Shimizu (2022) indicated that the crucial information needed to construct property price indices is the apartment's area, the building's age and its geolocation characteristics. Then, although adding other explanatory variables to the model would increase its precision, the effect on the index would be minimal. Similarly, Micallef (2022) studied the impact of including the geolocation of apartments at various levels of detail and did not detect any considerable differences between the two analysed approaches. Finally, Hill & Trojanek (2022) provided a comprehensive review of the methods of construction of HPIs to show that as long as the hedonic methods are used, the results are relatively robust to the choice of the index variant. In this context, only slight differences between the HPIs obtained using the pooled time-dummy method and the rolling time-dummy method (which allows for parameter changes in time) were noticed.

To date, the issue of the robustness of HPIs has been studied mainly in the context of the housing sales market, leaving the LTR market barely investigated. Moreover, there is a need to provide information on the sensitivity of HRIs to accounting for the phenomena which have been targeted in this dissertation – the pandemic-induced change of rent-setting factors, the utilisation of different quality signals and the compositional changes of the set of explanatory variables used for hedonic modelling. Nevertheless, it may be expected that, similarly to the housing market and the results presented by Diewert & Shimizu (2022), the HRIs would also show minor sensitivity to the discussed issues.

H3: Agents and landlords who list apartments for rent on the LTR market tend to overstate their declarations regarding the quality of the apartments.

Since access to complete information concerning the LTR market phenomena is hindered, the market cannot be considered fully efficient (Cheng et al., 2015, 2020). The problem is rooted in the information asymmetry (Akerlof, 1970). Yet, sellers may reduce it and send signals to reveal the information on the quality of the offered product to buyers (Spence, 1973, 2002). Recently, the first phase of signalling the quality of apartments for LTR rent has been done most often via listings posted on online platforms. A typical listing includes:

- title,
- information about the rent,
- a table grouping structural characteristics of the apartment,
- a declaration of the offered apartment quality chosen by the seller from a list of pre-specified quality labels, which may be referred to as *a direct textual signal of quality*,
- a textual description of an apartment – *an indirect textual signal of quality,*

- photos covering the interior and exterior of the apartment and the building in which it is located – *an indirect visual signal of quality*.

It has been documented that disclosing non-full information to some types of consumers of complex products might be beneficial for sellers (Hotz & Xiao, 2013). It has also been shown for the housing market listings where including less information may have both positive and negative trade-offs (Benefield et al., 2011; Bian et al., 2021). Goodwin et al. (2014) argued that when increasing the information load of the listing, the probability of enclosing the information that would discourage the potential buyer from arranging an in-person visit to the apartment also rises. They found that the carefully prepared descriptions may impact the selling price, time-on-market and the probability of concluding the transaction. To better understand the true meaning of apartments' descriptions, Goodwin et al. (2018) constructed dictionaries of words with positive, negative, or neutral meanings in the reality of the housing market in the USA. However, Haag et al. (2000) found that some sellers' remarks on the listing platforms may be classified as hype because certain strictly positive comments were associated with lower sales prices. Altogether, it may be hypothesised that the informational advantages resulting from information asymmetry may encourage housing sellers to ill-inform potential customers about the quality of apartments. To my knowledge, the sellers' listing strategies concerning soft quality signalling have not been studied in housing sales or the LTR market. Thus, the H3 hypothesis has been stated.

H4: The Wordscores algorithm may be used to extract quality signals from textual descriptions included in the LTR market listings.

H5: The underspecification of hedonic models of the LTR market in terms of housing quality leads to obtaining inferior statistical properties of the models and imprecise estimates.

Nowak & Smith (2017), Liu et al. (2020) and Nowak et al. (2021) argued that real estate hedonic models that do not include quality-related explanatory variables might not only be incomplete but also biased. To mitigate the problem, they transformed listing texts into tokens that reflect the presence or absence of each word in each listing. With the LASSO algorithm, they selected the most important tokens to form dictionaries of words to proxy for the omitted, quality-reflecting variables in hedonic models of the housing sales market. They showed that adding the selected tokens as explanatory variables in the hedonic models improved their statistical properties. Similarly, Shen & Ross (2021) used texts to construct an unsupervised machine-learning algorithm based on natural language processing to measure the apartment's "uniqueness", which should reflect the unobserved quality of an apartment and its market power compared to properties in its neighbourhood. They showed that adding the uniqueness as an explanatory variable absorbed a significant part of the model's residual variance. Seo et al. (2020) focused more directly on soft quality and studied descriptions in housing listings to specify seven phrases related to soft quality. First, the phrases appearing in the descriptions were manually assigned to the groups. Then, focusing on the presence or absence of the phrases from particular groups in listings, the impact of providing some specific information concerning the quality of apartments on selling prices has been proven. Other researchers utilised sellers' signals of quality and arranged them on the ordinal scale to include as one explanatory variable in the hedonic model (Brunauer et al., 2010; de Wit & Van der Klaauw, 2013; Trojanek & Gluszak, 2022). Finally, Olszewski et al. (2017) and Hahn et al. (2018) focused on direct textual signals of quality and implemented quality-related dummy variables, thus not specifying the differences between the assessment levels.

Even though the variables reflecting soft quality have proven to be significant determinants of house prices and rents and elevated the quality of hedonic models in all of the mentioned studies, researchers have most often constructed models that did not take account of the apartments' soft quality (e.g. Efthymiou & Antoniou, 2013; Löchl & Axhausen, 2010; Sieger & Weber, 2023; Tomal, 2020; Tomal & Helbich, 2022). One of the reasons for this is the rarity of data containing direct signals and the problematic processing of indirect signals. However, methods to extract quality information from housing listings have been recently developed. One approach is to use machine-learning visual quality assessment methods based on the photos of apartments. In this manner, Poursaeed et al. (2018) estimated the luxury level of apartment photos, which may represent soft quality. Using crowdsourcing, the researchers trained the dataset by comparing pairs of housing photos in terms of quality. Finally, the luxury level value was assigned to each room type in each listed apartment. Although computationally demanding and resource-costly, the method has successfully reduced the median error in the mass appraisal model.

The other approach entails utilising the information included in listings' textual descriptions. Although focused strictly on the needs of hedonic modelling, the earlier discussed methods developed by Nowak & Smith (2017), Liu et al. (2020), Nowak et al. (2021), Shen & Ross (2021) and Seo et al. (2020) have proven to be successful in extracting the soft-quality-related information. Even though, to my best knowledge, the supervised machine-learning methods (that require training datasets of the labelled data) have not yet been used in housing quality research, the already developed methods offer a wide range of practical applications. One of those is the Wordscores algorithm (Laver et al., 2003), designed for extracting textual sentiment[2] from political statements. Its main idea is to construct a dictionary of words with the assigned sentiment scores based on the already classified texts. Then, the dictionary is used to read non-classified texts and automatically estimate their overall sentiment. Contrary to the regression-based techniques, the Wordscores builds on the bag-of-words approach that utilises the content of whole texts without selecting only statistically significant words or phrases. Moreover, the method does not pre-specify the sign of correlation of the used words with the target value. It is essential given the results of the study by Haag et al. (2000), based on which it may be inferred that some strictly positive words may be linked with a low quality of apartments. Furthermore, the method allows the phenomenon represented in the classified texts to be targeted directly. It contrasts the methods used in the previous quality-related studies that focused primarily on reducing errors of hedonic models, not on the soft quality itself. Finally, the method has proven to be successful in sentiment analysis, yet simple, compared with the already developed and empirically tested textual and visual analysis approaches.

In the H3 hypothesis, it has been suspected that direct soft-quality signals are incomplete or biased. Thus, knowing the vital role of apartment quality, one should see a high potential in analysing indirect signals – either visual or textual. Yet, given the information asymmetry in the housing market, there is no method for ideally assessing the quality of the housing listed for LTR rent. It can be assumed that to get the most objective picture of the quality signal of the listings, one should carefully study the photos of each listed apartment and assess them using clear and unified guidelines. Still, it should be considered resource-costly and time-consuming, limiting the beneficial characteristics of listing data, particularly during economic shocks. Additionally, the quality assessment conducted this way would aim instead at assigning the listed apartments to quality-related groups rather than assessing the quality signal's intensity. In contrast, the supervised machine-learning method – Wordscores, would

[2] Sentiment – a latent variable to indicate the semantic orientation of text.

require evaluating the quality of the listed apartments only once. Later, the constructed dictionary may be used for further, repetitive studies. Importantly, all the observations would be quality assessed using the same rules, which may reduce the influence of unavoidable subjectivity of individual soft quality assessments and lead to the higher comparability of the generated quality assessments. Additionally, utilising the Wordscores approach will allow for assessing the intensity of each soft-quality signal, which may have beneficial properties for hedonic modelling. Based on all of these characteristics, the H4 hypothesis has been formed.

The crucial role of soft quality in shaping housing prices has been repeatedly highlighted. Yet, it may be argued that in the sales market, the purchased apartments are often renovated quickly after concluding the transaction. Thus, the importance of their soft quality in the moment of transaction might be lower than in the rental market, where a tenant does not desire to implement any significant changes to the apartment quality. Thus, it may be suspected that the importance of soft quality for hedonic modelling in the LTR market might be even higher than that shown in the sales market. By specifying the H5 hypothesis, this research aims to find the consequences of disregarding soft quality considerations in hedonic models. It should also help to understand the information load of different kinds of quality signals concerning their impact on the results of hedonic models.

H6: During the pandemic, the rent-setting factors in the LTR market changed.

The literature on the rent-setting factors in the LTR market has to date concerned mainly the Chinese market, which is rapidly developing because of the increasing “floating population” – workers migrating temporarily within the country (Cui et al., 2018; Li et al., 2019; Liu et al., 2022; Zhan et al., 2023; Zhang et al., 2019). A different strand of research has been devoted to exploring the impact of energy efficiency on rents in Germany (Cajias et al., 2019; Hahn et al., 2018; Kholodilin et al., 2017; Sieger & Weber, 2023). Other notable studies analysed the market of Switzerland (Baranzini & Ramirez, 2005; Crespo & Grêt-Regamey, 2013; Löchl & Axhausen, 2010), Greece (Efthymiou & Antoniou, 2013), the United Kingdom (McCord et al., 2014), the United States (Sirmans et al., 1989), and the Netherlands (Tomal & Helbich, 2023).

Among the Central and Eastern European countries, the research targeted, to the best of my knowledge, only the Polish market, focusing primarily on the rent-setting factors (Tomal, 2020; Tomal & Helbich, 2022, 2023) or HRIs (Gluszak & Trojanek, 2024; Trojanek & Gluszak, 2022). The apartment characteristics that contribute to the formation of rents have often been divided into structural and locational. As for the variables conventionally included in the hedonic models, it has been proven that floor area positively impacts rents, as does the number of rooms. However, the latter has not always been statistically significant. Then, depending on the approach taken, the type of the building in which the apartment is located, its age, and construction technology are important rent determinants. Every additional floor of the building has proven to decrease rents, while the higher location of the apartment in the building and the presence of an elevator are priced positively (Tomal, 2020; Tomal & Helbich, 2023). Furthermore, the additional spaces – garage, basement and usable room were analysed. When studied together, the features have little or no significance in Krakow (Poland) (Tomal, 2020; Tomal & Helbich, 2022). The presence of a small garden terrace has been shown to have a high impact on rents in Zurich (Switzerland) (Löchl & Axhausen, 2010), similarly to furnishing studied in Nanjing (China) (Zhang et al., 2019). Regarding locational characteristics, the distance from the building in which the apartment is located to the city centre was most often included in models, showing a negative influence on rents. At the same time, the distances to the nearest park, university, school or water reservoir have proven to affect rents negatively (Tomal, 2020), while the distance to university and the closest public transport stop revealed their negative relation with rents (Tomal & Helbich, 2023).

There is also one apartment characteristic, which to date has not been analysed in the context of the LTR market, namely the location of apartments in renovated or revitalised tenement houses. As described in the example of Wroclaw by Marcinkowska et al. (2015), most of the tenements in Poland were built from brick, using similar construction solutions from the second half of the 19th century until World War II. In 2021, around 15% of all apartments in Poznan were located in such buildings (GUS, 2023). Although their maintenance was careful before the war, in the post-war times of socialism, the necessary repairs were not provided, and the buildings have often deteriorated. Nevertheless, the increased demand for apartments in tenements has been noticed in Poland recently (Marcinkowska et al., 2015). Therefore, renovations of tenements have been taking place, and some of them are being revitalised. In Poland, revitalising a historical building should be perceived as an adaptation to meet requirements similar to those imposed on newly constructed ones (Terlikowski, 2013). Thus, it may be hypothesised that the valuation of renovated and non-renovated tenements might be drastically different, and the issue should be included in hedonic models of the LTR market.

Marona & Tomal (2020, 2023) discussed the impact of the pandemic on tenants' preferences. The studies conducted in two phases of the pandemic in Krakow found that most surveyed real estate agents signalled the change in their clients' attitudes. Moreover, the authors argued that it is highly likely that the changes will last longer than the pandemic and will become permanent. Among them, the increased demand of tenants for apartments with access to balconies or private gardens was indicated, highlighting the change in how to spend free time. It was confirmed in the survey study in Italy (Guglielminetti et al., 2021), where increased interest in apartments with private gardens was found. On the other hand, Nanda et al. (2021) argued that as homes had to adapt to new roles – providing space for efficient work from home, the demand for an additional, separate room has increased. Mouratidis (2021) considered the increased need for larger, high-quality apartments that allow for comfortable leisure and performing the work-related tasks. It was also highlighted by agents surveyed by Marona & Tomal (2020).

Concerning locational characteristics of housing, Nanda et al. (2021), Gallent & Madeddu (2021) and Liu & Su (2021) argued that changes in working patterns had reduced the importance of access to city centres, where business premises were often located, which Tomal & Helbich (2022) empirically confirmed for the first phases of the pandemic. Gamal et al. (2023) and Tomal & Helbich (2022) found the decline in demand for apartments in dense, multi-unit buildings and linked it to the need to internalise the risk of infection. Similarly, Guglielminetti et al. (2021) showed that the stay-at-home orders in Italy may have increased demand for less congested areas. Then, Tomal & Helbich (2022) found that the distance from university buildings was gradually losing importance because of the introduction of online studies. Nanda et al. (2021) and Mouratidis (2021) discussed the recreational needs that can be satisfied by access to urban green areas, where the risk of infection is relatively lower. It was supported by the survey study of Krakow by Noszczyk et al. (2022), who added that during the pandemic, visits to green areas had a crucial role in citizens' mental health. Lastly, Broitman (2023) showed the growing willingness to live near urban green areas but warned that the increased prices of such located housing might result in the displacement of low-income residents and "ecological gentrification".

However, the quantitative evidence of the micro-level impact of the pandemic on the LTR market should still be considered limited. Based on the qualitative studies, particularly Marona & Tomal (2020, 2023) and Nanda et al. (2021), it has been hypothesised that during the pandemic, the valuation of the availability of a balcony/terrace and private garden increased, together with the marginal price paid for the availability of an additional room for remote working or studying. For the locational characteristics, it has been suspected that the value of proximity to urban green areas increased, and the value of proximity to university buildings decreased.

H7: During the pandemic, rent levels in the STR and LTR markets declined, but the reactions varied spatially.

H8: During the pandemic, rent levels in the LTR market declined unevenly across quality-related market segments.

In the theoretical model by DiPasquale & Wheaton (1992), the demand for space is assumed to come from tenants and owners who occupy their properties. Then, it relies on the current rent level and exogenous economic factors like number of households or their income. On the contrary, supply is linked with the market for assets, as it depends on real estate asset prices and construction costs. Although the model's framework focuses on individual economic shifts, the net effect of multiple simultaneous changes should reflect the combination of individual impacts. Within the framework, the events accompanying the pandemic – the return of students to their family homes, the suspension of temporary labour migration, and the slowdown of economic activity may be perceived as factors that decrease the demand for space, pushing the LTR rents down. Similarly, the interest rate reduction should contribute to a long-run rent decline. Finally, converting some STR apartments to LTR purposes should constitute an exogenous shift in the housing supply, driving rents down. As a result, rent reductions should be expected during the pandemic.

The empirical validation of economic theory is often problematic because of the data availability problems. Therefore, simple aggregation methods are usually utilised to measure price changes – mean or median. However, they may lead to false conclusions, especially during market turbulence. Relying on mean or median, the periods in which relatively more apartments of higher quality or larger floor area, hence of higher rent, were rented might be wrongly considered as periods of a general increase in rent levels. To mitigate the problem, researchers use matching methods to construct price indices to find consecutive observations of transactions (or eventually listings) of the same properties. The most common approach of this kind is the repeat sales method proposed by Bailey et al. (1963) (BMN). In the housing sales market, specific apartments are the subject of transactions more often, while some may be transacted only once. Therefore, finding multiple transactions of the same apartment might be problematic, and excluding single-transaction apartments from the analytical dataset may lead to sample selection bias (Diewert, 2009). Yet, this should not be the case for the rental market, where apartments are transacted more frequently. For the paired micro-level observations, the matching methods demand no data other than individual rents for the apartments and periods of rent observation.

However, the rental values of apartments tend to depreciate over time as their soft quality deteriorates and apartments are subject to renovations. As a result, disregarding quality considerations and using matching models may also be considered oversimplification and lead to false conclusions. This argument should be particularly valid in the case of LTR apartments, which are subject to transactions less frequently than STR apartments. Their quality may considerably differ between two transactions. Therefore, one should focus primarily on quality-adjusted methods invariant to changes in the structure of the analysed sample between the studied periods, such as hedonic regression methods.

Hedonic research on HRIs of the LTR market has been to date scarce and limited to a few scientific articles – Gluszak & Trojanek (2024), Trojanek & Gluszak (2022), Micallef (2022), Kuk et al. (2021), Eichholtz et al. (2012), and Hoffmann & Kurz (2002). Gluszak & Trojanek (2024) and Trojanek & Gluszak (2022) studied the Polish LTR market and provided the quality-adjusted HRIs of the biggest Polish cities. In both articles, the authors constructed hedonic models that took into account quality in both the hard and soft senses. As a result, the geolocation of apartments has proven to be among the most critical, non-alterable characteristics of apartments that contribute to their hard quality. On the other hand, variables representing soft quality also proved to be highly important. Because of the particular significance of the location of apartments and their soft quality, it may be suspected that the LTR market may be segmented according to these characteristics. Different groups of consumers, who react differently to economic shocks, may target each segment. Knowing that the pandemic shock affected specific supply and demand groups unevenly, one might suspect that location- and quality-related market segment reactions may also differ. In this context, it can be hypothesised that rent declines were greatest in districts where the scale of conversion of apartments from the STR market to the LTR market was the largest, i.e. city centres. The STR apartments are generally of higher quality than the LTR apartments (Hill et al., 2023). Thus, their inflow into the LTR market should increase the high-quality market supply, reducing rents. On the other hand, the drop in student demand and the suspension of labour migration should result in a decline in the low-quality market.

Yet, probably because of restricted access to data with a time window of the required length, the impact of the pandemic on LTR rents has rarely been studied. In this case, the research concerning the housing sales market should not be used as a reference point. The reason is that negative changes in interest rates during the pandemic may have increased the relative demand for home ownership while reducing the propensity to rent. As apartment ownership and rental may be considered substitute goods, the usefulness of information on changes in the housing sales market would be limited for the LTR market. Then, one should focus solely on the studies of the LTR market. In the first two phases of the pandemic, researchers (using quality-unadjusted methods) pointed at either no significant changes in the Austrian market (Kadi et al., 2020) or sharp declines in Krakow (Tomal & Marona, 2021). In the Polish market, the study by Trojanek et al. (2021), which is a part of this dissertation, should be considered the first to target the construction of HRI in Poland. Its results have been presented in Section I.7. Then, the study by Trojanek & Gluszak (2022), using a similar methodology (quantile regression) added further information on the rent changes during the pandemic in Krakow and Warsaw, valuing them in overall at -13.3%, -14.2% and -12.1% for Krakow and -13.7%, -12.3% and -11.4% for Warsaw, respectively for 25th, 50th and 75th percentile of the conditional distribution of rents.

Boto-García (2022) was, to the best of my knowledge, the first to discuss the changes in the STR market rents using the quality-adjusted, hedonic framework. Although the division into professional and non-professional hosts presented in the study shed light on the heterogeneous price response to the pandemic shock by different STR market agents, only the second phase of the pandemic was analysed (from June 2020 to July 2021), with no reference to the pre-pandemic times. The greatest STR rent declines have been noticed in October, November, and December 2020. Still, it has not been possible to indicate whether some part of the change was not a result of standard, seasonal market behaviour. Thus, the authors encouraged future researchers to study the seasonal differences in the STR market pricing. Another notable study conducted by Hill et al. (2023) for the Sydney's (Australia) market using hedonic methods indicated that the STR market rents were constantly falling in 2016-2018. They argued that the price change is consistent with the scale of transformation of STR apartments from the LTR to the STR market purposes.

Finally, Cheung's (2023) study of the New Zealand's STR market should be considered the only one to approach the topic of rent indices of the STR market both before and during the pandemic shock. Although, in principle, the study did not focus on the empirical examination of the pandemic shock and provided the seasonally unadjusted time series, the constructed rent index did not show signs of a decrease in STR rents during the pandemic. From the methodological point of view, since the study utilised the data of the same structure and obtained from the same source as the data used in this dissertation, it may be considered a good reference point. The data concerned STR apartment rental transactions concluded using the most popular STR market platform – Airbnb. In the constructed hedonic model, the geolocational variables included on the neighbourhood level have proven statistically significant for shaping rents. Nevertheless, the author highlighted the problem of imprecision of the possessed information about the location of apartments. Thus, he could not fully control one of the essential features of STR apartments (geolocation) and had limited information about their characteristics (lack of soft quality information), so he emphasised the advantages of repeat sales methods. In this case, instead of using imprecise information about the apartment's location, one may pair the subsequent observations of the same apartments using the precise identifiers generated by the listing platform. As a result, in the repeat sales model, the author focused on the need to take into account the term structure of the STR market transactions to obtain more reliable results concerning rent change dynamics. Nevertheless, it may be argued that utilising the locational data to provide the results on the different spatial segments of the market may reveal further, previously unknown information about the STR market rent changes. It may be suspected that the reactions of STR rent levels were stronger during the pandemic in those segments that suffered the most from the drop in demand induced by the restrictions imposed on international tourism, i.e. in city centres.

H9: Changes in the STR market rents cause changes in the LTR market rents.

Even though the pre-pandemic volume of the STR market has been considered harmful to housing affordability, in the recent five years, the STR market has developed even further, particularly in Poland (Eurostat, 2024c). In Warsaw, the number of guest nights transacted via the most popular STR booking platforms (Airbnb, Booking, Expedia Group, Tripadvisor) increased from 2.5 million in 2018 to over 4 million in 2023. The corresponding values for Krakow equalled 2.8 and 3.7 million and for Poznan – 0.38 and 0.66 million. Although the base numbers were considerably higher in Barcelona and Paris, they have also increased, yet to a lesser extent – from 12 million in 2018 to 12.7 million in 2023 in Barcelona and from 10.1 to 12.2 million in Paris.

The recent literature has just started to measure and understand the linkages between the STR and LTR markets. Barron et al. (2020) argued that the growth of the STR market volume comes solely at the expense of the LTR market supply, without an influence on the total housing supply. Then, according to the studies conducted on local markets by Horn & Merante (2017), Benítez-Aurioles & Tussyadiah (2020), Garcia-López et al. (2020), Barron et al. (2020), and Chaves Fonseca (2024), the presence of STR rentals exerts a positive effect on LTR rents. It has also been confirmed in the study of multiple international markets by Reichle et al. (2023). Then, Mozo Carollo et al. (2024) showed that the effect is stronger in mid-sized cities, while Lee & Kim (2023) added that the influence is unequal when it comes to different types of housing listed in the STR market and that there may occur spillover effects of the STR market growth to the neighbouring regions. Finally, Hill et al. (2023) compared price indices of the STR and LTR markets in the years 2015-2018 to suggest that from 2016 the markets' rent levels were diverging from each other, which is consistent with the increase in supply in the STR market at the cost of supply in the LTR market.

Kadi et al. (2020) argued that among the possible incentives to enter the STR market, there are mainly higher rental income (in comparison with the LTR market), greater flexibility and, in many cases, favourable regulatory treatment (which stems from the STR market deregulation rather than from the optimistic approach of governments). Hill et al. (2023), on the example of Sydney's market, have proven the financial profitability of the conversion of the apartment from the LTR to the STR market. They argued that the average STR rent premium over LTR may be from 55% to 131%. The higher values were obtained for areas more attractive to tourists and for more expensive and larger apartments. However, Horn & Merante (2017) indicated that the STR market's profits are often overestimated because STR properties tend to remain vacant longer than expected. Nevertheless, renting on the STR market is more challenging to track than the formally regulated LTR market. As a result, it also attracts tax-evading-oriented property owners, who aim to earn even higher financial premiums this way. Unluckily, the outflow of apartments from the supply of residential housing into the STR market cannot be quickly supplemented by the increase in housing supply (Barron et al., 2020; Lee, 2016).

It may be suspected that a sudden rise in tourist interest in a given region (e.g. induced by a marketing campaign) may increase demand for STR rentals. This, in turn, would trigger the growth of the STR rents and lead to a higher rent premium over LTR rental. At the same time, the willingness of apartment owners to switch their properties from the LTR to the STR market would also increase, reducing the LTR market supply. On the other hand, the effects of a negative demand shock on the STR market should be considered the opposite. Nevertheless, some apartment owners might not be interested in an instant switch between the markets, but they would incorporate the changed opportunity cost into the expected rent level. Thus, although part of the adjustment of the LTR market rents to the STR market demand shock would result from supply changes, the other part may be related to the change in relative rents. Yet, both types of reactions may be unevenly distributed in time and space.

It may be argued that to understand the behaviour of rents in the LTR market, one should consider the changes in the STR market. Additionally, to increase comprehension of the housing rental submarkets and knowledge of their interdependency, one should analyse separately the price- and supply-related phenomena. In this regard, the H9 hypothesis targeted the relationship between the submarkets' rental prices, expecting that the STR market rent level may be used to understand and predict LTR market rents and thus can be considered its cause.

H10: During the pandemic, the decrease in the supply of apartments in the STR market contributed to the decline in rents in the LTR market.

Although multiple studies have already analysed the undesirable impact of the growth of the STR market on the LTR market rents, little has been said about the relation in the opposite direction. However, one might have recently observed two phenomena that led to a contraction of the STR market. First, multiple governments (either national or local) have decided to regulate the market to mitigate its negative externalities, as summarised by von Briel & Dolnicar (2021). Seiler et al. (2024) found that introducing the legislative ban on the STR market in Irvine (California, the USA) caused LTR rents to decline by 3% in the two years following the introduction. Moreover, they found a larger decline in the market segments in which properties in the STR and LTR markets showed the biggest similarities and in the STR-dense regions. On the other hand, Chaves Fonseca (2024) indicated that besides the decrease in STR listings, the ban introduced in Santa Monica (California, the USA) did not significantly impact the LTR rent levels. In this regard, von Briel & Dolnicar (2021) argued that the long-term impact of the restrictions would be limited in each case, as the STR market participants in the tourist areas are resilient and determined enough to adjust to any regulations introduced.

Secondly, the pandemic reversed the upward trend of the STR market volume because of the frozen tourist movement. Based on the STR_T dataset described in Section I.6.3.3. and GUS (2023), the pre-pandemic share of the STR market in two central districts of Warsaw – Srodmiescie and Wola equalled at least[3] 1.5% of all inhabited apartments in these districts, regardless of whether they were owned or rented. In contrast, the share in the other districts (treated jointly) amounted to 0.1%. Then, after the pandemic outbreak, the shares declined in November 2020 to 0.48% and 0.04%, respectively. Based on Eurostat (2024c), the number of guest nights spent in the STR apartments declined between 2019 and 2020 by 54% in Warsaw, 41% in Poznan, 65% in Krakow, 67% in Paris and 74% in Barcelona. Guglielminetti et al. (2023) argued that the decline in demand and supply of STR apartments should be attributed to the restrictions imposed by governments rather than to the fear of spreading infection. Boto-García (2022) showed the heterogeneity of price responses to the changing market situation, dividing the STR hosts into professionals and non-professionals, highlighting that the former decreased rents to a greater extent. On the other hand, Demir et al. (2024) showed that the effects of the pandemic might have been unequal because of the preference switch towards entire homes, in which it was possible to maintain social distancing. In the same spirit, Filieri et al. (2023) pointed out that post-pandemic consumers prefer

[3] The phrase "at least" is used in this case since there was a non-negligible number of apartments rented on the STR market, which were not included in the STR_T database that represents only the whole apartments rented via the most popular STR platform – Airbnb.

accommodations in rural areas and of higher quality. Sainaghi & Chica-Olmo (2022) argued for the increased role of hosts' credibility. Finally, Gossen & Reck (2021) and Bresciani et al. (2021) highlighted the declining value of shared accommodation and hotel rooms compared to the whole apartments, which are most similar to those rented in the LTR market.

Still, little has been known about the strength of the impact of the STR market supply changes on the LTR market rents during the pandemic. Among the current studies, only Mozo Carollo et al. (2024) and Batalha et al. (2022) targeted this issue. The former, based on the example of the mid-sized city (San Sebastian in Spain), showed that the density of STR listings influences LTR rents. Still, the effect was remarkably stronger in the pre-pandemic period (from July 2019 to March 2020) than during the first phases of the pandemic (April to December 2020). Then, Batalha et al. (2022) estimated that in Lisbon, the relocation of STR properties to the LTR market contributed to the decrease of the LTR rents by 4.1% – around a third of the decline in LTR rents at that time. Nevertheless, none of the studies discussed the spatial differentiation of changes in LTR rents. This, in turn, may be linked to the spatially unequal changes in the volume of the STR apartments and spatially unequal shares of STRs in the total housing supply. Finally, it should be noted that the studies mentioned above were conducted after the study by Trojanek et al. (2021), which constitutes part of this dissertation and should be considered pioneering in testing the relation reflected in the H10 hypothesis.

I.6. Research methodology and data sources

I.6.1. Hedonic price theory

Apartments for rent are heterogeneous goods comprising distinct characteristics. Moreover, there are no two identical apartments. Yet, they are being traded (rented) on the same market[4], where the variation in rents may be attributed to the differences between the apartment's characteristics, either tangible or intangible. The primary goal of hedonic methods is to infer the information concerning the non-observable valuations of the factors that contribute to rent differences based on the market prices of individual transactions. The valuations are inseparably linked with consumers' choices in the market – with their willingness to pay for particular housing characteristics. Although the hedonic analysis may be regarded as a two-step process, for the needs of this dissertation, only its first part, which focuses on the implicit prices of housing features, will be discussed. The second part, related to analysing the demand functions, is more complex, both technically and concerning the data requirements (Taylor, 2003).

Although the concept of price differentiation, which is based on the inseparable characteristics of goods, was studied almost a hundred years ago by Waugh (1928), it was Rosen (1974) who established the conceptual basis of hedonic price theory, using the microeconomic theory of consumer demand by Lancaster (1966). The framework outlined the relations between consumer preferences, their willingness-to-pay for whole products, as well as their particular characteristics, and the process of reaching market equilibrium. In this section, the process of bidding and offering the products has been presented from the perspective of tenants and landlords who may be considered players in the market. In the following subsections, the considerations of tenants' and landlords' choices in the LTR market have been studied following the works by Taylor (2003) and Day (2001).

I.6.1.1. Tenants' side

Within the hedonic model's framework, tenants choose an optimal combination of two products. First, Z should be considered the differentiated product with overall $\underline{z}$ features. In this regard we will discuss apartments for LTR rent that comprise of multiple characteristics

[4] For the purposes of this theoretical discussion, it has been assumed that apartments are rented only on the LTR market, thus rejecting the possibility to choose, whether to rent an apartment on the STR or on the LTR market. Moreover, it has been assumed that there are no new apartments entering the market.

and the values of $z_1, z_2, z_3, \dots, z_n$ portray quantity of a particular characteristic in the analysed apartment. The values concern all the apartment features, both structural and locational. Then, x represents jointly all other, non-differentiated goods, which will be obtained by tenant for the money left from his budget. Assuming the LTR market is a perfectly competitive one, the monthly rent, which will be the subject of further considerations, is established in a joint interaction of multiple landlords and multiple tenants. This way, the price schedule of rents is formed and, because of a multitude of agents in the market, it is treated as given (i.e. exogenous) by each tenant. As one assumes the perfect information, the equilibrium rental function $P(\underline{z})$ is a representation of characteristics of the apartment and may be referred as hedonic price function. Then, tenants have no impact on the rent they pay, other than the choice of apartment characteristics they want to rent through the rental transaction. Yet, they have a perfect knowledge on their preferences over all goods and the preferences distinguish them from other tenants. Using this information, one may specify the utility function for tenant j (with unique characteristics s_j), which is unequal with functions of other tenants:

$$U^j(x, z_1, z_2, z_3, \dots, z_n; s_j) \tag{1}$$

In this case, y – the budgetary line of a tenant will take the non-linear form (counter to the usual consumer choice problem, where the budgetary line has been linear):

$$y^j = x + P(\underline{z}) \tag{2}$$

Since tenants strive to maximize their utility of consumption of goods x and renting apartments $\underline{z}$ within their budgetary lines, for each characteristic z_i the chosen consumption structure will satisfy the Equation 3. It refers to the marginal rate of substitution between each characteristic of apartment and the x goods:

$$\frac{\partial P}{\partial z_i} = \frac{\partial U / \partial z_i}{\partial U / \partial x} \tag{3}$$

Then, θ may be defined as a bid function of each tenant that refers to the optimal bid tenant will make for apartment Z, in order to maintain the U_0 utility level in the case of constant budget of y_0 and changing apartment characteristics. Next, we can use θ to substitute x in the utility function:

$$U^j(y_0 - \theta, \underline{z}; s_j) \equiv U_0$$

(4)

In this regard, the bid function will be represented as:

$$\theta^j = \theta(\underline{z};\ y_0, U_0^j, s^j)$$

(5)

Then, by maximising the utility function presented in Equation 4, one will find that the marginal bid for z_i will be equal to the marginal rate of substitution between z_i and x. As a result, in tenant's optimum, the marginal bid placed for z_i, to which we may refer as θ_{zi} (equalling $\frac{\partial \theta}{\partial z_i}$) should be equal to the marginal rent related with that characteristic – P_{zi} (equaling $\frac{\partial P(\bar{Z})}{\partial z_i}$). Finally, U_{z_i}/U_x will represent the slope of the bid function and will mirror the indifference curve for this goods.

I.6.1.2. Landlords' side

In the LTR market, tenants pay rent $P(\underline{z})$ to individual landlords for the flow of utility that they derive from the rented apartment characteristics within the period of time – in Poland it is most often one month. Landlords have to incur particular costs to rent an apartment in the LTR market – cost of the purchase of the property, cost of maintaining the soft quality of apartment, and cost of investments that aim to change the hard quality of apartments (when possible). For the needs of comparability, the costs are hereby understood as costs per period of time, thus representing e.g. monthly instalments, or discounted investment costs. Then, the cost function of supplying the apartment with $\underline{z}$ characteristics by k landlord, may be presented as:

$$c^k(\underline{z}; \hat{P}(\hat{z}), \tilde{z}, r^k)$$

(6)

where:

- $\hat{P}(\hat{z})$ refers to the fixed price paid for the property,
- $\tilde{z}$ describes the characteristics of the apartment that are provided costlessly to the landlord, appear without further investments and are beyond landlords' control, e.g. the floor area of the apartment, number of bedrooms, neighbourhood's pollution level,

- r defines any other parameters important to cost formation, unique for each landlord and reflecting, e.g. his ability to provide necessary repairs in the apartment on his own, instead of using costly external services offered on the market.

Yet, it may be assumed that it is possible for landlords to change the characteristics of properties through investments, thus the cost function may be generalised. Combining it with the notion of profit per period – π and $P(\underline{z})$ – a rental price schedule, which is exogenous to landlords (because of the multitude of other landlords and tenants in the market) and depends solely on the characteristics of the offered apartment, one achieves:

$$\pi(\underline{z};\tilde{z},r^k) = P(\underline{z}) - c(\underline{z};\tilde{z},r^k) \tag{7}$$

Then, one may introduce the Rosen's offer function, which is a landlord's counterpart to tenant's bid and refers to the amount of money a landlord will accept for the set of offered $\underline{z}$ characteristics, other factors being constant:

$$\varphi^k = \varphi(\underline{z};\tilde{z},r^k,\pi_0) = \pi_0 + c(\underline{z};\tilde{z},r^k) \tag{8}$$

Next, for a given price schedule in the market, landlords search for the combination of apartment characteristics, offering which will lead to achieving the highest possible profit. Similarly to the considerations of tenants' side, landlords would be willing to accept the marginal price for z_i equalling the marginal cost of supplying the characteristic.

I.6.1.3. Rental market equilibrium

The bid and offer prices, together with the equilibrium $P(\underline{z})$ rental price schedule may be discussed using Figure 3. It should be noted that $P(\underline{z})$ relies solely on the changes in the quantity of z_i apartment characteristic. In the housing market, hedonic price functions for particular housing characteristics most often increase at a decreasing rate, but may well take any other form. In the Figure 3, we may see the concave bid functions of two consumers – θ^1 and θ^2. Focusing on the first consumer with bid function θ^1, one may see that exactly the same quantity (Z_1^{1*}) may be achieved for three different rent levels represented by – $P(z_1^{1*})$, $P'(z_1^{1*})$ and $P''(z_1^{1*})$. Then, achieving the same quantity of characteristic and spending less money (thus having more money to spend on other goods), will result in a higher utility level. Therefore, higher utility will be achieved for the lower located bid functions. Knowing that the rental price schedule $P(\underline{z})$ is given, tenant will achieve the highest possible utility using the bid function that will be tangent to the hedonic price function. For tenant 1, it will be in Z_1^{1*} for $P(Z_1^{1*})$, while for tenant 2 in Z_1^{2*} for $P(Z_1^{2*})$.

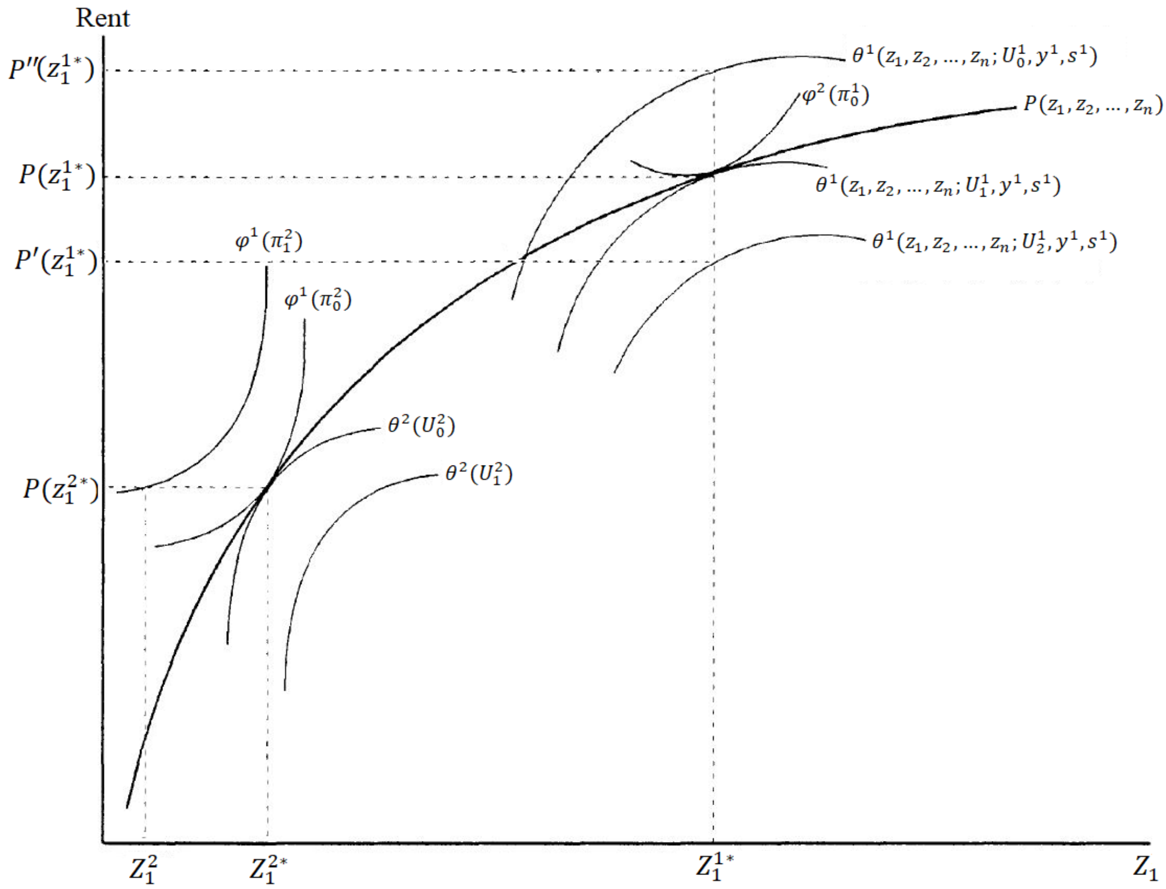


Figure 3. Bid and offer functions, as well as hedonic price function

Source: Taylor (2003), extended and adapted for the needs of this dissertation.

On the other hand, offer functions are convex, as presented on the example, where functions φ^1 and φ^2 represent two different landlords. For landlord 1 with bid function φ^1 the same value of rent $P(Z_1^{2*})$, may be achieved in both presented functions $\varphi^1(\pi_1^2)$ and $\varphi^1(\pi_0^2)$. The former function results in offering smaller quantity of Z_1^2 (counter to the latter function resulting in offering Z_1^{2*} for this rent level), thus it provides higher profit to the landlord. Although it is a scenario desired by landlords, they have to adjust to the given price schedule. Therefore, the landlord's optimum will establish in Z_1^{2*} for $P(Z_1^{2*})$.

Finally, one may argue that as landlords and tenants are rental price takers, the hedonic function should be treated as an envelope of all equilibrium interactions of tenants and landlords, who either bid or offer the differentiated product (Taylor, 2003). Then, there is a further question, which unfortunately requires separate theoretical analysis for each characteristic of apartments for rent – what is the functional form of rental price schedule for a given characteristic? Taylor (2003) presented the most often used functional forms and highlighted that the linear form should be used with the biggest caution, as rather non-linear functions should be expected for most housing attributes. In this case, the semi-log functional form (in which logarithms of rents depend on the levels of housing attributes) has either been used most often in the earlier hedonic studies of the housing market and has a most intuitive interpretation. Thus, the semi-log functional forms of hedonic models have been used further in this dissertation.

I.6.2. Methods

I.6.2.1. Hedonic methods

The Ordinary Least Squares (OLS) has been a baseline hedonic method used to empirically test most of the stated hypotheses. To construct HRI, the time-dummy approach (Hill, 2004) has been selected. The structure of the baseline model with a logarithm of a rent per month as a dependent variable (thus representing the rental price schedule, as discussed in Section I.6.1) may be specified as:

$$lnR_i = \beta_0 + \sum_{j=1}^{J} \beta_j C_{i,j} + \sum_{\mathrm{k}=2}^{K} \gamma_k D_{i,k} + u_i \tag{9}$$

where R_i is a rent for an i-th apartment, $C_{i,j}$ represents a value of j-th characteristic of an i-th apartment, β_j is a parameter reflecting the estimated marginal price of j-th characteristic, $D_{i,k}$ is a time dummy indicating whether an i-th apartment was rented (or listed for rent) in k-th period, γ_k is the estimated parameter reflecting change in rents in k-th period (compared to the base period) and u_i represents the model's error. Then, HRI_k value for each period (with period $k = 1$ as a base) may be calculated by exponentiation of the estimated γ_k coefficient.

However, OLS has certain disadvantages, among which heteroskedasticity of the error term and sensitivity to extreme values and outliers are often mentioned. In some cases, the issues might be dealt with by using the heteroskedasticity robust variance estimator (White, 1980) or by excluding highly influential observations (e.g. using Cook's distance (Cook, 1977)). Both mentioned problems of OLS may be mitigated also by the use of Quantile Regression approach (QR) (Koenker & Bassett, 1978). Additionally, the method allows modelling any conditional quantile of the dependent variable, which might represent price-related segments of the market. However, the method provides results based on optimising algorithms, which makes it numerically demanding, and it may be sometimes difficult to achieve model convergence for all quantiles. Finally, there may be a problem with estimating confidence intervals of QR parameters, especially while using small samples. Although solutions to this issue have been established (Tarr, 2012), they make modelling harder to interpret and test using conventional statistical tools than it is in the case of OLS. The QR model may be specified as:

$$ln\,R_i = X_{ik}\beta_{\theta k} + \varepsilon_{\theta i}, \text{ with } Q_\theta(ln\,R_i|X_{ik}) = X_{ik}\beta_{\theta k} \tag{10}$$

where: R_i is an apartment's rent, X_{ik} is a vector of independent variables (including time-dummies), θ refers to the estimated regression quantile, $\beta_{\theta k}$ is a vector of coefficients for the observations of the dependent variable's θth quantile, $\varepsilon_{\theta i}$ is an error term and $Q_\theta\ (\ln R_i|X_{ik})$ represents the θth quantile of a dependent variable $\ln R_i$ given X_{ik}.

Because of problems with availability of data concerning the LTR market, this dissertation relies primarily on listing data. Although they are rich in information on the characteristics of apartments, the locations of the observed apartments have often been specified imprecisely. This has often hindered the possibility to take account of spatial dependency between apartments in hedonic models and resulted in the use of non-spatial hedonic methods (OLS or QR) to test most of the hypotheses. Yet, the information on location of individual apartments have been still included in the models. It has taken the form of adding dummy variables indicating the district in which apartments were located or extending the model's structure with variables reflecting the distance of apartments to public amenities. Nevertheless, for testing some hypotheses a higher precision of information on geolocation of apartments has been required. In those cases, only the observations with a precisely specified location have been selected and the spatial hedonic models have been constructed. One of those, used and described extensively in Hebdzyński (2024b), is the Spatial Error Model (SEM). Its main assumption is that in addition to the parameters modelled in the OLS (Equation 9), the error term is also modelled, and:

$$u = \lambda Wu + e \text{ , while } e \sim N(0, \sigma^2 I_n) \text{ ,} \tag{11}$$

where λ is a parameter of autocorrelation, W is a matrix of spatial weights, Wu is a spatially lagged error of the model, and e is an independent error.

Finally, in Trojanek et al. (2021) the Multiscale Geographically Weighted Regression method (MGWR) (Fotheringham et al., 2017) has been used. The MGWR model allows for estimating both global parameters (that are equal for all locations) and local ones (that are different in each location). It may be specified as:

$$\ln R_i = \sum_{g=1}^{q} a_g\, x_{ig}(a) + \sum_{l=q+1}^{p} b_l(u_i, v_i) x_{il}(b) + \varepsilon_i\,, \qquad i = 1,2,\dots,n \tag{12}$$

where R_i is an apartment's rent, u_i and v_i are the latitude and longitude of apartment's geolocation, a_g is a global parameter, b_l is a local parameter, x_{ig} is an apartment's characteristic related to a global parameter, x_{il} is an apartment's characteristic related to a local parameter, ε_i is an error term.

I.6.2.2. Repeat sales method to calculate price indices

BMN – the repeat sales method introduced by Bailey et al. (1963) has been used to calculate a repeat sales rent index (RSI) based on apartments that have been transacted at least two times during the observation period. In the context of this dissertation, it has been used only for the analysis of the STR market rents, the motivation for which has been presented in Section I.5. The average daily rent for the apartment in a given month has been considered one transaction, but only if the apartment was rented for at least five days of the month. Then, matching the consecutive observations would allow us to construct the monthly RSI. Following Hill & Trojanek (2022), the repeat sales model may be specified as:

$$\ln(R_{im'}) - \ln(R_{im}) = \beta_{m'} - \beta_m + \ln U_{imm'} \tag{13}$$

where $R_{im}$is an i-th apartment's average daily rent in month m. Then, m' refers to the second month, in which the apartment was rented for at least five days ($m' > m$), parameters β reflect the estimated rent levels in consecutive periods and $U_{imm'}$ is an error term. Finally, the RSI is calculated by exponentiation of the β parameters.

I.6.2.3. Wordscores algorithm

In Hebdzyński (2023), the Wordscores algorithm by Laver et al. (2003) has been adapted and calibrated for extracting textual quality signals from housing listings. In short, the steps of the algorithm are as follows:

1. The training dataset should be prepared. Each observation should contain a textual description of the listed apartment and the assigned visual quality label. The quality assessment should be done using uniform methodology. For the needs of this dissertation and article Hebdzyński (2023), it has been done based on photos attached to listings, following the procedure suggested by the National Bank of Poland (2022) (it has been discussed in detail in Section I.6.3.1.). Three soft quality classes are distinguished: low, medium and high.
2. To reduce the noise of the analysis, numbers, special characters, and shortest words (≤ 4 characters) should be removed from each textual description in each listing.
3. All remaining words in all listings should be replaced with their base forms (lemmas), with the use of a morphological dictionary. For Polish listings, the dictionary by Miłkowski (2016) might be used.
4. The occurrences of each lemma in the descriptions of apartments of a given visual quality should be counted:

$$n_j = h_j + m_j + l_j, \tag{14}$$

   where n_j is the number of occurrences of the j-th lemma in all observations, h_j, m_j and l_j refer to the number of occurrences of the j-th lemma in the descriptions of apartments of high, medium and low visual quality. Lemmas representing the least frequently used words should be excluded – in Hebdzyński (2023) the threshold has been set at $n_j < 0{,}25\% * X$, where X refers to the number of all observations in the training dataset.
5. Each occurrence of a lemma in the description of a low visual quality apartment is treated as a negative signal of the lemma, while occurrence in the description of a high-visual-quality apartment – as a positive signal. Occurrences in the descriptions of medium-visual-quality apartments are considered neutral signals. Then, the quality signal of the lemma would be calculated as:

$$LEMMA\_SIGNAL_j = \frac{h_j * (+1) + m_j * 0 + l_j * (-1)}{n_j} \tag{15}$$

   A list of lemma-*LEMMA_SIGNAL* pairs may be referred to as a dictionary.
6. Having the already calculated dictionary, to extract textual quality signal from any LTR listing one should first pre-process the apartment description by conducting steps 2 and 3 of the

algorithm. Then, the *LEMMA_SIGNAL* scores from the dictionary should be assigned to each lemma in the listing. The lemmas not appearing in the dictionary should be excluded. The final, numerical representation of the listing's textual quality signal – the Wordscore value, may be calculated as an average of all *LEMMA_SIGNAL* scores of the remaining lemmas.

I.6.2.4. Granger causality test

Based on the assumptions made by Granger (1969), the macroeconomic variable y_1 may be considered a cause of y_2 variable if it is possible to forecast more accurately current values of y_2 using past values of y_1 than without using them, *ceteris paribus*. First, following Rubaszek (2012), one may define the vector autoregressive model in which:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} A_{p(11)} & A_{p(12)} \\ A_{p(21)} & A_{p(22)} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{16}$$

where $y_t = [y_{1t}\ y_{2t}]$ is a vector of two stationary time series, p is a model's lag, A_p refers to the unknown parameter describing the influence of lagged y_t on its current value (A_0 is an intercept), and ϵ_t is a white noise process with the expected value of 0 and Σ covariance matrix. Then, the tested hypothesis of the Granger causality test takes the form of:

$$\mathrm{H_0} : \Lambda_{1 \leq p \leq P} A_{p(12)} = 0 \tag{17}$$

In the case of rejecting H_0 one may assume that y_2 Granger-causes y_1. However, to test the simultaneous dependency, one should inspect values of Σ covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{18}$$

and test the following hypothesis:

$$\mathrm{H_0} : \Sigma_{12} = 0 \tag{19}$$

I.6.3. Data used

I.6.3.1. LTR listings

This dissertation utilised three independently collected datasets of listings of LTR apartments.

LTR_L_1 – Web-scraped LTR listings concerning the market of Poznan

The LTR listings used to study most of the stated hypotheses concerned the market of Poznan and were web-scraped quarterly (between the 10th and 20th day of March, June, September and December) from two leading Polish apartment listing websites – Otodom.pl and Gratka.pl, from June 2019 to December 2022. The dataset contains information on the listings' providers (whether they were landlords or housing brokers), the listed rents of individual apartments, and a wide range of apartment characteristics. The included housing features have been described in detail in Hebdzyński (2024a, 2024b). The data have been cleaned of duplicate observations – if the listing of the same apartment reappeared in adjacent periods (within the 6-month range), only the last observation has been retained. This way, the listed rent would be as close as possible to the final, transacted rent. The data gaps were filled by exploring the descriptions and photos of listings. It allowed obtaining information on multiple apartment characteristics, which are often not signalled through the structured listings' forms, among others on building maintenance, apartment furnishing, availability of a balcony or a private garden. Furthermore, based on expert market knowledge, the apartments with extreme rent or floor area were excluded, similarly to those deemed unrepresentative, e.g. multi-storey apartments and the ones with gardens bigger than 75 m^2. The incorrect listings, including highly improbable or contradictory information, were excluded. The minor mistakes that sellers made in the listing process were corrected.

Particular attention was paid to the issue of the soft quality of apartments. Direct textual quality signals were based on the sellers' declarations. Then, indirect visual signals of quality were obtained in the process of individual visual assessment of the quality of apartments solely based on photos attached to the listings. In the process, I followed the quality-assessment instructions by the National Bank of Poland (2022), which is an institution responsible for monitoring prices in the Polish housing market, hence gathering the micro-level data in the law-regulated process. The attributes of the apartment taken into account in the quality assessment were floors, fixtures, doors, walls, and kitchen & bathroom equipment.

A high-quality label was assigned to functionally arranged apartments, finished with good materials and revealing low exploitation. Those requiring renovation were considered low quality, while the rest – medium quality. Finally, full textual descriptions of apartments were processed using the Wordscores algorithm to obtain the proxy of indirect textual quality signals.

The exact location of the listed apartments was rarely provided. Thus, the information that it was, e.g. located “opposite to the XYZ restaurant” or “on a corner of X and Y streets”, was used. In the absence of more precise information, the geolocation of the centre of the street where the apartment was located was used. The observations were categorised into classes according to the precision of information: up to 500, 250, 125 or 0 meters (the last value was assigned to the exact locations). The ones with a lower precision were excluded from the dataset.

Given all the above features, the dataset should be considered exceptionally detailed compared to the listing data used in most hedonic housing market studies. Nevertheless, the constraints of a single-city study area, relatively low data collection frequency, and short observation period before the pandemic created the need to use two additional datasets. They have a similar structure but differ in terms of the precision of the information and the scope of available characteristics of apartments for LTR rent.

LTR_L_2 – LTR listings concerning the market of Poznan obtained from the listing platform

This dataset has been obtained directly from the owner of the Otodom.pl listing platform – OLX Group. Its time range does not entirely overlap with the period of the pandemic in Poland, as it covers the period of November 2020 - May 2023. However, this dissertation has used this dataset to test the technical hypotheses, which are not directly related to the pandemic. The included data are of monthly intervals, but they have been converted to quarterly to obtain results comparable with other datasets. Yet, the higher frequency of data has allowed pairing the higher number of listings from this dataset with transactional data (described in Section I.6.3.2.). However, as the listings’ textual descriptions are not available in this dataset, it has not been possible to benefit from exploring information included there. It applies to the lack of knowledge on the soft quality of apartments and having less precise information about their location and other characteristics. Still, the dataset has been cleaned and extended using expert knowledge, where possible, similarly to the *LTR_L_1* dataset.

LTR_L_3 – Web-scraped LTR listings concerning the market of Warsaw

The data was web-scraped monthly from the second biggest listing website in Poland – Gratka.pl, between August 2015 and December 2020 by the Supervisor of this dissertation – prof. Radosław Trojanek. Besides the housing characteristics conventionally used in the hedonic studies of the housing market, it also includes information on the soft quality of apartments (obtained from direct and indirect textual signals). Then, information about the geolocation of the listed apartments was provided with different precision. For the needs of the study by Trojanek et al. (2021), for the time range of January 2017 - December 2020, the information on the possibly most precise geolocation has been used. It required constructing a database containing information about street names and house numbers and, in the case of their unavailability, the geolocation of the centre of the street. It has been used to calculate distances from the apartment to the city centre and major public amenities. To allow for the comparison of the LTR market with the STR market in Hebdzyński & Trojanek (2024), for the time range of August 2015 - December 2020, the geo-locational information has been used on the less-detailed district level. Nevertheless, based on the given geo-location of the apartments and expert knowledge, the information on the building types (their age and technology) was supplemented to improve the dataset's quality. Finally, the dataset was cleaned similarly to the datasets *LTR_L_1* and *LTR_L_2*.

I.6.3.2. LTR transactions paired with listings (*LTR_T*)

The transactional database includes 197 observations of LTR rental transactions concluded in Poznan between Q1 2021 and Q2 2023, sourced from BaRN database (National Bank of Poland, 2022). Only those observations were selected, for which it was possible to pair the transaction with the corresponding listing in the *LTR_L_2* database. In particular, the observations of apartments located on the same street and the same floor of the building, with the same floor area and the number of rooms, were considered observations of the same apartment. It was assumed that the listed rent is always equal to or higher than the transacted rent, as in Horowitz (1992). Finally, the information on apartment characteristics from both sources was merged. Thus, apart from the complete set of apartment characteristics, each entry in the database included two prices – listed and transacted, and two dates – a quarter of the listing and a quarter of the transaction.

I.6.3.3. STR transactions (*STR_T*)

The dataset concerning the STR market was obtained from AirDNA company and extended using expert knowledge of the market. It is a complete dataset on STR rental transactions of apartments in Warsaw, concluded via the Airbnb internet platform. The dataset includes apartments available for rent for at least one day between August 2015 and December 2020. The dataset has two dimensions. First, the apartment's characteristics include, among others, apartment ID, location (at the district level and in the form of proxy coordinates), type, size, and capacity. Moreover, the performance of each apartment in each month has been provided, including occupancy rate, average daily rent, number of reservation days, and number of days of availability. The information on the quality of the offered apartments has been included in the dataset in the form of variables representing ratings of cleanliness, location and host, and the overall rating. Yet, that information is provided only for the last observed month of the apartment performance. Similarly, there is no information on the historical changes in the availability of housing amenities, cancellation policy, etc. As a result, none of this information can be used for historical analysis of the STR market prices.

Finally, some necessary data filtering has been done to target the precise segment of the STR market to calculate its rent index. First, only the apartments intended solely for short-term stays have been selected. Then, the observations for which it was impossible to assign them to the existing building with a precision of up to 200 metres have been excluded. Finally, the observations with contradictory information and the ones without any guest review have been considered unrepresentative and hence removed from the database.

I.6.4. Analytical steps

Table 1 presents the simplified analytical procedures used to test the research hypotheses, together with the data used and their scope. The complete descriptions of analytical steps have been included in the articles and working paper that constitute the dissertation.

Table 1. Simplified overview of analytical procedures and the scope of data used

Hypothesis	Analytical procedure	Data used and their scope

H1
hypothesis targeted in article **A1**
	1. Comparison of coefficients of variables achieved in OLS models built based on:
	• Transactional data paired with listings (LTR_T), using:
	  ○ Transacted rent as a dependent variable and the date of the transaction to construct HRI
	  ○ Listed rent as a dependent variable and the date of listing to construct HRI
	• All listing data (LTR_L_2) – listed rent as a dependent variable and the date of listing to construct HRI
	
	2. Calculating average percentage differences of coefficients achieved in:
	• QR models built based on all listing data for selected conditional quantiles of the distribution of LTR rents
	• An OLS model built on transactional data
		$\text{Dataset}_{A1.1}$:
		• Source: LTR_T
		• City: Poznan
		• Time: Q1 2021 – Q2 2023
		• N_{obs}: 197
		
		$\text{Dataset}_{A1.2}$:
		• Source: LTR_L_2
		• City: Poznan
		• Time: Q4 2020 – Q2 2023
		• N_{obs}: 9,186

H2 hypothesis targeted in articles **A1**, **A3** and **A4**	Article 1: Comparison of HRI built based on median QR models: • Using information on apartment geolocation in different forms and at different levels of precision (A1) • Using the full or reduced set of explanatory variables (A1) Article 3: Comparison of HRI built based on OLS and median QR models: • Using different approaches to take account of soft quality of apartments (A3) Article 4: Comparison of HRI built based on SEM models: • Disregarding the shock-induced variability of rent-setting factors (A4) • Allowing for the shock-induced variability of rent-setting factors (A4)	Dataset_{A1}: • Source: LTR_L_2 • City: Poznan • Time: 11/2020 – 05/2023 • N_{obs}: 9,186 Dataset_{A3}: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2022 • N_{obs}: 8,248 Dataset_{A4}: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2021 • N_{obs}: 4,342

Hypothesis	Analytical procedure	Data used and their scope
H3 hypothesis targeted in article **A2**	Comparison of the structure of soft quality signals sent by different types of sellers via LTR listings. Direct textual signals have been compared with indirect visual signals. Additionally, attention has been paid to the listings for which no direct quality signal was sent.	Dataset$_{A2}$: • Source: LTR_L_1 • City: Poznan • Time: 09/2020 – 06/2021 • N_{obs}: 1,441
H4 hypothesis targeted in article **A2**	Verification of consistency of soft quality signals sellers send via listings of apartments for LTR rent. The indirect textual signals that have been obtained with the use of the Wordscores algorithm have been compared with indirect visual signals. It has been done to decide whether the signals show similarity and may be used as substitute or complementary sources of information on apartment quality.	Dataset$_{A2}$: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 09/2020 • N_{obs}: 4,044
H5 hypothesis targeted in article **A3**	Comparison of statistical properties of OLS and QR models that include explanatory variables reflecting different types of soft quality signals sent by sellers via listings of apartments for LTR rent. Assessment of the differences between HRIs obtained based on those models. The following types of soft quality signals have been analysed: • No signal • Direct textual signals • Indirect visual signals • Indirect textual signals	Dataset$_{A3}$: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2022 • N_{obs}: 8,248
H6 hypothesis targeted in article **A4**	1. Construction of an SEM model to capture the major rent-setting factors. The model has included the variables reflecting the location of the apartment in the revitalised/non-revitalised tenement buildings, which have not been analysed in previous studies of the rental market. 2. Construction of an SEM model including variables representing the interaction of a dummy variable reflecting whether the apartment was listed in the pandemic period and variables reflecting specific housing characteristics. The group of analysed characteristics consisted of the availability of an additional room for remote working or studying, the availability of a small private garden, the availability of a balcony, the proximity to green areas, and the proximity to university buildings.	Dataset$_{A4}$: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2021 • N_{obs}: 4,342

Hypothesis	Analytical procedure	Data used and their scope
H7 hypothesis targeted in studies **A4**, **A5** and **WP1**	Article 4: Construction of an SEM model to obtain information on quarterly HRI changes in Poznan during the pandemic. The value of HRI for the last pre-lockdown period (March 2020) has been compared with the lowest value for the pandemic period. Article 5: Construction of QR models (for 25th, 50th and 75th percentiles) to obtain information on monthly HRI changes during the first phase of the pandemic in Warsaw (from March to December 2020). Construction of an MGWR model to get information on spatial variation of rent changes. The pre-lockdown (June 2019 – March 2020) and post-lockdown (April 2020 – December 2020) rent levels have been compared. Working paper 1: Construction of seasonally adjusted monthly rent indices of the LTR market (using OLS) and STR market (using OLS and BMN) of Warsaw. The changes in rent levels amidst the pandemic have been compared, taking into account the following segmentation of sub-markets by location: • Whole Warsaw • Only central districts (Srodmiescie and Wola) • Only non-central districts (the remaining districts)	Dataset$_{A4}$: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2021 • N_{obs}: 4,342 Dataset$_{A5}$: • Source: LTR_L_3 • City: Warsaw • Time: 01/2017 – 12/2020 • N_{obs}: 41,264 Dataset$_{WP1}$: • Source: STR_T • City: Warsaw • Time: 08/2015 – 12/2020 • N_{obs}: 99,110
H8 hypothesis targeted in article **A3**	Comparison of HRIs constructed based on median QR models for different soft quality segments of the LTR market. The segmentation has been conducted according to arguably the most reliable signals sent via apartment listings – indirect visual signals. To enhance the precision of the models, the explanatory variable reflecting indirect textual quality signals has also been included. The rent levels and their dynamics have been analysed.	Dataset$_{A3}$: • Source: LTR_L_1 • City: Poznan • Time: 06/2019 – 12/2022 • N_{obs}: 8,248

Hypothesis	Analytical procedure	Data used and their scope
H9 hypothesis targeted in working paper **WP1**	Granger causality test of monthly time series of rents – HRI of the LTR market and RSI of the STR market. The causality has been analysed in three dimensions: • Influence of rents in the STR market on rents in the LTR market • Influence of rents in the LTR market on rents in the STR market • Simultaneous dependency of rents in the STR and LTR markets	$Dataset_{WP1.1}$: • Source: LTR_L_3 • City: Warsaw • Time: 08/2015 – 12/2020 • N_{obs}: 117,124 $Dataset_{WP1.2}$: • Source: STR_T • City: Warsaw • Time: 08/2015 – 12/2020 • N_{obs}: 99,110
H10 hypothesis targeted in article **A5**	Construction of QR models of the LTR market of Warsaw (for the 25th, 50th and 75th percentile of the conditional distribution of rents). Models for three periods have been constructed: • Pre-lockdown (January 2017 – March 2020) • Pandemic (April 2020 – December 2020) • Whole analytical period (January 2017 – December 2020) Each model has included an explanatory variable reflecting the volume of supply of STR apartments in a given month in the district in which the listed apartment was located. The obtained coefficient and its statistical significance have been analysed to decide on the hypothesised relation and its statistical significance.	$Dataset_{A5.1}$: • Source: LTR_L_3 • City: Warsaw • Time: 01/2017 – 12/2020 • N_{obs}: 41,264 $Dataset_{A5.2}$: • Source: STR_T • City: Warsaw • Time: 01/2017 – 12/2020 • N_{obs}: 56,283

I.7. Research results and discussion

This section summarises the obtained results and confronts them with the stated hypotheses. The detailed descriptions of the findings have been included in the articles and working paper that constitute this dissertation.

H1: The LTR market listing data may act as a proxy of transactional data in terms of their usefulness for constructing hedonic models.

The analysis of the hedonic models' results used to study the hypothesis has been divided into two parts. The first part compares the results obtained using listings and transactional data and concerns the models' coefficients that reflect the rent-setting factors. The second part compares the listings-based and transactions-based HRIs.

First, in Hebdzyński (2024c), it has been shown that the differences between the coefficients of hedonic models based on transactions and listings are small (the average percentage difference has equalled 7.4%), provided that the calculations are made on the same sample of apartments but in various stages of the rental process. One of the reasons for the little discrepancy is that in the studied period the analysed LTR market in Poznan proved to be of high liquidity. The final, transacted rents were rarely negotiated, being, on average, only 2% lower than the listed ones. Although still moderate, the differences between the transactions-based and all-listings-based models' coefficients have been larger. For all independent variables, signs have agreed. As for the comparison of the magnitude of their impact on the dependent variable – for all explanatory variables, except one, the differences have not exceeded 20%. For the locational variable it has reached almost 50%. However, it should be noted that in the utilised dataset of listings (LTR_L_2), the precision of information on apartments' geolocation has been relatively low, which may have negatively affected the accuracy of the results. Additionally, a small number of observations available in the transactional dataset made it impossible to discuss more than one locational variable.

It has been ensured that the higher discrepancy has not been rooted in the fact that some listed apartments were transacted faster than others because, in the dataset, only the last listing of each apartment has been included. It has also been shown that the problem has not originated from the difference between the height of the listed and transacted rent. Then, it has been argued that the primary source of difference may be in the quality structure of the market reflected in all listings. At first, it was suspected that the quality differences may result from the fact that listing platforms are rarely free of charge, and they are dominated by listings provided by real estate professionals, who take a commission for their work. As a result, the online listing platforms may overrepresent a higher market segment in relation to the whole market's quality structure. However, based on the results of QR models, it has been shown that the highest compliance of the transactions-based and listings-based models coefficients has been reached for the 55th-75th percentile of the conditional distribution of listed rents. This leads to the conclusion that the analysed transactional data represent an even higher market segment than the listing data. This may result from the process of gathering data on the rental market in Poland, which leaves no responsibility on non-professional landlords to report precise data on rental transactions.

In the second part of the analysis, it has been shown that as long as the HRIs have been obtained from models calculated based on observations of the same apartments but using their listed or transacted rents, the indicated dynamics of the HRIs have been almost identical. However, the time range of HRIs has been short, hence the possibility to provide a technical comparison of indices' co-movement has been limited. Yet, the dynamics of indices have been studied using relatively simple techniques. In eight out of nine quarters analysed, the signs of the dynamics indicated by the HRIs have been the same, and the correlation of the indices' dynamics has been regarded as strong (0.83).

However, the transactions-based HRI has revealed two short-term peaks, that the all-listings-based HRI has not detected. This may be rooted in the opposite nature of listings and transactions, especially in the short run. For instance, amid a negative demand shock, if the demand for low-quality apartments rose, there would be an increased share of low-quality apartments in the transactional data and a decreased share of observations of low-quality apartments in the periodically collected listing data. Then, the transactional models would be better suited to the more turbulent, lower-quality segment of the market. At the same time, the drop in the number of available low-quality apartment listings would result in a worse fit of the listings-based hedonic model to this market segment. As a result, the short-run changes in the market would be reflected by the listings-based HRIs only if we prepared separate models for quality segments; otherwise, the listings-based HRIs may be expected to flatten the transacted market dynamics.

Next, the QR models' analysis indicated that the transactional HRI has been most similar to the HRIs obtained in the all-listings-based models, representing the 60th-80th percentile of the conditional distribution of rents. Thus, also in this context, the transactional data proved to represent (on average) a higher market segment than listing data. This has been consistent with the results obtained for coefficients of apartments' characteristics.

It should be concluded that the listings-based rent-setting factors have proven similar to the transactions-based ones for most of the studied housing characteristics. The higher discrepancy of the coefficient achieved for locational variable may be, at least to some extent, attributed to the imprecision of the analysed listing data. Concerning the HRIs, it may be assumed that in the medium term, the listings- and transactions-based HRIs point at the same dynamics of LTR rents, yet in the short run, they may deviate from each other. The main reason for this situation may be seen in the quality differences between the datasets. Thus, it has been indicated that to study short-term changes in the LTR market rents, one should construct separate indices for the quality-related market segments or construct the quality-weighted HRIs.

The first part of the analysis extends the studies by Shimizu et al. (2016), and Kolbe et al. (2021), who discussed micro-level data reflecting different stages of the sales process and the implications of their use for hedonic modelling. Although it is based on the limited number of the LTR market transactions, it introduces the topic to a new segment of the residential real estate market, which is much more data-restricted, not only in Poland but also worldwide. Then, the second part of the analysis extends the results obtained for the LTR market by Micallef (2022) on the high compatibility of HRIs obtained with the use of listings and transactional data and adds validity to the listings-based HRIs for Polish biggest cities presented by Trojanek and Gluszak (2022) and Gluszak & Trojanek (2024). The study shows that under certain conditions, listings may be used as substitutes for transactional data, especially when the transactional data also have multiple weaknesses. Finally, although there is no evidence to reject the H1 hypothesis, one should be aware of the indicated differences in the quality structure of the datasets of listings and transactions. In this case, the issue of housing quality requires a particular attention. Nevertheless, the obtained results contribute to the discussion that it is possible to use the listings-based indices as a proxy of transactional indices (Ahlfeldt et al., 2023; Ardila et al., 2021; Lyons, 2019; Shimizu et al., 2022; Wang et al., 2020; Anenberg & Laufer, 2017), counter to the studies that treat them only as a source of supplementary information (Knight et al., 1994, 1998; Kolbe et al. 2021).

H2: Hedonic rent indices of the LTR market are robust to minor changes in the selection of the explanatory variables in hedonic models.

First, in Hebdzyński (2024c), the HRIs obtained based on hedonic models that differed in explanatory variables have been confronted. It has been shown that only the HRI built based on the model not including information on apartments' floor area, has noticeably diverged from the one built based on the model that was considered the best in the study.

Then, Hebdzyński (2024c) checked the robustness of HRIs to accounting for apartment geolocation using distinct types of information. Although the model including information at the district level has shown to be superior to the model including only one locational explanatory variable (reflecting distance to the city centre), the HRIs constructed based on both models have behaved equally.

In Hebdzyński (2024a), the effect of the implementation of soft quality considerations on the course of HRIs has been studied using OLS and median QR models. The model that utilised indirect textual signals of apartments' soft quality has been considered statistically superior and constituted a base for comparison with other models. In the model that has not taken account of the soft quality of apartments, the difference has reached the maximum level of 5.1% within the analysed time range of 3.5 years. On the other hand, the model utilising direct quality signals has also shown a difference, albeit smaller. Nevertheless (when confronted with the HRI obtained in the best model), the correlation of HRIs' quarter-on-quarter (q/q) dynamics has exceeded the value of 0.9. Finally, the models that included information on the soft quality of apartments obtained from indirect visual signals of quality were nearly identical (dynamics' correlation of 0.95).

In the last part of the analysis, in Hebdzyński (2024b), the sensitivity of the indicated dynamics of HRIs to the changes in the rent-setting factors during the pandemic has been examined. In the comparison of two HRIs obtained from SEM models that either disregarded the pandemic-induced variability of rent-setting factors or allowed for it, the correlation of q/q dynamics equalled over 0.995. Hence, the sensitivity should be considered extremely low. A similar conclusion may be drawn from the visual analysis of HRI levels. They have been almost the same, indicating a minimal overestimation of rents portrayed by the HRI obtained from the model disregarding the pandemic-induced changes, which has been considered inferior in terms of statistical properties.

Altogether, the results confirm that the conclusions from Diewert and Shimizu (2022) also hold in the case of the LTR market and that the HRIs are not robust only to major compositional changes in hedonic models. There are specific minimum requirements of hedonic models, after which the sensitivity of HPIs to the selection of additional variables in the model becomes low. In this manner, neither the way of accounting for geolocation nor the decision on which type of quality signal would be included in the hedonic model as an explanatory variable impact noticeably the obtained dynamics of HRI. Only the HRI built based on the model that disregarded the issue of soft quality has been shown to deviate from the statistically superior model, albeit the difference has not been substantial. Concerning the minor sensitivity of HRIs to accounting for changes in the rent-setting factors induced by the pandemic, the results are in line with the study by Hill & Trojanek (2022). They noticed only slight differences between the HPIs obtained using the standard time-dummy method and the rolling time-dummy method (which allows for parameter changes in time).

The results agree with the H2 hypothesis. Yet, it should be noted that constructing more complete hedonic models and using the variables at higher levels of precision, especially those concerning soft quality, is advised to achieve estimates of higher accuracy. However, in the studied cases, relatively short periods were analysed. Thus, the above conclusions should not be generalised for the long-run analysis, in which the beneficial properties of complete, well-quality-adjusted models would be supposedly more visible.

H3: Agents and landlords who list apartments for rent on the LTR market tend to overstate their declarations regarding the quality of the apartments.

In Hebdzyński (2023), the declarations of sellers concerning the soft quality of the listed apartments, which take the form of direct textual signals, have been confronted with the possibly most reliable soft quality signals sent via listings – indirect visual signals. It has been shown that the textually signalled “high quality” has rarely corresponded with the visual quality. The same has applied to the textual signals of “good quality” that have corresponded mostly with medium or low visual quality. This means that either sellers were consistent in overstating the quality of individual apartments for LTR rent, or the quality assessment scale they adopted was much more optimistic than it was specified in the guidelines by the National Bank of Poland (2022). However, the latter is unlikely, knowing that most of the listings were posted by real estate professionals. Therefore, there is no reason to reject the H3 hypothesis.

Finally, it should be added that the Polish listing platforms offer no possibility to signal directly, textually low quality, so low-quality apartments are mostly being left unlabelled or mislabelled as “good quality” ones. Thus, the quality structure of listings, for which the direct textual quality signal was sent, has been compared with the quality structure of observations with quality not signalled textually. The noticed information gap as for the direct textual quality signals has been substantial and has covered more than half of all observations. However, the quality structure of the observations with signalled and non-signalled quality has been similar.

The study results are novel in documenting the size and specificity of information gap as for certain types of quality signals in listings and inspecting the quality-related information disclosure strategies of sellers. Based on that, it contributes to the subject literature, particularly to the studies by Benefield et al. (2011), Hotz & Xiao (2017) and Bian et al. (2021).

H4: The Wordscores algorithm may be used to extract quality signals from textual descriptions included in the LTR market listings.

In Hebdzyński (2023), the steps of the Wordscores algorithm have been presented, and the most appropriate variant of the method to extract indirect textual quality signals from LTR apartment listings has been selected. Then, the signals have been compared with the possibly most reliable ones – indirect visual quality signals. It has been shown that the consistency of signals has amounted to 71% and has increased as the textual quality signals became clearer, up to 83%. The share of observations for which the contradictory information has been noticed (e.g. a low-visual-quality signal and a high-textual-quality signal or *vice versa*) has been negligible. The consistency has been particularly high for medium- and high-visual-quality apartments and limited for low-visual-quality apartments. It may be attributed to the fact, that the low quality of apartments for rent often manifests itself in a high degree of wear and tear or in their obsolete design, which may be harder to reflect in texts, but would be evident while looking at the photos. Moreover, sellers might be particularly unwilling to advertise even the lowest-quality properties using negative wording.

Despite the relatively small number of observations of listings posted by landlords, compared with the number of listings posted by real estate brokers, the analysis of the consistency of textual signals sent by different agents has revealed interesting patterns. The consistency has proven to be slightly higher for landlords’ listings – 72% than for the listings posted by brokers – 70%. Though small, the difference may be rooted in the fact that landlords are more interested in finding a tenant who would generate stable financial flows, take care of the apartment, and stay there for longer. Thus, private owners of apartments would have more incentives to represent the reality as it is and to send consistent quality signals. At the same time, brokers are willing to conclude the transaction quickly, as their compensation relies most often solely on the fact of concluding a transaction, not on the quality of tenants or the length of their residence in the apartment. For this reason, they may use advertising tricks, e.g. describing the apartment as higher quality than it is, which results in the lower consistency of textual and visual signals.

Knowing the high compliance of indirect textual and visual quality signals, one may conclude that they may be considered substitute sources of information for studying the rental market phenomena. Moreover, the Wordscores algorithm may help to bridge the information gaps detected in the case of direct signals of quality. It may be used as a decent proxy of the listed apartments' soft quality in the LTR market studies. The method has proven to be accurate yet straightforward, compared with the already developed and empirically tested approaches to textual and visual analysis (Nowak & Smith, 2017; Liu et al., 2020; Nowak et al., 2021; Shen & Ross, 2021; Seo et al., 2020; Poursaeed et al., 2018). It assigns the quality scores using exactly the same rules for all studied observations, thus leading to the higher comparability of the generated quality assessments. Additionally, it measures the intensity of the soft quality of each listing, which may be beneficial for hedonic modelling. All the above findings are in line with the H4 hypothesis.

H5: The underspecification of hedonic models of the LTR market in terms of housing quality leads to obtaining inferior statistical properties of the models and imprecise estimates.

In Hebdzyński (2024a), it has been shown that hedonic models of the LTR market that take account of soft quality are better fitted than those without quality-related variables. The improvement in the goodness-of-fit has been slight for the model utilising direct textual quality signals, while the models using indirect quality signals as explanatory variables have been found to be superior. Yet, it has been shown that the lower coefficient of determination of the model that has included sellers' direct textual signals of soft quality should be associated rather with the fact that the quality was declared only for a minority of observations (as discussed for the H3 hypothesis) than with a low information load of these declarations.

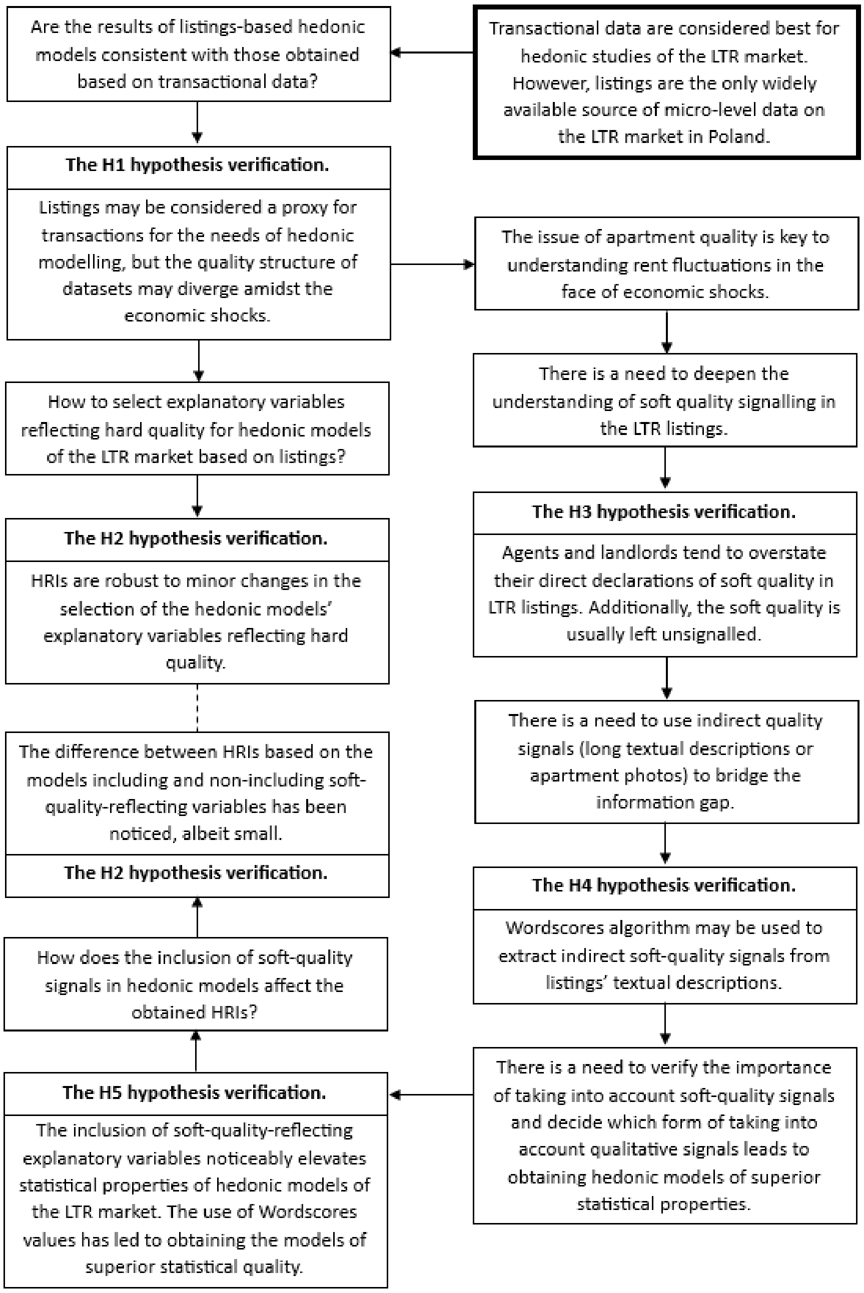


Figure 4. Logical structure of the first five hypotheses

Source: own elaboration

However, it has been noticed that instead of choosing explanatory variables reflecting apartments' soft quality, the selection of observations may be more important. It can be inferred that one needs to be careful when excluding observations for which no quality was declared from hedonic modelling. Although it has been earlier shown that the quality structure of observations with directly signalled and non-signalled soft quality may be regarded as similar, the fluctuations of rent levels indicated based on both subsamples have proven to be different. Thus, the decision whether to send a quality signal may be not random for housing sellers. In this case, resigning from implementing the quality considerations may be better than excluding the observations with non-signalled quality, as this approach may introduce bias to the results.

Finally, it may be concluded that the underspecification of hedonic models in terms of soft quality leads to obtaining inferior statistical properties of the models. Concerning precision of estimates – the models including soft quality considerations have proven to be statistically superior, but the difference in the indicated course of HRI has been low, albeit noticeable (as discussed for the H2 hypothesis). Thus, as hypothesised, it may be regarded that the quality considerations increase the precision of estimates. In this regard, the dissertation extends the research conducted on the housing sales market – Luchtenberg et al. (2019) by providing further evidence on the similarity of specific quality signals and Seo et al. (2020) by showing how they translate to the results of analyses of the LTR market.

Figure 4 presents logical structure along with the results of the first five hypotheses, the purpose of which has been to increase comprehension of listing data and hedonic models based on them.

H6: During the pandemic, the rent-setting factors in the LTR market changed.

In Hebdzyński (2024b), first, the rent-setting factors in the LTR market in Poznan have been established for the pre-pandemic and pandemic period. Based on the SEM model, the apartment area, together with its number of rooms and the soft quality of the interior, should be considered the major structural rent-setting factors. Then, the availability of a balcony, a designated parking space, and apartment furnishings have been shown to increase rents. On the other hand, the presence of a private garden has been considered insignificant in the rent formation process. Particular attention has been paid to building types. The apartments located in revitalised tenements have proven to be 4.2% more expensive than the ones in blocks of flats, *ceteris paribus*, while the marginal price of the location in a high-quality apartment building has been lower and estimated at 3.8%. Thus, the location in historical buildings that have undergone revitalisation may be considered to provide a rental price premium. At the same time, the apartments in renovated but non-revitalised tenements also have shown to be more expensive than in blocks – by around 3.8%, while those in non-renovated tenements – by 1.4%. As a result, the type of the building, the renovation performed, and its scale have been indicated as crucial factors in determining rents at the level of individual apartments.

Moreover, in line with the expectations, all the analysed locational variables have been shown to impact rents negatively – the larger the distance from the city centre, university buildings or green areas, *ceteris paribus*, the cheaper the apartments in Poznan. A similar direction of influence has been achieved for the proximity to public transport (tram), but, similarly to Krakow (Tomal, 2020), it has shown no significance. It may indicate that the relationship between public transport and rents is more complex and requires more data with greater precision to study multiple modes of transport and their both positive and negative externalities.

Concerning the changes in rent-setting factors during the pandemic, as hypothesised, the availability of a balcony or terrace increased rents by 2.8%, which was amplified during the pandemic by an additional 1.8 p.p. With regard to the availability of a small private garden, its impact on rents did not prove to be significant for the pre-pandemic period. However, the weak significance of change during the pandemic has been found. Thus, the overall impact of this variable on rents during the pandemic may be assumed to lie within the range of 1.5-5.4%. It confirms the findings of Marona & Tomal (2020, 2023) and Guglielminetti et al. (2021) and shows that home-related leisure increased its value for tenants amidst the health crisis and lockdowns. As for the valuation of an additional room for remote working or studying, the hypothesised change discussed by Nanda et al. (2021) has not been proven. It may mean that the increased need for an extra room for work-related purposes could have been balanced by the need for having more spacious rooms for better relaxation and general well-being, as supposed by Guglielminetti et al. (2021) and Mouratidis (2021).

Regarding locational factors, the impact of proximity to urban green areas on rents has been valued at -0.23 p.p. for every additional 100 m (which has proven to be insignificant), while the additional discount during the pandemic has been measured at an additional -0.39 p.p. (which has showed statistical significance). The noticed change aligns with the survey study by Noszczyk et al. (2022) and confirms the unique role assigned to urban green areas during health-related crises and lockdowns. Moving to the analysis of the distance from university buildings, in the model conducted on the full-time range, an additional 1 km, *ceteris paribus*, has proven to significantly decrease rents by 0.9%. However, the impact was not significantly different from zero during the pandemic period. This implies that factors other than proximity to university buildings were valued at that time. It confirms the findings by Tomal & Helbich (2022) for Krakow and shows that the students' transition to remote or hybrid education was quickly reflected also in changes in the LTR market valuation of housing characteristics.

Finally, it should be concluded that most of the suspected changes in rent-setting factors during the pandemic have been evidenced, which complies with the H6 hypothesis.

H7: During the pandemic, rent levels in the STR and LTR markets declined, but the reactions varied spatially.

Based on the SEM model presented in Hebdzyński (2024b), it may be inferred that the decline in LTR rents in Poznan amounted to 6.7% in March 2021, compared to the pre-pandemic level. The estimated value agrees with the expectations resulting from the theoretical analysis of the shock at the macro level (DiPasquale & Wheaton, 1992) and with the earlier discussed empirical evidence (Kuk et al., 2021; Tomal & Marona, 2021; Trojanek & Gluszak, 2022). Similarly, concerning the market of Warsaw, Trojanek et al. (2021) have shown that during the first nine months of the pandemic, LTR rents declined by 7-9%. The largest decrease has been noticed for the 25th percentile – for the cheapest market segment. Moreover, using the MGWR model, it has been shown that the drop in LTR rents varied spatially. The largest changes occurred in areas with the greatest nominal outflow of apartments from the STR market, i.e. in the central districts. The scale of decrease has been estimated to lie within -8.7% and -12.4%. At the same time, in some peripheral parts of the city, the negative change was lower than 5%.

Yet, in Hebdzyński & Trojanek (2024) it has been shown that during the studied period, i.e. August 2015 - December 2020, the dynamics of the STR market volume[5] were almost identical in the central districts of Warsaw that are often targeted by tourists (Srodmiescie & Wola) and in the non-central ones (the rest – 16 districts). It applied both to the pre-pandemic dynamic growth and to the drastic contraction of the market during the COVID times. Nevertheless, on

[5] In the discussed empirical study, the STR market has been understood as the market for STR rental of full apartments, transacted via the most popular online platform used for concluding transactions of this kind – Airbnb. Thus, the obtained results should be considered only a proxy of the results concerning the whole STR market.

average, the volume of the STR market was 3.2 times larger in the centrally located districts. Thus, the outflow of the apartments was much bigger there in nominal terms.

On the other hand, the rent levels on the spatially divided market segments behaved noticeably differently during the pandemic. In its first wave, the seasonally adjusted RSI of the STR market has indicated a 12.7% decline in rents in the central districts and a 9.2% decline in the non-central ones. Then, since the pandemic-related restrictions were remarkably loosened in the summer of 2020, the demand and rent levels recovered. In central districts, the scale of the rebuilt equalled 8.6 p.p. in July 2020, 4.1% below the pre-pandemic level. Still, in the STR market, rents revived to an even larger scale, growing by 10.8 p.p., reaching a higher level than before the crisis outbreak. Yet, the second wave of the pandemic that started in October 2020 brought the rent level in the central districts to the earlier lows in November 2020 – to a value 12.4% lower compared to February 2020. For the non-central districts, the respective decrease was slightly lower and reached -6.2%.

The documented discrepancy between spatial segments may be linked to the increased preference for less congested areas discussed by Filieri et al. (2023) for the STR market and Guglielminetti et al. (2021) for the residential housing market. Additionally, the pandemic-induced limits imposed on the possibility to enter main tourist attractions, often located in city centres, may have resulted in a drop in the attractiveness of central districts perceived by tourists. Finally, the economic crisis might have directed consumers to the market's cheaper, more peripheral segments, resulting in a relative change in the demand for STRs.

In Hebdzyński & Trojanek (2024), the LTR market changes have been discussed using OLS models with the same division – for central and non-central districts of Warsaw. In both cases, the seasonally adjusted LTR rents were decreasing steadily and reached their lowest levels in November 2020. Yet, similarly to the changes documented for the STR market, the changes were more profound in the city's central districts, amounting to -13.3%, while the corresponding value for the rest of Warsaw equalled -8.1%. Nevertheless, in both spatial segments, the increase in rent levels in December 2020 has been noticed, finally reaching the level of -5.9% and -7.6%, respectively, compared to the rent levels from February 2020.

All the discussed studies' results comply with the H7 hypothesis. This way, the dissertation has opened the discussion on the uneven impact of economic shocks on rents in the spatially segmented LTR and STR markets. In this context, it particularly extends recent considerations on rent changes in the LTR market during the pandemic by Kuk et al. (2021) and Tomal &

Marona (2021), and in the STR market by Boto-García (2022) and Cheung (2023). Moreover, it extends the studies showing the spatial differentiation of rent-setting factors (Tomal & Helbich, 2022; Crespo & Grêt-Regamey, 2013) by indicating that also the changes in rents should be considered spatially unequal in times of economic shocks.

H8: During the pandemic, rent levels in the LTR market declined unevenly across quality-related market segments.

In Hebdzyński (2024a), the models for the quality-related segments of the LTR market of Poznan have been constructed to show that the magnitude of rent changes during the pandemic was uneven and depended on the analysed segment. In the first wave of the pandemic, the rent decline first occurred in the high-quality segment and amounted to -7.6% q/q in March 2020. This was supposedly fuelled by the cross-border traffic restrictions that affected international tourism and led to the withdrawal of high-quality apartments from the STR market to list them for the long term. The higher pace of change in this segment may have also been related to the travel restrictions introduced earlier by other national governments and the landlords' expectations that the situation would not improve shortly. Then, in the next quarter, the decrease in rent levels was particularly visible for the low-quality market segment, often targeted by students or external labour migrants. The former were either terminating their rental agreements, expecting further lockdowns they wanted to spend in their family homes, or renegotiating the existing contracts. Concerning migrants – their inflow was slowed because of the cross-border restrictions. As a result, the rent levels declined q/q by 7.2%. At the same time, the rents of medium-quality apartments did not change much. Next, in September 2020, rents were stable, possibly because of tenants' and landlords' will to await the period of uncertainty about the further development of the pandemic. Yet, after the peak of COVID-19 cases in the second wave of the pandemic, in December 2020, the listed rents decreased in all quality segments. Although the low-quality HRI dropped the most, it was also the first to recover in early 2021. The considerable increase in rents recorded in all segments in September 2021 may indicate this date as the moment when the impact of the pandemic shock on the Polish rental market expired.

In total, the decline in rents was the largest for the low-quality segment where it amounted to -16.3% in December 2020 compared to the first observed period after the introduction of lockdown in Poland (in late March 2020). On the other hand, the high-quality market reacted first to the pandemic shock, but the reaction of its rents was not the strongest and reached the level of -10.6%. The decrease in rents of high-quality apartments was fuelled by the inflow of apartments from the STR market and by the economic slowdown that lowered the current and anticipated income of consumers. Yet, the relative preferences of tenants for higher quality might have increased because of the longer time spent at home, as discussed by Mouratidis (2021) and Marona & Tomal (2020). This could have been a factor that levelled the declines in rents in the high-quality market while deepening the decreases in the low-quality market. Lastly, the medium-quality segment’s rent level was affected to the lowest extent . Compared to February 2020, the maximum decline in rents equalled there -9.2% in March 2021. The obtained values point to bigger changes during the pandemic than it has been indicated in Hebdzyński (2024b) for the market of Poznan. Thus, the analysis of quality-related segments has not only revealed the varying scale of changes in market segments but also has captured their uneven timing regarding the response to the pandemic shock, which affected the obtained results.

It may be inferred that the presented results are in line with the H8 hypothesis. In this way, the study has contributed to extending the literature on the process of rental price adjustment in the LTR market to the pandemic shock (Kuk et al., 2021; Tomal & Marona, 2021). Notably, the process could have been uneven concerning the timing and scale of changes. Thus, the usefulness of analysing quality segments might be particularly high amidst the economic shocks that target specific tenants or landlords in the market.

H9: Changes in the STR market rents cause changes in the LTR market rents.

In Hebdzyński & Trojanek (2024), changes in rent levels in the STR and LTR markets in Warsaw have been analysed to extend the knowledge on the price-related channel of transmission of shocks from one sub-market to another. Using the Granger causality test, it has been shown that fluctuations in the STR market rents may not be considered to Granger-cause changes in the LTR market rents. The simultaneous dependency has also been tested, and similarly, no significant relationship has been found. Then, the earlier obtained rent indices for the spatial segments of the STR and LTR market were used to verify the local dependency. Yet, the relation of rents has proven to be significant neither in the case of the central districts of Warsaw nor for the non-central ones.

The obtained results contrast with the stated hypothesis. They suggest that the adverse impact of the STR market on the LTR market should be considered primarily in terms of the relation between the supply of apartments for STR rent and LTR rent level, as shown in the most recent studies (Chaves Fonseca, 2024; Reichle et al., 2023; Mozo Carollo et al., 2024). In this case, the LTR market will be indifferent to fluctuations in STR rents as long as they do not translate to higher or lower market supply.

From the analytical perspective, it means that to understand the process of withdrawing the apartments from the LTR market and using them for the STR market needs (thus decreasing the affordability of residential housing), one should focus on more than just price-related factors. In this case, studying rent premium, as shown by Hill et al. (2023), may be more helpful, as it combines two reflections of the demand for STRs – the achieved rent and the occupancy rate. Finally, one should conclude that as the LTR market rents are insensitive to sole changes of the STR market rents, thus they may be disregarded from the perspective of monitoring and supervision of the LTR market.

H10: During the pandemic, the decrease in the supply of apartments in the STR market contributed to the decline in rents in the LTR market.

In Trojanek et al. (2021), it has been documented that the COVID-related restrictions imposed on domestic and international tourism induced a massive withdrawal of apartments from the STR market, which was unequal across the districts of Warsaw. In the article, the elasticity of rents to the STR market supply changes has been studied using a QR model. The results indicate that a 1% change in STR market volume leads to a 0.031% change in rents. Notably, based on the median quantile results, the elasticity was higher in the period of the LTR market rent growth (before the pandemic). In this regard, 1% change in the STR market volume before the pandemic led to about 0.0322% change in LTR rents, compared to the decreased figure of 0.0219% in the period of the pandemic-induced STR market contraction.

The research results agree with the stated hypothesis and provide evidence of the impact of the STR market volume on the LTR market rents. It has proven to be persistent in times of the growing economy and during times of market contraction. Yet, the magnitude of the effect may differ, as it has been indicated for the pre-pandemic and pandemic periods. Nevertheless, it has been proven that the STR market volume is a factor which should be considered in the analyses aimed at studying fluctuations of the LTR market rent levels.

To my knowledge, it has been, to date, the only research for the CEE countries that targeted the issue of the impact of the STR market volume on LTR rents. In this regard, it complements the international literature on the topic (Horn & Merante, 2017; Benítez-Aurioles & Tussyadiah, 2020; Garcia-López et al., 2020; Barron et al., 2020; Chaves Fonseca, 2024; Reichle et al., 2023; Mozo Carollo et al., 2024; Lee & Kim, 2023). Finally, the study was the first to analyse the relationship between the STR market volume and LTR rents during the contraction of the STR market. Therefore, it has contributed to the formation of the knowledge in this study area, later confirmed by Boto-Garcia (2022), and extended concerning the empirical analysis of the efficiency of introducing the legislative bans on the STR market (Seiler et al., 2024; Chaves Fonseca, 2024).

I.8. Conclusions

The main goal of this dissertation has been to show the multidimensional impact of the COVID-19 pandemic on the housing rental market while highlighting the crucial role of information on housing quality in the analytical process. The results of the hypotheses testing and their relations with the studied subject's literature have been discussed extensively in Section I.7. and in the articles and working paper that form the dissertation. This chapter aims to aggregate the conclusions and present their bigger picture to prove or disprove the stated thesis and outline the dissertation's limitations and future research directions.

The comprehension of similarities and differences between the results of hedonic models that base on listings and transactions has been crucial for understanding the further achieved results. In Hebdzyński (2024c), it has been documented that the results of hedonic models of the LTR market based on listing data show high compliance with the results obtained based on scarce transactional data. It has been outlined that in certain situations, listings are not only a decent substitute for transactions but also may be considered superior. Additionally, it has been argued that since the obtained coefficients of the constructed hedonic models have been consistent, the listings-based coefficients may be regarded as a proxy of tenants' revealed preferences for particular housing characteristics, conventionally obtained based on transactional data. This way, the dissertation has contributed to the revealed preference theory (Samuelson, 1948) and hedonic price theory (Rosen, 1974) by showing that rental price schedule functions modelled based on listings may be considered a good approximation of the ones modelled based on market transactions, which should reflect the markets' equilibrium. It has laid the ground for the assumption that the later obtained results of listing-based hedonic models concerning rent-setting factors and rent indices may be a decent representation of the transacted market phenomena.

Nevertheless, it should be noted that certain discrepancies have also been detected between the results of models based on transactions and listings, particularly concerning rent indices. They have been attributed to the differences in the quality structure of the studied datasets. In Hebdzyński (2024a), it has been discussed that their quality structure may diverge from each other even further in the face of economic shocks that affect unevenly particular supply and demand groups in the LTR market. Then, conducting a separate analysis of quality-related market segments should guarantee higher robustness of hedonic studies' results.

Thus, the issue of apartment quality has been considered a key to deepening the understanding of rental price fluctuations, particularly amidst the economic shocks.

First, focusing on the quality signalling strategies, in Hebdzyński (2023), it has been shown that agents and landlords who list apartments for LTR rent tend to overstate their soft quality declarations that reflect apartment condition and design, taking advantage of information asymmetry (Akerlof, 1970). Additionally, they often decide not to declare soft quality at all, which creates an information gap. Yet, they indirectly signal quality by describing the apartments an posting their pictures. In Hebdzyński (2023), the Wordscores algorithm (Laver et al., 2003) has been adapted in a novel way to extract indirect textual quality signals from listings, which have shown high compliance with the visual signals. This way, the dissertation has contributed to the signalling theory (Spence, 1973, 2002), in the housing market context recently extended by the studies of information disclosure strategies (Benefield et al., 2011; Bian et al., 2021).

Next, the quality-related considerations have been moved to Rosen's framework of hedonic models to empirically verify the importance of accounting for soft quality of apartments in the models of the LTR market. It has been shown that although accounting for soft quality does not change the indicated course of HRIs substantially, compared to the models that disregard soft quality, it may affect the precision of estimates. Moreover, the usage of merely the observations for which the quality has been declared directly may introduce bias to the results. Thus, to maximise the accuracy of hedonic models, one must bridge the gaps in information on the soft quality of apartments. In this regard, the use of Wordscores values as explanatory variable approximating the soft quality of apartments has led to obtaining the models of superior statistical properties.

Then, in Hebdzyński (2024b), it has been shown that during the pandemic, the LTR market rent-setting factors have changed, which, based on Hebdzyński (2024c), proxies changes of tenants' revealed preferences. In Poznan, the increased valuation of leisure-related housing features (the availability of a balcony and a private garden) has been found for the pandemic. Yet, a similar change has not been proven for the marginal price paid for a separate room for remote working or studying. Concerning the locational characteristics, amidst the pandemic, the value of the proximity to urban green areas has increased, while the significance of the distance to university buildings has diminished. The obtained results should be treated as an empirical contribution to the hedonic price theory (Rosen, 1974) and the revealed preference theory (Samuelson, 1948), showing adjustments in consumer choices and market valuations of certain housing features to the pandemic shock.

The scale of the pandemic decline in rents in Poznan has been estimated in Hebdzyński (2024b). However, the change indicated for the whole city has been lower than the changes documented in Hebdzyński (2024a) for the separate quality-related market segments. It has been found that the low-quality segment was affected the most, while the medium-quality segment was affected the least. In the high-quality segment, rents fell even before the pandemic lockdown was introduced in Poland. This may have been indirectly caused by the lockdowns and restrictions on international tourism gradually imposed by other countries, the timing of which varied, and which resulted in the conversion of many high-quality apartments (earlier rented to tourists) from the STR market to the LTR market. Therefore, as a reason for the discrepancy between the results of analyses of the whole market and the quality-related market segments, one should see the unequal timing of changes in segments. Hence, the study has shown that the quality signals may be used for the LTR market segmentation to widen the scope of the analyses and enhance the comprehension of market phenomena. It has contributed to the extension of the literature that presented rent indices as universal for the local LTR markets, e.g. Gluszak & Trojanek (2024) and Trojanek & Gluszak (2022).

Additional analysis was necessary to understand how the LTR market reacts to changes in the STR market. In Trojanek et al. (2021), it has been shown that in Warsaw, the changes in the size of the STR market influence the LTR market rents. Although the effect was weaker during the pandemic, when the STR market contracted, it still was significant. At the time of publication of the article, there was no study analysing the relationship between the size of the STR market and LTR rents in the reality of the STR market contraction. Therefore, the article has contributed to the formation of knowledge in this study area, later confirmed by Boto-Garcia (2022) and extended by Seiler et al. (2024) and Chaves Fonseca (2024).

In Trojanek et al. (2021), it has also been indicated that the impact of the STR market on LTR rents may be spatially unequal because, during the pandemic, the strongest LTR rent decreases have been noticed in the STR-dense districts. Therefore, the spatial differentiation of the rental market of Warsaw has been studied in Hebdzyński & Trojanek (2024). It has been shown that although the size of the STR market in the centrally located districts was greater than in the non-central ones, the dynamics of the STR market volume were similar in years

2015-2020. Yet, in the non-central districts, the pandemic-induced decline in STR rents was noticeably weaker, and the recovery process was stronger.

The LTR market rents decreased steadily from the beginning of the pandemic until November 2020. Similarly to the STR market, the changes were more profound in the central districts of the city, yet, the HRI for December 2020 has shown higher value for the city centre. As a result, the dissertation has opened the discussion on the uneven magnitude and timing of the impact of economic shocks on rents in the spatially segmented LTR and STR markets. This way it has added to the research by Filieri et al. (2023) for the STR market and Tomal & Helbich (2022) for the LTR market that discussed the pandemic-induced increase of tenants' preferences for lower-congested areas, thus contributing to hedonic price theory (Rosen, 1974).

Finally, to extend the understanding of the STR and LTR market dependency, the causal relationship has been tested by Hebdzyński & Trojanek (2024). It has been proven that fluctuations of the STR market rents may not be considered to Granger-cause changes in the LTR market rents, neither on the whole city level nor on the level of spatial segments of Warsaw. The results indicate that the markets' interdependence should be considered primarily in terms of the relationship between the supply of apartments for STR rent and the LTR rent level. Thus, monitoring rental price changes in the STR market is not significant from the perspective of studying the LTR rent changes.

Based on the above presented results and conclusions, the dissertation answered both research questions, thus confirming the thesis formulated.

<u>The additional contribution</u>

Although the dissertation, in principle, targeted the LTR market in the face of the pandemic shock, the robustness of the proposed methods has been tested on the extended scope. As a result, several additional contributions of the dissertation's studies may be found. They should be regarded as proof of the universality of the proposed methods and approaches for analysing the housing market.

First, in Hebdzyński (2023), the sellers' information disclosure strategies, information gaps, compliance of quality signals provided via listings, and the accuracy of the Wordscores approach have been tested not only on the LTR market but also on the housing sales market.

This has been done to provide a picture of the studied phenomena for Poland's whole residential housing segment.

Then, in Hebdzyński (2024a), the analysis of the quality-related LTR market segments was conducted for both the pandemic shock and the one related to the outbreak of the war in Ukraine. It has been shown that in both cases, the quality-related segments were affected unevenly. Thus, it may be argued that the proposed market segmentation approach may be used to analyse any future shock that influences the LTR market.

Finally, in Hebdzyński (2024b), the impact of both pandemic-related and war-related shocks on the LTR market has been considered with the same methodological approach. As a result, knowledge on the processes of adjustment of rent-setting factors and rent levels to shocks has been extended. Additionally, the current state of the market and its rent-determining factors that mirror consumer choices has been outlined.

Policy implications

The LTR market is a crucial yet underdeveloped part of the housing market in Poland and in most of the CEE countries. In this regard, the methodological progress resulting from the improved knowledge of quality signalling should contribute to widening the scope of future hedonic analyses and their higher precision, thus improving market information and market efficiency. Moreover, the advance in understanding the results of the studies that are based on imperfect listing data may contribute to the more effective transfer of knowledge on the market from the scientific literature to rental market entities. It should be noted that although the novel analytical approaches to study the market with the use of listing data have been tested locally, the presented methods may be applied to the specificity of rental markets worldwide. The added value of using these approaches in each case would depend on the quality and availability of transactional data. Nevertheless, some discussed properties of listing data would make them superior to transactional data in all cases, contributing to the universality of the presented approaches. From the perspective of market monitoring and supervision, the extended analytical techniques should provide an enhanced comprehension of the phenomena that occur in the market and may be considered an essential input to the national macroeconomic stability programmes. Finally, a deep understanding of the interdependency between the STR and LTR markets is of the utmost importance for conducting a sustainable residential housing policy and imposing only the necessary regulations on STR rentals. It is needed in order to reconcile goals related to the promotion of short-term tourism, which may boost the economic growth of the regions, with the desire to limit its negative externalities.

Yet, improving the analytical toolkit has been stimulated by the particular need to indicate and interpret changes that took place in the rental market in the face of the COVID-19 pandemic. Even though its impact on the global economies has been levelled, infectious diseases pose a further threat to social and economic life. By drawing conclusions from the multidimensional impact of the pandemic evidenced in this dissertation, private and public entities may obtain information on the market adjustment processes to respond quicker and with less uncertainty to similar shocks in the future. Moreover, it has been shown that the pandemic has induced structural changes in the market. Thus, the provided updated knowledge of the current state of the post-pandemic rental market may be particularly useful for supply-side entities. In this regard, the knowledge of tenants' choices and their (proxied) revealed preferences obtained using quantitative methods may be used by real estate professionals to provide products well-suited to the needs of today’s tenants.

<u>Research limitations and future directions of research</u>

Though very promising, the dissertation’s results have some limitations that relate mainly to the constraints of the methods used for the market analysis in particular articles or working paper. They should be addressed in future rental market studies. It may be done using the proposed studies’ extensions.

Regarding the studies targeting quality signals, Hebdzyński (2023) and Hebdzyński (2024a) have utilised visual quality assessments, which have been done by one person, thus including a non-avoidable inconsistency and subjectivity. However, it may be suspected that the additional variance would lower the achieved results on the accuracy of the Wordscores method and the estimated consistency of quality signals. Thus, comparing Wordscores dictionaries built based on single-person and multiple-person quality assessments should lead to extracting quality signals with a higher accuracy. It may result in obtaining the hedonic models of even better statistical properties compared to those that disregard soft quality.

Then, one of the drawbacks of the study Hebdzyński & Trojanek (2024), which attempts to define the causal relationship between the STR and LTR market, should be seen in the impossibility of providing a profound quality adjustment. Therefore, the quality-constant repeat sales method (BMN) has been used to construct the rent index instead of hedonic methods. Supposedly, utilising texts of STR listings with the use of Wordscores (or similar textual methods) may significantly improve the quality of the results. On the other hand, in Hebdzyński & Trojanek (2024) and Trojanek et al. (2021), the analysis of the pandemic has been restricted to its first two waves and only one city, because of the limited availability of data. Therefore, the interdependency of the STR and LTR market should be reinvestigated using information from a more extended time and spatial range. Notably, in Trojanek et al. (2021), it has been argued that the impact of the STR market supply level on LTR rents diminished during the first months of the pandemic. In this regard, it should be verified whether the proven relation between the STR and LTR markets also holds for the post-pandemic times.

Moreover, in Hebdzyński (2024c), the similarity of results of listings- and transactions-based hedonic models has been tested on the local LTR market of Poznan using a particularly small set of transactional data. Thus, the analysis needs to be re-run to confirm the obtained results. There is a general problem of unavailability of rental market transactional data, which is impossible to avoid in the current market reality in Poland. However, the size of the analytical dataset may be increased by targeting a larger Polish market or the market from another CEE country (similar to the LTR market size and structure to the Polish one), for which transactional and listing data will be available. It may enable the re-run of the hedonic analysis utilising more advanced econometric methods, thus obtaining more stable and precise estimates. Additionally, it may allow exploring further similarity of results of listings- and transactions-based hedonic models (on both micro and macro level) as they may depend on the phase of the business cycle. The analysis may be also extended to provide information on the possible delay of HRIs. More specifically, the listing phase is always a phase preceding the transaction; hence, it may be suspected that HRIs of housing listings may be leading HRIs of housing transactions, which would be important from the perspective of market supervision. This may add further validity to the analysis of the listing data.

Finally, the LTR market in Poland is a place of widely accessible listing data and scarce transactional data. It has been argued that none of them fully represents the quality structure of the rental market. Still, although on a limited scale, public institutions do gather transactional data for market monitoring and supervision. Thus, one should strive to not only understand to what extent listing data may be used as substitutes for transactional data but also attempt to combine information from the two sources. In this regard, the Heckman correction method (Heckman, 1976) may be used to overcome the problems of non-random sample selection while utilising information from both types of data. It should be considered a major challenge to understand better the phenomena in the LTR market, including those resulting from super-shocks, such as the pandemic.

I.9. Acknowledgements

My time at the PUEB Doctoral School was definitely not easy, but it was very productive and, finally, satisfying. I met numerous inspiring people who not only helped me enhance my scientific work but also contributed to my personal development.

First of all, I would like to thank my Supervisor – dr hab. Radosław Trojanek, prof UEP, for his continuous, exceptional support during the whole process of preparation of this doctoral dissertation. He was a demanding supervisor, not accepting mediocrity and encouraging me to aim high. Still, he was able to praise my successes and offered his unlimited availability to help me find my way in the scientific world. Lastly, I am particularly grateful for his efforts to curb my desire to work late at night or to discuss my scientific problems even during my vacation.

Moreover, I would like to express my gratitude to all those who helped me to develop this dissertation, among others to teachers at the PUEB Doctoral School, professors and doctors from the Department of Microeconomics and Institute of Economics, colleagues and supervisors from the National Bank of Poland, journals' editors, articles' anonymous reviewers and attendees of the conferences that I have been to. Then, I cannot forget about my friends I met in the Doctoral School, especially Ania, Kasia and Cezary, who were incredible companions in fortune and misfortune and helped me to find the balance between scientific and everyday life.

On the other hand, I would like to thank my family, who have shaped me as a person and given me relief from all the challenges associated with my doctoral studies – my mom Ewa, brother Mariusz and sister Agnieszka. I would especially like to express my endless appreciation to my grandfather Henryk, who taught me logical reasoning processes, and my grandmother Róża, who always encouraged me to read and write. All these have proven to be extremely useful during the preparation of this dissertation.

Last but definitely not least, the completion of this dissertation would not have been possible without my fiancée Klaudia, who never doubted, offered me unconditional support, and patiently tolerated my countless hours of physical or mental absence. Although I take the blame for all the imperfections of this dissertation, its completion should be considered our mutual success.

I.10. List of articles and studies that constitute the doctoral dissertation

1. Hebdzyński, M. (2024c). Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data?, *Bank i Kredyt, 55*(4). https://bankikredyt.nbp.pl/content/2024/04/bik_04_2024_05m.pdf

2. Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures?, *Journal of Housing and the Built Environment, 38*, 2399–2425. https://doi.org/10.1007/s10901-023-10043-z

3. Hebdzyński, M. (2024a). Are hedonic models really quality-adjusted? The role of apartment quality in hedonic models of housing rental market. *Real Estate Management and Valuation*, *32*(2), 46-57. https://doi.org/10.2478/remav-2024-0014

4. Hebdzyński, M. (2024b). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies, 17*(2), 298-323. https://doi.org/10.14254/2071-8330.2024/17-2/16

5. Trojanek, R., Gluszak, M., Hebdzyński, M. & Tanas, J. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

6. Hebdzyński, M., Trojanek, R. (2024). *Is there a relationship between rents in the short- and long-term housing rental markets? A case study of Warsaw*. Working paper.

I.11. Bibliography

Ahlfeldt, G. M., Heblich, S., & Seidel, T. (2023). Micro-geographic property price and rent indices. *Regional Science and Urban Economics*, *98*. https://doi.org/10.1016/j.regsciurbeco.2022.103836

Akerlof, G. A. (1970). The Market for "Lemons": Qualty Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, *84*, 488–500. https://doi.org/10.1016/b978-0-12-214850-7.50022-x

Anenberg, E., & Laufer, S. (2017). A more timely house price index. *Review of Economics and Statistics*, *99*(4), 722–734. https://doi.org/10.1162/REST_a_00634

Ardila, D., Ahmed, A., & Sornette, D. (2021). Comparing ask and transaction prices in the Swiss housing market. *Quantitative Finance and Economics*, *5*(1), 67–93. https://doi.org/10.3934/qfe.2021004

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association*, *58*(304), 933–942. https://doi.org/10.1080/01621459.1963.10480679

Baranzini, A., & Ramirez, J. V. (2005). Paying for quietness: The impact of noise on Geneva rents. *Urban Studies*, *42*(4), 633–646. https://doi.org/10.1080/00420980500060186

Barron, K., Kung, E., & Proserpio, D. (2020). The effect of home-sharing on house prices and rents: Evidence from Airbnb. *Marketing Science*, *40*(1), 23–47. https://doi.org/10.1287/mksc.2020.1227

Batalha, M., Gonçalves, D., Peralta, S., & Pereira dos Santos, J. (2022). The virus that devastated tourism: The impact of covid-19 on the housing market. *Regional Science and Urban Economics*, *95*, 103774. https://doi.org/10.1016/j.regsciurbeco.2022.103774

Benefield, J. D., Cain, C. L., & Johnson, K. H. (2011). On the Relationship Between Property Price, Time-on-Market, and Photo Depictions in a Multiple Listing Service. *Journal of Real Estate Finance and Economics*, *43*(3), 401–422. https://doi.org/10.1007/s11146-009-9219-6

Benítez-Aurioles, B., & Tussyadiah, I. (2021). What Airbnb does to the housing market. *Annals of Tourism Research*, *90*, 103108. https://doi.org/10.1016/j.annals.2020.103108

Bian, X., Contat, J. C., Waller, B. D., & Wentland, S. A. (2021). Why Disclose Less Information?

Toward Resolving a Disclosure Puzzle in the Housing Market. *Journal of Real Estate Finance and Economics*. https://doi.org/10.1007/s11146-021-09824-6

Boros, L., Dudás, G., & Kovalcsik, T. (2020). The effects of COVID-19 on airbnb. *Hungarian Geographical Bulletin*, *69*(4), 363–381. https://doi.org/10.15201/hungeobull.69.4.3

Boto-García, D. (2022). Heterogeneous price adjustments among Airbnb hosts amid COVID-19: Evidence from Barcelona. *International Journal of Hospitality Management*, *102*, 103169. https://doi.org/10.1016/j.ijhm.2022.103169

Bresciani, S., Ferraris, A., Santoro, G., Premazzi, K., Quaglia, R., Yahiaoui, D., & Viglia, G. (2021). The seven lives of Airbnb. The role of accommodation types. *Annals of Tourism Research*, *88*, 103170. https://doi.org/10.1016/j.annals.2021.103170

Broitman, D. (2023). "Passive" Ecological Gentrification Triggered by the Covid-19 Pandemic. *Urban Planning*, *8*(1), 312–321. https://doi.org/10.17645/up.v8i1.6015

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *Journal of Real Estate Finance and Economics*, *41*(4), 390–411. https://doi.org/10.1007/s11146-009-9177-z

Bryx, M., Sobieraj, J., Metelski, D., & Rudzka, I. (2021). Buying vs. Renting a Home in View of Young Adults in Poland. *Land, 10*(11). https://doi.org/10.3390/land10111183

Byrne, M. (2020). Towards a political economy of the private rental sector. *Critical Housing Analysis*, *7*(1), 103–113. https://doi.org/10.13060/23362839.2020.7.1.507

Cajias, M., Fuerst, F., & Bienert, S. (2019). Tearing down the information barrier: the price impacts of energy efficiency ratings for buildings in the German rental market. *Energy Research & Social Science*, *47*, 177–191. https://doi.org/10.1016/j.erss.2018.08.014

Centrum AMRON. (2020). *Studenci na rynku nieruchomości*. https://www.amron.pl/strona.php?tytul=studenci-na-rynku-nieruchomosci

Chaves Fonseca, C. (2024). Short-term rentals and residential rents: evidence from a regulation in Santa Monica. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-01-2024-0001

Cheng, P., Lin, Z., & Liu, Y. (2020). Price Discovery with Heterogeneous Sellers in Real Estate. *Journal of Real Estate Research*, *42*(3), 341–364. https://doi.org/10.1080/08965803.2020.1840898

Cheng, P., Lin, Z., Liu, Y., & Seiler, M. J. (2015). The benefit of search in housing markets.

Journal of Real Estate Research, *37*(4), 597–621. https://doi.org/10.1080/10835547.2015.12091430

Cheung, K. S. (2023). Airbnb pricing and term structure: A temporal analysis of omitted variable bias and repeat sales method as remedies. *Tourism Economics*, *30*(5), 1330–1338. https://doi.org/10.1177/13548166231191551

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Crespo, R., & Grêt-Regamey, A. (2013). Local hedonic house-price modelling for urban planners: Advantages of using local regression techniques. *Environment and Planning B: Planning and Design*, *40*(4), 664–682. https://doi.org/10.1068/b38093

Cui, N., Gu, H., Shen, T., & Feng, C. (2018). The impact of micro-level influencing factors on home value: A housing price-rent comparison. *Sustainability (Switzerland)*, *10*(12), 1–23. https://doi.org/10.3390/su10124343

Czerniak, A., & Rubaszek, M. (2018). The Size of the Rental Market and Housing Market Fluctuations. *Open Economies Review*, *29*(2), 261–281. https://doi.org/10.1007/s11079-017-9452-1

Day, B. H. (2001). *The theory of hedonic markets: obtaining welfare measures for changes in environmental quality using hedonic market data*. Economics for the Environment Consultancy.

de Wit, E. R., & Van der Klaauw, B. (2013). Asymmetric information and list-price reductions in the housing market. *Regional Science and Urban Economics*, *43*(3), 507–520. https://doi.org/10.1016/j.regsciurbeco.2013.03.001

Demir, E., Zoğal, V., Yelsiz, M. Ş., & Emekli, G. (2024). Mapping the effects of COVID-19 on Airbnb: Evidence from European cities. *Anatolia*, *35*(1), 135–162. https://doi.org/10.1080/13032917.2023.2254780

Diewert, W. E., & Shimizu, C. (2022). Residential Property Price Indexes: Spatial Coordinates Versus Neighborhood Dummy Variables. *Review of Income and Wealth*, *68*(3), 770–796. https://doi.org/10.1111/roiw.12534

Diewert, W. E. (2009). The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions. pp 87-116 in W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (Ed.), Price and Productivity Measurement: *Volume 1 – Housing*. Trafford Press.

DiPasquale, D., & Wheaton, W. C. (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework. *Journal of the American Real Estate and Urban Economics Association*, *V20*(1), 191–197. https://doi.org/10.1111/1540-6229.00579

Dolnicar, S., & Zare, S. (2020). COVID19 and Airbnb – Disrupting the Disruptor. *Annals of Tourism Research*, *83*(102961). https://doi.org/10.1016/j.annals.2020.102961

Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, *52*, 1–22. https://doi.org/10.1016/j.tra.2013.04.002

Eichholtz, P., Straetmans, S., & Theebe, M. (2012). The Amsterdam rent index: The housing market and the economy, 1550–1850. *Journal of Housing Economics*, *21*(4), 269–282. https://doi.org/10.1016/j.jhe.2012.09.001

European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank. (2013). *Handbook on Residential Property Prices (RPPIs): 2013 edition*. OECD Publishing. https://doi.org/10.2785/34007

Eurostat. (2024a). *Distribution of population by tenure status, type of household and income group - EU-SILC survey*. https://doi.org/10.2908/ILC_LVHO02

Eurostat. (2024b). *Overcrowding rate by age, sex and poverty status - total population - EU-SILC survey*. https://doi.org/10.2908/ilc_lvho05a

Eurostat. (2024c). *Guest nights spent at short-stay accommodation offered via collaborative economy platforms*. https://doi.org/10.2908/TOUR_CE_OAN3

Eurostat. (2024d). *Real GDP growth rate*. https://doi.org/10.2908/TEC00115

Filieri, R., Milone, F. L., Paolucci, E., & Raguseo, E. (2023). A big data analysis of COVID-19 impacts on Airbnbs' bookings behavior applying construal level and signaling theories. *International Journal of Hospitality Management*, *111*, 103461. https://doi.org/10.1016/j.ijhm.2023.103461

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

FSB (Financial Stability Board), IMF (International Monetary Fund). (2009). *The Financial Crisis and Information Gaps: Report to the G-20 Finance Ministers and Central Bank Governors*.

Gallent, N., & Madeddu, M. (2021). Covid-19 and London's Decentralising Housing Market–What are the Planning Implications? *Planning Practice and Research*, *36*(5), 567–577.

https://doi.org/10.1080/02697459.2021.1964782

Gamal, A., Rohmah, L., & Muhyi, M. M. (2023). Housing preference shifting during COVID-19 pandemic in Indonesia. *Journal of Urban Management*, *12*(3), 268–283. https://doi.org/10.1016/j.jum.2023.05.002

Garcia-López, M. À., Jofre-Monseny, J., Martínez-Mazza, R., & Segú, M. (2020). Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona. *Journal of Urban Economics*, *119*(June 2019). https://doi.org/10.1016/j.jue.2020.103278

Gluszak, M., & Trojanek, R. (2024). War in Ukraine, the refugee crisis, and the Polish housing market. *Housing Studies*. https://doi.org/10.1080/02673037.2024.2334822

Goodwin, K. R., Waller, B. D., & Weeks, H. S. (2018). Connotation and Textual Analysis in Real Estate Listings. *Journal of Housing Research*, *27*(2), 93–106. https://doi.org/10.1080/10835547.2018.12092149

Goodwin, K., Waller, B., & Weeks, H. S. (2014). The Impact of Broker Vernacular in Residential Real Estate. *Journal of Housing Research*, *23*(2), 143–161. https://doi.org/10.1080/10835547.2014.12092089

Gossen, J., & Reck, F. (2021). The End of the Sharing Economy? Impact of COVID-19 on Airbnb in Germany. *Economic Research Guardian*, *11*(2), 255–269.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Grossman, S. J. (1981). Warranties and Private Disclosure About Product Quality. *Journal of Law and Economics*, *24*(3), 461–483. https://www.jstor.org/stable/725273

Guglielminetti, E, Loberto, M., & Mistretta, A. (2023). The impact of COVID-19 on the European short-term rental market. *Empirica*, *50*(3), 585–623. https://doi.org/10.1007/s10663-023-09576-x

Guglielminetti, E., Loberto, M., Zevi, G., & Zizza, R. (2021). Living on my Own: The Impact of the Covid-19 Pandemic on Housing Preferences. *Bank of Italy Occasional Paper No. 627*. https://doi.org/10.2139/ssrn.3891671

GUS (Główny Urząd Statystyczny). (2023). *Narodowy Spis Powszechny 2021 - Mieszkania zamieszkane wg okresu budowy budynku*.

Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2017). Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study. *Journal of Travel Research*, *57*(3), 342–359. https://doi.org/10.1177/0047287517696980

Haag, J., Rutherford, R., & Thomson, T. (2000). Real Estate Agent Remarks: Help or Hype? *Journal of Real Estate Research*, *20*(1–2), 205–216. https://doi.org/10.1080/10835547.2000.12091024

Hahn, J., Hirsch, J., & Bienert, S. (2018). Does "clean" pay off ? Housing markets and their perception of heating technology. *Property Management*, *36(5)*, 575–596. https://doi.org/10.1108/PM-08-2017-0051

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, *5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures? *Journal of Housing and the Built Environment, 30,* 2399-2425. https://doi.org/10.1007/s10901-023-10043-z

Hebdzyński, M. (2024a). Are Hedonic Models Really Quality-Adjusted? The Role of Apartment Quality in Hedonic Models of Housing Rental Market. *Real Estate Management and Valuation*, *32*(2), 46–57. https://doi.org/10.2478/remav-2024-0014

Hebdzyński, M. (2024b). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies*, *17*(2), 298–323. https://doi.org/10.14254/2071-8330.2024/17-2/16

Hebdzyński, M. (2024c). Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data? *Bank i Kredyt*, *55*(4), 519–550. https://bankikredyt.nbp.pl/content/2024/04/bik_04_2024_05m.pdf

Hebdzyński, M., & Trojanek, R. (2024). *Is there a relationship between rents in the short- and long-term housing rental markets? A case study of Warsaw*. Working paper.

Heckman, J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models Annals of Economic and Social Measurement, 5(4), 475-492. https://www.nber.org/chapters/c10491

Hill, R. J. (2004). Constructing Price Indexes across Space and Time: The Case of the European Union. *American Economic Review, 94*(5): 1379–1410. https://doi.org/ 10.1257/0002828043052178

Hill, R. J., Pfeifer, N., & Steurer, M. (2023). The Airbnb rent premium and the crowding-out of long-term rentals. *Journal of Housing Economics*, *61*, 101935. https://doi.org/10.1016/j.jhe.2023.101935

Hill, R. J., & Trojanek, R. (2022). An evaluation of competing methods for constructing house price indexes: The case of Warsaw. *Land Use Policy*, *120*, 106226. https://doi.org/10.1016/j.landusepol.2022.106226

Hoffmann, J., & Kurz, C. (2002). Rent Indices for Housing in West Germany. *European Central Bank Working Papers Series*, *116*.

Horn, K., & Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. *Journal of Housing Economics*, *38*, 14–24. https://doi.org/10.1016/j.jhe.2017.08.002

Horowitz, J. L. (1992). The role of the list price in housing markets: Theory and an econometric model. *Journal of Applied Econometrics*, *7*(2), 115–129. https://doi.org/10.1002/jae.3950070202

Hotz, V. J., & Xiao, M. (2013). Strategic information disclosure: The case of multiattribute products with heterogeneous consumers. *Economic Inquiry*, *51*(1), 865–881. https://doi.org/10.1111/j.1465-7295.2010.00340.x

Iyer, S. R., & Simkins, B. J. (2022). COVID-19 and the Economy: Summary of research and future directions. *Finance Research Letters*, *47*(PB), 102801. https://doi.org/10.1016/j.frl.2022.102801

Kadi, J., Schneider, A., & Seidl, R. (2020). Short-term rentals, housing markets and covid-19: Theoretical considerations and empirical evidence from four austrian cities. *Critical Housing Analysis*, *7*(2), 47–57. https://doi.org/10.13060/23362839.2020.7.2.514

Kholodilin, K. A., Mense, A., & Michelsen, C. (2017). The market value of energy efficiency in buildings and the mode of tenure. *Urban Studies*, *54*(14), 3218–3238. https://doi.org/10.1177/0042098016669464

Kholodilin, K. A., & Rieth, M. (2023). Viral shocks to the world economy. *European Economic Review*, *158*, 104526. https://doi.org/10.1016/j.euroecorev.2023.104526

Knight, J. R., Sirmans, C. F., & Turnbull, G. K. (1994). List price signaling and buyer behavior in the housing market. *The Journal of Real Estate Finance and Economics*, *9*(3), 177–192. https://doi.org/10.1007/BF01099271

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50.

https://doi.org/10.2307/1913643

Kolbe, J., Schulz, R., Wersing, M., & Werwatz, A. (2021). Real estate listings and their usefulness for hedonic regressions. *Empirical Economics, 61, 3239–3269*. https://doi.org/10.1007/s00181-020-01992-3

Kuk, J., Schachter, A., Faber, J. W., & Besbris, M. (2021). The COVID-19 Pandemic and the Rental Market: Evidence From Craigslist. *American Behavioral Scientist, 65*(12). https://doi.org/10.1177/00027642211003149

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. https://www.jstor.org/stable/1828835

Łaszek, J., Augustyniak, H., & Olszewski, K. (2021). The development of the rental market in Poland. In R. Sobiecki, J. Łaszek, & K. Olszewski (Eds.), *Real Estate at Exposure. New Challenges, Old Problems*. Warsaw School of Economics.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, *97*(2). https://doi.org/10.1017/S0003055403000698

Lee, D. (2016). How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harvard Law & Policy Review*, *10*, 229–253.

Lee, S., & Kim, H. (2023). Four shades of Airbnb and its impact on locals: A spatiotemporal analysis of Airbnb, rent, housing prices, and gentrification. *Tourism Management Perspectives*, *49*, 101192. https://doi.org/10.1016/j.tmp.2023.101192

Li, H., Wei, Y. D., & Wu, Y. (2019). Analyzing the private rental housing market in Shanghai with open data. *Land Use Policy*, *85*, 271–284. https://doi.org/10.1016/j.landusepol.2019.04.004

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies*, *8*(1), 1–41. https://doi.org/10.1093/rcfs/cfy009

Liu, C. H., Nowak, A. D., & Smith, P. S. (2020). Asymmetric or Incomplete Information about Asset Values? *Review of Financial Studies*, *33*(7), 2898–2936. https://doi.org/10.1093/rfs/hhz096

Liu, Q., Zhao, P., Xiao, Y., Zhou, X., & Yang, J. (2022). Walking Accessibility to the Bus Stop: Does It Affect Residential Rents? The Case of Jinan, China. *Land, 11*(6). https://doi.org/10.3390/land11060860

Liu, S., & Su, Y. (2021). The impact of the COVID-19 pandemic on the demand for density:

Evidence from the U.S. housing market. *Economics Letters*, *207*, 110010. https://doi.org/10.1016/j.econlet.2021.110010

Löchl, M., & Axhausen, K. W. (2010). Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, *3*(2), 39–63. https://doi.org/10.5198/jtlu.v3i2.117

Luchtenberg, K. F., Seiler, M. J., & Sun, H. (2019). Listing Agent Signals: Does a Picture Paint a Thousand Words? *Journal of Real Estate Finance and Economics*, *59*(4), 617–648. https://doi.org/10.1007/s11146-018-9674-z

Lyons, R. C. (2019). Can list prices accurately capture housing price trends? Insights from extreme markets conditions. *Finance Research Letters*, *30*, 228–232. https://doi.org/10.1016/j.frl.2018.10.004

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*(4), 445–465. https://doi.org/10.1002/jae.788

Marcinkowska, E., Gawron, K., & Rejment, M. (2015). Technical and economic aspects of revitalization of down-town tenement-houses in Wrocław. *Journal of Civil Engineering and Management*, *21*(8), 1036–1045. https://doi.org/10.3846/13923730.2015.1027259

Marona, B., & Tomal, M. (2020). The COVID-19 pandemic impact upon housing brokers' workflow and their clients' attitude: Real estate market in Krakow. *Entrepreneurial Business and Economics Review*, *8*(4), 221–232. https://doi.org/10.15678/EBER.2020.080412

Marona, B., & Tomal, M. (2023). The impact of the Covid-19 pandemic on the operations of real estate agencies and the attitudes of their clients in Poland. *International Entrepreneurship Review*, *9*(1), 7–16. https://doi.org/10.15678/IER.2023.0901.01

McCord, M., Davis, P. T., Haran, M., McIlhatton, D., & McCord, J. (2014). Understanding rental prices in the UK: a comparative application of spatial modelling approaches. *International Journal of Housing Markets and Analysis*, *7*(1), 98–128. https://doi.org/10.1108/IJHMA-09-2012-0043

Micallef, B. (2022). Hedonic indices for registered rents in Malta. *The Annual Malta Residential Rental Study Second Edition*, Ministry for Social and Affordable Accomodation, Housing Authority

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications.

The Bell Journal of Economics, *12*(2), 380–391. https://doi.org/10.2307/3003562

Miłkowski, M. (2016). *Polimorfologik 2.1*. https://github.com/morfologik/polimorfologik/releases/tag/2.1

Mouratidis, K. (2021). How COVID-19 reshaped quality of life in cities : A synthesis and implications for urban planning. *Land Use Policy*, *111*(September), 105772. https://doi.org/10.1016/j.landusepol.2021.105772

Mozo Carollo, I., Morandeira-Arca, J., Etxezarreta-Etxarri, A., & Izagirre-Olaizola, J. (2024). Is the effect of Airbnb on the housing market different in medium-sized cities? Evidence from a Southern European city. *Urban Research & Practice*, *17*(2), 260–279. https://doi.org/10.1080/17535069.2023.2179419

Nanda, A., Thanos, S., Valtonen, E., Xu, Y., & Zandieh, R. (2021). Forced homeward: The COVID-19 implications for housing. *Town Planning Review*, *92*(1), 25–31. https://doi.org/10.3828/tpr.2020.79

Nasreen, Z., & Ruming, K. (2022). Struggles and opportunities at the platform interface: tenants' experiences of navigating shared room housing using digital platforms in Sydney. *Journal of Housing and the Built Environment*, *37*(3), 1537–1554. https://doi.org/10.1007/s10901-021-09909-x

National Bank of Poland. (2022). *Badanie rynku nieruchomości - Instrukcja badania rynku wtórnego*. https://ssl.nbp.pl/publikacje/rynek_nieruchomosci/Instrukcja-badania-rynku-wtornego-2022.pdf

National Bank of Poland. (2023). *Sytuacja życiowa i ekonomiczna migrantów z Ukrainy w Polsce w 2023 roku*. https://nbp.pl/wp-content/uploads/2024/01/raport_migranci_z-Ukrainy_2023.pdf

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, *78*, 185–193. https://doi.org/10.1016/j.ijsu.2020.04.018

Noszczyk, T., Gorzelany, J., Kukulska-Kozieł, A. & Hernik, J. (2022). *The impact of the COVID-19 pandemic on the importance of urban green spaces to the public*. *113*(February 2022). https://doi.org/10.1016/j.landusepol.2021.105925

Nowak, A. D., Price, B. S., & Smith, P. S. (2021). Real Estate Dictionaries Across Space and Time. *Journal of Real Estate Finance and Economics*, *62*(1), 139–163.

https://doi.org/10.1007/s11146-019-09740-w

Nowak, A., & Smith, P. (2017). Textual Analysis in Real Estate. *Journal of Applied Econometrics*, *32*(4), 896–918. https://doi.org/10.1002/jae

Olszewski, K., Waszczuk, J., & Widłak, M. (2017). Spatial and Hedonic Analysis of House Price Dynamics in Warsaw, Poland. *Journal of Urban Planning and Development*, *143*(3), 04017009. https://doi.org/10.1061/(asce)up.1943-5444.0000394

Polityka Insight. (2022). *Kwartalnik mieszkaniowy. Raport o sytuacji na rynku mieszkań w I kwartale 2022 roku*. https://www.politykainsight.pl/_resource/multimedium/20305820

Poursaeed, O., Matera, T., & Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications*, *29*(4), 667–676. https://doi.org/10.1007/s00138-018-0922-2

Reichle, P., Fidrmuc, J., & Reck, F. (2023). The sharing economy and housing markets in selected European cities. *Journal of Housing Economics*, *60*, 101914. https://doi.org/10.1016/j.jhe.2023.101914

Renigier-Biłozor, M., Janowski, A., Walacik, M., & Chmielewska, A. (2022). Human emotion recognition in the significance assessment of property attributes. *Journal of Housing and the Built Environment*, *37*(1), 23–56. https://doi.org/10.1007/s10901-021-09833-0

Ronald, R. (2008). *The ideology of home ownership: Homeowner societies and the role of housing*. https://doi.org/10.1057/9780230582286

Ronald, R., & Kadi, J. (2018). The Revival of Private Landlords in Britain's Post-Homeownership Society. *New Political Economy*, *23*(6), 786–803. https://doi.org/10.1080/13563467.2017.1401055

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Rubaszek, M. (2012). *Modelowanie polskiej gospodarki z pakietem R*. Oficyna Wydawnicza - Szkoła Główna Handlowa.

Rubaszek, M., & Czerniak, A. (2017). Preferencje Polaków dotyczace struktury własnosciowej mieszkan: Opis wyników ankiety. *Bank i Kredyt*, *48*(2), 197–233. https://bankikredyt.nbp.pl/content/2017/02/BIK_02_2017_04Rubaszek.pdf

Rubaszek, M., & Rubio, M. (2020). Does the rental housing market stabilize the economy? A micro and macro perspective. *Empirical Economics*, *59*(1), 233–257. https://doi.org/10.1007/s00181-019-01638-z

Rubaszek, M., Stenvall, D., & Uddin, G. S. (2024). Rental market structure and housing dynamics: An interacted panel VAR investigation. *International Journal of Finance & Economics*. https://doi.org/10.1002/ijfe.2937

Sainaghi, R., & Chica-Olmo, J. (2022). The effects of location before and during COVID-19: Impacts on revenue of Airbnb listings in Milan (Italy). *Annals of Tourism Research*, *96*, 103464. https://doi.org/10.1016/j.annals.2022.103464

Samuelson, P. A. (1948). Consumption Theory in Terms of Revealed Preference. *Economica*, *15*(60), 243. https://doi.org/10.2307/2549561

Seiler, M. J., Siebert, R. B., & Yang, L. (2024). Airbnb or not Airbnb? That is the question: How Airbnb bans disrupt rental markets. *Real Estate Economics*, *52*(1), 239–270. https://doi.org/https://doi.org/10.1111/1540-6229.12440

Seo, Y., Im, J., & Mikelbank, B. (2020). Does the Written Word Matter? The Role of Uncovering and Utilizing Information from Written Comments in Housing Ads. *Journal of Housing Research*, *29*(2), 133–155. https://doi.org/10.1080/10527001.2020.1849929

Shen, L., & Ross, S. L. (2021). Information Value of Property Description: A Machine Learning Approach. *Journal of Urban Economics*, *121*. https://doi.org/10.1016/j.jue.2020.103299

Shimizu, C., Nishimura, K. G., & Watanabe, T. (2016). House prices at different stages of the buying/selling process. *Regional Science and Urban Economics*, *59*, 37–53. https://doi.org/10.1016/j.regsciurbeco.2016.04.001

Sieger, L., & Weber, C. (2023). Inefficient markets for energy efficiency ? – The efficiency premium puzzle in the German rental housing market. *Energy Policy*, *183*(January), 113819. https://doi.org/10.1016/j.enpol.2023.113819

Sirmans, S., Sirmans, C., & Benjamin, J. (1989). Determining Apartment Rent: The Value of Amenities, Services and External Factors. *Journal of Real Estate Research*, *4*(2), 33–43. https://doi.org/10.1080/10835547.1989.12090581

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics*, *87*, 355–374. https://doi.org/10.1016/b978-0-12-214850-7.50025-5

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, *92*(3), 434–459. https://doi.org/10.1257/00028280260136200

Tarr, G. (2012). Small sample performance of quantile regression confidence intervals. *Journal of Statistical Computation and Simulation*, *82*(1), 81–94.

https://doi.org/10.1080/00949655.2010.527844

Taylor, L. O. (2003). The Hedonic Model. *A Primer on Nonmarket Valuation*, 331–393.

Terlikowski, W. (2013). Revitalization capacity of public utility buildings. *Materiały Budowlane*, *6*(490), 48–50.

Tomal, M. (2020). Modelling housing rents using spatial autoregressive geographically weighted regression: A case study in Cracow, Poland. *ISPRS International Journal of Geo-Information*, *9*(6). https://doi.org/10.3390/ijgi9060346

Tomal, M., & Helbich, M. (2022). The private rental housing market before and during the COVID-19 pandemic: A submarket analysis in Cracow, Poland. *Environment and Planning B: Urban Analytics and City Science*, *49*(6), 1646–1662. https://doi.org/10.1177/23998083211062907

Tomal, M., & Helbich, M. (2023). A spatial autoregressive geographically weighted quantile regression to explore housing rent determinants in Amsterdam and Warsaw. *Environment and Planning B: Urban Analytics and City Science*, *50*(3), 579–599. https://doi.org/10.1177/23998083221122790

Tomal, M., & Marona, B. (2021). The impact of the covid-19 pandemic on the private rental housing market in poland: What do experts say and what do actual data show? *Critical Housing Analysis*, *8*(1), 24–35. https://doi.org/10.13060/23362839.2021.8.1.520

Tomczyk, E., & Widłak, M. (2010). Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy. *Bank i Kredyt*, *41*(1), 99–128. https://bankikredyt.nbp.pl/content/2010/01/bik_01_2010_05_art.pdf

Trojanek, R. (2018). *Teoretyczne i metodyczne aspekty wyznaczania indeksów cen na rynku mieszkaniowym*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.

Trojanek, R., & Gluszak, M. (2022). Short-run impact of the Ukrainian refugee crisis on the housing market in Poland. *Finance Research Letters*, *50*(103236). https://doi.org/10.1016/j.frl.2022.103236

Trojanek, R., Gluszak, M., Tanas, J., & Hebdzyński, M. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis*, *8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

von Briel, D., & Dolnicar, S. (2021). The evolution of Airbnb regulation - An international longitudinal investigation 2008–2020. *Annals of Tourism Research*, *87*, 102983. https://doi.org/10.1016/j.annals.2020.102983

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A*, *50*(6), 1147–1170. https://doi.org/10.1177/0308518X18778038

Waugh, F. V. (1928). Quality Factors Influencing Vegetable Prices. *Journal of Farm Economics*, *10*(2), 185–196. https://www.jstor.org/stable/1230278

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934

World Economic Forum. (2024). *The Global Risks Report*. https://www.weforum.org/publications/global-risks-report-2024/

Zhan, D., Xie, C., Zhang, J., & Meng, B. (2023). Investigating the Determinants of Housing Rents in Hangzhou, China: A Spatial Multilevel Model Approach. *Applied Spatial Analysis and Policy*, *16*(4), 1707–1727. https://doi.org/10.1007/s12061-023-09530-1

Zhang, S., Wang, L., & Lu, F. (2019). Exploring housing rent by mixed geographically weighted regression: A case study in Nanjing. *ISPRS International Journal of Geo-Information*, *8*(10). https://doi.org/10.3390/ijgi8100431

ZPP (Związek Przedsiębiorców i Pracodawców). (2021). *Podsumowanie Lockdown-u w Polsce*.

I.12. List of figures and tables

Figures

Tables

Part II. Articles and studies that constitute the doctoral dissertation

Article 1	**Hebdzyński, M. (2024c). Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data?, *Bank i Kredyt, 55*(4). https://bankikredyt.nbp.pl/content/2024/04/bik_04_2024_05m.pdf**

Bank i Kredyt 55(4), 2024, 519-550

BANK I KREDYT
BANK & CREDIT

Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data?

Michał Hebdzyński*

Submitted: 30 October 2023. Accepted: 26 March 2024.

Abstract

The problematic access to housing rental market transactional data forces the use of alternative sources. Using the observations of transactions and listings of apartments for rent located in Poznań, it was checked whether the fruits of hedonic models based on listings data may be treated as a proxy of consumers' revealed preferences. Moreover, the study answered whether listings and transactions-based hedonic rent indices show the same dynamics. Observing listings and transactions of the same apartments, it was concluded that the estimated determinants of listed rents may be considered a proxy of revealed preferences. Similarly, the listings-based hedonic rent indices proved to reflect well the dynamics indicated by market transactions. However, it is not the difference between the height of the listed and transacted rents that proved to be problematic in hedonic modelling, but the inequality of the quality structure of the analysed data types.

Keywords: rental market, transactions, listings, hedonic methods, price index

JEL: C51, C81, R21

* Narodowy Bank Polski, Regional Branch in Poznań; e-mail: michal.hebdzynski@nbp.pl; ORCID: 0000-0003-0354-054X.

The paper presents the personal opinions of the author and does not necessarily reflect the official position of Narodowy Bank Polski.

1. Introduction

A developed housing rental market not only stabilises the real estate market fluctuations, but may also contribute to the overall macroeconomic stability (Czerniak, Rubaszek 2018; Rubaszek, Rubio 2020). Moreover, it is considered an important factor for mobility in the labour market (Łaszek, Augustyniak, Olszewski 2021). However, the long-term rental market in Poland is among the least developed in the EU, as only 4.2% of households residing in Poland rent at market price (Eurostat 2024) and their strongly-prevailing preference for owning instead of renting has been noted (Rubaszek, Czerniak 2017; Bryx et al. 2021). Thus, it may rarely be considered a target tenure, being most often used by students, migrants and young adults, for whom renting is a transitional stage on the way to purchasing an apartment. Nevertheless, Rubaszek and Czerniak (2017) have found that after satisfying certain conditions, apartment rental may be treated in Poland as the more favourable tenure by a significant share of households. In this regard, in order to stimulate the much-needed growth of the rental market and to reveal its potential, the administrative authorities should take actions directed at increasing the professionalization of the market and striving to reduce the rent level.

To ensure healthy market development, housing policy should aim to meet future anticipated consumer needs and preferences. However, it has been noticed that the rent-setting factors have changed because of the recent market turbulences, as shown by Tomal and Helbich (2022) on the example of Cracow and the influence of the COVID-19 pandemic shock. The massive migrations that were a result of the Russian invasion of Ukraine in February 2022 have added further uncertainty to the knowledge of the market. Thus, studying the preferences of market participants and exploring methods of their timely monitoring is of utmost interest to state entities, institutional investors, developers, and individuals.

However, measuring consumer preferences is problematic, as these are non-observable. Multiple approaches have been developed to capture their current state. First, one can measure the stated preferences, as described by Timmermans, Molin and van Noortwijk (1994), and Brown (2003). This approach assumes that consumers know well what housing characteristics they would account for in the purchasing/renting process and what their relative importance is. This kind of analysis is most often based on survey studies, conjoint analysis, or hierarchical models. Although flexible and able to give direct, ready-to-use answers, the method is problematic to validate or reproduce. Moreover, it is sensitive to the structure of the questions asked, the measurement unit and the sample selection. Finally, because of the high cost of surveying, achieving representative results is often impossible.

The second approach is to study revealed preferences. According to Paul Samuelson's revealed preference theory (1938, 1948), the purchasing behaviours of consumers reveal the utility that they assign to goods. It is based on the assumption that housing purchases are utility-maximizing. Knowing the transacted prices of housing with their particular characteristics one can estimate the marginal prices that consumers paid for each of them. Then, an average price of characteristic would act as a proxy of a revealed preference. The decomposition is most often conducted using a hedonic model (Lancaster 1966; Rosen 1974), where the coefficients of the obtained models represent numerically the revealed preferences. The main disadvantage of the approach lies in the fact that a consumer who wants to purchase/rent an apartment has to choose from the current housing supply, which does not necessarily include the structures preferred by him/her (Boyle 2003). Among the advantages of the approach, one should mention its applicability and replicability. Counter to the stated preferences analysis, it does not require a costly surveying process. However, the specific type of cross-sectional data of individual observations of apartments sold/rented on a given market in a given time, together with their possibly wide range of characteristics, is needed.

The relations between stated housing preferences and the revealed ones have been studied by Hasanzadeh, Kyttä and Brown (2019), and Vasanen (2012), who proved their consistency. Moreover, Earnhart (2002) argued that stated and revealed preferences are in line only in the case of some apartment characteristics. What is most promising, he found that combining the revealed and stated information leads to the best understanding of the phenomena that drive the housing decisions. However, because of the long process of designing the stated preference study and surveying, this approach would be of little help in tracking market changes amid economic shocks. On the contrary, the revealed preferences studies can be provided for practically any time interval, which indicates their high usability for timely market analysis. Furthermore, the goals that guide the process of preparation of a reliable model of revealed consumer preferences are compliant with the requirements of the model aimed at tracking price movements of the market. Hence, based on the micro-level models, researchers are constructing hedonic price indices, which guarantee that the quality is held fixed when measuring changes in prices between two periods. It is also the approach suggested by international organizations (European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank 2013).

The factor that limits the possibility of studying the revealed preferences of consumers and price movements on the rental market is the sparsity of the publicly available datasets, which is common in European countries. For this purpose, the information on micro-level transactions would constitute the most reliable data type for modelling. This situation may be encountered in Poland, where a high share of the market transactions is concluded without an intermediary of housing brokers. Thus, the transactional data gathered in the confidential, law-regulated process by Narodowy Bank Polski are scarce. Moreover, the ones gathered are obtained with a significant time lag. The problematic access to transactional data forces the usage of alternative sources, out of which housing listings are the most popular. They are easily accessible with the use of web-scraping algorithms and, most often, are rich in information about housing characteristics. Moreover, efficient methods of extracting additional information from apartment descriptions (Nowak, Smith 2017; Hebdzyński 2023) or photos (Poursaeed, Matera, Belongie 2018) have recently emerged. However, one should also be aware that listings represent only the supply side of the housing market and cannot always be regarded as representative (Beręsewicz 2015; Nasreen, Ruming 2022).

This research has aimed to answer three research questions testing the utility of listings data to study revealed preferences and price movements in the housing rental market. First, it has been asked whether, because of the imperfect nature of listings data, one should refer to the fruits of the hedonic models as supply-side rent-setting factors (which will be little informative) or whether they may be treated as a proxy of consumers' revealed preferences in the Samuelson's spirit. It has been studied whether the results of hedonic models obtained based on listings data are in line with those obtained on transactional data. Secondly, it has been checked whether the hedonic price indices obtained based on listings and transactional data have the same dynamics. Finally, recently, much attention has been placed on developing methods of hedonic modelling that have advantageous statistical properties. Nevertheless, they have been more data-demanding as well as technically sophisticated. Thus, this study has asked whether the improvement of statistical quality translates into a change in the course of a hedonic price index. In particular, it has been checked how sensitive the hedonic models' results are to the changing composition of explanatory variables and the method of analysis.

The above research questions have been studied for Q4 2020–Q2 2023 using the dataset of 197 transactions and 9,234 listings of apartments for rent located in multi-family buildings in Poznań, the capital of the Greater Poland region. With over 540,000 citizens,[1] it is the fifth biggest city in Poland, which serves as a business, academic and tourist centre. Based on those, the Ordinary Least Squares (OLS) and Quantile Regression (QR) (Koenker, Bassett 1978) hedonic models have been constructed.

The study contributes to the literature in three ways. First, it adds to the revealed preference theory (Samuelson 1938, 1948) and develops the understanding of the results of hedonic methods applied to the specificity of the housing rental market. In this regard, it validates the recent studies of Tomal and Helbich (2022, 2023) on the micro-level rent-determining factors, which have been obtained based on listings data. Additionally, the paper extends the studies by Shimizu, Nishimura and Watanabe (2016), and Kolbe et al. (2021), who discussed the regression coefficients obtained using a hedonic approach based on transactional and listings data of the housing sales market. In this context, it introduces the topic to a segment of the residential real estate market, which is much more data-restricted.

Secondly, it is the first study that compares hedonic indices of transacted and listed rents gathered for the same apartments in the different stages of the renting process. It allows us to assess what part of the difference in the indicated course of the hedonic rent index should be attributed to the difference between the listed and the transacted price and how much stems from the quality differences between the datasets used. The research adds to Micallef (2022), who compared the movement of transacted and listed rents, and adds validity to the results by Trojanek and Gluszak (2022) and Trojanek et al. (2021), who studied the listings-based rent indices in Polish cities. The obtained results contribute to the discussion whether it is possible to use the listings-based indices as a sufficient proxy of transactional indices (Ahlfeldt, Heblich, Seidel 2023; Ardila, Ahmed, Sornette 2021; Lyons 2019; Shimizu, Nishimura, Watanabe 2016; Wang, Li, Wu 2020) or whether they should be treated as a source of supplementary information (Knight, Sirmans, Turnbull 1994, 1998; Kolbe et al. 2021). The issue has so far only been studied in the housing sales market.

Finally, the study extends Diewert and Shimizu (2022), who found that there are some housing characteristics whose inclusion in the hedonic model is crucial and adding other explanatory variables would have a minimal impact on the course of the obtained hedonic index. This study empirically verifies the adequacy of this finding for the housing rental market and adds to the discussion by Hill and Trojanek (2022) and Micallef (2022) on the differences between the outcomes of different modelling approaches used to obtain price indices.

The structure of the paper is as follows. Section 2 reviews the literature on the use of information from listings to study rent-setting factors and to construct hedonic rent indices. Section 3 presents the methodological approach chosen to answer the research questions and the data used. Section 4 shows the findings, and Section 5 discusses the results. Finally, Section 6 concludes, describing the study limitations and outlining the field for further research.

[1] Statistical Office in Poznań, https://poznan.stat.gov.pl/en/poznan/.

2. Literature review

To date, in hedonic studies of the housing rental market, researchers have refrained from calling the coefficients of hedonic models "revealed preferences", presumably because of imperfections of the used data. As long as the supply-side listings data have been analysed, they have most often decided to use phrases with a less direct meaning. Li, Wei and Wu (2019), and Zhang, Wang and Lu (2019) used the phrase with a rather econometric, technical interpretation – "determinant of housing rent". Similarly, Efthymiou and Antoniou (2013) wrote about "factors that determine rents", Löchl and Axhausen (2010) referred to "coefficients", and Brunauer et al. (2010) considered the "effect of" multiple housing characteristics on rents. Among studies that go a bit further are Crespo and Grêt-Regamey (2013), and Tomal (2020), who wrote about "tenants' willingness-to-pay". Finally, Tomal and Helbich (2022) referred to the results of hedonic models primarily as "marginal prices of characteristics" but sometimes called them "preferences".

However, the authors who analysed rental transactions also did not always use direct wording to interpret the results more meaningfully. In this context, some authors used phrases connected solely with the statistical relations – McCord et al. (2014) studied "rental price determinants", and Micallef (2022) referred to the "implied elasticity". As for the studies, which interpreted the obtained estimates in a way similar to the assumptions of revealed preferences, Sirmans, Sirmans and Benjamin (1989) provided "the value of amenities and services", and Baranzini and Ramirez (2005) called the hedonic coefficients most directly, as "preferences" or "implicit prices".

To decide whether it is appropriate to call the listings-based hedonic coefficients revealed preferences, the listings- and transactions-based results should be compared. To our knowledge, this kind of research has not yet been provided for the rental market. For the housing sales market, the number of empirical studies has also been limited and has pointed at inconsistent conclusions. Shimizu, Nishimura and Watanabe (2016) conducted a micro-level research of Tokyo house prices at four stages of the sales process, starting from the initial asking price and ending with the final transaction price reported by the buyer. The models were obtained using quantile regression, and based on separate datasets, the authors did not find large differences between the coefficients. The discrepancies of estimates for floor space, age of the building and two commuting-related neighbourhood variables were no bigger than 20%. Secondly, Kolbe et al. (2021) analysed the willingness-to-pay functions constructed separately for each housing feature, but dependent on all estimated coefficients. Counter to Shimizu, Nishimura and Watanabe (2016), the authors argued that based on the model explaining house prices in Berlin, the functions for the *age* variable differed substantially. At the same time, functions for *area* could be considered similar but not identical.

The methodological approach to constructing housing price indices (HPI) has been well-established (European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank 2013; Trojanek 2018; Widłak, Tomczyk 2010) and tested for its performance (Hill et al. 2022; Hill, Trojanek 2022). However, the literature concerning hedonic rent indices (HRI) constructed on transactional data has been scarce. To our knowledge, it has been limited to one paper by Micallef (2022), who analysed rent movements across the regions of Malta. The author used OLS to show that even with a small set of explanatory variables, it is possible to obtain highly interpretable results. As for the listings-based studies, they have been conducted on the Polish rental market by Trojanek et al. (2021), and Trojanek and Gluszak (2022). In both papers, the authors utilised the quantile regression

(QR) method, which mitigates the impact of outliers and enables a calculation of the price indices for different quantiles of the conditional distribution of the dependent variable. It allows to deepen the interpretation of the results, especially in times of considerable market changes.

In order to decide on the possibility of using listings data as a proxy of transactional data for the construction of HRIs of housing, one should ideally compare information that represents:

I. The targeted type of market – housing rental market

II. The same geographical location

III. The same time range

IV. The same individual apartments but at different stages of the selling/renting process.

To our best knowledge, no study has so far met all criteria. Micallef (2022) satisfied three of them (I, II, III) to prove that HRIs of listings and transactions are highly correlated in the market of Malta. However, he argued that to understand the indices' relation better, their timing differences and co-movements in the business cycle, it is necessary to use a longer time series. Shimizu, Nishimura and Watanabe (2016) (who satisfied conditions II, III, and IV) studied distributions of house prices in Tokyo collected at subsequent stages of the buying process from independent sources. They found the price distributions unequal and attributed the differences to the quality structure of datasets. To focus on the pure differences between price distributions, they proposed two approaches. The first is to use only observations that relate to the same apartments, for which information from both the listing and transaction phases is available. Secondly, one can conduct the quality adjustment, as proposed by Machado and Mata (2005), based on the quantile hedonic regression. After utilising both approaches, they detected only minor differences between the distributions of listed and transacted prices. Although they concluded that list prices may be utilised to construct HPIs, they argued that in order to do so, quality adjustment is needed. Finally, Knight, Sirmans and Turnbull (1994) (who satisfied conditions II, III, and IV) in the study of Baton Rouge (Louisiana, the USA) found that even though list prices prove to Granger-cause sales prices, they are least informative at peaks and troughs of the business cycle, i.e. in times when timely, precise indices are most needed. Nevertheless, they have proven that list HPIs may be successfully used to predict future housing prices.

Other researchers have satisfied only two conditions (II and III). Ahlfeldt, Heblich and Seidel (2023) argued that because of the low accessibility of transactional data in Germany, the listings-based indices may be used as a decent indicator of market trends. They found listings- and transaction-based indices to be positively correlated and that in the years 2007–2016, the indicated price trends were very similar. Anenberg and Laufer (2017) (based on data from the biggest cities of the USA) argued that listings-based indices are accurate, leading indicators of the level of transaction prices. Similarly, for the Irish market, using the Granger causality test, Lyons (2019) found that even during market turmoil, list price indices can be used as leading indicators of the state of the sales market. On the contrary, Ardila, Ahmed and Sornette (2021) found that in Switzerland, the listed and transacted prices do not Granger-cause each other. However, for various market segments, they have proven to be co-integrated. Thus, given the low availability of transactional data in Switzerland, listings may be considered their suitable substitutes.

Finally, the minimum requirements of the hedonic models have been outlined. Diewert and Shimizu (2022) have pointed out that the most critical information to construct property price indexes is the apartment's floor area, the age of the building and its geolocation characteristics. Then, although adding other explanatory variables to the model would increase its precision, the effect on the index would be minimal. Similarly, Micallef (2022) studied the effects of taking different approaches to including apartments' geolocation at two levels of detail and did not detect any considerable differences. Finally, Hill and Trojanek (2022) provided a comprehensive review of the methods of construction of HPIs to show that as long as the hedonic methods are used, the results are relatively robust to the choice of the index variant.

3. Data and methodology

3.1. Data

Listings

The listings dataset has consisted of 9,186 observations of apartments listed for long-term rent, located in multi-family buildings in Poznań (Poland), posted online in Otodom.pl apartment listing platform from November 2020 to May 2023 (source: OLX Group). The observations have been grouped according to the quarter of listing, where quarters have been understood as three-month periods ending in February (Q1), May (Q2), August (Q3) and November (Q4). If the same listing reappeared in the same or adjacent periods, only the last observation was retained in the dataset. This aimed at ensuring that the analysed listed rents would be as close as possible to the transacted rents.

Transactions (paired with listings)

The transactional database consisted of 197 observations of private long-term rental transactions of apartments located in multi-family buildings in Poznań from Q1 2021 to Q2 2023 (source: BaRN, NBP 2023). Only those observations were selected for which it was possible to pair the transaction with the corresponding listing (in particular, the observations of apartments located on the same street and the same floor of the building, with the same floor area and the number of rooms were considered observations of the same apartments). It was assumed that the listed price is always equal to or higher than the transacted price, as in Horowitz (1992). Because of the popularity of the listing platform and the specificity of the possessed transactional data (sourced mainly from real estate agents who are dominant entities that list apartments via Otodom.pl), it was possible to pair most of the gathered transactions. Finally, the information from both sources was merged. Thus, apart from the complete set of apartment characteristics, each entry included two prices – listed and transacted, and two dates – quarter of listing and quarter of transaction.

Based on Figure 1, the datasets may be generally considered balanced. However, for Q2 and Q3 2021, the shares of observations in the transactional dataset were considerably higher than in the dataset of listings. It should be attributed to the temporarily higher activity of some particular real estate agents who provided the data on transactions, rather than to the market situation. The low share of listings in Q4 2020 resulted from the fact that the observations were available only for one month of the quarter.

3.2. Variables

All the variables used in the study are presented in Table 1. The originally prepared variable *ROOMS* (indicating the number of rooms in the apartment) shows a high correlation with the *AREA* variable, which is undesirable in econometric modelling. Thus, the variable is transformed into *ROOM_INT* (room intensity). For calculation of the distance to the city centre (*DIST_CC*), if the exact address is not specified in the listing, the address is set to the middle of the declared street of location. Apartments located on streets longer than 2 km are excluded from the analysis.

3.3. Hedonic methods used

A hedonic model with logged dependent variable and dummies indicating the period of rent observation may be written as (Hill, Trojanek 2022; Tomczyk, Widłak 2010):

$$\ln R = \beta_0 + \sum_{j=1}^{J} \beta_j C_j + \sum_{i=2}^{I} \gamma_i D_i + \varepsilon \tag{1}$$

where:

R – the rent for an apartment in Polish złoty (PLN),
C_j – a matrix of independent apartments' characteristics,
D_i – time dummy variables,
β_j – estimated prices of characteristics,
γ_i – the vector of coefficients of time dummy variables that reflect change of prices.

Then, the hedonic rent index for period i (with period l as a base) may be obtained by exponentiation of the estimated γ coefficients:

$$\frac{R_i}{R_l} = \exp(\gamma_i) \tag{2}$$

In the Ordinary Least Squares (OLS), coefficients are estimated to predict the values of rents, for which the sum of squared residuals would be minimal. This method, in its basic form, is not robust to often appearing heteroscedasticity and outliers. Although the former issue may be approached by using heteroscedasticity robust variance estimator (White 1980) (all the OLS models presented in this paper would be constructed this way), the latter issue is hard to be avoided using OLS. As a solution, one can use the Quantile Regression (QR) method (Koenker, Bassett 1978), which allows to gain control over outliers and the problem of heteroscedasticity because no assumption about the distribution of residuals is made (Waldmann 2018; Widłak, Nehrebecka 2011). Moreover, it enables to model any quantile of the conditional distribution of the dependent variable. However, it provides results based on optimising algorithms, which makes it numerically demanding, and sometimes, it may not be possible to achieve model convergence for all quantiles. Finally, there may be a problem with estimating confidence intervals of QR parameters, especially with small sample sizes. Although solutions to this issue have been established (Tarr 2012), they make modelling complicated and more difficult to interpret than the results of OLS.

3.4. Analytical steps

Comparison of transactions- and listings-based models

1. Constructing a transactions-based OLS model with all explanatory variables mentioned in Table 1 (N = 197). The natural logarithm of transacted rent is selected as the dependent variable, and the transaction date is used to construct time dummies. Because of the small number of observations, the simple approach to account for the geolocation of observations is used (to retain a possibly high number of degrees of freedom of the model). It takes the form of the inclusion of an explanatory variable indicating the distance from an apartment to the city centre (*DIST_CC*). Next, the least statistically significant variable is removed, and the model is re-estimated until no insignificant variable remains (for $\alpha = 0.1$). The aim is to compare the coefficients only for those variables for which the coefficient is estimated with a certain precision. Finally, the *MOD_1* model is achieved.

2. Constructing *MOD_2* – OLS model, which is equivalent to *MOD_1* (as for its composition), but the natural logarithm of the listed rent is selected as the explanatory variable, and the date of listing is used to construct time dummies.

3. Constructing *MOD_3_DIST* – OLS model, which is equivalent to *MOD_2*, but is based on all listings (N = 9186).

4. Comparison of coefficients of explanatory variables achieved in *MOD_1*, *MOD_2* and *MOD_3_DIST* that have been constructed using the same methodological approach.

5. Comparison of hedonic indices obtained based on *MOD_1*, *MOD_2* and *MOD_3_DIST*.

Searching for the listings-based QR model that fits best to the results of transactions--based model

This phase aimed to find whether transactions-based OLS model and listings-based QR models represent the same price-related market segment. The decision was taken based on the similarity of the variables' coefficients and HRIs.

1. Constructing *MOD_4_Q* – QR models for selected conditional quantiles (Q) of distribution of the dependent variable, based on all listings (N = 9186).

2. For each *MOD_4_Q* model, calculating the absolute percentage deviation of the obtained estimates from the estimates of *MOD_1*. The differences are calculated separately for housing characteristics and the quarter-quarter dynamics of HRIs.

Searching for the best way to include geolocation in non-spatial hedonic models

1. Constructing *MOD_3_SUB* – OLS model as in *MOD_3_DIST*, but instead of the distance variable, dummies for subdistricts (*SUB_N*) of Poznań are included. Next, *MOD_3_ALL* is constructed, incorporating both approaches at once.

2. Comparison of the three models using Bayesian information criterion – BIC (Schwarz 1978) – lower BIC indicates higher model quality. Then, VIF values are calculated – to avoid collinearity of explanatory variables, values above five should raise concerns.

Verifying sensitivity of the obtained results of hedonic rent indices

1. Constructing the best possible listings-based median QR hedonic model *MOD_5* for all listings ($N = 9186$). First, all of the variables mentioned in Table 1 are used. Next, the least statistically significant variable (for $\alpha = 0.1$) is removed, and the model is re-estimated until no insignificant variable remains. The model is constructed using three forms of inclusion of geolocation – *MOD_5_DIST* (including distance variable – *DIST_CC*), *MOD_5_SUB* (including dummies for subdistricts – *SUB_N*) and *MOD_5_ALL* (including both approaches). The aim is to check how the method of accounting for geolocation in hedonic models affects the course of HRI.

2. Constructing *MOD_5_SIMPLE* model, as in the best *MOD_5* (using the approach to geolocation indicated as best in the previous section) but including only the representation of floor area, age of the building and geolocation (as in Diewert, Shimizu 2022). Constructing *MOD_5_SIMPLE_NO_AREA* model as in *MOD_5_SIMPLE* but without the variable representing floor area and *MOD_5_ONLY_AREA* model, in which area and time dummies are the only explanatory variables. The aim of all models is to check how the choice of explanatory variables affects the course of HRIs.

3. Constructing *MOD_5_OLS* model, as in the best *MOD_5,* but using OLS. The aim of the model is to check how the change in the estimation method affects the course of HRIs.

Checking the course of hedonic rent indices obtained for different quantiles

1. Constructing *MOD_6_Q25*, *MOD_6_Q50*, *MOD_6_Q75* QR models for price-related market segments. They are equivalent to the best *MOD_5* but calculated for the 25^{th}, 50^{th} (median) and 75^{th} percentile of the conditional distribution of the dependent variable. Because of the problem with reaching convergence of the model, some explanatory variables are removed (*TERRACE*, *PARKING_SPACE*, *DISHWASH*).

4. Findings

4.1. Comparison of transactions- and listings-based models

The models presented in Table 2 explain between 68.9% and 78.5% of the variance of rents, even though they are based on a small number of explanatory variables. Others are excluded because of the lack of proven statistical significance (in the *MOD_1* model), which is harder to achieve in the models based on a small number of observations. Thus, the exclusion of some variables should be treated rather as an impossibility to determine precisely the scale of their impact on rent level than as the lack of the meaning for rent determination.

The coefficients obtained in *MOD_1* and *MOD_2* are mostly in line. For no variable the difference exceeds 20% and the average percentage difference equals 7.4%. The differences between coefficients in *MOD_1* and *MOD_3_DIST* are also relatively small but noticeable. For all independent variables, signs agree. As for the magnitude of impact on the dependent variable – for four variables, the differences are no bigger than 20%, while for the distance variable *DIST_CC,* it reaches almost 50%. The mean rents in *MOD_1* and *MOD_2* differ by around 2%, indicating that the final transacted rents were in the period of analysis rarely negotiated. The higher mean rent in *MOD_3_DIST* is probably rooted in the lower share of listings for periods where the general rent level was lower (see Figure 1).

As presented in Figure 2, listings-based *MOD_2* indicates almost the same dynamics of rents as the transactions-based *MOD_1*. For only 1 out of 9 analysed periods, the indicated sign of change is different, and the correlation of the achieved indices is very high. As for the comparison of *MOD_1* and *MOD_3_DIST*, although the two models in 8 out of 9 cases point in the same direction of rent movement, and the correlation of the two indices is very high, one can see some discrepancy in two periods – Q4 2021 and Q4 2022. However, quickly after the periods of discrepancy, both models show again almost the same height of index.

4.2. Searching for the listings-based QR model that fits best to the results of transactions-based model

In the left panel of Figure 3, the average difference between the coefficients obtained in *MOD_4_Q* and *MOD_1* varies across the conditional rent distribution. The lowest values of approx. 19% are achieved for the 55^{th} – 75^{th} percentile of distribution (the value for the 95^{th} percentile is treated as an artefact). Thus, it should be concluded that the coefficients of transactions-based *MOD_*1 are most similar to those obtained while explaining the 55^{th} – 75^{th} percentile of the conditional distribution of rents in the listings-based *MOD_4_Q*.

Although more volatile, the HRI results presented in the right panel of Figure 3 also show that the slightest differences are achieved for the 60^{th} – 80^{th} percentile. Then, it should be interpreted that the rent dynamics indicated by the transactions-based *MOD_1* are most similar to those obtained while explaining the 60^{th} – 80^{th} percentile of the dependent variable in *MOD_4_Q* calculated based on all listings. However, the differences in the rent dynamics are smaller across the entire dependent variable distribution than it is for the rent determinants' coefficients.

4.3. Searching for the best way to include geolocation in non-spatial hedonic models

Table 3 presents the results of three hedonic models built with a different approach to include geolocation. Among the first two models – *MOD_3_DIST* and *MOD_3_SUB* the latter shows a higher R^2 coefficient and lower value of BIC. Thus, including dummies for subdistricts should be considered superior to including distance variable reflecting distance to the city centre. Although *MOD_3_ALL* (that has included both approaches) achieves an even lower (hence better) value of BIC, the highest VIF value among the studied variables reaches the level of 8.7, which may point to the problem of multicollinearity of location variables. Thus, among the three approaches, *MOD_3_SUB* has been considered best in the further analytical steps.

4.4. Verifying sensitivity of the obtained results of hedonic rent indices

According to the results of *MOD_5_SUB* presented in Table 4, two variables – *GARRET* and *GROUND_FLOOR* have been removed from the final version of the model because of the indicated statistical insignificance. Other variables show the expected sign of influence on rents. As presented in Figure 4, HRIs behave in line regardless of the estimation method chosen (OLS or QR) and the approach to account for geolocation (distance to the city centre, dummies reflecting subdistricts). Figure 5 shows that in terms of hedonic model composition, only the HRI that is based on the model not including information about floor area deviates from the best listings-based model (*MOD_5_SUB*).

4.5. Checking the course of HRIs obtained for different quantiles

From Figure 6 it may be implied that the indices behave differently for the selected conditional quantiles of the dependent variable. Since the beginning of the Russian invasion of Ukraine, the cheapest apartments (25th percentile – *MOD_6_Q25*) have proven to increase their rents to a higher degree than the most expensive ones (75th percentile – *MOD_6_Q75*). The difference persisted until the last analysed period.

5. Discussion

The main aim of this research was to test the utility of listings to analyse the revealed preferences of consumers in the housing rental market and to measure market price trends. First, it has been shown that the differences between the coefficients of hedonic models based on transactions and listings are small, provided that the calculations are made on the same group of apartments. Moreover, the rental market in Poznań has proven to be of high liquidity. The final transacted rents were rarely negotiated in the analysis period, as they were, on average, only 2% lower than the listed ones. However, the differences between the coefficients of the transactions-based model and the all-listings-based model were larger, especially for the distance variable. We ensured that the discrepancy was not rooted in the fact that some listed apartments were transacted faster than others because, in the analysis, only the last listing of each apartment has been included. It has also been shown that the problem did not originate from the difference between the height of the listed and transacted rent. Then, it may be hypothesised that the source of difference was in the quality structure of the market reflected in listings. One probable reason is that although online listing platforms represent the majority of the market supply, they are rarely free of charge, thus, they are dominated by listings provided by real estate professionals. The rest is advertised privately, often through social media groups dedicated to, among others, students or migrants. Because of their financial constraints, they are primarily interested in cheaper apartments of low or medium quality. As a result, the online listing platforms may underrepresent the lower segment of the market in relation to the whole market's quality structure. However, based on Quantile Regression, it has been shown that the highest compliance of the coefficients of transactions-based and listings-based models is reached for the 55th – 75th percentile of the conditional distribution of listed rents. This leads to the conclusion that the analysed transactional data represent an even higher market segment than the listings data. This may result from the process of gathering data on the rental market in Poland, which leaves no responsibility on private landlords to report data on rental transactions (other than the height of the transacted rent). Although it may be suspected that the listings data represent the market supply, which is not the ideal representation of the actual market structure, it has been shown that the transactional data gathered in the law-regulated process may be even further from it.

The second part of the study focused on the course of hedonic rent indices. As long as the HRI calculation was conducted on the same groups of apartments, the differences between transactions--based and listings-based HRIs were minimal. Next, the QR analysis indicated that the obtained transactional HRI was closest to the all-listings-based models that represented the 60^{th} – 80^{th} percentile of the conditional distribution of rents. Thus, with regard to the HRI, the transactional data proved to represent (on average) the higher market segment than all listings data. This is consistent with the results obtained for coefficients of apartments' characteristics. However, the transactions-based index revealed two short-term peaks, which the all-listings-based index did not detect. This may be rooted in the opposite nature of listings and transactions, especially in the short term. For instance, amid a negative demand shock, if the demand for low-quality apartments rose, there would be an increased share of low-quality apartments in the periodically collected transactional data and a decreased share of observations of low-quality apartments in the listings data. Then, the transactional models would be better suited to the more turbulent, lower-quality segment of the market. At the same time, the drop in the number of available low-quality apartment listings would result in a worse fit of the listings--based hedonic model to this market segment. Then, the short-term market changes would be reflected in listings-data only if we prepared separate models for quality segments; otherwise, the listings--based HRIs would be expected to flatten the real market dynamics. The pace of adjustment of the market to unilateral shocks that hit one quality segment of the market would then be reflected in the time needed for listings- and transactions-based HRIs to level out. In the example of the Polish rental market, the pace may be estimated at two quarters.

Lastly, the efficiency of the hedonic approaches to analyse the rental market and the sensitivity of their results were studied. It was shown that the non-spatial model utilising dummies for subdistricts performed slightly better than the model accounting for only the distance to the city centre. Nevertheless, both approaches showed comparable results of HRIs. The same applied to choosing between OLS and median QR. The sensitivity of HRIs proved to be elevated solely in the case of excluding the variable representing the floor area of the apartment. Thus, although other explanatory variables tested in this research proved to enhance the quality of hedonic models, their impact on the course of HRI was not found to be decisive.

6. Conclusions

It was revealed that the issue of concern in the process of the rental market analysis should be neither the difference between the listed and transacted rents nor the choice of analytical approach, but the inequality of the quality structure of analysed types of data. However, the Polish rental market is still a black box, as the actual quality structure of the market is not precisely known. Thus, to proceed with the development of knowledge about consumer preferences and market price trends, it is needed to conduct research that would approximate the market structure. Then, it would be possible to weigh observations and obtain a more representative HRI. The other solution would be to construct separate models for quality-related market segments regardless of their share in the whole market supply. The dynamics of the hedonic models constructed for several price-related market segments proved to noticeably differ as these are targeted by different consumers. It may be hypothesised that the differentiation would be even higher for the submarkets of various apartment sizes or apartments in different conditions.

It was shown that although more reliable, the scarce transactional data gathered in Poland in the law-regulated process represent the segments of the market that may be further from the actual market structure than listings data. Then, relying on them for analysing preferences or studying HRIs may introduce a bigger bias. This should be another argument in the discussion on the advantageous characteristics of listing data and the possibility of utilising them as a source of information about the rental market. Moreover, the listings-based rent determinants proved similar to the transaction-based revealed preferences for almost all studied housing characteristics, excluding distance variable. Thus, although one should treat with reserve the calculated coefficients of geolocation variables, the estimates obtained in the listings-based hedonic models of the rental market may be considered a suitable proxy of revealed preferences. As these have not been revealed but are close to the revealed ones, when referring to them, it is suggested to use the phrase "proxied preferences".

Finally, one should be aware of the study limitations. Firstly, the low development of the rental market in Poland results itself in a small volume of market transactions and available data entries. Then, the shortage is further exacerbated by the difficult access to transactional data. As a result, some part of the proven difference between the transactions-based and listings-based models could have been wrongly assigned in this study to the imperfections of listings. In reality, it might have been rooted in the specificity of analysed transactions. The small number of observations forced us to use relatively simple econometric methods to obtain transactions-based results comparable with the ones obtained on the more complete listings-based models. Therefore, the study and its conclusions should be considered introductory to the topic and require further testing.

References

Ahlfeldt G.M., Heblich S., Seidel T. (2023), Micro-geographic property price and rent indices, *Regional Science and Urban Economics*, 98, DOI: 10.1016/j.regsciurbeco.2022.103836.

Anenberg E., Laufer S. (2017), A more timely house price index, *Review of Economics and Statistics*, 99(4), 722–734, DOI: 10.1162/REST_a_00634.

Ardila D., Ahmed A., Sornette D. (2021), Comparing ask and transaction prices in the Swiss housing market, *Quantitative Finance and Economics*, 5(1), 67–93, DOI: 10.3934/qfe.2021004.

Baranzini A., Ramirez J.V. (2005), Paying for quietness: the impact of noise on Geneva rents, *Urban Studies*, 42(4), 633–646, DOI: 10.1080/00420980500060186.

Beręsewicz M. (2015), On the representativeness of internet data sources for the real estate market in Poland, *Austrian Journal of Statistics*, 44(2), 45–57, DOI: 10.17713/ajs.v44i2.79.

Boyle K.J. (2003), Introduction to revealed preference methods, in: P.A. Champ, K.J. Boyle, T.C. Brown (eds), *A Primer on Nonmarket Valuation. The Economics on Non-Market Goods and Resources*, Springer, DOI: 10.1007/978-94-007-0826-6_8.

Brown T.C. (2003), Introduction to stated preference methods, in: P.A. Champ, K.J. Boyle, T.C. Brown (eds), *A Primer on Nonmarket Valuation. The Economics on Non-Market Goods and Resources,* Springer, DOI: 10.1007/978-94-007-0826-6_4.

Brunauer W.A., Lang S., Wechselberger P., Bienert S. (2010), Additive hedonic regression models with spatial scaling factors: an application for rents in Vienna, *Journal of Real Estate Finance and Economics*, 41(4), 390–411, DOI: 10.1007/s11146-009-9177-z.

Bryx M., Sobieraj J., Metelski D., Rudzka I. (2021), Buying vs. renting a home in view of young adults in Poland, *Land*, 10, 1183, DOI: 10.3390/land10111183.

Crespo R., Grêt-Regamey A. (2013), Local hedonic house-price modelling for urban planners: advantages of using local regression techniques, *Environment and Planning B: Planning and Design*, 40(4), 664–682, DOI: 10.1068/b38093.

Czerniak A., Rubaszek M. (2018), The size of the rental market and housing market fluctuations, *Open Economies Review*, 29(2), 261–281, DOI: 10.1007/s11079-017-9452-1.

Diewert W.E., Shimizu C. (2022), Residential property price indexes: spatial coordinates versus neighborhood dummy variables, *Review of Income and Wealth*, 68(3), 770–796, DOI: 10.1111/roiw.12534.

Earnhart D. (2002), Combining revealed and stated data to examine housing decisions using discrete choice analysis, *Journal of Urban Economics*, 51(1), 143–169, DOI: 10.1006/juec.2001.2241.

Efthymiou D., Antoniou C. (2013), How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece, *Transportation Research Part A: Policy and Practice*, 52, 1–22, DOI: 10.1016/j.tra.2013.04.002.

European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank (2013), *Handbook on Residential Property Prices (RPPIs) – 2013 edition*, OECD Publishing, DOI: 10.2785/34007.

Eurostat (2024), *Distribution of population by tenure status, type of household and income group – EU-SILC survey*, DOI: 10.2908/ILC_LVHO02.

Hasanzadeh K., Kyttä M., Brown G. (2019), Beyond housing preferences: urban structure and actualisation of residential area preferences, *Urban Science*, 3(1), 21, DOI: 10.3390/urbansci3010021.

Hebdzyński M. (2023), Quality information gaps in housing listings: Do words mean the same as pictures?, *Journal of Housing and the Built Environment,* 38, 2399–2425, DOI: 10.1007/s10901-023-10043-z.

Hill R.J., Scholz M., Shimizu C., Steurer M. (2022), Rolling-time-dummy house price indexes: window length, linking and options for dealing with low transaction volume, *Journal of Official Statistics*, 38(1), 127–151, DOI: 10.2478/jos-2022-0007.

Hill R.J., Trojanek R. (2022), An evaluation of competing methods for constructing house price indexes: the case of Warsaw, *Land Use Policy*, 120, 106226, DOI: 10.1016/j.landusepol.2022.106226.

Horowitz J.L. (1992), The role of the list price in housing markets: theory and an econometric model, *Journal of Applied Econometrics*, 7(2), 115–129, DOI: 10.1002/jae.3950070202.

Knight J.R., Sirmans C.F., Turnbull G.K. (1994), List price signaling and buyer behavior in the housing market, *The Journal of Real Estate Finance and Economics*, 9(3), 177–192, DOI: 10.1007/BF01099271.

Knight J.R., Sirmans C.F., Turnbull G.K. (1998), List price information in residential appraisal and underwriting, *The Journal of Real Estate Research*, 15(1/2), 59–76, http://www.jstor.org/stable/24886871.

Koenker R., Bassett G. (1978), Regression quantiles, *Econometrica*, 46(1), 33–50, DOI: 10.2307/1913643.

Koenker R., Machado J.A.F. (1999), Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association,* 94(448), 1296–1310.

Kolbe J., Schulz R., Wersing M., Werwatz A. (2021), Real estate listings and their usefulness for hedonic regressions, *Empirical Economics*, 61, 3239–3269, DOI: 10.1007/s00181-020-01992-3.

Lancaster K.J. (1966), A new approach to consumer theory, *Journal of Political Economy*, 74(2), 132–157, https://www.jstor.org/stable/1828835.

Li H., Wei Y.D., Wu Y. (2019), Analyzing the private rental housing market in Shanghai with open data, *Land Use Policy*, 85 (April), 271–284, DOI: 10.1016/j.landusepol.2019.04.004.

Löchl M., Axhausen K.W. (2010), Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Journal of Transport and Land Use*, 3(2), 39–63, DOI: 10.5198/jtlu.v3i2.117.

Lyons R.C. (2019), Can list prices accurately capture housing price trends? Insights from extreme markets conditions, *Finance Research Letters*, 30, 228–232, DOI: 10.1016/j.frl.2018.10.004.

Łaszek J., Augustyniak H., Olszewski K. (2021), The development of the rental market in Poland, in: R. Sobiecki, J. Łaszek, K. Olszewski (eds), *Real Estate at Exposure. New Challenges, Old Problems*, Warsaw School of Economics.

Machado J.A.F., Mata J. (2005), Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, 20(4), 445–465, DOI: 10.1002/jae.788.

McCord M., Davis P.T., Haran M., McIlhatton D., McCord J. (2014), Understanding rental prices in the UK: a comparative application of spatial modelling approaches, *International Journal of Housing Markets and Analysis,* 7(1), 98–128, DOI: 10.1108/IJHMA-09-2012-0043.

Micallef B. (2022), Hedonic indices for registered rents in Malta, in: *The Annual Malta Residential Rental Study. Second Edition*, Ministry for Social and Affordable Accomodation, Housing Authority.

NBP (2023), *Badanie rynku nieruchomości. Instrukcja badania rynku wtórnego*, Narodowy Bank Polski, https://ssl.nbp.pl/publikacje/rynek_nieruchomosci/Instrukcja-badania-rynku-wtornego-2022.pdf.

Nasreen Z., Ruming K. (2022), Struggles and opportunities at the platform interface: tenants' experiences of navigating shared room housing using digital platforms in Sydney, *Journal of Housing and the Built Environment*, 37(3), 1537–1554, DOI: 10.1007/s10901-021-09909-x.

Nowak A., Smith P. (2017), Textual analysis in real estate, *Journal of Applied Econometrics,* 32(4), 896–918, DOI: 10.1002/jae.

Poursaeed O., Matera T., Belongie S. (2018), Vision-based real estate price estimation, *Machine Vision and Applications*, 29(4), 667–676, DOI: 10.1007/s00138-018-0922-2.

Rosen S. (1974), Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of Political Economy*, 82(1), 34–55, DOI: 10.1086/260169.

Rubaszek M., Czerniak A. (2017), Preferencje Polaków dotyczące struktury własnościowej mieszkań: opis wyników ankiety, *Bank i Kredyt,* 48(2), 197–234.

Rubaszek M., Rubio M. (2020), Does the rental housing market stabilize the economy? A micro and macro perspective, *Empirical Economics*, 59(1), 233–257, DOI: 10.1007/s00181-019-01638-z.

Samuelson P. (1938), A note on the pure theory of consumer's behaviour, *Economica*, 5(17), 61–71, DOI: 10.2307/2548836.

Samuelson P. (1948), Consumption theory in terms of revealed preference, *Economica*, 15(60), 243–253, DOI: 10.2307/2549561.

Schwarz G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6(2), 461–464, http://www.jstor.org/stable/2958889.

Shimizu C., Nishimura K.G., Watanabe T. (2016), House prices at different stages of the buying/selling process, *Regional Science and Urban Economics*, 59, 37–53, DOI: 10.1016/j.regsciurbeco.2016.04.001.

Sirmans S., Sirmans C., Benjamin J. (1989), Determining apartment rent: the value of amenities, services and external factors, *Journal of Real Estate Research*, 4(2), 33–43, DOI: 10.1080/10835547.1989.12090581.

Tarr G. (2012), Small sample performance of quantile regression confidence intervals, *Journal of Statistical Computation and Simulation*, 82(1), 81–94, DOI: 10.1080/00949655.2010.527844.

Timmermans H., Molin E., van Noortwijk L. (1994), Housing choice processes: stated versus revealed modelling approaches, *Netherlands Journal of Housing and the Built Environment*, 9(3), 215–227, DOI: 10.1007/BF02496997.

Tomal M. (2020), Modelling housing rents using spatial autoregressive geographically weighted regression: a case study in Cracow, Poland, *ISPRS International Journal of Geo-Information*, 9(6), DOI: 10.3390/ijgi9060346.

Tomal M., Helbich M. (2022), The private rental housing market before and during the COVID-19 pandemic: a submarket analysis in Cracow, Poland, *Environment and Planning B: Urban Analytics and City Science*, 49(6), 1646–1662, DOI: 10.1177/23998083211062907.

Tomal M., Helbich M. (2023), A spatial autoregressive geographically weighted quantile regression to explore housing rent determinants in Amsterdam and Warsaw, *Environment and Planning B: Urban Analytics and City Science*, 50(3), 579–599, DOI: 10.1177/23998083221122790.

Tomczyk E., Widłak M. (2010), Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy, *Bank i Kredyt*, 41(1), 99–128.

Trojanek R. (2018), *Teoretyczne i metodyczne aspekty wyznaczania indeksów cen na rynku mieszkaniowym*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.

Trojanek R., Gluszak M. (2022), Short-run impact of the Ukrainian refugee crisis on the housing market in Poland, *Finance Research Letters*, 50(103236), DOI: 10.1016/j.frl.2022.103236.

Trojanek R., Gluszak M., Tanas J., Hebdzyński M. (2021), The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw, *Critical Housing Analysis*, 8(1), 72–84, DOI: 10.13060/23362839.2021.8.1.524.

Vasanen A. (2012), Beyond stated and revealed preferences: The relationship between residential preferences and housing choices in the urban region of Turku, Finland, *Journal of Housing and the Built Environment*, 27(3), 301–315, DOI: 10.1007/s10901-012-9267-8.

Waldmann E. (2018), Quantile regression: a short story on how and why, *Statistical Modelling*, 18(3–4), 203–218, DOI: 10.1177/1471082X18759142.

Wang X., Li K., Wu J. (2020), House price index based on online listing information: the case of China, *Journal of Housing Economics*, 50 (May), 101715, DOI: 10.1016/j.jhe.2020.101715.

White H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48(4), 817–838, DOI: 10.2307/1912934.

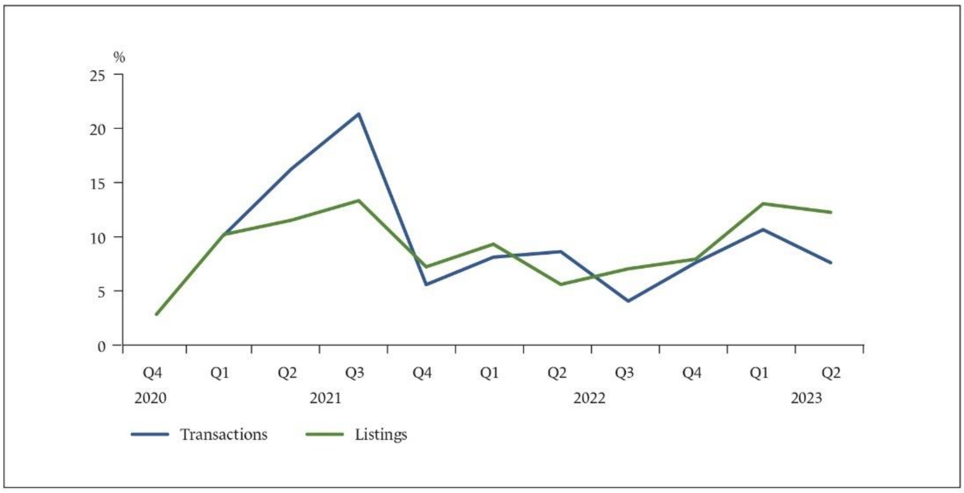
Widłak M., Nehrebecka N. (2011), Wykorzystanie regresji kwantylowej w analizie zróżnicowania cen mieszkań, *Wiadomości Statystyczne*, 5(600).

Widłak M., Tomczyk E. (2010), Measuring price dynamics: evidence from the Warsaw housing market, *Journal of European Real Estate Research*, 3(3), 203–227, DOI: 10.1108/17539261011094722.

Zhang S., Wang L., Lu F. (2019), Exploring housing rent by mixed geographically weighted regression: a case study in Nanjing, *ISPRS International Journal of Geo-Information*, 8(10), DOI: 10.3390/ijgi8100431.

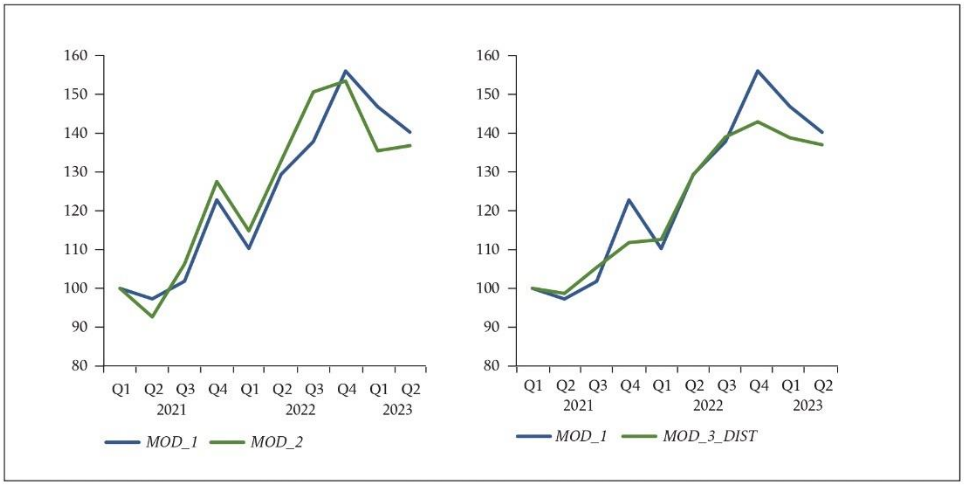
Appendix

Figure 1
Shares of observations from selected periods in analytical datasets



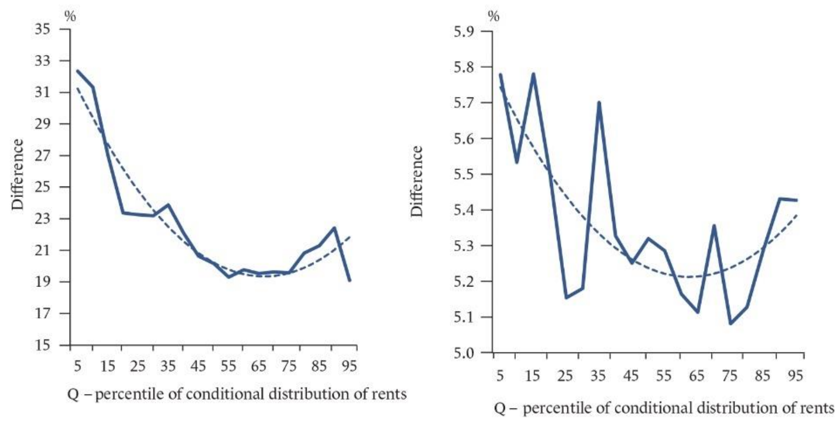
Source: own elaboration based on transactional data from BaRN (NBP 2023) and on listings data from Otodom.pl (OLX Group). Numbers of observations are presented in Table 5.

Figure 2
Hedonic rent indices based on OLS models (Q1 2021 = 100)



Source: own elaboration.

Figure 3
The average percentage difference between the coefficients obtained in *MOD_4_Q* and *MOD_1*

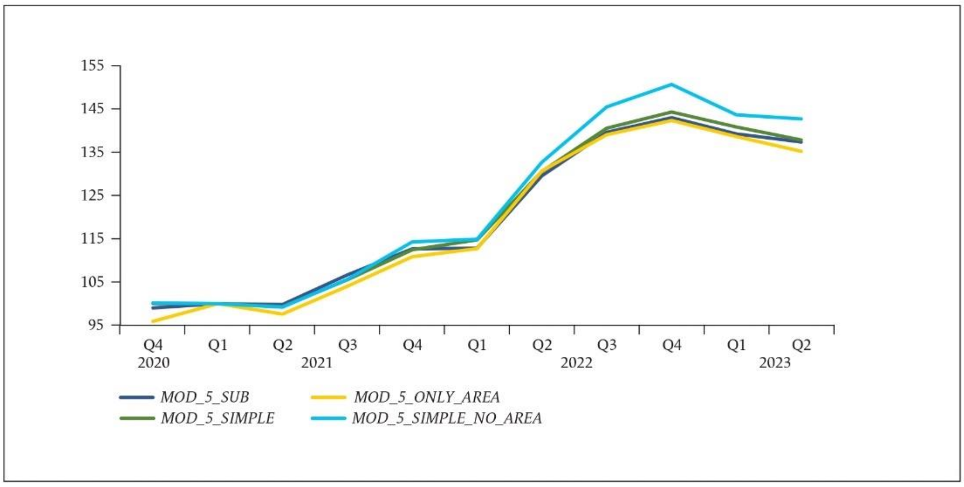


The left panel represents the deviation of coefficients for apartment characteristics. The right panel represents the deviation of the q-q dynamics of HRIs. The dashed line represents the polynomial trend line.
Source: own elaboration.

Figure 4
Sensitivity of hedonic rent indices to the estimation method and the approach to account for geolocation (Q1 2021 = 100)

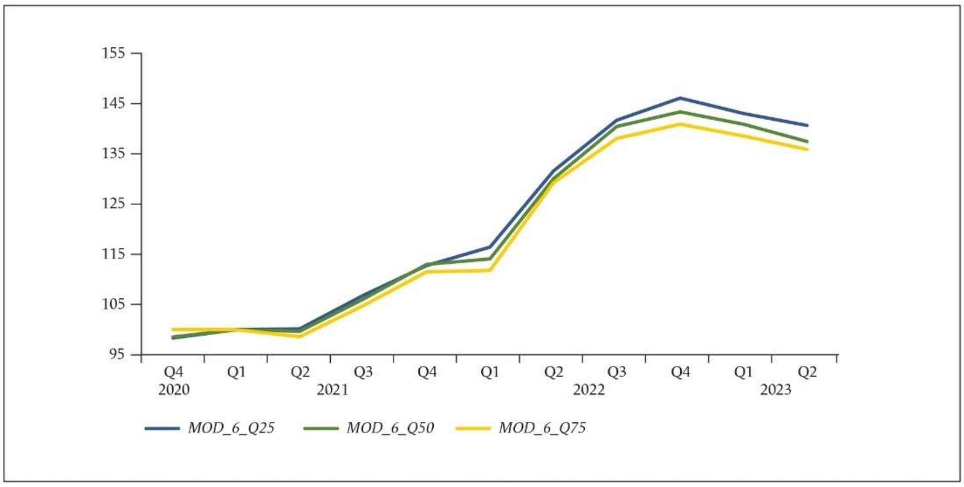
Source: own elaboration.

Figure 5
Sensitivity of hedonic rent indices to the composition of explanatory variables (Q1 2021 = 100)



Source: own elaboration.

Figure 6
Hedonic rent indices for different conditional quantiles of the dependent variable (Q1 2021 = 100)



Full models' results available in Table 9.

Source: own elaboration.

Table 1
Variables used in the research for the dataset of listings (upper values) and transactions (lower values)

Variable	Description	Min	Avg	Max	Share of 1's (in %)
AREA	floor area of the apartment (in the logarithmic form)	15 17.7	47 46.7	150 105.1	
ROOM_INT	room "intensity" – rooms per 1 m^2 of the apartment	0.01 0.01	0.04 0.04	0.08 0.07	
TO_1945	1 – if the building in which the apartment is located was built in 1945 or earlier 0 – otherwise				10.4 18.8
FROM_1946_TO_2004	1 – if the building in which the apartment is located was built between 1946 and 2004 0 – otherwise				26.5 35.5
FROM_2005	1 – if the building in which the apartment is located was built in 2005 or later 0 – otherwise				63.1 45.7
GROUND_FLOOR	1 – if the apartment is located on the ground floor 0 – otherwise				12.1 15.2
GARRET	1 – if the apartment is located on the highest floor of the building 0 – otherwise				16.0 18.8
*TERRACE**	1 – if there is a terrace in the apartment 0 – otherwise				4.5
PARKING_SPACE	1 – if there is access to the designated parking space 0 – otherwise				47.7 24.5
AIR_COND	1 – if there is air conditioning in the apartment 0 – otherwise				5.4 2.5
DISHWASH	1 – if there is a dishwasher in the apartment 0 – otherwise				45.9 15.7
DIST_CC	distance to the city centre [in km]	0.05 0.29	3.12 2.95	10.98 9.05	
*SUB_N**	1 – if the apartment is located in the *N*-th (out of 36) subdistrict of Poznań (division as in BaRN (NBP 2023) 0 – otherwise				

* Variables available only for the dataset of listings.

Source: own elaboration.

Table 2
Simplified results of the transactions- and listings-based OLS models

Variable	*MOD_1*	*MOD_2*	*MOD_3_DIST*
	Paired transactions	**Paired listings**	**All listings**
	Dependent variable: transacted rent	Dependent variable: listed rent	Dependent variable: listed rent
	Mean (PLN) = 1951.3	Mean (PLN) = 1992.2	Mean (PLN) = 2161.4
	N = 197	N = 197	N = 9186
	R^2 = 0.777	R^2 = 0.785	R^2 = 0.689
	Coefficient	**Coefficient**	**Coefficient**
ln_AREA	0.538***	0.545***	0.586***
TO_1945	0.147***	0.138**	0.124***
FROM_2005	0.214***	0.210***	0.198***
PARKING_SPACE	0.096***	0.077**	0.078***
DIST_CC	-0.050***	-0.046***	-0.028***
TIME-DUMMIES	YES	YES	YES
CONSTANT	5.316***	5.327***	5.133***

*** for $P \leq 0.01$; ** for $P \leq 0.05$; * for $P \leq 0.1$.
Full results available in Table 6.

Source: own elaboration.

Table 3
OLS models with different approaches to account for geolocation

	MOD_3_DIST	*MOD_3_SUB*	*MOD_3_ALL*
	Approach to geolocation: distance variable *DIST_CC*	**Approach to geolocation: dummies for subdistricts**	**Approach to geolocation: both approaches**
BIC	-4710.8	-4712.4	-4761.4
R^2	0.69	0.70	0.70
Highest VIF value	1.4 (for *FROM_2005* variable)	1.5 (for *SUB_25* variable)	8.7 (for *DIST_CC* variable)

Full models' results available in Table 7.

Source: own elaboration.

Table 4
Simplified results of the best *MOD_5* model

Variable	*MOD_5_SUB* Median quantile regression Approach to geolocation: dummies for subdistricts Pseudo $R^2 = 0.501$ N = 9186 Coefficient
ln_AREA	0.537***
ROOM_INT	4.075***
TO_1945	0.104***
FROM_2005	0.170***
TERRACE	0.023***
PARKING_SPACE	0.051***
AIR_COND	0.075***
DISHWASH	0.062***
GARDEN	0.012*
TIME-DUMMIES	YES
SUB_N (dummy variables for subdistricts)	YES
CONSTANT	5.194***

Full results of all *MOD_5* models available in Table 8.
Pseudo R^2 (an equivalent of R^2 for QR) has been calculated as described by Koenker and Machado (1999).

Source: own elaboration.

Table 5
Numbers of observations from selected periods in analytical datasets

	Q4 2020	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022	Q2 2022	Q3 2022	Q4 2022	Q1 2023	Q2 2023	Total
Transactions	–	20	32	42	11	16	17	8	15	21	15	197
Listings	260	911	1059	1225	663	856	514	646	729	1198	1125	9186

Source: own elaboration based on transactional data from BaRN (NBP 2023) and listings data from Otodom.pl (OLX Group).

Table 6
Results of the transactions and listings-based OLS models

Variable	*MOD_1*	*MOD_2*	*MOD_3_DIST*
	Paired transactions	Paired listings	All listings
	Coefficient	Coefficient	Coefficient
ln_AREA	0.538***	0.545***	0.586***
TO_1945	0.147***	0.138**	0.124***
FROM_2005	0.214***	0.210***	0.198***
PARKING_SPACE	0.096***	0.077**	0.078***
DIST_CC	-0.050***	-0.046***	-0.028***
Q4 2020		-0.10	-0.01
Q1 2021			
Q2 2021	-0.03	-0.08	-0.01
Q3 2021	0.02	0.06	0.05***
Q4 2021	0.21***	0.24***	0.11***
Q1 2022	0.10	0.14**	0.12***
Q2 2022	0.26***	0.28***	0.26***
Q3 2022	0.32***	0.41***	0.33***
Q4 2022	0.44***	0.43***	0.36***
Q1 2023	0.38***	0.30***	0.33***
Q2 2023	0.34***	0.31***	0.32***
CONSTANT	5.316***	5.327***	5.133***
Number of observations	197	197	9186
Dependent variable	transacted rent	listed rent	listed rent
Mean rent (PLN)	1951.3	1992.2	2161.4
R^2	0.777	0.785	0.689

*** for $P \leq 0.01$; ** for $P \leq 0.05$; * for $P \leq 0.1$.

Source: own elaboration.

Table 7
Results of *MOD_3* models

Variable	MOD_3_DIST	MOD_3_SUB	MOD_3_ALL
	Coefficient	Coefficient	Coefficient
ln_AREA	0.586***	0.584***	0.584***
TO_1945	0.124***	0.109***	0.106***
FROM_2005	0.198***	0.196***	0.197***
PARKING_SPACE	0.078***	0.077***	0.076***
DIST_CC	-0.028***		-0.025***
Q4 2020	-0.01	-0.01	-0.01***
Q1 2021			
Q2 2021	-0.01	-0.01	-0.01***
Q3 2021	0.05***	0.05***	0.05***
Q4 2021	0.11***	0.11***	0.11***
Q1 2022	0.12***	0.12***	0.12***
Q2 2022	0.26***	0.26***	0.26***
Q3 2022	0.33***	0.33***	0.33***
Q4 2022	0.36***	0.36***	0.36***
Q1 2023	0.33***	0.33***	0.33***
Q2 2023	0.32***	0.32***	0.32***
SUB_1		-0.08***	0.07**
SUB_2		-0.14***	-0.05***
SUB_3		-0.15***	-0.05*
SUB_4		-0.23***	-0.07
SUB_5		-0.13***	-0.07***
SUB_6		-0.27***	-0.01
SUB_7		-0.13***	-0.04**
SUB_8		-0.12***	-0.03*
SUB_9		-0.13***	0.01
SUB_10		-0.14***	0.06
SUB_11		-0.08***	0.03
SUB_12		-0.06***	0.00
SUB_13		-0.22***	-0.12***
SUB_14		-0.08***	-0.01
SUB_15		-0.11***	-0.08***
SUB_16		-0.15***	-0.04**
SUB_17		-0.09***	-0.05***
SUB_18		0.00	0.05*
SUB_19			

Table 7, cont'd

Variable	*MOD_3_DIST*	*MOD_3_SUB*	*MOD_3_ALL*
	Coefficient	Coefficient	Coefficient
SUB_20		-0.16***	-0.09***
SUB_21		-0.19***	-0.05*
SUB_22		-0.16***	-0.06
SUB_23		-0.17***	-0.03
SUB_24		-0.16***	-0.08***
SUB_25		-0.08***	-0.04***
SUB_26		-0.14***	-0.08***
SUB_27		-0.09***	0.02
SUB_28		-0.15***	-0.08***
SUB_29		-0.08***	-0.04***
SUB_30		-0.09	0.09
SUB_31		-0.18***	-0.05***
SUB_32		-0.14***	-0.08***
CONSTANT	5.133 ***	5.155***	5.170***
Number of observations	9186	9186	9186
BIC	-4710.8	-4712.4	-4761.4
R^2	0.689	0.698	0.700
Highest VIF value	1.4 (for *FROM_2005* variable)	1.5 (for *SUB_25* variable)	8.7 (for *DIST_CC* variable)

*** for P ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 0.1.
Numbers (*N*) corresponding to subdistricts used in the construction of *SUB_N* dummy variables: 1 – Antoninek-Zieliniec--Kobylepole and Szczepankowo-Spławie-Krzesinki, 2 – Chartowo, 3 – Dębiec, 4 – Fabianowo-Kotowo, 5 – Główna, 6 – Głuszyna, 7 – Górczyn, 8 – Grunwald, 9 – Junikowo, 10 – Krzyżowniki-Smochowice, 11 – Ławica, 12 – Łazarz, 13 – Naramowice, 14 – Ogrody, 15 – Ostrów Tumski-Śródka-Zawady-Komandoria, 16 – Podolany, 17 – Rataje, 18 – Sołacz, 19 – Stare Miasto, 20 – Starołęka-Minikowo-Marlewo, 21 – Strzeszyn, 22 – Świerczewo, 23 – Umultowo, 24 – Warszawskie-Pomet-Maltańskie, 25 – Wilda, 26 – Winiary, 27 – Wola, 28 – Żegrze, 29 – Jeżyce, 30 – Kwiatowe, 31 – Piątkowo, 32 – Winogrady, 33 – Kiekrz, 34 – Krzesiny-Pokrzywno-Garaszewo, 35 – Morasko-Radojewo. *SUB_19* (Stare Miasto) was used as a base variable. For construction of *SUB_1* two subdistricts have been merged because of the small number of observations from the subdistrict Szczepankowo-Spławie-Krzesinki. For *SUB_33*, *SUB_34* and *SUB_35* there was no observation in the analytical dataset.

Source: own elaboration.

Table 8
Results of *MOD_5* models

Variable	*MOD_5 _SUB*	*MOD_5 _DIST*	*MOD_5 _SIMPLE*	*MOD_5 _ONLY_AREA*	*MOD_5_SIMPLE _NO_AREA*	*MOD_5 _OLS*
	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
ln_AREA	0.54***	0.53***	0.57***	0.56***		0.58***
ROOM_INT	4.07***	3.68***				3.62***
TO_1945	0.10***	0.12***	0.11***		0.05***	0.10***
FROM_2005	0.17***	0.17***	0.21***		0.22***	0.18***
PARKING_SPACE	0.05***	0.05***				0.06***
AIR_COND	0.08***	0.08***				0.10***
DISHWASH	0.06***	0.06***				0.06***
GARDEN	0.01*	0.01**				0.02***
TERRACE	0.02***	0.02***				0.04***
DIST_CC		-0.03***				
Q4 2020	-0.01	-0.02	0.00	-0.04***	0.00	-0.01
Q1 2021						
Q2 2021	0.00	0.00	-0.01	-0.02**	-0.01	-0.01
Q3 2021	0.06***	0.06***	0.05***	0.04***	0.05***	0.05***
Q4 2021	0.12***	0.12***	0.12***	0.10***	0.13***	0.12***
Q1 2022	0.12***	0.12***	0.14***	0.12***	0.14***	0.12***
Q2 2022	0.26***	0.26***	0.27***	0.27***	0.28***	0.26***
Q3 2022	0.33***	0.34***	0.34***	0.33***	0.37***	0.33***
Q4 2022	0.36***	0.36***	0.37***	0.35***	0.41***	0.35***
Q1 2023	0.33***	0.33***	0.34***	0.33***	0.36***	0.33***
Q2 2023	0.32***	0.32***	0.32***	0.30***	0.36***	0.31***
SUB_1	-0.09**		-0.07***		-0.02	-0.09***
SUB_2	-0.13***		-0.12***		-0.09***	-0.16***
SUB_3	-0.14***		-0.12***		-0.15***	-0.18***
SUB_4	-0.27***		-0.23***		-0.20*	-0.25***
SUB_5	-0.14***		-0.11***		-0.18***	-0.14***
SUB_6	-0.26***		-0.20***		-0.27**	-0.29***
SUB_7	-0.13***		-0.11***		-0.10***	-0.13***
SUB_8	-0.11***		-0.11***		-0.10***	-0.13***
SUB_9	-0.11***		-0.08***		-0.10***	-0.13***
SUB_10	-0.17***		-0.14***		0.08***	-0.18***
SUB_11	-0.09***		-0.06***		-0.02	-0.10***
SUB_12	-0.08***		-0.07***		-0.05***	-0.06***

Table 8, cont'd

Variable	*MOD_5 _SUB*	*MOD_5 _DIST*	*MOD_5 _SIMPLE*	*MOD_5_ONLY _AREA*	*MOD_5_SIMPLE _NO_AREA*	*MOD_5 _OLS*
	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
SUB_13	-0.20***		-0.18***		-0.18***	-0.23***
SUB_14	-0.08***		-0.06***		-0.07***	-0.10***
SUB_15	-0.07***		-0.09***		-0.11***	-0.10***
SUB_16	-0.16***		-0.13***		-0.15***	-0.17***
SUB_17	-0.08***		-0.07***		-0.06***	-0.10***
SUB_18	-0.03		-0.02		0.00	-0.01
SUB_19						
SUB_20	-0.16***		-0.12***		-0.15***	-0.19***
SUB_21	-0.14***		-0.15***		-0.10***	-0.19***
SUB_22	-0.15*		-0.12***		-0.18***	-0.17***
SUB_23	-0.22***		-0.20***		-0.09***	-0.19***
SUB_24	-0.14***		-0.16***		-0.15**	-0.15***
SUB_25	-0.08***		-0.07***		-0.09***	-0.09***
SUB_26	-0.14***		-0.13***		-0.19***	-0.15***
SUB_27	-0.10***		-0.09***		-0.06**	-0.11***
SUB_28	-0.14***		-0.13***		-0.02	-0.17***
SUB_29	-0.06***		-0.06***		-0.04***	-0.08***
SUB_30	-0.17		-0.14		-0.60	-0.09***
SUB_31	-0.16***		-0.17***		-0.15***	-0.18***
SUB_32	-0.13***		-0.12***		-0.12***	-0.15***
CONSTANT	5.14***	5.15***	5.22***	5.32***	7.34***	5.02***
Number of observations	9186	9186	9186	9186	9186	9186
Pseudo R^2 (for *MOD_5_OLS* – R^2)	0.501	0.493	0.464	0.358	0.237	0.722

*** for P ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 0.1.
The description of numbers (N) corresponding to subdistricts used in the construction of SUB_N dummy variables is presented in Table 7.

Source: own elaboration.

Table 9
Results of *MOD_6* models

Variable	*MOD_6_Q25*	*MOD_6_Q50*	*MOD_6_Q75*
	Coefficient	Coefficient	Coefficient
ln_AREA	0.53***	0.56***	0.60***
ROOM_INT	4.84***	4.38***	3.66***
TO_1945	0.08***	0.11***	0.11***
FROM_2005	0.20***	0.21***	0.20***
AIR_COND	0.08***	0.10***	0.14***
GARDEN	0.05***	0.05***	0.03***
Q4 2020	-0.02	-0.01	0.00
Q1 2021			
Q2 2021	0.00	0.00	-0.01
Q3 2021	0.07***	0.06***	0.05***
Q4 2021	0.12***	0.12***	0.11***
Q1 2022	0.15***	0.13***	0.11***
Q2 2022	0.27***	0.26***	0.26***
Q3 2022	0.35***	0.34***	0.32***
Q4 2022	0.38***	0.36***	0.34***
Q1 2023	0.36***	0.34***	0.33***
Q2 2023	0.34***	0.32***	0.31***
SUB_1	-0.10***	-0.08***	-0.06
SUB_2	-0.13***	-0.12***	-0.16***
SUB_3	-0.16***	-0.16***	-0.17***
SUB_4	-0.20**	-0.29***	-0.22**
SUB_5	-0.11***	-0.14***	-0.15***
SUB_6	-0.23***	-0.26***	-0.32***
SUB_7	-0.10***	-0.12***	-0.14***
SUB_8	-0.12***	-0.11***	-0.13***
SUB_9	-0.08***	-0.10***	-0.14***
SUB_10	-0.09**	-0.19***	-0.19***
SUB_11	-0.05***	-0.08***	-0.11***
SUB_12	-0.09***	-0.07***	-0.08***
SUB_13	-0.20***	-0.20***	-0.23***
SUB_14	-0.08***	-0.09***	-0.13***
SUB_15	-0.08***	-0.09***	-0.09**
SUB_16	-0.11***	-0.14***	-0.19***
SUB_17	-0.09***	-0.08***	-0.10***

Table 9, cont'd

Variable	MOD_6_Q25	MOD_6_Q50	MOD_6_Q75
	Coefficient	Coefficient	Coefficient
SUB_18	-0.03	-0.03	0.02
SUB_19			
SUB_20	-0.16***	-0.16***	-0.17***
SUB_21	-0.19***	-0.16***	-0.18***
SUB_22	-0.13**	-0.15***	-0.17***
SUB_23	-0.20***	-0.24***	-0.30***
SUB_24	-0.12***	-0.13***	-0.20***
SUB_25	-0.08***	-0.08***	-0.09***
SUB_26	-0.15***	-0.12***	-0.13***
SUB_27	-0.08**	-0.09***	-0.12***
SUB_28	-0.13***	-0.14***	-0.18***
SUB_29	-0.07***	-0.06***	-0.09***
SUB_30	-0.11	-0.17**	-0.16***
SUB_31	-0.18***	-0.17***	-0.19***
SUB_32	-0.14***	-0.13***	-0.17***
CONSTANT	5.05***	5.05***	5.06***
Number of observations	9186	9186	9186
Pseudo R^2	0.689	0.698	0.700

*** for $P \leq 0.01$; ** for $P \leq 0.05$; * for $P \leq 0.1$.
The description of numbers (*N*) corresponding to subdistricts used in construction of *SUB_N* dummy variables is presented in Table 7.

Source: own elaboration.

Czynniki cenotwórcze czy ujawnione preferencje? Jak rozumieć wyniki modeli hedonicznych i indeksów hedonicznych stawek najmu mieszkań bazujących na danych ofertowych?

Streszczenie

Rozwinięty rynek najmu mieszkaniowego jest uznawany za czynnik, który przyczynia się do stabilności rynku nieruchomości mieszkaniowych i całej gospodarki. Rozwój metod analitycznych służących do analizy preferencji uczestników rynku powinien mieć więc kluczowe znaczenie dla świata nauki, podmiotów państwowych, inwestorów instytucjonalnych i osób prywatnych. Czynnikiem ograniczającym możliwość badania preferencji konsumentów i zmian cen na rynku najmu jest (powszechna w krajach europejskich) niedostateczna jakość danych gromadzonych na potrzeby analiz rynku. W tym kontekście pożądanym i najbardziej wiarygodnym rodzajem danych do modelowania byłyby informacje o indywidualnych transakcjach wynajmu mieszkań. Tego typu dane mogą zostać poddane dekompozycji hedonicznej nie tylko dla uzyskania numerycznych oszacowań ujawnionych preferencji konsumentów, ale także w celu sporządzenia odpornych na zmiany jakościowe hedonicznych indeksów cenowych. Problematyczny dostęp do transakcyjnych danych dotyczących rynku najmu nieruchomości mieszkaniowych o zadowalającej jakości powoduje jednak, że konieczne jest korzystanie z alternatywnych źródeł danych.

Celem niniejszego badania była więc odpowiedź na pytanie, czy – niezależnie od niedoskonałości danych ofertowych pochodzących z internetowych portali ogłoszeniowych – wyniki bazujących na nich modeli hedonicznych można traktować jako wyznacznik ujawnionych preferencji konsumentów. Starano się także sprawdzić, czy hedoniczne indeksy czynszów najmu uzyskane na podstawie danych ofertowych i transakcyjnych wskazują na tę samą dynamikę zmian cen na rynku. Do odpowiedzi na postawione pytania badawcze wykorzystano zbiór danych ofertowych i unikatowy zbiór danych transakcyjnych dotyczących długoterminowego wynajmu mieszkań zlokalizowanych w budynkach wielorodzinnych na terenie Poznania. Analizie za pomocą klasycznej metody najmniejszych kwadratów oraz regresji kwantylowej poddano obserwacje z okresu IV kwartał 2020 – II kwartał 2023 r.

Na podstawie obserwacji ofert i transakcji dotyczących dokładnie tych samych mieszkań (N = 197) stwierdzono, że oszacowane czynniki czynszotwórcze i hedoniczne indeksy czynszów ofertowych można uznać za dobre przybliżenie ujawnionych preferencji konsumentów i dobrą reprezentację dynamiki czynszów transakcyjnych. Porównanie niezależnych zbiorów danych – transakcyjnych (N = 197) oraz ofertowych (N = 9186) – ukazało jednak większe różnice, w szczególności w krótkim okresie. Co więcej, zauważono, że wyniki uzyskiwane na podstawie modeli hedonicznych wykazują się niewielką wrażliwością na użyty wariant modelowania oraz na skład zmiennych objaśniających wykorzystanych w modelach. Można więc uznać, że ani różnica między wysokością czynszów ofertowych i transakcyjnych, ani wybór szczególnego wariantu modelu nie są głównymi źródłami niepewności w modelowaniu hedonicznym opartym na ofertach. W tym kontekście decydujący może być wpływ zróżnicowania między strukturą jakościową analizowanych danych a prawdziwą strukturą jakościową rynku najmu.

Wprawdzie nieliczne dane transakcyjne gromadzone w Polsce są bardziej wiarygodne, jednak pokazano, że prawdopodobnie jeszcze słabiej odzwierciedlają prawdziwą strukturę rynku najmu niż dane ofertowe. Poleganie na nich podczas analizy ujawnionych preferencji lub zmian cen może więc prowadzić do obciążenia wyników. Jest to kolejny argument w dyskusji na temat bilansu wad i zalet wykorzystania danych ofertowych zamiast danych transakcyjnych jako źródła informacji o rynku najmu.

Należy zdawać sobie sprawę z ograniczeń badania, które polegają głównie na małej liczbie dostępnych i analizowanych danych transakcyjnych. W tym przypadku część różnic pomiędzy modelem opartym na transakcjach a modelem opartym na ofertach mogła zostać błędnie przypisana niedoskonałości ofert. W rzeczywistości różnice te mogły jednak wynikać ze specyfiki analizowanych transakcji. Finalnie niewielka liczba obserwacji wymusiła zastosowanie stosunkowo prostych metod ekonometrycznych w celu uzyskania modeli transakcyjnych porównywalnych z modelami opartymi na danych ofertowych. Dlatego też badanie i wynikające z niego wnioski należy traktować jako wprowadzenie do tematu.

Słowa kluczowe: rynek najmu, transakcje, oferty, metody hedoniczne, indeksy cenowe

Article 2

Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures?, *Journal of Housing and the Built Environment, 38*, 2399–2425. https://doi.org/10.1007/s10901-023-10043-z

Journal of Housing and the Built Environment (2023) 38:2399–2425
https://doi.org/10.1007/s10901-023-10043-z

ARTICLE

Quality information gaps in housing listings: Do words mean the same as pictures?

Michał Hebdzyński[1]

Received: 22 November 2022 / Accepted: 28 April 2023 / Published online: 20 June 2023

Abstract
The incomplete or defective information on the listed housing quality may impair the market clearing processes, may lead to the adverse selection and would have its reflection in the data needed to perform market analyses. To pinpoint the existing information gaps, this paper inspects the quality-related information disclosure strategies of agents who list apartments for sale or rent in online listing platforms. The first part of the empirical study compares the direct declarations of the listed housing quality with the quality indirectly signaled via photos attached to listings and shows that there extists a discrepancy between them. Combined with the high share of observations, for which the full information on quality is not being disclosed it contributes to the existence of informations gaps. The second part of the study answers, whether indirect textual quality signals (descriptions of apartments), processed with the use of Wordscores—a supervised machine learning algorithm are consistent with visual quality signals. It has been documented, that the sales listings' quality signals have agreed in 63–90% of cases, depending on the model's variant. For the rental market, the descriptive quality signals have matched the visual ones in 71–83% of cases. Moreover, it has been shown that among rental listings the consistency has been higher for listings posted by landlords, while for sales listings—by brokers. Finally, it may be argued that to fill the existing information gaps one may utilize the often-unused information conveyed via textual quality-signalling channel.

Keywords Residential real estate · Rental market · Quality · Signalling theory · Supervised machine learning · Textual analysis

1 Introduction

According to Eurostat (2021) a growing share of the EU citizens inhabits apartments located in multi-family buildings, reaching as much as 46.3% in 2020. This market's structure may be considered diversified as buyers and sellers satisfy there not only residential but also investment needs. However, an access to the full micro-level information needed for market understanding and supervision is difficult, if at all possible, thus

Michał Hebdzyński
Michal.Hebdzynski@phd.ue.poznan.pl

1 Department of Microeconomics, Poznan University of Economics and Business, Poznan, Poland

the market may not be considered fully efficient (Cheng et al., 2015, 2020). The problem may be rooted in the information asymmetry (Akerlof, 1970) as the sell side of the housing market transaction (real estate broker or apartment owner) has an information advantage over the buy side (buyer or tenant) concerning the knowledge on the true quality of the offered product.

The importance of quality in the housing context has been studied with the use of hedonic models (Lancaster, 1966; Rosen, 1974) where the quality-related variables have repeatedly proven to be key determinants of apartment prices (de Wit & van der Klaauw, 2013; Olszewski et al., 2017; Trojanek et al., 2021; Trojanek & Głuszak, 2022). Moreover, based on the micro-level hedonic models, researchers have been able to calculate hedonic price indices (invariant to changes in the quality structure of the analytical samples between the periods of analysis). These are recommended by European Commission, Eurostat, Organisation for Economic Co-operation and Development and World Bank (2013) to be computed by state institutions to measure changes in prices of real estate. Finally, the quality has been indicated as the most important attribute that contributes to housing attractiveness (Renigier-Biłozor et al., 2022). However, the approach of consumers to apartment quality has been recently changing, which has been induced by the COVID-19 pandemic (Marona & Tomal, 2020). Therefore, the issue of apartment quality should be considered important from both individual and institutional perspective.

To reduce information asymmetry, sellers would send signals (Spence, 1973, 2002) that reveal the information on the true quality to buyers. In this case, a signal may be defined as a way to communicate the product's observable but alterable characteristics. Recently, the signalling of quality of apartments for sale or rent has been done most often via listings posted in the online platforms. A typical listing contains:

- A title,
- Information about the price or rent,
- A table grouping structural characteristics of the apartment,
- A declaration on the offered apartment quality chosen by its seller from a drop-down list of pre-specified quality labels, which may be referred as *a direct textual signal of quality*,
- A textual description of an apartment—*a descriptive, indirect textual signal of quality* (as this way the quality is signalled by a multitude of expressions used to describe the apartment),
- Photos covering the interior and exterior of an apartment and the building, in which it is located—*an indirect visual signal of quality*.

It has been proven that the understanding of textual signals included in listings is the same for real estate sellers and buyers (Luchtenberg et al., 2019). Grossman (1981) and Milgrom (1981) in their pioneering papers showed that sellers have to provide buyers with the information on the product's quality as its non-disclosure would signal low quality. However, the literature suggests the existence of some configurations of consumer preferences that may incentivise sellers to signal quality not fully (Hotz & Xiao, 2013). Moreover, disclosing more photos to housing listings has been found to have both positive and negative trade-offs (Benefield et al., 2011) and for some segments of the market disclosing less information in listings might be beneficial (Bian et al., 2021). It may be argued that when increasing the information load of the listing, the probability of enclosing the information that would discourage the potential buyer to visit the apartment in person also rises (Goodwin et al., 2014).

It would be possible to identify the true quality of the listed apartments but the cost of the procedure would be high. The ideal quality assessment would demand a huge amount of time needed for an in-person visit to see each apartment and require a specialistic knowledge. Thus, buyers on the housing market are prone to using heuristics and are subject to anchoring effects (Bucchianeri & Minson, 2013; Tversky & Kahneman, 1974). Luchtenberg et al. (2019) examined the importance of pictures and words used as tools to signal the quality of homes listed for sale and concluded that words mean more than pictures in terms of attracting potential customers. In the contrary, Seiler et al. (2012) proved that when searching for home with the use of online listing platforms, buyers view first the apartment photos and only later the quantitative descriptions and sellers' remarks sections. Moreover, Renigier-Biłozor et al. (2022) have proven, that the apartment quality has been responsible for the largest fluctuations of detected emotions among all the apartment attributes represented in apartment photos. Based on that, it may be suspected that both textual and visual signals expressing housing quality may have a role in the anchoring process. Thus, in order to sell (or rent) for higher prices, the apartment owners (or landlords) may be interested in overstating the quality of the listed apartments.

Recently, when search processes in the housing market have been dominated by online listing platforms, the discussed phenomena may have two-dimensional implications. First, the incomplete or supposedly overstated information on housing quality provided by sellers in listings may lead to far-from-realistic anchoring and may create a platform for abuse of ill-informed buyers. Then, because of a significant time cost to arrange an in-person visit to multiple apartments, either a problem of adverse selection may arise or the market clearing processes may be impaired. Moreover, the imperfect knowledge of the buyers, driven by information asymmetry, might have an impact on price levels in the housing sector. It has been pointed by Łaszek et al. (2016) on the example of the Polish primary market. Moreover, Qiu et al. (2020) have proven, that as a result of a weak information, certain socio-economic groups of buyers pay higher prices for similar housing than the better-informed ones.

Secondly, the scarce, incomplete or defective information would have its reflection in the housing market data needed to perform analyses (Renigier-Biłozor et al., 2019). Transactional data of acceptable quality is usually available with a significant time lag, which hinders the possibility for a prompt reaction of administrative and supervisory authorities to a rapidly changing market situation. The weaknesses and challenges of the real estate market in the context of data have been presented on the example of Poland by Brzezicka (2021). Her study highlights that the gathered data is often fragmented because of the shortage of publicly available databases of housing transactions. Additionally, an access to micro data from the housing market is severly restricted, as they are a source of competitive advantage (in the case of private entities) or are covered by statistical confidentiality (in the case of public entities). In this market environment, the importance of the listings-originated data has been recently growing, even though the process of seachring for an apartment with the use of online listing platforms has multiple weaknesses that later affect also the quality of datasets. Among those, Nasreen and Ruming (2022) have found the misrepresentation of property condition and Beręsewicz (2015, 2019) has pointed that the micro-data concerning Polish housing market cannot be fully regarded as representative. However, listings data have shown an ability to provide an up-to-date view of the market situation, hence they may be used as a proxy of transactional data or their supplement (Boeing et al., 2020). What is more, the studies indicate the high accuracy of the listings-based price indices and their beneficial anticipatory nature (Anenberg & Laufer, 2017; Ardila et al., 2021; Wang, 2021). Finally, listings data is easily accessible with the use of web-scraping algorithms.

This paper intends to inspect the quality-related information disclosure strategies of sellers (owners, landlords or brokers) who list individual apartments for sale or rent in the online listing platforms. It aims to answer, what is the scale of non-full information disclosure and to empirically verify the hypothesis that housing sellers tend to overstate their direct declarations of the listed housing quality. Both issues may contribute to the creation of the unwanted information gap. Then, the paper checks, whether the visual signals sent by sellers are consistent with the textual ones, so that the information gap may be decreased by combining the earlier unused textual and visual information. It has been done with the novel use of the supervised machine learning, dictionary-based Wordscores algorithm (Laver et al., 2003). To verify the correctness of the method a unique dataset of housing listings has been prepared. It contains over 7000 Polish-language sales and rental listings of apartments located in Poznań, Poland, gathered between 06/2019 and 03/2021 labelled in terms of the listed apartment quality. The whole analysis has been conducted separately for the two most popular types of housing tenure in Poland—buying an apartment or renting it on a private market for a long-term (contracts for 6 months or more).

The research contributes to the literature in four ways that have implications for the study fields' development. First, it extends the previous studies on the information disclosure in the housing market (Benefield et al. (2011), Bian et al. (2021), Hotz and Xiao (2013)) by documenting the size of information gap as for the sellers' direct textual signals of apartment quality in listings. This kind of signals may be used to filter thousands of listings in the process of searching for an apartment to buy or rent, thus their absence may significantly prolong the market search process and impair the market clearing. Moreover, to date researchers (who construct their hedonic models with the use of micro-level, listings-originated data of housing market transactions) have most often utilized only the observations, for which the information on quality has been provided in the form of direct textual quality signal. The paper warns that this approach may introduce bias to the analyses.

Secondly, to fight the implications of the information gap, the study tests the utility of the Wordscores algorithm to automatically extract signals of apartments' quality from their textual descriptions included in listings. The fruits of the method may be used both to facilitate the processes of housing search and to enhance the analytical properties of datasets. In this manner, the study extends the results of Seo et al. (2020) (who studied the impact of the presence of quality-related information included in listings on house prices) by showing a way to calculate an independent measure of apartment quality.

Thirdly, using the calculated Wordscores, the paper verifies the consistency of textual and visual quality signals sent via sales and rental listings by owners of the apartments and real estate brokers. The study aims to answer whether the visual and textual signals carry the same information load, and therefore can be used interchangeably. This would noticeably ease the analytical processes that are currently resource- and technically-demanding because they utilize the visual data (as e.g. the one proposed by Poursaeed et al. (2018)). In this regard, it contributes to the signalling literature and extends the results of Luchtenberg et al. (2019).

Fourthly, to my best knowledge this is the first study to compare the performance of textual models of sale and rental markets and to inspect the similarities and differences in listing strategies used on them. Both submarkets interact to great extent as their supplies originate in the same housing stock. Thus, they are suspected to be a subject to similar phenomena related with housing quality. The contribution to the development of efficient, comparable tools of their analysis would constitute an additional value added of the research.

The structure of the paper is as follows. Section 2. discusses the ways of perceiving and signalling housing quality together with a review of methods for extracting quality-related signals from apartments' listings. Section 3. presents the data and methods used for measuring the consistency of textual and visual quality signals. In this regard, the steps of the Wordscores algorithm and the data used are introduced. Section 4. describes the empirical results and the performed robustness checks. Finally, Sect. 5. concludes and outlines the scope of further analyses.

2 Literature review

2.1 Perception of housing quality

Researchers consider that the quality of individual housing consists of four main components: apartment building quality, neighborhood quality, social and economic surrounding quality and apartment unit quality (Brkanić, 2017). The former three can be expressed relatively well with objective factors and with the use of information that may be considered *hard information* in the Liberti and Petersen (2019) definition. In this case, the publicly available information may be used e.g. building age, its construction technology or the building's location in relation to public utility facilities. However, the issue of apartment unit quality is more problematic. Some researchers find that it is being reflected by the structural apartment's characteristics, which can be presented objectively (Elsinga & Hoekstra, 2005; James, 2007; Tomal, 2022), e.g. number of rooms, presence of balcony, presence of AC or type of heating. Others focus also on more challenging to quantify information on apartment finishing, condition, design or adaptability (Kain & Quigley, 1970; Kim et al., 2005; Le et al., 2016; Renigier-Biłozor et al., 2022). Obtaining it would require having access to the ultra-precise data for each apartment or would entail personal inspection of apartments by specialized agents. Thus the objective judgements of apartment unit quality need to be substituted with more subjective ones, based on a more general look at the apartments (e.g. on their photos or descriptions). In this regard, the quality should be considered *soft information* (Liberti & Petersen, 2019) and this understanding of quality may be referred as *soft quality*. It means that the knowledge of a person providing the assessment, his/her motivations, and the context of the process are essential factors in understanding the meaning of soft quality assessments.

The official, universal guidance on how the housing quality should be measured is provided rarely. Among exceptions, there are the systems, which have been applied nationally in France (Qualitel, 2021), Switzerland (OFL, 2015) or Japan (Housing Performance Indication System and Long-life quality certification (Fujihira, 2017)). The more common approach is to prepare general guidelines set for a given purpose and it has been taken in Poland, by the National Bank of Poland. As an institution responsible for monitoring housing market phenomena, which may affect the country's financial stability, it gathers (from private and public entities) the micro-level data on transactions on the secondary residential real estate market. To ensure compatibility of data it has released the instruction (NBP, 2022), in which it defines the desired data structure. One part of it specifies how the soft housing quality should be measured. The main attributes of apartment one need to take into consideration are windows, floors, installations, doors, walls, kitchen & bathroom equipment. The high-quality label should be assigned to apartments that are functionally arranged, finished with good quality materials, and with a low degree of exploitation. The medium-quality apartments are those whose technical condition allows moving in without significant additional investments. Lastly, the low-quality ones are those that are eligible for renovation.

2.2 Targeting apartment quality in housing listings

The signals concerning quality in its hard sense are most often included in listings in the structured form of a table with the offered apartment's characteristics. Thus, they are relatively clear and little prone to manipulation. Thinking about soft quality signals sent by sellers, they might be considered twofold—as direct and indirect ones. A direct signal—a declaration of quality selected from a pre-specified drop-down list may be interpreted without any hardship. The bigger problem arises when it comes to indirect signals that are sent in the form of apartment's photos or its textual description. This kind of signals are much more detailed but also harder to interpret, because of their structure. However, they are often the only sources of information on soft quality provided by sellers.

The most precise approach to extracting information on soft quality from indirect signals would be to personally inspect each listing's photos and description and assign qualitative or ordinal labels to each of them. In order to provide as objective assessment as possible, the person providing an evaluation should follow strict rules set by experts in the field and, ideally, be an expert himself. However, the cost of obtaining information this way is very high, as it requires both a qualified workforce and a massive amount of time. For this reason, one should pay attention to the automatic, machine learning (ML) methods of extracting information on housing quality. They may limit the resources needed and reduce the unwanted bias and inconsistency of subjective human assessments. One may divide ML methods into unsupervised and supervised ones (Bonaccorso, 2018). Unsupervised ML models are used mostly for clustering or dimensionality reduction of big data. No labelled data is required to train this kind of models, which however makes it difficult to verify precisely the models' outcomes. In the case of supervised ML methods, which are designed for classification or regression purposes, it is easier to validate the results, but the process of obtaining training sets with labelled data may be time consuming, costly and prone to bias (Algaba et al., 2020).

One approach to automatically extract information on the quality of apartments is to use the ML-supported visual quality assessment methods based on the photos of apartments. Gastaldo et al. (2013) argued that mimicking human's perception of quality might be more effective than trying to understand the sophisticated human's process of assessing quality. Thus, it is better to focus on extracting photos' patterns than on modelling non-linear human decisions. In this case, the unsupervised ML methods may be used to effectively cluster similar images. However, they have huge data requirements and the tendency to overfit data. Finally, the clusters' labelling or grading in terms of quality would be anyway a subject to the expert, subjective decision. Therefore, the semi-supervised ML methods that require a minor share of labelled observations combined with majority of unlabelled ones may be seen as a compromise. In this manner, Poursaeed et al. (2018) estimated the luxury level of apartment photos and the obtained metrics may be considered a representation of soft quality. With crowdsourcing, the researchers trained the dataset by subjectively comparing pairs of photos (representing similar objects) in terms of quality. Finally, one of eight levels of luxury has been assigned to each type of room (kitchen, bathroom, bedroom etc.) in each offered apartment. Although computationally demanding and resource costly, the method has proven effective in reducing the median error in the mass appraisal model.

The other approach entails utilizing the information included in listings' textual descriptions. Algaba et al. (2020) pointed that supplying soft information explored from textual content is becoming a standard addition in economics and finance to conventionally used hard data. In this regard, the measure of sentiment (a latent variable to indicate the semantic orientation of text) has been often used as a parameter in economic modelling. Goodwin et al. (2014) studied the language used in real estate listings and its impact on the selling-price, time-on-market and the probability to be sold. They concluded that the performance of housing listings might be enhanced by a careful preparation of its textual descriptions. To understand better the true meaning of apartments' descriptions Goodwin et al. (2018) surveyed a nationwide sample of the US's respondents to construct the dictionaries of words with positive, negative or neutral meaning in the reality of the housing market. However, Haag et al. (2000) found, that some sellers' remarks in the listings platforms may be classified as hype, because certain strictly positive comments were associated with lower sales prices. Thus, using lexicons of word's sentiment based on the declarations (obtainted through surveys) may lead to false conclusions.

In some textual studies, the housing quality has been targeted indirectly. Shen and Ross (2021) used texts to construct an unsupervised ML algorithm (based on natural language processing) to measure the apartment's *uniqueness*, which may be interpreted twofold—as the unobserved quality of an apartment and as the market power of an apartment in relation to the properties in its neighborhood. The usage of *uniqueness* as a variable in the hedonic model absorbed a significant part of the model's residual variance. Nowak and Smith (2017), Liu et al. (2020) and Nowak et al. (2021) argued that real estate hedonic models that do not include quality-related explanatory variables might not only be incomplete, but also biased. To mitigate the problem, they transformed listings texts into tokens that reflect the presence or absence of each word in each listing. With the LASSO algorithm, they formed static and dynamic dictionaries of words to proxy for the omitted, quality-reflecting variables in hedonic models of housing prices. They proved that the addition of the selected tokens as explanatory variables in the hedonic models contributes to the improvement of their statistical properties.

As for the researchers that targeted the issue of quality directly Seo et al. (2020) studied descriptions of housing to form seven quality-related groups of phrases (in the form of bigrams—pairs of words). Among them, one can name mainly: physical housing quality, housing size, its accessibility or the building condition. The phrases were computationally extracted from the US's Multiple Listing Services platform and manually assigned to the pre-specified groups to form tokens. In the study researchers focused rather on the presence or absence of the specific (only positive) information on quality in listings than on calculating or estimating the quality level of apartments. As a result, the statistically significant impact of the presence of quality-related information on selling prices of housing has been proven.

Even though to my best knowledge the fully supervised ML methods have not been used in housing quality-related research yet, the already developed methods offer a wide range of practical applications. One of those is the Wordscores algorithm (Laver et al., 2003), which has been developed for extracting textual sentiment from political statements. The main idea behind it is to construct a dictionary of words with the assigned scores, based on the already classified texts. Then, the dictionary is being used for automatic reading of virgin, non-classified texts—as a result, each of them is provided with a sentiment score. Contrary to the regression-based techniques, the Wordscores builds on the bag-of-words approach that utilizes the content of whole texts without selection of only statistically significant words. Moreover, the method does not pre-specify the sign of correlation of the used words with the target value. It is important, knowing the results obtained by Haag et al. (2000), based on which it may be inferred that some words considered strictly positive may be linked with a low quality of apartments. Furthermore, the method allows targeting directly the phenomenon represented in the classified texts. It stands in contrast to the methods used in the quality-related studies that focused primarily on reducing errors of hedonic models. Finally, the method has proven to be effective in sentiment analysis, yet simple, compared with the already developed and empirically tested approaches to textual and visual analysis. Therefore, I hypothesize its high utility for extracting information on quality for the needs of the housing market analyses.

3 Methodology

3.1 Consistency of signals: direct textual vs indirect visual signals of quality

To document the scale of information gap and to verify the hypothesis that housing sellers' direct textual signals of quality tend to be overstated the unique dataset of individual listings has been used. It contains 1441 listings of apartments for rent from the period 09/2020–06/2021 and 1567 listings of secondary market apartments for sale from the period 09/2020–06/2021 sourced from the two most popular online housing listing platforms in Poland—www.otodom.pl and www.gratka.pl.

As a first step, one soft quality class has been assigned to each apartment listing—low, medium or high, based exclusively on the attached photos. This measure, obtained according to NBP (2022) guidelines (discussed in Sect. 2.1.), would be later referred as *a visual signal of quality* of apartment. For sales listings, all the observations, which were not possible to be categorized with the three mentioned quality classes were excluded. The apartments labelled as "for renovation" or "for refreshment" have been automatically regarded as low-quality ones and in their case the consistency of signals has not been measured. For rental listings, it has been assumed that an offered apartment should be ready to enter by a tenant without a need for prior renovation. The direct textual signalling of quality was done by sellers by selecting the most appropriate quality label from the following—for sales listings: "high quality", "good quality", "freshly renovated", "to enter", "for renovation", "for refreshment", and for rental listings: "high quality", "good quality", "freshly renovated". Finally, the visual signals of quality have been confronted with the sellers' direct textual signals of quality to verify their consistency.

3.2 Consistency of signals: indirect textual (descriptive) vs indirect visual signals of quality—Wordscores algorithm

The second part of the analysis has aimed at measuring the consistency of indirect textual and indirect visual quality signals. For extracting the indirect textual quality signals from texts (later referred as apartment's *descriptive quality*) the Wordscores algorithm has been used. The analytical dataset has covered 7872 Polish-language listings of secondary market apartments for sale and rent located in Poznań, Poland, gathered quarterly between 06/2019 and 03/2021 from two most popular online listing platforms in Poland. Similarly to the observations discussed in Sect. 3.1., each listing has been labelled with the quality class corresponding with the visual quality signal sent by the seller (later referred as apartment's *visual quality*). Moreover, for 6796 listings it has been indicated whether the listing has been posted by the real estate broker (84% of the tagged observations) or by the owner of the apartment (16% of the tagged observations).

The listings without photos that would allow assigning a visual quality label, without textual descriptions, duplicates, together with 5% of the shortest observations among the remaining ones have been excluded in the preliminary stage to ensure that all the descriptions in the analytical sample are unique and valid. Finally, the analytical sample has been divided into training and test set. In the baseline scenario, the test set has been formed by the observations from the last quarter. The initial structure of the analytical samples has been presented in Table 1.

3.2.1 Building dictionaries

The steps of the Wordscores algorithm have been the same for rental and sales listings, although the calculations have been conducted independently. For clarity, they have been presented on the sample listing's textual description:

PL: "Mam do zaoferowania na wynajem 2-pokojowy apartament z balkonem o powierzchni 40 m2 zlokalizowany w prestiżowej lokalizacji." (ENG: "I am offering for rent a 2-room apartment with a balcony, of [the area] 40 m2, located in a prestigious location.").

1. Numbers, special characters, and words no longer than four characters have been removed from each listing's description. The exclusion of the shortest words has aimed at reducing the noise of the analysis, as in Polish this group consists mainly of pronouns and conjunctions—the words excluded this way should carry no significant quality signal. Finally, the Polish listing platforms impose high word limits, thus the usage of abbreviations in not common.
2. Using the morphological dictionary of the Polish language (Miłkowski, 2016) all words in all listings (training and test sets) have been replaced with their lemmas. The lemmatization process converts inflected grammatical form of a word to its basic form (lemma). In this manner, the words "[he] renovates", "[she] renovated" or "[they are] renovating" would be replaced with their infinitive form "[to] renovate". All the words that have not been found in the morphological dictionary have been excluded from further analysis. Then the sample listing's description would take the following form:

Table 1 The initial structure of the analytical sample: baseline scenario. *Source*: own calculations

	Number of observations	Low visual quality apartments	Medium visual quality apartments	High visual quality apartments	Periods analysed
Rental listings					
Training set	3127	420 (13.4%)	1950 (62.4%)	757 (24.2%)	06/2019–06/2020 (5 quarters)
Test set	917	160 (17.4%)	507 (55.3%)	250 (27.3%)	09/2020 (1 quarter)
Sales listings					
Training set	2993	775 (25.9%)	1339 (44.7%)	879 (29.4%)	12/2019–12/2020 (5 quarters)
Test set	835	253 (30.3%)	272 (32.6%)	310 (37.1%)	03/2021 (1 quarter)

PL: zaoferować / wynajem / pokojowy / apartament / balkon / powierzchnia / zlokalizować / prestiżowy / lokalizacja (ENG: to offer / rent / room / apartment / balcony / area / locate / prestigious / location).

3. The occurrences of each lemma in descriptions of apartments of a given visual quality have been counted:

$$n_j = h_j + m_j + l_j, \tag{1}$$

where n_j is the number of occurrences of the j-th lemma in all training set observations, h_j, m_j and l_j refer to the number of occurrences of the j-th lemma in the descriptions of apartments of respectively high-, medium- and low-visual-quality.

4. Each occurrence of a lemma in the description of a low-visual-quality apartment has been treated as a negative sentiment of the lemma. In the contrary, each occurrence of a lemma in the description of a high-visual-quality apartment has been treated as a positive sentiment of the lemma. The occurrences in the medium-visual-quality apartments' descriptions have been treated as neutral. Then, the *WORD_SIGNAL* score—the quality sentiment of the single (j-th) lemma may be calculated as:

$$WORD\_SIGNAL_j = \frac{h_j * (+1) + m_j * 0 + l_j * (-1)}{n_j} \tag{2}$$

As a result, a list of lemma-*WORD_SIGNAL* pairs has been obtained, which may be referred as *a dictionary*. The base—*DICT_0%* version of the dictionary has covered all the lemmas that appeared at least once in all the training set listings' descriptions. Later, the base version has been restricted to leave only the lemmas, which appeared at least:

$D * 3,127$ times—for the dictionary of the rental listings,
$D * 2,993$ times—for the dictionary of the sales listings.
Finally, eight variants of the dictionary, for:

$$D = \{0\%;0.25\%;0.5\%;1\%;2.5\%;5\%;10\%;20\%\}.$$

labelled as *DICT_D* have been obtained. The higher the D-value, the more restrictive the dictionary has become, hence it has included only the more frequently used lemmas. The neutral *WORD_SIGNAL* score for the rental listings' dictionary (for the lemma that would theoretically appear exactly once in every listing, regardless of the listed apartment quality) equals:

$$WORD\_SIGNAL_{neutral,rental} = \frac{420 * (-1) + 1950 * 0 + 757 * (+1)}{3127} = 0.108 \tag{3}$$

and is dependent on the training set's quality structure. It means that all lemmas with a *WORD_SIGNAL* score close to 0.108 may be treated as neutral, and the more the *WORD_SIGNAL* score diverges from this value, the stronger the positive or negative quality signal represented by the lemma. Table 2 presents some selected lemmas from the *DICT_0%* dictionary for rental listings.

Table 2 Some selected lemmas included in the rental market DICT_0% dictionary. *Source*: own calculations

No	Lemma	English meaning	Number of appearances—low visual quality	Number of appearances—medium visual quality	Number of appearances—high visual quality	Sum of appearances	WORD_SIGNAL score
1	luksusowy	Luxurious	0	19	78	97	0.80
2	zaprojektować	To design	2	33	97	132	0.72
3	niepowtarzalny	Unique	0	11	28	39	0.72
4	architektura	Architecture	0	16	34	50	0.68
5	prestiżowy	Prestigious	3	37	77	117	0.63
6	mieszkanie	Apartment	1376	6850	2756	10,982	0.13
7	najem	Rental	404	1909	789	3102	0.12
8	lodówka	Refrigerator	205	1043	402	1650	0.12
9	korytarz	Corridor	179	891	327	1397	0.11
10	student	Student	57	157	18	232	– 0.17
11	niski	Low	93	237	32	362	– 0.17
12	wykładzina	Floor covering	17	10	10	37	– 0.19
13	meblościanka	Wall unit	10	7	0	17	– 0.59

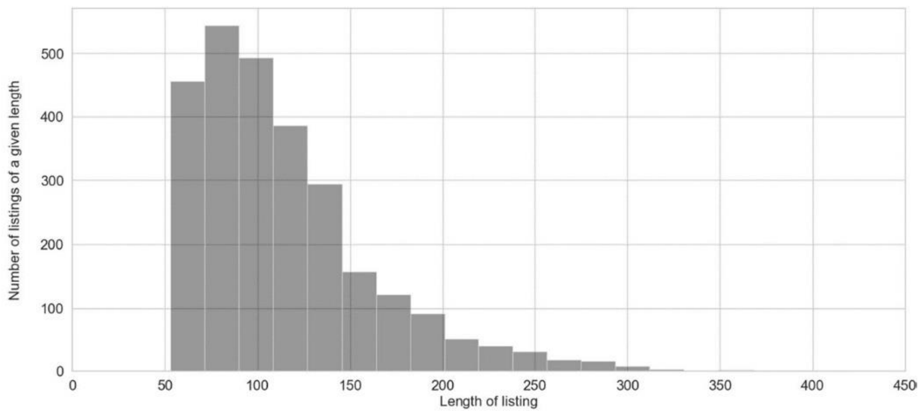


Fig. 1 Histogram of the length of the rental listing (measured as a number of lemmas included). *Source*: own calculations

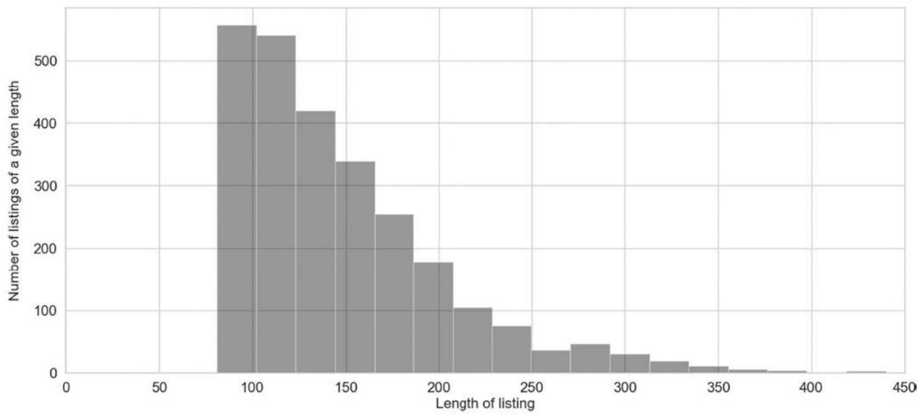


Fig. 2 Histogram of the length of the sales listing (measured as a number of lemmas included). *Source*: own calculations

3.2.2 Calculation of wordscores

The listing length has been defined as an absolute number of lemmas (not necessarily unique) that appeared in the listing and have been included in the *DICT_1%* dictionary. Although all of the listings from the initial training sets have contributed to the dictionaries, it has been assumed that using the shortest listings for the models' calibration and testing may lead to false conclusions. Thus, all observations shorter than an average listing length decreased by its standard deviation have been excluded. In the case of the rental observations, the cut-off point has been set at 53 lemmas and for the sales observations at 81 lemmas. Finally, the training set for the rental market analysis has included 2717 observations (12.0% of low, 62.2% of medium and 25.8% of high visual quality), while for the sales market analysis it has been 2633 observations (25.3% of low, 44.2% of medium and 30.5% of high visual quality). Figures 1 and 2 present histograms of listing lengths for the training sets.

As a next step, the *WORD_SIGNAL* score from a dictionary has been assigned to each lemma in each observation from the training set. Based on the scores of all lemmas included in the corresponding version of the dictionary, the Wordscore—*DESC_SIGNAL* has been calculated to represent the strength of the listing's descriptive quality signal. Six variants related to the measures of *WORD_SIGNAL* scores' distribution within individual listing have been chosen to answer, which of them signals best the apartment's quality. Thus, each *DESC_SIGNAL* score has been calculated as 1st decile (labelled as *DE1*), 1st quartile (*Q1*), median (*MED*), average (*AVG*), 3rd quartile (*Q3*) and 9th decile (*DE9*) of all *WORD_SIGNAL* scores of lemmas included in the observation. Finally, the *DESC_SIGNAL* score for each listing from the training set has been calculated in 48 variants—for all combinations of 8 dictionary variants and 6 distributional variants. They reflect different approaches to measure the overall quality signalled via whole descriptions of apartments. The calculation's scheme has been presented in Fig. 3.

All observations from the training set for which the $DESC\_SIGNAL\_TR_{i,d,v}$ has been calculated according to the rules of the certain variant would be jointly referred as models, labelled as $DESC\_SIGNAL\_TR\_MODEL_{d,v}$. *TR* refers to the training set, *i* indexes the training set observations, *d* refers to the variant of the dictionary and *v* refers to the distributional variant of the calculation.

3.2.3 Models' calibration phase

The models' calibration (based on the training set observations) has aimed at obtaining the threshold values γ and δ needed to change the scale of the obtained measures of descriptive quality from continuous to ordinal (low, medium and high quality). The threshold values would be later used in the model's testing and validation phase. The phase has proceeded as follows:

$$\begin{cases} if\ DESC\_SIGNAL\_TR_{i,d,v} > \gamma_{d,v} \Rightarrow \hat{q}_{i,d,v} = high \\ if\ DESC\_SIGNAL\_TR_{i,d,v} < \delta_{d,v} \Rightarrow \hat{q}_{i,d,v} = low \\ if\ \delta_{d,v} \leq DESC\_SIGNAL\_TR_{i,d,v} \leq \gamma_{d,v} \Rightarrow \hat{q}_{i,d,v} = medium \end{cases}, \quad (4)$$

$$\begin{cases} if\ \hat{q}_{i,d,v} = q_i \Rightarrow A_{i,d,v} = 1 \\ if\ \hat{q}_{i,d,v} \neq q \Rightarrow A_{i,d,v} = 0 \end{cases}, \quad (5)$$

where $q_{i,d}$ is the visual quality of the listed apartment, $\hat{q}_{i,d,v}$ is the model's measure of the apartment's descriptive quality in the ordinal form and $A_{i,d,v}$ is the test parameter checking, whether the descriptive quality matches the visual quality of the apartment. The following

ENG	to offer	rent	room	apartment	balcony	area	locate	prestigious	location		
PL	zaoferować	wynajem	pokojowy	apartament	balkon	powierzchnia	zlokalizować	prestiżowy	lokalizacja		
	↓	↓	↓	↓	↓	↓	↓	↓	↓		
WORD_SIGNAL SCORE	0.10	0.12	0.49	0.60	0.13	0.16	0.23	0.63	0.17	→	OVERALL DESC_SIGNAL SCORE
NUMBER OF OCCURENCES	30	885	87	795	1,957	1,751	474	117	1,856		

Fig. 3 The scheme for calculating the overall DESC_SIGNAL quality scores, based on the sample listing's textual description. *Source*: own calculations

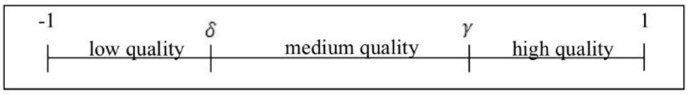


Fig. 4 Visualisation of the training set's calibration procedure. *Source*: own elaboration

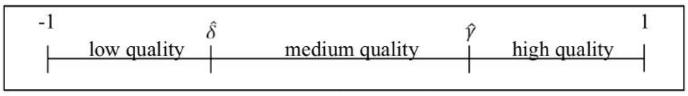


Fig. 5 Visualization of the conversion of the test set's DESC_SIGNAL quality measures from continuous into ordinal scale. *Source*: own elaboration

optimization task has been set for each $DESC\_SIGNAL\_MODEL_{d,v}$ in order to retrieve optimal values of parameters $\gamma_{d,v}$ and $\delta_{d,v}$, for which the model-estimated signals of descriptive quality would match the visual quality signals to the highest degree:

$$\bigwedge_{d \in D} \bigwedge_{v \in V} f(A) = \sum_{i=1}^{I} A_{i,d,v} \Rightarrow max \tag{6}$$

The optimization procedure has been conducted with the use of an evolutionary algorithm (Powell & Batt, 2008). Figure 4 presents the visualisation of the calibration phase procedure.

3.2.4 Models' testing phase

First, the test set observations have been subject to the procedure described in Sect. 3.2.2. for training sets. The test set for the rental market analysis has included 767 observations (14.3% of low, 56.2% of medium and 29.5% of high visual quality), while for the sales market analysis it has been 719 observations (27.4% of low, 33.1% of medium and 39.5% of high visual quality). The overall $DESC\_SIGNAL\_TE_{i,d,v}$ of each test set (TE) observation has been calculated in 48 model's variants, which would be referred as $DESC\_SIGNAL\_TE\_MODEL_{d,v}$. Then, each listing's estimated descriptive quality signal measure has been converted from continuous into ordinal scale based on the threshold parameters $\hat{\gamma}_{d,v}$ and $\hat{\delta}_{d,v}$ obtained for the corresponding training set model's variant (as presented in Fig. 5). Finally, the visual quality signals and estimated descriptive quality signals have been confronted with the use of contingency matrices.

The testing phase has aimed at reducing the scope of search for the models' best variants. As a result, the combinations of dictionary variant and distributional variant of the model that have shown the highest consistency of signals have been selected. The models' variants with the smallest number of cardinal inconsistencies have been of the particular value. A cardinal inconsistency has been defined as assigning the descriptive quality label "high" to the low-visual-quality apartment or the label "low" to the high-visual-quality apartment.

3.2.5 Models' validation phase

In the baseline scenario, the last available quarter has been selected as the test set, while the observations from remaining five quarters have constituted the training set. To provide robust results the whole algorithm has been rerun, every time selecting one different period as the test and the remaining five periods as the training set (sixfold validation). Then, the contingency matrices summarizing the results of all six models have been constructed to select the single variant of the model for which the quality signals have been consistent to the highest degree.

It may be suspected that signals' inconsistency would be higher for listings' with *DESC_SIGNAL* scores closer to the threshold values $\hat{\gamma}_{d,v}$ and $\hat{\delta}_{d,v}$ than to the ones more distant, which may be considered less clear signals. To find, whether the consistency of visual and descriptive quality signals differs depending on the descriptive signal's strength, the ranges of inconclusiveness have been introduced. The parameter $\alpha_{p,d,v}$ ($\alpha > 0$) has aimed at providing the ordinal descriptive quality measure only to some restricted group of observations. Then:

$$\begin{cases} if\ DESC\_SIGNAL\_TE_{i,d,v} > \hat{\gamma}_{d,v} + \alpha_{p,d,v} \Rightarrow \hat{q}_{i,d,v} = high \\ if\ \hat{\gamma}_{d,v} + \alpha_{p,d,v} \geq DESC\_SIGNAL\_TE_{i,d,v} \geq \hat{\gamma}_{d,v} - \alpha_{p,d,v} \Rightarrow \hat{q}_{i,d,v} = no\ decision \\ if\ \hat{\delta}_{d,v} + \alpha_{p,d,v} < DESC\_SIGNAL\_TE_{i,d,v} < \hat{\gamma}_{d,v} - \alpha_{p,d,v} \Rightarrow \hat{q}_{i,d,v} = medium \\ if\ \hat{\delta}_{d,v} + \alpha_{p,d,v} \geq DESC\_SIGNAL\_TE_{i,d,v} \geq \hat{\delta}_{d,v} - \alpha_{p,d,v} \Rightarrow \hat{q}_{i,d,v} = no\ decision \\ if\ DESC\_SIGNAL\_TE_{i,d,v} < \hat{\delta}_{d,v} - \alpha_{p,d,v} \Rightarrow \hat{q}_{i,d,v} = low \end{cases} \tag{7}$$

$$p = \frac{\sum_{i=1}^{I} ND_{i,d,v}}{I}, \tag{8}$$

$$\begin{cases} if\ \hat{q}_{i,d,v} \in \{high, medium, low\} \Rightarrow ND_{i,d,v} = 0 \\ if\ \hat{q}_{i,d,v} = no\ decision \Rightarrow ND_{i,d,v} = 1 \end{cases}. \tag{9}$$

where p refers to the percentage of all observations, for which the visual and descriptive quality signals have been compared (otherwise, the model has not provided a descriptive quality measure—"no decision" and the consistency of signals has not been considered) and $ND_{i,d,v}$ is the test parameter checking whether the descriptive quality label has been assigned or not. The parameter $\alpha_{p,d,v}$ has been estimated using the evolutionary algorithm for $p \in \{0.9;0.8;0.7;0.6;0.5;0.4;0.3;0.2;0.1\}$, while for $p = 1$ the ordinal descriptive quality labels have been assigned to all observations (the situation is then equal to the standard scenario). The scheme of the restricting procedure has been presented in Fig. 6.

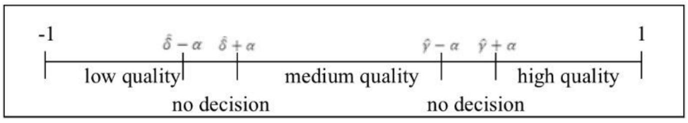


Fig. 6 Visualization of the models' restricting procedure after introduction of ranges of inconclusiveness. *Source*: own elaboration

3.2.6 Models for agents: quality signals sent by brokers vs quality signals sent by owners

The last part of the study has aimed to answer whether the consistency of signals sent by real estate brokers differs from the consistency of signals sent by owners of apartments. Knowing the best *d*—dictionary and *v*—distributional variants of the models for sales and rental listings one can extend *Eq.* 5 to calculate the number of consistent signals for specific apartment quality groups (s) and listed by specific agents (b). Then:

$$\bigwedge_{b\in B}\bigwedge_{q_i\in S}\begin{cases} if\ q_i = \hat{q}_{i,d,v} \Rightarrow A_{i,d,v,b,s} = 1 \\ if\ q_i \neq \hat{q}_{i,d,v} \Rightarrow A_{i,d,v,b,s} = 0 \end{cases} \tag{10}$$

where $B = \{broker;private;not\ indicated\}$ and $S = \{low;medium;high\}$.

Subsequently, the number of listings for which the descriptive and visual quality have proven to be consistent may be defined as:

$$\bigwedge_{b\in B}\bigwedge_{s\in S} C_{b,s} = \sum_{i=1}^{I} A_{i,d,v,b,s} \tag{11}$$

and $T_{d,v,s,b}$ may be defined as a number of all listings of *s*—standard, posted by *b*—agent and calculated according to the rules of *d*—dictionary and *v*—distributional variant of the model. In this regards, *G*—the consistency of the model for each agent may be calculated as:

$$\bigwedge_{b\in B} G_b = \sum_{s\in S}\frac{C_{b,s}}{T_{d,v,b}} \tag{12}$$

However, the consistency of the model for agents calculated this way would be dependent on the quality structure of listings posted by each type of agent. To account for the differences and to ensure the comparability of the results the H-value has been introduced. It represents the consistency of the model of agents weighted with the use of the quality structure of the full dataset (used in the validation phase). Its final form may be presented as:

$$\bigwedge_{b\in B} H_b = \sum_{s\in S}\frac{C_{b,s}}{T_{d,v,b,s}} * \frac{T_{d,v,s}}{T_{d,v}} \tag{13}$$

4 Findings

4.1 Consistency of signals: direct textual vs indirect visual signals of quality

In Tables 3 and 4 the direct textual signals of apartments' quality have been confronted with the visual quality signals. In rental listings, the textually signalled high quality has rarely corresponded with the visual quality. The same has applied to the textual signals of good quality that have corresponded mostly with medium or low visual quality. It may mean that the direct textual signals are overstated, but may be well a result of a different

Table 3 Comparison of direct textual and indirect visual signals of quality of apartments listed for rent. *Source*: own calculations

Apartments listed for rent				
Direct textual signals of quality		Indirect visual signals of quality		
Signal	Number of listings	Number of listings	Signal	Share (%)
High quality	275	86	High quality	31
		186	Medium quality	68
		3	Low quality	1
Good quality	147	5	High quality	3
		88	Medium quality	60
		54	Low quality	37
Freshly renovated	48	4	High quality	8
		39	Medium quality	81
		5	Low quality	10
Non-defined	971	181	High quality	19
		660	Medium quality	68
		130	Low quality	13

Table 4 Comparison of direct textual and indirect visual signals of quality of apartments listed for sale. *Source*: own calculations

Apartments listed for sale				
Direct textual signals of quality		Indirect visual signals of quality		
Signal	Number of listings	Number of listings	Signal	Share (%)
High quality	157	132	High quality	84
		24	Medium quality	15
		1	Low quality	1
Good quality	95	4	High quality	4
		51	Medium quality	54
		40	Low quality	42
Freshly renovated	14	6	High quality	43
		8	Medium quality	57
		0	Low quality	0
To enter	308	142	High quality	46
		128	Medium quality	42
		38	Low quality	12
For renovation / for refreshment	195	0	High quality	0
		0	Medium quality	0
		195	Low quality	100
Non-defined	798	183	High quality	23
		407	Medium quality	51
		208	Low quality	26

(much more optimistic) scale adopted by sellers than by NBP (2022). The more serious problem is that the Polish listing platforms offer no possibility to signal directly, textually low quality, so low-quality apartments are mostly being left unlabelled or mislabelled as "good quality" ones.

In sales listings, the textually signalled high quality has most often corresponded with the visual signal of high quality and textually signalled good quality has corresponded with low or medium visual quality. The number of apartments declared textually as "freshly renovated" has been insignificantly low. Finally, the textual quality signal "to enter", intuitively should refer only to the medium- and high-visual-quality apartments, however, it has also been used for some low-quality ones. Nevertheless, the consistency of signals has been much higher than for the apartments listed for rent.

Table 5 presents the quality structure of listings, for which the direct textual quality signal has been sent compared with the quality structure of observations with quality non-signalled textually. The noticed information gap as for the direct textual quality signal has been substantial and has covered more than half of all observations. Although for the rental listings it has been found that the quality structure of the observations with signalled and non-signalled quality is similar, for the sales listings the difference has been considerable. It may introduce bias to the often-performed analyses that base only on the observations with textually, directly signalled quality.

4.2 Consistency of signals: indirect textual (descriptive) vs indirect visual signals of quality—Wordscores algorithm

The results obtained in the testing phase presented in Table 6 have pointed to the highest consistency of signals in the Wordscores models that have based on relatively little restricted variants of the dictionary ($D \leq 2.5\%$). For sales listings the most consistent models have based on average, 3rd quartile or 9th decile of *WORD_SIGNAL* scores of words included in listings and for rental listings—median, average or 3rd quartile of *WORD_SIGNAL* scores.

Table 7 presents the results of the validation phase for the models of the highest consistency obtained in the testing phase. For sales listings the highest consistency has equalled 64% for the $DESC\_SIGNAL\_TE\_MODEL_{0\%,DE9}$, which means that visual quality signals are most consistent with the textual signals conveyed by the 10% of words with the highest quality-sentiment score. However, the model that based on the average sentiment of words ($DESC\_SIGNAL\_TE\_MODEL_{0\%,AVG}$) has given similar results. As for the dictionary variants, all analysed variants have given comparable results, but the highest consistency of

Table 5 Comparison of the quality structure of listings, for which the direct textual quality signal has been sent with the quality structure of observations with non-signalled quality. *Source*: own calculations

Visual quality of apartment	Rental listings		Sales listings	
	Direct textual signal of quality present (%)	Direct textual signal of quality absent (%)	Direct textual signal of quality present (%)	Direct textual signal of quality absent (%)
High quality	20	19	37	23
Medium quality	67	68	27	51
Low quality	13	13	36	26

Table 6 The consistency of indirect textual (descriptive) and visual quality signals based on the Wordscores models—testing phase results. *Source*: own calculations

Sales listings consistency of signals: testing phase							Rental listings consistency of signals: testing phase						
	DE1	Q1	MED	AVG	Q3	DE9		DE1	Q1	MED	AVG	Q3	DE9
DICT_0%	0.54	0.56	0.55	0.60	0.57	0.62	DICT_0%	0.64	0.66	0.69	0.70	0.69	0.67
DICT_0.25%	0.54	0.55	0.54	0.59	0.57	0.61	DICT_0.25%	0.63	0.67	0.69	0.70	0.69	0.67
DICT_0.5%	0.56	0.57	0.55	0.60	0.58	0.61	DICT_0.5%	0.66	0.68	0.69	0.70	0.69	0.67
DICT_1%	0.56	0.55	0.54	0.60	0.58	0.61	DICT_1%	0.65	0.66	0.69	0.70	0.69	0.66
DICT_2.5%	0.54	0.54	0.54	0.62	0.57	0.62	DICT_2.5%	0.65	0.66	0.69	0.70	0.70	0.67
DICT_5%	0.54	0.55	0.51	0.60	0.58	0.60	DICT_5%	0.64	0.65	0.66	0.68	0.66	0.66
DICT_10%	0.50	0.52	0.54	0.59	0.56	0.58	DICT_10%	0.62	0.64	0.67	0.70	0.63	0.65
DICT_20%	0.47	0.51	0.51	0.56	0.53	0.55	DICT_20%	0.61	0.62	0.62	0.69	0.64	0.64

Table 7 The consistency of indirect textual (descriptive) and visual quality signals based on the Wordscores models—validation phase results. *Source*: own calculations

Sales listings consistency of signals: validation phase				Rental listings consistency of signals: validation phase			
	AVG	Q3	DE9		MED	AVG	Q3
DICT_0%	0.63	0.60	0.64	DICT_0%	0.69	0.70	0.69
DICT_0.25%	0.63	0.60	0.63	DICT_0.25%	0.69	0.71	0.69
DICT_0.5%	0.63	0.60	0.63	DICT_0.5%	0.69	0.70	0.69
DICT_1%	0.63	0.60	0.63	DICT_1%	0.68	0.70	0.69
DICT_2.5%	0.63	0.60	0.63	DICT_2.5%	0.68	0.70	0.69

signals has been obtained for the unrestricted version of the dictionary (*DICT_0%*). For the rental analysis, the *AVG* distributional variant has been dominant and the highest consistency has been achieved for $DESC\_SIGNAL\_TE\_MODEL_{0.25\%,AVG}$, amounting to 71%.

Tables 8 and 9 present the results of the analysis with the introduced ranges of inconclusiveness. Only the variants for which the highest consistency has been detected in the first stage of the validation phase have been presented. In Fig. 7 the results have been presented on a graph.

For rental listings, the consistency of indirect textual and visual quality signals for medium- and high-visual-quality apartments has been high and has increased with the decreasing *p*-percentage of most clear textual signals. However, for low-visual-quality apartments the consistency has been low. It may be rooted in the specificity of rental listings, in which sellers may have incentives to exaggerate and pretend that the quality is higher than it is indicated by the attached photos to achieve the highest possible price. It stands in line with the results presented in Sect. 4.1 concerning sellers' over-optimism. Moreover, the low quality of apartments for rent often manifests itself in a high degree of wear and tear of the apartment or in its obsolete design, which are harder to reflect in texts. Finally, the rental model in the standard scenario have detected only 4 (out of 3482 signals compared which equals around 0.1%) cardinal inconsistencies. This number has dropped even more after introducing ranges of inconclusiveness.

For sales listings the consistency of signals for medium-quality apartments has been high, whereas in the case of high-quality ones it has been very high. The consistency of signals for low-quality apartments has been relatively low, but higher than for the rental listings. It may be rooted in the fact that in the sales market some customers may see the low quality as an advantage. They supposedly do not want to pay extra for the apartment of medium quality, because either way they intend to renovate it prior to moving in. Therefore, sellers do not refrain from describing the quality as low in the textual descriptions or from signalling the low quality using specific vocabulary (i.e. "apartment with potential" used as a synonym for "apartment for renovation"). Finally, comparing two variants of the sales model, the similar results have been achieved in all but one category. The $DESC\_SIGNAL\_TE\_MODEL_{0\%,DE9}$ detected 45/3712 cardinal inconsistencies, (which equaled 1.2%), whereas the $DESC\_SIGNAL\_TE\_MODEL_{0\%,AVG}$ detected only 30/3718 ones (0.8%), which is considerably less. Therefore, although the $DESC\_SIGNAL\_TE\_MODEL_{0\%,AVG}$ variant has shown little less consistency, it has been considered as the best variant of the sales model.

Overall, when introducing the ranges of inconclusiveness, the models showed consistency of signals of up to 83% for the rental listings and 90% for the sales ones. The increasing consistency (following the decrease of p-percentage of all signals compared) is

Table 8 Contingency matrices of the Wordscores models that have shown the highest consistency of signals—results of the validation phase with ranges of inconclusiveness. *Source*: own calculations

			SALES—DESC_SIGNAL_TE_MODEL (0%, AVG)			SALES—DESC_SIGNAL_TE_MODEL (0%, DE9)			RENTAL—DESC_SIGNAL_TE_MODEL (0.25%, AVG)		
			Visual quality signal								
P (%)			Low	Medium	High	Low	Medium	High	Low	Medium	High
100	Model estimated descriptive quality signal	Low	469	242	10	478	241	16	49	27	0
		Medium	483	1062	327	465	1044	306	384	1888	408
		High	20	273	826	29	292	841	4	204	518
70		Low	291	122	5	273	134	6	11	4	0
		Medium	295	768	210	317	747	180	279	1515	232
		High	12	187	709	19	196	727	1	83	314
40		Low	156	45	2	94	36	2	3	1	0
		Medium	123	358	92	167	406	87	144	941	101
		High	9	110	590	13	109	571	1	38	163
10		Low	14	3	1	2	0	0	2	0	0
		Medium	0	0	0	0	0	0	29	207	18
		High	2	30	321	2	31	336	0	14	79

P—the percentage of all signals compared

Table 9 The consistency of indirect textual (descriptive) and indirect visual quality signals based on the Wordscores models that have shown the highest consistency of signals—results of the validation phase with ranges of inconclusiveness. *Source*: own calculations

P (%)	SALES—DESC_SGINAL_TE_MODEL (0%, AVG)	SALES—DESC_SIGNAL_TE_MODEL (0%, DE9)	RENTAL—DESC_SIGNAL_TE_MODEL (0.25%, AVG)
100	0.63	0.64	0.71
70	0.68	0.67	0.75
40	0.74	0.72	0.80
10	0.90	0.91	0.83

P—the percentage of all signals compared

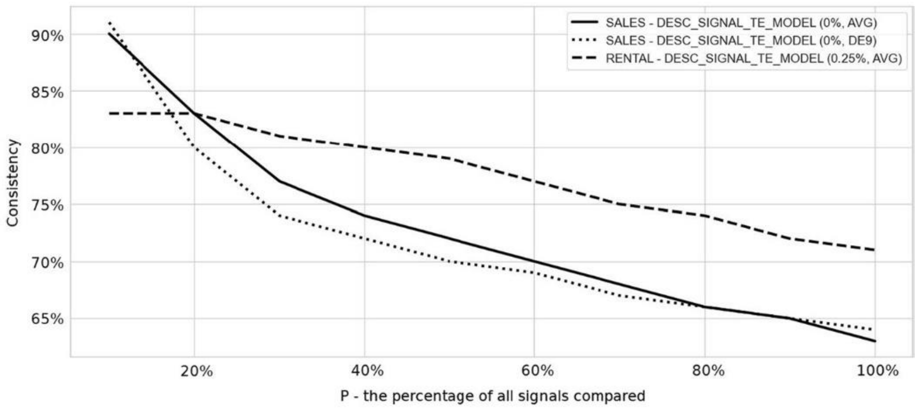


Fig. 7 The consistency of indirect textual (descriptive) and indirect visual quality signals based on the Wordscores models that have shown the highest consistency of signals—results of the validation phase with ranges of inconclusiveness. *Source*: own calculations

particularly important, because it confirms not only the high consistency of indirect textual and visual signals but also proves the increasing consistency of signals along with the increasing clarity of signals.

4.3 Consistency of signals: models for agents

Despite the relatively small number of observations of listings posted by owners of the apartments (only 764 rental and 325 sales listings) compared with the number of listings posted by brokers, the results of models for agents reveal interesting patterns. For rental listings, the consistency (in the form of the quality-corrected H-value) of quality signals has proven to be slightly higher for private listings—72% than for the listings posted by brokers—70%. It may be rooted in the fact that private landlords are more interested in finding the tenant, who not only would generate financial flows, but also would take care of the apartment and would stay in the apartment for longer. Thus, private landlords would have more incentives to represent the reality as it is and to point at the consistent quality standard with the use of both descriptive and visual signals. At the same time, brokers are willing to conclude the transaction quickly, as their commission is not dependent on the time needed to find a suitable tenant. For this reason, they may use advertising tricks e.g. describe the apartment as the one of the higher quality than it really is, which results in the lower consistency of descriptive and visual signals.

As for sales listings, the results have been the opposite—the higher consistency of signals has been found for listings posted by brokers—63% than for the listings posted by the owners of the apartments—60%. In this case the brokerage agency might be interested in providing more reliable descriptions than in tricking the potential customer. In the era of information spreading rapidly on the Internet, apart from focusing only on finding a buyer quickly, brokers would also like to gain a good reputation. In contrast, the private sellers of apartments do not need to take care of their reputation as sellers, because they conclude only a single transaction. Thus, they may be more interested in sending the conflicting quality signals, hoping it will enable to achieve higher sales price.

5 Conclusions

In the research conducted on online sales and rental listings of apartments the information disclosure strategies of sellers have been inspected and the compliance of the three types of apartment soft quality signals: direct textual signals, indirect textual (descriptive) signals and indirect visual signals have been verified. In the first part of the empirical study, it has been documented, that there exists some discrepancy between direct textual signals of housing quality provided by apartments' sellers and the visual quality signals. It has been found that the direct textual signals sent in rental listings have diverged from the visual signals, which could have been caused by a different adopted scale, much more optimistic than the NBP (2022) scale adopted as a reference point. Thus, in the case of apartments for rent, direct textual quality signals may be considered overstated by sellers. The direct textual signals of quality for the sales listings have been mainly in line with the visual quality assessments, which has been in the contrary to the initial assumptions. However, for sales listings, it has been found that there exists a considerable difference of the quality structure of apartments for which sellers declared quality directly and the ones left without declaration. All the mentioned phenomena, combined with the observed very high percentage of observations without directly textually signalled quality (for both rental and sales listings) contribute to the creation of the information gap. Its presence may dampen the efficiency of the market search processes and introduce bias to the increasingly more popular analyses based on sales and rental market listings data. Therefore, to achieve possibly representative and unbiased results of the analyses, one needs to both correct the over-optimistic direct textual quality signals and address the issue of quality of the observations for which no direct textual signal has been provided.

The second part of the study has focused on checking, whether the indirect textual quality signals, processed with the use of the simple supervised ML algorithm—Wordscores are consistent with the visual quality signals. It has been documented, that the sales listings' descriptive and visual quality signals have agreed in 63–90% of cases, depending on the model's variant. For the rental market, the descriptive quality signals matched the visual ones in 71–83% of cases. As a result, it may be concluded that both types of signals show considerable consistency. What is more, the consistency of quality signals for low-quality apartments may reveal the use of different listing strategies. The wording used by sellers in rental listings aims at hiding the negative aspects of apartments, which is being done by refraining from using the words with negative quality sentiment. Therefore, for apartments for rent the low-visual-quality signals have most often corresponded with the medium-quality descriptive signals. At the same time, in the sales market, the low quality may be sometimes considered as an advantage and sellers would refrain less from describing apartments with the wording with negative sentiment. Thus, for apartments for sale the low-visual-quality signals have much more often corresponded with low-quality descriptive signals. It may be also a reason, why the descriptive signals regarded as most clear have been more consistent with the visual signals in the sales listings' analysis than in the rental one.

Moreover, it has been shown, that for rental listings the consistency of signals has been higher for listings posted by private landlords than for the ones posted by brokers. For sales listings it has been the opposite. The obtained results may be rooted in different motivations of both types of agents that drive them, while concluding the sales or rental agreement. Although the differences in consistency of signals sent by brokers and by the owners of the apartments have been relatively small and the numbers of observations have been limited, the results show some interesting patterns, which may constitute a base for further research.

Knowing the high compliance of textual and visual quality signals one may use them interchangeably to enhance the search processes and elevate the quality of the analytical datasets. Then, the descriptive quality measure, which is the fruit of the Wordscores method (*DESC_SIGNAL*), may be used as a variable in the market related studies. It has been proven that to obtain the measure that would be most consistent with the visual quality signal one should utilize information from all or almost all words used in the listed apartment description. It means that in the case of extracting the quality signals the Wordscores model may be treated as a real "bag-of-words" model, in which literally all the words included in the description matter.

Though very promising, the proposed approach to extract textual quality signals from housing listings and the achieved results of signals' consistency generate some issues that should be addressed in further studies. While the analyzed visual quality assessments have been done according to the official guidelines set by the National Bank of Poland, they also include some non-avoidable subjectivity. It may affect the visual quality assessments that have been used to train the Wordscores models on and may disturb the models' results. However, it may be suspected, that the additional variance would rather lower the achieved results on the consistency of signals, thus the consistency may be in reality even higher than it has been proven in this paper.

Moreover, Liu et al. (2020) and Loughran and McDonald (2011) showed that the dictionaries of words are highly localized and Goodwin et al. (2018) proved that the interpretation of real estate wording depends on multiple social and economic factors. As a result, the outcomes of the Wordscores models obtained with the use of the dictionary build on Poznań's data may differ in the specificity of other local markets and demand further testing. The same applies to the influence of the market conditions on the wording used in listings, which has not been tested in this paper, as it requires access to data of a longer time horizon.

Finally, it is advisable to both further extend the knowledge on the quality signals sent via listings and to develop methods of utilizing the textual information on quality. Combined with the promotion of creation of central databases of publicly available information on the real estate market transactions, this would increase market transparency and would enable more accurate and effective market monitoring.

Funding The National Science Centre of Poland partly supported this work under Grant number 2021/43/B/HS4/01213. The author has no other relevant financial or non-financial interests to disclose. The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*, 488–500. https://doi.org/10.1016/b978-0-12-214850-7.50022-x

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys, 34*(3), 512–547. https://doi.org/10.1111/joes.12370

Anenberg, E., & Laufer, S. (2017). A more timely house price index. *Review of Economics and Statistics, 99*(4), 722–734. https://doi.org/10.1162/REST

Ardila, D., Ahmed, A., & Sornette, D. (2021). Comparing ask and transaction prices in the Swiss housing market. *Quantitative Finance and Economics, 5*(1), 67–93. https://doi.org/10.3934/qfe.2021004

Benefield, J. D., Cain, C. L., & Johnson, K. H. (2011). On the relationship between property price, time-on-market, and photo depictions in a multiple listing service. *Journal of Real Estate Finance and Economics, 43*(3), 401–422. https://doi.org/10.1007/s11146-009-9219-6

Beręsewicz, M. (2015). On the representativeness of internet data sources for the real estate market in Poland. *Austrian Journal of Statistics, 44*(2), 45–57. https://doi.org/10.17713/ajs.v44i2.79

Beręsewicz, M. (2019). Correlates of representation errors in internet data sources for real estate market. *Journal of Official Statistics, 35*(3), 509–529. https://doi.org/10.2478/jos-2019-0022

Bian, X., Contat, J. C., Waller, B. D., & Wentland, S. A. (2021). Why disclose less information? Toward resolving a disclosure puzzle in the housing market. *Journal of Real Estate Finance and Economics*. https://doi.org/10.1007/s11146-021-09824-6

Boeing, G., Wegmann, J., & Jiao, J. (2020). Rental housing spot markets: How online information exchanges can supplement transacted-rents data. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X20904435

Bonaccorso, G. (2018). *Machine learning algorithms* (2nd ed.). Packt Publishing.

Brkanić, I. (2017). Housing quality assessment criteria. *Elektronički Časopis Građevinskog Fakulteta Osijek, 14*, 37–47. https://doi.org/10.13167/2017.14.5

Brzezicka, J. (2021). The real-time problem in real estate market research. *World of Real Estate Journal, 2*(116), 61–74. https://doi.org/10.14659/WOREJ.2021.106.05

Bucchianeri, G. W., & Minson, J. A. (2013). A homeowner's dilemma: Anchoring in residential real estate transactions. *Journal of Economic Behavior and Organization, 89*, 76–92. https://doi.org/10.1016/j.jebo.2013.01.010

Cheng, P., Lin, Z., & Liu, Y. (2020). Price discovery with heterogeneous sellers in real estate. *Journal of Real Estate Research, 42*(3), 341–364. https://doi.org/10.1080/08965803.2020.1840898

Cheng, P., Lin, Z., Liu, Y., & Seiler, M. J. (2015). The benefit of search in housing markets. *Journal of Real Estate Research, 37*(4), 597–621. https://doi.org/10.1080/10835547.2015.12091430

de Wit, E. R., & Van der Klaauw, B. (2013). Asymmetric information and list-price reductions in the housing market. *Regional Science and Urban Economics, 43*(3), 507–520. https://doi.org/10.1016/j.regsciurbeco.2013.03.001

Elsinga, M., & Hoekstra, J. (2005). Homeownership and housing satisfaction. *Journal of Housing and the Built Environment, 20*(4), 401–424. https://doi.org/10.1007/s10901-005-9023-4

European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank. (2013). *Handbook on Residential Property Prices (RPPIs) : 2013 edition*. OECD Publishing. https://doi.org/10.2785/34007

Eurostat. (2021). *EU-SILC user database description*. https://doi.org/10.2907/EUSILC2004-2019V.2
Fujihira, K. (2017). Background: Existing Japanese systems related to sustainable housing. *Sustainable Home Design by Applying Control Science*. https://doi.org/10.5772/intechopen.71326
Gastaldo, P., Zunino, R., & Redi, J. (2013). Supporting visual quality assessment with machine learning. *EURASIP Journal on Image and Video Processing*. https://doi.org/10.1186/1687-5281-2013-54
Goodwin, K. R., Waller, B. D., & Weeks, H. S. (2018). Connotation and textual analysis in real estate listings. *Journal of Housing Research, 27*(2), 93–106. https://doi.org/10.1080/10835547.2018.12092149
Goodwin, K., Waller, B., & Weeks, H. S. (2014). The impact of broker vernacular in residential real estate. *Journal of Housing Research, 23*(2), 143–161. https://doi.org/10.1080/10835547.2014.12092089
Grossman, S. J. (1981). Warranties and private disclosure about product quality. *Journal of Law and Economics, 24*(3), 461–483.
Haag, J., Rutherford, R., & Thomson, T. (2000). Real estate agent remarks: Help or hype? *Journal of Real Estate Research, 20*(1–2), 205–216. https://doi.org/10.1080/10835547.2000.12091024
Hotz, V. J., & Xiao, M. (2013). Strategic information disclosure: The case of multiattribute products with heterogeneous consumers. *Economic Inquiry, 51*(1), 865–881. https://doi.org/10.1111/j.1465-7295.2010.00340.x
James, R. N. (2007). Multifamily housing characteristics and tenant satisfaction. *Journal of Performance of Constructed Facilities, 21*(6), 472–480. https://doi.org/10.1061/(asce)0887-3828(2007)21:6(472)
Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association, 65*(330), 532–548. https://doi.org/10.1080/01621459.1970.10481102
Kim, S. S., Yang, I. H., Yeo, M. S., & Kim, K. W. (2005). Development of a housing performance evaluation model for multi-family residential buildings in Korea. *Building and Environment, 40*(8), 1103–1116. https://doi.org/10.1016/j.buildenv.2004.09.014
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157.
Łaszek, J., Olszewski, K., & Waszczuk, J. (2016). Monopolistic competition and price discrimination as a development company strategy in the primary housing market. *Critical Housing Analysis, 3*(2), 1–12. https://doi.org/10.13060/23362839.2016.3.2.286
Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*. https://doi.org/10.1017/S0003055403000698
Le, L. H., Ta, A. D., & Dang, H. Q. (2016). Building up a system of indicators to measure social housing quality in Vietnam. *Procedia Engineering, 142*, 116–123. https://doi.org/10.1016/j.proeng.2016.02.021
Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies, 8*(1), 1–41. https://doi.org/10.1093/rcfs/cfy009
Liu, C. H., Nowak, A. D., & Smith, P. S. (2020). Asymmetric or Incomplete Information about Asset Values? *Review of Financial Studies, 33*(7), 2898–2936. https://doi.org/10.1093/rfs/hhz096
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65. https://doi.org/10.2469/dig.v41.n2.20
Luchtenberg, K. F., Seiler, M. J., & Sun, H. (2019). Listing agent signals: Does a picture paint a thousand words? *Journal of Real Estate Finance and Economics, 59*(4), 617–648. https://doi.org/10.1007/s11146-018-9674-z
Marona, B., & Tomal, M. (2020). The COVID-19 pandemic impact upon housing brokers' workflow and their clients' attitude: Real estate market in Krakow. *Entrepreneurial Business and Economics Review, 8*(4), 221–232. https://doi.org/10.15678/EBER.2020.080412
Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics, 12*(2), 380–391.
Miłkowski, M. (2016). *Polimorfologik 2.1*. https://github.com/morfologik/polimorfologik/releases/tag/2.1
Nasreen, Z., & Ruming, K. (2022). Struggles and opportunities at the platform interface: Tenants' experiences of navigating shared room housing using digital platforms in Sydney. *Journal of Housing and the Built Environment, 37*(3), 1537–1554. https://doi.org/10.1007/s10901-021-09909-x
National Bank of Poland. (2022). *Badanie rynku nieruchomości - Instrukcja badania rynku wtórnego*. https://ssl.nbp.pl/publikacje/rynek_nieruchomosci/Instrukcja-badania-rynku-wtornego-2022.pdf
Nowak, A. D., Price, B. S., & Smith, P. S. (2021). Real estate dictionaries across space and time. *Journal of Real Estate Finance and Economics, 62*(1), 139–163. https://doi.org/10.1007/s11146-019-09740-w
Nowak, A., & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics, 32*(4), 896–918. https://doi.org/10.1002/jae
Office fédéral du logement OFL. (2015). *Concevoir, évaluer et comparer des logements - Système d'évaluation de logements SEL*.

Olszewski, K., Waszczuk, J., & Widłak, M. (2017). Spatial and hedonic analysis of house price dynamics in Warsaw, Poland. *Journal of Urban Planning and Development, 143*(3), 04017009. https://doi.org/10.1061/(asce)up.1943-5444.0000394

Poursaeed, O., Matera, T., & Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications, 29*(4), 667–676. https://doi.org/10.1007/s00138-018-0922-2

Powell, S. G., & Batt, R. J. (2008). *Modeling for insight: A master class for business analysts*. Wiley.

Qiu, L., Tu, Y., & Zhao, D. (2020). Information asymmetry and anchoring in the housing market: A stochastic frontier approach. *Journal of Housing and the Built Environment, 35*(2), 573–591. https://doi.org/10.1007/s10901-019-09701-y

Qualitel. (2021). *Le Référentiel Qualité - Certification Renovation*.

Renigier-Biłozor, M., Janowski, A., Walacik, M., & Chmielewska, A. (2022). Human emotion recognition in the significance assessment of property attributes. *Journal of Housing and the Built Environment, 37*(1), 23–56. https://doi.org/10.1007/s10901-021-09833-0

Renigier-Biłozor, M., Janowski, A., & D'Amato, M. (2019). Automated Valuation Model based on fuzzy and rough set theory for real estate market with insufficient source data. *Land Use Policy*. https://doi.org/10.1016/j.landusepol.2019.104021

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55. https://doi.org/10.1086/260169

Seiler, M. J., Madhavan, P., & Liechty, M. (2012). Toward an understanding of real estate homebuyer Internet search behavior: An application of ocular tracking technology. *Journal of Real Estate Research, 34*(2), 211–241. https://doi.org/10.1080/10835547.2012.12091333

Seo, Y., Im, J., & Mikelbank, B. (2020). Does the written word matter? The role of uncovering and utilizing information from written comments in housing ads. *Journal of Housing Research, 29*(2), 133–155. https://doi.org/10.1080/10527001.2020.1849929

Shen, L., & Ross, S. L. (2021). Information value of property description: A machine learning approach. *Journal of Urban Economics*. https://doi.org/10.1016/j.jue.2020.103299

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*, 355–374. https://doi.org/10.1016/b978-0-12-214850-7.50025-5

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review, 92*(3), 434–459. https://doi.org/10.1257/00028280260136200

Tomal, M. (2022). Exploring the meso-determinants of apartment prices in Polish counties using spatial autoregressive multiscale geographically weighted regression. *Applied Economics Letters, 29*(9), 822–830. https://doi.org/10.1080/13504851.2021.1891194

Trojanek, R., Gluszak, M., Tanas, J., & Hebdzynski, M. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1016/0022-1031(70)90052-1

Wang, B. (2021). How does COVID-19 affect house prices? A cross-city analysis. *Journal of Risk and Financial Management, 14*(2), 47. https://doi.org/10.3390/jrfm14020047

Article 3 **Hebdzyński, M. (2024a). Are hedonic models really quality-adjusted? The role of apartment quality in hedonic models of housing rental market. *Real Estate Management and Valuation, 32*(2), 46-57. https://doi.org/10.2478/remav-2024-0014**

www.degruyter.com/view/j/remav

ARE HEDONIC MODELS REALLY QUALITY-ADJUSTED? THE ROLE OF APARTMENT QUALITY IN HEDONIC MODELS OF HOUSING RENTAL MARKET

Michał Hebdzyński[1]*

[1] *Department of Microeconomics, Poznan University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: michal.hebdzynski@phd.ue.poznan.pl, ORCID: 0000-0003-0354-054X*

** Corresponding author*

ARTICLE INFO

Keywords:

rental market, quality, signaling, hedonic methods, price index

JEL Classification:

C55, C81, D82, R31

ABSTRACT

In micro-level data concerning the housing market, the apartment quality may be signaled via textual statements or the attached descriptions/photos. It may be done using hard information related to the easier-to-measure structural characteristics or soft information related to the apartment condition and design - *soft quality*. This paper checks whether the choice of the approach to handling the issue of soft quality of apartments influences the properties of hedonic models and the course of hedonic rent indices. The study shows that hedonic models that account for soft quality have better statistical properties than those without soft-quality-related variables. Among them, the models that include the information on quality extracted from descriptions of apartments prove to be the best. Still, considerable differences in the indicated course of hedonic rent indices have not been detected. However, the paper concludes that utilizing information on apartments' soft quality may be crucial to understanding the market adjustment process to economic shocks. It has been proven that the price reaction of the Poznań (Poland) rental market to the COVID-19 pandemic and the Russian aggression on Ukraine has been diversified in the quality-related market segments.

Citation:

Hebdzyński, M. (2024). Are hedonic models really quality-adjusted? The role of apartment quality in hedonic models of housing rental market. *Real Estate Management and Valuation, 32*(2), 46-57. https://doi.org/10.2478/remav-2024-0014

1. Introduction

A developed housing rental market may not only contribute to the stability of the real estate market but also plays a role in maintaining the stability of the whole economy (Czerniak & Rubaszek, 2018; Rubaszek & Rubio, 2020). However, it has recently been subject to two economic shocks of global importance. First, the COVID-19 pandemic that was spread throughout Europe starting in early 2020. The date of expiration of the pandemic shock can be assumed to be the same as the date of the emergence of another one, i.e. the Russian invasion of Ukraine on 24.02.2022, which resulted in an unprecedented inflow of Ukrainian refugees across Europe, among which 1 million applied for temporal refugee protection in Poland (Office for Foreigners, 2023).

At the core of the rental market monitoring and supervision, one can place the analysis of price movements. European Commission, Eurostat, Organisation for Economic Co-operation and Development and World Bank (2013) recommend constructing quality-adjusted indices of housing prices with the use of hedonic methods developed by Rosen (1974) based on Lancaster's (1966) theory of consumer demand. Their main idea is that prices of heterogenic goods may be divided into implicit prices of attributes. This is rooted in the revealed preference theory (Samuelson, 1938), stating that purchases made by utility-maximizing entities reveal the utility assigned to individual product features. One can use micro-level cross-sectional data to conduct an econometric decomposition of prices. Based on the micro-level hedonic model, the market's macro-level hedonic price index (HPI) may be obtained. This would aim to provide a time series adjusted to the changing

eISSN: 2300-5289

quality of the rented or sold apartments.

To design a quality adjustment of the hedonic model, one must first establish the targeted definition of quality. Focusing on the quality of individual apartments, some understand it as the representation of structural characteristics – e.g. number of rooms, age of the building, presence of a balcony (Elsinga & Hoekstra, 2005; James, 2007). The information about them may be regarded as objective, hard information (Liberti & Petersen, 2019). Others understand quality more as an apartment's finishing, condition, design or adaptability (Kain & Quigley, 1970; Kim et al., 2005). This kind of soft information depends more on the person providing the quality assessment, and is thus more subjective, even if the assessor is an expert in the field.

In hedonic models of the housing market, the representation of quality based on soft information (*soft quality*) has been included in different forms. Some researchers used variables based on the ordinal-scale assessments of brokers or other listing providers (de Wit & Van der Klaauw, 2013; Trojanek & Gluszak, 2022; Trojanek et al., 2021). Others (Olszewski et al., 2017) implemented quality-reflecting dummy variables and thus did not specify *a priori* the differences between the assessment levels. Although, to the best of our knowledge, none of the researchers who utilized the supervised machine learning methods have targeted the issue of quality directly, some have approached the quality-related information contained in text data. Shen and Ross (2021) used natural language processing to extract the measure of uniqueness from property descriptions (that reflected both the housing quality and its market power) to include it in the hedonic model. Seo et al. (2020) focused on the indirect inclusion of quality by enriching the structure of hedonic models with tokens reflecting the presence or absence of positive quality descriptors. Finally, Liu et al. (2020) and Nowak and Smith (2017) used the LASSO algorithm to form the dictionaries of words that have proven to constitute a proxy for the omitted, quality-reflecting variables. Although only Trojanek et al. (2021) targeted the rental market, quality-reflecting variables have proven to be significant determinants of house prices and elevated the quality of hedonic models in all of the mentioned studies.

Apart from choosing a desired understanding of quality and the form of its inclusion in the hedonic model, one should remember that the housing market is a place full of information asymmetries. This means that the selling side of the transaction possesses the informational advantage over the buying side in terms of the true quality of the offered product (Akerlof, 1970). To reduce the informational distance between the two parties, sellers send signals that aim to reveal the quality of the product (Spence, 1973). It has been argued that the product's quality has to be signaled as this would otherwise indicate its low quality (Grossman, 1981; Milgrom, 1981). However, Hotz and Xiao (2013) found that, to reach some customers, it is more profitable for sellers not to signal quality fully, as confirmed by Bian et al. (2021) for the housing sales market. Lastly, Hebdzyński (2023) did not find a considerable difference in the quality structure of listings of apartments for rent with signaled and non-signaled quality. However, he did show that the quality declared by listing' providers might be overstated.

In Poland, as is common for the EU countries, transactional data gathered by state institutions are scarce, confidential and available with a significant lag. Moreover, they focus solely on hard information, rarely containing any signals of soft quality. Considering the above, one can exploit alternative sources, such as listings data. Because of the growing popularity of online listing platforms, the data are widely accessible and, more often than not, very detailed. However, they cannot be regarded as fully representative even if they cover most of the market supply (Beręsewicz, 2019). Nevertheless, hedonic indices of house prices that rely on listing data may be successfully used as a proxy (Ardila et al., 2021; Lyons, 2019; Shimizu et al., 2016) or supplementary information (Kolbe et al., 2021) to transactional indices. Moreover, they include soft quality information in addition to hard information. According to the division proposed by Hebdzyński (2023), this may take the form of *direct textual signal* of quality (direct declaration by landlord/agent, in a pre-defined form, standardized by the listing platform), *indirect textual signal* (conveyed via the whole textual description of the apartment) or *indirect visual signal* (conveyed via photos attached to the listing). Because of the rarity of data containing direct signals and the problematic processing of indirect signals, researchers have often constructed hedonic models that did not account for the apartments' soft quality. However, the methods of obtaining quality-related information from housing listings have been recently developed (Hebdzyński, 2023; Nowak & Smith, 2017; Poursaeed et al., 2018; Shen & Ross, 2021). Therefore, it may be asked whether it is still justified to call the models that do not account for soft quality in any way "quality-adjusted".

This research aims to answer whether accounting for the variables reflecting soft quality in hedonic models reveals previously unknown information on the housing rental market phenomena, influences the statistical properties of hedonic models and the course of hedonic rent indices (HRI). It may be argued that although in the housing sales market, the purchased apartments are often renovated quickly after the sales transaction, in the rental market, a tenant does not desire to implement any significant changes to the quality of the apartment. Thus, we hypothesize that soft quality's role in shaping individual apartments' rents might be pivotal. In this regard, the hedonic model with no soft quality variable has been compared with three approaches to utilize the quality signals provided in listings. Then it was checked whether the price tendencies obtained for separate, quality-related segments are in line with the tendencies of the price-related segments. Finally, the analysis of the course of HRIs amid the recent economic shocks was presented.

The unique dataset of listings of apartments for rent located in multi-family buildings in Poznań (Poland) has been used to answer the research questions. It contains 8,248 observations from the period June 2019 – December 2022. As regards the analytical methods, the Ordinary Least Squares (OLS) and Quantile Regression (Koenker & Bassett, 1978) approaches to constructing hedonic models were chosen.

The study contributes in three ways. First, it adds to the literature concerning signaling theory by empirically verifying the linkages of various types of quality signals with the overall value of individual apartments for rent. The results extend the studies by Hebdzyński (2023), who analyzed the compatibility of visual and textual quality signals and Seo et al. (2020), who studied how the inclusion of textual quality signals translates to prices of apartments listed for sale.

Secondly, the paper inspects the fluctuations of rent levels amid the pandemic-related and war-related shock and contributes to a better understanding of the economy's reaction to exogenous shocks. Although both situations have already been studied – by Trojanek et al. (2021) for the pandemic and by Trojanek and Gluszak (2022) for the war-related refugee crisis in Poland - none of the studies have provided the indices for quality-related segments. This paper argues that this distinction conveys essential, previously unknown information.

Thirdly, the present work contributes to the further development of efficient analytical methods. It proposes methods of using listings data to explore areas that are impossible to reach with conventional transactional data. As a result, the work levels the quality distance between the harder-to-reach transactional data and the easily available listings data. Moreover, it adds understanding to the sensitivity of hedonic indices to the selection of variables and the method used. Hence, it complements the research of Diewert and Shimizu (2022), who presented the minimal requirements for a satisfactory hedonic model, as well as Hill and Trojanek (2022), who compared the outcomes of multiple hedonic modelling approaches.

The structure of the rest of the paper is as follows. *Section 2* describes the methodological approach to answering the research questions and the data used, *Section 3* outlines the results, while *Section 4* provides a discussion of their interpretation. Finally, *Section 5* concludes the manuscript.

2. Material and methods

2.1. Data

The analytical dataset contains 8,248 observations of listings of apartments for long-term rent located in multi-family buildings in Poznań (Poland), listed in two of the top Polish online listing platforms, Otodom.pl and Gratka.pl. Both data sources were found to be almost identical regarding the type of information included in listings, which were gathered quarterly from June 2019 to December 2022 in the middle of the last month of a quarter. Observations without a textual description, without attached photos, and ones for which it was impossible to determine the geolocation of the listed apartment with an accuracy of one kilometer, were removed from the dataset. If the same apartment was listed multiple times in one period or listed in adjacent periods, only the last observation was left in the dataset. Duplicates were identified based on the assumption that two listings that refer to apartments located on the same street, with the same floor area (rounded to integers), the same number of rooms, located on the same floor and listed in the same or adjacent periods may be considered listings of the same apartment. Only those observations for which information on all the variables presented in Table 1 was available were retained in

eISSN: 2300-5289

the dataset. The variables marked with * were available in only one data source – Gratka.pl, and were gathered in a limited time range, from September 2020 to December 2022 (N=1791).

Table 1

Descriptions and basic statistics of the variables used in the research

VARIABLE	DESCRIPTION	MIN	AVG	MAX	COUNT OF 1'S
RENT (dependent variable)	rent for the apartment [in PLN]	650	1961.5	9800	
AREA	floor area of the apartment [in m^2]	15	47.5	150	
ROOMS_INTENSITY	rooms per 1 m^2 of the apartment	0.011	0.043	0.08	
MODERN_APART	1- if the apartment is located in the modern apartment building, 0- otherwise				2162
RESTORED_TENEMENT	1- if the apartment is located in the restored tenement building, 0- otherwise				590
FULL_FURNISHING	1- if the apartment is fully or almost fully equipped with furniture, 0- otherwise				7096
BALCONY	1- if there is a balcony/terrace in the apartment, 0- otherwise				6066
PARKING	1- if there is access to the designated parking space (included in the listed rent), 0- otherwise				1660
GARDEN	1- if there is an access to the private garden, 0- otherwise				257
DIST_CC	distance of the apartment to the city center [in km]	0.021	2.982	10.916	
DIST_GREEN	distance of the apartment to the nearest urban green area [in km]	0.002	0.268	0.996	
DIST_LAKE	distance of the apartment to the nearest lake [in km]	0.044	2.226	6.406	
Q_VISUAL_HIGH	1- if the indirect visual signal of quality indicated that the quality of the apartment is high, 0- otherwise				2052
Q_VISUAL_MEDIUM	1- if the indirect visual signal of quality indicated that the quality of the apartment is medium, 0- otherwise				5425
Q_VISUAL_LOW	1- if the indirect visual signal of quality indicated that the quality of the apartment is low, 0- otherwise				771
* *Q_DIR_TEXTUAL_HIGH*	1- if the direct textual signal of quality indicated that the quality of the apartment is high, 0 - otherwise				439
* *Q_DIR_TEXTUAL_GOOD*	1- if the direct textual signal of quality indicated that the quality of the apartment is good, 0- otherwise				172
* *Q_DIR_TEXTUAL_RENOVATED*	1- if the direct textual signal of quality indicated that the apartment is freshly renovated, 0- otherwise				51
* *Q_DIR_TEXTUAL_NOSIGNAL*	1- if the direct textual signal of quality has not been sent by the listing provider, 0- otherwise				1129
Q_INDIR_TEXTUAL	indirect textual quality signal of the apartment based on the Wordscores algorithm (continuous variable)	-0.014	0.165	0.335	

Source: own elaboration.

The indirect visual signals of quality have been obtained in the process of individual visual assessment of the quality of apartments solely based on photos attached to the listings, following the quality-assessment instruction by the National Bank of Poland (2023) (an institution responsible for the monitoring of prices on the Polish housing market). The attributes of the apartment accounted for in the quality assessment were floors, fixtures, doors, walls, and kitchen & bathroom equipment. The high-quality label was assigned to functionally arranged apartments, finished with good materials and revealing low exploitation. Those in need of renovation or refreshment were considered low quality, while the rest – medium quality.

The variables representing direct textual quality signals are based on the declaration of the listing provider and were available in only one data source (Gratka.pl). For each listing, the person who was responsible for posting it on the listing platform could choose one quality class out of a finite set of categories, i.e. – "high quality", "good quality" or "freshly renovated", or choose to leave the quality undeclared.

The variable reflecting indirect textual quality signals has been obtained from textual descriptions of the listed apartments using the Wordscores algorithm (Laver et al., 2003). The method has been adapted and calibrated for extracting textual quality signals from housing listings by Hebdzyński (2023). It has demonstrated high accuracy with the human-made visual assessments of the apartment photos attached to listings. The steps of the Wordscores algorithm have been conducted as follows:

1. The training set was defined; it contained 3,127 observations of apartments listed for rent between June 2019 and June 2020. Each observation included a textual description of the apartment together with the assigned indirect visual quality label.
2. To reduce the noise of the analysis, numbers, special characters, and shortest words (≤4 characters) were removed from each description in the training and analytical dataset.
3. All words in all listings were replaced with their base forms (lemmas), with the use of a morphological dictionary (Miłkowski, 2016).
4. For the training set, the occurrences of each lemma in the descriptions of apartments of a given visual quality were counted:

$$n_j = h_j + m_j + l_j, \quad (1)$$

where n_j is the number of occurrences of the j-th lemma in all observations, h_j, m_j and l_j refer to the number of occurrences of the j-th lemma in the descriptions of apartments of high-, medium- and low-visual-quality. Lemmas representing the least frequently used words (with $n_j < 8$) were excluded.
5. Each occurrence of a lemma in the description of a low-visual-quality apartment from the training set was treated as a negative signal of the lemma, while occurrence in the description of a high-visual-quality apartment – as a positive signal. Occurrences in the descriptions of medium-visual-quality apartments' were treated as neutral signals. Therefore, the quality signal of the lemma may be calculated as:

$$LEMMA\_SIGNAL_j = \frac{h_j*(+1)+m_j*0+l_j*(-1)}{n_j} \quad (2)$$

Then, a list of lemma-*LEMMA_SIGNAL* pairs may be referred to as a dictionary.
6. The *LEMMA_SIGNAL* score from a dictionary was assigned to each lemma in each observation from the analytical dataset. The final numerical representation of the listing's textual quality signal – *TEXTUAL_SIGNAL* was calculated as an average of all *LEMMA_SIGNAL* scores.

2.2. Hedonic methods used

First, Ordinary Least Squares (OLS) hedonic models with time-dummies and the logarithm of rent for an apartment as the dependent variable were constructed. However, OLS has some disadvantages, among which heteroskedasticity of the error term and sensitivity to extreme values and outliers are often mentioned. The former issue was dealt with in this paper by using the heteroskedasticity robust variance estimator (White, 1980), but the latter is more problematic to handle within OLS. In such a case, the use of Quantile Regression (QR) (Koenker & Bassett, 1978), which mitigates both mentioned problems of OLS, should be considered superior. Additionally, it allows modeling any conditional quantile of the dependent variable. Then the models for selected quantiles may be referred as models representing price-related segments of the market (and the results for median quantile QR model are equivalent to the OLS's results). However, QR models are more difficult to test and compare using conventional statistical tools than OLS models. The QR model may be specified as:

$$ln\, R_i = X_{ik}\beta_{\theta k} + \varepsilon_{\theta i} \quad (3)$$
$$with\ Q_\theta(ln\, R_i|X_{ik}) = X_{ik}\beta_{\theta k} \quad (4)$$

where: R_i is an apartment's rent, X_{ik} is a vector of independent variables, θ is an estimated regression quantile, $\beta_{\theta k}$ is a vector of coefficients for the observations of the dependent's variable's θth quantile, $\varepsilon_{\theta i}$is an error term and $Q_\theta\,(\ln R_i|X_{ik})$ represents the θth quantile of a dependent variable $\ln R_i$given X_{ik}.

Lastly, the OLS models were compared using the BIC information criterion (Schwarz, 1978) and R^2, whereas the QR models were compared with the use of pseudo R^2 (Koenker & Machado, 1999).

2.3. Analytical steps

In each step, the statistical properties of the specific versions of hedonic models and hedonic rent indices built based on the models' results were compared.

2.3.1. Models with three approaches to accounting for quality

As noted in Section 2.1., the direct textual quality signals were available in only one source of data (Gratka.pl) and in a limited timeframe (from September 2020 to December 2022). Thus, the first part of the research was conducted on a restricted dataset to compare models that include all three approaches to account for apartment quality and the hedonic model containing no quality-related variables. Then, as some listings providers decided not to directly signal quality textually, it was checked whether the hedonic price indices built based on the observations for which the quality signal had been sent show different price tendencies from those for which it had not been sent.

1) Construction of OLS models including:
   - no variable reflecting quality signals – model *A_1*,
   - variables reflecting direct textual quality signals – model *B_1*,
   - variables reflecting indirect visual quality signals – model *C_1*,
   - variable reflecting indirect textual quality signals – model *D_1*.
2) Construction of OLS models for two subsets of observations, for which the signal of direct textual quality:
   - had been sent – model *E_1*,
   - had not been sent – model *F_1*.

2.3.2. Models with two approaches to accounting for quality

As a second step, the models on the full dataset (for two data sources and the entire time period) were calculated using two different modeling approaches. The models compared two approaches to account for apartment quality and the hedonic model with no quality-related variables. OLS and median (50^{th} percentile) QR models were constructed, and included:

- no variable reflecting quality signals – models *A_2* and *A_3_Q50*,
- variables reflecting indirect visual quality signals – models *C_2* and *C_3_Q50*,
- variable reflecting indirect textual quality signals – models *D_2* and *D_3_Q50*.

2.3.3. Models for price-related and quality-related segments

Finally, the models on the full dataset divided into price-related and quality-related market segments have been constructed using two modeling approaches.

1) Construction of QR models for price-related segments. The share of low-quality apartments in the analytical dataset amounted to 9%, with the share of high-quality ones totaling 25% (based on the distribution of indirect visual quality signals). Thus, the models (*D_3_Q9*, *D_3_Q42*, *D_3_Q75*) were constructed for the 9^{th}, 75^{th} and 42^{nd} conditional percentile of the dependent variable – where the 42^{nd} percentile reflects the midpoint between the 9^{th} and 75^{th} percentile.
2) Construction of median QR models for quality-related market segments – *D_4_L* (for low quality), *D_4_M* (for medium quality), *D_4_H* (for high quality). The observations have been divided according to indirect visual quality signals. However, the indirect textual quality signal measure has been retained in the model to account for quality differences within the given segment.

3. Results

3.1. Models with three approaches to accounting for quality

Based on the results presented in Table 2, it should be noted that regardless of the approach to accounting for quality, the models that do so are statistically better than the model with no quality-reflecting variables. Looking at R^2 and BIC, the *D_1* model (the inclusion of indirect textual quality signal) has been superior. The difference in the percentage of explained variance between the worst and the best models - *A_1* and *D_1* has amounted to 11 p.p. Apart from the variable *Q_DIR_TEXTUAL_RENOVATED,* all the quality-related variables proved to be statistically significant in explaining rents. Interestingly, the direct textual signal of “good quality” was negatively linked with rents (when no quality declaration was taken as a base).

Looking at the left panel of *Fig. 1.*, it can be seen that, although the course of HRI is very similar for all approaches, the *A_1* model showed the highest rise in rents after the outbreak of the war in Ukraine, potentially overestimating its scale. From the right panel of Fig. 1., it can be inferred that HRIs obtained from subsets divided according to the availability of the direct quality signal behave differently, especially in the times of the war in Ukraine.

3.2. Models with two approaches to accounting for quality

The results of the models constructed on the full dataset (*Table 3*) confirm that the models utilizing quality signals are noticeably better than those not including quality-related variables. Using indirect textual quality signals (*D_2* and *D_3_Q50*) provided the best statistical properties. From Fig. 2., it may be inferred that OLS and QR models without quality-reflecting variables (*A_2* and *A_3_Q50*) slightly underestimated the drop in rents during the pandemic and overestimated the rise in rents due to the war in Ukraine. Nevertheless, the differences in HRIs between the two estimation methods and the approaches to accounting for quality were very small.

Table 2

Coefficients and basic statistics obtained in OLS models with different approaches to accounting for apartment quality

MODEL	*A_1*	*B_1*	*C_1*	*D_1*	*E_1*	*F_1*
EXPLANATORY VARIABLE	COEFF	COEFF	COEFF	COEFF	COEFF	COEFF
AREA	0.012 ***	0.012 ***	0.012 ***	0.012 ***	0.012 ***	0.012 ***
ROOMS_INTENSITY	3.52 ***	3.65 ***	3.55 ***	3.49 ***	3.50 ***	3.77 ***
MODERN_APART	0.15 ***	0.14 ***	0.08 ***	0.03 ***	0.14 ***	0.13 ***
RESTORED_TENEMENT	0.10 ***	0.09 ***	0.05 ***	0.03	0.06 *	0.10 ***
FULL_FURNISHING	0.07 ***	0.07 ***	0.04 ***	0.04 ***	0.04 *	0.09 ***
BALCONY	0.05 ***	0.04 ***	0.03 ***	0.02 **	0.05 ***	0.04 ***
PARKING	0.07 ***	0.06 ***	0.05 ***	0.04 ***	0.04 **	0.08 ***
GARDEN	0.15 ***	0.14 ***	0.07 ***	0.05 ***	0.07 **	0.20 ***
DIST_CC	-0.028 ***	-0.018 ***	-0.014 ***	-0.010 ***	-0.018 ***	-0.020 ***
DIST_GREEN	-0.204 ***	-0.190 ***	-0.111 ***	-0.070 ***	-0.131 ***	-0.224 ***
DIST_LAKE	-0.010 **	-0.008 *	-0.007	-0.007 *	-0.018 ***	0.000
Q_DIR_TEXTUAL_GOOD		-0.09 ***			-0.12 ***	
Q_DIR_TEXTUAL_RENOVATED		0.03				
Q_DIR_TEXTUAL_HIGH		0.08 ***			0.05 **	
Q_VISUAL_HIGH			0.23 ***			
Q_VISUAL_LOW			-0.19 ***			
Q_INDIR_TEXTUAL				3.48 *		
TIME_DUMMIES	YES	YES	YES	YES	YES	YES
CONSTANT	6.68 ***	6.70 ***	6.72 ***	6.20 ***	6.75 ***	6.68 ***
NOBS	1791	1791	1791	1791	662	1129
R^2	0.68	0.70	0.78	0.80	0.77	0.66
BIC	-692.3	-772.6	-1334.5	-1495.5		

**** for P ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 0.1.*

Source: own elaboration.

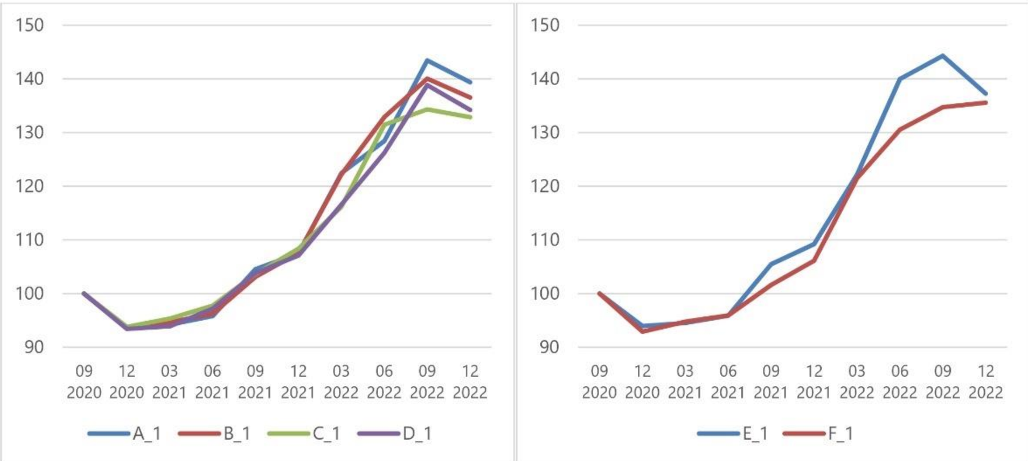


Fig. 1. Comparison of hedonic rent indices obtained based on OLS models with different approaches to accounting for apartment quality (09 2020=100). *Source*: own elaboration.

Fig. 2. Comparison of hedonic rent indices obtained based on OLS (left panel) and median QR (right panel) models with different approaches to accounting for apartment quality (12 2019=100). *Source*: own elaboration.

Table 3

Coefficients and basic statistics obtained in OLS and median QR models with different approaches to accounting for apartment quality

MODEL	A_2	C_2	D_2	A_3_Q50	C_3_Q50	D_3_Q50
EXPLANATORY VARIABLE	COEFF	COEFF	COEFF	COEFF	COEFF	COEFF
AREA	0.012 ***	0.012 ***	0.012 ***	0.012 ***	0.012 ***	0.011 ***
ROOMS_INTENSITY	3.75 ***	3.65 ***	3.81 ***	3.98 ***	3.66 ***	3.79 ***
MODERN_APART	0.16 ***	0.08 ***	0.04 ***	0.15 ***	0.08 ***	0.04 ***
RESTORED_TENEMENT	0.10 ***	0.05 ***	0.03 ***	0.10 ***	0.06 ***	0.04 ***
FULL_FURNISHING	0.08 ***	0.05 ***	0.04 ***	0.09 ***	0.05 ***	0.04 ***
BALCONY	0.07 ***	0.05 ***	0.04 ***	0.07 ***	0.05 ***	0.04 ***
PARKING	0.08 ***	0.05 ***	0.04 ***	0.08 ***	0.05 ***	0.04 ***
GARDEN	0.07 ***	0.03 ***	0.01	0.07 ***	0.03 **	0.02
DIST_CC	-0.021 ***	-0.015 ***	-0.011 ***	-0.021 ***	-0.016 ***	-0.012 ***
DIST_GREEN	-0.145 ***	-0.081 ***	-0.046 ***	-0.139 ***	-0.073 ***	-0.040 ***
DIST_LAKE	-0.002	-0.003	-0.002	-0.002	-0.004 *	-0.001
Q_VISUAL_HIGH		0.20 ***			0.19 ***	
Q_VISUAL_LOW		-0.18 ***			-0.19 ***	
Q_INDIR_TEXTUAL			3.13 ***			3.05 ***
TIME_DUMMIES	YES	YES	YES	YES	YES	YES
CONSTANT	6.66 ***	6.70 ***	6.22 ***	6.67 ***	6.71 ***	6.24 ***
NOBS	8248	8248	8248	8248	8248	8248
R^2	0.68	0.77	0.79	0.43	0.52	0.55
BIC	-4251.7	-7005.4	-7675.7			

**** for P ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 0.1.*

Source: own elaboration.

3.3. Models for price-related and quality-related segments

The results in Fig. 3 show that the short-term HRIs obtained for the different price-related segments point to almost identical directions and magnitude of changes. Although more volatile, the results for quality-related segments indicate that the direction and scale of fluctuations have differed, depending on the analyzed segment. The rent dynamics have been presented in Fig. 4 and Fig. 5.

Fig. 3. Comparison of hedonic rent indices obtained in QR models for price-related (left panel) and quality-related (right panel) market segments (12 2019=100). *Source*: own elaboration.

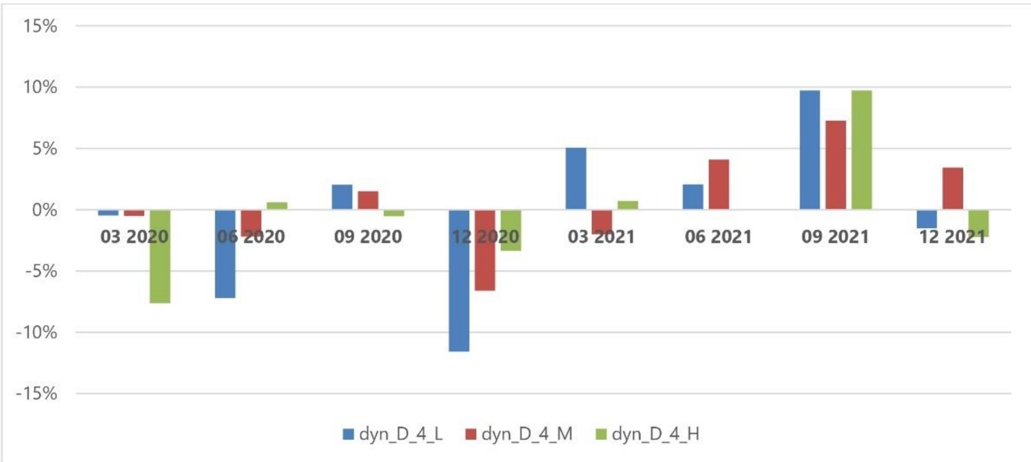


Fig. 4. Quarter-to-quarter dynamics of hedonic rent indices based on median QR models for quality-related segments during the COVID-19 pandemic. *Source*: own elaboration.

4. Discussion

Based on the results of multiple hedonic approaches, it was found that models that account for soft quality are better fitted than those without quality-related variables. The difference in the goodness-of-fit was small for the model that utilized direct textual quality signals, while the models with indirect quality signals were found to be superior for both OLS and median QR. However, the weaker determination of the *B_1* model should be associated with the fact that the quality was declared directly by the listing provider for only a small number of observations, rather than with the low information load of these declarations. This can be inferred from the R^2 coefficient, which for the *E_1* model was close to the values obtained in the best models, while for the *F_1* model - was the worst of all OLS models.

Although the implementation of quality considerations increased the quality of models, this had little effect on the course of HRIs presented in Fig. 1 (left panel) and Fig. 2. Only a slight overvaluation of the index in selected periods was observed for both OLS and median QR models. Regardless of the approach taken, the HRIs were

eISSN: 2300-5289

nearly identical, pointing to the fact that neither the additional variables used nor the chosen analytical method change the course of HRI. What matters more is the selection of observations. From the right panel of Fig. 2, it can be inferred that one needs to be very careful when excluding observations for which no quality was declared from hedonic modeling. Hence, in this process, it is better to resign from implementing the quality considerations than to exclude the observations with non-signaled quality. The subsamples of observations with signaled and non-signaled soft quality showed considerably different price variability. Thus, the decision whether or not to send a signal does not appear to be random.

The use of indirect textual and visual signals of quality has enabled the study of quality-related market segments. In Fig. 3, the differences in HRI for price-related segments have been negligible, while the HRIs for quality-related segments have differed. Hence, based on the HRIs built for quality-related market segments, one can provide a deepened analysis of the recent turbulences in the Polish rental market. Fig. 4. shows that the drop in rent first occurred in the high-quality segment at the beginning of the pandemic. This was supposedly fueled by the cross-border traffic restrictions that affected tourism and the withdrawal of high-quality apartments from the short-term rental market to list them for long-term rent. In the next quarter, the decrease in rent levels was visible for low-quality apartments, as students were terminating their rental agreements and returning to their family homes to stay there during lockdowns. At the same time, the rents of medium-quality apartments did not change much. Afterwards, after the autumn peak of COVID-19 cases in December 2020, listed rents decreased in all quality segments. Although the low-quality HRI dropped the most, it was also the first to rebuild itself in early 2021. The significant increase in rents observed in all segments in September 2021 may indicate this date as the time when the influence of the pandemic shock on the Polish rental market leveled.

Finally, Fig. 5 shows the unprecedented growth of rents for low-quality apartments in the first months of the war in Ukraine. In March 2022, this might be attributed solely to the increased demand driven by Ukrainian refugees. However, in September 2022, this coincided with seasonally increased student demand, resulting in the growth of rents higher than 25% q-q, being corrected only a little in December 2022. The growth of rents for high-quality apartments occurred quicker than for medium-quality apartments. Although both segments were not primarily the target of refugees, the owners of high-quality apartments may have had the ability to translate the increasing costs more quickly to asked rents (because consumer inflation and mortgage installments were rising).

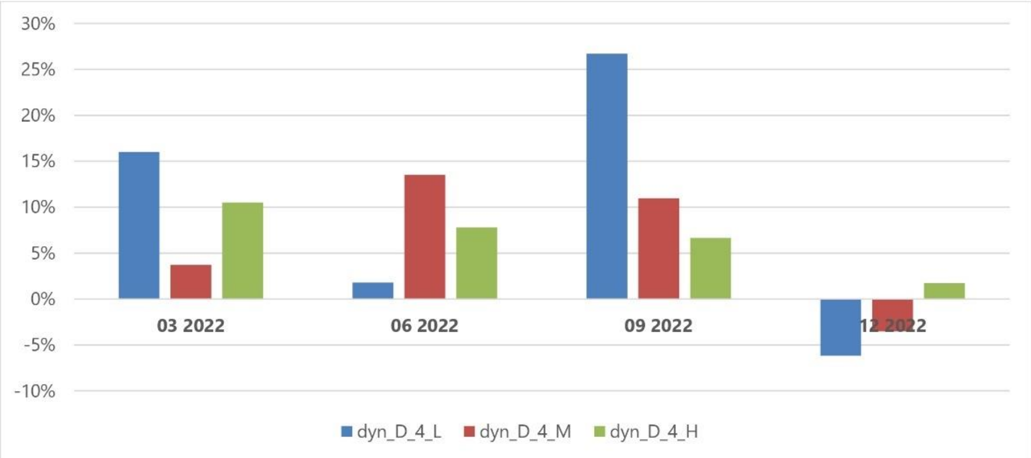


Fig. 5. Quarter-to-quarter dynamics of hedonic rent indices based on median QR models for quality-related segments during the war in Ukraine.
Source: own elaboration.

5. Conclusions

The study aimed to answer the question of whether it is justified to refer to hedonic models that do not account in any way for soft quality of apartments as "quality-adjusted". It has been shown that implementing quality-related variables to hedonic models enhances their statistical properties, and models that include the measure of quality extracted from textual descriptions of apartments or their photos have proven to be the best. Moreover, it has been documented that running the hedonic analysis exclusively on observations where quality is directly declared in listings may lead to biased results. Even though some approaches have been shown to be statistically superior, only modest differences between the indicated courses of HRIs have been detected. This confirms the results obtained by other researchers for the sales market on the low sensitivity of HRIs to the choice of hedonic method and the selection of additional variables (to the conventionally included ones). Bearing in mind that hedonic models based on listings data are constructed primarily for the needs of HRIs, it should be concluded that the quality adjustment of hedonic models based on hard, structural data may be considered sufficient.

As the paper's main contribution, one should consider the finding that utilizing soft quality-related information may be crucial to understanding the market's price changes, especially in times of turbulent market structure changes. The study has shown that, amid the COVID-19 pandemic and the war in Ukraine, the rent levels in the housing rental market in Poland have changed considerably, and the price responses to the shock have been unequal in different quality-related segments of the market. This kind of information on the market adjustment to economic shocks could not be obtained from the more commonly performed analysis of price-related segments. In this manner, running only the hedonic analysis of the whole market may be considered an oversimplification. Thus, pursuing the development of methods that will enable capturing the quality-related information in housing listings would allow for expanding the scope of the analysis rather than contribute to achieving better price/rent indices.

It is noteworthy that the indirect textual quality signals obtained from the descriptions of apartments processed using the Wordscores algorithm have proven to be the most meaningful for explaining the volatility of housing rents. Thus, they may be used to divide the market into quality-related segments. Although the method first requires constructing a dictionary of words and their quality scores based on the quality-labeled listings, it allows the scale and frequency of the analysis to be considerably increased once the dictionary is prepared. It should then be considered efficient compared to the precise but time-consuming visual inspection of photos attached to listings or relying on the declarations of sellers, which are neither precise nor complete.

It should be noted that the study has been limited to only one local market and the discussion on the course of HRIs is based on a seasonally unadjusted time series, and should therefore be considered introductory to the topic. To mitigate its limitations, it is recommended to confirm the results as soon as data with a longer time and geographical scope become available.

Funding Sources

The National Science Centre of Poland has supported this work under Grant number 2021/43/B/HS4/01213.

References

Akerlof, G. A. (1970). The market for "Lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*, 488–500. https://doi.org/10.2307/1879431

Ardila, D., Ahmed, A., & Sornette, D. (2021). Comparing ask and transaction prices in the Swiss housing market. *Quantitative Finance and Economics*, 5(1), 67–93. https://doi.org/10.3934/QFE.2021004

Beręsewicz, M. (2019). Correlates of representation errors in internet data sources for real estate market. *Journal of Official Statistics, 35*(3), 509–529. https://doi.org/10.2478/jos-2019-0022

Bian, X., Contat, J. C., Waller, B. D., & Wentland, S. A. (2021). Why disclose less information? Toward resolving a disclosure puzzle in the housing market. *The Journal of Real Estate Finance and Economics, 66*, 443–486. https://doi.org/10.1007/s11146-021-09824-6

Czerniak, A., & Rubaszek, M. (2018). The size of the rental market and housing market fluctuations. *Open Economies Review, 29*(2), 261–281. https://doi.org/10.1007/s11079-017-9452-1

de Wit, E. R., & Van der Klaauw, B. (2013). Asymmetric information and list-price reductions in the housing market. *Regional Science and Urban Economics, 43*(3), 507–520. https://doi.org/10.1016/j.regsciurbeco.2013.03.001

Diewert, W. E., & Shimizu, C. (2022). Residential Property Price Indexes: Spatial Coordinates Versus Neighborhood Dummy Variables. *Review of Income and Wealth, 68*, 770–796. https://doi.org/10.1111/roiw.12534

Elsinga, M., & Hoekstra, J. (2005). Homeownership and housing satisfaction. *Journal of Housing and the Built Environment, 20*(4), 401–424. https://doi.org/10.1007/s10901-005-9023-4

European Commission. Eurostat. Organisation for Economic Co-operation and Development. World Bank. (2013). *Handbook on Residential Property Prices (RPPIs): 2013 edition*. OECD Publishing. https://doi.org/10.2785/34007.

 eISSN: 2300-5289

Grossman, S. J. (1981). Warranties and private disclosure about product quality. *The Journal of Law & Economics, 24*(3), 461–483. https://doi.org/10.1086/466995

Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures? *Journal of Housing and the Built Environment, 38*, 2399–2425. Advance online publication. https://doi.org/10.1007/s10901-023-10043-z

Hill, R. J., & Trojanek, R. (2022). An evaluation of competing methods for constructing house price indexes: The case of Warsaw. *Land Use Policy, 120*, 106226. https://doi.org/10.1016/j.landusepol.2022.106226

Hotz, V. J., & Xiao, M. (2013). Strategic information disclosure: The case of multiattribute products with heterogeneous consumers. *Economic Inquiry, 51*(1), 865–881. https://doi.org/10.1111/j.1465-7295.2010.00340.x

James, R. N., III. (2007). Multifamily housing characteristics and tenant satisfaction. *Journal of Performance of Constructed Facilities, 21*(6), 472–480. https://doi.org/10.1061/(ASCE)0887-3828(2007)21:6(472)

Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association, 65*(330), 532–548. https://doi.org/10.1080/01621459.1970.10481102

Kim, S. S., Yang, I. H., Yeo, M. S., & Kim, K. W. (2005). Development of a housing performance evaluation model for multi-family residential buildings in Korea. *Building and Environment, 40*(8), 1103–1116. https://doi.org/10.1016/j.buildenv.2004.09.014

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica, 46*(1), 33–50. https://doi.org/10.2307/1913643

Koenker, R., & Machado, J. A. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association, 94*(448), 1296–1310. https://doi.org/10.1080/01621459.1999.10473882

Kolbe, J., Schulz, R., Wersing, M., & Werwatz, A. (2021). Real estate listings and their usefulness for hedonic regressions. *Empirical Economics, 61*, 3239–3269. Advance online publication. https://doi.org/10.1007/s00181-020-01992-3 PMID:33462521

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy, 74*(2), 132–157. https://doi.org/10.1086/259131

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *The American Political Science Review, 97*(2). Advance online publication. https://doi.org/10.1017/S0003055403000698

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *The Review of Corporate Finance Studies, 8*(1), 1–41. https://doi.org/10.1093/rcfs/cfy009

Liu, C. H., Nowak, A. D., & Smith, P. S. (2020). Asymmetric or Incomplete Information about asset values? *Review of Financial Studies, 33*(7), 2898–2936. https://doi.org/10.1093/rfs/hhz096

Lyons, R. C. (2019). Can list prices accurately capture housing price trends? Insights from extreme markets conditions. *Finance Research Letters, 30*, 228–232. https://doi.org/10.1016/j.frl.2018.10.004

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics, 12*(2), 380–391. https://doi.org/10.2307/3003562

Miłkowski, M. (2016). *Polimorfologik 2.1.* https://github.com/morfologik/polimorfologik/releases/tag/2.1

National Bank of Poland. (2023). *Badanie rynku nieruchomości - Instrukcja badania rynku wtórnego* (eng. *Real estate market research - Secondary market research instructions*).Retrieved from https://ssl.nbp.pl/publikacje/rynek_nieruchomosci/Instrukcja-badania-rynku-wtornego-2022.pdf

Nowak, A., & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics, 32*(4), 896–918. https://doi.org/10.1002/jae.2550

Office for Foreigners. (2023). Obywatele Ukrainy w Polsce – aktualne dane migracyjne (eng. *Ukrainian citizens in Poland – current migration data*), Retrieved from https://www.gov.pl/web/udsc/obywatele-ukrainy-w-polsce--aktualne-dane-migracyjne

Olszewski, K., Waszczuk, J., & Widłak, M. (2017). Spatial and Hedonic Analysis of House Price Dynamics in Warsaw, Poland. *Journal of Urban Planning and Development, 143*(3), 04017009. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000394

Poursaeed, O., Matera, T., & Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications, 29*(4), 667–676. https://doi.org/10.1007/s00138-018-0922-2

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55. https://doi.org/10.1086/260169

Rubaszek, M., & Rubio, M. (2020). Does the rental housing market stabilize the economy? A micro and macro perspective. *Empirical Economics, 59*(1), 233–257. https://doi.org/10.1007/s00181-019-01638-z

Samuelson, P. A. (1938). A Note on the pure theory of consumer's behaviour. *Economica, 5*(17), 61. https://doi.org/10.2307/2548836

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seo, Y., Im, J., & Mikelbank, B. (2020). Does the written word matter? The role of uncovering and utilizing information from written comments in housing ads. *Journal of Housing Research, 29*(2), 133–155. https://doi.org/10.1080/10527001.2020.1849929

Shen, L., & Ross, S. L. (2021). Information value of property description: A machine learning approach. *Journal of Urban Economics, 121*, 103299. Advance online publication. https://doi.org/10.1016/j.jue.2020.103299

Shimizu, C., Nishimura, K. G., & Watanabe, T. (2016). House prices at different stages of the buying/selling process. *Regional Science and Urban Economics, 59*, 37–53. https://doi.org/10.1016/j.regsciurbeco.2016.04.001

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics, 87*, 355–374. https://doi.org/10.2307/1882010

Trojanek, R., & Gluszak, M. (2022). Short-run impact of the Ukrainian refugee crisis on the housing market in Poland. *Finance Research Letters, 50*(103236), 103236. Advance online publication. https://doi.org/10.1016/j.frl.2022.103236

Trojanek, R., Gluszak, M., Hebdzynski, M. & Tanas, J. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838. https://doi.org/10.2307/1912934

Article 4	**Hebdzyński, M. (2024b). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies, 17*(2), 298-323. https://doi.org/10.14254/2071-8330.2024/17-2/16**

Hebdzyński, M. (2024). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies, 17*(2), 298-323. doi:10.14254/2071-8330.2024/17-2/16

Journal of International Studies

Centre of Sociological Research

Scientific Papers

Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland

Michał Hebdzyński
Department of Microeconomics,
Poznan University of Economics and Business,
Poland
michal.hebdzynski@phd.ue.poznan.pl
ORCID 0000-0003-0354-054X

Abstract. The COVID-19 pandemic that started in early 2020 and the Russian invasion of Ukraine in February 2022 caused multidimensional shocks to the housing market. Understanding their micro-level impact is crucial to optimizing future responses to similar shocks, designing sustainable urban and socio-economic policies, and investing. Based on the hedonic Spatial Error Model for the local housing rental market in Poland, during the pandemic, the valuation of the leisure-related apartment characteristics (the availability of a balcony and a private garden) increased. As tenants spent more time at home or in their neighbourhood, the proximity of housing to green areas became increasingly important, and the relevance of proximity to university buildings decreased. Then, amidst the war, a reluctance to use gas heating has been noticed. Combined with the observed price premium for the location of apartments in revitalised tenement houses, this means that the modernisation of the historic housing stock is not only ecologically desired, but also is capitalised in the achieved rents. The rent change throughout the pandemic has been estimated at -6.7%, while during the war-related crisis, at +29.7%. Finally, low sensitivity of hedonic rent indices to the detected changes in rent-setting factors has been found.

Received: April, 2023
1st Revision: May, 2024
Accepted: June, 2024

DOI: 10.14254/2071-8330.2024/17-2/16

Keywords: housing market, rental market, revealed preferences, COVID-19 pandemic, war in Ukraine, hedonic models, heating technology, renovation, revitalisation, hedonic rent index.

JEL Classification: R21, R32, C21, C81, D12

1. INTRODUCTION

A developed, long-term housing rental market reduces the volatility of the real estate market and contributes to the overall economic stability (Rubaszek & Rubio, 2020). However, in 2023, only 4.2% of Polish households rented apartments at market prices, much below the EU average of 20.2% (Eurostat, 2024a). The low share of renters may be attributed, among others, to the sale of state-owned apartments following the systemic changes in the last decades of the 20th century (Ronald, 2008). Nevertheless, in Poland, the strong preference for owning instead of renting is still present, which has been documented by Rubaszek & Czerniak (2017) and Bryx et al. (2021). Even though the transition from the "ownership society" (Ronald, 2008) to the "generation rent" (Ronald & Kadi, 2018; Byrne, 2020) may trigger the development of this vital segment of the economy, it has not been the only reason for the recent interest in the housing rental market.

The COVID-19 pandemic spreading globally since early 2020 induced an unprecedented, multidimensional shock to worldwide economies (Iyer & Simkins, 2022; Kholodilin & Rieth, 2023) and can be considered a super-shock, initiating structural changes (Dolnicar & Zare, 2020). It also affected the Polish rental market, which was dominated by young couples, students (Polityka Insight, 2022), and migrants (Narodowy Bank Polski, 2023). First, universities switched to distance learning, and students often decided to terminate the rental agreements in the cities where they studied and return to their family homes (Centrum AMRON, 2020). Secondly, the imposed cross-border traffic restrictions limited the influx of external migrants and suspended international tourism. Thus, many apartments rented on a short-term basis were converted to long-term rental purposes, increasing the long-term market supply (Boros et al., 2020; Marona & Tomal, 2020). Moreover, introducing lockdowns aimed at preventing COVID-19 transmission resulted in GDP declines for Poland estimated at -2% in 2020 (Eurostat, 2024b). To stimulate the demand, the European central banks decreased interest rates (in Poland – from 1.5% in 2020-02 to 0.1% in 2020-05). Hence, the greater availability of mortgages encouraged people to buy apartments instead of renting and to invest in rental apartments.

It may be assumed that in Poland, the effects of the pandemic shock lasted until the first quarter of 2022. First, in March, the COVID-related restrictions reached the lowest level since the beginning of the pandemic (Hale et al., 2021). Secondly, in February, Russia invaded Ukraine, starting a war. The number of refugees fleeing to European countries has equalled 6 million (UNHCR, 2023), while 1 million Ukrainians applied for temporal protection in the neighbouring country – Poland (Office for Foreigners, 2023). In Polish biggest cities, the resulting growth of population ranged from 12% to 53% (Wojdat & Cywiński, 2022), exerting a drastic demand shock to the housing rental market. The prices of energy resources rocketed (Ari et al., 2022), together with the costs of commodities and materials used in the construction sector. The crisis in the energy market prompted national governments to rethink their energy policies (Żuk & Żuk, 2022), as the supply shock quickly translated into general inflationary pressure. Moreover, the European central banks were forced to raise interest rates (in Poland – from 2.25% in 2022-02 to 6.75% in 2022-12), which elevated the perceived risk in housing investments and reduced the creditworthiness of potential buyers of apartments. Hence, it directed the demand from housing sales to the housing rental market.

On a macro scale, based on the theoretical model of DiPasquale & Wheaton (1992), the pandemic should have caused a decrease in rents, while the war-related crisis – rent increase. Similar changes were indicated in the empirical studies of the pandemic (Kuk et al., 2021; Tomal & Marona, 2021; Trojanek et al., 2021) and the war-induced refugee crisis (Gluszak & Trojanek, 2024; Trojanek & Gluszak, 2022). On the other hand, the consequences of recent economic shifts at the micro level have yet to be completely understood and tested. Regarding the pandemic, researchers discussed its several implications for lifestyle and housing preferences (Gallent & Madeddu, 2021; Mouratidis, 2021; Nanda et al., 2021), which was supported by survey studies (Marona & Tomal, 2020, 2023; Noszczyk et al., 2022). However, the quantitative evidence has been limited to date (Tomal & Helbich, 2022; Gamal et al., 2023; Guglielminetti et al., 2021). Concerning the war-related shock, the energy crisis and its implications may have induced a preference switch for some heating types, which has yet to be analysed.

The importance of the rental market, its structural changes triggered by the economic shifts, and the ongoing social transformation constitute a need to understand the new market rules. Moreover, in 2024, infectious diseases and natural resource shortages continue to be listed among the top global risks in both short and long-term perspectives (World Economic Forum, 2024). Thus, knowledge of how the housing market adjusts to economic shocks will allow us to understand its current state better and respond more effectively to similar shocks in the future.

First, the research aims to establish the marginal prices tenants pay for specific micro-level housing features using hedonic methods developed by Rosen (1974) based on Lancaster's (1966) consumer theory. Assuming consumers reveal the utility they attribute to goods through purchasing behaviour (Samuelson, 1948), the obtained marginal prices can be regarded as their revealed preferences. Besides the apartment characteristics included conventionally in hedonic housing models, the marginal prices paid for specific heating types have been estimated, together with the valuation of different scales of renovations of tenement buildings.

In its main part, the study has evaluated the impact of the pandemic- and war-related shocks on the micro-level valuations of rental apartments' characteristics. It was hypothesised that during the pandemic, the valuation of the availability of a balcony/terrace and private garden increased, together with the marginal price paid for the availability of an additional room for remote working or studying. Moreover, the spatial characteristics were considered, hypothesising that the value of proximity to urban green areas increased, and the value of proximity to the university buildings decreased. The attitude towards heating the rented apartment using natural gas was examined for the war-related shock, expecting a growing aversion. Finally, hedonic rent indices (HRI) have been calculated to assess the rent changes in the face of the shocks and to inspect the sensitivity of HRI to the shock-induced variability of rent-setting factors.

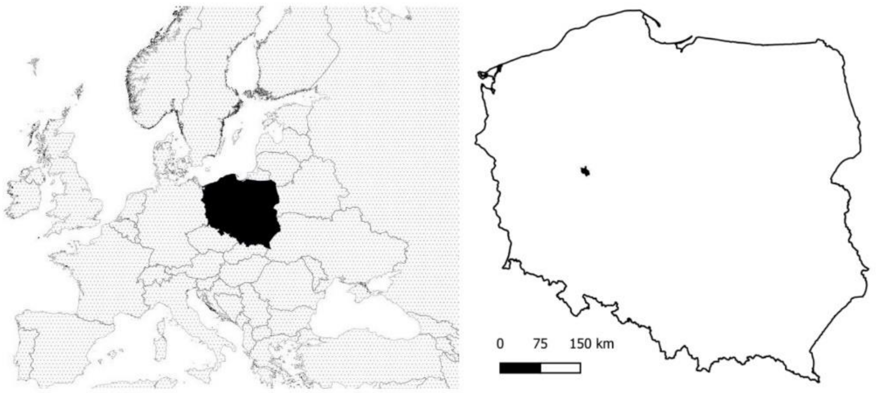


Figure 1. The location of Poland on the map of Europe (left panel). The location of Poznan on the map of Poland (right panel)

Source: own elaboration.

The hypotheses have been tested with the use of non-spatial (Ordinary Least Squares) and spatial hedonic models (Spatial Error Model) on the local market of Poznan – the fifth most populated Polish city, located in central Poland (its location has been presented in Figure 1). It serves not only as an academic destination but also is considered a business centre, attracting internal and external migrants. As a result, the rental market in Poznan is one of the most developed in Poland and constitutes a well-suited study area. Lastly, it was selected because of the author's knowledge of this market and the availability of data necessary to test the stated research hypotheses.

The research contributes to Rosen's (1974) hedonic price theory and adds to the literature on rent-setting factors, assessing the relationship between certain apartment characteristics and rents. In this regard, it extends the research of Hebdzyński (2024), who estimated the value of building renovations, and Sieger & Weber (2023) and Hahn et al. (2018), who described the impact of the type of apartment heating on rents. Then, it complements Tomal & Helbich's (2022) study of rent-setting factors' variability during the pandemic by testing hypotheses derived from qualitative studies – mainly Marona & Tomal (2020, 2023) and Nanda et al. (2021). As a result, it broadens the understanding of the adjustment process of consumer preferences to the pandemic and war-related shocks, which is important not only for investment purposes but also has implications for designing sustainable urban and socio-economic policies. In the case of hedonic rent indices, the study adds to Hill & Trojanek (2022), who compared different approaches to estimation and Hebdzyński (2024), who studied the sensitivity of HRI to the composition of variables included in hedonic models. The issue should be particularly important for public and private entities interested in tracking changes in rent levels using hedonic indices, which is the method recommended for this purpose by international institutions (European Commission, Eurostat, Organisation for Economic Co-operation and Development & World Bank, 2013).

The rest of the paper follows a structure: Section 2 reviews the literature on the topics covered by the study. Section 3 describes the data used and the chosen methodological approach to modelling. Section 4 outlines the empirical results and discusses the findings. Lastly, Section 5 concludes the manuscript, indicating its limitations and the field for further research.

2. LITERATURE REVIEW

2.1. Micro-level rent-setting factors

The existing literature on rent-setting factors concerns mostly the Chinese market, which is rapidly developing because of the increasing "floating population" – workers migrating temporarily within the country (Cui et al., 2018; Li et al., 2019; Liu et al., 2022; Zhan et al., 2023). A different strand of research has been devoted to establishing the impact of energy efficiency on rents in Germany (Cajias et al., 2019; Hahn et al., 2018; Kholodilin et al., 2017; Sieger & Weber, 2023). Other notable studies analysed the market of Switzerland (Baranzini & Ramirez, 2005; Crespo & Grêt-Regamey, 2013; Löchl & Axhausen, 2010), Greece (Efthymiou & Antoniou, 2013), the United Kingdom (McCord et al., 2014), the United States (Sirmans et al., 1989) and the Netherlands (Tomal & Helbich, 2023).

Among the Central and Eastern Europe countries, the research targeted to the best of our knowledge only the Polish market, focusing primarily on the rent-setting factors (Hebdzyński, 2024; Tomal, 2020; Tomal & Helbich, 2022, 2023) or on hedonic rent indices (Gluszak & Trojanek, 2024; Hebdzyński, 2024; Trojanek et al., 2021; Trojanek & Gluszak, 2022). Besides ordinary least squares (OLS), often used as a baseline approach in the hedonic analysis, authors employed a wide spectrum of methods to account for the Polish market specificity. Thus, their results will be discussed and used as a reference point. Gluszak & Trojanek (2024) constructed the micro-level hedonic models as a base to track the regional impact of the war-related migration on rents in five Polish biggest cities, utilizing a quasi-experimental variant of the difference-in-differences (DD) method. Hebdzyński (2024), based on the example of Poznan and Trojanek & Gluszak (2022), on the example of Warsaw and Krakow, treated the micro-level rent determinants as static in space, allowing them to differ across quality-related or price-related market segments using quantile regression (QR). Trojanek et al. (2021) added the spatial considerations on the geographically varying changes of rents in Warsaw amidst the early months of the pandemic and applied multiscale geographically weighted regression (MGWR). On the other hand, Tomal (2020) showed spatial autocorrelation of residuals in the OLS models for Krakow and explored rent determinants using spatial autoregressive model (SAR) and spatial autoregressive geographically weighted regression (GWR-SAR). Then, Tomal & Helbich (2022), in the study of Krakow, used geographically and temporally weighted regression (GTWR), which accounts for spatiotemporal non-stationarity of data. Finally, Tomal & Helbich (2023) used spatial autoregressive geographically weighted quantile regression (GWQR-SAR), allowing the model estimates for Warsaw to vary across space and the conditional distribution of rent levels.

The apartment characteristics that influence rents are often divided into structural and locational. As for the variables conventionally included in the hedonic models, it has been proven that floor area has a positive impact on rents, as does the number of rooms, although the latter did not always show statistical significance. Then, depending on the approach taken, the type of the building in which the apartment is located, its age or construction technology have proven to influence rents significantly. Tomal (2020) warned that when multiple approaches are used at once, the collinearity of explanatory variables might be present. Every additional floor of the building has proven to decrease rents, while the higher location of the apartment in the building and the presence of a lift are priced positively (Tomal, 2020; Tomal & Helbich, 2023). Furthermore, the additional spaces were analysed. Hebdzyński (2024) proved that the presence of the designated parking space, balcony and private garden (in some model configurations) individually increase rents in Poznan, but when analysed together, the features have proven to be of little or no significance in Krakow (Tomal, 2020; Tomal & Helbich, 2022). As for the apartment equipment, Hebdzyński (2024) showed that full furnishing may noticeably increase rent. Lastly, apartment quality has proven to be a significant determinant of rents, regardless of the form of its inclusion, which was in detail studied by Hebdzyński (2024).

Regarding locational characteristics, the distance from the building in which the apartment is located to the city centre was most often included in models, showing a negative influence on rents. However, in Krakow, in the model rich in variables reflecting distance to public amenities, this variable showed collinearity problems. At the same time, the distances to the nearest park, university, school or water reservoir have proven to affect rents negatively (Tomal, 2020). In Poznan (Hebdzyński, 2024), the distance to urban green areas has been found to decrease rents, while the distance to the lake has been insignificant. Finally, for Warsaw, based on Tomal & Helbich (2023), the distance to the university and to the nearest public transport stop showed statistical significance for the highest share of analysed observations, revealing their negative relation with rents.

There are also structural variables, rarely analysed at the micro level, which have been targeted in this research to improve understanding of both the current state of the market and the impact of recent economic shocks. The location of the apartments in the renovated tenement buildings in Poznan has proven to increase rents by 4% compared to blocks of flats and non-renovated tenements (Hebdzyński, 2024). However, to the best of our knowledge, the impact of the scale of renovation on rent has not been yet studied for the rental market. In 2021, 14.7% of all apartments in Poznan were located in buildings constructed before 1945 (GUS, 2024). Similar to the downtown area of Wroclaw (Poland) studied by Marcinkowska et al. (2015), most of Poznan's tenement houses were built using similar construction solutions from the second half of the 19th century until World War II, following the Industrial Revolution. Although their maintenance was careful before the war, in the post-war times of socialism, the necessary repairs were not provided, and the buildings have often deteriorated. Nevertheless, the increased demand for apartments in tenements has been noticed since the political transformation in Poland, which started in 1989 (Marcinkowska et al., 2015). Most recently, renovations of tenement houses have been taking place, and some of them are being revitalized. In the context of individual buildings in Poland, revitalisation should be perceived as a change or adaptation of the historical building to meet requirements similar to those imposed on newly constructed ones (Terlikowski, 2013). Although Bieda & Maniak (2024) did not find a clear effect of the revitalisation of an entire district in Krakow on apartment prices, it may be hypothesised that the relation is stronger on the level of individual buildings and depends on the scale of the renovation.

Secondly, the study has targeted the pricing of apartment heating types, which is connected with upgrading the existing stock of tenements across Poland. Originally, these buildings were heated with coal furnaces, but later, in cities, they have often been changed to gas-fuelled heating systems. Based on our analytical dataset (described in Section 3), they might be regarded as a dominant type of heating in rented apartments located in tenements in Poznan. Yet, the dataset shows that most apartments in revitalised tenements in Poznan have been connected to the district heating system. To date, studies of the rental markets have been mostly devoted to the more general issue of energy efficiency. Kholodilin et al. (2017) found that in Germany, it is capitalised in rents, but the value of future energy cost savings is larger than tenants' willingness to pay for better efficiency. Similarly, Cajias et al. (2019) found the rent premium for energy efficiency in the analysis of 403 local markets in Germany; however, it was not confirmed for metropolitan housing markets. Lastly, März et al. (2022) showed that although tenants pay a price premium for energy efficiency, it is small compared with other property features.

Hahn et al. (2018) divided heating systems into "brown", "standard", and "green" ones, where district heating was perceived as "green", apartments connected to central heating and heated with gas were treated as "standard", and those heated with coal or oil – "brown". The advantages of district heating were discussed by Mazhar et al. (2018), who considered it, among others, energy-efficient (thus cheaper), safer, environmentally friendly and space-efficient (as they require no bulky water boiler). Hahn et al. (2018) estimated the premium for using "green" technologies in Germany. They showed that it can be considered similar to the brown technology discount, equalling +/- 2.4% compared to the "standard" technologies. In the same model, the Authors also accounted for the energy performance of the apartments, which was the approach followed and extended by Sieger & Weber (2023), who estimated the marginal prices paid for particular heating types in the German market in years 2014-2020. District and central heating systems proved to be 0.4% - 0.8% more expensive, while electric heating – 3.7% - 5.5% cheaper than gas heating.

2.2. Micro-level impact of economic shocks

Marona & Tomal (2020, 2023) discussed the structural changes in preferences for apartment characteristics from the pandemic's beginning. The studies conducted in two phases of the pandemic in Krakow found that, respectively, 80% and 65% of surveyed real estate agents signalled the change in their clients' attitudes. Moreover, the Authors argued that it is highly likely that the changes will last longer than the pandemic and will become permanent. Among them, the increased demand of tenants for apartments with access to balconies or private gardens was indicated, highlighting the change in the ways of spending free time. It was confirmed in the survey study in Italy (Guglielminetti et al., 2021), where increased interest in apartments with private gardens was found.

On the other hand, Nanda et al. (2021) argued that as homes had to adapt to new roles – providing space for efficient work from home, the demand for an additional, separate room has increased. Mouratidis (2021) added considerations on the increased need for having larger, high-quality apartments that allow for comfortable leisure and enable to perform work-related tasks efficiently. It was also highlighted by agents surveyed by Marona & Tomal (2020). Moreover, Nanda et al. (2021), Gallent & Madeddu (2021), and Liu & Su (2021) argued that changes in working patterns reduced the importance of access to city centres, where business premises were traditionally located, which Tomal & Helbich (2022) empirically confirmed for the first phases of the pandemic. Gamal et al. (2023) and Tomal & Helbich (2022) found the decline in demand for apartments in dense, multi-unit buildings and linked it to the need to internalise the risk of the infection.

Similarly, Guglielminetti et al. (2021) showed that the stay-at-home orders in Italy may have increased demand for less congested areas.

Regarding other locational characteristics of apartments for rent, Tomal & Helbich (2022) found that the distance from the university buildings was gradually losing importance because of the introduction of online studies. Nanda et al. (2021) and Mouratidis (2021) discussed the recreational needs that can be satisfied by access to urban green areas, where the risk of infection was relatively lower. It was supported by the survey study of Krakow by Noszczyk et al. (2022), who added that during the pandemic, visits to green areas had a key role in citizens' mental health. Lastly, Broitman (2023) showed the growing willingness to live near urban green areas but warned that the increased prices of such located housing might result in the displacement of low-income residents and "ecological gentrification".

Concerning the preferences for heating types, we have found no study that linked them with economic shocks. However, one may see two reasons for the hypothesised increased aversion to gas heating after the beginning of the war in Ukraine. Firstly, one can name the rapid rise of natural gas prices. Ari et al. (2022) estimated that in 2022, the drastic change in the prices of fossil fuels contributed to the increase in the cost of living of European households by 7%, and the effect could persist until 2026. In Poland, the offered rents most often constitute only a fee for the property owner, not including any costs related to the use of the apartment, such as administrative fees and heating, electricity, or gas bills. On the other hand, as of 2021, around 80% of the natural gas consumed in Poland was imported (GUS, 2023), of which 56% was from Eastern Europe and Central Asia, mainly from the aggressor country – Russia (IGSMiE PAN, 2022). However, some gas supply contracts were terminated because of the tensions between Poland and Russia in the first months of the war in Ukraine (Balawender, 2022). As a result, tenants might have been concerned not only about the increase in gas costs that they would have to cover but also about possible shortages of this fuel.

Secondly, based on the pre-war survey, Rosak-Szyrocka & Żywiołek (2022) argued that Poles are generally little aware of the environmental damage caused by irresponsible energy consumption and of ways to save energy. After the outbreak of the war, Żuk & Żuk (2022) argued that countries could either strive to ensure the energy security or accelerate the energy transition. However, in the European Commission's (2023) study on climate change conducted more than a year from the beginning of the war, 43% of the surveyed Poles admitted that due to the war-induced energy crisis, the use of renewable energy sources should be increased, energy efficiency elevated and the transition to a green economy accelerated. Even if the share was below the EU average of 58%, it may be reflected in citizens' perceptions of heating types.

2.3. Macro-level impact of economic shocks

To study changes in rent level, one can use the theoretical model of DiPasquale & Wheaton (1992). In the four-quadrant analysis in which the real estate market is divided into the market for space and assets, the rent represents the former. It is situated at the core of the long-run analysis of the economic adjustment processes to economic shocks. In the model, the demand for space is assumed to come from tenants and owners who occupy their properties. Then, it relies on the current rent level and exogenous economic factors like number of households or their income. On the contrary, supply is linked with the market for assets, as it depends on real estate asset prices and construction costs. Although the model's framework focuses on individual economic shifts, the net effect of multiple simultaneous changes should reflect the combination of individual impacts.

Within the framework, the events accompanying the pandemic (described in Section 1) – the return of students to their family homes, the suspension of temporary labour migration, and the slowdown of economic activity may be perceived as factors that decrease the demand for space, pushing rents down.

Similarly, the interest rate reduction should contribute to a long-run rent decline. Finally, converting some short-term rental apartments to long-term rental purposes would constitute an exogenous shift in the housing supply, driving rents down. Therefore, rent reductions should be expected during the pandemic. It would be in line with the empirical studies for the early months of the pandemic – Trojanek et al. (2021) for Warsaw, Kuk et al. (2021) for the United States and Tomal & Marona (2021) for Krakow. Concerning the war-related economic shock, the influx of refugees drastically increased the number of households reporting demand for space. Moreover, because of the change in interest rates, an elevated number of Polish households showed interest in apartment rental, as they could not afford to take out mortgages. Thus, the demand pressure directed rents toward higher levels. From the supply side, the negative shift in the construction schedule of new apartments because of the rising costs of materials and energy resources also resulted in upward pressure on rents. Altogether, the observed demand and supply changes indicate that amidst the war in Ukraine, we should have experienced a rise in rent levels in Poland. The expected rent changes align with those indicated in the empirical study by Trojanek & Gluszak (2022) and Gluszak & Trojanek (2024) for Polish biggest cities in the first months of the war in Ukraine. Similar changes were reported by Balkan et al. (2018) regarding the impact of Syrian refugees on the Turkish rental market and Alhawarin et al. (2021), drawing on the experiences of Syrian migration to Jordan.

3. MATERIALS AND METHODS

3.1. Data

The research utilized a dataset of listings of apartments for long-term rental, located in multi-family buildings in Poznan. The listings were gathered quarterly, between 10th and 20th day of March, June, September and December, from 2019-06 to 2022-12, from two leading Polish apartment listing websites – Otodom.pl and Gratka.pl. The data were cleaned of duplicate observations – if the listing of the same apartment reappeared in adjacent periods (within 6 months range), only the last observation has been retained. This way, the listed rent would be as close as possible to the final, transacted rent. The data gaps were filled by exploring the descriptions and photos of listings. It allowed obtaining information on multiple apartment characteristics, which are often not signalled through the structured listings' forms, among others on building renovation, apartment furnishing, and availability of a balcony or a private garden. Then, based on expert market knowledge, the apartments with extreme rent or floor area were excluded, as well as those deemed unrepresentative, e.g., multi-storey apartments and those with gardens bigger than 75 m^2. It was assumed that the tenement buildings were built in Poznan until the 1960s and the newest buildings declared as tenements should be rather treated as blocks or apartment buildings (buildings designed to the highest standards and constructed from the best quality materials). Nevertheless, the average year of construction of the tenements included in the dataset was 1919. Regarding the heating type, the information provided in the listings was supplemented with expert knowledge, assuming that within the market of Poznan, the apartments located in high-quality apartment buildings and in the newest blocks are connected to the district heating system. Moreover, full textual descriptions of apartments were processed using the Wordscores algorithm (Hebdzyński, 2023, 2024; Laver et al., 2003) to obtain the proxy of apartment's quality.

The location of the listed apartments was rarely precisely indicated, thus the information that it is e.g., located "opposite to the XYZ restaurant" or "on a corner of" two streets was used. The observations were categorised into classes, according to the precision of information: up to 500, 250, 125 or 0 meters (the last value was assigned to the exact locations). The ones with lower precision were excluded from the dataset. As a result, the average precision of the location in the full dataset, which consisted of 7,768 observations, equalled 212 meters. Finally, in the study of the variability of rent-setting factors induced by the pandemic (which has targeted changes in marginal prices of locational factors), only the observations with the precision of up to 250 meters were used, which aimed at increasing the reliability of the results. Then, the average precision equalled up to 123 meters.

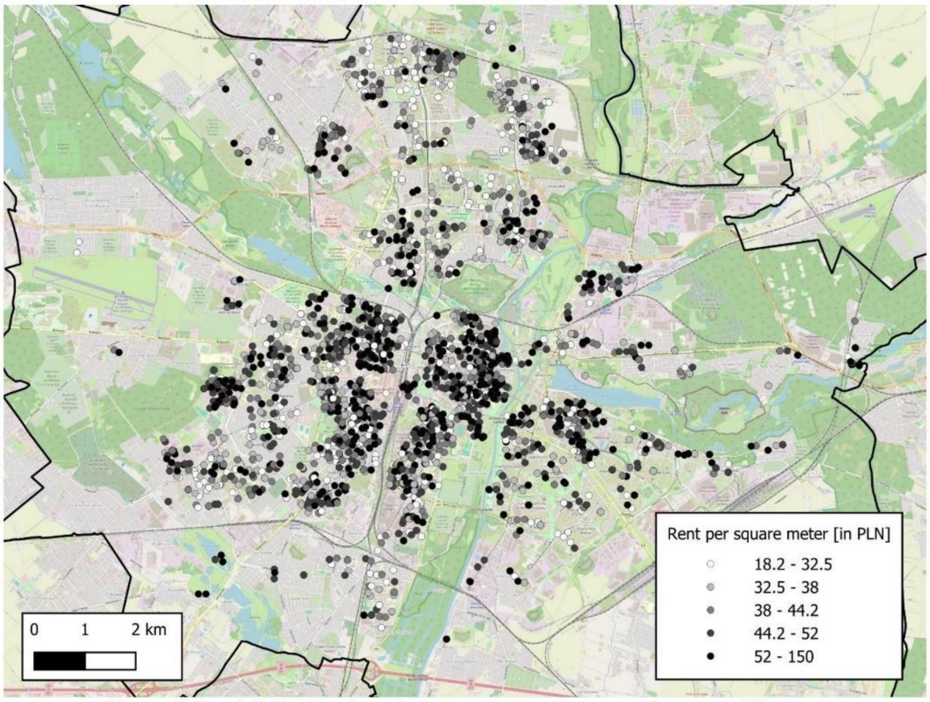


Figure 2. Spatial distribution of rent per square meter on the map of Poznan
Source: own elaboration based on OpenStreetMap

The listings have been used instead of transactions because of the impossibility of obtaining transactional data of the required quality, which is a frequent problem in European countries. In Poland, a high share of rental transactions is concluded without the intermediary of a real estate broker and there is no obligation to report them for purposes other than tax. Moreover, the data gathered by public institutions in the law-regulated process are not publicly available. That is why alternative data sources are often used, among which online listings are the most popular. Although they cannot be regarded as fully representative (Beręsewicz, 2015; Nasreen & Ruming, 2022) and they should be considered only a proxy of market transactions, they are superior regarding the availability of information on housing characteristics, which was utilised fully in this study. It has also been proven that listings-based hedonic indices of rents are highly correlated with transactional ones (Micallef, 2022) and may be considered accurate and leading indicators of transaction prices (Anenberg & Laufer, 2017).

3.2. Theory and calculation

The statistics of variables and tests discussed in this Section have been included in Appendix A.

To test the research hypotheses, the hedonic models were constructed using a logarithm of rent as a dependent variable and the variables presented in Table 1 as explanatory variables. The time-dummy method (Widłak & Tomczyk, 2010) was selected to construct HRI. The structure of the baseline hedonic model may be specified as:

$$lnR_i = \beta_0 + \sum_{j=1}^{J} \beta_j C_{i,j} + \sum_{k=2}^{K} \gamma_k D_{i,k} + u_i \tag{1}$$

where R_i is a rent for an i-th apartment, $C_{i,j}$ represents a value of j-th characteristic of an i-th apartment, β_j reflects the estimated marginal price of j-th characteristic, $D_{i,k}$ is a time dummy indicating whether an i-th apartment was listed in k-th period, γ_k is the estimated parameter reflecting change in rents in k-th period (compared to the base period) and u_i represents model's error.

Table 1

The variables used

Variable	Description
LN_RENT	dependent variable – natural logarithm of rent for the apartment [in PLN]
AREA	floor area of the apartment
ROOMS	number of rooms in the apartment
BLOCK	1 – for apartments located in blocks of flats, 0 – otherwise (base variable)
APART	1 – for apartments located in high-quality apartment buildings, 0 – otherwise
TENE_RENO	1 – for apartments located in renovated, non-revitalised tenements, 0 – otherwise
TENE_REVI	1 – for apartments located in revitalised tenements, 0 – otherwise
TENE_N_RENO	1 – for apartments located in non-renovated tenements, 0 – otherwise
QUALITY	soft quality of apartment obtained through processing of the listing description with the use of Wordscores algorithm, as in Hebdzyński (2023, 2024)
FURN	1 – for furnished apartments, 0 – otherwise
BALCONY	1 – for apartments with a balcony/terrace, 0 – otherwise
PARKING	1 – for apartments with access to the designated parking space, 0 – otherwise
GARDEN	1 – for apartments with access to the private garden (up to 75 m²), 0 – otherwise
DIST_CC	the distance of the apartment to the city centre [in km]
DIST_GREEN	the distance of the apartment to the closest urban green area [in km]
DIST_TRAM	the distance of the apartment to the closest tram station [in km]
DIST_UNI	the distance of the apartment to the closest university building [in km]
HEAT_DISTR *	1 – for apartments connected to the district heating, 0 – otherwise (base variable)
HEAT_GAS *	1 – for apartments heated by an individual gas furnace, 0 – otherwise
PRE-SHOCK	1 – for observations from the pre-shock period (2019-06 – 2020-03), 0 – otherwise
PAND	1 – for observations from the pandemic period (2020-06 – 2021-12), 0 – otherwise
WAR	1 – for observations from the war period (2022-03 – 2022-12), 0 – otherwise

* The information on heating type has been available for 5,672 out of 7,768 initially selected observations. *Source*: own elaboration.

First, an OLS model was estimated and outliers were excluded based on Cook's distance (Cook, 1977) using the threshold of $4/N$. Then, the restricted model (MOD_0) was tested. Based on the Author's knowledge of the market, the spatial similarity of rents was hypothesised. Thus, Moran's I test was performed, and positive spatial autocorrelation of the model's residuals was proven, pointing to residuals' clustering. However, Cliff & Ord (1970) argued that Moran's I test may give false positive results due to non-linear relationships between the variables or because of the omission of crucial regressors. Thus, multiple alternative model structures were tested in search of significant explanatory variables or their transformations (squares or logarithms). In this context, the results of RESET test of the model's specification (Ramsey, 1969) and AIC information criterion were considered in order to find the best model. The collinearity of the variables was checked using the VIF test to include only the non-collinear ones. Finally, the spatial distribution of the model's residuals was analysed visually.

As a result, no omitted variable suspected to influence the spatial distribution of residuals was found. Moreover, transforming explanatory variables to logarithmic forms did not lead to achieving better model results in terms of the RESET test and AIC. Therefore, the decision to use spatial methods was made. Due to the heteroscedasticity of residuals detected in the OLS model, its final version has been constructed using a heteroscedasticity-consistent variance estimator (White, 1980).

In the next step, the Lagrange Multiplier test statistics were calculated and proven statistically significant in all test variants. Therefore, it was decided that the MOD_1 model would be constructed in two variants – spatial error model (SEM) and spatial lag model (SLM), and the final decision regarding the estimation method and the way of constructing the weighting matrix would be made based on the AIC criterion. A matrix of spatial weights for neighbours in the proximity of 250 meters was selected as a starting one because it was the minimum value that exceeded the average precision of the location of the analysed apartments. Of all tested model variants, the SEM model, which used the matrix of proximity at a distance of 250 meters, proved to be the best. Hence, it was selected as a final approach to construct models. The main assumption of SEM is that in addition to modelling the parameters presented in Equation 2, the error term is also modelled, and:

$$u = \lambda W u + e \text{ , while } e \sim N(0, \sigma^2 I_n) \text{ ,} \tag{2}$$

where λ is a parameter of autocorrelation, W is a matrix of spatial weights for neighbours in the proximity of 250 meters, Wu is a spatially lagged error of the model, and e is an independent error.

To capture the rent-setting factors' variability, the models with interaction variables and restricted to specific time periods were constructed. The observations from the PRE-SHOCK period (pre-pandemic) were considered the control group, and the observations from the PAND or WAR period constituted the treatment group. Finally, the following models were prepared:

- MOD_0 – a baseline, non-spatial model, not accounting for changes in rent-setting factors,
- MOD_1 – a spatial model, not accounting for changes in rent-setting factors, based on full time range; MOD_2 – an analogous model, but restricted to observations with a known heating type (district or gas heating),
- MOD_3 – a spatial model based on the dataset restricted to the PRE-SHOCK (base period) and PAND periods, including interactions of PAND with: BALCONY, GARDEN, ROOMS, DIST_GREEN and DIST_UNI (the initially selected interaction of PAND and DIST_CC showed problem with collinearity, hence it was excluded from considerations); MOD_3_BASE – an analogous model with no interaction variables,
- MOD_4 – a spatial model based on the dataset restricted to the PRE-SHOCK (base period) and WAR periods, and observations with a known heating type (district or gas heating), including interaction of WAR with HEAT_GAS; MOD_4_BASE – an analogous model with no interaction variables.

For the needs of studying the pandemic-related change in rents, the MOD_3 models were used. As the MOD_4 models utilized non-randomly selected observations, no HRI was constructed based on them. Instead, the analysis of the rent changes amidst the war was conducted using the MOD_1 results. Lastly, the sensitivity of HRI to the changes in rent-setting factors was studied by analysing the differences between HRI derived from the MOD_3 and MOD_3_BASE models.

4. EMPIRICAL RESULTS AND DISCUSSION

4.1. Models' verification

The results of the discussed statistical tests and plots have been included in Appendix B.

Table 2

The models' results

	MOD_0	MOD_1	MOD_2	MOD_3	MOD_4
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
AREA	0.0079 ***	0.0080 ***	0.0080 ***	0.0086 ***	0.0083 ***
ROOMS	0.088 ***	0.086 ***	0.092 ***	0.080 ***	0.096 ***
PAND x ROOMS				-0.006	
APART	0.037 ***	0.032 ***	0.023 ***	0.037 ***	0.022 ***
TENE_RENO	0.019 *	0.018 *	0.020	0.037 ***	-0.023
TENE_REVI	0.040 ***	0.039 ***	0.027 **	0.041 ***	0.011
TENE_N_RENO	-0.001	-0.001	-0.001	0.014 *	-0.021
QUALITY	3.009 ***	2.901 ***	2.953 ***	2.858 ***	3.064 ***
FURN	0.045 ***	0.045 ***	0.049 ***	0.042 ***	0.053 ***
BALCONY	0.034 ***	0.033 ***	0.032 ***	0.027 ***	0.012
PAND x BALCONY				0.018 *	
PARKING	0.036 ***	0.034 ***	0.031 ***	0.035 ***	0.029 ***
GARDEN	0.016 **	0.011	0.012	-0.039	0.005
PAND x GARDEN				0.053 *	
DIST_CC	-0.0094 ***	-0.0103 ***	-0.0104 ***	-0.0101 ***	-0.0091 ***
DIST_GREEN	-0.055 ***	-0.058 ***	-0.063 ***	-0.026	-0.068 ***
PAND x DIST_GREEN				-0.049 **	
DIST_TRAM	-0.003	-0.001	-0.007	-0.006	-0.003
DIST_UNI	-0.009 ***	-0.009 ***	-0.006	-0.003	-0.009 *
PAND x DIST_UNI				-0.004	
HEAT_GAS			-0.011		0.012
WAR x HEAT_GAS					-0.040 *
TIME-DUMMIES	YES	YES	YES	YES	YES
CONSTANT	6.390 ***	6.409 ***	6.392 ***	6.394 ***	6.373 ***
METHOD	**OLS**	**SEM**	**SEM**	**SEM**	**SEM**
R^2	**0.828**				
AIC	**-10,173**	**-10,292**	**-7,595**	**-6,012**	**-3,211**
N	**7,391**	**7,391**	**5,381**	**4,342**	**2,344**

Full models' results have been included in Appendix C. *** for p-value < 0.01; ** for p-value < 0.05; * for p-value < 0.1. *Source*: own elaboration.

The results of the constructed models have been presented in Table 2. First, based on R^2 of the MOD_1 model (standard R^2 has equalled 0.828, while adjusted R^2 – 0.827), it should be inferred that the model explains most of the variation of rents, i.e. the major rent-setting factors were accounted for. Compared to the analogous spatial model – MOD_1, the latter should be considered superior as it shows a lower value of AIC. The visual analysis of the spatial distribution of model residuals did not reveal any clear clustering pattern. However, in Poznan's central districts, the absolute values of residuals seem to be bigger. Thus, the models' residuals were examined using tests of residuals' homoscedasticity (Breusch & Pagan, 1979) and the normality of their distribution (Jarque & Bera, 1980). Firstly, the null hypothesis of homoscedasticity has been rejected in all cases, which may also be caused by autocorrelation of residuals (Kopczewska, 2006). Secondly, most of the models have had problems with heavy tails. Yet, in the models that are based on the restricted dataset (models with interaction variables), the tails of the residual distribution either have been closer to the quantiles of normal distribution (MOD_3) or may be assumed to be normally distributed (MOD_4). Nevertheless, the detected heteroscedasticity may implicate the spatial variability of rent-setting factors. Thus, when interpreting the models' results, the focus should be placed on the changes in rent-setting factors rather than on the analysis of their precise levels.

4.2. Discussion

4.2.1. Micro-level rent determinants and their changes

First, it should be noted that the variables included in hedonic models have shown the expected direction of influence on rents in the MOD_1 model for the full-time range. The apartment area, together with its number of rooms, interior quality and furnishing should be considered the major structural rent-setting factors. Then, the availability of a balcony, designated parking space and private garden has been shown to increase rents. However, the last feature has been insignificant. In line with the expectations, all the analysed locational and distance variables impacted rents negatively – the bigger the distance from the city centre, university buildings or green areas ceteris paribus, the cheaper the apartments in Poznan. A similar direction of influence has been achieved for the proximity to public transport (tram), but, similarly to Krakow (Tomal, 2020), it has shown no significance. Supposedly, it may indicate that the relationship between public transport and rents is more complex and requires more data of greater precision that will allow us to study multiple modes of transport and their both positive and negative externalities.

More attention should be paid to the issue of building types. Based on the MOD_1 and MOD_2 models, the apartments located in revitalised tenements have proven to be 2.7 - 4% more expensive than the ones in blocks ceteris paribus, while the marginal price of the location in a high-quality apartment building has been lower and estimated at 2.4 - 3.3%. Thus, the location in historical buildings that have undergone revitalisation may be considered to provide a price premium. At the same time, the apartments in renovated but non-revitalised tenements also have shown to be more expensive than blocks – by around 2%, and the rents in non-renovated tenements have not statistically differed from the ones in blocks. Thus, although in the case of entire districts, revitalisation proved to be insignificant for price formation in Krakow (Bieda & Maniak, 2024), in Poznan, the performed renovation of the building and its scale have shown to be important factors for determining rents at the level of individual apartments. Importantly, in hedonic models for the full-time range, the heating types have shown no significant impact on individual rents, counter to the previous studies by Hahn et al. (2018) and Sieger & Weber (2023).

Secondly, most hypotheses concerning the changes in rent-setting factors during the pandemic have been confirmed by looking at the MOD_3 and MOD_3_BASE models. As hypothesised, the availability of a balcony or terrace increased rents by 2.8%, which was amplified during the pandemic by an additional 1.8 p.p. Concerning the availability of a small private garden, its impact on rents did not prove to be significant for the pre-pandemic period, however, the weak significance of change during the pandemic has been found. Thus, the overall impact of this variable on rents during the pandemic may be assumed to lie within the range of 1.5-5.4%. It confirms the findings of Marona & Tomal (2020, 2023) and Guglielminetti et al. (2021)

and shows that home-related leisure increased its value for tenants amidst the health crisis and lockdowns. As for the valuation of an additional room for remote working or studying, the hypothesized change discussed by Nanda et al. (2021) has not been proven. It may mean that the increased need for an extra room for work-related purposes could have been balanced by the need for having more spacious rooms for better relaxation and general well-being, as supposed by Guglielminetti et al. (2021) and Mouratidis (2021).

When it comes to locational factors, the impact of proximity to urban green areas on rents was significant in the joint analysis of the pre-pandemic and pandemic periods. Although the variable has been expressed in kilometres, its maximum value has been lower than one kilometre. Thus, it would be more intuitive to analyse the change in rents for 100 m of additional distance. In this regard, the impact should be estimated at -0.5 p.p. for every 100 m. However, when accounting for the pandemic-induced variability, the overall impact may be regarded as -0.23 p.p. (which has proven to be insignificant), while the additional discount during the pandemic has been measured at an additional -0.39 p.p. The noticed change is in line with the survey study by Noszczyk et al. (2022) and confirms the special role assigned to urban green areas during the health-related crisis and lockdowns. Moving to the analysis of the distance from university buildings, in the MOD_1 model conducted on the full-time range, an additional 1 km ceteris paribus has proven to significantly decrease rents by 0.9%. However, in the pandemic period the impact has shown to be not significantly different from 0. It implies that during the pandemic other factors than the proximity to university buildings were taken into account. It confirms the findings by Tomal & Helbich (2022) for Krakow and shows that the students' transition to remote or hybrid education was quickly reflected also in changes in the revealed preferences of tenants.

Finally, the analysis of the MOD_4 model has confirmed the hypothesized growth of the aversion towards gas heating after the outbreak of the war in Ukraine. Although earlier the marginal price of this feature was not statistically different from the value assigned to district heating, the relation changed amidst the war. The negative impact of gas heating on rents in the first four quarters of the war in Ukraine has been estimated at 2.7 - 3.9%. It may be considered a sign of change in the perception of this fuel in the times of the war-related increased volatility of prices of fossil fuels (Ari et al., 2022). Furthermore, it may be a symptom that the acceleration of energy transition discussed by Żuk & Żuk (2022), and studied by the European Commission (2023) is taking place on the micro level in Poland. Thus, it should be taken into account in the valuation of future investment processes, especially those connected with the renovations of the tenement buildings, in which gas heating is often used.

4.2.2. Hedonic rent indices and their sensitivity

Based on Figure 3 we may observe the dynamics of HRI in the pre-pandemic and pandemic periods, constructed with the use of estimates from the MOD_3 and MOD_3_BASE models. When comparing the models using the AIC criterion, it can be concluded that the MOD_3 model, which takes into account changes in the rent-setting factors, should be considered superior. Based on it, the mean absolute quarter-quarter dynamics of rents equalled 2.88%, while the mean absolute dynamics' difference – 0.19 p.p. Thus, although the changes in multiple rent-setting factors have been noticed, they have had little influence on the course of HRI in the short-term. This is in line with the study by Hill & Trojanek (2022), who noticed only slight differences between the apartment price indices obtained using the standard time-dummy method and the rolling-time dummy method (which allows for parameter changes in time). Moreover, it is consistent with the results by Hebdzyński (2024) on the little impact on the HRI of minor changes in the composition of hedonic models.

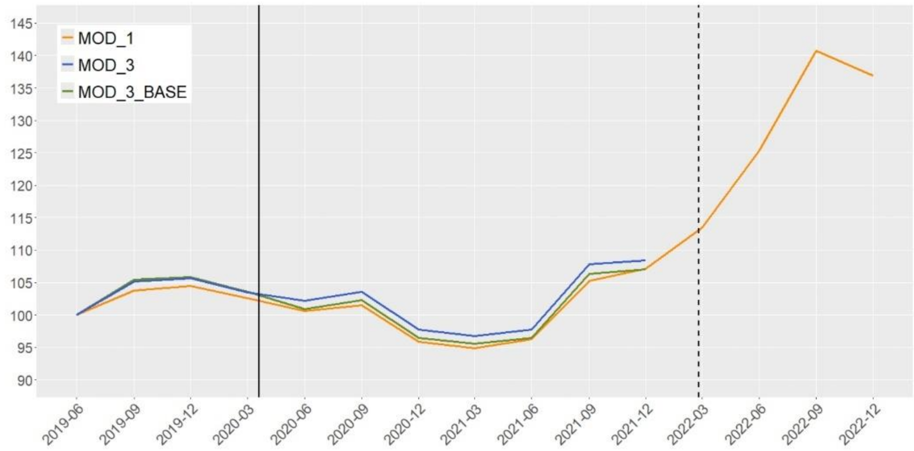


Figure 3. Hedonic rent indices (2019-06 = 100)

The solid vertical line represents the beginning of the pandemic in Poland (first lockdown – 2020-03-20); the dashed vertical line represents the beginning of the war in Ukraine (2022-02-24). *Source:* own elaboration.

Regarding the rent changes indicated by the models, it may be implied that rents in Poznan decreased by 6.7% during the pandemic, while for four quarters after the outbreak of the war the increase equalled 29.7%. The estimated changes in HRI agree with the expectations resulting from the theoretical analysis of both economic shocks at the macro level and with the earlier empirical evidence (Trojanek et al., 2021; Kuk et al., 2021; Tomal & Marona, 2021; Trojanek & Gluszak, 2022; Gluszak & Trojanek, 2024).

5. CONCLUSIONS

The study aimed to establish the factors determining rents at the micro level on the local rental market in Poznan, track their changes in the face of the recent economic shocks, and assess the sensitivity of hedonic rent indices to these changes. In the empirical part of the analysis, the hypotheses derived from the discussion papers and qualitative studies have been tested using spatial hedonic methods. It has been shown that in Poznan, individual apartment features influence rents similarly to other Polish markets, as described in previous research. However, in addition to the characteristics typically included in the hedonic models of housing, the revealed preferences for the types of buildings in which apartments are located have been examined. It has been shown that the pricing of tenements varies depending on whether the building has been renovated and on the scale of the renovation. Importantly, the location of the apartment in the revitalised tenement has been valued the most by tenants among all studied building types. In the next step, the changes in rent-setting factors have been considered. First, the structural characteristics of the apartment (the availability of a balcony or a private garden) were tested to find the increased valuation of leisure-related housing features during the pandemic. Yet, the similar change has not been proven for the marginal price paid for a separate room for remote working or studying. Concerning the valuation of locational characteristics amidst the pandemic, the strength of the impact of the proximity to urban green areas on rents has increased, while the significance of the distance to university buildings diminished. As for the shock caused by the war and the accompanying crisis in the energy market, it has been shown that it could also have had an impact on consumers at the micro level. Although earlier they valued similarly the district and gas heating, the latter has become a negative factor during the first year of the war in Ukraine. Finally, the hedonic rent indices have been estimated, pointing to the modest decrease in rents during the pandemic and the drastic increase amidst the war-related crisis. The models that allowed for the changes in rent-setting factors have proven to be superior, nevertheless, the impact of the changes on the achieved dynamics of the index has been assessed as small.

It can be expected that changes in preferences for particular housing characteristics reported by consumers since the beginning of COVID-19 will become permanent (Marona & Tomal, 2023). Furthermore, the energy crisis, which has been caused on the one hand by the war in Ukraine and on the other hand by the adjustment of energy policy at the national and international level, also has the potential to change consumers' attitudes. Apart from their reaction to the increased volatility of gas prices on the global markets, which may be of a temporary nature, the perception of the switch towards greener energy might have changed permanently. Even though the housing market has already responded to the described shocks on both the micro and macro levels, infectious diseases, international armed conflicts, and resource shortages continue to be listed among the most important global risks. Therefore, examining the micro-level impact of economic shocks related to health, war and natural resources may help private and public entities to optimize their reaction to future shocks. Moreover, the responsible urban planning decisions aimed at stimulating the development of the housing rental market, which is a crucial economic sector, should incorporate the post-shock preferences of citizens. It would be profitable from both the perspective of satisfaction of the city's residents (because of meeting their housing needs) and from the perspective of economic efficiency of investments. Particularly, providing changes in the heating systems of older, energy-inefficient buildings, is not only desired in a way to reduce the consumption of non-renewable fuels but also proves to be positively priced by tenants. The same applies to the revitalisation of the existing housing stock, which results in the rent premium for the location of the apartment in the historical building.

Although the study utilized the dataset of listings that fully reflects the potential of this data source, it should be noted that the obtained marginal prices have relied on the models built based on listings providers' declarations. Therefore, the results constitute only a proxy of the truly revealed preferences that are possible to be obtained based on transactional data. Furthermore, supposedly also due to the limited size of the Poznan market and the resulting moderate number of observations, in some cases the statistical significance of indicated changes has been relatively weak (p-values < 0.1). Additionally, the study considers only a short-term change of rent determinants, not attempting to answer the question of the durability of the changes. To answer the above weaknesses, it is recommended to repeat the study as soon as transactional data, representing longer time scope and including more observations would become available. On the other hand, the results of the statistical tests point to the possible spatial variability of rent determinants. Thus, studying rent-setting factors in different spatial regimes (e.g. centrally vs. non-centrally located apartments) might reveal more information on the analysed phenomena. However, the primary objective of the study was not to precisely estimate the marginal prices of housing features, but rather to indicate their relative importance and changes over time. Thus, the SEM method, which assumes spatial homogeneity of estimates, was used to facilitate the interpretability of the results. Yet, special care should be taken when making predictions based on the models' results. Finally, we believe that each of the identified micro-level changes may require separate investigation to fully understand its nature, which we recommend as an area of research for future studies.

ACKNOWLEDGEMENT

The National Science Centre of Poland has supported this work under Grant number 2021/43/B/HS4/01213. The author has no other relevant financial or non-financial interests to disclose. The datasets generated and analysed during the study are available from the author on reasonable request.

REFERENCES

Alhawarin, I., Assaad, R., & Elsayed, A. (2021). Migration shocks and housing: Short-run impact of the Syrian refugee crisis in Jordan. *Journal of Housing Economics, 53*(101761). https://doi.org/10.1016/j.jhe.2021.101761

Anenberg, E., & Laufer, S. (2017). A more timely house price index. *Review of Economics and Statistics, 99*(4), 722–734. https://doi.org/10.1162/REST_a_00634

Ari, A., Celasun, O., Mineshima, A., Arregui, N., Mylonas, V., Teodoru, I., Black, S., Zhunussova, K., Iakova, D., & Parry, I. (2022). Surging Energy Prices in Europe in the Aftermath of the War: How to Support the Vulnerable and Speed up the Transition Away from Fossil Fuels. *IMF Working Papers, 2022*(152). https://doi.org/10.5089/9798400214592.001

Balawender, G. (2022). Nieoficjalnie: Rosja zakręca Polsce kurek z gazem. *Rzeczpospolita*. https://energia.rp.pl/gaz/art36152141-nieoficjalnie-rosja-zakreca-polsce-kurek-z-gazem [access: 04-06-2024]

Balkan, B., Tok, E., Torun, H., & Tumen, S. (2018). Immigration, Housing Rents, and Residential Segregation: Evidence from Syrian Refugees in Turkey. *IZA Institute of Labor Economics Discussion Paper Series, 11611*. https://doi.org/10.2139/ssrn.3205899

Baranzini, A., & Ramirez, J. V. (2005). Paying for quietness: The impact of noise on Geneva rents. *Urban Studies, 42*(4), 633–646. https://doi.org/10.1080/00420980500060186

Beręsewicz, M. (2015). On the representativeness of internet data sources for the real estate market in Poland. *Austrian Journal of Statistics, 44*(2), 45–57. https://doi.org/10.17713/ajs.v44i2.79

Bieda, A., & Maniak, W. (2024). The Effect of Revitalisation on the Value of Residential Premises in Historic Urban Layouts. *Real Estate Management and Valuation, 32*(1), 95–113. https://doi.org/10.2478/remav-2024-0011

Boros, L., Dudás, G., & Kovalcsik, T. (2020). The effects of COVID-19 on airbnb. *Hungarian Geographical Bulletin, 69*(4), 363–381. https://doi.org/10.15201/hungeobull.69.4.3

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963

Broitman, D. (2023). "Passive" Ecological Gentrification Triggered by the Covid-19 Pandemic. *Urban Planning, 8*(1), 312–321. https://doi.org/10.17645/up.v8i1.6015

Bryx, M., Sobieraj, J., Metelski, D., & Rudzka, I. (2021). Buying vs. Renting a Home in View of Young Adults in Poland. *Land, 10*(11). https://doi.org/10.3390/land10111183

Byrne, M. (2020). Towards a political economy of the private rental sector. *Critical Housing Analysis, 7*(1), 103–113. https://doi.org/10.13060/23362839.2020.7.1.507

Cajias, M., Fuerst, F., & Bienert, S. (2019). Tearing down the information barrier: the price impacts of energy efficiency ratings for buildings in the German rental market. *Energy Research & Social Science, 47*, 177–191. https://doi.org/10.1016/j.erss.2018.08.014

Centrum AMRON. (2020). *Studenci na rynku nieruchomości 2020*. https://www.amron.pl/strona.php?tytul=studenci-na-rynku-nieruchomosci [access: 04-06-2024]

Cliff, A. D., & Ord, K. (1970). Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography, 46*(sup1), 269–292. https://doi.org/10.2307/143144

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics, 19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Crespo, R., & Grêt-Regamey, A. (2013). Local hedonic house-price modelling for urban planners: Advantages of using local regression techniques. *Environment and Planning B: Planning and Design, 40*(4), 664–682. https://doi.org/10.1068/b38093

Cui, N., Gu, H., Shen, T., & Feng, C. (2018). The impact of micro-level influencing factors on home value: A housing price-rent comparison. *Sustainability, 10*(12), 1–23. https://doi.org/10.3390/su10124343

DiPasquale, D., & Wheaton, W. C. (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework. *Real Estate Economics, 20*, 181–198. https://doi.org/10.1111/1540-6229.00579

Dolnicar, S., & Zare, S. (2020). COVID19 and Airbnb – Disrupting the Disruptor. *Annals of Tourism Research, 83*(102961). https://doi.org/10.1016/j.annals.2020.102961

Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice, 52*, 1–22. https://doi.org/10.1016/j.tra.2013.04.002

European Commission, Eurostat, Organisation for Economic Co-operation and Development, World Bank. (2013). *Handbook on Residential Property Prices (RPPIs) : 2013 edition.* OECD Publishing. https://doi.org/10.2785/34007

European Commission. (2023). *Climate Change, Special Eurobarometer 538.* https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=88207 [access: 04-06-2024]

Eurostat. (2024a). *Distribution of population by tenure status, type of household and income group - EU-SILC survey.* https://doi.org/10.2908/ILC_LVHO02

Eurostat. (2024b). *Real GDP growth rate.* https://doi.org/10.2908/TEC00115

Gallent, N., & Madeddu, M. (2021). Covid-19 and London's Decentralising Housing Market–What are the Planning Implications? *Planning Practice and Research, 36*(5), 567–577. https://doi.org/10.1080/02697459.2021.1964782

Gamal, A., Rohmah, L., & Muhyi, M. M. (2023). Housing preference shifting during COVID-19 pandemic in Indonesia. *Journal of Urban Management, 12*(3), 268–283. https://doi.org/10.1016/j.jum.2023.05.002

Gluszak, M., & Trojanek, R. (2024). War in Ukraine, the refugee crisis, and the Polish housing market. *Housing Studies*, https://doi.org/10.1080/02673037.2024.2334822

Guglielminetti, E., Loberto, M., Zevi, G., & Zizza, R. (2021). Living on my Own: The Impact of the Covid-19 Pandemic on Housing Preferences. *Bank of Italy Occasional Paper No. 627.* https://doi.org/10.2139/ssrn.3891671

GUS. (2023). *Energia (Energy).* https://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5485/1/11/1/energia_2023.pdf

Hahn, J., Hirsch, J., & Bienert, S. (2018). Does "clean" pay off? Housing markets and their perception of heating technology. *Property Management, 36*(5), 575–596. https://doi.org/10.1108/PM-08-2017-0051

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour, 5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures? *Journal of Housing and the Built Environment, 38*, 2399–2425. https://doi.org/10.1007/s10901-023-10043-z

Hebdzyński, M. (2024). Are Hedonic Models Really Quality-Adjusted? The Role of Apartment Quality in Hedonic Models of Housing Rental Market. *Real Estate Management and Valuation, 32*(2), 46–57. https://doi.org/10.2478/remav-2024-0014

Hill, R. J., & Trojanek, R. (2022). An evaluation of competing methods for constructing house price indexes: The case of Warsaw. *Land Use Policy, 120*, 106226. https://doi.org/10.1016/j.landusepol.2022.106226

IGSMiE PAN. (2022). *Gospodarka surowcami mineralnymi w Polsce w latach 2012 – 2021.* https://min-pan.krakow.pl/wydawnictwo/wp-content/uploads/sites/4/2023/06/Rocznik-GSM-2022.pdf

Iyer, S. R., & Simkins, B. J. (2022). COVID-19 and the Economy: Summary of research and future directions. *Finance Research Letters, 47*(PB), 102801. https://doi.org/10.1016/j.frl.2022.102801

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5

Kholodilin, K. A., Mense, A., & Michelsen, C. (2017). The market value of energy efficiency in buildings and the mode of tenure. *Urban Studies, 54*(14), 3218–3238. https://doi.org/10.1177/0042098016669464

Kholodilin, K. A., & Rieth, M. (2023). Viral shocks to the world economy. *European Economic Review, 158*, 104526. https://doi.org/10.1016/j.euroecorev.2023.104526

Kopczewska, K. (2006). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN* (Wydanie I). CeDeWu.

Kuk, J., Schachter, A., Faber, J. W., & Besbris, M. (2021). The COVID-19 Pandemic and the Rental Market: Evidence From Craigslist. *American Behavioral Scientist, 65*(12) https://doi.org/10.1177/00027642211003149

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy, 74*(2), 132–157. https://www.jstor.org/stable/1828835

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review, 97*(2). https://doi.org/10.1017/S0003055403000698

Li, H., Wei, Y. D., & Wu, Y. (2019). Analyzing the private rental housing market in Shanghai with open data. *Land Use Policy, 85*, 271–284. https://doi.org/10.1016/j.landusepol.2019.04.004

Liu, Q., Zhao, P., Xiao, Y., Zhou, X., & Yang, J. (2022). Walking Accessibility to the Bus Stop: Does It Affect Residential Rents? The Case of Jinan, China. *Land 11*(6). https://doi.org/10.3390/land11060860

Liu, S., & Su, Y. (2021). The impact of the COVID-19 pandemic on the demand for density: Evidence from the U.S. housing market. *Economics Letters, 207*, 110010. https://doi.org/10.1016/j.econlet.2021.110010

Löchl, M., & Axhausen, K. W. (2010). Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use, 3*(2), 39–63. https://doi.org/10.5198/jtlu.v3i2.117

Marcinkowska, E., Gawron, K., & Rejment, M. (2015). Technical and economic aspects of revitalization of down-town tenement-houses in Wrocław. *Journal of Civil Engineering and Management, 21*(8), 1036–1045. https://doi.org/10.3846/13923730.2015.1027259

Marona, B., & Tomal, M. (2020). The COVID-19 pandemic impact upon housing brokers' workflow and their clients' attitude: Real estate market in Krakow. *Entrepreneurial Business and Economics Review, 8*(4), 221–232. https://doi.org/10.15678/EBER.2020.080412

Marona, B., & Tomal, M. (2023). The impact of the Covid-19 pandemic on the operations of real estate agencies and the attitudes of their clients in Poland. *International Entrepreneurship Review, 9*(1), 7–16. https://doi.org/10.15678/IER.2023.0901.01

März, S., Stelk, I., & Stelzer, F. (2022). Are tenants willing to pay for energy efficiency? Evidence from a small-scale spatial analysis in Germany. *Energy Policy, 161*, 112753. https://doi.org/10.1016/j.enpol.2021.112753

Mazhar, A. R., Liu, S., & Shukla, A. (2018). A state of art review on the district heating systems. *Renewable and Sustainable Energy Reviews, 96*, 420–439. https://doi.org/10.1016/j.rser.2018.08.005

McCord, M., Davis, P. T., Haran, M., McIlhatton, D., & McCord, J. (2014). Understanding rental prices in the UK: a comparative application of spatial modelling approaches. *International Journal of Housing Markets and Analysis, 7*(1), 98–128. https://doi.org/10.1108/IJHMA-09-2012-0043

Micallef, B. (2022). Hedonic indices for registered rents in Malta. *The Annual Malta Residential Rental Study Second Edition.* Ministry for Social and Affordable Accomodation, Housing Authority, https://rentregistration.mt/wp-content/uploads/2022/06/Residential-Rental-Study-2nd-edition.pdf

Mouratidis, K. (2021). How COVID-19 reshaped quality of life in cities: A synthesis and implications for urban planning. *Land Use Policy, 111*, 105772. https://doi.org/10.1016/j.landusepol.2021.105772

Nanda, A., Thanos, S., Valtonen, E., Xu, Y., & Zandieh, R. (2021). Forced homeward: The COVID-19 implications for housing. *Town Planning Review, 92*(1), 25–31. https://doi.org/10.3828/tpr.2020.79

Narodowy Bank Polski. (2023). *Sytuacja życiowa i ekonomiczna migrantów z Ukrainy w Polsce w 2023 roku.* https://nbp.pl/wp-content/uploads/2024/01/raport_migranci_z-Ukrainy_2023.pdf

Nasreen, Z., & Ruming, K. (2022). Struggles and opportunities at the platform interface: tenants' experiences of navigating shared room housing using digital platforms in Sydney. *Journal of Housing and the Built Environment, 37*(3), 1537–1554. https://doi.org/10.1007/s10901-021-09909-x

Noszczyk, T., Gorzelany, J., & Kukulska-Kozieł, A., Hernik, J. (2022). The impact of the COVID-19 pandemic on the importance of urban green spaces to the public. *Land Use Policy, 113*, 105925. https://doi.org/10.1016/j.landusepol.2021.105925

Office for Foreigners. (2023). *Obywatele Ukrainy w Polsce – aktualne dane migracyjne.* https://www.gov.pl/web/udsc/obywatele-ukrainy-w-polsce--aktualne-dane-migracyjne [access: 04-06-2024]

Polityka Insight. (2022). *Kwartalnik mieszkaniowy. Raport o sytuacji na rynku mieszkań w I kwartale 2022 roku.* https://www.politykainsight.pl/_resource/multimedium/20305820

Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

Ronald, R. (2008). *The ideology of home ownership: Homeowner societies and the role of housing.* https://doi.org/10.1057/9780230582286

Ronald, R., & Kadi, J. (2018). The Revival of Private Landlords in Britain's Post-Homeownership Society. *New Political Economy, 23*(6), 786–803. https://doi.org/10.1080/13563467.2017.1401055

Rosak-Szyrocka, J., & Żywiolek, J. (2022). Qualitative Analysis of Household Energy Awareness in Poland. *Energies, 15* (6). https://doi.org/10.3390/en15062279

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55. https://doi.org/10.1086/260169

Rubaszek, M., & Czerniak, A. (2017). Preferencje Polaków dotyczace struktury własnosciowej mieszkan: Opis wyników ankiety. *Bank i Kredyt, 48*(2), 197–233.

Rubaszek, M., & Rubio, M. (2020). Does the rental housing market stabilize the economy? A micro and macro perspective. *Empirical Economics, 59*(1), 233–257. https://doi.org/10.1007/s00181-019-01638-z

Samuelson, P. A. (1948). Consumption Theory in Terms of Revealed Preference. *Economica, 15*(60). https://doi.org/10.2307/2549561

Sieger, L., & Weber, C. (2023). Inefficient markets for energy efficiency ? – The efficiency premium puzzle in the German rental housing market. *Energy Policy, 183*, 113819. https://doi.org/10.1016/j.enpol.2023.113819

Sirmans, S., Sirmans, C., & Benjamin, J. (1989). Determining Apartment Rent: The Value of Amenities, Services and External Factors. *Journal of Real Estate Research, 4*(2), 33–43. https://doi.org/10.1080/10835547.1989.12090581

Terlikowski, W. (2013). Revitalization capacity of public utility buildings. *Materiały Budowlane, 6*(490), 48–50.

Tomal, M. (2020). Modelling housing rents using spatial autoregressive geographically weighted regression: A case study in Cracow, Poland. *ISPRS International Journal of Geo-Information, 9*(6). https://doi.org/10.3390/ijgi9060346

Tomal, M., & Helbich, M. (2022). The private rental housing market before and during the COVID-19 pandemic: A submarket analysis in Cracow, Poland. *Environment and Planning B: Urban Analytics and City Science, 49*(6), 1646–1662. https://doi.org/10.1177/23998083211062907

Tomal, M., & Helbich, M. (2023). A spatial autoregressive geographically weighted quantile regression to explore housing rent determinants in Amsterdam and Warsaw. *Environment and Planning B: Urban Analytics and City Science, 50*(3), 579–599. https://doi.org/10.1177/23998083221122790

Tomal, M., & Marona, B. (2021). The impact of the covid-19 pandemic on the private rental housing market in poland: What do experts say and what do actual data show? *Critical Housing Analysis, 8*(1), 24–35. https://doi.org/10.13060/23362839.2021.8.1.520

Trojanek, R., & Gluszak, M. (2022). Short-run impact of the Ukrainian refugee crisis on the housing market in Poland. *Finance Research Letters, 50*(103236). https://doi.org/10.1016/j.frl.2022.103236

Trojanek, R., Gluszak, M., Tanas, J., & Hebdzynski, M. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

UNHCR. (2023). *Ukraine Refugee Situation.* https://data2.unhcr.org/en/situations/ukraine/location?secret=unhcrrestricted [access: 04-06-2024]

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817–838. https://doi.org/10.2307/1912934

Widłak, M., & Tomczyk, E. (2010). Measuring price dynamics: evidence from the Warsaw housing market. *Journal of European Real Estate Research, 3*(3), 203–227. https://doi.org/10.1108/17539261011094722

Wojdat, M. & Cywiński, P. (2022). *Urban hospitality: unprecedented growth, challenges and opportunities.* https://metropolie.pl/fileadmin/user_upload/UMP_raport_Ukraina_ANG_20220429_final.pdf

World Economic Forum. (2024). *The Global Risks Report 2024.* https://www.weforum.org/publications/global-risks-report-2024 [access: 04-06-2024]

Zhan, D., Xie, C., Zhang, J., & Meng, B. (2023). Investigating the Determinants of Housing Rents in Hangzhou, China: A Spatial Multilevel Model Approach. *Applied Spatial Analysis and Policy, 16*(4), 1707–1727. https://doi.org/10.1007/s12061-023-09530-1

Żuk, P., & Żuk, P. (2022). National energy security or acceleration of transition? Energy policy after the war in Ukraine. *Joule, 6*(4), 709–712. https://doi.org/10.1016/j.joule.2022.03.009

APPENDIX A

Table A1

Basic statistics of the variables in the database (N=7,768)

Variable	MIN	AVG	MAX	Count of 1's	VIF
RENT (dependent variable)	700	1,927	6,000		
AREA	15	46.8	150		2.95
ROOMS	1	1.99	4		2.93
BLOCK				5,186	
APART				2,027	1.44
TENE_RENO				196	1.10
TENE_REVI				377	1.22
TENE_N_RENO				686	1.36
QUALITY	-0.014	0.165	0.335		1.52
FURN				6,709	1.04
BALCONY				5,712	1.49
PARKING				1,526	1.07
GARDEN				234	1.03
DIST_CC	0.021	2.969	10.916		2.49
DIST_GREEN	0.001	0.269	0.996		1.13
DIST_TRAM	0.005	0.462	6.478		1.58
DIST_UNI	0.017	1.005	9.141		1.95
HEAT_DISTR				5,238 (out of 5,672)	
HEAT_GAS				434 (out of 5,672)	1.69
PRE-SHOCK				1,666	
PAND				4,601	
WAR				1,501	

VIF values were calculated based on the results of the MOD_0 model. VIF for HEAT_GAS was calculated based on a model analogous to the MOD_0 model but restricted to observations with a known heating type (N=5,672). *Source*: own elaboration.

Table A2

Test statistics for the MOD_0 model

Type of test	Test statistics	p-value of test statistics
Moran	0.0468	< 0.001
RESET	2.7478 (df1=30, df2=7331)	< 0.001
Breusch-Pagan	202.5 (df=29)	< 0.001
Lagrange multiplier (LMERR)	149.59 (df=1)	< 0.001
Lagrange multiplier (LMLAG)	125.89 (df=1)	< 0.001
Lagrange multiplier (RLMERR)	77.53 (df=1)	< 0.001
Lagrange multiplier (RLMLAG)	53.83 (df=1)	< 0.001

Source: own elaboration.

Table A3

Values of AIC achieved in the process of selection of the best spatial variant of the MOD_1 model

Spatial modelling approach	Weighting matrix distance, m	AIC
SEM	250	-10,291.69
SLM	250	-10,282.61
SEM	375	-10,280.86
SLM	375	-10,278.77
SEM	500	-10,276.52
SLM	500	-10,285.16
SEM	600	-10,264.44
SLM	600	-10,271.60

Source: own elaboration.

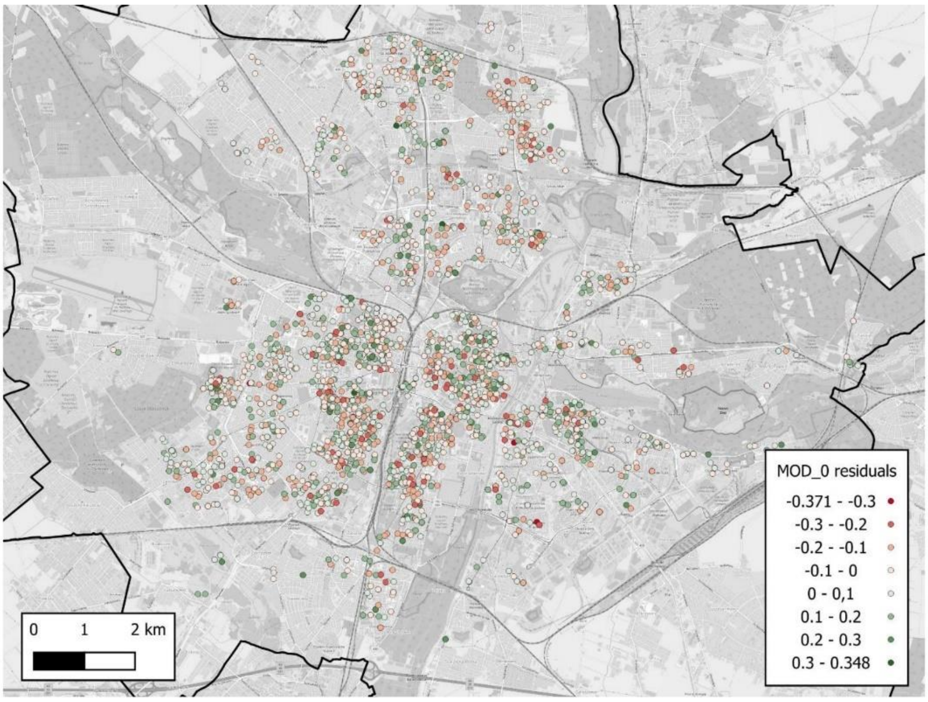


Figure A3. Spatial distribution of the MOD_0 model residuals

Source: own elaboration based on OpenStreetMap.

320

APPENDIX B

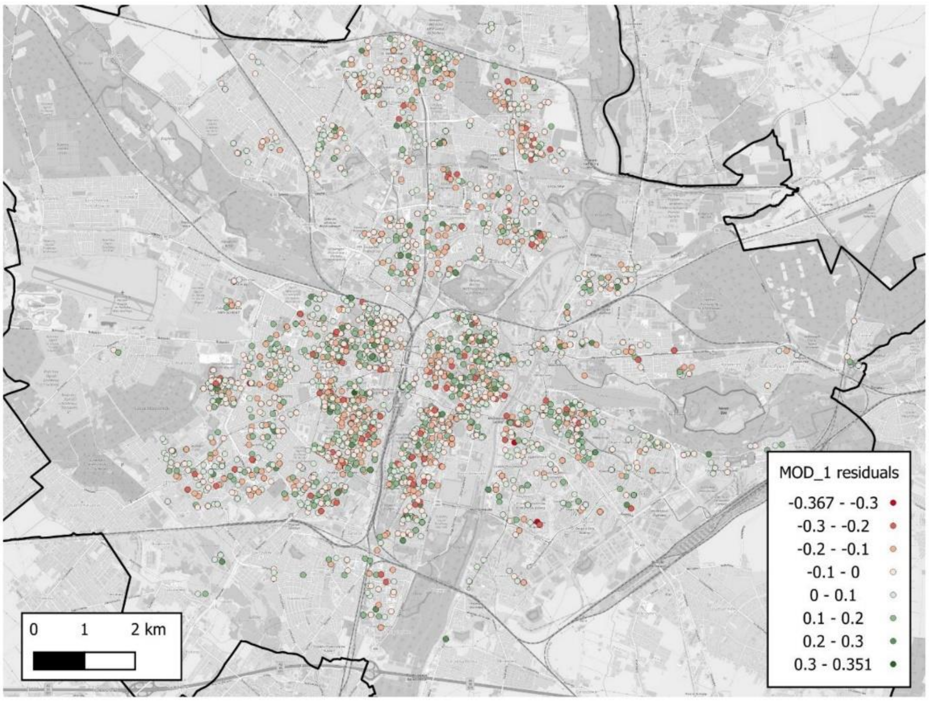


Figure B1. Spatial distribution of the MOD_1 model residuals
Source: own elaboration based on OpenStreetMap.

Table B1

Test statistics for spatial models

Type of test	Tested model	Test statistics	p-value of test statistics
Breusch-Pagan	MOD_1	221.4 (df=29)	< 0.001
Breusch-Pagan	MOD_2	147.1 (df=30)	< 0.001
Breusch-Pagan	MOD_3	164.2 (df=30)	< 0.001
Breusch-Pagan	MOD_4	59.6 (df=24)	< 0.001
Jarque-Bera	MOD_1	38.4 (df=2)	< 0.001
Jarque-Bera	MOD_2	21.7 (df=2)	< 0.001
Jarque-Bera	MOD_3	16.6 (df=2)	< 0.001
Jarque-Bera	MOD_4	2.1 (df=2)	0.349

Source: own elaboration.

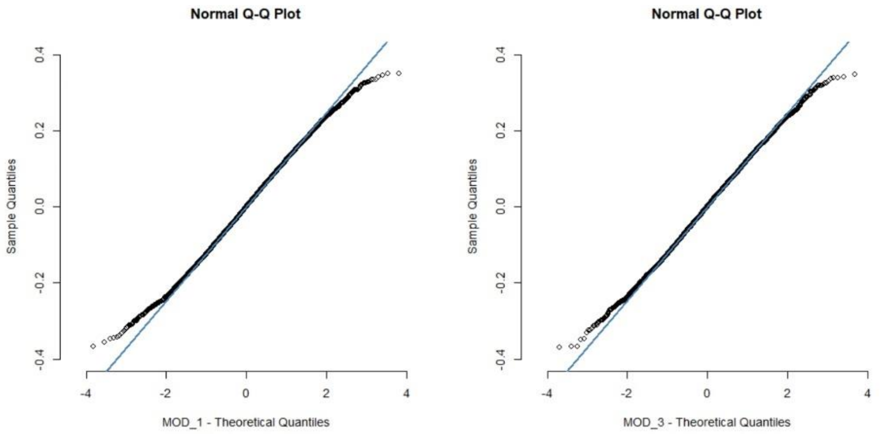


Figure B2. Distribution of selected models' residual versus quantiles of normal distribution

Source: own elaboration.

APPENDIX C

Table C1

Full results of models

	MOD_0	MOD_1	MOD_2	MOD_3	MOD_3_BASE	MOD_4	MOD_4_BASE
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
AREA	0.0079 (0.000)	0.0080 (0.000)	0.0080 (0.000)	0.0086 (0.000)	0.0086 (0.000)	0.0083 (0.000)	0.0083 (0.000)
ROOMS	0.088 (0.000)	0.086 (0.000)	0.092 (0.000)	0.080 (0.000)	0.076 (0.000)	0.096 (0.000)	0.096 (0.000)
PAND x ROOMS				-0.006 (0.359)			
APART	0.037 (0.000)	0.032 (0.000)	0.023 (0.000)	0.037 (0.000)	0.037 (0.000)	0.022 (0.000)	0.022 (0.000)
TENE_RENO	0.019 (0.021)	0.018 (0.067)	0.020 (0.162)	0.037 (0.004)	0.038 (0.004)	-0.023 (0.424)	-0.325 (0.246)
TENE_REVI	0.040 (0.000)	0.039 (0.000)	0.027 (0.011)	0.041 (0.000)	0.042 (0.000)	0.011 (0.489)	0.012 (0.446)
TENE_N_RENO	-0.001 (0.823)	-0.001 (0.787)	-0.001 (0.890)	0.014 (0.077)	0.014 (0.080)	-0.021 (0.184)	-0.021 (0.184)
QUALITY	3.009 (0.000)	2.901 (0.000)	2.953 (0.000)	2.858 (0.000)	2.858 (0.000)	3.064 (0.000)	3.049 (0.000)
FURN	0.045 (0.000)	0.045 (0.000)	0.049 (0.000)	0.042 (0.000)	0.041 (0.000)	0.053 (0.000)	0.052 (0.000)
BALCONY	0.034 (0.000)	0.033 (0.000)	0.032 (0.000)	0.027 (0.003)	0.041 (0.000)	0.012 (0.126)	0.012 (0.000)
PAND x BALCONY				0.018 (0.082)			
PARKING	0.036 (0.000)	0.034 (0.000)	0.031 (0.000)	0.035 (0.000)	0.035 (0.000)	0.029 (0.000)	0.029 (0.000)
GARDEN	0.016 (0.020)	0.011 (0.202)	0.012 (0.189)	-0.039 (0.168)	0.005 (0.647)	0.005 (0.715)	0.005 (0.721)
PAND x GARDEN				0.053 (0.089)			
DIST_CC	-0.0094 (0.000)	-0.0103 (0.000)	-0.0104 (0.000)	-0.0101 (0.000)	-0.0101 (0.000)	-0.0091 (0.002)	-0.0092 (0.000)
DIST_GREEN	-0.055 (0.000)	-0.058 (0.000)	-0.063 (0.000)	-0.026 (0.227)	-0.064 (0.000)	-0.068 (0.000)	-0.068 (0.000)
PAND x DIST_GREEN				-0.049 (0.028)			
DIST_TRAM	-0.003 (0.440)	-0.001 (0.863)	-0.007 (0.226)	-0.006 (0.413)	-0.006 (0.424)	-0.003 (0.666)	-0.003 (0.695)
DIST_UNI	-0.009 (0.000)	-0.009 (0.009)	-0.006 (0.121)	-0.003 (0.633)	-0.006 (0.156)	-0.009 (0.061)	-0.009 (0.058)
PAND x DIST_UNI				-0.004 (0.409)			
HEAT_GAS			-0.011 (0.187)			0.013 (0.487)	-0.010 (0.461)
WAR x HEAT_GAS						-0.399 (0.058)	
TD_201909	0.035 (0.000)	0.037 (0.002)	0.041 (0.004)	0.050 (0.000)	0.053 (0.000)	0.041 (0.004)	0.041 (0.004)
TD_201912	0.043 (0.000)	0.044 (0.000)	0.041 (0.000)	0.055 (0.000)	0.056 (0.000)	0.041 (0.000)	0.041 (0.000)
TD_202003	0.025 (0.001)	0.025 (0.001)	0.022 (0.018)	0.034 (0.000)	0.035 (0.000)	0.016 (0.102)	0.015 (0.112)
TD_202006	0.007 (0.318)	0.006 (0.439)	0.008 (0.349)	0.022 (0.177)	0.009 (0.288)		
TD_202009	0.015 (0.045)	0.015 (0.047)	0.018 (0.051)	0.035 (0.030)	0.023 (0.011)		
TD_202012	-0.042 (0.000)	-0.042 (0.000)	-0.033 (0.000)	-0.023 (0.154)	-0.036 (0.000)		
TD_202103	-0.052 (0.000)	-0.053 (0.000)	-0.052 (0.000)	-0.033 (0.044)	-0.046 (0.000)		
TD_202106	-0.039 (0.000)	-0.038 (0.000)	-0.030 (0.001)	-0.023 (0.158)	-0.036 (0.000)		
TD_202109	0.052 (0.000)	0.051 (0.000)	0.054 (0.000)	0.075 (0.000)	0.062 (0.000)		
TD_202112	0.069 (0.000)	0.069 (0.000)	0.064 (0.000)	0.081 (0.000)	0.068 (0.000)		
TD_202203	0.126 (0.000)	0.126 (0.000)	0.130 (0.000)			0.133 (0.000)	0.131 (0.000)
TD_202206	0.228 (0.000)	0.225 (0.000)	0.222 (0.000)			0.223 (0.000)	0.221 (0.000)
TD_202209	0.342 (0.000)	0.342 (0.000)	0.338 (0.000)			0.335 (0.000)	0.332 (0.000)
TD_202212	0.314 (0.000)	0.314 (0.000)	0.315 (0.000)			0.321 (0.000)	0.318 (0.000)
CONSTANT	6.390 (0.000)	6.409 (0.000)	6.392 (0.000)	6.394 (0.000)	6.405 (0.000)	6.373 (0.000)	6.378 (0.000)
Method	**OLS**	**SEM**	**SEM**	**SEM**	**SEM**	**SEM**	**SEM**
R^2	**0.828**						
AIC	**-10,172.8**	**-10,292.0**	**-7,595.1**	**-6,012.4**	**-6,010.5**	**-3,211.8**	**-3,210.2**
AIC of an OLS equivalent		**-10,172.8**	**-7,488.2**	**-5,942.0**	**-5,940.7**	**-3,194.1**	**-3,192.9**
N	**7,391**	**7,391**	**5,381**	**4,342**	**4,342**	**2,344**	**2,344**

P-values have been included in parentheses ("0.000" indicates p-values lower than 0.0005).
Source: own elaboration.

Article 5

Trojanek, R., Gluszak, M., Hebdzyński, M. & Tanas, J. (2021). The COVID-19 pandemic, Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis, 8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

Critical Housing Analysis
Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

The COVID-19 Pandemic, Airbnb and Housing Market Dynamics in Warsaw

Radoslaw Trojanek (corresponding author)

Poznań Univeristy of Economics and Business, Poland
radoslaw.trojanek@ue.poznan.pl

Michal Gluszak

Cracow University of Economics, Poland
gluszakm@uek.krakow.pl

Michal Hebdzynski

Poznań Univeristy of Economics and Business, Poland
michal.hebdzynski@phd.ue.poznan.pl

Justyna Tanas

WSB University in Poznan, Poland
justyna.tanas@wsb.poznan.pl

Abstract: *In this study, we analyse the impact of COVID-19 on house rents and prices in Warsaw, the capital of Poland. Hedonic indexes indicate a slight increase in prices (ca. 1.2%) and a substantial drop in long-term rents (ca. -7.7%) between March 2020 and December 2020. The largest decline in rents occurred in centrally located neighbourhoods, which was largely due to the inflow of new housing supply from the short-term rental market (the Airbnb Warsaw market shrank by almost 30% in December 2020 y/y). Using hedonic methods, we show the effect of the shrinking Airbnb market on the drop in long-term rents. The study indicates the elasticity of rents with respect to Airbnb supply, with a 1% change in Airbnb listings leading to a 0.031% change in rents.*

Keywords: housing market; Airbnb; COVID-19.

Introduction

The economic impact of the COVID-19 pandemic has been widely discussed in popular and social media; however, it has not been comprehensively investigated in the economic literature. In general, we hypothesised that the COVID-19 pandemic has a significant and negative impact on house rents due to a decrease in demand (higher unemployment rate, labour market uncertainty, lack of students) and an increase in supply (inflows from the short-term rental sector). The effect of the COVID-19 pandemic on house prices is not clear from the economic perspective for several reasons—both market and policy-related. Some factors may have a negative impact on house prices (decreased rental incomes, economic uncertainty), whereas others suggest the reverse effect (drop in interest rates). House prices tend to be sticky in the down-market as owners are reluctant to sell. Overall, the exact short-term net impact of the COVID-19 pandemic is yet to be thoroughly evaluated. To the best of our knowledge, this study is one of the first attempts to examine the relationship between Airbnb and long-term rent responses to the COVID-19 pandemic.

Literature review

Due to data limitations, few papers evaluate the consequences of pandemics on the housing market. An Italian study suggests that COVID-19 has had a significant impact on housing prices. Authors predict a 4–6% short-run price decrease from 2020 to 2021 (Del Giudice, De Paola and Del Giudice 2020). Early results from Marona and Tomal (2020) show several demand-side adjustments that have occurred due to COVID-19 in Poland. The linkages between pandemics, credit risk, and policies and mortgage lending have been discussed through the example of the Chinese housing market (Su, Cai, Qin, Tao and Umar 2021). Another recent study evaluated the links between monetary policy and house prices in 31 countries in the post-Covid era (Apergis 2021). The impact of COVID-19 on house prices has been reported in two recent studies. Using a quasi-experimental approach Qian, Qiu and Zhang (2021) found that house prices in China decreased by 2.5% in communities with confirmed COVID-19 cases. An Australian study indicates that the number of confirmed Covid-19 cases had a significant impact on housing returns. Housing returns were not affected by policy interventions (lockdown orders) (Hu, Lee and Zou 2021). Concerning the impact of the Covid-19 pandemic on long-term residential rental rates, researches that have been conducted indicate both no significant changes (Kadi, Schneider and Seidl 2020) but also sharp declines (Tomal and Marona 2021).

Many studies suggest that there is a substitution effect between short-term rental (for example, Airbnb) and the housing rental market (Benítez-Aurioles and Tussyadiah 2020). A US-based study revealed that Airbnb listings increase significantly with rents and house prices (Barron, Kung and Proserpio 2018). Other studies, conducted in Boston (Horn and Merante 2017), New York (Wachsmuth and Weisler 2018), San Francisco and Los Angeles (Lee 2016), confirmed also this connection between Airbnb and rents. The impact of short-term rental on housing rents and prices has been demonstrated based on empirical data from popular tourist destinations outside the US—Barcelona (Garcia-López, Jofre-Monseny, Martínez-Mazza and Segú 2020) and Taiwan (Chang 2020). The opposite effect should manifest when a share of short-term rental apartments increases the supply of rental housing. We investigated this effect using hedonic price models.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Data and method

Data

In Poland, there are no official monthly statistics on housing prices and rents. The information on transactions in the Land Registry is delayed by around 6–8 months, rendering it impossible to pinpoint up-to-date changes in the Warsaw housing market. Moreover, rental agreements are confidential, and there is no available official information on actual rents that can be used in empirical research. Considering these limitations, we decided to use offers, which may be an adequate substitute when transaction data are not available (Anenberg and Laufer 2017; Lyons 2019). Despite potential caveats, using offer information allows us to obtain up-to-date results.

The data on asking prices (167,875 observations) and asking rents (112,963 observations) were collected for the purpose of our research from advertising portals every month from January 2017 to December 2020. A more detailed description of the dataset formation can be found in Trojanek (2021). The information on Airbnb was purchased from the AirDNA company (information on the monthly performance of 56,283 active apartments).

Method

The research strategy consists of the following steps. To estimate the indexes and elasticity of rents to Airbnb supply, we used the quantile regression proposed by Koenker and Bassett (Koenker and Bassett 1978), whereby it is possible to assess the various quantile functions of the conditional variable distribution. We decided to use symmetric (quantile 0.5) and asymmetric (quantiles 0.25 and 0.75) weighting. Monthly housing price and rent indexes in 2017–2020 were estimated using the time-dummy hedonic method with following the equation:

$$lnP = \beta_0 + \sum_{j=1}^{K} \beta_j C_j + \sum_{i=2}^{t} \gamma_i D_i + \varepsilon$$

where D_i is a zero-one time variable and C_j denotes the apartment's characteristics. To control for changes in the structure and quality of apartments for sale/rent in monthly periods and thereby determine the dynamics, we used information on the location (district), size, and age of apartments (set of variables $\boldsymbol{C}_{\mathrm{j}}$).

The hedonic model (Lancaster 1966; Rosen 1974) was used to investigate the influence of Airbnb supply on long-term rents. The research was conducted for three periods: 2017–2020, before the lockdown (2017 to March 2020) and after (April 2020–December 2020). A regression analysis of the apartment's rent was conducted on a set of independent variables. We controlled for size, age, quality of the apartment, distance variables, time dummies, and Airbnb supply (monthly number of active apartments in the district). The Airbnb market is centralised in Warsaw; 90% of apartments are located in 7 districts, with 70% in Srodmiescie and Wola. These neighbourhoods were the basis for further investigation.

As a robustness check, to explore the phenomenon even further, we applied spatial econometrics. The plausible hypothesis was that rent dynamics differed significantly over space; thus, we investigated rent changes with a multiscale geographically weighted regression

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

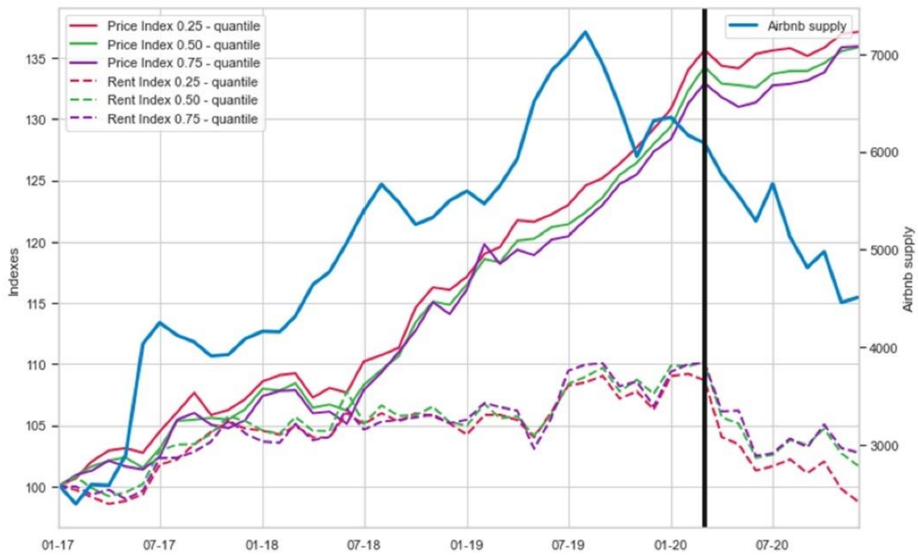
(MGWR) model (Fotheringham, Yang and Kang 2017). Based on the data from June 2019 to December 2020, we investigated the spatial distribution of rent changes, dividing the observation into two groups - before and after the lockdown. Using the estimated coefficients for each observation, we produced a map showing the spatial differentiation of rent changes in Warsaw. Additionally, a map of Airbnb market activity from 2019 to 2020 (the difference in the number of active apartments) was prepared using kernel density estimation.

Results and discussion

The first confirmed case of the SARS-CoV-2 infection in Poland was officially reported on 4 March 2020. By 31 December 2020, a total of 1,294,878 confirmed cases and 28,554 deaths had been reported (Hale, Webster, Petherick, Phillips and Kira 2020). COVID-19 related restrictions were introduced between 10-15 March 2020, which included the closing of schools and universities and a ban on international travel. Lockdown restrictions were tightened on 25 March (See Appendix A for details and a chronology).

We estimated the quantile hedonic regression models to construct quality-controlled housing rent and price indexes in Warsaw. The hedonic indexes and Airbnb supply are presented in Figure 1.

Figure 1: Quantile hedonic house price, rent indexes and Airbnb supply for Warsaw (Jan 2017–Dec 2020)



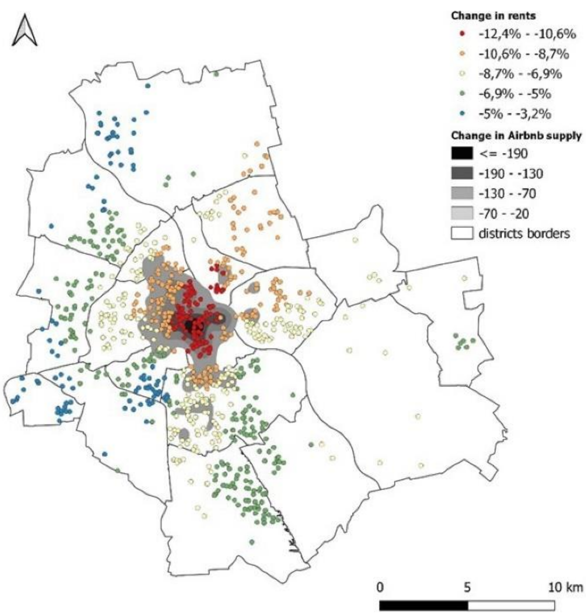
Note: The vertical line indicates the introduction of lockdown in Poland.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

The empirical estimates indicate that house prices and rents increased at a different pace before lockdown. We observed a 10% rent increase from January 2017 to March 2020. The COVID-19 pandemic significantly affected the rental market. Throughout the nine months following the beginning of the lockdown in March 2020, rents in Warsaw decreased by 7–9%, depending on the quantile. Decreases were particularly strong in the 0.25 quantile. Compared to rents, house prices in Warsaw rose dynamically before the pandemic (by 34–35% from January 2017 to March 2020) and were significantly less affected by the COVID-19 lockdown. Initially, we observed a small adjustment at the beginning of the pandemic. In April and May 2020, house prices decreased by 1–2% but later continued to rise steadily for the remainder of 2020.

Since April 2020 the Polish rental market has been severely hit by the COVID-19 pandemic. The short-term rent decrease since the pandemic started can be partly attributed to supply adjustments. The inflow of apartment stock from the short-term rental sector has been heavily affected by various COVID-19-related restrictions imposed on domestic and international travel. The restrictions imposed on domestic and international tourism induced a massive withdrawal of apartments from the short-term market. Some of them have been adapted for long-term market purposes; others have been offered on a mid-term basis (agreements for around six months), suggesting landlords' intention to reintroduce the apartments into the short-term market as soon as the pandemic subsides. The Airbnb market shrank by almost 30% in December 2020 (y/y) to 4510 apartments in Warsaw. We produced a map of rent price changes for long-term rentals and a map of the spatial outflow of apartments from the Airbnb market during the COVID-19 restrictions (Fig 2).

Figure 2: Long-term rents and Airbnb supply changes from 2019 to 2020



Source: Authors.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

The drop in rents varied spatially. The largest declines occurred in areas with the greatest outflow of apartments from the Airbnb market. At the same time, the transition to remote learning encouraged students to stay in the family home and terminate any long-term rental agreement they had in the cities in which they had been studying before the pandemic. Those factors contributed to both an additional increase in supply on the long-term market, which may have contributed to the bigger decrease in apartments in the 0.25 quantile.
We studied the elasticity of rents with respect to Airbnb supply using a quantile hedonic regression model. The partial results (full results and description of variables in Table B1 and B2 of Appendix B) of the estimation are presented in Table 1.

The results of the estimates for 2017–2020 confirm the positive relationship between the rents and Airbnb supply. The study indicates that a 1% change in Airbnb listings leads to a 0.031% change in rents. It is worth noting that the elasticity (median quantile) was higher in the period of rent growth (before the pandemic)—about 0.0322 compared to the decreased figure of 0.0219.

Our research results provide preliminary conclusions on the impact of COVID-19 on the residential real estate market during the first months of the pandemic in Poland on the example of Warsaw. We found a significant decline in long-term rents in the rental market, which was undoubtedly influenced by many economic and social factors. However, this decline was exacerbated by the Airbnb market's collapse and an increase in the supply of apartments for rent. The result provides further support for the argument found in the literature that there is a relationship between short-term and long-term rental markets (Barron, Kung and Proserpio 2018; Benítez-Aurioles and Tussyadiah 2020; Horn and Merante 2017; Lee 2016). The study suggests that the impact is persistent not only in the growing economy but also during the recession. This is indicated by the spatial variation in rent declines (the most significant drop in central districts) and by the elasticity that was determined, which indicates a positive relationship between rents and Airbnb supply. Concerning housing prices, the results are not in line with previous studies (Qian, Qiu and Zhang 2021), as we did not see a decline. However, it should be noted that, compared to recent years, the increase slowed significantly.

Compared to other European countries, in Poland only limited housing policy measures were introduced to support tenants and owners and they focused mostly on mortgage forbearance (OECD 2020). On the other hand, the Monetary Policy Council decreased interest rates to a historical minimum, which provided a stimulus for housing investments by limiting mortgage costs. Moreover, the decrease in interest rates made housing investment even more attractive because it has been historically proven to be a good hedge against inflation. Nonetheless, the economic literature suggests that the impact of monetary policy depends on the housing cycle and is weaker when house prices are high (Bluwstein, Brzoza-Brzezina, Gelain and Kolasa 2020). Recent empirical evidence shows that the effects of monetary expansion during the Covid-19 pandemic were relatively limited (Apergis 2021). None of those measures is directly controlled within this particular study; however, investigating the effect of different housing policies on housing prices or rents and comparing European countries' outcomes is interesting for future research.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Table 1: Partial estimation results

	January 2017—December 2020				January 2017—March 2020				April 2020—December 2020			
	0.25Q	**0.50Q**	**0.75Q**	**OLS**	**0.25Q**	**0.50Q**	**0.75Q**	**OLS**	**0.25Q**	**0.50Q**	**0.75Q**	**OLS**
lnairbnb	0.0303 (0.0015)	0.0310 (0.0014)	0.0244 (0.0016)	0.0297 (0.0014)	0.0308 (0.0016)	0.0322 (0.0015)	0.0247 (0.0018)	0.0306 (0.0015)	0.0183 (0.0046)	0.0219 (0.0044)	0.0233 (0.0046)	0.0228 (0.0040)
characteristics controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
area effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
time effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	41246	41246	41246	41246	35824	35824	35824	35824	5422	5422	5422	5422
R2				0.8184				0.8170				0.8156

Source: Authors.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Conclusion

There are few economic studies on the impact of pandemics on the housing market, and none has addressed the role of the short-term rental market. This paper helps fill this gap by investigating the dynamics of housing prices and rents in Warsaw before and during the Covid-19 pandemic. Using hedonic indexes, we found a slight increase in prices and a substantial drop in long-term rents between March 2020 and December 2020. The shrinking Airbnb market increased the supply in the long-term rental market, which at least partially explains the rent decrease. This is indicated by the spatial variation in rent declines and by the elasticity identified, which indicates a positive relationship between rents and Airbnb supply.

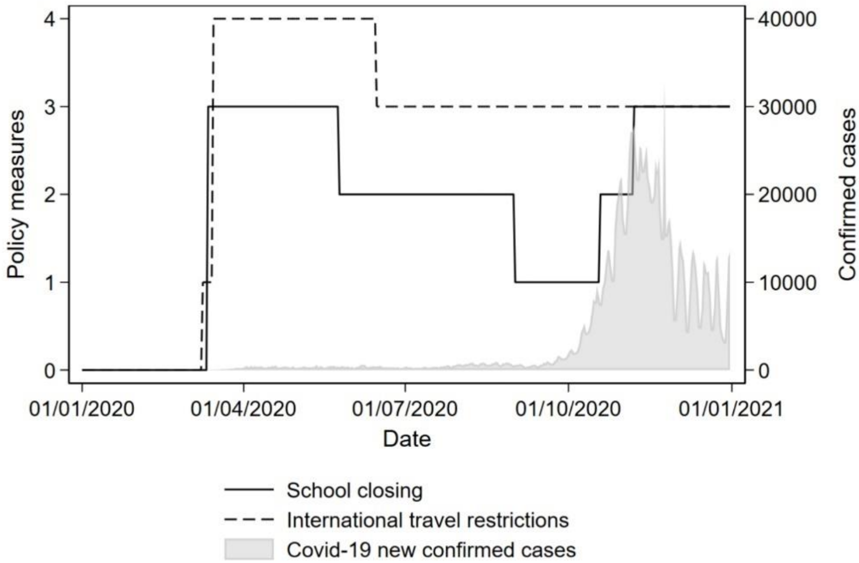
Funding

This study is based on research project financed by the National Science Centre of Poland (grant no. 2017/27/B/HS4/01848).

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Appendix A

Figure A1: Covid-19 epidemics and lockdown restrictions in Poland



Note: School closing (Ordinal scale): 0 - no measures; 1 - recommend closing or all schools open with alterations; 2 - require closing (only some levels or categories); 3 - require closing all levels; International travel restrictions (for foreign travellers) (Ordinal scale): 0 - no restrictions; 1 - screening arrivals; 2 - quarantine arrivals from some or all regions; 3 - ban arrivals from some regions; 4 - ban on all regions or total border closure.

Source: Authors based on Hale et al. (2020).

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Appendix B

Table B1: Regression estimation results

	January 2017 - December 2020				January 2017- March 2020				April 2020 – December 2020			
	0.25Q	**0.50Q**	**0.75Q**	**OLS**	**0.25Q**	**0.50Q**	**0.75Q**	**OLS**	**0.25Q**	**0.50Q**	**0.75Q**	**OLS**
Const	7.0818	7.1488	7.3370	7.1012	7.0969	7.1538	7.3570	7.1158	7.0173	7.0874	7.1793	7.0122
	(0.0183)	(0.0163)	(0.0191)	(0.0161)	(0.0194)	(0.0172)	(0.0206)	(0.0172)	(0.0555)	(0.0510)	(0.0530)	(0.0468)
age	-0.0068	-0.0074	-0.0076	-0.0049	-0.0070	-0.0073	-0.0076	-0.0050	-0.0056	-0.0074	-0.0074	-0.0046
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
age2	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
area	0.0153	0.0167	0.0181	0.0172	0.0150	0.0165	0.0179	0.0169	0.0167	0.0180	0.0200	0.0188
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
lnairbnb	0.0303	0.0310	0.0244	0.0297	0.0308	0.0322	0.0247	0.0306	0.0183	0.0219	0.0233	0.0228
	(0.0015)	(0.0014)	(0.0016)	(0.0014)	(0.0016)	(0.0015)	(0.0018)	(0.0015)	(0.0046)	(0.0044)	(0.0046)	(0.0040)
lncc	-0.0800	-0.0960	-0.1305	-0.1125	-0.0799	-0.0944	-0.1302	-0.1111	-0.0894	-0.1088	-0.1359	-0.1232
	(0.0024)	(0.0021)	(0.0025)	(0.0021)	(0.0026)	(0.0023)	(0.0027)	(0.0023)	(0.0065)	(0.0060)	(0.0063)	(0.0055)
lnsubway	-0.0179	-0.0168	-0.0179	-0.0123	-0.0191	-0.0181	-0.0191	-0.0137	-0.0066	-0.0060	-0.0082	-0.0024
	(0.0014)	(0.0012)	(0.0012)	(0.0012)	(0.0015)	(0.0012)	(0.0013)	(0.0012)	(0.0043)	(0.0038)	(0.0036)	(0.0035)
lnuga	-0.0043	-0.0069	-0.0130	-0.0093	-0.0037	-0.0070	-0.0139	-0.0091	-0.0100	-0.0066	-0.0065	-0.0115
	(0.0012)	(0.0011)	(0.0012)	(0.0011)	(0.0013)	(0.0011)	(0.0013)	(0.0011)	(0.0035)	(0.0033)	(0.0033)	(0.0030)
quality	0.0456	0.0462	0.0417	0.0520	0.0458	0.0473	0.0427	0.0521	0.0469	0.0418	0.0340	0.0504
	(0.0015)	(0.0013)	(0.0016)	(0.0013)	(0.0016)	(0.0014)	(0.0018)	(0.0014)	(0.0041)	(0.0038)	(0.0038)	(0.0035)
q2	-0.0173	-0.0131	-0.0079	-0.0068	-0.0162	-0.0121	-0.0070	-0.0071				
	(0.0045)	(0.0041)	(0.0049)	(0.0040)	(0.0044)	(0.0040)	(0.0049)	(0.0040)				
q3	0.0417	0.0393	0.0488	0.0389	0.0412	0.0398	0.0505	0.0390				
	(0.0054)	(0.0049)	(0.0059)	(0.0049)	(0.0053)	(0.0048)	(0.0059)	(0.0048)				
q4	0.0408	0.0347	0.0344	0.0394	0.0402	0.0343	0.0356	0.0394				
	(0.0040)	(0.0037)	(0.0044)	(0.0036)	(0.0040)	(0.0036)	(0.0044)	(0.0036)				
q5	0.0275	0.0216	0.0269	0.0273	0.0264	0.0217	0.0264	0.0270				
	(0.0043)	(0.0040)	(0.0047)	(0.0039)	(0.0043)	(0.0039)	(0.0048)	(0.0039)				
q6	0.0102	-0.0021	0.0058	0.0079	0.0105	-0.0014	0.0051	0.0071				
	(0.0056)	(0.0051)	(0.0061)	(0.0051)	(0.0056)	(0.0051)	(0.0062)	(0.0050)				
q7	0.0346	0.0206	0.0212	0.0309	0.0345	0.0201	0.0210	0.0298				
	(0.0058)	(0.0053)	(0.0063)	(0.0052)	(0.0057)	(0.0052)	(0.0064)	(0.0052)				
q8	0.0390	0.0145	0.0062	0.0225	0.0389	0.0136	0.0056	0.0217				
	(0.0073)	(0.0066)	(0.0079)	(0.0066)	(0.0072)	(0.0066)	(0.0080)	(0.0065)				
q9	0.0324	0.0204	0.0163	0.0210	0.0309	0.0197	0.0151	0.0197				
	(0.0087)	(0.0080)	(0.0095)	(0.0079)	(0.0087)	(0.0079)	(0.0096)	(0.0078)				
q10	0.0253	0.0148	0.0021	0.0230	0.0245	0.0131	0.0042	0.0212				
	(0.0079)	(0.0073)	(0.0087)	(0.0072)	(0.0079)	(0.0072)	(0.0088)	(0.0072)				
q11	0.0418	0.0385	0.0394	0.0337	0.0400	0.0368	0.0380	0.0320				
	(0.0079)	(0.0072)	(0.0086)	(0.0071)	(0.0078)	(0.0071)	(0.0087)	(0.0071)				
q12	0.0283	0.0224	0.0206	0.0207	0.0276	0.0216	0.0194	0.0190				
	(0.0093)	(0.0085)	(0.0101)	(0.0084)	(0.0092)	(0.0083)	(0.0102)	(0.0083)				
q13	0.0437	0.0276	0.0218	0.0320	0.0419	0.0257	0.0227	0.0307				
	(0.0086)	(0.0078)	(0.0093)	(0.0077)	(0.0085)	(0.0077)	(0.0094)	(0.0077)				
q14	-0.0214	-0.0200	-0.0286	-0.0272								
	(0.0075)	(0.0068)	(0.0081)	(0.0068)								
q15	-0.0249	-0.0173	-0.0078	-0.0218					-0.0012	0.0072	0.0124	0.0055
	(0.0078)	(0.0071)	(0.0085)	(0.0070)					(0.0106)	(0.0100)	(0.0103)	(0.0091)
q16	-0.0381	-0.0208	-0.0042	-0.0185					-0.0207	0.0029	0.0199	0.0034
	(0.0050)	(0.0046)	(0.0055)	(0.0046)					(0.0094)	(0.0087)	(0.0088)	(0.0080)
N	41246	41246	41246	41246	35824	35824	35824	35824	5422	5422	5422	5422
R2				0.8184				0.8170				0.8156

Source: Authors.

Table B2: A description of the variables

Symbol	Description
q1, …., q16	16-time dummy variables used in the global model. If the dwelling was sold in a given year-quarter, it takes the value 1; otherwise, it takes 0.
area	area of dwelling
age	age of the building in years
uga	distance to the nearest urban green area in metres
subway	distance to the nearest subway station
cc	distance to the city centre in metres
airbnb	monthly number of active apartments in the district in which the apartment was offered for long-term rent
quality	quality of the apartment (values 1-5)

Source: Authors.

References

Anenberg, E., S. Laufer 2017. 'A More Timely House Price Index.' *Review of Economics and Statistics* 99(4): 722–734. DOI: 10.1162/REST_a_00634.

Apergis, N. 2021. 'The role of housing market in the effectiveness of monetary policy over the Covid-19 era.' *Economics Letters* 200: 109749. DOI: 10.1016/j.econlet.2021.109749.

Barron, K., E. Kung, D. Proserpio 2018. 'The Sharing Economy and Housing Affordability: Evidence from Airbnb.' Pp. 5 in *EC '18: Proceedings of the 2018 ACM Conference on Economics and Computation.* DOI: 10.1145/3219166.3219180.

Benítez-Aurioles, B., I. Tussyadiah 2020. 'What Airbnb does to the housing market.' *Annals of Tourism Research*, published online: December 13, 2020. 2020. DOI: 10.1016/j.annals.2020.103108.

Bluwstein, K., M. Brzoza-Brzezina, P. Gelain, M. Kolasa 2020. 'Multiperiod Loans, Occasionally Binding Constraints, and Monetary Policy: A Quantitative Evaluation.' *Journal of Money, Credit and Banking* 52(7): 1691–1718. DOI: 10.1111/jmcb.12689

Chang, H.-H. (2020). 'Does the room sharing business model disrupt housing markets? Empirical evidence of Airbnb in Taiwan.' *Journal of Housing Economics* 49: 101706. DOI: 10.1016/j.jhe.2020.101706.

Del Giudice, V., P. De Paola, F. P. Del Giudice 2020. 'COVID-19 infects real estate markets: Short and mid-run effects on housing prices in Campania region (Italy).' *Social Sciences* 9(7): 114. DOI: 10.3390/SOCSCI9070114.

Fotheringham, A. S., W. Yang, W. Kang 2017. 'Multiscale Geographically Weighted Regression (MGWR).' *Annals of the American Association of Geographers* 107(6): 1247–1265. DOI: 10.1080/24694452.2017.1352480.

Garcia-López, M.-À., J. Jofre-Monseny, R. Martínez-Mazza, M. Segú 2020. 'Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona.' *Journal of Urban Economics* 119: 103278. DOI: 10.1016/j.jue.2020.103278.

Hale, T., S. Webster, A. Petherick, T. Phillips, B. Kira 2020. *Oxford COVID-19 Government Response Tracker, Blavatnik School of Government.*

Horn, K., M. Merante 2017. 'Is home sharing driving up rents? Evidence from Airbnb in Boston.' *Journal of Housing Economics* 38: 14–24. DOI: 10.1016/j.jhe.2017.08.002.

Hu, M. R., A. D. Lee, D. Zou 2021. 'COVID-19 and Housing Prices: Australian Evidence with Daily Hedonic Returns.' *Finance Research Letters*, published online: February 7, 2021. DOI: 10.1016/j.frl.2021.101960.

Kadi, J., A. Schneider, R. Seidl 2020. 'Short-term rentals, housing markets and covid-19: Theoretical considerations and empirical evidence from four austrian cities.' *Critical Housing Analysis* 7(2): 47–57. DOI: 10.13060/23362839.2020.7.2.514.

Volume 8 | Issue 1 | 2021 | 72-84
Available online at www.housing-critical.com
https://dx.doi.org/10.13060/23362839.2021.8.1.524

Koenker, R., G. Bassett 1978. 'Regression Quantiles.' *Econometrica* 46(1): 33–50. DOI: 10.2307/1913643.

Lancaster, K. J. 1966. 'A New Approach to Consumer Theory.' *Journal of Political Economy* 74(2): 132–157. Retrieved February 28, 2021, from http://www.jstor.org/stable/1828835.

Lee, D. 2016. 'How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations.' *Harvard Law & Policy Review* 10 (1): 229–253. Retrieved February 28, 2021, from https://harvardlpr.com/wp-content/uploads/sites/20/2016/02/10.1_10_Lee.pdf.

Lyons, R. C. 2019. 'Can list prices accurately capture housing price trends? Insights from extreme markets conditions.' *Finance Research Letters* 30: 228–232. DOI: 10.1016/j.frl.2018.10.004.

Marona, B., M. Tomal 2020. 'The COVID-19 pandemic impact upon housing brokers' workflow and their clients' attitude: Real estate market in Krakow.' *Entrepreneurial Business and Economics Review* 8(4): 221–232. DOI: 10.15678/EBER.2020.080412.

OECD 2020. *Housing amid Covid-19: Policy responses and challenges*. Retrieved February 28, 2021, from: https://www.oecd.org/coronavirus/policy-responses/housing-amid-covid-19-policy-responses-and-challenges-cfdc08a8/.

Qian, X., S. Qiu, G. Zhang 2021. 'The impact of COVID-19 on housing price: Evidence from China.' *Finance Research Letters*, published online: January 27, 2021. DOI: 10.1016/j.frl.2021.101944.

Rosen, S. 1974. 'Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.' *Journal of Political Economy* 82(1): 34–55. Retrieved February 28, 2021, from http://www.jstor.org/stable/1830899.

Su, C.-W., X.-Y. Cai, M. Qin, R. Tao, M. Umar 2021. 'Can bank credit withstand falling house price in China?' *International Review of Economics & Finance* 71: 257–267. DOI: 10.1016/j.iref.2020.09.013.

Tomal, M., B. Marona 2021. 'The Impact of the COVID-19 Pandemic on the Private Rental Housing Market in Poland : What Do Experts Say and What Do Actual Data Show ?' *Critical Housing Analysis* 8(1): 24–35. DOI: 10.13060/23362839.2021.8.1.520.

Trojanek, R. 2021. 'Housing Price Cycles in Poland – the Case of 18 Provincial Capital Cities in 2000–2020.' *International Journal of Strategic Property Management* 25(4): 332-345. DOI: 10.3846/ijspm.2021.14920.

Wachsmuth, D., A. Weisler 2018. 'Airbnb and the rent gap: Gentrification through the sharing economy.' *Environment and Planning A: Economy and Space* 50(6): 1147–1170. DOI: 10.1177/0308518X18778038.

Working paper 1	**Hebdzyński, M., Trojanek, R. (2024).** ***Is there a relationship between rents in the short- and long-term housing rental markets? A case study of Warsaw*****. Working paper.**

Title:

Is there a relationship between rents in the short- and long-term housing rental markets? A case study of Warsaw

Michał Hebdzyński[1], Radosław Trojanek[2]

[1]Department of Microeconomics, Poznan University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: michal.hebdzynski@phd.ue.poznan.pl, ORCID: 0000-0003-0354-054X
[2]Department of Microeconomics, Poznan University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: radoslaw.trojanek@ue.poznan.pl, ORCID: 0000-0002-8614-9484

Abstract: The affordability of housing to buy or rent for the long term (LTR) is often hindered by the development of the short-term housing rental market (STR). In principle, it should boost the tourist potential of cities and bring economic gains. Yet, its negative externalities for the housing sector may outweigh the benefits, as multiple residential apartments are withdrawn from the LTR market to be used for the needs of tourist accommodation. Based on the example of Warsaw, the capital of Poland, the research aimed to answer whether, apart from the already proven adverse relation between the STR market volume and LTR rents, there exists also a relation between their rent levels. It was hypothesised that fluctuations in STR rents induce changes in the LTR market rents, constituting a channel of transmission of shocks between the sub-markets. However, in the Granger causality analysis, the relation has been proven, neither for the whole local rental market nor its spatial segments. Nevertheless, the hedonic and repeat sales indices have revealed that rents fluctuate unevenly in the central and non-central districts of the city. It was particularly visible during the COVID-19 pandemic when rents in the non-central districts proved to be more robust. Thus, analysing spatial segments should be recommended to increase comprehension of rental market phenomena, particularly amidst the economic shocks.

Keywords: housing market; rental market; short-term rental; Airbnb, price index, COVID-19

JEL codes: C22, C55, R32

1. Introduction

The long-term housing rental market (LTR) covers agreements that satisfy housing needs and are typically concluded for 12 months. It has been proven that a developed LTR market stabilises the real estate market fluctuations, contributes to the overall macroeconomic stability (Rubaszek et al., 2024; Rubaszek & Rubio, 2020) and is an important factor for mobility in the labour market (Łaszek et al., 2021). On the other hand, the affordability of housing to rent for the long term has been recently hindered by the growth of the short-term rental market (STR), i.e. contracts that are used mainly for tourism (Guttentag et al., 2017) and last for a few days. Before the COVID-19 pandemic, in 2018, 2-5% of all apartments in the central districts of Warsaw were listed for at least one day in the STR platforms (Śledziewska et al., 2019). In Barcelona or Paris, the share equalled over 2% of all housing units (Garcia-López et al., 2020). Nevertheless, a similar situation may be encountered in any other European city that experiences a high tourist demand.

In principle, the STR market enhances resource efficiency by renting unused or vacant housing space. Moreover, it makes tourism more affordable and directs the increased money flow from tourists into tourism-related neighbourhoods and businesses (Barron et al., 2020), thus decreasing local unemployment. However, there are also multiple disadvantages of the uncontrolled growth of the STR market. Firstly, tourists disturb local communities as they are less prone to adhere to social rules. Secondly, the deregulation of the STR rentals attracts tax-evading-oriented property owners. Finally, one should name gentrification (Wachsmuth & Weisler, 2018) or hotelisation of whole city districts, which results in an outflow of apartments from the residential housing stock towards the supply of short-term tourist accommodation, which cannot be quickly supplemented by the increase of housing supply (Barron et al., 2020; Lee, 2016).

The recent literature has only recently started understanding the relationship between the STR and LTR markets. Barron et al. (2020) argued that the growth of the STR market volume comes solely at the expense of the LTR market supply, without an influence on the total housing supply. Then, according to the studies conducted on local markets by Horn & Merante (2017), Benítez-Aurioles & Tussyadiah (2020), Garcia-López et al. (2020), Barron et al. (2020), and Chaves Fonseca (2024) the presence of STR rentals exerts a positive effect on LTR rents. It has been confirmed in the study of multiple international markets by Reichle et al. (2023).

Mozo Carollo et al. (2024) showed that the effect is stronger in mid-sized cities, while Lee & Kim (2023) added that the influence is unequal when it comes to different types of housing listed in the STR market and that there may occur spillover effects of the STR market growth to neighbouring regions.

However, one might have recently observed two phenomena that led to a contraction of the STR market. First, multiple governments (either national or local) have decided to regulate the market to mitigate its negative externalities, as summarised by von Briel & Dolnicar (2021). The authors argued that the long-term impact of the restrictions would be limited in each case, as the STR market participants in the tourist areas are resilient and determined enough to adjust to any regulations introduced. Secondly, the COVID-19 pandemic that started in early 2020 influenced social and economic reality (Nicola et al., 2020), including the housing market (Dolnicar & Zare, 2020). Among them, the imposed cross-border traffic restrictions and lockdowns (in Poland, the first lockdown was introduced in late March 2020) suspended international tourism and resulted in the conversion of many apartments rented on the STR market to the LTR market purposes, increasing the LTR market supply (Boros et al., 2020; Marona & Tomal, 2020). Based on Eurostat (2024a), the number of guest nights spent in the STR apartments declined between 2019 and 2020 by 54% in Warsaw, 65% in Krakow, 67% in Paris and 74% in Barcelona. Demir et al. (2024) showed that the effects of the pandemic on the STR market might have been unequal because of the preference switch towards entire homes, in which it was possible to maintain social distancing. In the same spirit, Filieri et al. (2023) indicated that the post-pandemic STR market consumers prefer accommodations of higher quality and located in rural areas. Additionally, Guglielminetti et al. (2021) outlined the increased preference for housing in less congested areas. Finally, Gossen & Reck (2021) and Bresciani et al. (2021) highlighted the declining value of shared accommodation and hotel rooms to full apartments, similar to the ones rented in the LTR market.

Trojanek et al. (2021) studied changes in the supply of STR apartments in Warsaw to prove that a 1% change in the STR market volume led to a 0.032% change in the LTR rents before the pandemic. Yet, during the pandemic-induced market contraction, the strength of the impact was weaker and equalled 0.022%. Similarly, Mozo Carollo et al. (2024), based on the example of a mid-sized city (San Sebastian, Spain), showed that the density of the STR listings influences the LTR rents. However, the effect was remarkably stronger in the pre-pandemic period than during its first phases. Then, Batalha et al. (2022) estimated that in Lisbon, the relocation of STR properties to the LTR market contributed to the decrease of LTR rents by 4.1% – around a third of the total decrease of LTR rents at that time. Lastly, Trojanek et al. (2021) showed that the largest changes in LTR rents were noticed in districts where the outflow of apartments from the STR market was the greatest.

Even though the pre-pandemic volume of the STR market has been considered harmful to housing affordability, the STR market has developed even further in the recent five years, particularly in Poland (Eurostat, 2024a). In Warsaw, the number of guest nights transacted via the most popular STR booking platforms (Airbnb, Booking, Expedia Group, Tripadvisor) increased from 2.5 million in 2018 to over 4 million in 2023. The corresponding values for Krakow equalled 2.8 and 3.7 million and for Poznan – 0.38 and 0.66 million. Although the base numbers were considerably higher in Barcelona and Paris, they have also increased, yet to a lesser extent – from 12 million in 2018 to 12.7 million in 2023 in Barcelona and from 10.1 to 12.2 million in Paris.

Kadi et al. (2020) argued that among the incentives to enter the STR market, there are mainly higher rental income (in comparison to the LTR market), greater flexibility and, in many cases, favourable regulatory treatment (which stems from the STR market deregulation rather than from the optimistic approach of governments). Hill et al. (2023), on the example of Sydney's market, have proven the financial profitability of the conversion of the apartment from the LTR to the STR market. They argued that the average STR rent premium over LTR may equal 55% to 131%. The higher values were obtained for areas that were more attractive to tourists and for more expensive and larger apartments. Additionally, it should be noted that renting on the STR market is more challenging to track than renting on the formally regulated LTR market. As a result, it attracts also the tax-evading-oriented property owners, who aim to earn even higher financial premiums this way.

Following the literature, one may infer that to understand the behaviour of rents in the economically desired LTR market, the changes in the STR market volume should be considered. Yet, to increase comprehension of the interdependency between the STR and LTR markets (hereinafter referred to as rental market "sub-markets"), one should analyse separately the market price- and supply-related phenomena. It is based on the assumption that a rise in tourist interest in a given region may increase demand for STR rentals. This, in turn, would trigger the growth of the STR rents and lead to a higher rent premium over LTR rental. At the same time, the willingness of apartment owners to switch their properties from the LTR to the STR market would also increase, negatively affecting the LTR market volume. On the other hand, the effects of a negative demand shock on the STR market should be considered the opposite. Nevertheless, some apartment owners might not be interested in an instant switch between the markets, but they would incorporate the changed opportunity cost into the expected rent level. Thus, although part of the adjustment of the LTR market rents to the STR market demand shock would result from changes in the market size, which has been recently studied in the scientific literature, the other part may be related to the change in relative rents. To our knowledge, the latter issue has not been tested yet.

Supposedly because of the scarcity of data concerning both the STR and LTR markets, researchers to date have not constructed rent indices and have not formally studied the sub-markets' dependency in terms of their rent levels. Yet, inspecting this relation may be crucial to understanding the process of transmission of apartments between the markets, which should be linked with their relative rent levels that affect rent premiums (Hill et al., 2023). This, in turn, lies at the root of the problematic decline in housing affordability caused by the development of the STR market. Moreover, to date, the only studies that targeted the issue of rent changes in the STR market took a tourism industry perspective and analysed the whole market supply (Boto-García, 2022; Cheung, 2023). Yet, to understand the linkages between the LTR and STR markets, there is a need to analyse only this part of the STR market supply, which may be successfully used for the purposes of both sub-markets. Without a complete picture of the sub-markets’ dependency, providing suitable policy implications to mitigate the STR market's negative externalities might be difficult. It is particularly important as the recent regulations of the STR market have often shown minimal impact (Chaves Fonseca, 2024; von Briel & Dolnicar, 2021).

On the example of Warsaw, the capital of Poland, which is also the largest Polish STR market (Eurostat, 2024a), the research aims to answer whether apart from the already proven relation between the STR market volume and LTR rents there exists also a relation of their rent levels. It has been hypothesised that there is a general, city-wide causal relation in which the fluctuations in STR rents induce the changes in the LTR market rents. Furthermore, it has been suspected that the strength of the relationship may be differentiated spatially with regard to the density of the STR market apartments.

To test the stated hypotheses, the monthly rent indices of the STR and LTR markets in the period August 2015 - December 2020 have been constructed for the whole city, as well as for its central and non-central districts. Hedonic methods (Rosen, 1974) have been used for the LTR market, while for the STR market – repeat sales methods (Bailey et al., 1963). Then, the achieved rent indices have been seasonally adjusted to conduct the Granger causality analysis (Granger, 1969), based on the Vector Autoregressive (VAR) model.

Firstly, the study contributes to the literature documenting the recent changes in the STR and LTR rents. It is the first study that provides the STR market rent index for the European market of a longer time scope and the first globally concerning the spatial differentiation of STR rent changes. In this context, it shows the unequal rent fluctuations in spatial segments of the STR market that reflect the STR-dense and STR-sparse districts. To date, the subject's literature has been limited to just two scientific papers this study aims to extend – Boto-García (2022), who discussed differentiated changes in STR rents among the professional and non-professional hosts and Cheung (2023), who focused on the term structure of STR transactions. For the LTR market, the research extends recent considerations on the existence of the market's diversified segments – price-related ones, as shown by Trojanek et al. (2021) and Trojanek & Gluszak (2022), and the quality-related ones, studied by Hebdzyński (2024a).

Secondly, to our knowledge, it is the first attempt to confront rental price fluctuations in the STR and LTR markets. In this way, it adds to the recent studies that targeted the relation between the STR market volume and the LTR market rents (e.g. Reichle et al. (2023)) and extends particularly the ones that also include the considerations on the pandemic (Batalha et al., 2022; Mozo Carollo et al., 2024; Trojanek et al., 2021). Finally, it should be considered as an additional factor in understanding the economic basis of the transmission of apartments between the sub-markets, to which the foundations have been laid by Kadi et al. (2020) and Hill et al. (2023).

The rest of the paper is as follows: Section 2 presents the literature review on rent changes in the STR and LTR markets and the approaches used to measure them. Section 3 describes the utilised datasets and methods. Section 4 discusses the obtained results, while Section 5 concludes by indicating the study's limitations and the field for further research.

2. Literature review

The processes of studying the rental market phenomena are hindered by data availability. Therefore, simple aggregation methods are often used to measure rental price changes (as in

Kuk et al. (2021) or Tomal & Marona (2021)). However, they may lead to incomplete or even false conclusions, especially in times of market turbulence. Relying on mean or median, the periods in which relatively more apartments of higher quality or larger floor area, hence of higher rent, were rented might be wrongly considered as periods of a general increase in rent levels. To mitigate the problem, researchers use matching methods to construct price indices, in which they have to find consecutive observations of transactions (or eventually listings) of the same properties. The most common approach of this kind is the repeat sales method proposed by Bailey et al. (1963). In the housing sales market, certain apartments are subject to transactions more often, while some may be transacted only once. Therefore, finding multiple transactions of the same apartment might be problematic. Moreover, excluding single-transaction properties from the analytical dataset may lead to sample selection bias (Diewert, 2009). Yet, this should not be the case for the rental market, where apartments are transacted more frequently. For the paired micro-level observations, the matching methods demand no data other than individual rents for the apartments and periods of rent observation. However, the rental values of apartments tend to change over time as their quality deteriorates and apartments are subject to renovations. As a result, disregarding quality considerations and using matching models may also be considered oversimplification and lead to false conclusions. This argument should be particularly valid in the case of LTR apartments, which are subject to transactions less frequently than STR apartments, and their quality may considerably differ between the two transactions. Therefore, when possible, one should focus primarily on the quality-adjusted methods invariant to changes in the structure of the analysed sample between the studied periods, such as hedonic regression methods.

Hedonic methods are a numerical representation of the hedonic price theory developed by Rosen (1974) based on Lancaster's (1966) theory of consumer demand. Its central assumption is that the price of a heterogeneous good (as is an apartment) may be presented as a function of its non-separable attributes. Then, using the econometric methods, one may decompose prices into the marginal prices of particular housing features. Finally, based on the micro-level hedonic models, one may construct hedonic price indices (HPI) or hedonic rent indices (HRI) (Hill & Trojanek, 2022; Widłak & Tomczyk, 2010; Trojanek, 2018), which international institutions recommend for the supervision and monitoring of the housing market (European Commission, Eurostat, Organisation for Economic Co-operation and Development & World Bank, 2013).

To construct hedonic models to study the rental market one should use cross-sectional data. For this purpose, the data on micro-level transactions and the broadest possible information on individual apartments' characteristics should be considered the most reliable. However, transactional data of the quality required for hedonic modelling are hardly available. This situation may be encountered in Poland, where a high share of rental transactions is concluded without the intermediary of real estate brokers, with no obligation to report them to public institutions for purposes other than tax. Additionally, access to transactional data is severely restricted because they are a source of competitive advantage (in the case of private entities) or are covered by statistical confidentiality (in the case of public entities).

The problematic access to transactional data forces the usage of alternative sources, out of which housing listings are the most popular. This kind of data is easily accessible with the use of web-scraping algorithms, and their timeliness is among their most important advantages over transactional data that often incorporate a time lag between the moment of transaction and its reporting. Additionally, listings are rich in information on housing attributes required for hedonic modelling. Although the mentioned characteristics position them as a viable source of information on the housing market (Ahlfeldt et al., 2023; Anenberg & Laufer, 2017), also in times of economic turmoil (Lyons, 2019), one should remember that they represent only the market's supply side rents, which may divert from the market equilibrium. Thus, using them as a proxy of transactional data incorporates additional uncertainty (Nasreen & Ruming, 2022).

Supposedly because of the limited data availability, hedonic research on the recent changes in rents in the LTR market has been scarce and limited to just a few scientific articles. Yet, many of them targeted the Polish market – Gluszak & Trojanek (2024), Hebdzyński (2024a, 2024b, 2024c), Trojanek & Gluszak (2022), Trojanek et al. (2021). Among the rest, one may find the research by Micallef (2022) and Kuk et al. (2021). Concerning the analysis of the LTR market segments, Trojanek et al. (2021) and Trojanek & Gluszak (2022) showed that LTR rents may be diversified in the price-related segments, which they tested using a quantile regression approach (Koenker & Bassett, 1978). On the other hand, Hebdzyński (2024a) utilised several types of quality signals sent via listings to show the existence of quality-related market segments. Finally, Trojanek et al. (2021), who used spatial hedonic methods, showed that rent changes may be diversified not only between the local markets (as shown by Gluszak & Trojanek (2024)) but also within them.

Boto-García (2022) was, to the best of our knowledge, the first to construct the rent index of the STR market, for which he used the quality-adjusted, hedonic framework. In the paper, the division of the STR market into professional and non-professional was presented to document the heterogeneous rental price response to the pandemic shock by different STR market agents. He argued that during the worst months of the pandemic, in which tourism was severely limited, professional hosts decided on a greater reduction of rents. However, it was not possible to indicate whether the results were not disturbed by a typical, seasonal market behaviour because the time window shorter than one year was studied. Thus, the authors encouraged future researchers to study also the seasonal differences in the STR market pricing. Another notable study was conducted by Hill et al. (2023) for the market of Sydney (Australia). Using hedonic methods, the authors indicated that the STR rents constantly fell between 2016 and 2018. They argued that the change in rents is consistent with the scale of transformation of apartments from the LTR to the STR market purposes.

Finally, Cheung's (2023) study of the New Zealand’s STR market should be considered the only one to approach the topic of rent indices in a longer time window – between 2016 and 2021. From the methodological point of view, since the study utilised the data of the same structure and obtained from the same source as the dataset used in this article, it may be treated as a reference point. The data concerned STR apartment rental transactions concluded using the most popular STR market platform – Airbnb. In the constructed hedonic model, the locational explanatory variables included on the neighbourhood level have proven to be statistically significant for shaping rents. Nevertheless, the author highlighted the problem of imprecision regarding the possessed information about the location of apartments. Thus, not being able to fully control for one of the most important features of STR apartments (geolocation) and having limited information about their characteristics, he emphasised the advantageous properties of repeat sales methods, which have been selected as an analytical tool for the study. In this case, instead of using imprecise information about the location of the apartment, one may pair the subsequent observations of the same apartments using the precise identifiers generated by the listing platform. Finally, the author focused on the need to account for the term structure of the STR market transactions to obtain more reliable results concerning rent dynamics.

3. Material and methods

3.1. Data

3.1.1. Web-scraped LTR market listings

The data was web-scraped monthly from the second biggest listing website in Poland – Gratka.pl, between August 2015 and December 2020. It contains information on the following characteristics of apartments located in Warsaw: monthly rent, floor area of the apartment, its interior quality, age of the building, in which it was located, and its construction technology. It also includes information about the listed apartments' geolocation, which is provided at the district level. For most apartments, it has also been specified with a higher precision, yet to allow for comparing the LTR market with the STR market, it has been used on the less-detailed district level. Nevertheless, based on the given geo-location of the apartments and expert knowledge, the information on the building types (their age and technology) was supplemented to improve the dataset's quality. Finally, the dataset was cleaned of the duplicated observations, the ones deemed unrepresentative and those with extreme values of housing characteristics. Finally, the whole dataset consisting of 117,124 observations has been divided into subsets regarding the spatial division of the market into central districts – namely Wola and Srodmiescie (42,604 observations) and non-central ones – rest of the districts (74,520 observations).

3.1.2. STR market transactions

The dataset on the STR market was obtained from AirDNA company and extended using expert knowledge of the market. It is a complete dataset on STR rental transactions of apartments located in Warsaw, concluded via the Airbnb internet platform. The dataset includes apartments available for rent for at least one day between August 2015 and December 2020. The dataset has two dimensions. First, the apartment's characteristics include, among others, apartment ID, location (at the district level and in the form of proxy coordinates), type, size, and capacity. Moreover, the performance of each apartment in each month was provided, including occupancy rate, average daily rent, number of reservation days, and number of days of availability. The information on the quality of the offered apartments has been included in the dataset in the form of variables representing cleanliness, location, host, or overall rating. However, that information has been provided only for the last observed month of the apartment performance. Similarly, there is no information on the historical changes in the availability of housing amenities, cancellation policy, etc. As a result, none of these attributes of STR transactions can be used for historical analysis of the STR market prices.

Finally, some necessary data filtering has been done to target the precise segment of the STR market to calculate its rent index. First, only the apartments intended solely for short-term stays have been selected. Then, the observations for which it was impossible to assign them to the existing building with a precision of up to 200 metres have been excluded. Finally, the observations with contradictory information and those without any guest review have been considered unrepresentative and hence removed from the database. As a result, the final database contained 99,109 observations. They have been divided into two groups – central districts of Warsaw, where most STR apartments were located (74,904 obs.) and non-central ones, including the other 16 districts (24,205 obs.). The average daily rent for the apartment in a given month has been considered one STR transaction, but only if the apartment was rented for at least five days of the month.

3.2. Methods used

3.2.1. Repeat sales method to calculate price indices

BMN – the repeat sales method introduced by Bailey et al. (1963) has been used to calculate a repeat sales rent index (RSI) based on apartments transacted at least two times during the observation period. In the context of this study, it has been used only for the analysis of the STR market rents, the motivation for which has been described in Section 2. Then, matching the consecutive observations would allow the monthly repeat sales index to be constructed. Following Hill & Trojanek (2022), the repeat sales model may be specified as:

$$\ln(R_{im'}) - \ln(R_{im}) = \beta_{m'} - \beta_m + \ln U_{imm'} \tag{1}$$

where R_{im} is an i-th apartment's average daily rent in month m. Then, m' refers to the second month, in which the apartment was rented for at least five days ($m' > m$), parameters β reflect the estimated rent levels in consecutive periods, and $U_{imm'}$ is an error term. The RSI is calculated by exponentiation of the β parameters.

For the calculated RSIs, the seasonal adjustments have been conducted using TRAMO/SEATS procedure (Gómez & Maravall, 1997). Finally, the real values of indices have been obtained by deflating the rent indices by monthly HICP values (Eurostat, 2024b).

3.2.2. Hedonic methods

The Ordinary Least Squares (OLS) with a logarithm of rent as a dependent variable has been a baseline hedonic method used to test the stated hypotheses empirically. To construct HRI, the time-dummy approach (Hill, 2004) has been selected. The structure of the baseline model may be specified as:

$$lnR_i = \beta_0 + \sum_{j=1}^{J} \beta_j C_{i,j} + \sum_{\mathrm{k}=2}^{K} \gamma_k D_{i,k} + u_i \tag{2}$$

where R_i is a rent for an i-th apartment, $C_{i,j}$ represents a value of j-th characteristic of an i-th apartment, β_j is a parameter reflecting the estimated marginal price of j-th characteristic, $D_{i,k}$ is a time dummy indicating whether an i-th apartment was rented (or listed for rent) in k-th period, γ_k is the estimated parameter reflecting change in rents in k-th period (compared to the base period) and u_i represents model's error. Then, HRI_k value for each period (with period $k = 1$ as a base) may be calculated by exponentiation of the estimated γ_k coefficient. The seasonal adjustment and inflation correction have been conducted in line with the procedure described in Section 3.2.1. for RSI.

3.2.3. Granger causality test

Based on the assumptions made by Granger (1969), the macroeconomic variable y_1 may be considered a cause of y_2 variable if it is possible to forecast more accurately current values of y_2 using past values of y_1 than without using them, *ceteris paribus*. First, following Rubaszek (2012), one may define the vector autoregressive model in which:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} A_{p(11)} & A_{p(12)} \\ A_{p(21)} & A_{p(22)} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{3}$$

where $y_t = [y_{1t} \;\; y_{2t}]$ is a vector of two stationary time series, p is a model's lag, A_p refers to the unknown parameters describing the influence of lagged y_t on its current vales (A_0 is an intercept), and ϵ_t is a white noise process with the expected value of 0 and Σ covariance matrix. The stationarity testing has been conducted using the ADF test (Dickey & Fuller, 1979). Then, the tested hypothesis of the Granger causality test takes the form of:

$$H_0 : \Lambda_{1 \le p \le P} A_{p(12)} = 0 \tag{4}$$

In the case of rejecting H_0, one should assume that y_2 Granger-causes y_1. However, to test the simultaneous dependency, one should inspect the values of Σ covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{5}$$

and test the following hypothesis:

$$H_0 : \Sigma_{12} = 0 \tag{6}$$

3.3. Analytical procedure

A. Monthly time series of rents have been calculated. In the hedonic models, based on which the HRIs have been obtained, the explanatory variables have been included that reflected apartment floor area and squared floor area, age of the building, construction technology, apartment condition, location in the district and date of the last listing. For the STR market, RSIs have been constructed. For each sub-market, three models have been constructed that took into account the following spatial segmentation:
   - Whole Warsaw (LTR_ALL, STR_ALL)
   - Only central districts – Srodmiescie and Wola (LTR_CEN, STR_CEN)
   - Only non-central districts – all the remaining districts (LTR_NCE, STR_NCE)

B. The indices have been seasonally adjusted. Based on them, the discussion on fluctuations in STR and LTR rents has been provided.

C. For the needs of performing Granger-causality analysis, the indices have been corrected for inflation. ADF tests for stationarity of time series have been conducted, and (if necessary) the time series' differences have been calculated.

D. Granger-causality tests have been performed, comparing the indices' fluctuations in three dimensions:
   - Influence of rents in the STR market on rents in the LTR market
   - Influence of rents in the LTR market on rents in the STR market

- Simultaneous dependency of rents in the STR and LTR markets

E. Additionally, to ensure robustness, three HRIs of the STR market have been calculated, including the explanatory variables reflecting the number of bedrooms, maximum number of guests, number of bathrooms, date of observation, and location in the district. Finally, the additional RSI analysis was conducted, using exclusively the observations of STR apartments that “survived” the pandemic-induced contraction of the STR market.

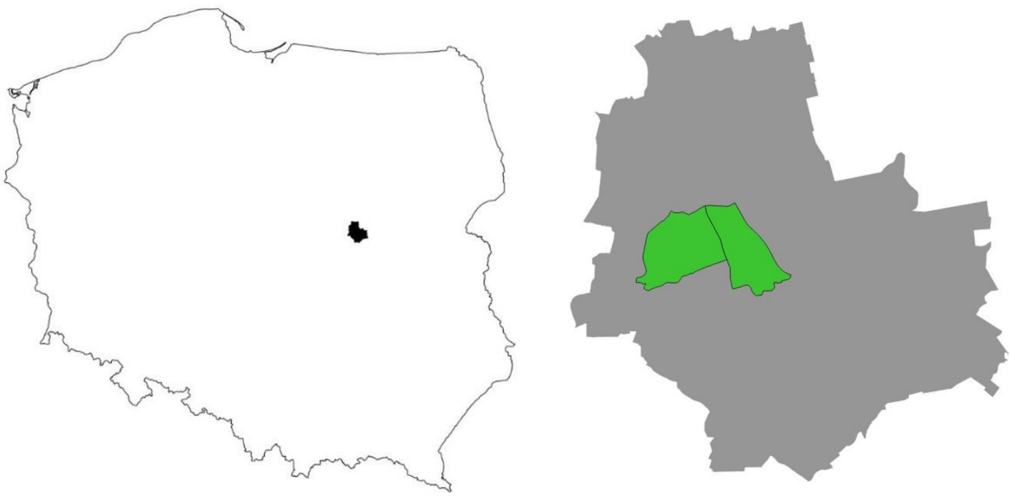


Figure 1. Location of Warsaw on the map of Poland (left panel) and location of Srodmiescie and Wola on the map of Warsaw (right panel)
Source: own elaboration

4. Empirical results and discussion

4.1. Fluctuations in rent levels

Based on the HRIs presented in Figure 2, it may be inferred that during the analysed period, LTR rents increased faster in the central districts of Warsaw than in the non-central ones. This may have been caused by the unequal distribution of the STR market growth. Thus, the size of the STR market has been studied in division into spatial segments, as presented in Figure 3. Noteworthy, the central and non-central STR market volume dynamics were almost identical. It applies both to the pre-pandemic growth and the drastic contraction of the market during the COVID times. Nevertheless, on average, the volume of the STR market was 3.2 times larger in the centrally located districts, thus the pandemic-induced outflow of apartments was much bigger there in nominal terms. Concerning the LTR market changes during the pandemic, rents decreased steadily in central and non-central districts and reached their lowest levels in

November 2020. The changes were more profound in the city centre, amounting to -13.3%, while the corresponding value for the rest of Warsaw equalled -8.1%. Nevertheless, in both spatial segments, the rent increase in December 2020 has been noticed, finally reaching the level of -5.9% and -7.6%, respectively, compared to the rent levels from February 2020.

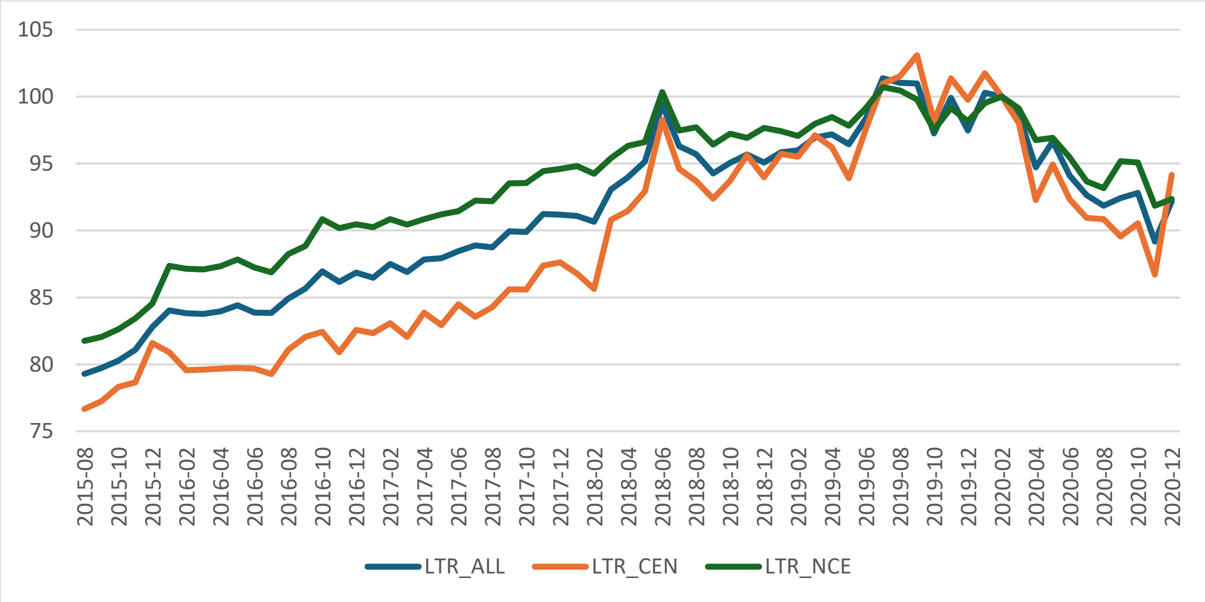


Figure 2. Seasonally adjusted hedonic rent indices of the LTR market of Warsaw (2020-02 = 100)

Source: own elaboration

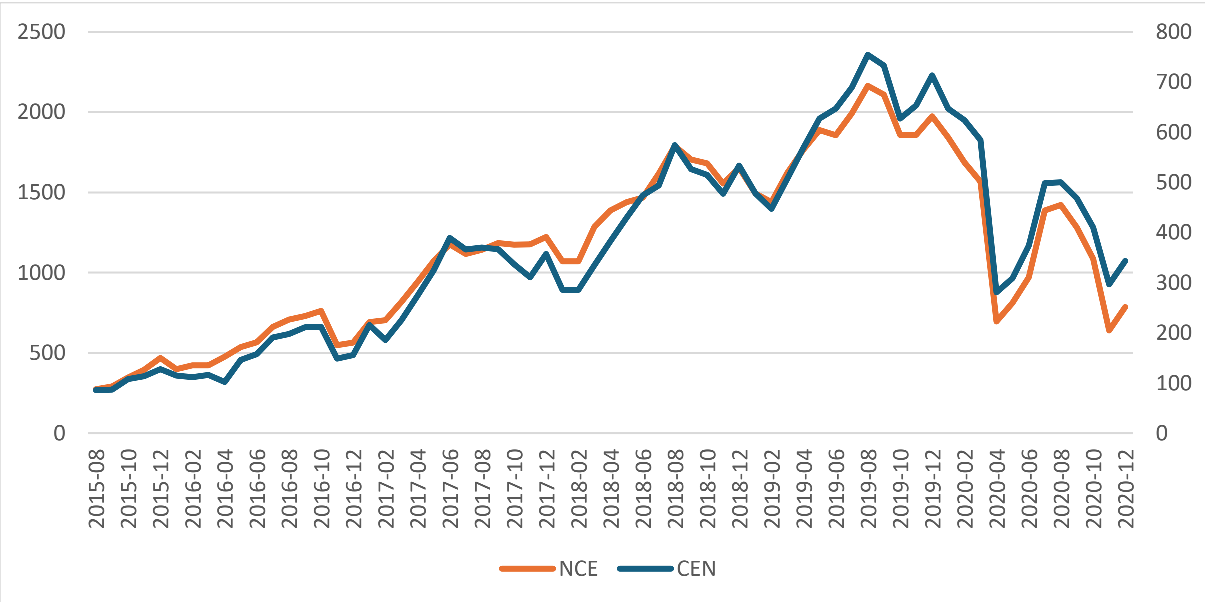


Figure 3. Size of the STR market in the centre of Warsaw (CEN, left axis) and in its non-central districts (NCE, right axis) based on the dataset of the STR market transactions. Transactions have been selected as discussed in Section 3.1.2.

Source: own elaboration

To better understand the fluctuations, one may combine the data on the number of inhabited apartments in particular districts of Warsaw (GUS, 2023) with the volume of the STR market. Based on it, the pre-pandemic share of the STR market in Srodmiescie and Wola equalled at least 1.5% of all 132.549 apartments in these districts, regardless of whether they were owned or rented. In contrast, the respective share in the other districts (treated jointly) amounted to 0.1% of 696.878 apartments. The shares should be considered lower limits – in reality, they were undoubtedly higher, as there was a non-negligible number of apartments rented on the STR market, which were not included in the analysed database that represents only the apartments rented via the Airbnb platform. Altogether, in the central districts, not only could a higher number of apartments have been converted from the STR to the LTR market, but the LTR market was also of a relatively smaller size. Therefore, rent changes were more profound in the city centre.

Concerning STR rents, presented in Figure 4, one may see that the dynamics of rents in the central and non-central districts have been similar, and only temporarily has the RSI achieved a higher index value for central Warsaw. Yet, the magnitude of the pandemic-induced changes in spatial segments might be considered unequal. In its first wave, the STR rents declined by 12.7% in the central districts and 9.2% in the non-central ones. Then, since the pandemic-related restrictions were remarkably loosened in the summer of 2020, the demand and rent levels recovered. In central districts, the scale of the rebuilt equalled 8.6 p.p. in July 2020, to be 4.1% below the pre-pandemic level. Still, in the non-central STR market, rents revived to an even larger scale, growing by 10.8 p.p. and reaching a higher level than before the crisis' outbreak. However, the second wave of the pandemic that started in October 2020 brought the rent level in the central districts to the earlier lows in November 2020 – to a value 12.4% lower compared to February 2020. For the non-central districts, the respective decrease was slightly lower and reached -6.2%.

Finally, in Figure 5, the HRI and RSI of the whole STR market in Warsaw have been compared to check the robustness of the results. Although the indices have confirmed the dynamics of rent level during the last observed year, the long-run course of indices has considerably differed. It may be attributed to the relatively low share of the variance of rents, which has been explained by the hedonic model (R^2=0.27) that may negatively affect the obtained results. Lastly, the difference between the STR_ALL index and the RSI index estimated using exclusively the observations of apartments that "survived" the pandemic has been negligible.

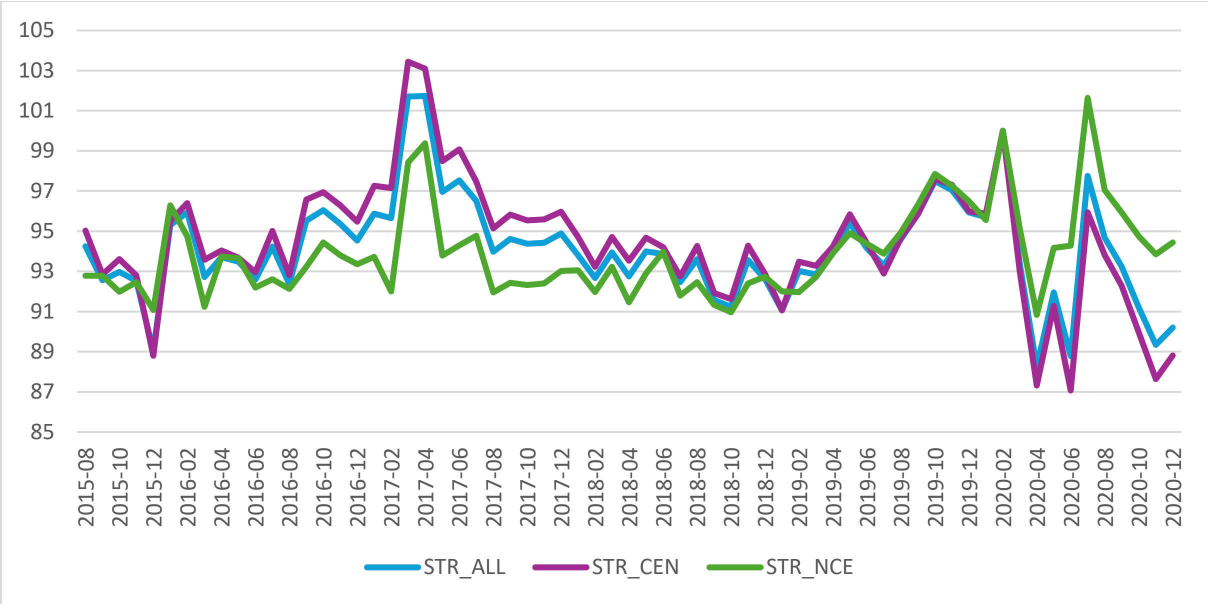


Figure 4. Seasonally adjusted hedonic rent indices of the STR market in Warsaw (2020-02 = 100)

Source: own elaboration

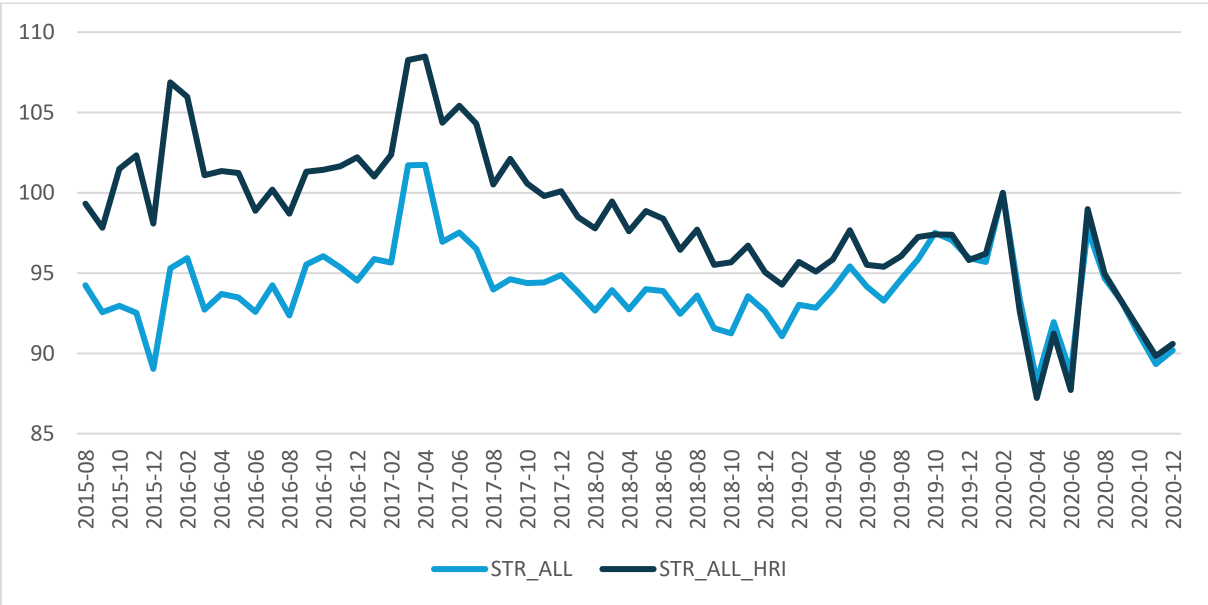


Figure 5. Seasonally adjusted hedonic rent index (STR_ALL_HRI) and repeat sales index (STR_ALL) of the STR market in Warsaw (2020-02 = 100)

Source: own elaboration

4.2. Analysis of causality

First, based on the ADF test, the indices of monthly real rents for both the STR and LTR markets have proven to be non-stationary, integrated of order 1. Therefore, the first differences of the logarithmic time series have been used for the Granger causality analysis. The results of the analysis have been presented in Table 1.

Hypothesis:			P-value of F-Test
STR_ALL	→ does not Granger-cause →	LTR_ALL	0.7222
LTR_ALL	→ does not Granger-cause →	STR_ALL	0.7569
LTR_ALL	← no instantaneous causality between →	STR_ALL	0.1901
STR_CEN	→ does not Granger-cause →	LTR_CEN	0.8301
LTR_CEN	→ does not Granger-cause →	STR_CEN	0.7502
LTR_CEN	← no instantaneous causality between →	STR_CEN	0.2749
STR_NCE	→ does not Granger-cause →	LTR_NCE	0.9526
LTR_NCE	→ does not Granger-cause →	STR_NCE	0.9476
LTR_NCE	← no instantaneous causality between →	STR_NCE	0.1665

Table 1. Results of Granger causality analysis

Source: own elaboration

Based on the presented results, in which none of the hypotheses has been rejected, it should be inferred that fluctuations in the STR market rents may not be considered to Granger-cause changes in the LTR market rents. The simultaneous dependency was also tested, and similarly, no significant relationship was found. It applies to all spatial levels of the analysis – the whole Warsaw, its central and non-central districts. Finally, the analysis was re-run on the nominal rent indices and on levels of the dependent variable (for non-logarithmic values). Still, the results indicated no Granger causality between the time series in the observed period.

5. Conclusions

The article aims to extend the knowledge of price-related phenomena in housing rental sub-markets. Firstly, it has been shown that throughout the analysed period, rent levels in Warsaw varied with regard to the spatial segment of the city. The LTR rents were increasing faster in the city centre before the pandemic, but the shock reversed the trend and contributed to the more profound decline in rents in the central districts. Although the pre-pandemic dynamics of the STR market rents have been comparable in spatial segments, the STR market rents in the non-central districts of Warsaw have proven to be more robust to the pandemic shock. In this context, the study adds to the literature that highlights the unequal responses of particular segments of the LTR market to economic shocks, particularly Trojanek et al. (2021) and Hebdzyński (2024a). Moreover, it adds to the discussion on the pandemic-induced changes in LTR rents that have been studied by Trojanek et al. (2021), Hebdzyński (2024a, 2024b), Kuk et al. (2021) and Tomal & Marona (2021).

In the main part of the study, the fluctuations in the STR and LTR market rents have been analysed to contribute to the knowledge of the price-related channel of transmission of shocks from one sub-market to another. The results obtained using the Granger causality test suggest that the adverse impact of the STR market on the LTR market should be considered in terms of the relation between the supply of apartments for STR rent and LTR rent level, as shown in the most recent studies (Chaves Fonseca, 2024; Mozo Carollo et al., 2024; Reichle et al., 2023). In this case, the LTR market is indifferent to fluctuations in STR rents as long as they do not translate to a higher or lower market supply.

From the analytical perspective, it means that to understand the process of withdrawing the apartments from the LTR market to use them for the STR market needs (thus decreasing the affordability of residential housing), one should focus on more than just price-related factors. In this case, studying rent premium, as shown by Hill et al. (2023), might be more helpful, as it combines two reflections of the demand for STRs – the achieved rent and the occupancy rate. Finally, one should conclude that based on the results of Granger causality test, the LTR market rents are insensitive to sole changes in the STR market rents. Thus they might be disregarded from the perspective of monitoring and supervision of the LTR market.

Nevertheless, one should remember that the Granger-causality test is only one way to test the relationship between two time series. Therefore, other methods of causality testing should be used to provide a final answer to the questions stated in this article. Among the limitations of the article, one should see the impossibility of providing a profound quality adjustment, which has also been discussed in a similar context by Cheung (2023). As a result, the quality-constant repeat sales method has been used to construct the STR rent index instead of hedonic methods. This strategy may pose certain risks, particularly in the long run. Thus, extending the information on the quality of apartments listed for STR rent should be considered a primary goal to improve the credibility of the results. Finally, the analysis was conducted for a relatively short time range. Thus, the interdependency of the sub-markets should be reinvestigated using information from an extended time range. It will allow us to better understand the sub-markets' price adjustment to economic shocks and to conduct a more robust time series analysis.

Bibliography:

Ahlfeldt, G. M., Heblich, S., & Seidel, T. (2023). Micro-geographic property price and rent indices. *Regional Science and Urban Economics*, *98*. https://doi.org/10.1016/j.regsciurbeco.2022.103836

Anenberg, E., & Laufer, S. (2017). A more timely house price index. *Review of Economics and Statistics*, *99*(4), 722–734. https://doi.org/10.1162/REST_a_00634

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association*, *58*(304), 933–942. https://doi.org/10.1080/01621459.1963.10480679

Barron, K., Kung, E., & Proserpio, D. (2020). The effect of home-sharing on house prices and rents: Evidence from Airbnb. *Marketing Science*, *40*(1), 23–47. https://doi.org/10.1287/mksc.2020.1227

Batalha, M., Gonçalves, D., Peralta, S., & Pereira dos Santos, J. (2022). The virus that devastated tourism: The impact of covid-19 on the housing market. *Regional Science and Urban Economics*, *95*, 103774. https://doi.org/10.1016/j.regsciurbeco.2022.103774

Benítez-Aurioles, B., & Tussyadiah, I. (2020). What Airbnb does to the housing market. *Annals of Tourism Research*, *90*, 103108. https://doi.org/10.1016/j.annals.2020.103108

Boros, L., Dudás, G., & Kovalcsik, T. (2020). The effects of COVID-19 on airbnb. *Hungarian Geographical Bulletin*, *69*(4), 363–381. https://doi.org/10.15201/hungeobull.69.4.3

Boto-García, D. (2022). Heterogeneous price adjustments among Airbnb hosts amid COVID-19: Evidence from Barcelona. *International Journal of Hospitality Management*, *102*, 103169. https://doi.org/10.1016/j.ijhm.2022.103169

Bresciani, S., Ferraris, A., Santoro, G., Premazzi, K., Quaglia, R., Yahiaoui, D., & Viglia, G. (2021). The seven lives of Airbnb. The role of accommodation types. *Annals of Tourism Research*, *88*, 103170. https://doi.org/10.1016/j.annals.2021.103170

Chaves Fonseca, C. (2024). Short-term rentals and residential rents: evidence from a regulation in Santa Monica. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-01-2024-0001

Cheung, K. S. (2023). Airbnb pricing and term structure: A temporal analysis of omitted variable bias and repeat sales method as remedies. *Tourism Economics*, *30*(5), 1330–1338. https://doi.org/10.1177/13548166231191551

Demir, E., Zoğal, V., Yelsiz, M. Ş., & Emekli, G. (2024). Mapping the effects of COVID-19 on Airbnb: Evidence from European cities. *Anatolia*, *35*(1), 135–162. https://doi.org/10.1080/13032917.2023.2254780

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.2307/2286348

Diewert, W. E. (2009). The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions. pp 87-116 in W. E. Diewert, B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (Ed.), Price and Productivity Measurement: Volume 1 – Housing. Trafford Press.

Dolnicar, S., & Zare, S. (2020). COVID19 and Airbnb – Disrupting the Disruptor. *Annals of Tourism Research*, *83*(102961). https://doi.org/10.1016/j.annals.2020.102961

European Commission, Eurostat, Organisation for Economic Co-operation and Development & World Bank. (2013). *Handbook on Residential Property Prices (RPPIs): 2013 edition*. OECD Publishing. https://doi.org/10.2785/34007

Eurostat. (2024a). *Guest nights spent at short-stay accommodation offered via collaborative economy platforms*. https://doi.org/10.2908/TOUR_CE_OAN3

Eurostat. (2024b). *HICP - monthly data (index)*. https://doi.org/10.2908/PRC_HICP_MIDX

Filieri, R., Milone, F. L., Paolucci, E., & Raguseo, E. (2023). A big data analysis of COVID-19 impacts on Airbnbs' bookings behavior applying construal level and signaling theories. *International Journal of Hospitality Management*, *111*, 103461. https://doi.org/10.1016/j.ijhm.2023.103461

Garcia-López, M. À., Jofre-Monseny, J., Martínez-Mazza, R., & Segú, M. (2020). Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona. *Journal of Urban Economics*, *119*(June 2019). https://doi.org/10.1016/j.jue.2020.103278

Gluszak, M., & Trojanek, R. (2024). War in Ukraine, the refugee crisis, and the Polish housing market. *Housing Studies*. https://doi.org/10.1080/02673037.2024.2334822

Gómez, V., & Maravall, A. (1997). Programs TRAMO and SEATS: instructions for the user. *Banco de España - Servicio de Estudios*, *9805*.

Gossen, J., & Reck, F. (2021). The End of the Sharing Economy? Impact of COVID-19 on Airbnb in Germany. *Economic Research Guardian*, *11*(2), 255–269.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Guglielminetti, E., Loberto, M., Zevi, G., & Zizza, R. (2021). Living on my Own: The Impact of the Covid-19 Pandemic on Housing Preferences. *Bank of Italy Occasional Paper No. 627*. https://doi.org/10.2139/ssrn.3891671

GUS (Główny Urząd Statystyczny). (2023). *Narodowy Spis Powszechny 2021 - Mieszkania zamieszkane wg okresu budowy budynku*.

Guttentag, D., Smith, S., Potwarka, L., & Havitz, M. (2017). Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study. *Journal of Travel Research*, *57*(3), 342–359. https://doi.org/10.1177/0047287517696980

Hebdzyński, M. (2024a). Are Hedonic Models Really Quality-Adjusted? The Role of Apartment Quality in Hedonic Models of Housing Rental Market. *Real Estate Management and Valuation*, *32*(2), 46–57. https://doi.org/10.2478/remav-2024-0014

Hebdzyński, M. (2024b). Effects of the COVID-19 pandemic and the war in Ukraine on the local housing rental market in Poland. *Journal of International Studies*, *17*(2), 298–323. https://doi.org/10.14254/2071-8330.2024/17-2/16

Hebdzyński, M. (2024c). Price-setting factors or revealed preferences? How to understand the results of hedonic models and hedonic indices of the housing rental market that base on listings data? *Bank i Kredyt*, *55*(4), 519–550.

Hill, R., J. (2004). Constructing Price Indexes across Space and Time: The Case of the European Union. *American Economic Review, 94*(5): 1379–1410.

Hill, R. J., Pfeifer, N., & Steurer, M. (2023). The Airbnb rent premium and the crowding-out of long-term rentals. *Journal of Housing Economics*, *61*, 101935. https://doi.org/10.1016/j.jhe.2023.101935

Hill, R. J., & Trojanek, R. (2022). An evaluation of competing methods for constructing house price indexes: The case of Warsaw. *Land Use Policy*, *120*, 106226. https://doi.org/10.1016/j.landusepol.2022.106226

Horn, K., & Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. *Journal of Housing Economics*, *38*, 14–24. https://doi.org/10.1016/j.jhe.2017.08.002

Kadi, J., Schneider, A., & Seidl, R. (2020). Short-term rentals, housing markets and covid-19: Theoretical considerations and empirical evidence from four austrian cities. *Critical Housing Analysis*, *7*(2), 47–57. https://doi.org/10.13060/23362839.2020.7.2.514

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Kuk, J., Schachter, A., Faber, J. W., & Besbris, M. (2021). The COVID-19 Pandemic and the Rental Market: Evidence From Craigslist. *American Behavioral Scientist, 65*(12). https://doi.org/10.1177/00027642211003149

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. https://www.jstor.org/stable/1828835

Łaszek, J., Augustyniak, H., & Olszewski, K. (2021). The development of the rental market in Poland. In R. Sobiecki, J. Łaszek, & K. Olszewski (Eds.), *Real Estate at Exposure. New Challenges, Old Problems*. Warsaw School of Economics.

Lee, D. (2016). How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harvard Law & Policy Review*, *10*, 229–253.

Lee, S., & Kim, H. (2023). Four shades of Airbnb and its impact on locals: A spatiotemporal analysis of Airbnb, rent, housing prices, and gentrification. *Tourism Management Perspectives*, *49*, 101192. https://doi.org/10.1016/j.tmp.2023.101192

Lyons, R. C. (2019). Can list prices accurately capture housing price trends? Insights from extreme markets conditions. *Finance Research Letters*, *30*, 228–232. https://doi.org/10.1016/j.frl.2018.10.004

Marona, B., & Tomal, M. (2020). The COVID-19 pandemic impact upon housing brokers' workflow and their clients' attitude: Real estate market in Krakow. *Entrepreneurial Business and Economics Review*, *8*(4), 221–232. https://doi.org/10.15678/EBER.2020.080412

Micallef, B. (2022). Hedonic indices for registered rents in Malta. *The Annual Malta Residential Rental Study Second Edition*, Ministry for Social and Affordable Accomodation, Housing Authority

Mozo Carollo, I., Morandeira-Arca, J., Etxezarreta-Etxarri, A., & Izagirre-Olaizola, J. (2024). Is the effect of Airbnb on the housing market different in medium-sized cities? Evidence from a Southern European city. *Urban Research & Practice*, *17*(2), 260–279. https://doi.org/10.1080/17535069.2023.2179419

Nasreen, Z., & Ruming, K. (2022). Struggles and opportunities at the platform interface: tenants' experiences of navigating shared room housing using digital platforms in Sydney. *Journal of Housing and the Built Environment*, *37*(3), 1537–1554.

https://doi.org/10.1007/s10901-021-09909-x

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, *78*, 185–193. https://doi.org/10.1016/j.ijsu.2020.04.018

Reichle, P., Fidrmuc, J., & Reck, F. (2023). The sharing economy and housing markets in selected European cities. *Journal of Housing Economics*, *60*, 101914. https://doi.org/10.1016/j.jhe.2023.101914

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Rubaszek, M. (2012). *Modelowanie polskiej gospodarki z pakietem R*. Oficyna Wydawnicza - Szkoła Główna Handlowa.

Rubaszek, M., & Rubio, M. (2020). Does the rental housing market stabilize the economy? A micro and macro perspective. *Empirical Economics*, *59*(1), 233–257. https://doi.org/10.1007/s00181-019-01638-z

Rubaszek, M., Stenvall, D., & Uddin, G. S. (2024). Rental market structure and housing dynamics: An interacted panel VAR investigation. *International Journal of Finance & Economics*. https://doi.org/10.1002/ijfe.2937

Śledziewska, K., Gyódi, K., Hardy, W., & Mazur, J. (2019). *Airbnb w Warszawie. Charakterystyka rynku i wyzwania dla miasta*. https://www.delab.uw.edu.pl/wp-content/uploads/2020/09/DELab_airbnb_prezentacja.pdf

Tomal, M., & Marona, B. (2021). The impact of the covid-19 pandemic on the private rental housing market in poland: What do experts say and what do actual data show? *Critical Housing Analysis*, *8*(1), 24–35. https://doi.org/10.13060/23362839.2021.8.1.520

Trojanek, R. (2018). *Teoretyczne i metodyczne aspekty wyznaczania indeksów cen na rynku mieszkaniowym*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.

Trojanek, R., & Gluszak, M. (2022). Short-run impact of the Ukrainian refugee crisis on the housing market in Poland. *Finance Research Letters*, *50*(103236). https://doi.org/10.1016/j.frl.2022.103236

Trojanek, R., Gluszak, M., Tanas, J., & Hebdzyński, M. (2021). The COVID-19 pandemic , Airbnb and housing market dynamics in Warsaw. *Critical Housing Analysis*, *8*(1), 72–84. https://doi.org/10.13060/23362839.2021.8.1.524

von Briel, D., & Dolnicar, S. (2021). The evolution of Airbnb regulation - An international longitudinal investigation 2008–2020. *Annals of Tourism Research*, *87*, 102983. https://doi.org/10.1016/j.annals.2020.102983

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A*, *50*(6), 1147–1170. https://doi.org/10.1177/0308518X18778038

Widłak, M., & Tomczyk, E. (2010). Measuring price dynamics: evidence from the Warsaw housing market. *Journal of European Real Estate Research*, *3*(3), 203–227. https://doi.org/10.1108/17539261011094722