MISCELLANEA

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 25th, 50th (median) and 75th 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 (25^{th} percentile – *MOD_6_Q25*) have proven to increase their rents to a higher degree than the most expensive ones (75^{th} 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 55^{th} – 75^{th} 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.

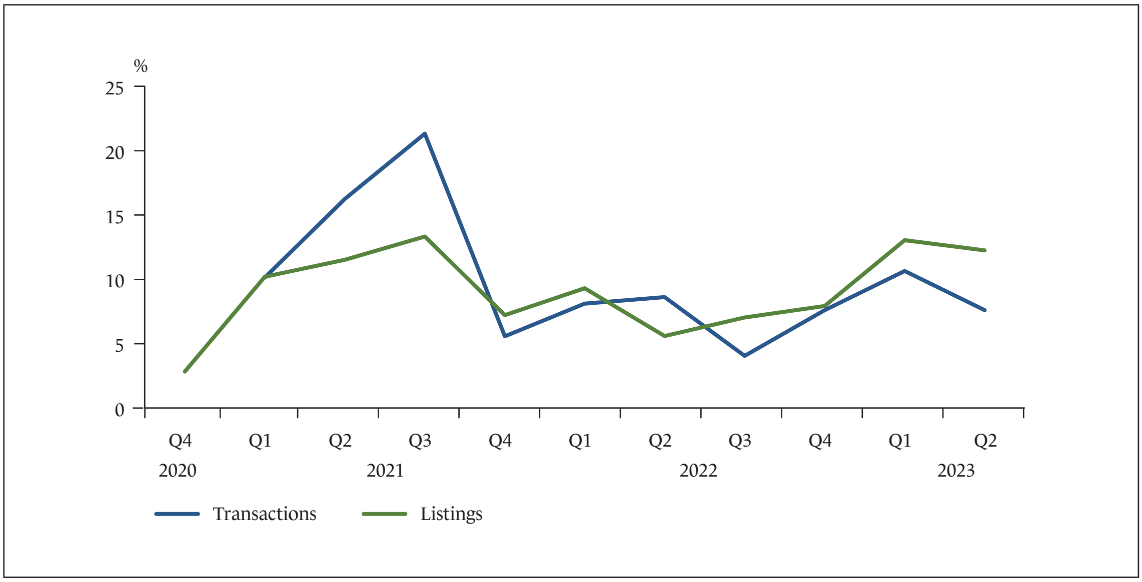
Widłak M., Nehrebecka N. (2011), Wykorzystanie regresji kwantylowej w analizie zróżnicowania cen mieszkań, *Wiadomości Statystyczne*, 5(600).

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.

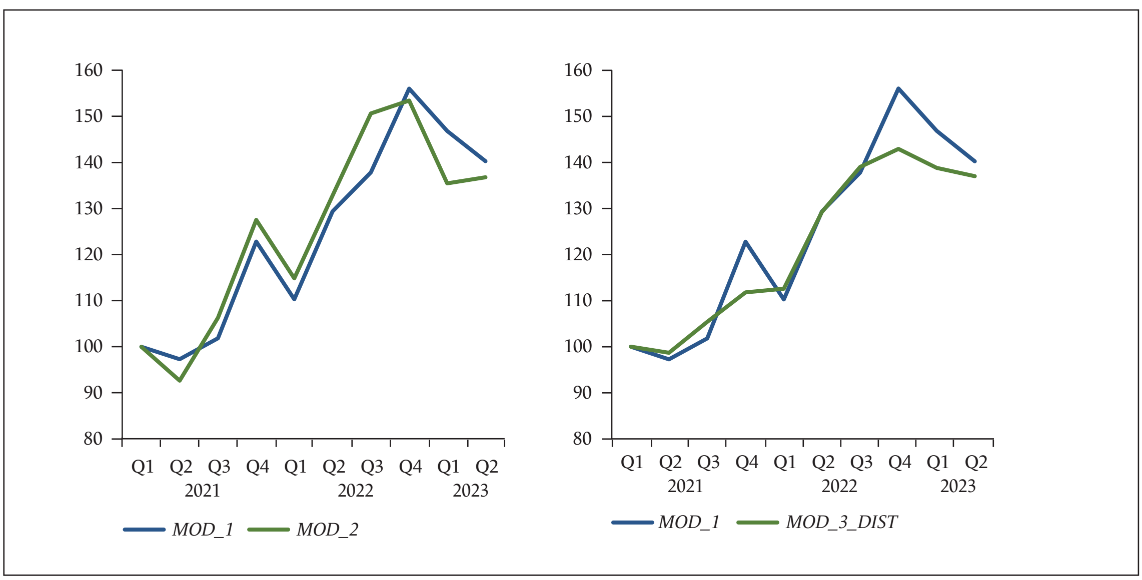
Appendix

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*

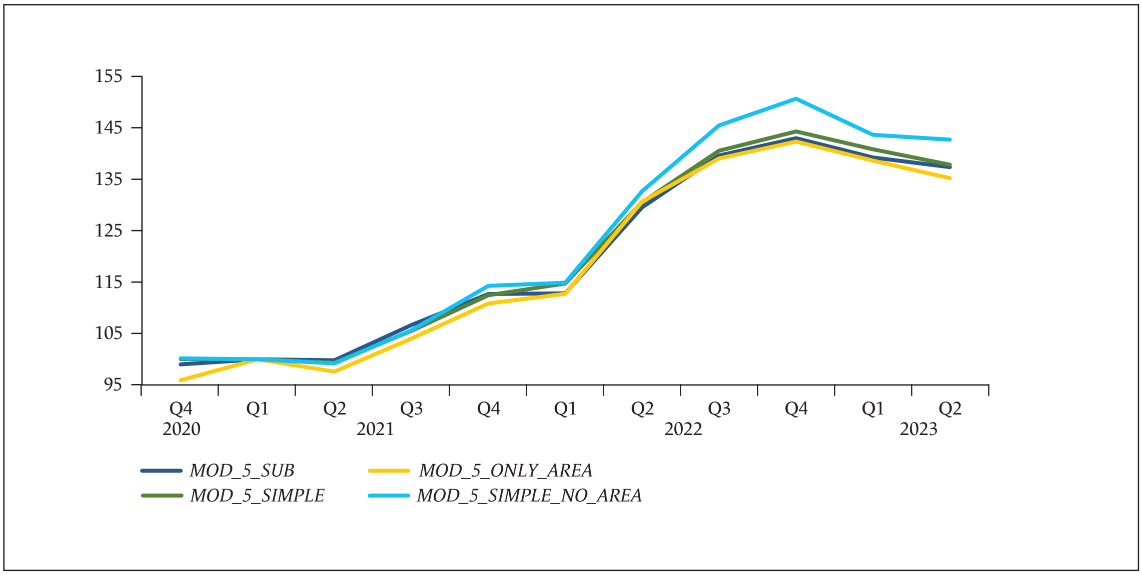
The left panel represents the deviation of coefficients for apartment characteristics. The right panel represents the deviation of the q-q dynamics of HRIs. The dashed line represents the polynomial trend line.

Source: own elaboration.

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

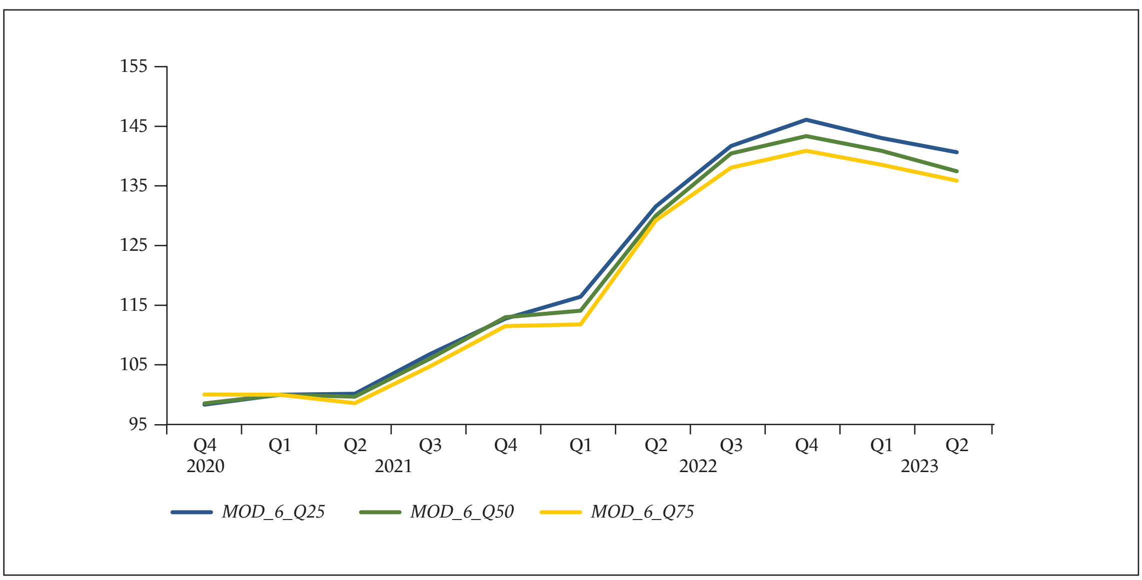
Source: own elaboration.

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 ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 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 ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 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 ≤ 0.01; ** for P ≤ 0.05; * for P ≤ 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