

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.hebdzyński@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

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 & Gluszek, 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 Gluszek (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

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	
<i>RENT (dependent variable)</i>	rent for the apartment [in PLN]	650	1961.5	9800		
<i>AREA</i>	floor area of the apartment [in m ²]	15	47.5	150		
<i>ROOMS_INTENSITY</i>	rooms per 1 m ² of the apartment	0.011	0.043	0.08		
<i>MODERN_APART</i>	1- if the apartment is located in the modern apartment building, 0- otherwise					2162
<i>RESTORED_TENEMENT</i>	1- if the apartment is located in the restored tenement building, 0- otherwise					590
<i>FULL_FURNISHING</i>	1- if the apartment is fully or almost fully equipped with furniture, 0- otherwise					7096
<i>BALCONY</i>	1- if there is a balcony/terrace in the apartment, 0- otherwise					6066
<i>PARKING</i>	1- if there is access to the designated parking space (included in the listed rent), 0- otherwise					1660
<i>GARDEN</i>	1- if there is an access to the private garden, 0- otherwise					257
<i>DIST_CC</i>	distance of the apartment to the city center [in km]	0.021	2.982	10.916		
<i>DIST_GREEN</i>	distance of the apartment to the nearest urban green area [in km]	0.002	0.268	0.996		
<i>DIST_LAKE</i>	distance of the apartment to the nearest lake [in km]	0.044	2.226	6.406		
<i>Q_VISUAL_HIGH</i>	1- if the indirect visual signal of quality indicated that the quality of the apartment is high, 0- otherwise					2052
<i>Q_VISUAL_MEDIUM</i>	1- if the indirect visual signal of quality indicated that the quality of the apartment is medium, 0- otherwise					5425
<i>Q_VISUAL_LOW</i>	1- if the indirect visual signal of quality indicated that the quality of the apartment is low, 0- otherwise					771
* <i>Q_DIR_TEXTUAL_HIGH</i>	1- if the direct textual signal of quality indicated that the quality of the apartment is high, 0 - otherwise					439
* <i>Q_DIR_TEXTUAL_GOOD</i>	1- if the direct textual signal of quality indicated that the quality of the apartment is good, 0- otherwise					172
* <i>Q_DIR_TEXTUAL_RENOVATED</i>	1- if the direct textual signal of quality indicated that the apartment is freshly renovated, 0- otherwise					51
* <i>Q_DIR_TEXTUAL_NOSIGNAL</i>	1- if the direct textual signal of quality has not been sent by the listing provider, 0- otherwise					1129
<i>Q_INDIR_TEXTUAL</i>	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:

$$\ln R_i = X_{ik} \beta_{\theta k} + \varepsilon_{\theta i} \quad (3)$$

$$\text{with } Q_{\theta}(\ln R_i | X_{ik}) = X_{ik} \beta_{\theta k} \quad (4)$$

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 θ th quantile, $\varepsilon_{\theta i}$ is an error term and $Q_{\theta}(\ln R_i | X_{ik})$ represents the θ th 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 (50th 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

9th, 75th and 42nd conditional percentile of the dependent variable – where the 42nd percentile reflects the midpoint between the 9th and 75th 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 $P \leq 0.01$; ** for $P \leq 0.05$; * for $P \leq 0.1$.

Source: own elaboration.

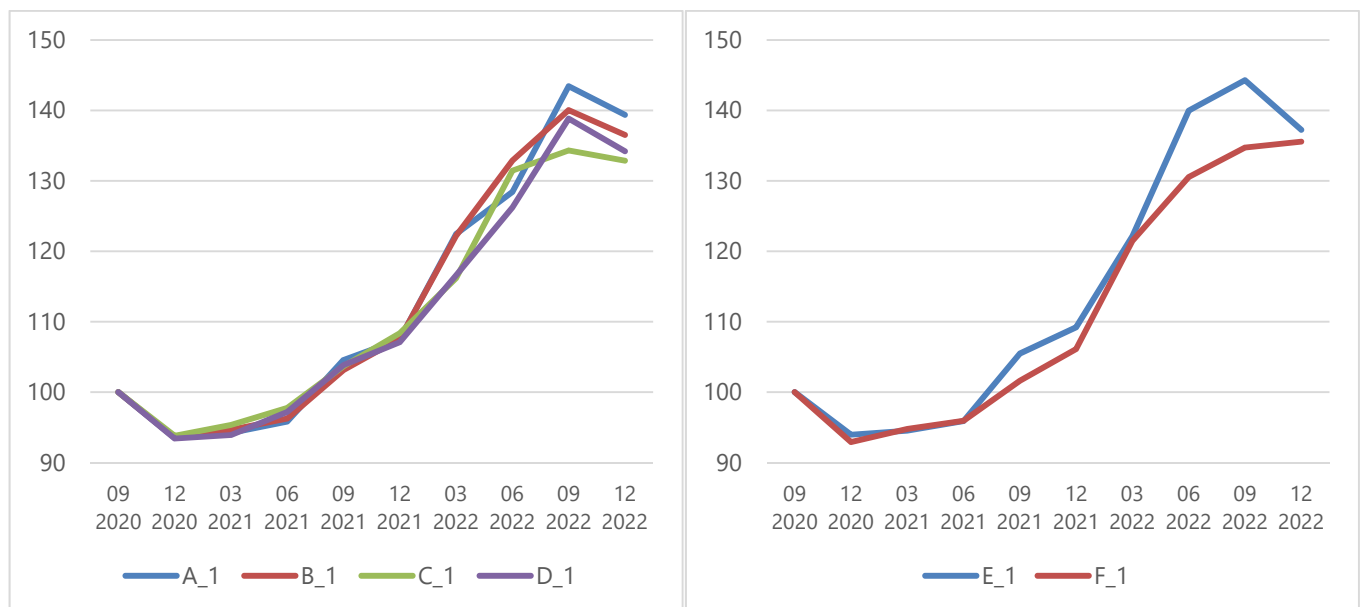


Fig. 1. Comparison of hedonic rent indices obtained based on OLS models with different approaches to accounting for apartment quality (09 2020=100). Source: own elaboration.

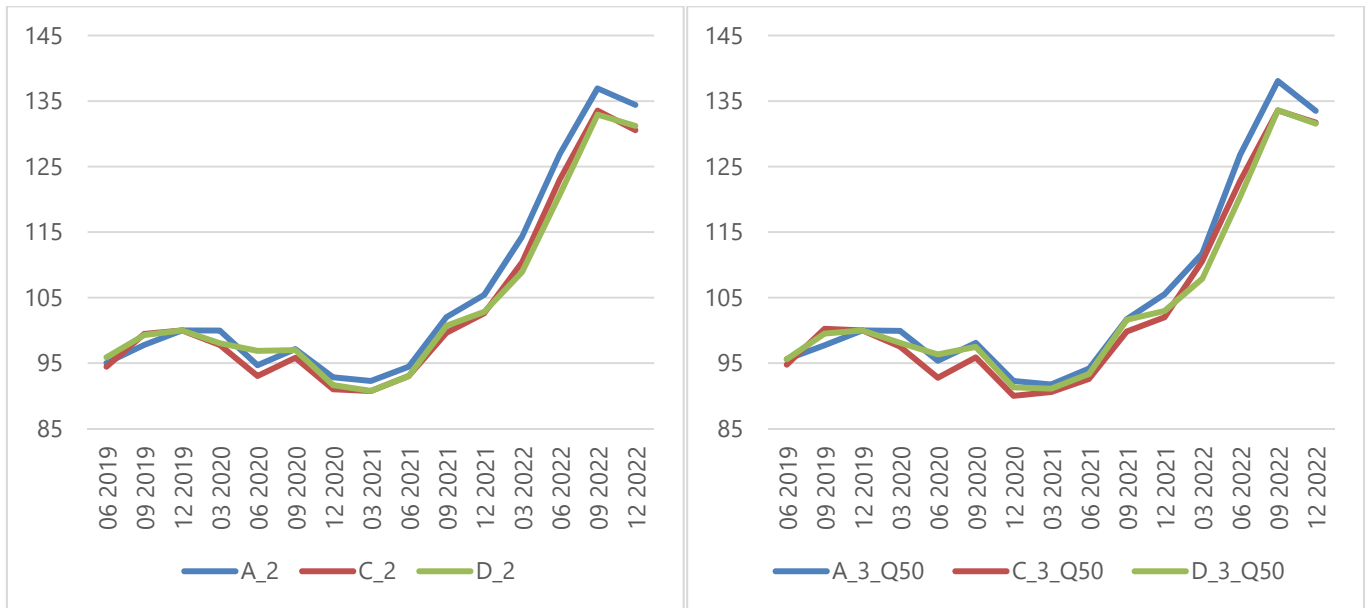


Fig. 2. Comparison of hedonic rent indices obtained based on OLS (left panel) and median QR (right panel) models with different approaches to accounting for apartment quality (12 2019=100). *Source:* own elaboration.

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

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.

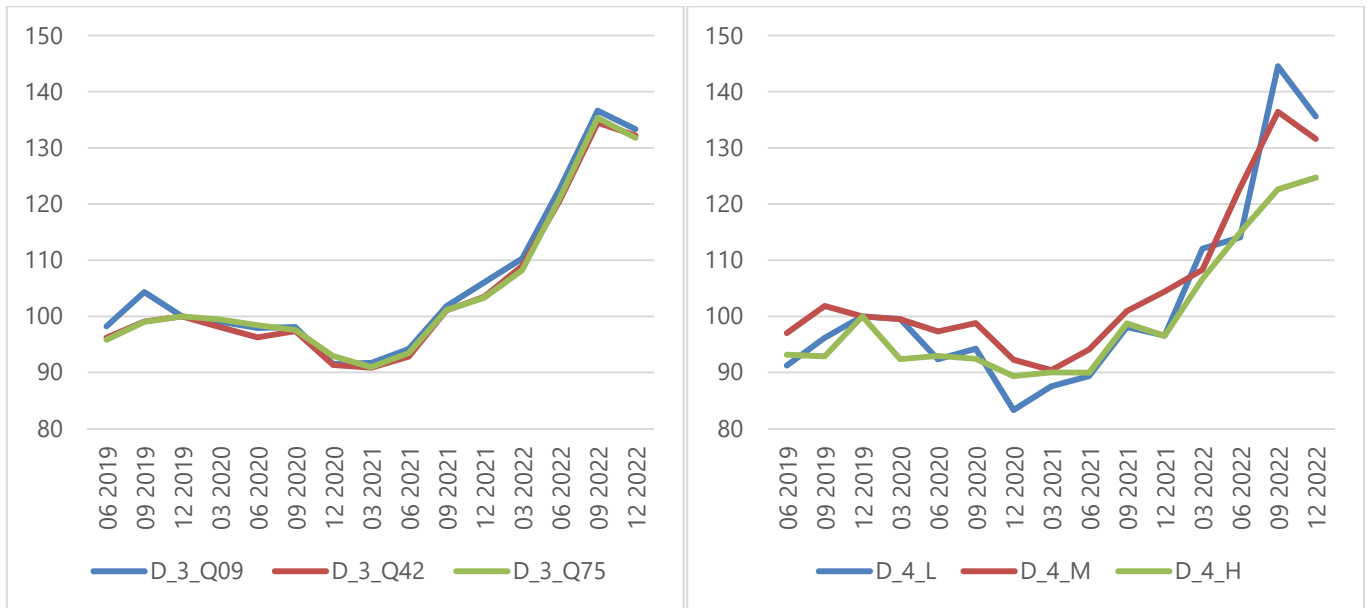


Fig. 3. Comparison of hedonic rent indices obtained in QR models for price-related (left panel) and quality-related (right panel) market segments (12 2019=100). Source: own elaboration.

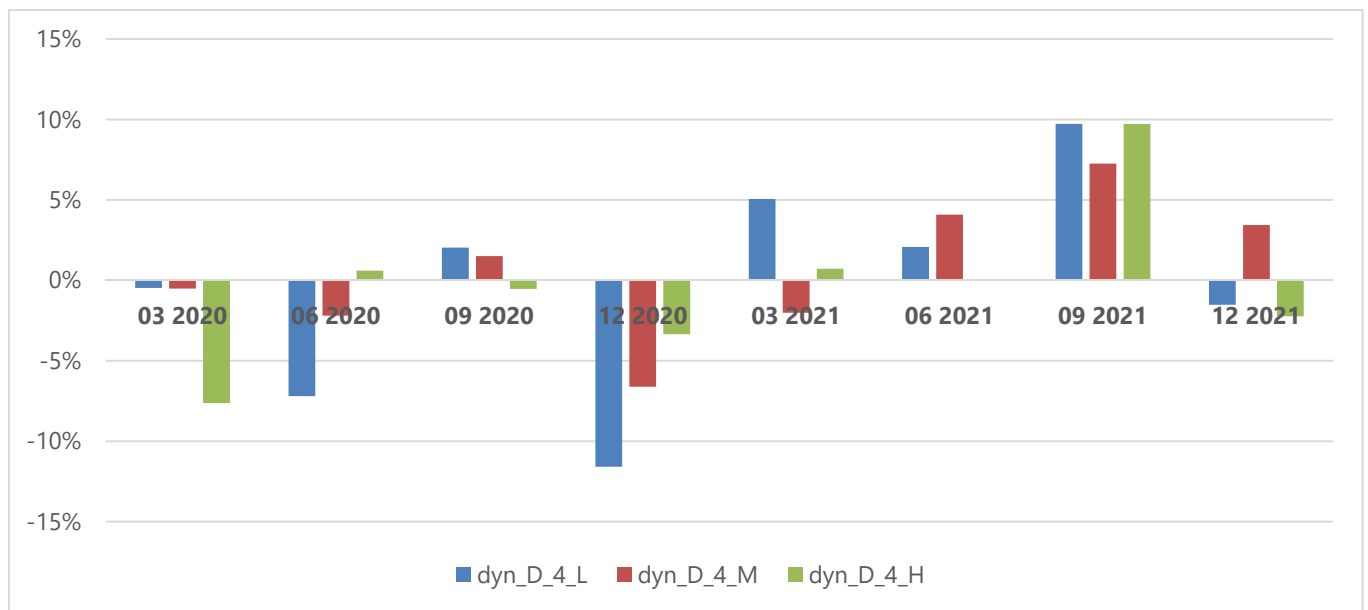


Fig. 4. Quarter-to-quarter dynamics of hedonic rent indices based on median QR models for quality-related segments during the COVID-19 pandemic. Source: own elaboration.

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

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).

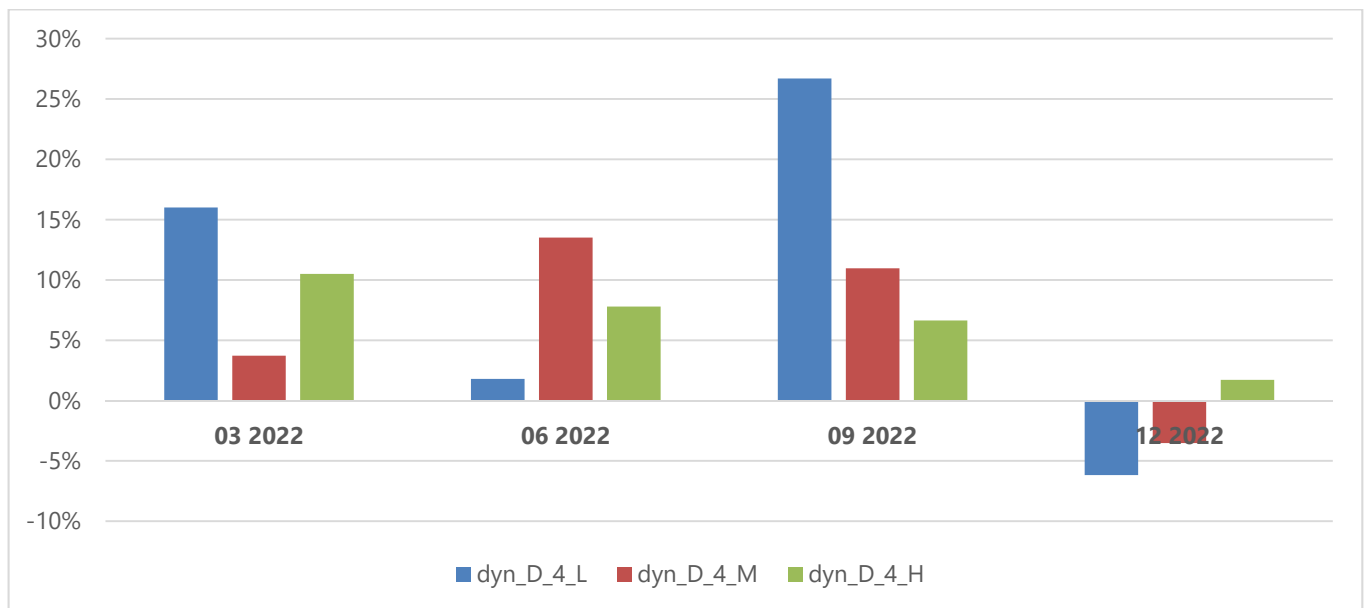


Fig. 5. Quarter-to-quarter dynamics of hedonic rent indices based on median QR models for quality-related segments during the war in Ukraine. Source: own elaboration.

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 (RPPi): 2013 edition*. OECD Publishing. <https://doi.org/10.2785/34007>

- 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., & Trojane, 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\(472\)](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](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>
- Trojane, R., & Gluszek, 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>
- Trojane, R., Gluszek, 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>
- 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>