

THE HETEROGENEOUS PRICING EFFECT OF TEXTUAL SENTIMENT IN REAL ESTATE LISTINGS: EVIDENCE FROM KRAKOW, POLAND

Bartosz Wilczyński¹, Joanna Węgrzyn^{1*}

¹ Department of Real Estate and Investment Economics, Krakow University of Economics, Krakow, Poland, (JW) email: wegrzynj@uek.krakow.pl, ORCID: 0000-0003-3327-6313

* Corresponding author

ARTICLE INFO	ABSTRACT
Keywords: residential real estate; online property listings; sentiment analysis; hedonic modelling; textual information; housing prices	Textual information contained in online property listings has become an increasingly important source of market signals, yet the economic value of linguistic features remains insufficiently understood. This study examines whether sentiment expressed in residential property descriptions is associated with asking prices and whether this relationship depends on the informational context of the listing. Drawing on signalling theory and information asymmetry, the study argues that sentiment should be viewed as a conditional signal rather than a universal price premium. The paper analyzes residential property listings from Krakow, Poland, and applies hedonic pricing models incorporating textual sentiment measures and interaction terms. The results show that the relationship between sentiment and price is heterogeneous rather than uniform. Positive sentiment is associated with higher asking prices particularly in contexts characterized by greater information uncertainty, including tenement buildings. The findings further indicate that seller type influences the level of sentiment expressed in listings but does not significantly moderate its price effect. Overall, the study contributes to the literature by showing that textual sentiment functions as a context-dependent market signal, with its value determined by the informational characteristics of the property and listing environment.
JEL Classification: C55, D82, R31	
Citation: Ahead of Print	
Wilczyński, B., & Węgrzyn, J. (2026). The heterogeneous pricing effect of textual sentiment in real estate listings: Evidence from Krakow, Poland. <i>Real Estate Management and Valuation</i> , [Ahead of print]. https://doi.org/10.2478/remav-2026-0037	

1. Introduction

Real estate listings are not neutral documents. They are marketing instruments in which sellers deploy language, tone, and emotional register to communicate property quality, negotiate perceived value, and attract buyers in a market characterized by fundamental information asymmetry. Over the past two decades, the application of text mining to property listings has demonstrated that language measurably affects market outcomes: certain word choices correlate with transaction prices (Levitt & Syverson, 2008); the rhetorical style of professional agents differs systematically from that of private sellers (Pryce & Oates, 2008); and sentiment scores computed from listing descriptions contribute to the predictive accuracy of hedonic price models (Lawani et al., 2019; Ploessl & Just, 2024).

Yet the dominant framing of this literature treats the pricing effect of textual sentiment as a global, average phenomenon: a listing with a higher positive sentiment score commands a higher price, *ceteris paribus*. This framing assumes that the emotional valence of a property description affects buyers' willingness to pay in the same direction and to the same degree across all property types, building categories, and listing contexts. This assumption has not been empirically tested.

This paper demonstrates empirically that the pricing effect of textual sentiment in real estate listings is heterogeneous and context-dependent. These findings carry theoretical significance beyond their empirical content. They reveal that the economic value of textual sentiment in real estate is not a universal feature of all listings but rather an information-contingent premium: sentiment functions as a pricing signal specifically

where other objective information channels are limited, ambiguous, or insufficient to resolve buyer uncertainty. In terms of Spence's (1973) signalling model, positive textual sentiment is a credible and priced signal in high-asymmetry contexts; in low-asymmetry contexts, where structural attributes speak for themselves, it adds no independent pricing information.

This argument contributes to the literature in three ways. First, the paper determines whether the offers listed by agents differ due to their emotional tone from offers listed by individual owners. Second, it goes beyond the basic finding that "sentiment affects prices" and shows under what conditions this effect actually holds - a more informative contribution than reporting a single average effect. Third, it investigates which characteristics of the property and the listing shape how sentiment translates into price: sentiment is not valued the same way in every listing, and what determines its value is the amount of information available about the property.

The paper proceeds as follows. Section 2 reviews the theoretical and empirical literature on information asymmetry, signalling, sentiment analysis in property markets, and hedonic models with textual attributes. Section 3 describes the data and empirical methodology. Section 4 presents the results. Section 5 discusses the findings in relation to the theoretical framework. Section 6 concludes the paper.

2. Literature Review

2.1. Information Asymmetry and Signalling in Real Estate Markets

Real estate transactions are paradigm instances of markets with asymmetric information. Sellers possess detailed knowledge of the property's condition, history, and defects that buyers cannot fully observe prior to purchase. In Akerlof's (1970, 2023) canonical formulation, such asymmetry can cause markets to unravel or clear at quality-discounted prices, as buyers rationally discount their offers to account for the possibility of adverse selection. The extent of this quality discount - and whether it is overcome - depends critically on the information transmission mechanisms available to market participants.

Spence (1973) formalized signalling as the mechanism by which informed sellers communicate otherwise unobservable quality to uninformed buyers through observable, costly actions. For a signal to be informative in a separating equilibrium, it must be differentially costly: high-quality sellers must be able to bear the cost of signalling more easily than low-quality

sellers. In real estate markets, pricing, staging, photography, and - as this paper argues - the emotional register of listing language all function as potential signalling instruments.

The implications of information asymmetry for agent-intermediated real estate markets have been extensively studied. Levitt and Syverson (2008) provided compelling evidence that agents possess an informational advantage over their clients: homes owned by real estate agents sell at higher prices than comparable homes sold by clients on whose behalf agents act, suggesting that agents deploy insider knowledge for their own benefit. Pryce and Oates (2008), extending this analysis to the linguistic dimension, demonstrated that agents have developed a professional rhetorical style that varies over time and across regions - a style whose consistency and market-calibration reflects the informational advantage accumulated through repeated market participation.

Early evidence that listing language itself may influence market outcomes was provided by Haag et al. (2000), who showed that textual comments included in Multiple Listing Service (MLS) advertisements were associated with both asking prices and time on the market. Importantly, they also observed that not all expressions convey meaningful information, distinguishing informative descriptions from "hype"-language that attracts attention but contributes little to buyers' assessment of property quality. Bond and Devine (2016) further demonstrated that specific informational cues embedded in listing descriptions, such as references to environmentally sustainable features, can generate measurable market premiums, with LEED-certified buildings achieving the largest rental premium.

More recently, the information asymmetry framework has been applied to digital real estate markets. Shen and Ross (2021) showed that textual and visual cues in online listings function as complementary information channels: buyers exploit multiple signals simultaneously to update their prior beliefs about property quality. Hebdzyński (2023) demonstrated that text-based features improve hedonic price predictions in the Polish residential market, consistent with the view that listing language carries information beyond what is captured by structural attributes alone. The theoretical upshot is that any textual attribute of a real estate listing - including its emotional tone - is potentially informative about property quality under conditions of information asymmetry, provided that buyers can decode the signal and that their decoding affects their willingness to pay.

2.2. Textual Sentiment as a Pricing Signal in Property Markets

The concept that the emotional tone of property descriptions affects market outcomes is not new. Pryce and Oates (2008) established that the rhetoric of agents is neither arbitrary nor neutral, but is instead purposive and adapted to market conditions. Beck et al. (2023) showed that property descriptions become more positive in tone as the volume of competing listings increases, suggesting that sentiment is a strategic response to competitive market pressure rather than an objective property attribute - directly consistent with the signaling view.

However, the methodological and conceptual foundation for applying sentiment analysis to real estate market outcomes was laid by research in financial markets. Tetlock (2007) demonstrated that media pessimism scores predict subsequent market returns, establishing a precedent for textually extracted sentiment as a carrier of price-relevant information. Antweiler and Frank (2004) reached similar conclusions for internet message boards, while Baker and Wurgler (2006) formalized investor sentiment as a measurable driver of asset prices. Building on this foundation, sentiment analysis has been applied in real estate research to social media, news media, platform reviews, and market sentiment indices. However, relatively few studies explicitly analyze the sentiment of online property listings or classified advertisements, indicating that this remains a largely underexplored area of real estate research (Węgrzyn, 2026).

The direct application to residential real estate was developed by Lawani et al. (2019), who explicitly embedded sentiment scores into a classical hedonic price model, recovering a statistically significant positive implicit price for positive sentiment. Ploessl and Just (2024), using news-based sentiment indicators for the German residential market, showed that aggregate market sentiment significantly predicts price growth. Hebdzyński (2023) demonstrated similar effects for the Polish market. What unites these contributions is their implicit assumption that the pricing effect of sentiment is an aggregate, market-level phenomenon captured by a single coefficient. The present paper relaxes this assumption and tests whether the sentiment premium is homogeneous or heterogeneous across property contexts.

Differences in listing narratives are expected to be particularly pronounced when comparing professional real estate agents with private sellers. Professional agents typically operate within institutionalized

environments, adhere to established industry conventions, and rely on standardized marketing practices. Private sellers, in contrast, face fewer formal constraints and tend to emphasize different property attributes. As a result, listing descriptions can be understood as strategic forms of communication that reflect distinct objectives and incentives rather than as arbitrary textual accounts.

For instance, Nowak et al. (2021) show that agents' property descriptions can be used to account for spatial and temporal heterogeneity in housing markets. Similarly, Robertson and Doig (2010) argue that real estate agents may shape market dynamics through specific linguistic cues, including urgency-inducing expressions such as "be in quick." In line with this perspective, prior research emphasizes the active role of agents in shaping home-buying behavior, noting that "it is clear that agents are not simply information intermediaries but are active market makers" (Levy et al., 2008). These findings suggest that sentiment in listings may partly reflect deliberate linguistic strategies rather than purely descriptive content, and that its expression can differ systematically between professional and private sellers.

While agents' listing language appears to reflect strategic and market-oriented considerations, sentiment expression among private sellers may instead be driven by behavioral biases, particularly the endowment effect (Thaler, 1980; Tomal, 2024). Emotional attachment to a property may lead to more positively charged language in listings. However, this mechanism primarily explains the generation of sentiment rather than its direct impact on pricing. The endowment effect can therefore be interpreted as a common-cause mechanism, whereby both higher sentiment and higher asking prices are jointly driven by seller attachment.

The empirical literature reviewed above establishes three well-supported claims: (1) listing language is non-neutral and strategically deployed; (2) sentiment scores carry price-relevant information; and (3) hedonic models can incorporate textual attributes and recover their implicit prices. What the literature does not establish is whether the implicit price of sentiment is homogeneous across the property market or varies systematically with the informational context of the transaction.

This gap is consequential for both theory and empirical practice. Theoretically, if the pricing effect of sentiment is context-dependent - operating primarily where information asymmetry is highest - then the

informational channel through which sentiment affects price is supported, and the signaling interpretation is strengthened. Empirically, hedonic models that include sentiment as a single, unconditional predictor will recover a coefficient that is an average across heterogeneous contexts. If the true effect is concentrated in tenement buildings and near zero for standard blocks, the average coefficient is biased toward zero and misrepresents the effect as universal. This paper fills this gap by estimating both a baseline and an interaction-augmented hedonic model, and by interpreting the pattern of significant interactions through the lens of information asymmetry and signaling theory.

3. Material and methods

3.1. Research Design and Identification Strategy

The study employs a cross-sectional observational design based on a census of residential property listings active on a single date in the Krakow housing market. The identification strategy relies on the hedonic pricing method (Rosen, 1974): regression of log listing price on structural, locational, and textual attributes. Three model types are estimated: (1) a baseline OLS hedonic model without sentiment as an explanatory variable; (2) an extended model with a sentiment score (*sent_norm*) as an additional explanatory variable; (3) an interaction hedonic model augmenting the baseline with sentiment-by-property-characteristic interaction terms.

The interaction terms are specified in accordance with the building type dummies and seller type. The purpose is to test by which of these characteristics the pricing effect of sentiment is moderated and whether the pattern of moderation is consistent with the information asymmetry interpretation.

All models use the logarithm of asking price as the dependent variable, enabling coefficient interpretation as percentage effects.

3.2. Data Collection

Data were collected from *otodom.pl*, the principal residential property listing portal in Poland, on April 18th, 2024. The data cover all apartments listed for sale in the Krakow metropolitan area. The original dataset comprised 4,199 listings. Each observation records listing price, floor area, building type, floor level, finishing standard, amenity indicators (balcony, car

park, lift, gated area, additional space, separate kitchen, virtual tour), number of photographs, geographic location, seller category, and the full text of the listing description.

The geographic focus on Krakow is motivated by its position as Poland's second-largest city and one of the most active residential markets in the country. Two seller categories are identifiable under Polish real estate licensing law: private vendors (13% of listings), agents (87%).

3.3. Data Pre-processing

Listings missing price, floor area, or location were excluded. Duplicate listings - the same property advertised by multiple vendors - were resolved by retaining one randomly selected observation per property, following the approach proposed by Trojanek (2021). Properties with total floor area exceeding 200 m² were excluded as outliers. The final analytical sample comprises 3,100 valid cases.

Listing descriptions were pre-processed using the OpenAI GPT API (model: *gpt-3.5-turbo-0125*) to resolve orthographic errors, non-standard word segmentation, and glued or spaced-out phrases. This step is necessary given the unvalidated, user-generated nature of online listing text in Polish.

Then texts were tokenized using the *spaCy* NLP library with the Polish language pipeline (*pl_spacy_model_morfeusz*, large model). Stop words were removed using *spaCy*'s built-in lexicon, supplemented with a domain-specific list constructed via TF-IDF weighting to identify high-frequency. Named Entity Recognition (NER) was applied to mask personal identifiers, allowing for the exclusion of sensitive data (Tuora & Kobylinski, 2019). Frequently occurring but semantically uninformative words, such as conjunctions and prepositions, were removed following Ploessl and Just (2021).

Polish is a highly inflected language requiring lemmatization for accurate sentiment scoring. This process was implemented to convert word forms into their base forms (Lula & Wójcik, 2011; Rizun et al., 2018) and was carried out using *spaCy*'s part-of-speech tagging combined with the *Morfeusz 2* lemmatization module, which selects contextually appropriate lemmas (Tuora & Kobylinski, 2019).

A brief overview of all variables is presented in Table 1.

Table 1

Table of variables

	variable	N	Definition	Min	Mean	Max	SD
1	ln_price	3100	Logarithm of price	12.2	13.69	15.36	0.44
2	ln_unit_price	3100	Logarithm of the unit price	8.39	9.72	10.71	0.25
3	liv_area	3100	Floor area (m2)	25	57.34	210	25.92
4	dist_cent	3100	Distance to the city centre (km)	0.04	4.1	11.18	2.37
5	ln_dist_cent	3100	Logarithm of distance to the city centre.	-3.19	1.18	2.41	0.77
6	no_pict	3100	Number of pictures	0	12.37	20	4.94
7	no_pict_pa	3100	Number of pictures per floor area	0	0.24	0.77	0.11
13	stand	3100	Standard (4 levels: ready to move in, to finish, to renovate, ask)				
15	type_build_4	3100	Type of building (4 levels: "block" short for block of flats, "precast" ~ precast large-panel block of flats, "apartment" ~ higher standard block of flats, "tenement" ~ mostly historic buildings)				
16	floor_lev	3100	Floor level (4 levels: first floor, last floor, floor in the middle, NA – information not available)				
17	who_sells_2	3100	Who sells (2 levels: private owner, real estate agent)				
18	balcony	3100	Balcony (2 levels: 1- yes, 0-information not available)				
20	car_park	3100	Car park or a parking lot (2 levels: 1- yes, 0-information not available)				
21	lift	3100	Lift in the building (2 levels: 1- yes, 0-information not available)				
22	gated_area	3100	Building on the gated area (2 levels: 1- yes, 0-information not available)				
23	add_space	3100	Additional space. e.g., basement (2 levels: 1- yes, 0-information not available)				
24	sep_kitch	3100	Separate kitchen (2 levels: 1- yes, 0-information not available)				
25	virt_tour	3100	Virtual tour (2 levels: 1- yes, 0-information not available)				

Source: own study.

3.4. Sentiment Analysis

Sentiment scores were computed using the sentimentPL package, which provides PyTorch-based models for Polish-language sentiment regression. The models are built on the Allegro HerBERT transformer architecture - a BERT-variant pre-trained on a large Polish-language corpus - and fine-tuned on the CLARIN-PL PolEmo 2.0 sentiment dataset. Each listing description was scored to yield a continuous sentiment index in the range $[-1, 1]$, where -1 represents the most negative and $+1$ the most positive emotional connotation. For modeling purposes, sentiment was normalized to $[0, 1]$ (variable: sent_norm): $N = 3,099$, mean = 0.70, SD = 0.18.

The HerBERT-based approach was preferred over dictionary-based methods for two reasons: (1) Polish is a morphologically complex language for which standard sentiment lexica have limited coverage and accuracy; and (2) property-domain vocabulary is substantially divergent from the general-language corpora on which dictionary lexica are calibrated.

3.5. Hedonic Models

Three hedonic model specifications were estimated to examine the relationship between listing characteristics, textual sentiment, and property prices. Model 1 represents the baseline hedonic specification and includes only structural and listing-related characteristics to identify the traditional determinants of listing prices. Model 2 extends the baseline model by

incorporating the overall sentiment score (sent_norm) as an additional explanatory variable, allowing for the estimation of the unconditional effect of sentiment on property prices. Model 3 further extends Model 2 by introducing interaction terms between sentiment and selected property characteristics. This specification is designed to test whether the relationship between sentiment and prices is heterogeneous across different types of properties and information environments.

4. Results

4.1. Sentiment analysis

To address the first research question, which examines whether property listings prepared by real estate agents differ in their emotional tone from listings published by private sellers, the study first compares sentiment levels across seller groups. This analysis allows us to determine whether professionally mediated listings are characterized by a systematically different use of emotional language compared with privately posted offers.

To examine whether property listings prepared by real estate agents differ in their emotional tone from listings published by private sellers, a comparison of sentiment scores between the two seller groups was conducted. Due to the potential inequality of variances between groups, the Welch two-sample t-test was applied. This test evaluates whether the mean sentiment score differs significantly between privately listed properties and agent-mediated listings without assuming equal variance across groups.

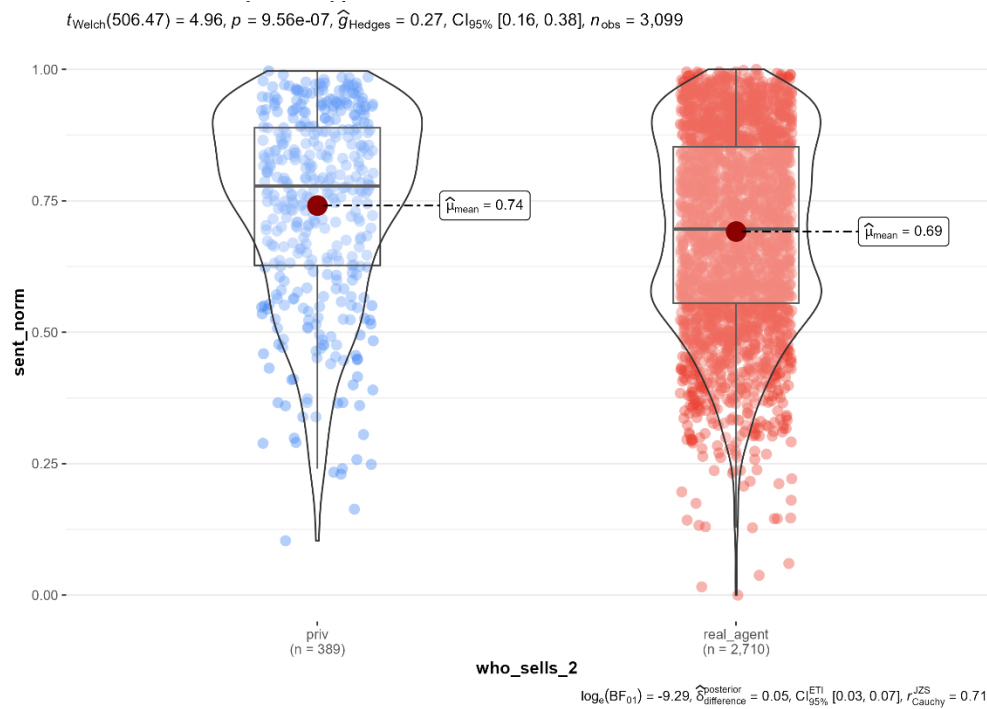


Fig. 1. Sentiment score by seller type. Source: own study.

The results indicate a statistically significant difference in sentiment levels between the two groups ($t = 4.96, p < 0.001$). Private sellers achieved a higher average sentiment score ($M = 0.741$) than real estate agents ($M = 0.691$). The estimated mean difference was 0.049 points, with a 95% confidence interval ranging from 0.030 to 0.069, confirming that the difference is consistently positive in the sample. These findings suggest that privately created listings are characterized by a more positive emotional tone compared with listings prepared by professional intermediaries. Therefore, the results provide evidence that seller type is associated with systematic differences in the emotional framing of online property descriptions.

4.2. Hedonic models

The hedonic modeling procedure was conducted in three stages. First, a baseline hedonic model was estimated using ordinary least squares (OLS), with the logarithm of listing price as the dependent variable. The baseline specification included conventional structural and listing-related characteristics of properties: logarithm of distance to the city centre, logarithm of living area, seller type, virtual tour availability, number of photographs per square meter, finishing standard, building type, balcony, parking availability, lift, floor level, and separate kitchen indicator.

Second, an extended hedonic model was estimated by incorporating the sentiment score (sent_norm) as an

additional explanatory variable. This specification was used to assess whether the textual characteristics of property descriptions provide additional explanatory power beyond traditional hedonic attributes.

To obtain a more parsimonious specification, variable selection was performed using a stepwise procedure based on the Bayesian Information Criterion (BIC).

To examine whether the relationship between listing sentiment and property prices is heterogeneous across different property characteristics, an interaction model was specified. The model included interaction terms between the sentiment score and selected structural, locational, and listing-related attributes of properties. Given the relatively large number of potential interaction effects and the risk of overparameterization, a regularization-based variable selection procedure was applied using the Least Absolute Shrinkage and Selection Operator (LASSO).

The LASSO procedure performs variable selection by adding an L1 penalty term to the ordinary least squares objective function. This penalty constrains the absolute values of regression coefficients and shrinks less informative coefficients towards zero. Variables with coefficients equal to zero are excluded from the model, whereas variables with non-zero coefficients are retained as potentially relevant predictors. The amount of shrinkage is controlled by the tuning parameter λ , which determines the balance between model

complexity and predictive performance. In this study, λ was selected using 10-fold cross-validation, and the one-standard-error rule was applied to obtain a more parsimonious model specification.

The LASSO selection procedure was applied to a model containing the main effects of property characteristics as well as interaction terms between sentiment and selected attributes. This approach was chosen because the presence of multiple interaction terms may lead to multicollinearity and overfitting, making traditional stepwise selection procedures less suitable. Unlike stepwise approaches based solely on statistical criteria such as AIC or BIC, LASSO performs simultaneous estimation and variable selection by penalizing model complexity, which is particularly

useful when a relatively large set of correlated predictors and interaction effects is considered, see (Nowak & Smith, 2017).

Following the LASSO selection step, the variables retained by the regularization procedure were included in a post-selection ordinary least squares (OLS) regression model. The post-LASSO OLS estimation was used to obtain unbiased coefficient estimates and conventional statistical inference, including the direction, magnitude, and significance of the identified effects. This modeling strategy allows for the identification of relevant heterogeneous effects of sentiment while reducing the risk of overfitting and improving the interpretability of the final hedonic price model

Table 2

Models specification			
term	OLS base	BIC model	Model int
(Intercept)	10.299*** (0.055)	10.198*** (0.050)	10.273*** (0.049)
log(dist_cent)	-0.182*** (0.006)	-0.182*** (0.006)	-0.153*** (0.022)
log(liv_area)	0.887*** (0.011)	0.889*** (0.011)	0.882*** (0.015)
who_sells_2real_agent	-0.031** (0.010)		-0.032** (0.010)
virt_tour0	-0.017 (0.012)		
no_pict_pa	0.174*** (0.036)	0.192*** (0.035)	0.274** (0.104)
standask	-0.029** (0.009)	-0.031*** (0.009)	-0.026** (0.009)
standfin	-0.072*** (0.012)	-0.074*** (0.012)	-0.073*** (0.012)
standreno	-0.146*** (0.013)	-0.148*** (0.013)	-0.145*** (0.013)
type_build_4apartment	0.109*** (0.009)	0.107*** (0.009)	0.130*** (0.038)
type_build_4precast	-0.033** (0.013)	-0.037** (0.013)	-0.065 (0.056)
type_build_4tenement	0.013 (0.013)	0.015 (0.013)	-0.165*** (0.048)
balcony	0.054*** (0.008)	0.054*** (0.008)	0.053*** (0.008)
car_park0	-0.005 (0.007)		
lift	0.065*** (0.007)	0.065*** (0.007)	0.067*** (0.007)
floor_levfirst	-0.031*** (0.009)	-0.030*** (0.009)	-0.029*** (0.009)
floor_levlast	-0.043*** (0.009)	-0.041*** (0.009)	-0.040*** (0.008)
sep_kitch0	0.046*** (0.008)	0.043*** (0.008)	
sent_norm		0.067*** (0.018)	
log(dist_cent) × sent_norm			-0.047

			(0.031)
log(liv_area) × sent_norm			0.012
			(0.015)
no_pict_pa × sent_norm			-0.135
			(0.143)
type_build_4apartment × sent_norm			-0.029
			(0.051)
type_build_4precast × sent_norm			0.045
			(0.075)
type_build_4tenement × sent_norm			0.281***
			(0.068)
sep_kitch0 × sent_norm			0.067***
			(0.010)
Num.Obs.	3034	3033	3033
R2	0.833	0.833	0.837
R2 Adj.	0.832	0.832	0.836
AIC	81344.5	81312.8	81249.6
BIC	81458.8	81415.1	81388.0
Log.Lik.	897.509	897.811	935.423
F	884.544	1003.628	737.019
RMSE	0.18	0.18	0.18

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: own study.

The results presented in Table X demonstrate that incorporating textual characteristics of property listings extends the explanatory power of traditional hedonic price models. The baseline OLS specification confirms the importance of conventional property attributes: living area and location remain the strongest determinants of price, while renovation standard, building type, balcony availability, elevator access, and floor level significantly differentiate property values.

The second model, selected using the BIC criterion, introduces the overall sentiment score and shows a positive and statistically significant association between listing sentiment and property prices ($\beta = 0.067$, $p < 0.001$). This result indicates that textual information contained in property descriptions provides additional explanatory power beyond structural and locational characteristics. However, this average effect assumes that sentiment operates similarly across all property types and listing contexts, which may conceal important differences in how buyers interpret descriptive language.

To examine whether the relationship between sentiment and price is heterogeneous, the third specification incorporates interactions between sentiment and selected property and listing characteristics. The interaction model provides evidence that sentiment does not have a uniform relationship with property prices. Compared with the BIC specification, the interaction model achieves improved model fit (R^2 increases from 0.833 to 0.837 and AIC decreases from 81312.8 to 81249.6), suggesting that allowing sentiment effects to vary

across property contexts provides additional explanatory value.

Following the model selection procedure, only a limited set of interaction terms was retained in the final specification. These include interactions between normalized sentiment and the logarithm of distance to the city centre, the logarithm of living area, the number of pictures per advertisement, building types, and the absence of a separate kitchen. All remaining interaction terms were excluded because they did not improve model fit according to the adopted selection criterion.

Notably, the model selection procedure did not retain any interaction between sentiment and seller type. This suggests that the association between listing sentiment and transaction prices does not differ systematically across seller categories once property characteristics and other listing attributes are controlled for.

The strongest evidence of heterogeneity concerns building type. The interaction between sentiment and tenement buildings is positive and statistically significant ($\beta = 0.281$, $p < 0.001$), indicating that sentiment is more strongly associated with prices for this category of properties. This suggests that descriptive language may play a particularly important role for properties whose value is not fully captured by standard hedonic characteristics and where narrative framing may contribute to their market positioning.

A significant interaction is also observed for kitchen configuration ($\beta = 0.067$, $p < 0.001$), indicating that the association between sentiment and price differs depending on this property characteristic. In the model

specification, the reference category corresponds to properties with a separate kitchen ($\text{sep_kitch} = 1$), while the interaction term captures the difference associated with properties without a separate kitchen ($\text{sep_kitch} = 0$). Therefore, the coefficient does not represent the standalone effect of lacking a separate kitchen, but rather indicates that the relationship between sentiment and price differs between these two types of properties.

However, this result should be interpreted cautiously because the model does not control directly for building age due to a substantial number of missing observations for this variable. Since kitchen configuration is likely correlated with building age, the observed interaction may partly reflect broader differences between older and newer properties rather than the isolated effect of kitchen layout. Older buildings typically include separate kitchens, whereas newer developments more frequently feature open-plan layouts. Therefore, the identified heterogeneity may capture differences in how these segments of the housing market are positioned, perceived, and described in property listings, rather than a pure effect of kitchen configuration itself.

Overall, the findings go beyond the basic conclusion that sentiment affects prices by showing under which conditions this relationship is observed. Sentiment is not valued equally across all listings; rather, its economic relevance depends on the characteristics of the property and the informational context of the advertisement. Positive language appears to have greater importance when textual descriptions contribute additional meaning, help differentiate properties, or shape perceptions that cannot be fully explained by objective attributes alone. These results support the view that sentiment functions not merely as an emotional feature of listings, but as a contextual signal whose relationship with price depends on the type of property being presented and the information available to potential buyers.

5. Discussion

The empirical findings reported in Section 4 speak most directly to the theoretical framework outlined in the introduction, namely that the pricing relevance of textual sentiment in property listings is not a fixed or universal feature of the listing itself, but instead depends on the informational context in which it is embedded. Read through the lens of signaling theory and information asymmetry (Spence, 2002; Broxterman & Zhou, 2023) the results suggest that sentiment

operates as a conditional rather than an unconditional signal: its capacity to influence buyers' willingness to pay appears to be activated primarily in settings where objective, verifiable information about property quality is comparatively scarce or difficult to interpret. This interpretation is supported, first, by the observation that a single average coefficient for sentiment - as estimated in the BIC-selected model - conceals considerable heterogeneity once interaction terms are introduced. The improvement in model fit associated with the interaction specification indicates that the relationship between sentiment and price is not homogeneous across the sample, but instead varies systematically with the characteristics of the property and the listing. This finding extends prior work that has treated the sentiment-price relationship as an aggregate, market-level phenomenon (Lawani et al., 2019; Ploessl & Just, 2024; Hebdzyński, 2023) by suggesting that the average effect previously reported in the literature may itself reflect the pooling of contexts in which sentiment plays markedly different roles.

Second, the pronounced positive interaction between sentiment and tenement buildings is broadly consistent with the proposed information asymmetry mechanism (Zhou et al., 2015). Tenement properties, typically older and structurally heterogeneous, are likely to display greater variation in technical condition, renovation history, and other quality dimensions that are not easily captured by standard structural variables such as floor area or number of rooms. Under such conditions, buyers may have comparatively limited access to objective, verifiable cues, so textual descriptions may come to play a larger role in helping them form judgments about underlying property quality. This pattern is consistent with, although it does not by itself confirm, the view that sentiment functions as a substitute information channel where structural information is less informative.

Third, the significant interaction involving kitchen configuration warrants a more cautious reading (Kolbe et al., 2021). Because the model does not directly control for building age - a substantial share of observations lack this information - the kitchen-configuration variable may partly proxy for differences between older and newer housing segments rather than capturing an independent architectural effect. Separate kitchens are considerably more common in older buildings, whereas open-plan layouts are more typical of recent developments; consequently, the interaction term may reflect broader differences in how

older and newer segments of the housing stock are described and marketed, rather than an isolated consequence of kitchen layout itself. We therefore interpret this result as an indicator of information availability associated with a wider bundle of unobserved property characteristics, rather than as evidence of a discrete causal role for kitchen configuration.

Fourth, it is worth noting that the final interaction model does not include an interaction between sentiment and seller type. During the model selection procedure, this interaction was not retained because it did not improve model fit according to the adopted selection criterion. This combination of results points to a useful distinction between factors that shape the generation of sentiment in listings and factors that determine whether that sentiment is reflected in asking prices. Private sellers may use a more positively toned register than agents - plausibly reflecting emotional attachment to the property, consistent with an endowment-effect explanation (Tomal, 2024) - without this difference translating into a differential pricing response. One possible interpretation is that seller-specific differences in language style may serve purposes other than directly influencing price formation. Professional agents may deliberately adopt a more neutral or standardized communication style as part of market-oriented presentation practices, whereas private sellers may rely on more affective descriptions. Such differences in wording may influence other stages of the transaction process, such as attracting attention, encouraging enquiries, shaping perceived attractiveness, or affecting listing exposure duration, but these mechanisms fall outside the scope of the present study. Therefore, the absence of seller-type moderation should not be interpreted as evidence that sentiment differences between sellers are irrelevant, but rather that their economic consequences may operate through channels other than asking prices.

Taken together, these findings support the study's central theoretical contribution: textual sentiment in real estate listings functions as a conditional signal whose economic relevance is contingent on the level of information asymmetry associated with a given property and listing context, rather than as a uniform linguistic premium applying equally across the market. This reframing suggests that future hedonic models incorporating textual sentiment should, where possible, allow for such conditionality rather than imposing a single unconditional coefficient, since doing so may otherwise understate the effect where it is economically

most meaningful and overstate it where objective information already resolves most of the buyer's uncertainty.

6. Conclusions

This study examined not only whether textual sentiment expressed in online property listings is associated with asking prices, but also under what conditions such a relationship is observed. Using a cross-sectional sample of residential listings from the Krakow housing market and a hedonic pricing framework augmented with sentiment-by-characteristic interaction terms, the analysis indicates that sentiment does not exert a uniform effect on prices but is instead associated with prices in a manner that depends on the characteristics of the property and the informational environment of the listing.

Three conclusions follow from the results. First, the relationship between sentiment and price is heterogeneous rather than universal: a single average coefficient, of the kind reported in prior studies, appears to conceal meaningful variation once property-specific interactions are considered. Second, the economic relevance of sentiment appears concentrated in contexts - most notably tenement buildings - where textual information may help reduce buyer uncertainty and complement other, more limited sources of objective information about property quality. Third, it is property characteristics rather than seller identity that appear to determine when sentiment becomes relevant to price formation: although private sellers and real estate agents differ in the emotional tone of their listings, this difference in how sentiment is generated does not translate into a difference in how sentiment is valued by the market.

These findings should be interpreted in view of several limitations. The hedonic models estimated here do not directly control for building age because of substantial missing data for this variable, and the kitchen-configuration interaction discussed in Section 5 may therefore partly reflect broader, unobserved differences between older and newer segments of the housing stock rather than an independent effect of kitchen layout. Furthermore, the analysis relies on cross-sectional data drawn from a single date and on asking prices rather than realized transaction prices, which may diverge from final sale prices as a result of negotiation. These limitations suggest caution in generalizing the findings beyond the Krakow market and the specific listing period under study, and point to directions for future research, including the incorporation of building-age data, longitudinal listing

data, and matched transaction-price information.

Overall, this study contributes to the literature by reframing textual sentiment in property listings not as a simple, universal linguistic premium, but as a context-dependent market signal whose informational value is activated primarily under conditions of elevated information asymmetry. This perspective suggests that future research on textual attributes in hedonic price models may benefit from explicitly modeling the conditions under which such attributes are likely to matter, rather than assuming a constant effect across all listings.

Funding Sources

The publication presents the result of the Project financed from the subsidy granted to the Krakow University of Economics.

References

- Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500. <https://doi.org/10.2307/1879431>
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645–1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>
- Beck, J., Bono, H. R., & Toma, M. (2023). Working hard or hardly working? Competition's effect on real estate agent effort. *Journal of Housing Research*. Advance online publication. <https://doi.org/10.1080/10527001.2023.2265073>
- Bond, S., & Devine, A. (2016). Certification matters: Is green talk cheap talk? *The Journal of Real Estate Finance and Economics*, 52(2), 117–140. <https://doi.org/10.1007/s11146-015-9499-y>
- Broxterman, D., & Zhou, T. (2023). Information frictions in real estate markets: Recent evidence and issues. *The Journal of Real Estate Finance and Economics*, 66(2), 203–298. <https://doi.org/10.1007/s11146-022-09918-9> PMID:38625176; <https://doi.org/10.1007/s11146-022-09922-z>
- Haag, J., Rutherford, R., & Thomson, T. (2000). Real estate agent remarks: Help or hype? *Journal of Real Estate Research*, 20(1–2), 205–216. <https://doi.org/10.1080/10835547.2000.12091024>
- Hebdiński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures? *Journal of Housing and the Built Environment*, 38(4), 2399–2425. <https://doi.org/10.1007/s10901-023-10043-z>
- Kolbe, J., Schulz, R., Wersing, M., & Werwatz, A. (2021). Real estate listings and their usefulness for hedonic regressions. *Empirical Economics*, 61, 3239–3269. <https://doi.org/10.1007/s00181-020-01992-3> PMID:33462521
- Lawani, A., Reed, M. R., Mark, T., & Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, 75, 22–34. <https://doi.org/10.1016/j.regsciurbeco.2018.11.003>
- Levitt, S. D., & Syverson, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics*, 90(4), 599–611. <https://doi.org/10.1162/rest.90.4.599>
- Levy, D., Murphy, L., & Lee, C. K. C. (2008). Influences and emotions: Exploring family decision-making processes when buying a house. *Housing Studies*, 23(2), 271–289. <https://doi.org/10.1080/02673030801893164>
- Lula, P., & Wójcik, K. (2011). Sentiment analysis of consumer opinions written in Polish. *Economics and Management*, 16(1), 1286–1291.
- 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>
- Ploessl, F., & Just, T. (2024). News coverage vs sentiment: Evaluating German residential real estate markets. *International Journal of Housing Markets and Analysis*, 17(2), 395–417. <https://doi.org/10.1108/IJHMA-07-2022-0102>
- Ploessl, F., Just, T., & Wehrheim, L. (2021). Cyclicity of real estate-related trends: Topic modelling and sentiment analysis on German real estate news. *Journal of European Real Estate Research*, 14(3), 381–400. <https://doi.org/10.1108/JERER-12-2020-0059>
- Pryce, G., & Oates, S. (2008). Rhetoric in the language of real estate marketing. *Housing Studies*, 23(2), 319–348. <https://doi.org/10.1080/02673030701875105>
- Rizun, N., Taranenko, Y., & Waloszek, W. (2018). Improving the accuracy in sentiment classification in the light of modelling the latent semantic relations. *Information (Basel)*, 9(12), 307. Advance online publication. <https://doi.org/10.3390/info9120307>, <https://doi.org/10.20944/preprints201810.0338.v1>
- Robertson, K., & Doig, A. (2010). An empirical investigation of variations in real-estate marketing language over a market cycle. *Housing Theory and Society*, 27(2), 178–189. <https://doi.org/10.1080/14036090903159994>
- 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>
- Shen, L., & Ross, S. (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>
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374. <https://doi.org/10.2307/1882010>
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39–60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Tomal, M. (2024). The endowment effect and housing studies: The role of multiple reference points. *Critical Housing Analysis*, 11(2), 184–193. <https://doi.org/10.13060/23362839.2024.11.2.576>
- Trojanek, R. (2021). Housing price cycles in Poland – The case of 18 provincial capital cities in 2000–2020. *International Journal of Strategic Property Management*, 25(4), 332–345. <https://doi.org/10.3846/ijspm.2021.14920>
- Tuora, R., & Kobylinski, Ł. (2019). Integrating Polish language tools and resources in spaCy. In *Proceedings of the PP-RAI 2019 Conference* (pp. 210–214).
- Węgrzyn, J. (2026). Mining real estate data: A systematic review of text-based approaches. *Real Estate Management and Valuation*. Advance online publication. <https://doi.org/10.2478/remav-2026-0001>
- Zhou, X., Gibler, K., & Zahirovic-Herbert, V. (2015). Asymmetric buyer information influence on price in a homogeneous housing market. *Urban Studies (Edinburgh, Scotland)*, 52(5), 891–905. <https://doi.org/10.1177/0042098014529464>