REMV
REAL ESTATE MANAGEMENT AND VALUATION

Journal homepage:
www.remv-journal.com

PREDICTING HOUSING PRICE TRENDS IN POLAND: ONLINE SOCIAL ENGAGEMENT - GOOGLE TRENDS

Mirosław Bełej[1*]

[1] *Departament of Spatial Analysis and Real Estate Market, University of Warmia and Mazury in Olsztyn, ul. Oczapowskiego 2, 10-719 Olsztyn, Poland, e-mail: miroslaw.belej@uwm.edu.pl, ORCID: 0000-0001-8650-6990*

**Corresponding author*

ARTICLE INFO

Keywords:

calendar effects, behavioral finance, real estate market

JEL Classification:

R20, R30, R31

ABSTRACT

Various research methods can be used to collect housing market data and predict housing prices. The online search activity of Internet users is a novel and highly interesting measure of social behavior. In the present study, dwelling prices in Poland were analyzed based on aggregate data from seven Polish cities relative to the number of online searches for the keyword *dwelling* tracked by Google Trends, as well as several classical macroeconomic indicators. The analysis involved a vector autoregressive (VAR) model and the Granger causality test. The results of the study suggest that the volume of online searches returned by Google Trends is an effective predictor of housing price dynamics, and that unemployment and economic growth are important additional variables.

Citation:

Bełej, M. (2023). Predicting housing price trends in Poland: Online social engagement - Google Trends. *Real Estate Management and Valuation, 31*(4), 73-87. https://doi.org/10.2478/remav-2023-0032

1. Introduction

Society operates in the housing market (Muellbauer & Murphy, 2008; Smith et al., 2004), as everyone needs to meet the basic human need for housing, which provides people with a sense of security and stability. In addition to fulfilling this primary social function, housing is also a very important element of the social hierarchy because property ownership in a given segment of the housing market attests to the owner's social status. At present, the ownership of housing in a given location and neighborhood determines the owner's standing in the social structure, where financial status is a fundamental concern (tribalization). In this context, housing is not a typical market commodity (Musa & Yusoff, 2018) because the long-term use of housing leads to the development of an emotional bond between the property and its owner (safety, security, place attachment, identity), whereas other types of goods do not elicit such a response.

In an urban setting, location, type of residential estate, housing quality, and a neighborhood's general reputation are strongly correlated with housing prices. Housing prices are largely influenced by spatial, infrastructural, organizational, economic, and social factors (Ball, 2017; De Bruyne & Van Hove, 2006; Mallick & Mahalik, 2015). These dependencies have been studied in various contexts, including:

- government policy (Ball, 2017; Leung, 2004),
- economic growth (Bełej & Cellmer, 2014; Trojanek, 2010),
- interest rates (Arce & Lopez-Salido, 2006),
- labor market (Lewandowska-Gwarda, 2018),
- demography (De Bruyne & Van Hove, 2006),
- environment and climate (Zambrano-Monserrate & Ruano, 2019),
- infrastructure (Lu et al., 2021).

An understanding of the forces that drive changes in housing prices, or at least an attempt to come closer to such an understanding, facilitates predictions of the trajectory of the housing market in a given country. Predictions of housing market trends are a fundamental issue for political decision-makers, investors, as well as consumers who are faced with the challenge of purchasing housing. Housing market forecasts have been extensively researched (Banerjee & Ghosal, 2020; Bełej, 2021; Chen et al., 2014; Kishor & Marfatia, 2018; Cañizares Martínez et al., 2023) based on historical data and predictive analytics that provide information about housing price dynamics in

the future. In addition to classical factors that impact housing prices, the rapid development of an information society (Göhring, 2004; Hilty, 2002) and the rising popularity of online platforms that shape social behaviors should be also considered in such analyses. According to Bełej (2022), information and communications technologies (ICT) provide prospective property buyers with online information that can be accessed with the use of a laptop or a smartphone without leaving their homes. As a result, Internet users can compare the prices of similar properties, analyze the neighborhood, air pollution levels, sunlight exposure and safety in a given neighborhood. According to Yang et al.(2015), real-time online activity provides information about collective behaviors and the studied population's interests, fears, and intentions. In this context, Google Trends is a useful application for monitoring social sentiment. Google Trends is a free, open-access tool that provides information about trending keywords entered by users in the Google search engine. This valuable tool for analyzing the popularity of different products, topics, and events can be used to determine the relative search volume (RSV) and compare various keywords (Castelnuovo & Tran, 2017; Kulkarni et al., 2009; Limnios & You, 2021). Instead of providing absolute search volume numbers, Google Trends estimates the number of searches in a given place and time and calculates the RSV index.

The aim of this study was to perform a comprehensive analysis of the Polish housing market based on aggregate data from seven cities relative to online search activity tracked by Google Trends. In addition, several classical macroeconomic factors that drive housing price dynamics in the long term were also analyzed. The collected data were analyzed with the use of the vector autoregressive (VAR) model and the Granger causality test. The results were processed using software packages in the R environment.

2. Materials and Methods

2.1. Study Area and Data Description

The database of the National Bank of Poland (BaRN) covering the period of 2010-2022 was the main source of information about housing prices in Poland. The study period was not arbitrarily selected. The Polish real estate market boomed in 2007-2008, and housing prices increased rapidly. Real estate prices decreased significantly in late 2008 and began to stabilize in 2009, which is why 2010 was chosen as the starting point in the analysis. The changes on the Polish housing market have been discussed extensively in numerous research studies (Brzezicka et al., 2018; Dittmann, 2013; Głuszak, 2015; Heldak, 2017; Kokot, 2014; Rącka et al., 2017; Rącka & Khalil ur Rehman, 2018; Trojanek, 2021). December 2022 marks the end of the analyzed period due to the availability of market data at the time this article was written. Statistical data are usually published with a quarterly lag; therefore, these types of analyses are usually based on historical data.

The present study is based on a fixed price index which was calculated as the weighted average of housing prices quoted on local real estate markets in the Polish cities of Gdańsk, Gdynia, Łódź, Kraków, Poznań, Warsaw, and Wrocław. The price per square meter was used to compare dwelling prices on the secondary market. The volume of online searches returned by Google Trends was retrieved and displayed with the use of the *gtrendsR* package in R. An R script enables the user to minimize errors and select the search category. Code 1080 (Real Estate Listings) was selected for the needs of this study to determine the most popular keywords associated with real estate listings. Macroeconomic data published by GUS (the Central Statistical Office in Poland), includes the Gross Domestic Product (GDP), unemployment (U_BAEL), the number of issued building permits (H_poz), and the RSV index for the keyword "dwelling".

2.2. Methods

Forecasting is one of the most popular techniques in time series analyses (Davies & Brown, 1960; Hyndman, 2011; Makridakis et al., 2018; Petropoulos et al., 2022). Forecasts are developed to estimate the dynamics of the future prices of goods and services based on various factors, including macroeconomic indicators. Price forecasts can be used to make political, economic, business, and investment decisions (Makridakis et al., 2018; Marfatia et al., 2021; Sun et al., 2019). In a time series analysis of housing prices, forecasting techniques can be applied to identify factors that drive housing prices, and the resulting information can be used to increase the accuracy of the predictions.

In the current study, housing prices were forecast with the use of the VAR model and the Granter causality test. The VAR model is a statistical technique for predicting changes in multivariate time series, and is particularly useful for analyzing the relationships between multiple variables that influence each other

over time. In the VAR model, every variable is expressed by a linear combination of its lagged values and the lagged values of all other variables in the system (Lu, 2001; Wang & Ding, 2018). The VAR model accounts for the dynamic interactions between variables, and it allows the user to analyze the impact of changes in one variable on the remaining variables in the system. The VAR model is usually estimated by ordinary least squares (OLS) regression or maximum likelihood estimation (MLE). The VAR modeling process usually involves the following steps (Giannone et al., 2015; Lu, 2001):

- the number of lags is selected for all observation periods and used to forecast future values,
- the VAR model is estimated by determining the values of coefficients that best fit empirical data,
- a forecast is generated for the VAR model based on the estimated coefficients of lagged values of the modeled empirical variables.

A VAR model not only predicts the future values of the analyzed independent variable, but also returns the confidence intervals that give an indication of the degree of uncertainty associated with the predictions. Confidence intervals facilitate predictions of future changes by minimizing their risk. The mathematical formula of a VAR model that is widely used in the literature (Kusideł, 2000) is presented below:

$$x_t = A_0 D_t + A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_k x_{t-k} + e_t \qquad (1)$$

where:

x_t - vector of observations on the current values of all modeled variables, $x = [x_{1t} x_{2t} x_{3t} \ldots x_{nt}]^T$,

D_t - vector of deterministic components of equations,

A_0 - *matrix* of parameters with vector variables D_t that does not contain zero elements,

Ai - *matrix* of parameters with lagged variables of vector x_t that does not contain zero elements,

e_t - vectors of stationary random disturbances, $e_t = [e_{t1} e_{t2} e_{t3} \ldots e_{tn}]^T$.

One of the greatest strengths of VAR models is that they can be used to identify the cause-and-effect relationships between variables. This is because VAR models can be applied to estimate impulse response functions which provide information about the evolution of the modeled variables in reaction to a shock in one or more variables. Vector autoregressive models are powerful prediction tools, but they have certain limitations. Prediction accuracy is largely determined by the quality of empirical data, the model's adequacy, and the stability of the analyzed relationships over time. The accuracy of VAR models can be also undermined by sudden structural changes in data or data unavailability (Bacchiocchi & Fanelli, 2015; Thoma, 2008).

3. Results

3.1. Basic research

The observed variable were dwelling prices per square meter (PLN/m^2) between Q1 2010 and Q4 2022. The information about the prices of dwellings on the secondary market was obtained from the database of the NBP, and a weighted average was calculated based on the prices quoted on local markets in seven Polish cities: Gdańsk, Gdynia, Łódź, Kraków, Poznań, Warsaw, and Wrocław. Housing prices from seven cities were normalized (1 variable) to best reflect the variations in average dwelling prices in Poland. The rate of changes in housing prices in the analyzed period is presented in Figure 1.

As demonstrated in Figure 1, dwelling prices have been increasing steadily since 2017, from PLN 6600/m^2 (EUR 1413/m^2) to PLN 10,300/m2 (EUR 2206/m^2), marking an increase of 65% over a period of 6 years. The first decrease in nominal housing prices was observed in Q4 2022. It remains unknown whether this decline will mark the beginning of a cooling off period on the housing market, as predicted by a small number of researchers. However, the sudden change in market trends (which had remained stable since 2017) prompts debate over the problem. The fact that the upward trend (that began in 2013) in housing prices was replaced with a downward trend could be a minor market correction, and similar events were observed in 2015 and 2017. It should be noted, however, that inflation reached 17% in Q4 2022, whereas the NBP reference rate increased from 0.1% in May 2020 to 6.75% in December 2022. At the same time, millions of Ukrainians have migrated to Poland since the beginning of the Russian invasion of Ukraine in February 2022. The question that remains to be answered is whether such a large number of negative events will reverse the long-term upward trend in housing prices or even lead to a decrease in prices.

The study was divided into two main parts. The first part is based largely on the author's previous study entitled "Does Google Trends Show the Strength of Social Interest as a Predictor of Housing Price Dynamics?" (Bełej, 2022) to determine whether housing price predictions would change if the time series were expanded by 12 months. This is a

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

necessary measure because, in the cited study, December 2021 was the last month of the analyzed period; therefore, major macroeconomic changes in Poland that have occurred since that time and, above all, fundamental geopolitical changes (Russian invasion of Ukraine) had not been taken into consideration in the analysis. In the second part of the study, housing prices were predicted with the use of a VAR model involving more than one variable.

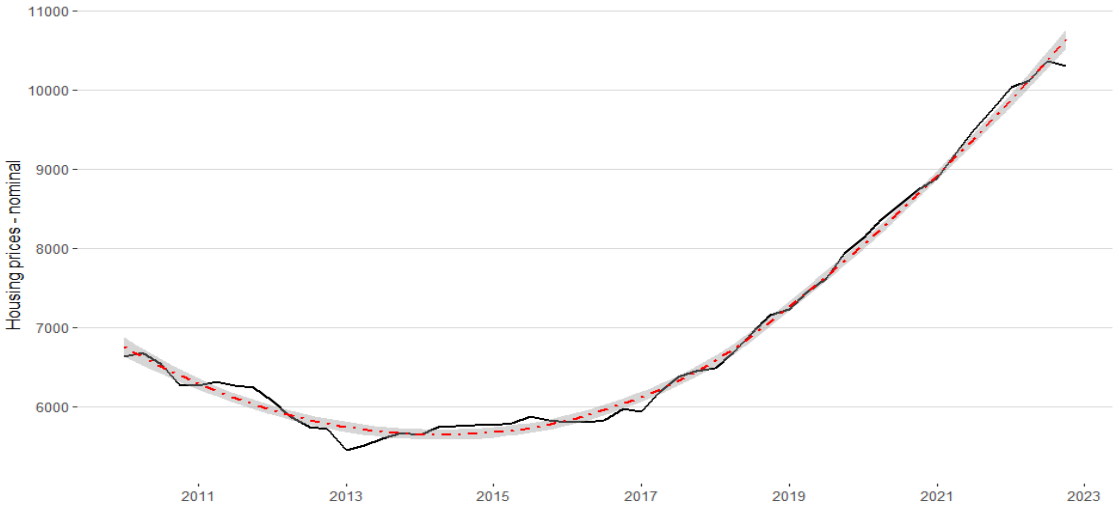


Fig. 1. Dwelling prices in Gdańsk, Gdynia, Łódź, Kraków, Poznań, Warsaw, and Wrocław (2010-2022) normalized based on the weighted average of the prices quoted on local real estate markets. *Source*: own elaboration based on NBP data (2023).

3.2. Vector autoregressive model - nominal housing prices and the RSV index

The relationship between average dwelling prices in seven Polish cities and the RSV index, calculated based on Google Trends data in the *gtrendsR* package in R, was analyzed in the first stage of the study. As previously mentioned, the RSV index was calculated based on the number of search queries for the Polish keyword *mieszkanie* (dwelling) in the Polish version of Google. The study was based on the assumption that the number of searches for the keyword *dwelling* is a measure of consumer interest in the housing market. Therefore, high values of the RSV index suggest that consumers have a high interest in searching for, buying, selling, or renovating dwellings. However, a sudden drop in the number of online searches for this keyword could be regarded as a predictor of a substantial decrease in the demand for housing. Therefore, in view of the law of inertia (Bełej, 2016; Bełej & Kulesza, 2012), dwelling prices can be expected to stabilize or decrease with a time lag. The study covered the period between Q1 2010 and Q4 2022. The RSV index of the keyword *dwelling* in Poland in Google Trends is presented in Figure 2.

Figure 2 indicates that the interest in the keyword "dwelling" has been increasingly steadily since 2010, which is generally consistent with the upward trend in dwelling prices observed during this period (Fig. 1). Seasonal variations in the RSV index of the analyzed keyword (Fig. 2) are a normal phenomenon on the housing market. In the following step, the RSV time series was smoothed with the LOESS function (red line), and smoothed values of the RSV were analyzed in subsequent parts of the analysis. The interest in the keyword "dwelling" decreased suddenly in 2020 (for the first time since 2010), which could be attributed to the fact that *COVID-19* was the most searched term on Google worldwide in March 2020. After a temporary decrease, an interest in the search phrase "dwelling" was restored to previous levels, but February 2020 (Russian invasion of Ukraine) brought a permanent decline in the RSV index of the analyzed keyword in Poland. These observations suggest that fluctuations in the RSV index of the keyword "dwelling" could play a role in predicting dwelling prices in Poland. If we assume that Internet activity is a measure of potential social behavior, then a rapid decrease in the number of searches for the keyword "dwelling" could point to a decline in demand for housing. As a result, such a significant decrease in public interest, which was associated with the geopolitical crisis and the ensuing changes in energy policy, could have contributed to the stabilization or decline in housing prices.

eISSN: 2300-5289 | Received 2023-07-26 | Revised 2023-08-16 | Accepted 2023-09-12

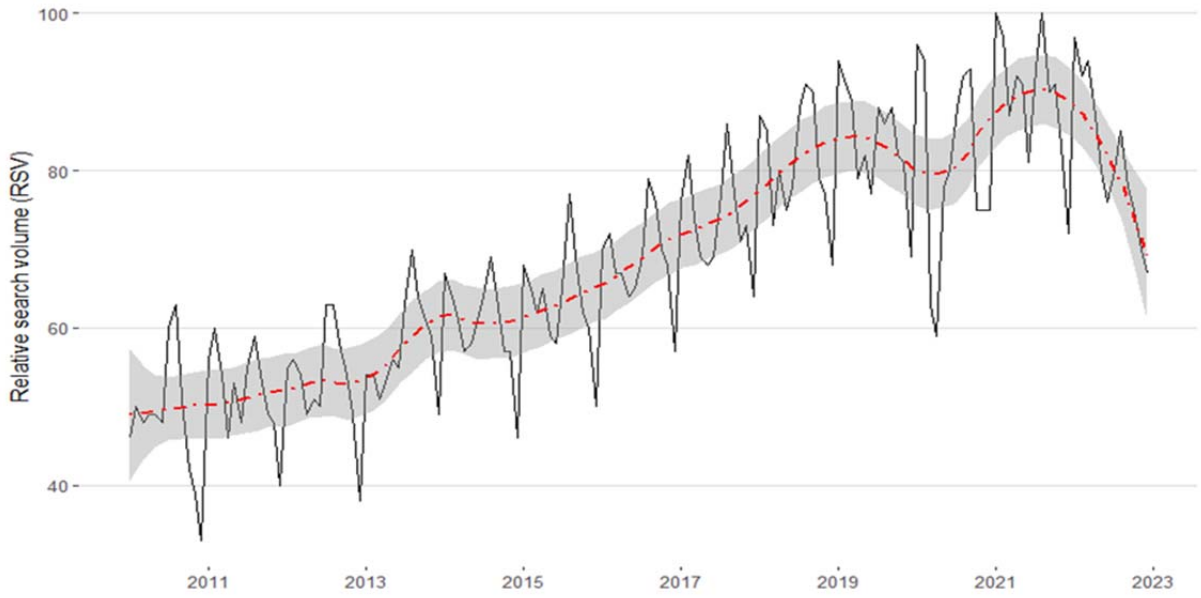


Fig. 2. Relative search volume (RSV) index of the keyword *dwelling* in Poland (2010-2022). *Source*: own elaboration.

In the first stage of the analysis examining the relationships between dwelling prices in seven cities (SHP_7) and the RSV index of the keyword "dwelling", the monthly RSV index was transformed into quarterly values. This approach was adopted because the NBP publishes information about housing prices on a quarterly basis. In the following steps, the logarithms of the studied variables were used according to the recommendations made by Canova (1999) and Drachal (2018). The lagged version of the examined time series had to be computed for the VAR model. This was accomplished with the use of the VARselection function in the "vars" package in R, which calculates the optimal lag based on various information criteria. The Akaike information criterion (AIC) is most widely used, and the Schwarz information criterion (SC) can also be applied to small samples. A lag that minimizes the value of AIC and SC is regarded as the optimal lag. The results of the analysis are presented in Table 1.

Table 1 indicates that all results were generated with a lag of AIC(1), HQ(1), SC(1), and FPE(1). Therefore, the lag level was set at 1 for the needs of the study. The roots of characteristic polynomials point to the stability and dynamics of the VAR model. In the analyzed case, the roots for variables SHP_7 and RSV were determined at 0.9793586 and 0.9214425, respectively. These statistics indicate that the VAR model is stable. The values of the VAR model were estimated in the next step, and the results are presented in Table 2.

The Granger causality test was conducted in the next stage of the VAR modeling process. In the Granger test, causality is the relationship between processes that generate data. Causality cannot be inferred directly from the VAR model, therefore, the cause-effect relationships between variables have to be determined. The Granger test was conducted in this stage of the VAR modeling process, and the results are presented in Table 3.

Table 1

Criteria for choosing the lag level in the VAR model

Criterion/Lags	1	2	3	4	5	6
AIC(n)	**-14.52905369**	-14.41459917	-14.31091331	-14.14699683	-14.05525192	-13.96828548
HQ(n)	**-14.34780453**	-14.17293362	-14.00883137	-13.7844985	-13.6323372	-13.48495438
SC(n)	**-14.03755599**	-13.7592689	-13.49175046	-13.16400141	-12.90842393	-12.65762493
FPE(n)	**4.91704 x10^{-7}**	5.5415 x10^{-7}	6.19981 x10^{-7}	7.40023 x10^{-7}	8.26499 x10^{-7}	9.252 x10^{-7}

Source: own elaboration.

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

Table2

VAR (1) model estimation

Criterion/Lags	Coefficient	Std. Err.	t-Value	*p*-Value	
RSV.l1	0.50956	0.14809	3.441	0.001324	**
SHP_7.l1	-0.04061	0.73973	-0.055	0.956480	
RSV.l2	0.46771	0.14122	3.312	0.001912	**
SHP_7.l2	-0.03622	0.73424	-0.049	0.960887	
const	0.78615	0.50217	1.566	0.124965	
sd1	0.19157	0.05350	3.580	0.000883	***
sd1	0.08362	0.02589	3.229	0.002412	**
sd3	0.21404	0.04394	4.871	1.61 x10-5	***
R^2	0.929				
Adjusted R^2	0.917				
F -statistic	78.1 on 7 and 42 DF, p-value: < 2.2 $x10^{-16}$				

Note: **p<0.01,*** p < 0.001.

Source: own elaboration.

Table 3

Results of the Granger causality test for the VAR model

	F-Test	*p*-Value	
SHP_7 → RSV	0.63487	p-value = 0.5325	
RSV → SHP_7	14.041	p-value = 5.485 $x10^{-6}$	***

Note: **p<0.01,*** p < 0.001.

Source: own elaboration.

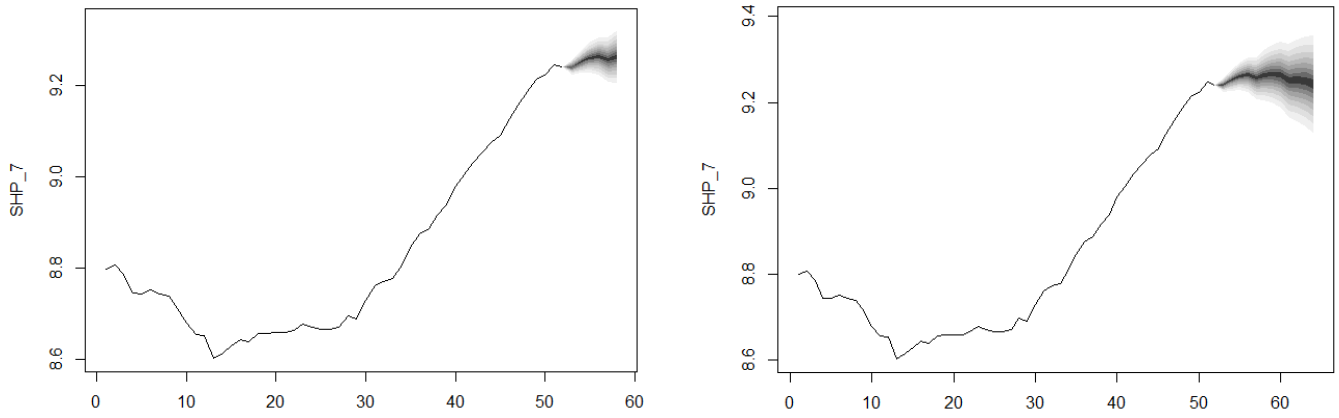


Fig. 3. Prediction of nominal quarterly dwelling prices (SHP_7) on the secondary market in Poland (2010-2022) based on the values of the relative search volume (RSV) index for a period of: a) 6 months, and b) 12 months. *Source*: own elaboration.

In Table 4, the *p*-value of Granger causality (RSV → SHP_7) is much lower than 0.05, which supports the hypothesis that the volume of online search queries is related to the actual dynamics of housing prices in Poland. Based on this observation, it can be assumed that the RSV index can be used to predict dwelling prices in Poland. The influence of only one indicator on housing prices was analyzed in the first part of the study. In the following stage, several macroeconomic indicators were considered, and the developed VAR model was applied to predict nominal dwelling prices (SHP_7) over a period of 6 months and 12 months after the end of the analyzed time series (Q4 2022). The results are presented in Figure 3.

Despite the presence of a single downward trend in the 6-month forecast, housing prices can be expected to stabilize. However, the prediction does not indicate a rapid decrease in prices as a result of the geopolitical crisis. The results of the 12-month prediction (covering the entire 2023) are more ambiguous. Based on the law of inertia (the classical principle governing the functioning of the housing

market), the sudden decrease in the value of the RSV index can be expected to exert a long-term effect on housing prices. Therefore, a decrease in the number of online searches for the keyword "dwelling" could lead to a minor drop in prices at the end of the year. This assumption holds true only in the absence of other macroeconomic shocks. At the time this article was written, various financial incentive schemes were being proposed to encourage property buyers to take mortgage loans. Such schemes could significantly influence the performance of the real estate market and, to a certain degree, minimize the negative impact of geopolitical factors, inflation or interest rates.

3.3. Vector autoregressive model - real value of housing and selected macroeconomic indicators

In this part of the study, the processes described in subsection 3.2 were repeated, but a larger number of macroeconomic variables was taken into consideration. Due to the high inflation in Poland in 2022, only the real value of dwellings was considered in the calculations. The nominal value is the current price of a commodity, which is usually expressed in domestic currency, whereas the real value is the nominal value adjusted for inflation. In countries where inflation has been low and stable for many years, this parameter is often disregarded in research. In Poland, however, inflation rose to 17% due to the geopolitical crisis and changes in domestic macroeconomic factors. Therefore, the real value of housing was used in the study to analyze price dynamics in greater detail. In this context, a decrease in the real prices of dwellings implies that the value of housing decreased relative to other goods and services. Despite the fact that the nominal prices of housing are currently stable, the availability of housing has decreased for many Polish consumers. Therefore, only the real prices of dwellings were taken into consideration in successive steps of the analysis. A similar approach has been used in other published studies.

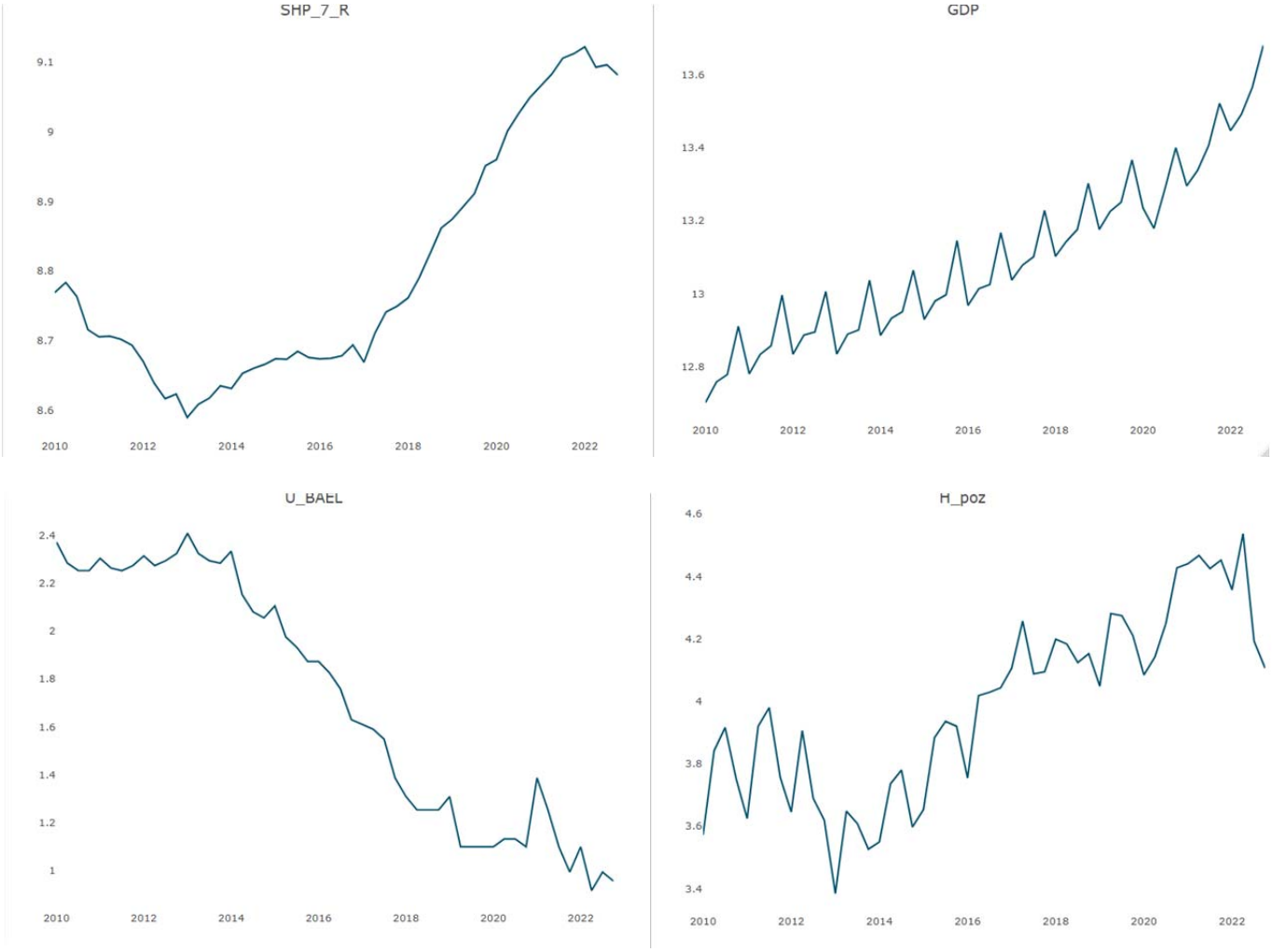


Fig.4. Macroeconomic indicators analyzed in the study (2010-2022). *Source*: own elaboration.

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

Table 4

Criteria for choosing the lag level in the VAR model

Criterion/Lags	1	2	3	4	5	6
AIC(n)	-3.127612	-3.184073	-3.165037	-3.216307	-3.321290	**-3.432955**
HQ(n)	-3.060599	-3.079831	-3.023565	-3.037606	-3.105360	**-3.179795**
SC(n)	**-2.948723**	-2.905801	-2.787383	-2.739270	-2.744870	-2.757153
FPE(n)	**2.679495 x10^{-14}**	3.91704 x10^{-14}	4.23415 x10^{-14}	6.99495 x10^{-14}	7.64243 x10^{-14}	7.97546 x10^{-14}

Source: own elaboration.

Table5

VAR model estimation

Criterion/Lags	Coefficient	Std. Err.	t-Value	*p*-Value	
H_poz.l1	-0.037667	0.019791	-1.903	0.06388	.
U_BAEL.l1	-0.045502	0.016833	-2.703	0.00987	**
GDP.l1	-0.089019	0.037535	-2.372	0.02237	*
RSV.l1	0.082036	0.032533	2.522	0.01556	*
SHP_7_R.l1	0.982559	0.032241	30.475	2.0 x10^{-16}	***
const	1.211032	0.420577	2.879	0.00624	**
sd1	0.013601	0.010375	1.311	0.19700	
sd1	0.002025	0.006412	0.316	0.75372	
sd3	0.015506	0.007634	2.031	0.04861	*
R^2	0.993				
Adjusted R^2	0.992				
F -statistic	78.1 on 7 and 42 DF, p-value: < 2.2 x10^{-16}				

Note: **p<0.01,*** p < 0.001.

Source: own elaboration.

The following variables were considered in the analysis: real quarterly price of housing per square meter (SHP_7_R), the Gross Domestic Product (GDP), unemployment (U_BAEL), the relative search volume (RSV) index, and the number of issued building permits (or the number of housing construction projects that do not require a permit) (H_poz). The values of the analyzed macroeconomic indicators are presented in Figure 4.

Structural breaks (OLS-CUSUM) can facilitate the analysis of time series for each variable in the VAR model, and they were examined in the following stage. The OLS-CUSUM statistical method is applied to identify structural breaks in a time series. In this procedure, the cumulative sum of OLS residuals over time is eliminated, and significant deviations from zero are identified. The presence of a significant deviation points to a structural break. Points above the red line in the OLS-CUSUM analysis are indicative of a significant structural break in data. As a result, the relationship between variables changes at this point, which can affect predictions of future trends. The results are presented in Figure 5.

The Granger causality test was conducted at this stage of the VAR modeling process. In the Granger test, causality is the relationship between processes that generate data. Causality cannot be inferred directly from the VAR model, therefore, the cause-effect relationships between variables have to be determined. The Granger test was then conducted at this stage of the VAR modeling process, and the results are presented in Table 6.

Granger causality (Table 6) was identified in the relationships between unemployment (U_BAEL), the GDP, and the RSV index in Google Trends vs. the remaining variables. Since the p-value of Granger causality is significantly below 0.05 for these variables, these three parameters influence other variables. This implies that changes in these three variables precede changes in the remaining variables, in particular the examined variable, i.e. the real quarterly price of housing per square meter (PLN/ m^2). Once again, these observations confirm that the RSV index of the keyword *dwelling* could be a significant predictor of changes in housing prices. The above implies that the influence exerted by social media, online videos, and other herd behaviors can be used in analyses of housing market trends. Six- and 12-month forecasts of real prices of dwellings quoted on the secondary market in Poland are presented in Figure 6.

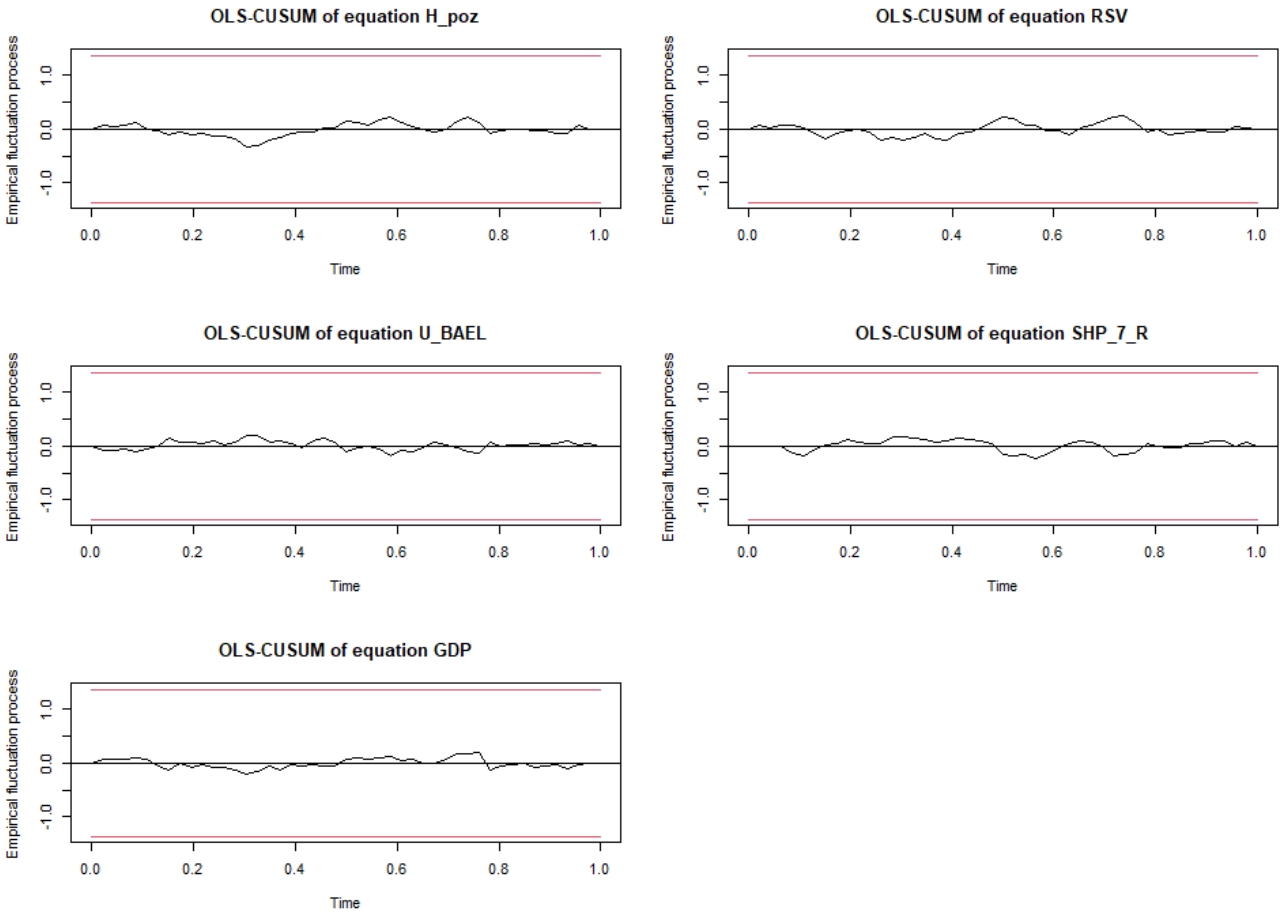


Figure5. The results of the test for structural breaks in the residuals of the VAR model. *Source*: own elaboration.

Table 6

Results of the Granger causality test for the VAR model

	F-Test	***p*-Value**	
H_poz → U_BAEL, GDP, RSV, SHP_7_R	1.1793	p-value = 0.3209	
U_BAEL → H_poz, GDP, RSV, SHP_7_R	2.8746	p-value = 0.02392	**
GDP → H_poz, U_BAEL, RSV, SHP_7_R	5.8759	p-value = 0.0001681	***
RSV → H_poz, U_BAEL, GDP, SHP_7_R	4.6958	p-value = 0.001193	***
SHP_7_R → H_poz, U_BAEL, GDP, RSV	1.619	p-value = 0.06759	

Note: **p<0.01,*** p < 0.001

Source: own elaboration.

The data in Figure 6 are more difficult to interpret than the results of the VAR estimation based on a single variable (Fig. 3). It should be noted that real values are better predictors of housing availability than nominal values. The analysis revealed a clear decrease in the real prices of housing in the last quarters. In contrast, the decrease in nominal prices seems to have been caused by a market correction. The use of two forecasting periods considerably minimizes this risk. According to the 6-month forecast, housing prices will decrease substantially in the first half of 2023. However, in the 12-month forecast, the predicted decrease is followed by an increase in real prices to the level noted in the last quarter of the analyzed period (Q4 2022). In this scenario, a sudden decrease in prices is not predicted, and real prices are expected to stabilize. It should be noted that the present study was based on historical data, and the demand-supporting shock associated with new mortgage incentive schemes was not taken into consideration in this analysis.

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

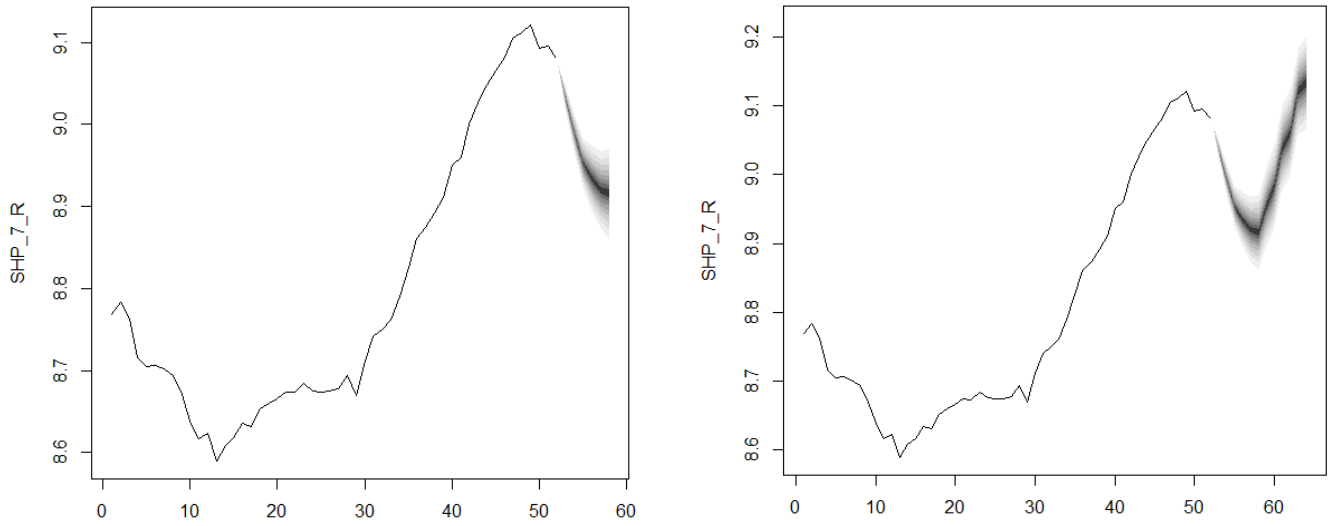


Figure6. a) Six-month forecast of housing prices; b) Twelve-month forecast of housing prices. *Source*: Own elaboration

4. Discussion

The significance of housing price forecasts was discussed at the beginning of this article, and the economic and social consequences of accurate predictions were emphasized. The availability of reliable price trend forecasts facilitates decision-making and plays a very important role on the housing market. Prospective buyers can plan property purchases over time to secure the lowest price, whereas property sellers can develop supply-side strategies to secure maximum returns. Therefore, accurate and reliable housing market forecasts are essential for all involved parties, including property buyers, owners, real estate brokers, investors, financial institutions, and decision-makers. The importance of housing market forecasts has been emphasized in numerous research studies (Billah, Hyndman, and Koehler 2005; Chatfield 2001; Chen et al. 2014; Kishor and Marfatia 2017; Cañizares Martínez et al. 2023; Wilson et al. 2002).

In the current study, dwelling prices were forecast with the use of the VAR model which analyzes the cause-effect relationships between economic variables, including housing prices and macroeconomic indicators. VAR models can be applied to estimate impulse response functions which provide information about the extent to which a shock to one variable affects the remaining variables in the model (Clements & Mizon, 1991; Min & Lee, 2019). For example, Henly and Wolman (2011) analyzed housing prices in the USA with the use of the VAR model and found that shocks to residential construction projects influenced the GDP during a period of economic recession. They demonstrated that employment in the housing industry began to decrease before the overall decline in employment. The cited study emphasized the significance of the relationship between the housing market, GDP, and employment during a recession. In another study, the VAR model was used to analyze the housing market in Saudi Arabia, and the results revealed that the house price index (HPI), inflation measured by the consumer price index (CPI), unemployment (UNEMP), and GDP per capita (GDPP) influenced housing prices. In addition, a Granger causality relationship was reported between prices and the above variables. These observations correspond with the results of the present study, where GDP and unemployment also affected housing prices. In this study, inflation was not a separate variable in the VAR model, but it was used indirectly in the second part of the study to analyze the real prices of housing. Guo and Unal (2011) relied on the structural vector autoregressive (SVAR) model to investigate housing prices in eight developed countries. In most of the examined countries, the wealth effects of housing and stock market wealth contributed to an increase in consumption. The cited authors concluded that a decrease in prices exerted a more negative effect on production than the stock market, which could lead to higher unemployment (Guo & Unal, 2011). The above findings also confirmed the applicability of the adopted research method and variables. However, social responses were not investigated.

This study makes a valuable contribution to the

eISSN: 2300-5289 | Received 2023-07-26 | Revised 2023-08-16 | Accepted 2023-09-12

existing knowledge because human behavior and the functioning of an information society were considered in the process of forecasting housing prices. The analysis revealed that housing price trends are strongly correlated with online search activity tracked by Google Trends (Castelnuovo & Tran, 2017; Limnios & You, 2021; Matias, 2013), which indicates that the study is timely and relevant. Google Trends statistics are rarely used by Polish researchers to predict housing market trends. Bulczak (2021) relied on Google Trends data, but the study analyzed the British housing market (2004-2014). Therefore, this article and the author's previous study (Bełej, 2022) are rare examples of such reports in Poland. Further research is needed to explore the applicability of Google Trends data for forecasting prices on the Polish real estate market. Due to the general scarcity of published articles, the present results cannot be compared with the findings of other Polish researchers, and they will be discussed based mainly in reference to the international literature.

A similar approach was adopted by Askitas (2016) who created a time series of weekly ratios of Google searches on buying and selling in the real estate category of Google Trends. The cited study revealed a strong correlation between the Case-Shiller home price index and the number of "buy" and "sell" searches online, which indicates that Google Trends data are useful for forecasting housing prices. According to Steegmans (2021), special attention should be paid to the correct usage of Google data because the smaller the sample size, the greater the sampling error. The cited author investigated the relationship between online search activity for mortgages and real housing market activity. Dietzel (2015) examined whether Google search volume data can be used as a leading indicator of consumer sentiment to predict turning points on the US housing market. The study demonstrated that the Google model was not always accurate in timing, but that it correctly predicted upcoming turning points in housing prices. In other studies, Google Trends was described as a reliable tool for forecasting housing prices (Castelnuovo & Tran, 2017; Dietzel, 2015; Huarng et al., 2020; Wu & Brynjolfsson, 2009). A somewhat different approach was adopted by Limnios and You (Limnios & You, 2021; Limnios & You, 2016) who combined linear pricing models for the housing market with Google Trends data to determine whether crowd-sourced search query data can improve the models' forecasting ability. They compared various performance measures of the models' dynamic forecasts and found that augmentation with Google Trend data did not significantly improve the models' forecasting performance. Despite the negative attitude regarding the use of Google Trends statistics in housing market research expressed by many scientists, this tool has numerous advantages, which is why it was applied in this study.

5. Conclusions

Housing price predictions play an important role in research. In the present study, quarterly (nominal and real) dwelling prices in Poland were forecast based on one social variable and several macroeconomic variables. The analysis was based on housing prices published in the BaRN database of the NBP, the RSV index of the keyword *dwelling* in Google Trends, and the macroeconomic indicators published by GUS. The study demonstrated that the online activity of an information society can affect the housing market and predictions of housing prices. The results indicate that the RSV index of keywords related to the housing market is a useful tool in real estate research.

Housing prices were forecast with the use of the VAR model and the Granger causality test. The study demonstrated that Google Trends data improved the quality of price forecasts on the Polish housing market. The results indicate that the RSV index could be a helpful tool for improving the accuracy of home value predictions, thus facilitating consumer and economic decision-making. The present findings can be used by investors to develop more effective investment strategies. However, housing forecasts should be based on a wide range of data, in particular macroeconomic indicators such as unemployment and the GDP.

The main limitation of this study is that it relies solely on data from the Polish housing market. As a result, any attempts to generalize the results to housing markets in other countries are burdened with a high risk. Macroeconomic indicators should be examined in a broader context to capture changes in the surroundings of the housing market with greater accuracy. Further research is needed to minimize these limitations. Diverse data from other countries should be examined to validate the observations made on the Polish housing market. The novelty and originality of this study stems from the fact that classical prognostic factors based on macroeconomic indicators were integrated with the online search activity of an information society as a non-classical measure.

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

Statement

A research paper on forecasting is most often published sometime after the realization of the forecast. The author is, therefore, clearly not influenced by events that have been realized in the macroeconomic environment, which may have influenced the behavior of the phenomenon under study and emerged after the research had been carried out.

Acknowledgements

The author expresses his sincere gratitude to the Journal Editor and the anonymous Reviewers who spent their valuable time providing constructive comments and assistance to improve the quality of this paper.

References

Arce, O., & Lopez-Salido, D. (2006). House prices, rents, and interest rates under collateral constraints. *Banco de Espana Research Paper No. WP-0610*.

Askitas, N. (2016). Trend-spotting in the housing market. *Cityscape (Washington, D.C.), 18*(2), 165–178.

Bacchiocchi, E., & Fanelli, L. (2015). Identification in structural vector autoregressive models with structural changes, with an application to US monetary policy. *Oxford Bulletin of Economics and Statistics, 77*(6), 761–779. https://doi.org/10.1111/obes.12092

Ball, M. (2017). *Housing policy and economic power: The political economy of owner occupation* (Vol. 828). Routledge., https://doi.org/10.4324/9780203472613

Banerjee, D., & Ghosal, A. (2020). Forecasting House Price with an Optimum Set of Features. In *Computational Intelligence in Pattern Recognition* (pp. 93–100). Springer Singapore., https://doi.org/10.1007/978-981-15-2449-3_7

Bełej, M. (2016). Ekonomia złożoności w badaniach rynków nieruchomości. *Studia Ekonomiczne / Polska Akademia Nauk. Instytut Nauk Ekonomicznych, 3*, 462–482.

Bełej, M. (2021). Housing price forecasting in selected polish cities during the COVID-19 pandemic. *Geomatics and Environmental Engineering, 15*(4), 59–80. https://doi.org/10.7494/geom.2021.15.4.59

Bełej, M. (2022). Does Google trends show the strength of social interest as a predictor of housing price dynamics? *Sustainability 2022, 14*(9), 5601. https://doi.org/10.3390/su14095601

Bełej, M., & Cellmer, R. (2014). The effect of macroeconomic factors on changes in real estate prices - response and interaction. *Acta Scientiarum Polonorum. Oeconomia, 13*(2), 5–16.

Bełej, M., & Kulesza, S. (2012). Modeling the Real Estate Prices in Olsztyn under Instability Conditions. Folia Oeconomica Stetinensia. https://doi.org/10.2478/v10031-012-0008-7

Billah, B., Hyndman, R. J., & Koehler, A. B. (2005). Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation, 75*(10), 831–840. https://doi.org/10.1080/00949650410001687208

Brzezicka, J., Wisniewski, R., & Figurska, M. (2018). Disequilibrium in the real estate market: Evidence from Poland. *Land Use Policy, 78*, 515–531. https://doi.org/10.1016/j.landusepol.2018.06.013

Bulczak, G. M. (2021). Use of Google Trends to Predict the Real Estate Market: Evidence from the United Kingdom. *International Real Estate Review, 24*(4), 613–631. https://doi.org/10.53383/100332

Cañizares Martínez, C., de Bondt, G. J., & Gieseck, A. (2023). Forecasting housing investment. *Journal of Forecasting, 42*, 543–565. Advance online publication. https://doi.org/10.1002/for.2946

Canova, F. (1999). Vector autoregressive models: Specification, estimation, inference, and forecasting. In *Handbook of Applied Econometrics Volume 1: Macroeconomics* (pp. 53–110). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1111/b.9780631215585.1999.00003.x

Castelnuovo, E., & Tran, T. D. (2017). Google it up! A google trends-based uncertainty index for the United States and Australia. *Economics Letters, 161*, 149–153. https://doi.org/10.1016/j.econlet.2017.09.032

Chatfield, C. (2001). *Time-Series Forecasting*. Chapman & Hall/CRC. https://books.google.com/books/about/Time_Series_Forecasting.html?hl=pl&id=PFHMBQAAQBAJ

Chen, N.-K., Cheng, H.-L., & Mao, C.-S. (2014). Identifying and forecasting house prices: A macroeconomic perspective. *Quantitative Finance, 14*(12), 2105–2120. https://doi.org/10.1080/14697688.2013.842650

Clements, M. P., & Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review, 35*(4), 887–917. https://doi.org/10.1016/0014-2921(91)90042-H

Davies, O. L., & Brown, R. G. (1960). Statistical Forecasting for Inventory Control. *Journal of the Royal Statistical Society. Series A (General), 123*(3), 348. https://doi.org/10.2307/2342487

De Bruyne, K., & Van Hove, J. (2006). Explaining the spatial variation in housing prices: An economic geography approach. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.947671

Dietzel, M. (2015). Sentiment-based predictions of housing market turning points with Google trends. *22nd Annual European Real Estate Society Conference*. https://doi.org/10.15396/eres2015_3

Dittmann, I. (2013). Primary and secondary residential real estate markets in Poland–analogies in offer and transaction price development. *Real Estate Management and Valuation, 21*(1), 39–48. https://doi.org/10.2478/remav-2013-0006

Drachal, K. (2018). Causality in the Polish Housing Market: Evidence from Biggest Cities. *Financial Assets and Investing, 9*(1), 5–20. https://doi.org/10.5817/FAI2018-1-1

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics, 97*(2), 436–451. https://doi.org/10.1162/REST_a_00483

Głuszak, M. (2015). Multinomial logit model of housing demand in Poland. *Real Estate Management and Valuation, 23*(1), 84–89. https://doi.org/10.1515/remav-2015-0008

Göhring, W. (2004). The Memorandum "Sustainable Information Society". In Minier, P. & Susini, A. (Hrsg.), *Sh@ring – EnviroInfo 2004*. http://enviroinfo.eu/sites/default/files/pdfs/vol110/0278.pdf

Guo, S., & Unal, U. (2011). VAR estimates of the housing and stock wealth effects: Cross-country evidence. *Economics Research Working Paper Series, 18*. https://digitalcommons.fiu.edu/economics_wps/18

Heldak, M. (2017). Changes in transaction prices of real estate and agricultural land in Poland in years 2004-2015. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, 17*(53), 77-84. https://doi.org/10.5593/sgem2017/53/S21.010

Henly, S. E., & Wolman, A. L. (2011). *Housing and the Great Recession: A VAR Accounting Exercise*. ERN: Business Fluctuations; Cycles. Topic.

Hilty, L. M. (2002). Sustainable development and the information society. In K. Brunnstein & J. Berleur (Eds.), Human choice and computers: Issues of choice and quality of life in the information society (pp. 305–315). https://doi.org/10.1007/978-0-387-35609-9_26

Huarng, K.-H., Hui-Kuang Yu, T., & Rodriguez-Garcia, M. (2020). Qualitative analysis of housing demand using Google trends data. *Ekonomska Istrazivanja, 33*(1), 2007–2017. https://doi.org/10.1080/1331677X.2018.1547205

Hyndman, R. J. (2011). Forecasting: An overview. In *International Encyclopedia of Statistical Science* (pp. 536–539). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_256

Kishor, N. K., & Marfatia, H. A. (2018). Forecasting house prices in OECD economies. *Journal of Forecasting, 37*(2), 170–190. https://doi.org/10.1002/for.2483

Kokot, S. (2014). The analysis of differences in residential property price indices. *Real Estate Management and Valuation, 22*(3), 14–27. https://doi.org/10.2478/remav-2014-0023

Kulkarni, R., Haynes, K. E., Stough, R. R., & Paelinck, J. H. P. (2009). Forecasting Housing Prices with Google Econometrics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1438286

Kusideł, E. (2000). Modele wektorowo-autoregresyjne–VAR–metodologia i zastosowania, [w:] B. Suchecki (red.). *Dane Panelowe i Modelowanie Wielowymiarowe w Badaniach Ekonomicznych, 3*.

Leung, C. (2004). Macroeconomics and housing: A review of the literature. *Journal of Housing Economics, 13*, 249–267. Advance online publication. https://doi.org/10.1016/j.jhe.2004.09.002

Lewandowska-Gwarda, K. (2018). Geographically weighted regression in the analysis of unemployment in Poland. *ISPRS International Journal of Geo-Information, 7*(1), 17. https://doi.org/10.3390/ijgi7010017

Limnios, A. C., & You, H. (2021). Can Google trends improve housing market forecasts? *Curiosity: Interdisciplinary Journal of Research and Innovation, 1*(2), 21987. https://doi.org/10.36898/001c.21987

Limnios, C., & You, H. (2016). Can Google trends actually improve housing market forecasts? *SSRN* Electronic Journal. https://doi.org/10.2139/ssrn.2886705

Lu, M. (2001). Vector autoregression (var) - An approach to dynamic analysis of geographic processes. *Geografiska Annaler. Series B, Human Geography, 83*(2), 67–78. https://doi.org/10.1111/j.0435-3684.2001.00095.x

Lu, X., Cai, D., & Zeng. (2021). Influencing factors of housing price differentiation based on the spatial quantile model: A case study of Wuhan City. *Progress in Geography, 40*(2), 283–292. https://doi.org/10.18306/dlkxjz.2021.02.009

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One, 13*(3), e0194889. https://doi.org/10.1371/journal.pone.0194889 PMID:29584784

Mallick, H., & Mahalik, M. K. (2015). Factors determining regional housing prices: Evidence from major cities in India. *Journal of Property Research, 32*(2), 123–146. https://doi.org/10.1080/09599916.2014.963642

Marfatia, H. A., Ji, Q., & Luo, J. (2022). Forecasting the volatility of agricultural commodity futures: The role of co-volatility and oil volatility. *Journal of Forecasting, 41*, 383–404. https://doi.org/10.1002/for.2811

Matias, Y. (2013). Nowcasting with Google Trends. *International Symposium on String Processing and Information Retrieval*, 4. https://doi.org/10.1007/978-3-319-02432-5_4

Milo, W., Łapińska-Sobczak, N., Malaczewski, M., Szafrański, G., Ulrichs, M., & Wośko, Z. (2010). Stabilność rynków finansowych a wzrost gospodarczy. Wydawnictwo Naukowe PWN.

Min, S.-O., & Lee, Y.-S. (2019). Korean Housing Market Dynamics: A VAR Analysis with Sign Restrictions. *Journal of Real Estate Analysis, 5*(1), 1–13. https://doi.org/10.30902/jrea.2019.5.1.1

Muellbauer, J., & Murphy, A. (2008). Housing markets and the economy: The assessment. *Oxford Review of Economic Policy, 24*(1), 1–33. https://doi.org/10.1093/oxrep/grn011

Musa, U., & Yusoff, W. (2018). Impact of Location and Dwelling Characteristics on Residential Property Prices/Values: A Critical Review of Literature. *International Journal of Advances in Management and Economics*, May. https://www.managementjournal.info/index.php/IJAME/article/view/409.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., . . . Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting, 38*(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Rącka, I., & Khalil ur Rehman, S. (2018). Housing Market in Capital Cities–the Case of Poland and Portugal. *Geomatics and Environmental Engineering, 12*(3), 75–87. https://doi.org/10.7494/geom.2018.12.3.75

Rącka, I., Palicki, S., Krajewska, M., Szopińska, K., & Kempa, O. (2017). Changes On the Housing Market of the Downtown Area in Selected Polish Cities. *Real Estate Management and Valuation, 25*(2), 79–90. https://doi.org/10.1515/remav-2017-0015

Smith, S. J., Easterlow, D., & Munro, M. (2004). Housing for health: Does the market work? *Environment and Planning A. Environment & Planning A, 36*(4), 579–600. https://doi.org/10.1068/a3672

Steegmans, J. (2021). The Pearls and Perils of Google Trends: A Housing Market Application. *Big Data, 9*(6), 443–453. https://doi.org/10.1089/big.2020.0183 PMID:34898271

Sun, S., Wei, Y., Tsui, K.-L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1–10. https://doi.org/10.1016/j.tourman.2018.07.010

Thoma, M. (2008). Structural change and lag length in VAR models. *Journal of Macroeconomics, 30*(3), 965–976. https://doi.org/10.1016/j.jmacro.2007.08.001

Trojanek, R. (2010). Dwelling's price fluctuations and the business cycle. *Economics & Sociology (Ternopil), 3*(2), 67–77. https://doi.org/10.14254/2071-789X.2010/3-2/7

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

Wang, L., & Ding, S. (2018). Vector autoregression and envelope model. *Stat (International Statistical Institute), 7*(1), e203. https://doi.org/10.1002/sta4.203

Wilson, I. D., Paris, S. D., Ware, J. A., & Jenkins, D. H. (2002). Residential property price time series forecasting with neural networks. *Knowledge-Based Systems, 15*(5–6), 335–341. https://doi.org/10.1016/S0950-7051(01)00169-1

Wu, L., & Brynjolfsson, E. (2009). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2022293

eISSN: 2300-5289 | © 2023 The Author(s) | Article under the CC BY license

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences of the United States of America, 112*(47), 14473–14478. https://doi.org/10.1073/pnas.1515373112 PMID:26553980

Zambrano-Monserrate, M. A., & Ruano, M. A. (2019). Does environmental noise affect housing rental prices in developing countries? Evidence from Ecuador. *Land Use Policy, 87*, 104059. https://doi.org/https://doi.org/10.1016/j.landusepol.2019.104059 https://doi.org/10.1016/j.landusepol.2019.104059