
Source of Transaction Price Data vs. Rates of Return on Housing Investment: Evidence from Poland

Submitted 15/08/25, 1st revision 29/08/25, 2nd revision 11/09/25, accepted 30/09/25

Sebastian Kokot¹, Krzysztof Janasz²

Abstract:

Purpose: The purpose of the research presented in this thesis is to estimate and assess the levels, trends and variation in returns on home purchases and rentals, as well as the impact on these phenomena of the data sources used

Design/Methodology/Approach: The paper thus addresses two problems related to the analysis of the housing market. The first is the trends in returns on investment in secondary market housing. The second is the reliability of housing price data published by various institutions. The research was conducted for 16 local markets - the largest cities in Poland. Statistical measures and charts were used for comparisons.

Findings: The analyses conducted prove that the prices of dwellings indicated by three different data sources vary. While in the case of prices published by the advertising service, the differences compared to other sources – CSO and NBP – are understandable, the significant differences between the transaction prices published by the two government institutions are inexplicable. The differences between the average rates of return obtained from different sources for individual cities are not large, rarely exceeding 0.5%. In terms of rates of return, the cities surveyed differed from each other to a much greater extent than the same cities differed in terms of data sources.

Practical implications: The paper shows that the data published by different institutions on average market prices of dwellings differ and that these differences are not subject to specific regularities. The indicated differences in prices do not translate into correspondingly significant differences in the levels of investment returns outside of incidental situations. However, such incidental distortions may be the reason for an inappropriate investment decision.

Originality value: The study is novel, as research on the explicitly addressed problem has not been undertaken to date. The study demonstrates that public datasets of market information should not be used for further analysis and research indiscriminately and that it is legitimate to evaluate them.

Keywords: Housing market, data quality, housing prices, rates of return, Poland.

JEL Codes: R10, R30.

Paper type: Research article.

¹University of Szczecin, Poland, ORCID: 0000-0001-7312-0984,
e-mail: sebastian.kokot@usz.edu.pl;

²University of Szczecin, Poland, ORCID: 0000-0002-3957-1797,
e-mail: krzysztof.janasz@usz.edu.pl;

Acknowledgement: *This research was co-financed by the Minister of Science under the "Regional Excellence Initiative.*



1. Introduction and Literature Review

Conducting real estate market analyses requires reliable and complete data. Problems arising from the legal, organisational and technical conditions for the recording of property prices mean that, as a result of a combination of many circumstances, published market data may be erroneous, unreliable, unreliable, outdated, and thus affect the quality of the analyses prepared on their basis. Real estate transactions are relatively rare.

The observation of prices is carried out in a specific way, as individual transactions concern diverse objects and, in addition, occur at irregular times (Kokot, 2015). The important function of the real estate sector in the national economy is indisputable, and the situation in the real estate market plays a significant role in the life of almost everyone, as the livelihood, occupational, infrastructural, cultural and other needs of society are met through real estate. Hence the need to generate reliable information on the real estate market situation, in particular on transaction prices.

These needs are being met by state institutions and commercial entities publishing data and analyses on the property market. From a societal point of view and due to the structure of the real estate market, the housing market is the main segment and it is to this market that this paper is devoted. Different countries have different systems for processing and sharing data on housing transaction prices and different institutions are responsible for these processes.

For example, in the UK, the Office for National Statistics publishes quarterly average house prices (1m² and total) in 9 regions and in London. In the UK, there is a government agency, HM Land Registry, which also publishes, among other things, average prices, as well as price indices and even data on individual transactions.

Reports from independent institutions such as NerdWallet on the property market are also available (Leonard, 2024). In Germany, on the other hand, the collection and processing of real estate market information is institutionally handled by the offices of the Commission of Experts (Gutachterschüsse) operating in the cadastral offices (Katasterämter) in each district (Kreis).

The information is anonymised and processed into, for example, price maps and other indicators. The results are publicly available and are additionally presented in annual reports. The legal basis for this is the German Building Code (Baugesetzbuch). In many countries, various institutions publish property price indices.

In Australia, among others, the Building Owners and Managers Association publishes the BOMA-Russell Property Index, the Australian Bureau of Statistics publishes the House Price Indices. In Canada, Statistics Canada publishes the New Housing Price Index.

In Hong Kong, the Ratings and Valuation Department publishes the Property Price Index. In Malaysia, the Ministry of Finance publishes the Malaysian House Price Index. In New Zealand, the Property Council of New Zealand publishes the Investment Performance Index. In Singapore, the Urban Redevelopment Authority publishes the URA Property Price Index (Lum, 2004). Averaged and reduced to index form, data on changes in house prices in individual European Union countries are published by Eurostat.

In Poland, data on average prices of flats on local markets are published by state institutions, such as the Central Statistical Office (CSO) and the National Bank of Poland (NBP), as well as by commercial institutions, such as urban.one (PriceR sp. z o.o.), morizon.pl (Grupa Morizon-Gratka Sp. z o.o.), analytics.otodom.pl (Grupa OLX sp. z o.o.). The main focus of these publications is also on the flat market, as among the various types of real estate, flats are the most frequently traded

We can broadly divide the sources of data on housing prices into 2 categories (Konowalczyk and Ramian, 2012):

- primary data - data collected by analysts at the point of recording individual market events,
- secondary data - data that have been collected as primary data by a specific entity and subsequently processed and made available in that form.

Other classifications may refer to subject and geographical divisions of properties and markets, e.g. the market for flats in a municipality, county, the market for flats as separate objects of ownership and the market for co-operative ownership rights to dwellings, the rental market (rental market), or types of prices (market prices, non-market prices, offer prices).

The primary source of data used for property market analyses should be primary data on market transactions. In Poland, data on transaction prices can be obtained from county offices (starostwo) within the structures of which there are geodetic documentation centres that maintain registers of property prices.

In practice, data from this source can be effectively used for analyses of selected local markets, as each starosty collects and can make available data on transactions registered in its area. An important shortcoming of data collected in price registers, especially from the point of view of analyses of the housing market, is that they do not include prices recorded in transactions of flats as objects of cooperative ownership rights to premises. Such information goes to housing cooperatives, which are not obliged to make it available to analysts, and obtaining complete data would require cooperation with all housing cooperatives operating on the local market.

In addition, data on real estate transactions can be found at tax offices, land and mortgage courts, notary offices and estate agency offices. However, they do not maintain systematic transaction databases. In practice, these are sources that are inaccessible to real estate analysts. Some of them - notary's offices and estate agents' offices have information only on those transactions which they themselves have handled.

The difficulty of obtaining complete primary data, the time-consuming nature of processing it and sometimes the high cost of obtaining it often lead analysts to use secondary sources, particularly publicly available sources. A drawback of data from such sources is the lack of insight into the mechanism of their generation. They also do not provide the freedom to regulate the scope of the data in terms of subject matter and geography.

Usually secondary data are in the form of statistical parameters (e.g., arithmetic mean) calculated for a specific period (e.g. a quarter) and geographical area (e.g., a city). As it has already been indicated, currently in Poland, secondary data on average residential prices can be obtained freely and free of charge from public sources such as the Central Statistical Office and the National Bank of Poland.

The CSO publishes databases on many areas of the country's social and economic life. From the point of view of the main subject of this work, data on median and average prices of 1m² dwellings are valuable. Such data are published quarterly for counties, broken down by primary and secondary market.

As indicated by the CSO, the presented data on sales volumes and prices of real estate were compiled on the basis of information obtained from Real Estate Price Registers maintained by county offices and presidents of cities with county rights, using information from notarial deeds. The data refers to sales of a market nature, i.e. sales on the free market (including, e.g., sales between related parties) and sales by tender.

The data refer to dwellings located in multi-family buildings (colloquially flats in blocks of flats) and do not refer to information on single-family houses. As such Real Estate Price Registers do not record transactions of co-operative ownership rights to premises, which in Poland are traded, so to speak, in parallel to premises

separated as real estate, these statistics do not take into account the trade in co-operative premises. As a result, the actual number of units sold is likely to be higher, and average prices do not entirely correctly reflect market conditions. Nevertheless, with such caveats, it seems that they could be successfully used to, for example, investigate price dynamics and differentiation. The data have been published since 2010.

Another useful source of data on the residential market in Poland for researchers is the National Bank of Poland, which publishes quarterly reports "on the situation on the residential and commercial real estate market in Poland" on its website as part of its cyclical analytical materials.

These publications are accompanied by the availability of the residential real estate price database comprising the average unit prices in selected cities and groups of cities, as well as the indices of price per 1 m⁽²⁾ of flats determined on their basis. The NBP's residential property price database is created by the submission of data by real estate agents and developers with the involvement of NBP District Branches.

The database constitutes valuable information material that can be used in analytical work in the area of this segment of the real estate market. It should be presumed that, in contrast to the CSO database, the data also include transactions of co-operative ownership rights to residential premises. In addition to the average transaction prices, the database also contains average offer prices, which gives an additional research field. On the other hand, the shortcomings of this database include the fact that it concerns the markets of only 17 large Polish cities. In addition, it is based on surveys, which may also weaken its relevance.

Unlike the CSO and NBP databases, the commercial databases contain bid price data. This is a significant difference and one should be aware of it when using them. In the presented study, comparative data from the otodomAnalytics database was used due to the largest temporal and territorial overlap with CSO and NBP databases. This database, available in the free version, publishes, among others, data on the average monthly offer price of 1m² flats in 40 cities as of January 2019.

Having at their disposal the above-characterised databases of average prices of 1m² flats, researchers and investors are faced with the dilemma of which one to use for market analyses and, consequently, whether or not to make an investment decision. The CSO and NBP databases are backed by the authority of state institutions and theoretically present the same, or at least very similar, information.

The differences lie in the way the input data is obtained. The CSO obtains them from price registers, while the NBP obtains them from surveys. In addition, the CSO base does not use data on transactions in co-operative ownership rights to dwellings. For these reasons, some minor and natural differences in published prices should be expected. The otodomAnalytics database concerns offers and not transactions.

This raises the question of whether the use of different databases of information on residential market prices can lead to different research conclusions and investment decisions? This article compares published prices and analyses their differences, as well as their impact on the magnitude of simple residential investment returns.

Thus, it is a kind of comparative evaluation of publicly available, from various sources, data sets on average prices of flats in Poland in terms of the possibility of their use for profitability analyses of investments in the housing market.

Analyses using such datasets on housing prices are conducted in many academic centres in Poland and concern various issues, including price indices (Tomczyk and Widłak, 2010; Kokot 2015), price convergence (Dittmann, 2014; Matysiak and Olszewski, 2019; Tomal, 2019), price anchoring (Kokot, 2023, Kokot and Doszyń, 2024), price factors (Leszczynski and Olszewski, 2017; Kokot, 2020; Melnychenko *et al.*, 2022), price forecasts (Belej, 2023), comparison and evaluation of price differentiation (Widłak and Nehrebecka, 2011; Dittmann, 2013a; Doszyń, 2023), housing availability (Matel and Marcinkiewicz, 2017; Czerniak *et al.*, 2022; Zakrzewska-Półtorak and Pluta, 2023).

Research on profitability from housing investments in an academic dimension is undertaken far less frequently and often in the form of reports (The Polish Real Estate Guide 2023). The profitability of the housing market is sometimes considered in different contexts.

The mainstream of research in this area concerns the susceptibility of investment returns to the volatility of various factors such as interest rates and expected prices (Tian, 2013), location and belonging to property clusters (Napoli *et al.*, 2017); tenant risk (Letdin *et al.*, 2022), population density (Fisher *et al.*, 2020), or local risk factors (Manganelli *et al.*, 2024).

The problem of input reliability in the context of investment decision-making is recognised in the work of Dittmann (2013b). The construction of rating of local markets as investment areas has been addressed as a subject of research on investing in the housing market (Altman and Rijken, 2004; Renigier-Biłozor *et al.*, 2018).

Other studies address the risks of investing and financing real estate investments (Gołabeska, 2018; Rybacka, 2024) and housing purchases as capital protection against inflation (Wolski, 2023). A somewhat more loosely related research area to the subject of this study is the estimation of the size of the rental market (Derkacz and Kohen, 2024).

A separate strand of research is the quality of real estate market data and, in fact, its impact on the research results obtained. Work on assessing various aspects of the quality of data used for property market analyses has also been undertaken (Ge and Harfield, 2006; Tomczyk and Widłak, 2010; Nawrocka, 2013; Konowalczyk, 2014).

The conclusions drawn from them are rather pessimistic, which nevertheless prompts one to pay great attention to the sources and quality of data used in real estate market research. During the literature studies, no research was encountered that involved a direct comparison of property price data from different sources (Konowalczyk, 2014), nor of the impact of the data source on the research results.

It is also worth mentioning that uncertain data quality does not only apply to the property market but is a general problem of public statistics, where many limitations of data acquisition and processing are revealed (Cook, 2004).

The aim of the research undertaken is to estimate and assess the levels, trends and variation in the rates of return on purchases and rentals of flats in secondary local markets in Poland's largest cities, as well as the impact on these phenomena of the data sources used.

2. Material and Methods

For the purposes of the study presented here, the following datasets were used:

1. Concerning housing prices:
 - On average quarterly prices per $m^{(2)}$ of dwellings sold in secondary market transactions published by the Central Statistical Office in the Local Data Bank. These data are compiled on the basis of information obtained from the Property Price Register (until 31 July 2021 the Property Price and Value Register) [C_{ti}^{CSO}]
 - On the average quarterly transaction prices of 1 m^2 flats on the secondary market from the National Bank of Poland's residential property price database, which is created by submitting data from real estate agents and developers with the involvement of NBP district branches. The database was downloaded from the NBP website [C_{ti}^{NBP}]
 - On average offer prices per 1 $m^{(2)}$ of secondary market flats published on OtodomAnalytics read for the months of February, May, August and November as the middle months, representing each quarter [C_{ti}^O]
2. Concerning rental income from flats - on average monthly rental offer prices of flats published on OtodomAnalytics read in the same way as data on prices of flats on this website [D_{ti}^O]. Unfortunately, due to the lack of publications on transactional rental prices, the study only used data on rental offer prices.

In addition, for the purpose of performing the calculations, information was needed on the average turnoverable area of dwellings in a given period in each city, which was calculated as the ratio of the total turnoverable area and the number of turnoverable dwellings. This data was obtained from the CSO publication in the Local Data Bank [P_{ti}].

The scope of all data covers the period from Q1 2019 to Q4 2022. This period has been determined by the common scope of data from the indicated sources - data from Otodom is available from 2019 and the latest data from the Central Statistical Office at the date of undertaking the research covered 2022. The data covers the cities of Bialystok, Bydgoszcz, Gdansk, Katowice, Kielce, Krakow, Lublin, Lodz, Olsztyn, Opole, Poznan, Rzeszow, Szczecin, Warsaw, Wroclaw, Zielona Gora.

The analytical work included:

1. Presentation, by means of line graphs, of average transaction prices of 1m² flats published by CSO and NBP and average offer prices of 1m² OtodomAnalytics recorded quarterly for particular surveyed cities. This initial stage of the research is aimed at preliminary assessment whether there are differences between data on prices of 1m² flats, whether they are similar in all cities, whether there are noticeable regularities in their course (e.g. whether prices from one source are always higher than prices from another source, whether data from one source show greater fluctuations, whether they are characterised by consistent dynamics, etc.).
2. Calculation of the relative differences between the prices of 1m² dwellings published by the CSO and the NBP in each city, which were determined from the formula:

$$R_{ti}^{rel} = \frac{C_{ti}^{CSO} - C_{ti}^{NBP}}{C_{ti}^{CSO}} \times 100\% \quad (1)$$

Where:

C_{ti}^{CSO} - the average price of 1m² dwellings in period t in the i-th city as published by the CSO,

C_{ti}^{NBP} - average price of 1m² dwellings in period t in the i-th city as published by the NBP.

This stage of the study was aimed at identifying the scale of differences in the levels of published prices. Formally, prices published by CSO and NBP are average transaction prices, so despite some methodological differences in their determination, similar levels should be expected. Differences in relation to offer prices from Otodom Analytics were not calculated due to the different, not directly comparable nature of these data (transaction prices vs offer prices). The relative differences between the prices of 1m² flats published by CSO and NBP were presented by means of a box-and-whisker chart, which included information on their basic structure parameters: arithmetic mean, median, arithmetic mean +/- standard deviation, minimum and maximum values.

3. Calculation of average simple rates of return as a ratio of annual residential

rental income to average prices based on data from individual sources using the following formulas:

– based on CSO data:

$$r_{ti}^{CSO} = \frac{D_{ti}^O \times 12 \times 0,95}{C_{ti}^{CSO} \times P_{ti}} \times 100\% \quad (2)$$

– based on NBP data:

$$r_{ti}^{NBP} = \frac{D_{ti}^O \times 12 \times 0,95}{C_{ti}^{NBP} \times P_{ti}} \times 100\% \quad (3)$$

– based on OtodomAnalytics data:

$$r_{ti}^O = \frac{D_{ti}^O \times 12}{C_{ti}^{CSO} \times P_{ti}} \times 100\% \quad (4)$$

Where:

D_{ti}^O - average monthly rental price in period t in the i-th city

P_{ti} - the average area of the dwelling in period t in the i-th city.

The returns determined on the basis of CSO and NBP data include a correction coefficient of 0.95 to reflect the effect of negotiations resulting in an adjustment of the offer price. The amount of the coefficient was determined after consultations with real estate agents represented by Bas Nieruchomości and Ciepłiński Nieruchomości offices.

The author is aware of the effects of such assumptions, in particular their impact on the results of the analyses and the related interpretation. This coefficient was not applied to the determination of the simple rate of return on the basis of data from OtodomAnalytics due to the fact that in this case both income (numerator) and prices (denominator) are bid data, i.e. subject to negotiation in the process of arriving at income and market prices.

The average rates of return were presented by means of line graphs. As can be seen, the study used the so-called simple rate of return formula, sometimes equated with ROI (return on investment). In fact, the total rate of return on real estate investment captures the increase in value of the property in addition to the rental income. However, it is difficult to estimate - requiring the use of forecasting tools that often fail anyway.

Furthermore, there are no official publications on such forecasts. It was therefore considered reasonable and sufficient for analytical and

comparative purposes to use the simple rate of return formula as the ratio of predicted average income to average prices. Income from real estate has been taken into account as gross income, i.e. without deducting losses that may arise from periodic tenant shortages and rent arrears, as well as operating expenses understood as property maintenance costs incurred by the owner.

4. Calculation of the basic parameters of the average return structure: arithmetic mean, standard deviation, coefficient of variation, minimum and maximum. The results obtained are presented numerically in a table and, in addition, the arithmetic means, arithmetic mean $\pm$ standard deviation, minimum and maximum are presented by means of box-and-whisker diagrams in two arrangements: by city and by period.
5. Conduct a cluster analysis of returns for cities calculated from data extracted from various sources using the Ward method. This is a traditional and frequently used method of extracting clusters. It is effective when the number of units is not very large and generally determines what the actual number of clusters is (Migdal-Najman and Najman, 2013). Ward's method was applied using Euclidean distances. This analysis identified clusters of cities that were similar in terms of the levels of returns obtained. In addition, obtaining equal clusters of rates of return calculated on the basis of different data sources will demonstrate the lack of influence of the data source on the level of rates of return in the context of belonging of a given city to a cluster of cities characterised by similar levels of obtained rates of return.

It was expected that the analyses carried out according to the scheme presented would enable the questions to be answered:

- What are the rates of return on investment in housing in the cities surveyed?
- Do the cities surveyed differ in terms of recorded levels of return?
- Can similar trends in the course of returns be observed in the cities studied?
- Does the source of price data affect the results of return calculations?

3. Results and Discussion

The average 1m² secondary market prices quoted in each city from Q1 2019 to Q4 2022 according to CSO, NBP and OtodomAnalytics publications are presented in Appendix A. Analysis of the charts contained therein leads to the following observations. In the period under study, in all cities, the prices of 1m² flats published by the three centres under consideration, although differing in levels, showed an increasing trend. Invariably, offer prices were higher than transaction prices published by both the CSO and the NBP, to varying degrees in different cities.

The most pronounced differences were recorded for large cities with high price levels: Gdańsk, Kraków, Warszawa and Wrocław, although they are also clear for Katowice, Kielce and Zielona Góra.

From the point of view of this study, however, it is important that differences between average transaction prices published by the CSO and the NBP become visible and, what is perhaps most important, it is difficult to find an unambiguous regularity.

There are cities where CSO prices are higher than NBP prices (Białystok, Kielce, Warsaw), as well as cities where the opposite is true (Bydgoszcz, Szczecin, Rzeszów, Zielona Góra). There are also cities where these prices intermingle or are generally similar. The greater variability of CSO prices is also noteworthy, which is particularly evident for Poznań, Szczecin, Rzeszów and Zielona Góra.

Figure 1 shows the relative differences between the prices of 1m² dwellings in each city published by the CSO and the NBP. The chart provides information on the arithmetic mean, median, mean $\pm$ standard deviation and the range of these differences. We can see that the average differences for most cities are no greater than $\pm 5\%$ and usually the arithmetic mean is close to the median, indicating that the distributions of calculated differences for individual cities are symmetrical. This level of differences can be justified by the extent of the source data and the way in which it was obtained by the CSO and the NBP.

However, average differences amounting to -12% for Rzeszów or even -7% for Szczecin seem to be already large enough to raise doubts concerning the reliability of at least one of those sources. Critical assessments are also prompted by the values of $\bar{x} + s$ and $\bar{x} - s$, which often exceed $\pm 5\%$, and there are also cities for which the values exceed -10%.

From the figure we also conclude that in individual cases the differences between the prices published by the CSO and the NBP may exceed 20 or even 30%, as indicated by the extreme values. Certainly, such large differences are unacceptable and may indicate procedural errors in the analyses underlying the publication of such data. It is to be expected that such large discrepancies in published housing price data will lead to divergent conclusions regarding the profitability of investment.

The average rates of return on investment in buying and renting a flat between Q1 2019 and Q4 2022 in the surveyed cities, calculated using data from the three sources considered, are presented in Appendix B. An analysis of the graphs contained therein shows that, regardless of the source of the data, in all cities the rates of return fell until the beginning of 2022, after which they began to rise. In general, the lowest levels of return rates are indicated by data from Otodom Analytics, while for some cities these levels were clearly lower than return rates calculated on the basis of CSO and NBP data (Gdańsk, Katowice, Kraków).

There are also cities for which these differences clearly fluctuated (Warsaw, Wrocław, Zielona Góra). However, the pace of these changes was different and, obviously, the levels of return rates in individual cities were different. Smaller differences, although also noticeable, concern rates of return calculated on the basis of data from CSO and NBP.

Interestingly, it happens that rates of return calculated on the basis of transaction prices published by one institution are closer to rates of return calculated on the basis of offer prices than to rates of return calculated on the basis of transaction prices published by the other institution (Poznań, Rzeszów). In order to assess the rates of return for individual cities and data sources more precisely, their structure parameters were calculated.

They are presented in Table 1. We can see that the differences between the levels of average rates of return are not large, rarely exceeding 0.5%, which indicates that despite significant differences in the levels of prices published by individual institutions, the differences in the levels of rates of return are not so large that an investment decision made on the basis of data from one source could be different from an analogous decision made on the basis of data from another source.

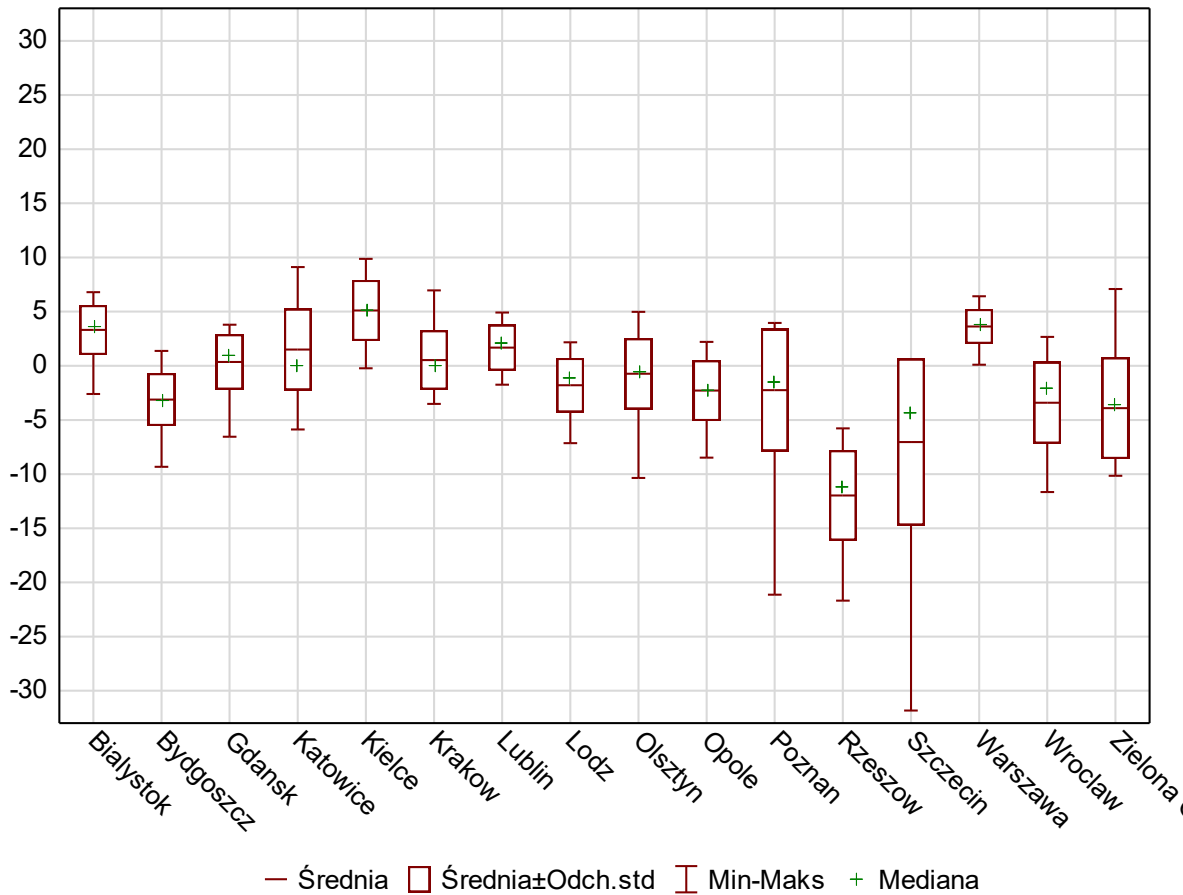
Higher levels of differences, however, result from the comparison of maximum levels of returns, where differences in excess of 1% already occur. Attention is also drawn to the relatively high values of volatility measures - standard deviations of returns and volatility coefficients.

The above observations are confirmed by an analysis of the graphics contained in Appendix C, where the arithmetic means, arithmetic means $\pm$ standard deviations and minimum and maximum values of the rates of return (%) in each city for the quarters of 2019-2022, calculated on the basis of CSO, NBP and OtodomAnalytics data, are presented in the form of graphs frame-slides.

The average rates of return determined based on the different data sources are similar, with the average rates of return determined from CSO and NBP data usually being closer to each other and the rates of return determined from OtodomAnalytics data usually being slightly lower. Rzeszów breaks out of this rule. However, the attention is drawn to the varied levels of standard deviations and extreme values. It should be noted that in cities for which relatively high levels of these measures were recorded, they are high regardless of the source of data from which they were determined.

In other words, there is no city for which the measures of variability determined on the basis of one source of data would be large, and on the basis of another source - small. Another observation concerns the asymmetry of the extreme values - the minimum in all cities and for all data sources is closer to the mean than the maximum.

Figure 1. Relative price differences of 1m² published by the CSO and the NBP on the secondary market in the surveyed cities [%].



Source: Own study.

Table 1. Return rate structure parameters

	CSO					NBP					OtodomAnalytics				
	¶	s	Vs	min	max	¶	s	Vs	min	max	¶	s	Vs	min	max
Białystok	5.90	0.60	10.16	4.97	7.16	6.10	0.53	8.70	5.25	7.30	5.71	0.49	8.64	4.70	6.71
Bydgoszcz	7.05	0.81	11.45	6.08	8.63	6.84	0.76	11.11	5.95	8.46	6.60	0.66	10.06	5.81	7.98
Gdańsk	5.83	0.70	11.99	5.02	7.07	5.85	0.62	10.66	4.98	6.92	5.19	0.47	9.06	4.56	5.96
Katowice	7.26	0.96	13.18	6.14	9.68	7.37	0.75	10.14	6.34	9.15	6.58	0.64	9.70	5.71	8.18
Kielce	5.98	0.71	11.87	5.06	7.71	6.30	0.71	11.25	5.25	8.13	5.60	0.58	10.28	4.70	7.01
Kraków	5.75	0.69	12.04	4.79	7.07	5.80	0.79	13.58	4.67	7.27	5.10	0.60	11.82	4.22	6.14
Lublin	6.62	0.81	12.31	5.34	8.33	6.73	0.83	12.38	5.60	8.49	6.55	0.76	11.58	5.46	8.17
Łódź	7.95	1.01	12.71	6.71	9.91	7.81	0.98	12.51	6.54	9.86	7.60	0.90	11.87	6.52	9.30
Olsztyn	5.42	0.45	8.24	4.69	6.38	5.38	0.39	7.33	4.94	6.08	5.29	0.40	7.54	4.75	5.97
Opole	6.45	0.59	9.10	5.60	7.32	6.31	0.51	8.11	5.40	7.22	6.19	0.52	8.36	5.31	7.07
Poznań	5.61	0.48	8.48	4.98	6.54	5.49	0.48	8.79	4.92	6.50	5.31	0.45	8.50	4.77	6.16
Rzeszów	6.29	0.79	12.59	5.03	7.72	5.62	0.71	12.60	4.75	6.99	5.60	0.66	11.82	4.82	7.10
Szczecin	6.88	1.25	18.20	4.86	10.16	6.43	1.04	16.14	4.49	8.07	6.27	0.98	15.60	4.41	7.88
Warsaw	7.22	0.96	13.25	5.85	8.76	7.49	0.94	12.51	6.11	9.11	7.00	0.81	11.53	5.76	8.37

Wrocław	6.44	0.76	11.80	5.44	8.01	6.22	0.64	10.21	5.22	7.57	5.95	0.60	10.01	5.19	7.04
Zielona	7.17	0.58	8.07	6.13	8.33	6.90	0.59	8.49	6.00	8.37	6.49	0.59	9.02	5.58	7.77

Source: *Own study.*

On the other hand, Appendix D also presents, by means of box-and-whisker charts, the arithmetic means, arithmetic means $\pm$ standard deviations and extreme values of returns (%) determined for each quarter of the years 2019-2022 in the surveyed cities. The purpose of including the data in this way is to try to identify quarters where one or all of the data sources show irregularities, which could be indicative of some kind of market turbulence.

As before, it can be seen that the returns determined from the Otodom Analytics data are slightly lower than the returns determined from the data from the other sources. However, it is interesting to note that from the beginning of the observations until around Q3 2021, the volatility of the returns decreases - we observe smaller and smaller standard deviations and increasingly close to the average minimum and maximum values, which are usually not characterised by a clear asymmetry.

The situation slowly starts to change from Q4 2021 onwards, and these changes are already very evident in Q3 and Q4 2022, when there is an increase in the means, standard deviations and extreme values - both minimum and maximum returns. The asymmetry of the extreme values, which was not noticeable before, is also evident. What is important, however, is that the same regularities apply to the returns determined from all of the data sources considered.

Appendix E presents the results of the cluster analysis of rates of return for cities calculated using data obtained from different sources using the Ward method. The results obtained for different data sources are not the same, which indicates that the source of data influences the obtained levels of rates of return and, consequently, may induce different potential investment decisions.

However, it should also be noted that the differences in clustering are not diametrical. Regardless of the source of data on residential prices, Łódź, Zielona Góra, Katowice, Warsaw and Bydgoszcz remained in one large cluster, and Poznań, Olsztyn, Rzeszów, Kraków and Gdańsk in the other. The cities that changed their affiliation within the largest separate cluster were Lublin, Szczecin, Wrocław, Opole, Kielce and Białystok.

Obviously, at the lower level of clusters there were more pronounced differences between them. Summing up this stage of the study, it should be stated that for 10 out of 16 cities included in the study, rates of return obtained in the studied period were similar enough, regardless of the source of data, to belong to the same clusters. On the other hand, for 6 cities, the source of data on residential prices influenced their belonging to the appropriate cluster. Therefore, it is not possible to unequivocally state that the data source has no impact on the level of obtained rates of return.

4. Summary and Findings

The conducted analyses prove that prices of flats indicated by the 3 data sources taken into account differ. While in the case of prices published by OtodomAnalytics, the differences with respect to the other sources - CSO and NBP - are understandable, as, unlike them, they concern offer prices, sometimes significant differences between transaction prices published by two important state institutions are inexplicable. The regularities concerning these differences cannot be clearly identified, as they vary from city to city.

The simple average housing yields depending on time, city and data source are in the range of 4.11 - 10.16%. The differences between simple average rates of return over the period under study for different cities according to different data sources range from 5.10 to 7.95%, which should be considered significant. These are average levels, so when considering specific investments both the levels of returns and the differences between them can be even greater.

The differences between the levels of average rates of return obtained using data from different sources for individual cities are not large, rarely exceeding 0.5%. In terms of the levels of rates of return, the cities studied differed significantly more than the same cities due to the source of the data. The variation in rates of return is therefore greater by city than by data source.

The general trends in the courses of rates of return in the surveyed cities are similar, but differ in the details and these differences are influenced by the data source. The source of the price data has an impact on the results of the return calculations, although this impact is smaller than is the case for prices alone. For 10 of the 16 cities studied, it was not so large that the cluster analysis qualified them into the same large cluster regardless of the source of the price data.

The obtained results lead to the formulation of a postulate for the Central Statistical Office and the National Bank of Poland, as important state institutions, to make an effort to improve the quality of their published data on housing transaction prices. The situation in this respect should be monitored by scientific circles by undertaking relevant research.

Diversification and classification of local markets in terms of the obtained levels of real estate market rates of return and improvement of techniques of obtaining reliable market data for such analyses also appear as interesting directions for future research, in the light of the obtained results.

References:

Altman, E.I., Rijken, H.A. 2004. How rating agencies achieve rating stability. *Journal of Banking and Finance*, 28(11), 2679-2714.

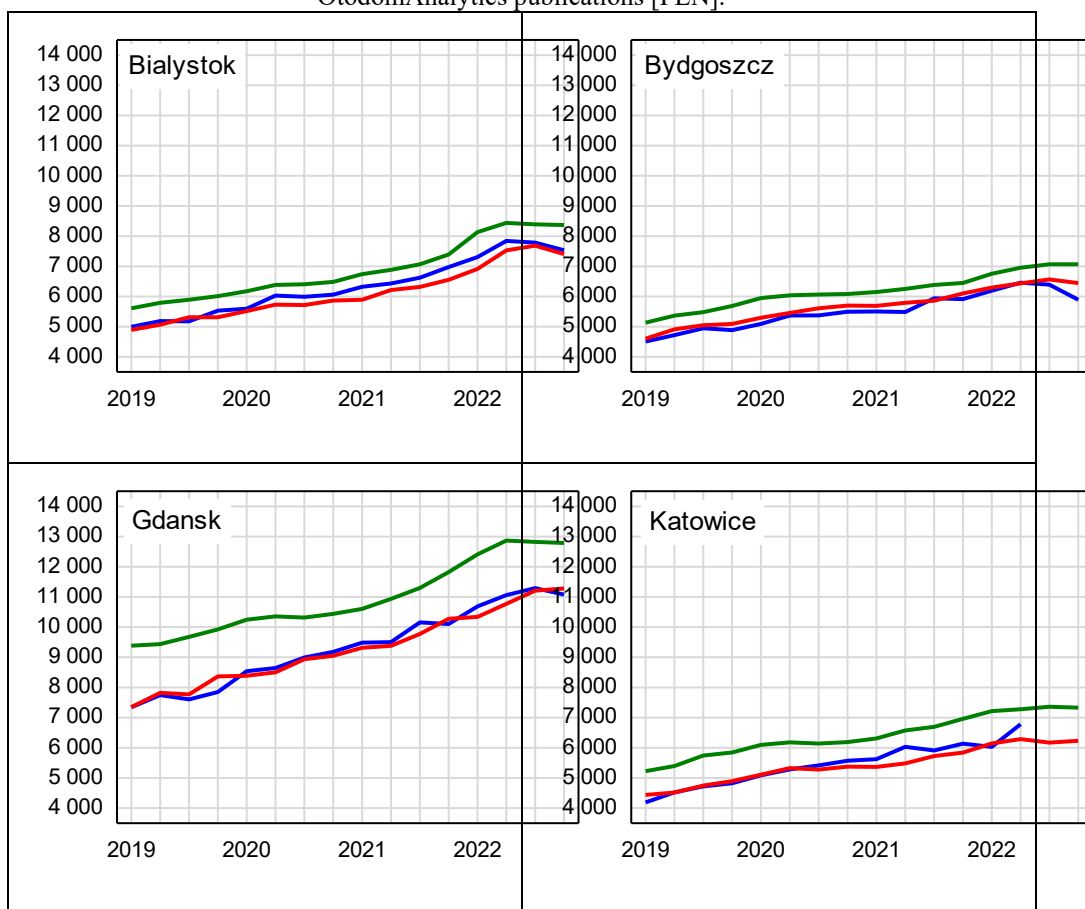
- <http://dx.doi.org/10.1016/j.jbankfin.2004.06.006>.
- Belej, M. 2023. Predicting Housing Price Trends in Poland: Online Social Engagement - Google Trends. *Real Estate Management and Valuation*, 31(4), 73-87. <https://doi.org/10.2478/remav-2023-0032>.
- Cook, L. 2004. Revisions to statistics: their role in measuring economic progress. *Economic Trends*, 603, 36-43.
- Czerniak, A., Czaplicki, M., Mokrogulski, M., Niedziółka, P. 2022. Dostępność finansowa mieszkań w krajach Europy Środkowo-Wschodniej na tle zmian parametrów polityki pieniężnej. Raport SGH i Forum Ekonomicznego, 85-108. <http://doi.org/10.33119/978-83-8030-563-2.2022.3.85.108>.
- Derkacz, A.J., Cohen, V. 2024. The size of the rental housing segment in Poland and its main determinants. *Real Estate Management and Valuation*, 32(1), 49-57. <https://doi.org/10.2478/remav-2024-0005>.
- Dittmann, I. 2013a. 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. <http://dx.doi.org/10.2478/remav-2013-0006>.
- Dittmann, I. 2013b. Podejmowanie decyzji inwestycyjnych na Rynkach nieruchomości mieszkaniowych w oparciu o prognozy cen mieszkań konstruowane na podstawie opinii ekspertów zużyciem rozkładu Weibulla (Making Investment Decisions on Housing Markets Based on Forecasts Housing Prices Constructed on the Basis of Expert Opinion Using the Weibull Distribution). *Studia Ekonomiczne*, 155, 162-174.
- Dittmann, I. 2014. Gamma konwergencja cen na lokalnych rynkach mieszkaniowych. *Studia Ekonomiczne*, 181, 195-207.
- Doszyń, M. 2023. Integration and cointegration of apartment prices on the primary and secondary market in Szczecin in the years 2006-2022. *Real Estate Management and Valuation*, 31(4), 36-44. <https://doi.org/10.2478/remav-2023-0028>.
- EY. 2023. The Polish Real Estate Guide. Retrieved from: https://assets.ey.com/content/dam/ey-sites/ey-com/pl_pl/topics/real-estate/ey-book_re_guide_2023.pdf.
- Fisher, G., Steiner, E., Titman, S., Viswanathan, A. 2020. How does property location influence investment risk and return. Available at: https://www.reri.org/research/files/2020_propertylocation_fisher_steiner_titman_viswanathan.pdf.
- Ge, X.J., Harfield, T. 2006. The Quality of Data and Data Availability for Property Research. Available at: https://www.prres.org/uploads/796/1571/Ge_Quality_Data.pdf.
- Gołabeska, E. 2018. Sieć ryzyka inwestycyjnego na rynku nieruchomości. Wydawnictwo Politechniki Białostockiej.
- Kokot, S. 2015. Przyczynek do badań nad indeksami cen nieruchomości. *Wiadomości Statystyczne*, 647(4), 1-12. <https://doi.org/10.5604/01.3001.0016.0850>.
- Kokot, S. 2020. Socio-economic factors as a criterion for the classification of housing markets in selected cities in Poland. *Real Estate Management and Valuation*, 28(3), 77-90. <https://doi.org/10.1515/remav-2020-0025>.
- Kokot, S. 2023. The Effect of Price Anchoring on the Housing Market Based on Studies of Local Markets in Poland. *Real Estate Management and Valuation*, 31(3), 44-57. <https://doi.org/10.2478/remav-2023-0020>.
- Kokot, S., Doszyń, M. 2024. Econometric modeling of average housing prices in local markets and the price anchoring effect. *Real Estate Management and Valuation*, 32(3), 116-126. <https://doi.org/10.2478/remav-2024-0029>.
- Konowalczuk, J. 2014. The corporate real estate market in public statistics in Poland. *Real*

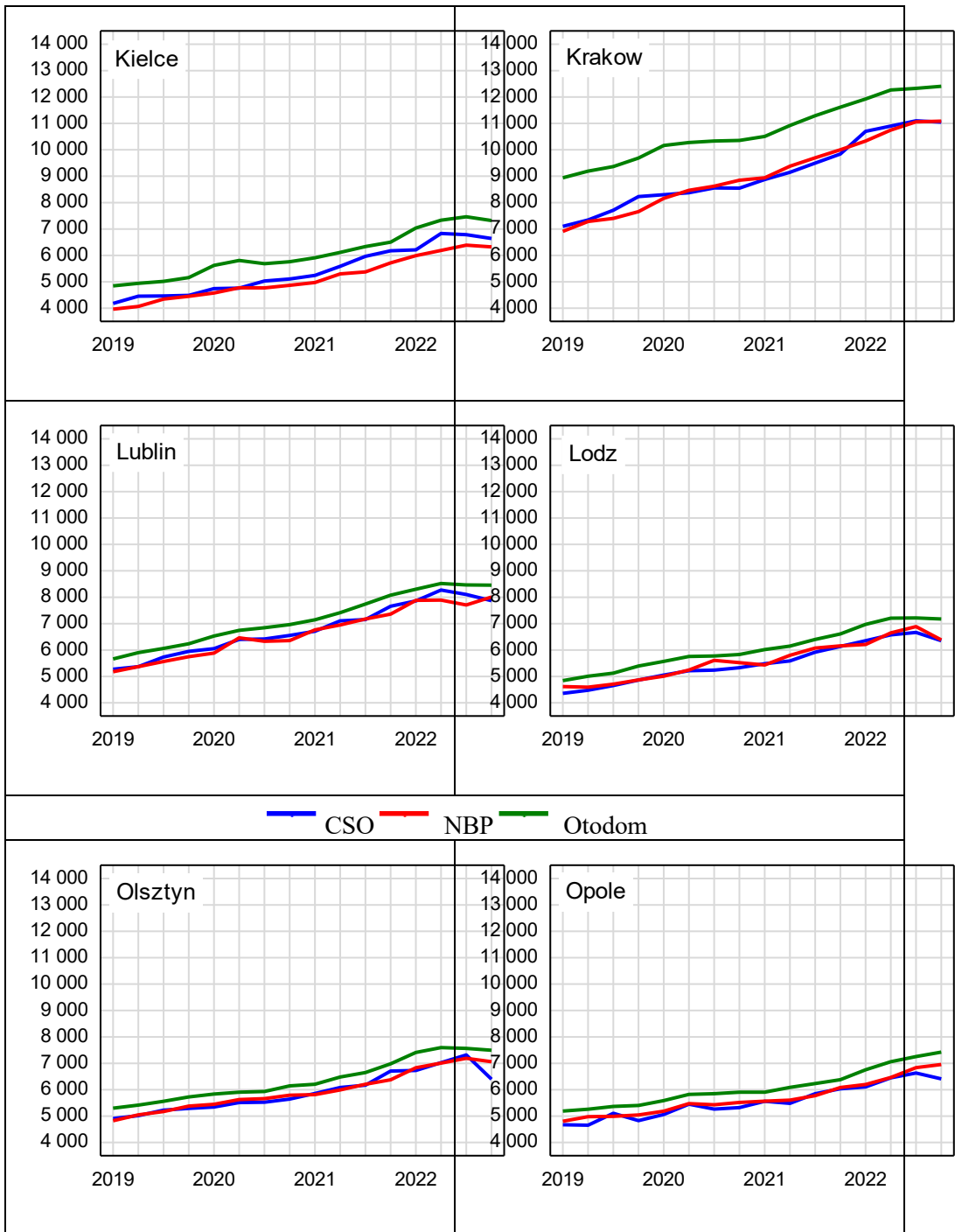
- Estate Management and Valuation 22(2), 41-51. <https://doi.org/10.2478/remav-2014-0016>.
- Konowalczuk, J., Ramian, T. 2012. Potrzeby i bariery informacyjne rynku nieruchomości z perspektywy wyceny nieruchomości w Polsce. *Studia i Materiały Towarzystwa Naukowego Nieruchomości*, 20(2), 105-118.
- Leszczyński, R., Olszewski, K. 2017. An analysis of the primary and secondary housing market in Poland: evidence from the 17 largest cities. *Baltic Journal of Economics*, 17(2), 136-151. <https://doi.org/10.1080/1406099X.2017.1344482>.
- Leonard, T. 2024. UK House Price Trends & Predictions December. <https://www.nerdwallet.com/uk/mortgages/house-prices/>.
- Letdin, M., Sirmans, G.S., Smersh, G.T., Zhou, T. 2022. An analysis of the determinants of retail capitalization rates. Available at: https://www.reri.org/research/files/determinants_of_retail_caprates.pdf.
- Lum, S.K. 2004 Property price indices in the Commonwealth – Construction methodologies and problems. *Journal of Property Investment & Finance*, 22, No. 1, 25-54.
- Manganelli, B., Anelli, D., Tajani, F., Morano, P. 2024. Capitalization rate and real estate risk factors: An analysis of the relationships for the residential market in the city of Rome (Italy). *Real Estate Management and Valuation*, 32(3), 101-115. <https://doi.org/10.2478/remav-2024-0028>.
- Matel, A., Marcinkiewicz, J. 2017. Analiza dostępności mieszkań w miastach wojewódzkich Polski z wykorzystaniem metody topsis affordability. *Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 475, 170-182. <https://doi.org/10.15611/pn.2017.475.15>.
- Matysiak, G.A., Olszewski, K. 2019. A panel analysis of Polish regional cities: residential price convergence in the primary market. NBP Working Paper, 316.
- Melnychenko, O., Osadcha, T., Kovalyov, A., Matskul, V. 2022. Dependence of Housing Real Estate Prices on Inflation as One of the Most Important Factors: Poland's Case. *Real Estate Management and Valuation*, 30(4), 25-41. <https://doi.org/10.2478/remav-2022-0027>.
- Migdal-Najman, K., Najman, K. 2013. Analiza porównawcza wybranych metod analizy skupień w grupowaniu jednostek o złożonej strukturze grupowej. *Zarządzanie i Finanse*, 3(2), 179-194.
- Napoli, G., Giuffrida, S., Trovato, M.R., Valenti, A. 2017. Cap rate as the interpretative variable of the urban real estate capital asset: A comparison of different sub-market definitions in Palermo, Italy. *Buildings*, 7(4), 80. <https://doi.org/10.3390/buildings7030080>.
- Nawrocka, E. 2013. Analiza i ocena jakości krajowych zasobów informacji o nieruchomościach w aspekcie ich wykorzystania do opracowania indeksu cen. *Studia Ekonomiczne*, 155, 82-96.
- Renigier-Biłozor, M., Wisniewski, R., Kaklauskas, A., Biłozor, A. 2018. Rating methodology for real estate markets – Poland case study. *International Journal of Strategic Property Management*, 18(2), 198-212. <https://doi.org/10.3846/1648715X.2014.927401>.
- Rybacka, J. 2024. Real Estate as Investment Instruments in Poland Between 2020 and 2023. *Annales Universitatis Mariae Curie-Skłodowska, sectio H – Oeconomia*, 58(2), 153-170. <http://dx.doi.org/10.17951/h.2024.58.2.153-170>.
- Tian, C.Y. 2013. Are capitalization rates really constant within housing markets? Available at SSRN 2282780. <https://doi.org/10.2139/ssrn.2282780>.
- Tomal, M. 2019. House Price Convergence on the Primary and Secondary Markets:

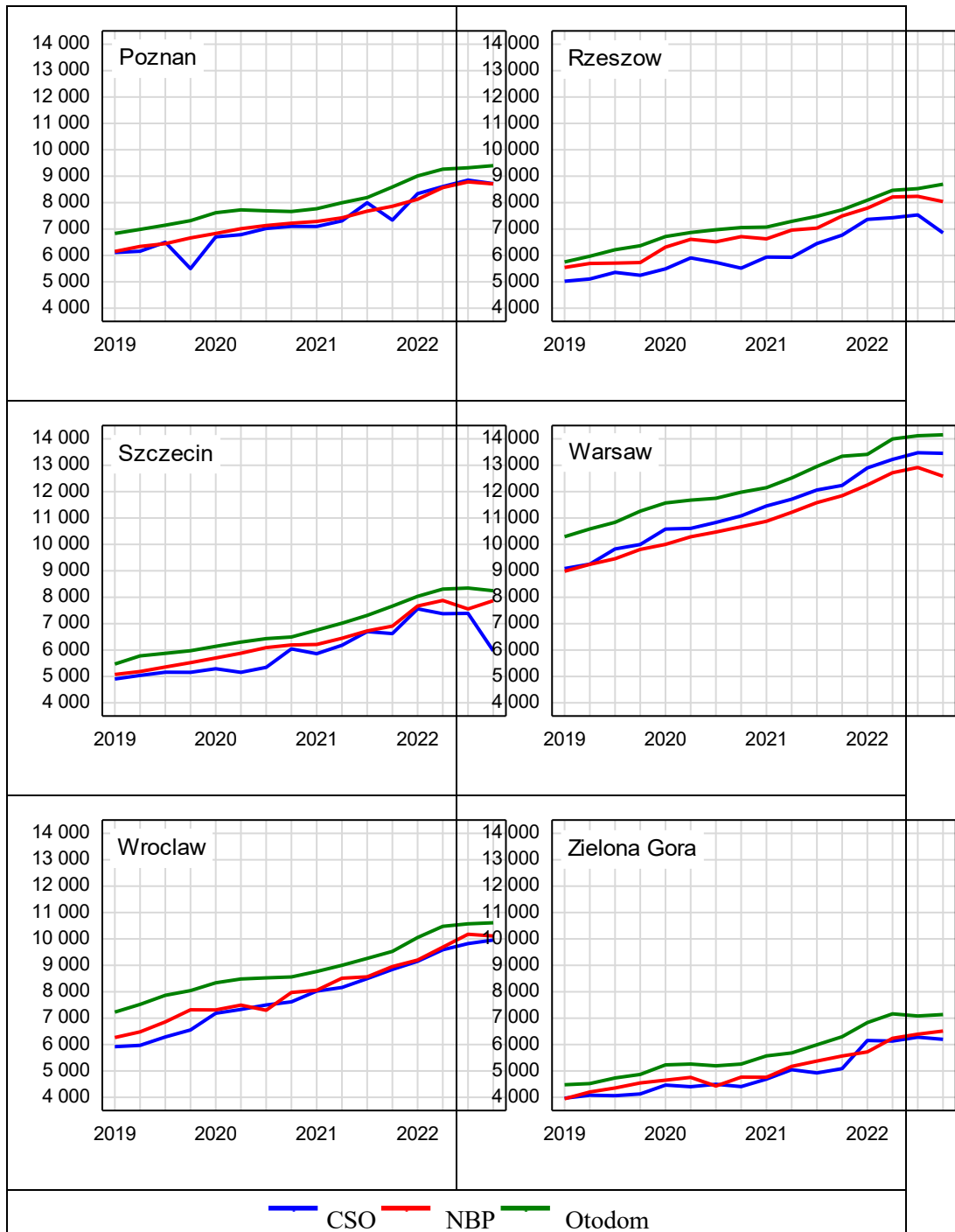
- Evidence from Polish Provincial Capitals. *Real Estate Management and Valuation*, 27(4), 62-73. <https://doi.org/10.2478/remav-2019-0036>.
- Tomczyk, E., Widłak, M. 2010. Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy. *Bank i Kredyt*, 41(1).
- Widłak, M., Nehrebecka, N. 2011. Wykorzystanie regresji kwantylowej w analizie różnicowania cen mieszkań. *Wiadomości Statystyczne. The Polish Statistician*, 600(5), 17-46.
- Wolski, R. 2023. Residential real estate as a potential hedge of capital against inflation. *Real Estate Management and Valuation*, 31(1), 36-42. <https://doi.org/10.2478/remav-2023-0004>.
- Zakrzewska-Półtorak, A., Pluta, A. 2023. Zmiany w zakresie atrakcyjności nieruchomości mieszkaniowych w polskich miastach w latach 2010-2020 ze szczególnym uwzględnieniem dostępności. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 67(3). 164-176. <https://doi.org/10.15611/pn.2023.3.15>.

Appendix A:

Average prices of 1m² on the secondary market according to CSO, NBP and OtodomAnalytics publications [PLN].



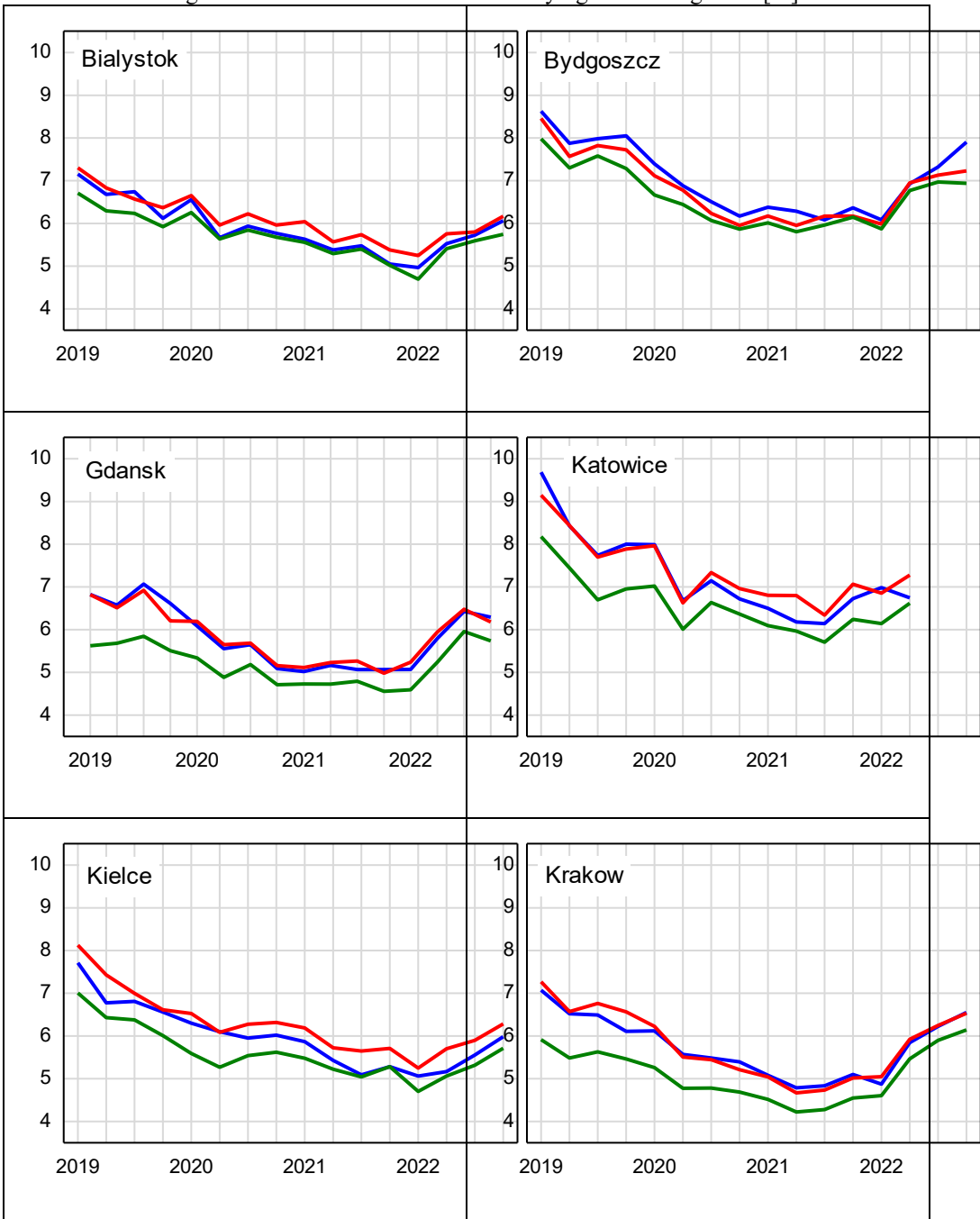


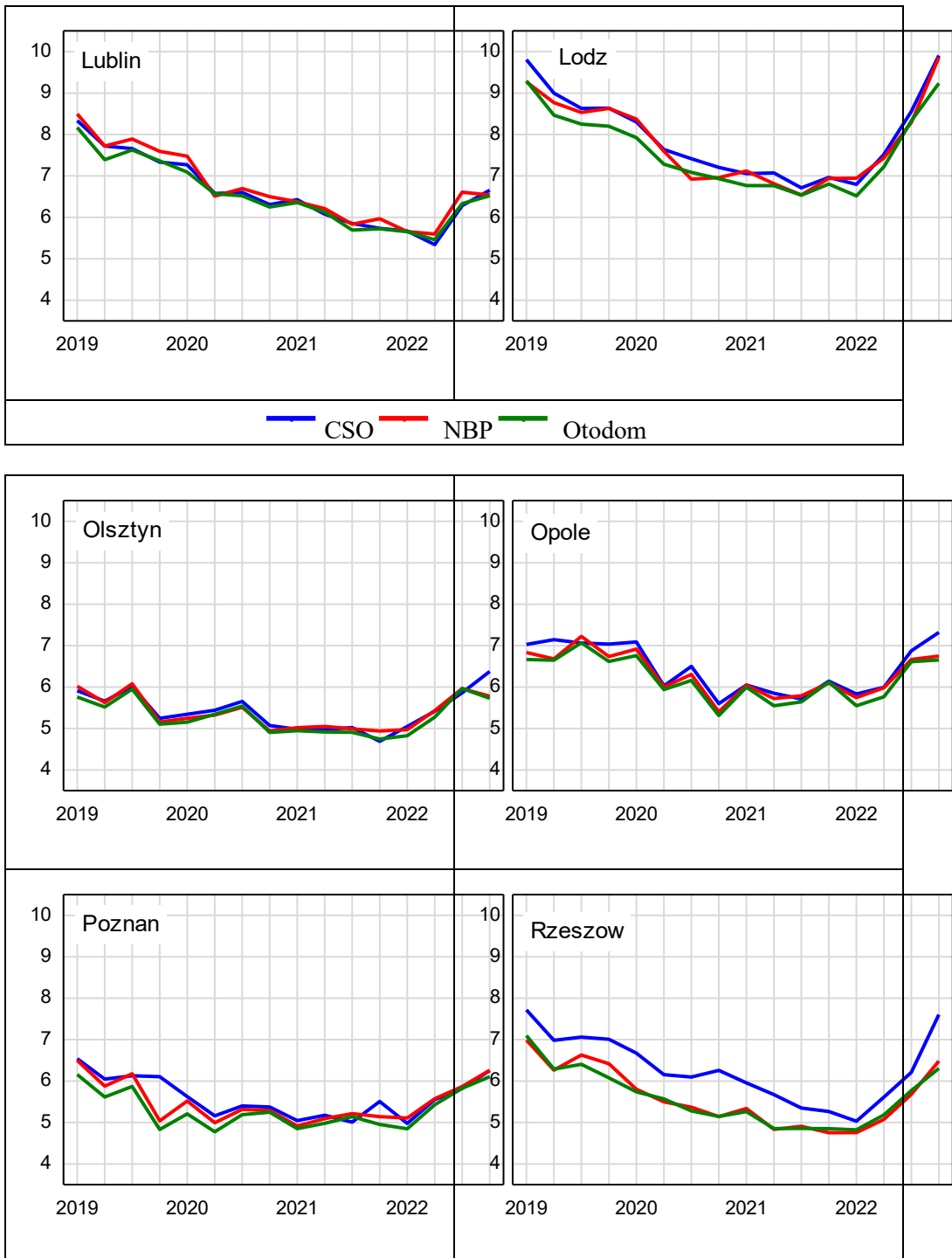


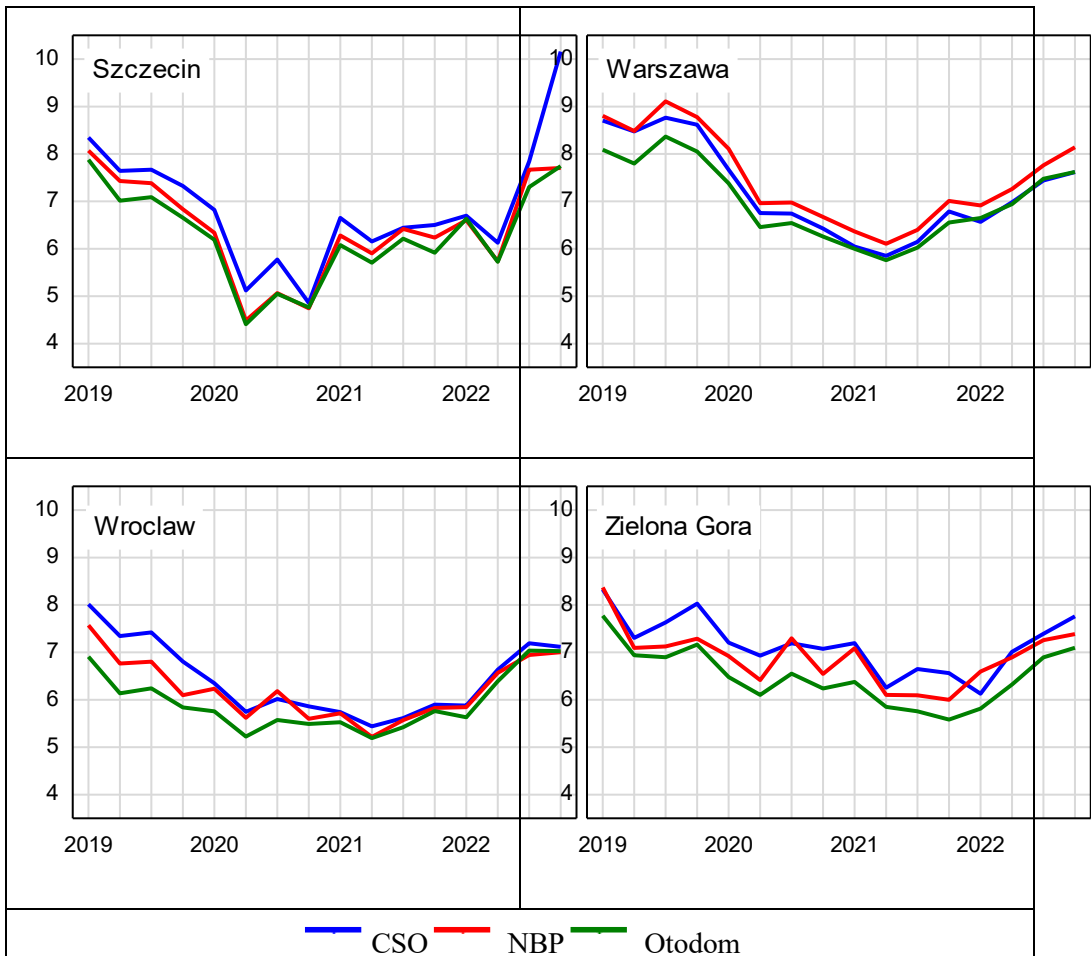
Source: Own study.

Appendix B:

Average rates of return on investment in buying and renting a flat [%].



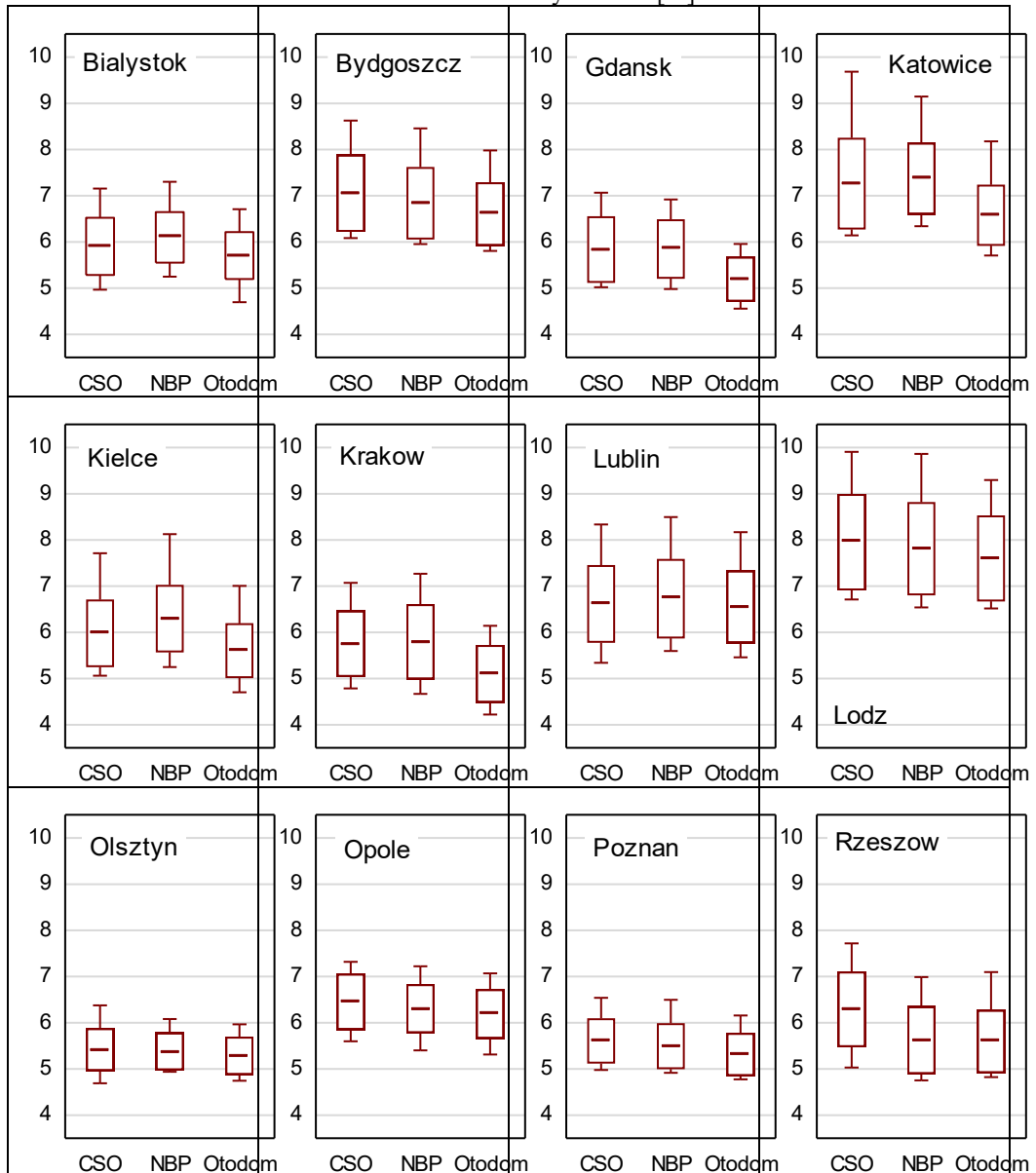


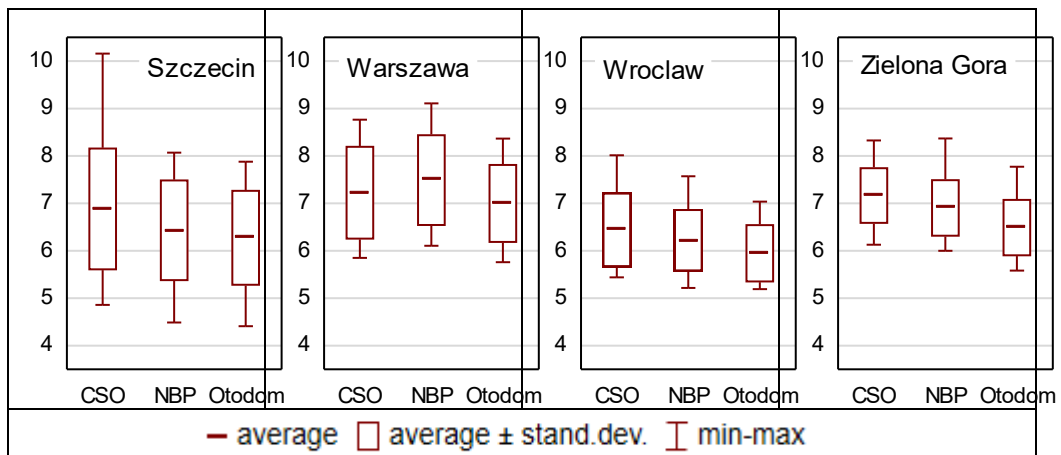


Source: Own study.

Appendix C:

Arithmetic averages, arithmetic averages $\pm$ standard deviations and minimum and maximum values of returns in each city for the quarters 2019-2022 calculated from CSO, NBP and OtodomAnalytics data [%].

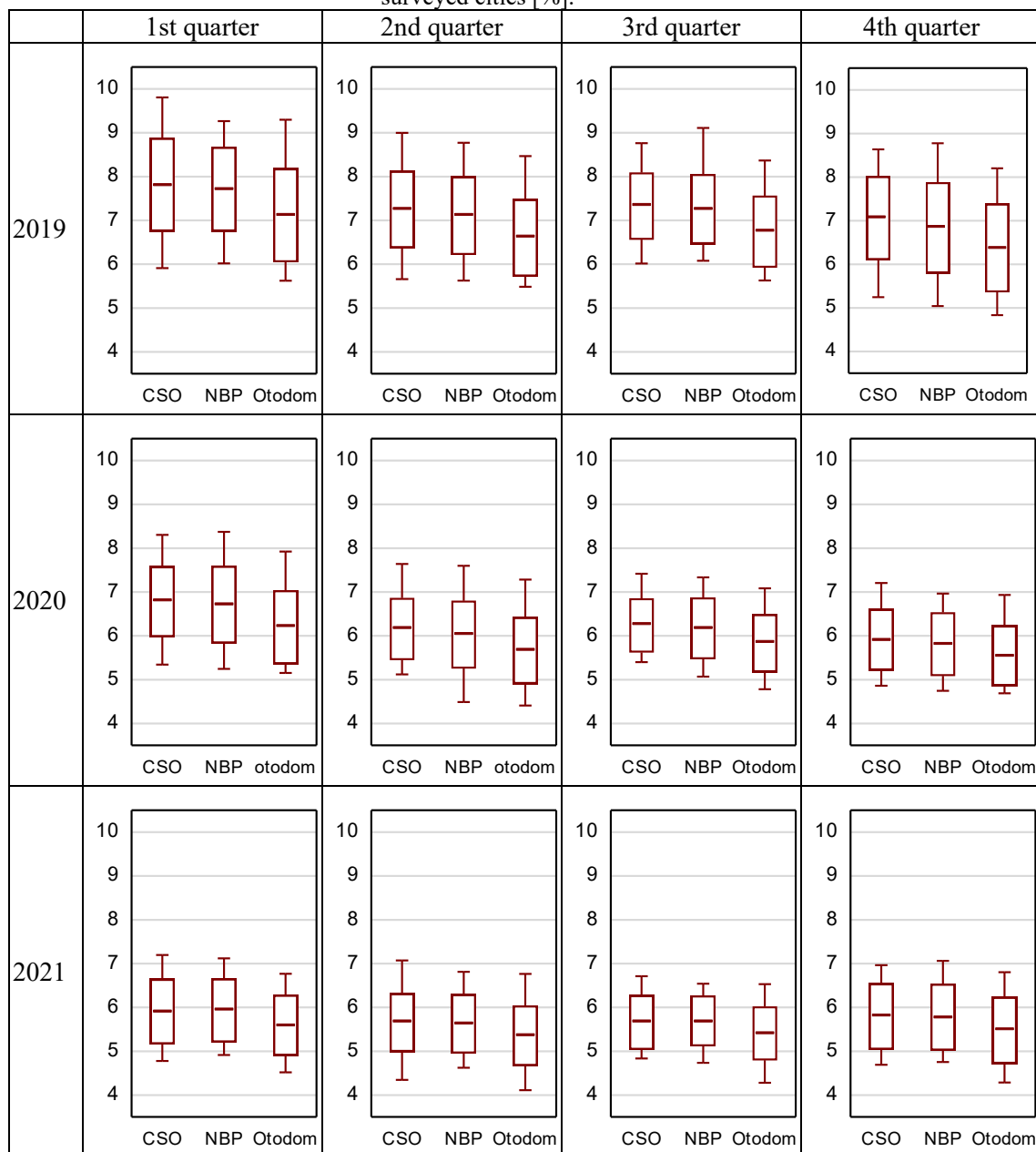


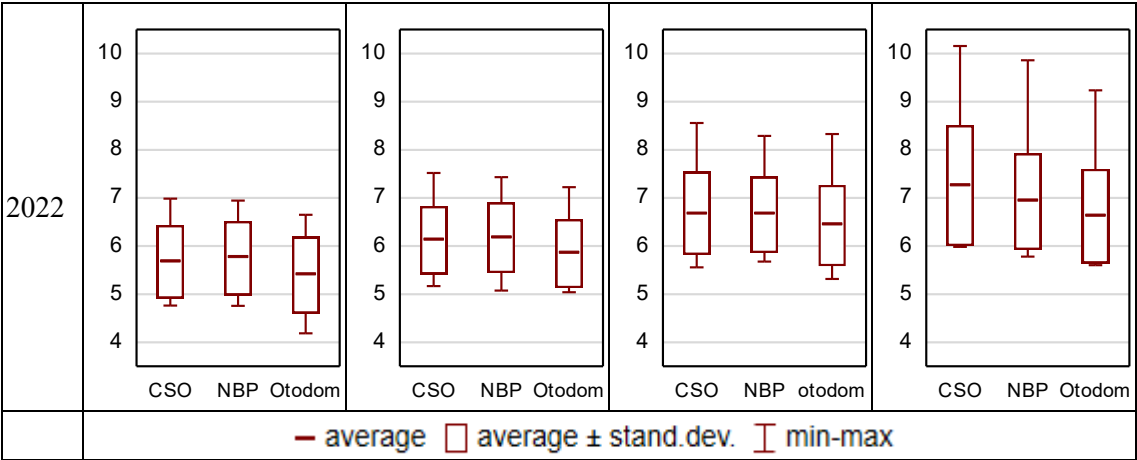


Source: Own study.

Appendix D:

Averages, standard deviations and ranges of returns (%) by quarter for 2019-2022 in the surveyed cities [%].

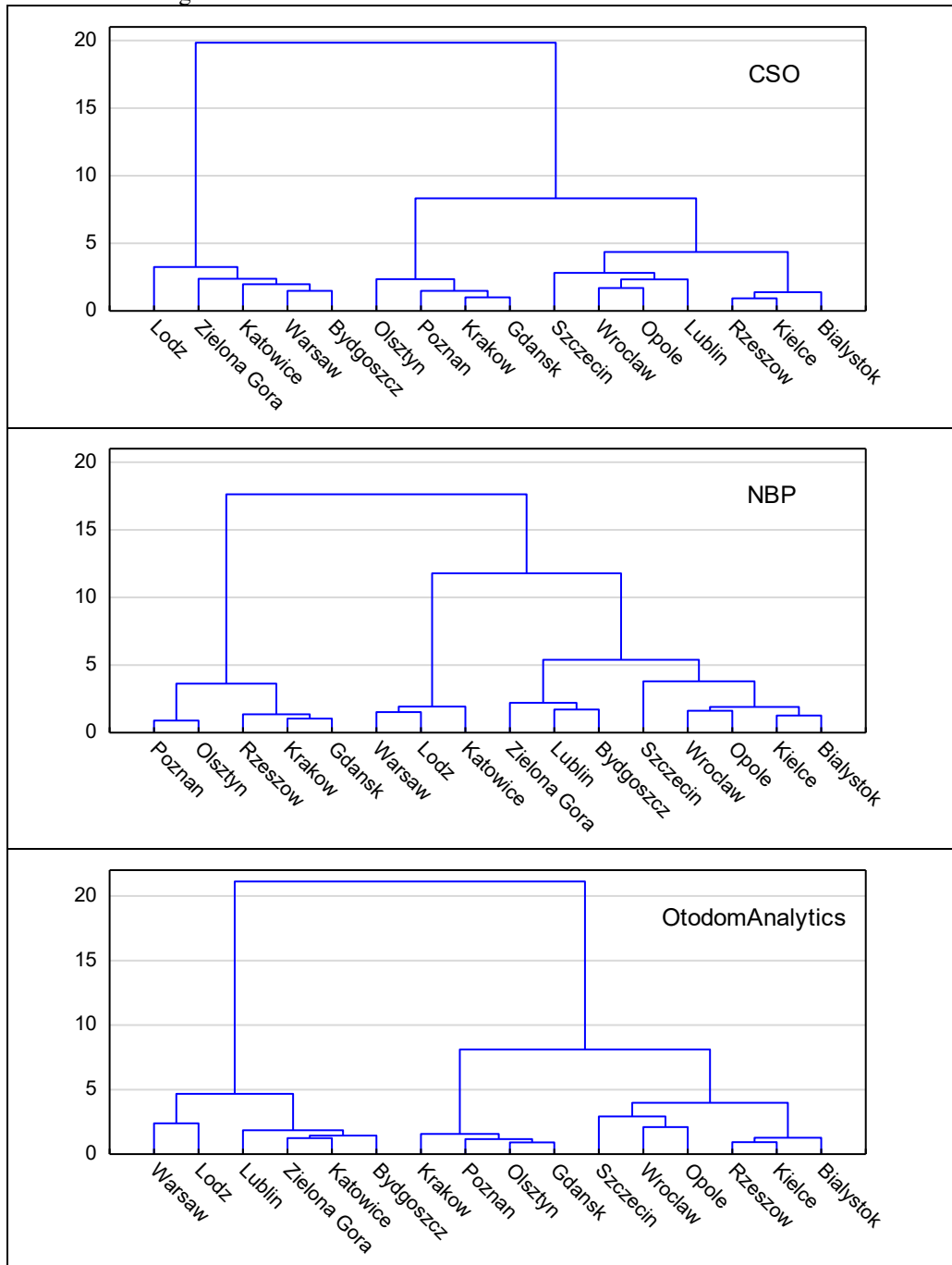




Source: Own study.

Appendix E:

Cluster analysis of returns calculated from CSO, NBP and OtodomAnalytics data using Ward's method.



Source: Own study.