Marek Kulczycki*

# A Survey of Preferences as a Basis for Setting Weights for Property Characteristics**

## 1. Introduction

According to until recently valid Standard III.7 *Wycena nieruchomości przy zastosowaniu podejścia porównawczego*, having presently status of currently interpretational note within Common National Principles of Valuation, weights of market characteristics are set on the basis of analysis of behavior and analysis of preferences of real estate market participants.

Analysis of behavior should be understood as analytical inference on the basis of registered real estate selling transactions. This way would be sufficient if for each and every real estate market limited both in area and in type, the sufficient number of transaction was given, and if these transactions had free-market character (i.e. full trust to price is justified) and finally if on the basis of notary act and field inspection could be possible to assign to each end every property being subject to transaction, numbers describing their market characteristics. Conditions afore mentioned cannot be always fulfilled, is such cases, it is worth using the second method i.e. a survey of preferences of potential buyers. Certainly, method this is not perfect either. Stressing its drawbacks, it is enough to mention about time consuming process of data collection. Also, a choice of respondents is worth discussing (respondent can be more or less a potential buyer). Despite afore mentioned obstacles, it is worth employing this method especially if the analyzed property market is atypical or if the intention is an atypical use of captured knowledge.

Real estate like every other thing can be described with indefinite number of characteristics. All these characteristics on real estate market analysis purposes can be divided into three parts:

1) group of legal characteristics – bundle of rights linked with real estate being analyzed and particular conditions and limitations in benefiting from these rights,

* Department of Geomatics, AGH University of Science and Technology, Krakow
** Self scientific research

2) group of development characteristics – characteristics linked with real and planned development including spatial development planning,
3) remaining physical and non-physical characteristics.

Taking a different criterion of division, in each group of characteristics aforementioned next two groups can be distinguished:

1) group of characteristics defining the market,
2) group of market characteristics – characteristics, which have a dominant influence on decisions made by market participants, by what they form (differentiate) the prices.

Defining the property market, choice of market characteristics and assigning weights to market characteristics are appraiser's activities which are repeated every time when new valuation file is needed. Repeating the same activities very often leads to a fatal routine, especially when the same set of market characteristics is applied to every task being solved and to every real estate market. Set of market characteristics selected properly for an apartment's market for the entire city cannot be applied to prediction of property value for a narrowed market like for example a single residential complex. The fact is, that all the apartments of the residential complex have the same location, surroundings or state and type of built-up, but it does not mean that all the apartments have the same value. The apartments of this residential complex are more or less attractive for potential buyer, thus have different values. The values will be precisely estimated only by this appraiser who knows the preferences of potential buyers.

The need of knowledge of buyers' preferences is equally high among other real estate market participants, e.g. real estate agents and developers. Developer, on the basis of this knowledge, can differentiate asking prices for newly built apartments and sell the investment in the most effective way. Real estate agents need only a couple of phrases to guess the liking of their customers.

The knowledge of buyers' preferences is getting more and more important in Poland. Some time ago, a huge excess of demand over supply caused, that potential buyer was forced to get rid of majority of preferences and buy this, what has not been bought yet. In very many cases, the purchase concerned a flat in not-existing-yet building but only in project. Presently, the rate of demand and supply is changing with benefit for buyers. In predictable future, we can meet situation that flats already built will be waiting for buyers empty and developers before starting every new investment will perform much more deep analysis of profitability, the analysis consisting of among others analysis of preferences of future customer. The buyer will be able to do a purchase according to his or her preferences.

## 2. The Ranking Method

One of the simplest ways of "measuring" customers' preferences is a survey, in which customers express their opinions by answering suitably formulated questions in an inquiry form.

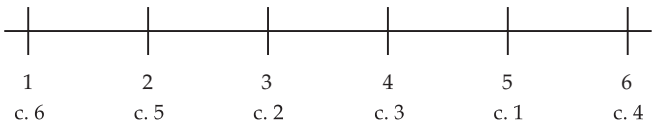
The inquiry form is a fundamental measuring tool. Each and every inquiry form should be prepared in a way to make answering questions easy for a respondent and simultaneously not to suggest answers. The questions should be understandable for everyone.

Striving for recognizing the preferences of the property purchaser, the most obvious question seems to be a question relying on arranging in order the characteristics of the property from the most to the least relevant, this is:

*Think a while and answer the question. What would have an importance to you while making a decision on buying an apartment. In an empty cells write down digits from 1 to 6, in such a way that digit 6 describes characteristic which is the most significant and digit 1 is the least significant.*

| characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 | characteristic 6 |
|---|---|---|---|---|---|
| 5 | 3 | 4 | 6 | 2 | 1 |

Looking at the scale of preferences of this respondent:




we can decipher that:

- the most relevant is characteristic no. 4;
- the least relevant is characteristic no. 6;
- significance of characteristic no. 3 in reference to characteristic 2 is less important than significance of characteristic no. 1 in reference to characteristic no. 5 (but is it really threefold?);
- significance of characteristic no. 5 in reference to characteristic no. 6, characteristic no. 2 in reference to characteristic no. 5, characteristic no. 3 in reference to characteristic 2 and also 1 to 3 and 4 to 1 are identical (but does respondent really think so??).

Despite correctness of afore mentioned findings, they are useless for setting weights for market characteristics because of their absolutely qualitative character.

The situation will change if more than only one respondent is asked afore mentioned question, the more representative sample the better. Questioning many respondents, on the basis of their individual scales of preferences, tabular listing of average answers is possible:

| characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 | characteristic 6 |
|---|---|---|---|---|---|
| 3.94 | 2.79 | 4.52 | 5.42 | 2.31 | 1.44 |

On its basis, construction of a new scale of preferences is possible:


from which we can find additionally that:

- in proportion to entire sample a single respondent showed the same preferences considering the importance of analyzed market characteristics while making a decision on buying a flat (exception characteristics no. 1 and no. 3, which changed the order);
- importance of neighboring characteristics no. 3 and characteristic no. 1 on a scale of preferences is much closer in eyes of respondents than (also neighboring) characteristics 1 and 2 (in this case it is almost relation 1:2.5).

On the basis of this scale of preferences it seems to be possible to set weights for market characteristics. Taking some assumptions, namely:

- sum of all weights expressed in percents equals 100%,
- relation of respective characteristics stays unchanged against relation between characteristics expressed through scale of preferences,

we obtain on the basis of following formula

$$k_i\,\% = \frac{p_i}{\Sigma} \tag{1}$$

where:

$p_i$ – position of i$^{th}$ characteristic on the scale of preferences,
$\Sigma$ – sum of values from scale of preferences calculated for all characteristics,

the following weights for characteristics:

| characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 | characteristic 6 |
|---|---|---|---|---|---|
| 19% | 14% | 22% | 27% | 11% | 7% |

We can mak next assumption: the least attractive characteristic according to scale of preferences takes the weight equal to $k_{min}$%.

After applying formula

$$\Delta_{0\rightarrow} = \frac{k_{\min}\%}{100\% - n \cdot k_{\min}\%} \cdot \Sigma_0 \tag{2}$$

or

$$\Delta_{\min\rightarrow} = \frac{k_{\min}\%}{100\% - n \cdot k_{\min}\%} \cdot (\Sigma - n \cdot p_{\min}) \tag{3}$$

where:

$k_{\min}$% – expected value of the weight for, according to respondents, the least attractive characteristic,

$n$ – number of characteristics,

$p_{\min}$ – minimal value from the scale of preferences (value of preference for the least attractive characteristic),

$\Sigma_0$ – sum of values from scale of preferences calculated for all the characteristics before translation ($p_{\min} = 0$),

$\Sigma$ – current sum of values from scale of preferences,

$\Delta_{0\rightarrow}$ – value of translation of scale of preferences (translation against 0), to get weight $X$% for the least attractive characteristic,

$\Delta_{\min\rightarrow}$ – value of translation of scale of preferences (translation against minimal value on the scale of preferences), to get weight $X$% for the least attractive characteristic,

we obtain rescaled weights for all characteristics:

| $X$% | $\Delta_{0\rightarrow}$ | characteristic 4 | characteristic 3 | characteristic 4 | characteristic 2 | characteristic 5 | characteristic 6 |
|---|---|---|---|---|---|---|---|
| 0% | 0.0000 | 33.8% | 26.1% | 21.2% | 11.5% | 7.4% | 0.0% |
| 4% | 0.6200 | 29.7% | 23.9% | 20.1% | 12.7% | 9.6% | 4.0% |
| 8% | 1.8123 | 25.6% | 21.6% | 19.0% | 14.0% | 11.8% | 8.0% |
| 12% | 5.0486 | 21.5% | 19.3% | 17.9% | 15.2% | 14.1% | 12.0% |
| 16% | 47.1200 | 17.4% | 17.0% | 16.8% | 16.5% | 16.3% | 16.0% |

Assessment of $k_{\min}$% may be dependent on number of respondents, who indicated the least attractive characteristic as still not the worst, but it is worth noting,

that assigning to the least attractive characteristic more and more weight causes "flattening" for the remaining characteristic i.e. fuzziness of preferences, what is depicted on the following chart.

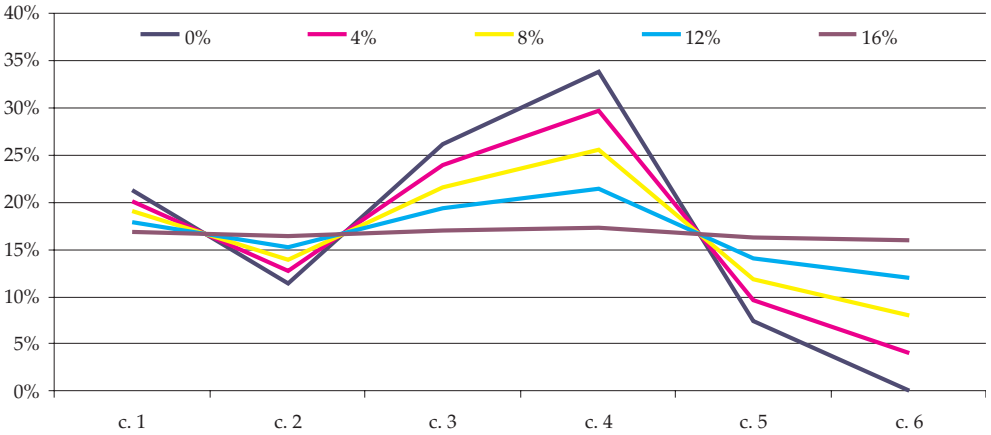



**Fig. 1.**

## The Thurstone's method

According to researches, a man can without any problems express an opinion on which of two products is better or which of two features of the product is more important. For majority of respondents, simultaneous arranging three products from the best to the worst will not be a problem either. The problem starts when more products or very complex products are subject to arranging and assessment of their attractiveness depends on many factors. Real estate or its single features can be for sure an example of very high complexity product. Thus, applying the aforementioned ranking method carries a danger that respondent who is not very concentrated, will give the answer which is not definitely reflecting individual preferences.

Therefore, the opinion that the only one just method to apply is Thurstone method (because of capability of getting much more precise results) is very often met in literature. Under this name, both the way of leading surveys (particularly the way of formulating questions) and the computational algorithm itself are hidden.

An inquiry form for this method (called also pairwise comparison method) consists of many questions. Each question is a composition of 2 real estate characteristics. Respondent is asked for an opinion on which property characteristic has more significance while making decision on buying an apartment, e.g.:

*Imagine that you are making a decision about purchasing one of two apartments. Both apartments have the same price. One apartment differs from another with only two characteristics – district and standard. The first apartment is located in the least attractive district but can be characterized by high standard. The second is located in a very attractive district but the standard is low. Which of these two apartments do you choose?*

| I am buying the first apartment | I am buying the second apartment |
|---|---|
| □ | □ |

A great virtue of such constructed inquiry form is, that (of course if respondent has both enough time and patience to answer all questions) we can mix questions and even use control questions (i.e. ask about the same pair of characteristics twice changing for example the order), by what we eliminate respondent's indecision and closer we are to precise knowledge of respondent's preferences. Introduction of weighting scheme for answers is also possible, for example by using 5 degree Likert's scale, thus answering identical question like this mentioned above the respondent marks one of five possible answers:

| definitely first | first | no opinion | second | definitely second |
|---|---|---|---|---|
| □ | □ | □ | □ | □ |

where answer *definitely first* or *definitely second* gets weight equal to 1.00, answer *first* or *second* gets weight for example 0.75, whilst answer *no opinion* as neutral gets weight equal to 0.00, or:

| first | rather first | no opinon | rather second | second |
|---|---|---|---|---|
| □ | □ | □ | □ | □ |

where answer *first* or *second* gets weigt equal to 1.00, answer *rather first* or *rather second* gets weight for example 0.25. Answer *no opinion* similarly to the aforementioned gets weight equal to 0.00.

Despite many virtues, an inquiry form prepared on the purpose of this survey has one essential drawback – filling out the entire inquiry form by one respondent requires a great deal of time.

The first step of calculation in Thurstone's method relies on pairwise comparison of analyzed market characteristics, i.e. listing of uniform characteristics' ranking, called also 1 – dimensional scale of preferences.

Let number $P(X_i, X_j)$ denotes respondent's ratio, for who characteristic $X_i$ is more important than characteristic $X_j$, thus

$$P(X_i, Y_j) = \frac{N_{Xi}}{N} \tag{4}$$

where:

$N_{Xi}$ – number of respondents preferring characteristic $X_i$ to characteristic $X_j$,

$N$ – total number of respondents,

$P(X_i, X_j)$ – empirical coefficient of preferences of pair of characteristics.

Coefficients $P(X_i, X_j)$ form an empirical matrix of preferences. On the main diagonal there are empty entries or assigned values of 0.5, what has no influence on final value. The values on both sides of the main diagonal sum up to unity.

In the second stage of calculations coefficients of preferences are transformed. We standardize obtained empirical results, using normal distribution tables. For each and every empirical value we search for theoretical value, which is positive if empirical value was greater than 0.5 and negative if it was lesser than 0.5. To fulfill the task we use cumulative normal distribution function $N(0,1)$

$$T(X_i, X_j) = F^{-1} P((X_i, X_j)) \tag{5}$$

where:

$F$ – cumulative normal distribution $N(0,1)$,

$T(X_i, X_j)$ – theoretical coefficient of preferences of pair of characteristics.

Obtained results are listed in a form of table of theoretical distribution of standardized variable. Coefficients $T(X_i, X_j)$ form a theoretical matrix of preferences.

In the third (last) stage, values of standardized variables (theoretical coefficients) are summed by rows (a sum for each characteristic)

$$T_i(X) = \sum_j T(X_i, X_j) \tag{6}$$

Numbers $T_i(X)$ are then normalized, i.e. divided by number of characteristics and enlarged or reduced by a constant number in such a way that, the worst characteristic gets value of "zero", thus

$$p_i = \frac{T_i(X)}{n} \pm \Delta \tag{7}$$

where:

$n$ – number of characteristics,

$\Delta$ – a translation on a scale of preferences necessary to get value "zero" for the least attractive characteristic.

On a basis of such constructed scale of preferences we can calculate the weights for market characteristics, using formula (1) and formulas (2) or (3).

Verification of the results is a very important issue while carrying on a survey and performing calculations. In some cases, verification can rely on a reconstruction of input data by means of obtained results. The lesser the differences between real data and results obtained from applied procedures the more valuable the results of a survey.

The Thurstone's method is one of the methods which enable such verification. Value of cumulative normal distribution $N(0,1)$ for difference $p_i - p_j$ enables to estimate the proportion of respondents, for who the characteristic $X_i$ is more important than characteristic $X_j$ while making a decision on buying a real estate. Verification is carried out by means of formula

$$\delta = \left| F(p_i - p_j) - P(X_i, X_j) \right| \tag{8}$$

Application of Thurstone's method requires unfortunately fulfilling some assumptions resulting from the properties of the function $F^{-1}$.

First, the method cannot be applied in case, when one of the characteristics is better or worse in proportion to another characteristic in the eyes of all respondents. We have then a situation, in which coefficient of preferences of pair of characteristics $P(X_i, X_j)$ takes value of 1 – value for which the function $F^{-1}$ is indefinite. The problem aforementioned can be limited by appropriate choice of market characteristics and by choosing large sample of respondents.

Another, very important limitation of applying Thurstone's method is the fact, that function $F^{-1}$ increases very rapidly, when its arguments tend to 1. From this reason, in case when values $P(X_i, X_j)$ are close to 1, obtained values of $p_i$ are overestimated.

Elimination of aforementioned limitations relies on replacing function $F^{-1}$ by another function. The function which increases slower and is definite on the interval [0;1]. These conditions are satisfied among others by two functions:

$$T(X_i, X_j) = \sin(P(X_i, X_j) - 0.5) \tag{9}$$

and

$$T(X_i, X_j) = \operatorname{tgh}(P(X_i, X_j) - 0.5) \tag{10}$$

Application of both functions leads to reasonable construction of scale of preferences. Obtained results are convergent to results obtained from function $F^{-1}$, and verification of results is carried out analogically to the manner aforementioned with a difference, that in a formula (8) cumulative normal distribution function $N(0,1)$ is replaced respectively by arcsin or arctgh function:

$$\delta = \left|(\arcsin(p_i - p_j) + 0.5) - P(X_i, X_j)\right| \tag{11}$$

or

$$\delta = \left|(\operatorname{arctgh}(p_i - p_j) + 0.5) - P(X_i, X_j)\right| \tag{12}$$

**AN EXAMPLE**

The following example presents an algorithm for Thurstone's method.

**Stage 1.** Response matrix created on the basis of inquiry form (independently for every single respondent).

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | – | 0 | 1 | 0 | 0 |
| characteristic 2 | 1 | – | 1 | 0 | 0 |
| characteristic 3 | 0 | 0 | – | 0 | 0 |
| characteristic 4 | 1 | 1 | 1 | – | 0 |
| characteristic 5 | 1 | 1 | 1 | 1 | – |

A response matrix with number of columns and rows equal to number of analyzed characteristics. Digit 1 on the intersection of row 1 and column 3 means, that for a particular respondent characteristic 1 has more significance than characteristic 3 while decision making process concerning a real estate purchase.

**Stage 2.** Tabular listing of results from all inquiry forms.

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | 0 | 142 | 154 | 91 | 133 |
| characteristic 2 | 45 | 0 | 113 | 27 | 84 |
| characteristic 3 | 33 | 74 | 0 | 23 | 73 |
| characteristic 4 | 96 | 160 | 164 | 0 | 133 |
| characteristic 5 | 54 | 103 | 114 | 54 | 0 |

Number 23 on the intersection of row 3 and column 4 contains total number of respondents who indicated a characteristic 3 as more important than characteristic 4. There are zero entries on the main diagonal and on the both sides of main diagonal the entries sum up to total number of respondents.

**Stage 3.** Transformation of the response matrix into empirical matrix of preferences.

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | 0 | 0.76 | 0.82 | 0.49 | 0.71 |
| characteristic 2 | 0.24 | 0 | 0.60 | 0.14 | 0.45 |
| characteristic 3 | 0.18 | 0.40 | 0 | 0.12 | 0.39 |
| characteristic 4 | 0.51 | 0.86 | 0.88 | 0 | 0.71 |
| characteristic 5 | 0.29 | 0.55 | 0.61 | 0.29 | 0 |

Empirical matrix of preferences results from applying formula (4) to every single entry of response matrix.

**Stage 4.** Transformation of preference matrix from empirical to theoretical form.

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | – | 0.70 | 0.93 | –0.03 | 0.56 |
| characteristic 2 | –0.70 | – | 0.26 | –1.06 | –0.13 |
| characteristic 3 | –0.93 | –0.26 | – | –1.16 | –0.28 |
| characteristic 4 | 0.03 | 1.06 | 1.16 | – | 0.56 |
| characteristic 5 | –0.56 | 0.13 | 0.28 | –0.56 | – |

The transformation is carried out with application formula (5) (or in particular cases with one of the formulas (9) or (10)) to empirical values from response matrix.

**Stage 5.** Construction of scale of preferences and computation of market characteristics' weights on its basis.

| | $T_i(X)/n$ | $p_i$ ($k_{\min} = 0\%$) | $k\%$ | $pi$ ($k_{\min} = 5\%$) | $k\%$ | $pi$ ($k_{\min} = 10\%$) | $k\%$ |
|---|---|---|---|---|---|---|---|
| characteristic 1 | 0.54 | 1.20 | 36% | 1.4164 | 32% | 1.8551 | 28% |
| characteristic 2 | –0.41 | 0.25 | 8% | 0.4702 | 11% | 0.9088 | 14% |
| characteristic 3 | –0.66 | 0.00 | 0% | 0.2193 | 5% | 0.6580 | 10% |
| characteristic 4 | 0.70 | 1.36 | 41% | 1.5801 | 36% | 2.0188 | 31% |
| characteristic 5 | –0.18 | 0.48 | 15% | 0.7003 | 16% | 1.1389 | 17% |

The construction of scale of preferences is carried out on the basis of formulas (6) and (7). Calculating weights for market characteristics requires applying formula (1) with taking into account formulas (2) and (3).

**Stage 6.** Verification of results.

Verification of results relies on reconstructing input data (empirical $P(X_i, X_j)$) by means of obtained results (difference of characteristics' entries on the scale of preferences ($p_i - p_j$)), with an appropriate application of formulas (8), (11), (12).

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | – | 0.95 | 1.20 | –0.16 | 0.72 |
| characteristic 2 | –0.95 | – | 0.25 | –1.11 | –0.23 |
| characteristic 3 | –1.20 | –0.25 | – | –1.36 | –0.48 |
| characteristic 4 | 0.16 | 1.11 | 1.36 | – | 0.88 |
| characteristic 5 | –0.72 | 0.23 | 0.48 | –0.88 | – |

Matrix of differences on the scale of preferences with using formulas (6) and (5).

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | – | 0.83 | 0.88 | 0.43 | 0.76 |
| characteristic 2 | 0.17 | – | 0.60 | 0.13 | 0.41 |
| characteristic 3 | 0.12 | 0.40 | – | 0.09 | 0.32 |
| characteristic 4 | 0.57 | 0.87 | 0.91 | – | 0.81 |
| characteristic 5 | 0.24 | 0.59 | 0.68 | 0.19 | – |

Reconstructed data with using formula (8), above.

| | characteristic 1 | characteristic 2 | characteristic 3 | characteristic 4 | characteristic 5 |
|---|---|---|---|---|---|
| characteristic 1 | – | –0.07 | –0.06 | 0.05 | –0.05 |
| characteristic 2 | 0.07 | – | 0.01 | 0.01 | 0.04 |
| characteristic 3 | 0.06 | –0.01 | – | 0.04 | 0.08 |
| characteristic 4 | –0.05 | –0.01 | –0.04 | – | –0.10 |
| characteristic 5 | 0.05 | –0.04 | –0.08 | 0.10 | – |

Matrix of differences by reconstructing the data with using formula (8). The average difference of reconstruction of the data is 5.00%.

## References

[1] Barbera J., Niemczyk A.: *Zachowanie konsumentów na runku.* Wydawnictwo AE, Kraków 1997.
[2] Churchill G.A.: *Badania marketingowe. Podstawy metodologiczne.* Wydawnictwo Naukowe PWN, Warszawa 2002.
[3] Czaja J.: *Metody i systemy określania wartości nieruchomości.* Wydawnictwa AGH, Kraków 1999.
[4] Dzudzewicz M.: *Modyfikacje metody Thurstone'a i ich zastosowanie w modelowaniu preferencji.* Poznań 2003.
[5] Kędzior Z., Karcz K.: *Badania marketingowe w praktyce.* Wydawnictwo PWE, Warszawa 2007.
[6] PFVA, *Krajowe Powszechne Zasady Wyceny.* Warszawa 2008.
[7] Rószkiewicz M.: *Metody ilościowe w badaniach marketingowych.* Wydawnictwo Naukowe PWN, Warszawa 2002.
[8] Sztumski J.: *Wprowadzenie do metod i technik badań społecznych.* Wydawnictwo „Śląsk", Katowice 1999.
[9] Walesiak M.: *Metody analizy danych marketingowych.* Wydawnictwo Naukowe PWN, Warszawa 1996.