Summary: This paper concerns the application of parametric and non-parametric regression methods in explaining the impact of advertising on consumers on the dairy market. Two specific goals were stated in this research. The first goal was to evaluate the impact of advertising components on its effectiveness operationalized as the AIDA marketing model. The second goal attempted at clustering consumers with respect to the effectiveness of subsequent components of advertising. The former was achieved using the parametric regression model, for the latter the regression tree model was applied. Among the components of advertising, the TV medium and music – factors that strongly attract attention – affects the behaviour of consumers of dairy products the most. Other important revealed factors turn out to be varying in significance depending on the regression method applied.

Keywords: regression method, linear regression, regression tree, advertising, dairy market.

DOI: 10.15611/ekt.2014.3.05

1. Introduction

From the marketing perspective, a complete description of the impact of advertising on consumers buying behaviour is an extremely difficult task. The reason is high complexity of both decision-making process and determinants of consumers behaviour. A common approach is to analyze a restricted subset of these aspects, with advertising as one of important factors remaining the subject of interest in the consumer-related research. The effectiveness of advertising is understood as the
extent to which advertising achieved the objective set [Pociecha 1996]. Compared to the evaluation of efficiency, defined as the ratio of expenditure on advertising to financial results, the estimation of effectiveness is focused on examining how advertising changes the attitude of a consumer towards a product or a product category. Research on the effectiveness of advertising allows for a broader perspective and a better overview of the condition of advertising in the entire industry. It also answers more general questions concerning the sensitiveness of consumers to advertising, optimal composition of marketing activities for a specific industry and optimization of costs of advertising. Additionally, the effectiveness of advertising has a significant impact on advertising efficiency and provides guidelines for the directions of future marketing activity. Research on the effectiveness of advertising, with its cognitive values, is also often carried out for exploratory purposes in the academic research [Maison 2007].

It is considered very hard to estimate the impact of advertising on consumers behaviour, hence the common approach is to focus on a specific aspect or a group of aspects of advertising. The groups of such components that are most commonly verified are:

- Forms of advertising [Ogilvy, Raphaelson 1982; Keller 1987; Lewiński 1999; Świątkowska 2007];

Based on these research studies, it is possible to create a full set of components that may be relevant when evaluating advertising effectiveness. One common component is TV – a medium with the strongest impact due to its range and dynamic communication based on audio and video features. On the other hand, the sensitivity of the components depends on many other factors, including product category. It is therefore not recommended to underestimate components with potentially low importance as they may reveal their importance when analyzed in a specific target group or localization.

Evaluation measures of the effectiveness of advertising are based on the analysis of different scopes.\(^1\) Although it is possible to consider these measures separately,
the analysis of the impact of advertising on consumers led to the formation of many detailed concepts that expand stages of changes in consumer attitudes to an advertisement. One of such concepts is the AIDA model proposed by Strong [Kotler 2005]. AIDA is a hierarchical model that assumes a cause-and-effect relationship between succeeding stages of consumer engagement with an advertisement and a decreasing number of the recipients of an advertisement for subsequent stages. An advertisement is presented to the highest number of consumers and catches their attention (A), the lower number of consumers show interest (I) and demonstrate desire (D), but only few of them – 5 to 15%, depending on the product category – take action (A), i.e. buy a product.

In due course of the research on the impact of advertising, the AIDA model was used as a base for various modification [Terelak 1998; Rudnicki 2004]. It is worth noting though that all of them share a common – derived from AIDA – scheme of sequential processing of information consisting of conscious perception of advertising, information processing and reaction, also known as “think-feel-do”.

The perception stage consists of the cognitive sphere, during which the consumer perceives an advertising message, directs his or her attention towards it and comprehends information contained (in various models known as: attention, definition, awareness, stay, perception, comprehension or interest). During the information processing stage the emotional component is strongly manifested when a consumer expresses his or her preference or confidence towards a product and his or her desire to own it while confronting content of an advertising message with his or her knowledge and conviction (look, desire, conviction, confidence, acceptance). The reaction stage completes the process of an direct impact on a consumer and consists of components related to a decision (volitional) and – in the case of purchase – also to an actual purchase and an after-purchase reaction (intention, stimulation, action, buy, satisfaction). The scheme described, with the AIDA model as the best-known instance, is widely used for an empirical verification of the effectiveness of advertising [Pociecha 1996; Łodziana-Grabowska 1996]; however, in most cases it has only a descriptive character. It seems legitimate therefore to make an attempt to include the AIDA scheme during statistical modelling.

Numerous studies [e.g. Maison 1998; Cline, Kellaris 1999; Jachnis 2007] show that a reaction to advertising in various product categories clearly differs between

---

2 Among others the following models were proposed: AICA (attention-interest-conviction-action), AICCA (attention-interest-comprehension-confidence-action), AIDCAS (attention-interest-desire-conviction-action-satisfaction), model DIPADA (definition-identification-proof-acceptance-desire-action), DAGMAR – ACCA Colley’s (awareness-comprehension-conviction-action), AAPIS (awareness-acceptance-preference-intention-stimulation) and EPCCA (exhibition-perception-comprehension-conviction-action).
groups of consumers. Therefore, an advertising message should take this aspect into account. A similar relationship is relevant to different reactions to advertising in groups of consumers that differ in the level and specificity of consumption. According to prior research [Chudzian 2013], consumers of dairy products are evidently grouped according to the frequency of dairy products consumption with the dissimilarity level suggesting a separate analysis of their reaction to advertising. In that context, the statistical modelling of the effectiveness of advertising should include all three presented concepts: the hierarchical nature, the diversity of the AIDA model and the variety of reactions to advertising.

2. Research methodology

The measurement of the effectiveness of advertising is highly complicated task due to the amount of factors influencing the process. Therefore, the task of measuring effectiveness is usually divided into smaller tasks concentrated on selected aspects and factors influencing effectiveness. Hence in this study two specific research goals were stated: to evaluate the impact of individual factors on the AIDA model value and to cluster consumers with respect to the effectiveness of subsequent components of advertising. The accomplishment of both research goals using a single method proves to be difficult; therefore, two methods were used: the classical multiple regression and the non-parametric regression.

The survey for this study was conducted using CAWI methodology. It was preceded by pilot surveys (6 trials on the total number of 650 interviewees), which allowed for a systematic construction and improvement of measurement methodology. The first batch of pilot surveys aimed at the exploration of the research area (390 interviewees) and the second was used to iteratively evaluate and improve accuracy and reliability of questionnaire (260 interviewees).

Sample selection for the final survey of the study was a two-stage process: initially the filtering criteria were specified and then the filtered population was randomly sampled as a second step. Due to the main goal of the study strictly related to consumer behaviour on the single market, the survey covered only people who declare consumption of dairy products. Due to the requirement of a wide range of dairy products and a wide range of advertising types, the research comprised of localization filtering. Nine biggest cities of Poland were selected with the agglomeration within the 30 km radius. According to the Polish Central Statistical Office data, which take into account geographical location and are based on the experience of research agencies, the research included the following cities: Warszawa, Poznań, Kraków, Tricity, Katowice, Białystok, Łódź, Lublin and Wrocław. Because of the high fragmentation of dairy production, such a selection allowed for including dairy product brands with substantial local importance.

In order to minimize the bias of random errors on the results of the statistical analysis, the survey was designed to cover at least 500 interviewees. During the
survey 630 questionnaires were obtained from dairy products consumers aged 15 or more who live in one of nine agglomerations. During initial data cleansing a share of samples was removed due to ambiguity, uncertainty and inconsistency of questionnaire answers. Finally, 550 questionnaires were approved for the analysis.

The analysis included an attempt to operationalize the AIDA model presented in the theoretical part of this paper and use it as a basis for the linear regression model and the regression tree model. The former allowed for examining which components affect the effectiveness of advertising the most. The latter – the non-parametric regression model – was used to detect segments of consumers with respect to their reaction to different components of advertising. The SPSS and R statistical software were used to conduct and evaluate the analysis and interpret the results.

3. Variables selection

In order to determine the impact of advertising on dairy products purchase, parametric and non-parametric regression methods were used. This allowed for evaluating the effectiveness of advertising by the optimization of the mixture of factors, media, goals and advertising forms.

For both models the effectiveness of advertising was specified as a dependent variable, constructed as a weighted sum of the strength of each stage derived from the theoretical AIDA model. It is worth noting that the frequency with which the AIDA model stages are observed decreases with successive stages – the most frequently observed stage is related to watching advertising, the least to the purchase under the influence of advertising. Simultaneously, the effectiveness of advertising increases within the subsequent stages – it is the highest when the process completes with a purchase and the lowest when it completes with only watching an advertisement. Since these effects influence the measurement of effectiveness, the weights for each stage forming the dependent variable were introduced in order to reduce these discrepancies. The detailed procedure for forming the AIDA variable is presented in Table 1. In order to allow for an easier interpretation of the AIDA variable, each forming variable was scaled to <0,1> range; hence, the values of AIDA also fall into this range.

In the population examined during the survey, the mean value of AIDA variables was 0.32, which could be either due to the highest weights for the least frequent stages (desire and action) or to low values of all the indicators forming the AIDA variable. It is also worth noting that the AIDA values above 0.5 occur among only 25% of the surveyed population. The mean AIDA was very close to the median (0.30). The AIDA variable tended to be highly dispersed – the standard deviation amounted to around 75% of the mean value.

---

3 Attention (A) was operationalized as a declaration of watching dairy products advertising, interest (I) – as a declaration that advertising generates interest in a product, desire (D) – as a positive attitude towards advertising, and action (A) – as a declaration of purchase under the influence of advertising.
Table 1. AIDA variable construction

<table>
<thead>
<tr>
<th>AIDA stage</th>
<th>Operationalization</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (attention)</td>
<td>I often watch advertisements of dairy products.</td>
<td>0.1</td>
</tr>
<tr>
<td>I (interest)</td>
<td>I’m interested in dairy products advertisements.</td>
<td>0.2</td>
</tr>
<tr>
<td>D (desire)</td>
<td>I like watching dairy products advertisements</td>
<td>0.3</td>
</tr>
<tr>
<td>A (action)</td>
<td>I happen to buy dairy products under the influence of advertisements.</td>
<td>0.4</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

Source: own research.

The distribution of the AIDA variable is presented on the histogram in Figure 1. It displays a high frequency of the lowest values and a very low frequency of the highest (above 0.8) values, which explains a relatively low mean value of the AIDA variable.

Based on the literature survey, the following factors were chosen that may potentially explain the degree of the effectiveness of advertising. These factors, forming a list of independent variables for regression models, were divided into four groups:

• $x_1$–$x_8$ – advertising objectives (information, demonstration – product application, persuasion, attention – attracting consumer attention, offer, acknowledgements – gratefulness for the previous purchase, recall, competitiveness – showing a differences between competing products);
• x9–x16 – forms of advertising (banners, animation, role-play from everyday life, role-play in a kitchen, role-play during a meal, outdoor role-play, production of a product, demonstration of product application);
• x17–x26 – advertising elements (music, humour, content, graphics, idea, realization, attractiveness, character, length, show time);
• x27–x36 – advertising media (TV, radio, press, Internet, outdoor, vehicles, packaging, point of sale, cinema, e-mail).

All independent variables took value from the <1;5> range with 1 standing for low intensity of an indicator and 5 for high intensity.

Type of the relationship between the AIDA variable and independent variables was examined as the next step. During this analysis some cases revealed a relationship of multinomial nature; therefore, the set of independent variables was extended with binomials of selected variables.

4. Results

4.1. Parametric regression model

Due to the nature of the study, the bidirectional stepwise regression was used to build a multiple regression model. As a result, the only independent variables that remained in the final model were those statistically significant (having a significant impact on the AIDA model). Due to the fact that it is not possible to explain the variability of the AIDA indicator solely using advertising factors, the study did not aim at building a model describing the whole variability of the AIDA but rather at evaluating effectiveness of various elements of advertising. Hence, it is reasonable to use bidirectional elimination to assess this effectiveness. After estimation using OLS, the linear regression model for dairy products consumer took the following form:

\[ AIDA = 0.051 * TV + 0.007 * MUSIC + 0.005 * CONTENT + 0.024 * BANNER + + 0.003 * OFFER + 0.022 * ANIMATION - 0.158 \]

All the factors included in the model have positive coefficients, which means that the increase in their values implies the increase in the effectiveness of advertising, while the level of the increase varies considerably. The standardized parameters were also determined in order to verify which of the selected factors have the greatest impact on the AIDA variable. The estimation results for both actual and standardized parameters are presented in Table 2.

The analysis of standardized regression coefficients Beta shows that the most important factors in the model are television media and music elements. Subsequently the effectiveness is affected by the amount of content of advertisement. The set of factors of slightly lesser importance consists of form-related elements (animation and banner) and the actual offer. In addition, each of the variables has a significant impact on the AIDA variable, as confirmed by high absolute values of \( t \)-statistic.
Table 2. Selected summary of the created model

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized coefficient</th>
<th>Standardized coefficients</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.158</td>
<td>0.042</td>
<td>-3.768</td>
<td>0.000</td>
</tr>
<tr>
<td>form ANIMATION</td>
<td>0.022</td>
<td>0.010</td>
<td>2.279</td>
<td>0.023</td>
</tr>
<tr>
<td>form BANNER</td>
<td>0.024</td>
<td>0.009</td>
<td>2.696</td>
<td>0.007</td>
</tr>
<tr>
<td>media TV</td>
<td>0.051</td>
<td>0.008</td>
<td>6.734</td>
<td>0.000</td>
</tr>
<tr>
<td>component CONTENT²</td>
<td>0.005</td>
<td>0.001</td>
<td>3.792</td>
<td>0.000</td>
</tr>
<tr>
<td>component MUSIC²</td>
<td>0.007</td>
<td>0.001</td>
<td>5.937</td>
<td>0.000</td>
</tr>
<tr>
<td>objective OFFER²</td>
<td>0.003</td>
<td>0.001</td>
<td>2.652</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Source: own research.

The resulting model accounts for 30% of the AIDA variable variance (coefficient of determination $R^2 = 0.300$). Due to the relatively large number of variables in the model, the coefficient of determination is biased; therefore, it is worth reporting also unbiased adjusted coefficient $R^2$, which takes the value of 0.293. It is also worth noting that relatively low values of both coefficients of determination stem from the fact that advertising is one of many factors influencing a purchase. In addition, the analysis of the values of both coefficients confirms that it is justified to apply the model to the whole population from which the analysed sample was drawn.

An additional analysis of variance was conducted to assess the quality of the model and the results are presented in Table 3. The high values of Snedecor’s $F$-test ($F_{6,543} = 38.857$) suggest rejection of the null hypothesis that standardized coefficients are equal to zero at level 2.2e-16, which proves the accuracy of the resulting model.

Table 3. The analysis of the variance table for regression

<table>
<thead>
<tr>
<th></th>
<th>Sum of squares</th>
<th>DF</th>
<th>Mean squares</th>
<th>F-test value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>9.605</td>
<td>6</td>
<td>1.601</td>
<td>38.857</td>
<td>2.2e-16</td>
</tr>
<tr>
<td>Residuals</td>
<td>22.370</td>
<td>543</td>
<td>0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31.975</td>
<td>549</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: own research.

---

4 Other factors (psychological, socio-cultural and economic) that may account for the remaining unexplained portion of the variance were not included in the regression analysis.

5 The difference between $R^2$ and $R_s^2$ below 0.01 confirms random selection of the sample and the possibility to generalize the results from the model to the entire population [Field 2005].
Additional diagnostic measurement was taken in order to rule out any systematic error. The normal probability-probability plot of standardized residuals of regression (see Figure 2) does not show alarming departure from 1:1 line. The plot of regression residuals versus fitted (predicted) values (see Figure 3) is disturbed for predicted values lower than 0.5, but it is only due to [0;1] range of values of dependent variable. The plot of Cook’s distance (see Figure 3) does not show an enormous number of outliers that could potentially influence the regression model.

**Figure 2.** Normal p-p plot of regression standardized residual
Source: own research.

**Figure 3.** Regression residuals versus fitted and Cook’s distance
Source: own research.
The analysis of the resulting model shows that an optimal composition of dairy products advertising should contain emotional elements building atmosphere (such as music or animation) and rational elements providing consumers with information about a current offer. Due to a relatively low involvement of consumers in the purchase of dairy products, it seems reasonable to use the elements of music and appropriate advertising media to create a vivid eye-catching message. A message presented in simple form (banners of simple animation) allows recipients to focus on a product and not a story. When attention is attracted, it becomes important to provide consumers with comprehensive information about a product and its qualities. Out of all the available advertising media, consumers of dairy products are the most susceptible to television advertising.

4.2. Non-parametric regression model

The presented linear model accounts for about 30\% of the variance of the AIDA variable and extracts a set of factors influencing the effectiveness of advertising. An additional regression analysis using C&RT\(^6\) [Breiman et al. 1998] was performed in order to verify their importance and identify other potential factors that have not found their way into the regression model due to the nonlinear (but perhaps not functional) relationship with the AIDA variable. The resulting model has the form of a regression tree.

At the beginning of the regression tree construction, all the independent variables from four groups were presented, i.e. advertising objectives, forms, components and media. Subsequent variables that affect the value of AIDA variable were selected using the C&RT algorithm – in the following steps of the algorithm, consumers were divided into groups with similar mean and the smallest possible variance of the AIDA variable. The division is based on the inspection of all independent variables and cut points through its values. Out of all possible combinations the variable and cut points were selected that minimized the variance of the AIDA variable. The tree was constructed as long as the subsequent divisions were resulting in the decrease of AIDA variance exceeding the preselected level. The extracted groups of consumers were characterized by the similar value of AIDA variable and similar opinions on the factors.

The quality of the regression tree is assessed using the mean squared error. For the resulting tree constructed for dairy products consumers, the mean squared error calculated using 10-fold cross-validation was relatively low and amounted to 0.047. The set of factors discriminating the AIDA variable again consisted of music, TV media and advertising in the form of animation. The structure of the regression tree was as follows: the first division of respondents was based on the opinion on

\(^6\) Detailed information about the algorithm can be found in *Classification and Regression Trees* by Breiman et al. [1998].
music elements in advertising. The respondents who definitely considered music an important factor in advertising (response 5) had an average value of the AIDA variable equal to 0.417, significantly higher than the mean for all consumers. Another

Figure 4. Regression tree

Source: own research.
important factor was TV advertising, which divided the respondents at the next level of the tree. The results obtained using the regression tree modelling are presented in Figure 4.

It is worth noting that the lowest impact of advertising on consumer behaviour ($AIDA = 0.111$) was observed in the group of consumers who do not pay special attention to music elements and; in addition, they do not consider TV advertising as particularly relevant for reaching dairy products. Advertising works best for people who definitely pay attention to music and do not reject TV advertising as such ($AIDA = 0.447$). Among these people it is also possible to distinguish those who consider advertising in the form of animation as important ($AIDA = 0.485$) and those for whom it is not relevant ($AIDA = 0.375$).

5. Conclusions

The present study of the effectiveness of advertising on the dairy products market used the AIDA model, widely applied for qualitative description. Due to its simplicity, it is possible to operationalize this model, which was also showed in the paper.

The division of the respondents into homogenous (with respect to the effectiveness of advertising) groups was achieved using the C&RT algorithm. In this non-parametric regression model the following factors were indicated as significantly affecting the AIDA variable: music, TV media and a factor associated with animation.

The analysis of the linear regression model shows that out of all the advertising media influencing the AIDA variable, consumers of dairy products are the most susceptible to TV advertising. The high intensity of the impact of this type of advertising is also confirmed by numerous studies [e.g. Budzyński 1999; Dołhasz 2006]. It allows for highly dynamic presentation and intensively attracts consumers attention, which seems particularly relevant in the case of low-engaging products where a decision process is highly simplified.

In the linear regression model, in addition to TV media, also music elements have a large impact on the effectiveness of advertising. Subsequently, the amount of information contained in an advertisement, form-related elements and actual offer affect the AIDA variable slightly less. At the general level, the results obtained are consistent with the publications of other authors [e.g. Maison 1998]. During the analysis with the AIDA-based approach, it is important to remember that each product category may hold specific structure of the model. Therefore, determining the precise mixture of components is possible only for a specific product category.

The results obtained with both methods are convergent yet do not overlap completely; therefore, it seems that it is justified to use both of them. The comparison of the results from the regression tree and the linear regression model allowed for formulating a thesis that music elements, TV media, and advertising in the form of animation increase the effectiveness of dairy products advertising; hence, they have a significant impact on consumers behaviour. Effective dairy product advertising
should therefore attract the interest of consumers through a creative and expressive large-scale medium.

References

ZASTOSOWANIE PARAMETRYCZNYCH I NIEPARAMETRYCZNYCH METOD REGRESJI W CELU OKREŚLENIA WPŁYWU SKŁADNIKÓW REKLAMY NA ZACHOWANIA KONSUMENTÓW

Streszczenie: Artykuł dotyczy zastosowania parametrycznych i nieparametrycznych metod regresji w celu wyjaśnienia wpływu reklamy na zachowania konsumentów na rynkuмелczarskim. W pracy przyjęto dwa główne cele. Pierwszym z nich była ocena wpływu elementów reklamy na rynku mleczarskim.
reklamy na jej efektywność zoperacjonalizowaną za pomocą marketingowego modelu AIDA. Drugi cel stanowiła próba podziału konsumentów na grupy ze względu na ich podatność na poszczególne komponenty przekazu reklamowego. Pierwszy z celów został osiągnięty poprzez zastosowanie parametrycznego modelowania regresyjnego; do osiągnięcia drugiego wykorzystano nieparametryczny model drzew regresyjnych. Spośród analizowanych składników reklamy najsilniejszym wpływem na zachowania konsumentów na rynku mleczarskim odznaczały się elementy silnie przyciągające uwagę: telewizja jako medium reklamowe oraz muzyka występująca w reklamie. Pozostałe istotne komponenty reklamy charakteryzowały się zróżnicowaną istotnością w zależności od zastosowanej metody regresji.

Słowa kluczowe: metoda regresji, regresja liniowa, drzewo regresyjne, reklama, rynek mleczarski.