Targeting through smartphone application in off-line retail: evidence from field experiment.

Mariia Okuneva, University of Kiel;
Dmitriy Potapov, Higher School of Economics

Abstract.
Purpose – This paper aims to investigate the effectiveness of personally targeted promotion via new communication media – branded mobile application developed by offline grocery retailer.
Design/methodology/approach – Randomized field experiment was conducted in partnership with Russian regional retail chain with total number of participants more than 10 thousands people. The treatment was an exposer of a discount offer for a particular customer as a pop-up smartphone notification and response – purchase incidents and purchase amount during promo campaign.
Findings – Our main findings illustrate heterogeneous influence of exposer over both advertisement content and customers’ characteristics. Moreover, we show and discuss that the effect can be even negative.
Practical implications – This study can help a retail chain to elaborate rules for individual targeting that assure more profits.
Originality/value – This paper seem to be the first one that investigates effectiveness of promotion via branded mobile application for off-line retailer in field experiment settings. Furthermore, the study answers not only the question about average promo effect, but examines response heterogeneity across customers RFM characteristics as far as category and discount depth of promoted goods.

Keywords: mobile targeting, randomized field experiment, mobile application, promo effectiveness.
Paper type – Research paper
Introduction.

In the era of information technology new marketing channels arise and open novel opportunities for both producers and retailers to make customers aware of brands, promotions and special offers. One of such channels is called ‘branded mobile apps’, which is defined (Bellman et al., 2011) as: “software downloadable to a mobile device which prominently displays a brand identity, often via the name of the app and the appearance of a brand logo or icon, throughout the user experience”.

According to Portio Research (Portio Research, 2013), mobile applications are becoming more popular: the number of people worldwide using mobile apps is forecast to rocket from 1.2 billion at the end of 2012 to 4.4 billion users by the end of 2017. This rapid growth of the market motivates companies to create ‘branded applications’ aimed at building long-term relationship with loyal customers and attracting those consumers who enjoy using this up-to-date media.

Previous researchers indicate that such apps have generated substantial interest among marketers, primarily because of their high level of user engagement and the positive impact this presumably has on a user's attitude toward the sponsoring brand. Bellman and others (Bellman at al., 2011) determine whether using popular mobile phone apps affects brand attitude and brand purchase intention. Their results show that using these apps has a positive persuasive impact, increasing interest in the brand and also the brand's product category. In more recent paper (Kim et al., 2015) authors provide quantified evidence of the effects of mobile app using customers' mobile application log data and transaction histories that are more improved measures of these behaviours.

Almost all these papers investigate mobile apps launched by the producer of international brands. Nevertheless we don’t know any study that concerns applications branded by off-line grocery retail chain as a part of their loyalty program. Such applications allow the user to have an access to their purchase history, create ‘to buy’ list, seek the information about special offers, leave feedback or get extra relevant information. Simultaneously they enable retailer to
make personalized offers (e.g. discounts for certain goods) for different customers taking into account their purchase history. This new media starts playing a significant role in off-line retail communication strategies, in particularly nine Russian top-10 retailers launched mobile apps by the March 2017.

Thus the aim of this paper is to analyse how effective is the targeting though mobile application owned by grocery retail chain assuming heterogeneity in response. For that purpose we utilize randomized field experiment in cooperation with grocery retail chain that operates in Russian off-line market. Our treatment was to send customers a push-notification informing about the current discount for one of the products sold in the store. As a desirable response we examine actual purchase behaviour measured by the number shopping trips and purchase amount.

To preview our results, we find out heterogeneous effect of advertising corresponding to both offer and customer characteristics. Moreover, we demonstrate there can be even negative treatment effect depending on the depth of discount mentioned in the offer (customers dislike small discounts). Furthermore we show that treatment effect is significantly diverse in interaction to behaviour characteristics of customers such as recency, frequency and monetary value. This combination of findings provides exact opportunities for practical implications.

The paper proceeds as follows. In section 2, we discuss prior relevant studies and formulate research questions. In section 3, we suggest the design of the field experiment and exploited model specifications in order to set up a framework for the analysis of personalized promotion. Empirical results are presented in section 4. The section 5 concludes the paper.

**Theoretical background**

This part of the paper aims to discuss three streams of relevant literature in marketing and economics that supplement each other. They are: effectiveness of promotion and advertising, field experiments and personalized marketing. For each of this stream we will especially focus on mobile applications studies. This analysis has the potential to highlight the under-researched areas and establish research question for this paper.
The effectiveness of advertising has become interesting for researchers and practitioners in marketing many years ago (Bagwell, 2008) and is covered in relation to TV (Lodish, 1995), store flyers (Gijsbrechts et al., 2003), coupon campaigns (Venkatesan et al., 2012), etc. The results reported in the literature are ambiguous. For example, Lodish in his work found out that only 49% of 360 advertising campaigns for different brands were statistically significant at the 20% level. The most relevant predecessors to our research are the papers exploring the effectiveness of the advertising provided via the Internet and mobile phones (SMS-messages, smartphone applications). The appearance of the Internet allowed to collect data about millions of users, target those who would respond more probably. It caused the development of recommendation systems, such as Amazon.com’s personalized book and music recommendation (Arora, 2008). While the advertisements on the web sites have become better and more personal, the main question about the effectiveness of online advertising has remained undiscovered. Lewis and Reiley (Lewis and Reiley, 2014b) have found that online advertisement leads to an increase of purchases by 5% (brick-and-mortar stores account for 93% of the growth). The authors underlined the complexity to find statistically significant effect of advertisement due to high variance of sales even with the use of large sample – 1.6 million of individuals.

Mobile marketing has its own advantages: SMS-messages can reach the person wherever he is that should make an ad more effective. Moreover, the customer can be targeted by time and even place (just near the store). In the article (Luo et al., 2013) the effectiveness of such practices is investigated using the sample of 12265 mobile phone users. The authors conducted large-scale randomized experiment and came to the conclusion that individually geographical and temporal targeting are effective, but simultaneous use of these two strategies can lead to different results. Merisavo (Merisavo et al., 2006) also identified the effectiveness of mobile advertising on the basis of field experiment and found that the effect of advertising varied across people with different content preferences and usage level of the mobile services. While X. Luo et al. (Luo et al., 2013)
consider only focal targeting, N. Fong, Z. Fang and X. Luo (Fong et al., 2015) go one step further and explore the effectiveness of competitive targeting. Researchers vary three factors in the experiment (location of customers, discount size and time) and conclude that medium discount (40%) is optimal for focal targeting, whereas deep discount (60%) should be used for competitive targeting.

We observed many articles about the effectiveness of advertising campaigns in different media, but did not give clear explanation when the ad is considered effective. The problem is that different metrics can be used for this purpose: weekly sales (Lewis and Reiley, 2014a, 2014b), purchase intent (Goldfarb and Tucker, 2011; Bart et al., 2014), purchase probability (Luo et al., 2013), attitude toward advertised product (Bart et al., 2014), store traffic – weekly number of receipts per store outlet – (Gijsbrechts et al. 2003), trip revenue (Venkatesan et al., 2012), average daily expenditure (Merisavo et al., 2006) or purchase rate (Fong et al., 2015). The choice of the metric should be based on the purpose of the research and availability of data.

To our knowledge, no empirical research exists about the effect of advertising campaign run through mobile application on sales of the retailer. At the same time, Bellman (Bellman, 2011) investigated the effect of mobile apps on brand attitude and brand purchase intention. He has found that highly relevant to the person and informative apps are characterized by greater values of purchase intention. In terms of the current study, it means that personalization of the advertisement content should lead to an increase in the effectiveness of the branded app. Another research devoted to branded mobile apps (Eunice, 2013) examines engagement attributes and entertainment features common to such applications.

Because of the endogeneity problem that arises with regard to the relationship between advertising and sales (simultaneous causality), omitted variable bias or selection bias (Bagwell, 2008), randomized experiments are used more often to measure the effectiveness of advertising than observational data. For instance, one of the earliest experiments in this field (Ackoff, 1975) was devoted to the effect of advertising on Budweiser beer sales. According to Levitt (Levitt,
2009), experiments with private entities will be more popular in future and they will be aimed at testing and extending current economic theories. However, field experiments have specific limitations and drawbacks. One of the issues is associated with randomization bias that presents a serious problem in research devoted to medical trials or other laboratory experiments. At the same time, Harrison (Harrison et al., 2008) and Levitt (Levitt, 2009) assert that randomization bias is not an important limitation in other types of experiments (such as field experiments).

Other relevant to this research articles are concentrated on personalized targeting. One-to-one marketing, targeting and personalization are the important concepts of customer relationship management. Targeting, or one-to-one marketing, refers to “setting marketing policy differentially for different customers or segments” (Dong, 2009). Personalization is the form of one-to-one marketing that can be described as the process of identifying the best match between marketing mix and customer’s preferences by the company (Arora, 2008). The work written by P. Rossi (Rossi et al., 1996) was among the first that underlined and quantified the effectiveness of direct targeting. Authors found that revenue associated with target couponing can exceed mass market couponing by 2.5 times. Another article (Ansari and Mela, 2003) is concentrated on the effect of content targeting. According to Ansari and Mela, content personalization of e-mail letters can lead to an increase of click-throughs by 62%.

Targeted offers will be effective in terms of response rates, only if information provided to a customer is perceived as relevant. The trouble is to articulate what kind of variables (characteristics of past consumer behavior, demographic information, special features of offer) enable the researcher to determine whom to target and with what sort of advertising campaign. For instance, is it more effective to advertise the product to those who buy it frequently or to the customers that have never bought it? Zhang and Wedel (Zhang and Wedel, 2009) discussed this issue and concluded that loyalty promotions (aimed at customers who bought the target good on the prior occasion) are more effective in
online stores than competitive promotions, offering products to those who didn’t buy them, while the opposite is true for offline stores explored in this work.

According to Bose and Chen (Bose and Chen, 2009) there are three main types of data used in direct marketing: external data (customers’ geographic, demographic and lifestyle characteristics), customers’ transaction records and feedback from consumers. Recency, Frequency and Monetary value (RFM) model, summarizing transaction data about buyers, is often used to select the customers that are worth targeting (Colombo, 1999). The most simple form of this framework assumes that response rate depends on following factors: how often the customer buys the product or visits the shop, how much the consumer spends on current and past transactions and how recently the last purchase has been made by a buyer.

The question about what products are more suitable for targeted promotions is rarely explored with the help of randomized experiments in the literature. We are familiar with two works in this field. Bart (Bart et al., 2014) proved that mobile display advertising of utilitarian products with higher level of involvement was more effective than advertising of hedonic goods with lower involvement in terms of consumers’ favorable attitudes and purchase intentions. Blake et al. (Blake et al., 2015) investigated product response heterogeneity but did not find significant difference in advertising effectiveness across various product attributes.

The main goal of our research is to understand the relationship between characteristics of customers, different content of the message and the effectiveness of grocery retailer targeting via branded mobile apps. In order to understand who to target and with which offers, we would like to answer the following questions:

1) How off-line FMCG retailer targeting provided via mobile application influence consequent customers purchase behavior?

2) Does this impact differ across characteristics of the message (discount depth and product category)?

3) Is this impact heterogeneous across behavioral (RFM) characteristics of customers?
Getting the answers for these questions can help us to provide customers of the retail chain with more relevant offers that is to create some effective rules for personalized targeting system.

**Methodology**

*Experimental Design*

To enable an unbiased estimate of the targeting effect, we conducted a field experiment performed in collaboration with the retail grocery chain. The chain consists of stores of different sizes located in the city of Perm (one in the top 15 largest cities in Russia with a population of more than 1 million people). The retailer developed the branded smartphone application about half a year before the experiment to extend their existing loyalty program. Almost 10 thousand customers from more than 300 thousand loyalty program participants installed the app before the experiment started. The application seem to be beneficial for both customers who are able to create purchase lists, use the app as a discount card, and browse their purchase history, and the retailer who can communicate through an additional channel. For the retailer in particular, the app provides the opportunity to send users push notifications. These notifications can contain different pieces of texts and links to app pages displaying particular products on discount. Products are selected from an existing two-week offline promotion campaign when almost six hundred goods are made available to all customers at discounted prices.

For the purpose of this study by the treatment we determine sending the app user with a different push-notification. We randomly assigned all app users (9956 customers) into control and treatment groups. Control group members (1418 customers) are not eligible to see any message. Treated buyers (8538 people) were divided into twelve groups and received a push notification about a discount on one of twelve products belonging to six product categories: coffee, tea, juice, dairy products, sweets and non-food. As far as one of our research questions concerns if treatment effect differs among product categories we selected categories with different consuming patterns, share of valet and inter-purchase cycle. In each category we chose two products (high price and low price) from a wide range of
commodities on which a discount was offered during the regular two-week promo campaign. After that we checked that discount depth of offers picked up for experiment varies significantly (from 10% to 32 %) Table 1 summarizes information about discounts on the promoted products and their prices before (P₀) and during (P₁) the campaign. The last column contains the number of treated customers assigned to each offer.

Table 1. Characteristics of Promoted Products and Number of Customers in Experimental Groups

There are two most common methods of randomization – pure randomization and stratification (e.g. in relation to different past purchase characteristics of customers). According to Miriam Bruhn and David McKenzie (Bruhn and McKenzie, 2009), “in samples of 300 or more, the different methods perform similarly”. As the size of the sample in our research exceeds 300 customers, we used a pure randomization method that balanced the characteristics of customers across treatment and control groups. As far as the users of the application were not aware that they participated in an experiment and that the data could be used for the research, our experiment can be classified as a “natural field experiment” according to Harrison and List (Harrison and List, 2004).

Available data describes individual customer purchases from the 1st of January 2014 until the 15th of March 2015. This period can be split into the 14 months before the promo campaign was run (the “pre-test” period) and two weeks after this date (the “post-test” period). We assure that customers were randomly assigned to groups by comparing differences in the most important historical variables – recency (days since last purchase), frequency (number of purchases during 14 months prior to the experiment) and monetary value (average basket amount in the pre-test period). As none of the differences are statistically significant, we assure that the allocation of customers to either control or treatment groups were close to purely random.
**Models and specifications**

In this subsection we will present and justify three models and their specifications utilized to answer our research questions.

**Models and dependent variables.**

The goal of the study is to assess the effectiveness of targeting via mobile application on subsequent sales. In literature different metrics of sales are suggested in similar research contexts: weekly sales, purchase intent, purchase probability, attitude toward promoted product, average daily expenditure or purchase rate. In this paper three different models will be used to test the effects.

In our case a smartphone push-up notification reminds a consumer about the retailer or informs him or her about new promoted products. We assume that this prompt will influence consumers’ desire to make an additional purchase or prevent them from switching to another store. Thus the first dependent variable we investigate is the number of purchase occasions in the period of the current promo campaign (two weeks) - $Q$. As this variable is discrete with an ordered metric ($Q_i=0,1,2,...$) and has large number of zero values (31% of app users didn’t complete any transaction) we use Zero-inflated Poisson regression for the analysis. The specification is as follows:

$$
\begin{align*}
Pr(Q_i = 0) &= w_i + (1 - w_i) \cdot e^{-\lambda_i} \\
Pr(Q_i = q) &= (1 - w_i) \cdot \frac{e^{-\lambda_i} \cdot \lambda_i^q}{q!}; q = 1, 2, ..., \\
\ln(\lambda_i) &= \beta X_i
\end{align*}
$$

(M1)

where parameter $\lambda_i$ is responsible for intensity of purchase occasion process and depends on the treatment that is discussed further in the section.

Additional purchase occasions induce an increase in the total sum spent by a consumer in the chain. Thus the second dependent variable is the amount of money paid in the period of the promo campaign – purchase amount $PA$. We will use two models to estimate treatment effect on purchase amount. First, we will use the classical linear model as many researches do to explore different effects of targeting.
\[ PA_i = \beta X_i + \varepsilon_i \]  

(M2)

Besides, we would like to take into account the left-censored nature of our dependent variable. We assume that the customer has a latent demand for goods, denoted by \( PA^* \), that is not expressed as a purchase until some constant threshold, denoted by \( \gamma \), is passed (Cameron and Trivedi, 2005). The basic idea is that we observe \( PA^* \) only when it exceeds a threshold. Furthermore, as expenditure data is better modelled as lognormal, we will deal with a special case of Tobit model for lognormal data with a nonzero threshold. The threshold equals the minimum uncensored value of \( \ln(PA^*_i) \). Maximum likelihood is used as an estimation method for this model.

\[
P_{Ai} = \begin{cases} 
PA^*_i, & \text{if } \ln(PA^*_i) > \gamma \\
0, & \text{if } \ln(PA^*_i) \leq \gamma 
\end{cases} \]  

(M3)

\[
PA^*_i = \exp(\beta X_i + \varepsilon_i)
\]

Average treatment effect of targeting.

Basically we pay attention to the average effect of targeting. The effectiveness of targeting via digital media (SMS-messages, online advertising, mobile display advertising) has been explored in several papers. For instance, Blake et al. (Blake et al., 2015) proved with the help of field experiments that eBay’s advertising on Google had a small and statistically insignificant effect on sales. Lewis and Reiley (Lewis and Reiley, 2014b) report a randomized field experiment that finds an increase in purchases of the retailer by 5% caused by advertising on Yahoo! This growing interest in the ads effect is associated with new technological opportunities that allow researchers to carry out large-scale field experiments and track important variables that are required for measurement of causal effect. Still, some challenges persist to correctly estimate the causal effect of advertising. As Lewis and Reiley (Lewis and Reiley, 2014b) mention, the effect of brand advertising is often diffuse and may be not as immediate as the effect of other types of advertising, such as direct mailing. One advertising campaign can be not enough to change purchase behavior of consumers.

Our targeting is special: the retailer, by informing customers about discounts
for some goods, advertises not the products, but rather it’s own brand increasing buyers’ intention to choose its store for a shopping trip. That is why it is really difficult to predict whether our targeting will be effective on average. Furthermore, the application is targeted at loyal customers who would like to know about all the discounts provided with the retailer. This can be the case that advertising might not change purchase habits of these people because they already value the retailer and are familiar with the take-off products offered in the chain.

To estimate the average effect of targeting we utilize the following specification:

$$\beta X_i = \beta_1 \cdot \text{Exposed}_i + \epsilon_i$$ (S1)

where $\text{Exposed}_i$ – a dummy variable that takes the value of 1 when the user is exposed to any offer.

In both models (M1) and (M3) $\beta_1$ is the coefficient of interest that is interpreted as the percentage difference in respective dependent variables between consumers who were exposed to the targeting and those who did not receive any message. $\beta_1$ in the model (M2) is interpreted as an absolute difference in purchase amount between exposed to targeting and control groups. Due to the randomized nature of our experiment $\beta_1$ coefficient can be explained only as an effect of the targeting campaign, while other factors (competitive or macroeconomic events) influence both treatment and control groups and don’t bias coefficient of interest.

**Treatment effect heterogeneity across discount depth.**

We have divided all the promoted products into three groups based on the discount depth of the offers: discounts under 20%, discounts between 20% and 30% and discounts above 30%. This discount variation is typical for grocery chains operating in Russia. Moreover, discount rates applied in our experiment cover almost the whole range of cut-offs available in the chain.

In the article by N. Fong (Fong, 2015) it is said that the medium discount (40%) is optimal for focal targeting and it leads to an increase in purchase rates. We consequently expect that making customers aware of higher discounts (30-32% in our experiment) will make customers purchase more. At the same time, as most
of the application’s users are loyal customers, they can be familiar with the average discount level in the retail chain. Sending such people the message containing information about small discounts may cause no extra intention to visit the chain.

We answer the question about influence of discount depth by jointly estimating the coefficients \( \beta_1, \beta_2 \) and \( \beta_3 \) in the following specification of models (M1)-(M3):

\[
\beta X_i = \beta_1 \times Discount1_i + \beta_2 \times Discount2_i + \beta_3 \times Discount3_i + \epsilon_i, \quad (S2)
\]

where \( Discount1_i \) – dummy variable that takes the value of 1 when the user is exposed to the advertisement of a product discount which is less than 20% and the value of 0 otherwise (control group serves as the baseline condition); \( Discount2_i \) – dummy for discounts in the range 20-29%; \( Discount3_i \) – dummy for discounts greater than 29%.

Treatment effect heterogeneity across different customers characteristics

One of the goals of the study is to measure if treatment effect is heterogeneous for different customers in correspondence to their behavioural characteristics. To light this question we employ widely used RFM (recency, frequency, monetary value) variables because they take into account major behavioural characteristics that can be extracted from transaction data.

Recency of the last purchase.

Experiments are rarely used in the literature to study heterogeneous effects of advertising/promotion in correspondence to behavioral characteristics of consumers. Two field experiments (Lewis and Reiley, 2014a; Johnson et al., 2014) explore heterogeneity of advertising effects in relation to demographic characteristics (age, gender) and location. At the same time, the only large scale field experiment that reports the dependence of advertising’s effect on recency of purchase was carried out by Blake et al. (Blake et al., 2015). The authors find a large and statistically significant effect of eBay advertising on consumers who have not purchased in over a year. The authors explain that this occurs due to an informative function of advertising: the effect is significant for those who do not remember about the offerings of eBay. Gonul and Shi (Gonul and Shi, 1998)
received the same results for direct mailing in non-experimental settings: researchers applied a structural model to the database of a national cataloger and found that it is not effective to mail to individuals at low recency levels.

To understand how the effect of targeting is influenced by the recency (how much time passed since last customer’s purchase) interaction, the following specification of models (M1)-(M3) is estimated:

\[ \beta X_i = \beta_1 * \text{Exposed}_i + \beta_2 * \text{Exposed}_i * \text{Recency}_i + \beta_3 * \text{Exposed}_i * \text{Recency}_i^2 + \beta_4 * \text{Recency}_i + \beta_5 * \text{Recency}_i^2 + \varepsilon_i, \]

where \( \text{Recency}_i \) – log of the number of days since the customer’s last purchase. The \( \beta_1 \) coefficient shows the effect of targeting for a customer who made a purchase in a store immediately before receiving the message (when Recency equals zero), \( \beta_2 \) and \( \beta_3 \) are parameters of interest.

Frequency of purchases.

The experiment conducted by Blake et al. (Blake et al., 2015) provides empirical evidence of the largest and most significant effect of eBay advertising on sales for consumers who have not completed eBay transactions in the year before the experiment. The authors explain this finding by the informative role of advertising. We suppose that in our experiment infrequent customers who downloaded the application will be highly affected by the presence of advertising because this will provide them with new and relevant information.

Gonul and Shi (Gonul and Shi, 1998) assert that it is not optimal to target those consumers who purchased many times from the catalog because such customers are likely to buy anyway. However, our case is a bit different because the branded mobile application is a ‘pull’ kind of media. It means that the users download the application if they feel that they need it. The application is created for loyal customers in order to build a long-term relationship. Consequently, we suspect that frequent customers are loyal and will be influenced by the targeting aimed at loyal customers. Actually, this also can prove to be an informative view of advertising because loyal customers download the application to know about all the discounts of the retailer, i.e. to get new information.
The analysis of the heterogeneous effect of targeting in relation to the frequency of purchases made by a consumer in 14 months prior to the experiment is performed by estimating the following specification:

$$\beta X_i = \beta_1 * \text{Exposed}_i + \beta_2 * \text{Exposed}_i * Frequency_i + \beta_3 * \text{Exposed}_i * Frequency_i^2 + \beta_4 * Frequency_i + \beta_5 * Frequency_i^2 + \epsilon_i$$

(S5)

where \( Frequency_i \) – the log of the expression: number of purchases for the 14-month period before the campaign divided by the lifetime of the consumer expressed in weeks. We define lifetime as the number of weeks between the first and the last transactions completed by the user. Similar to the (S4) specification we are interested in \( \beta_2 \) and \( \beta_3 \) estimates.

Monetary value (average basket amount).

While we are not familiar with the articles where the relationship between effectiveness of targeting and the monetary value of customers is discussed, we expect that the effect of message can vary across people with a different average basket amount.

For the sake of completeness, we estimate the following specification:

$$\beta X_i = \beta_1 * \text{Exposed}_i + \beta_2 * \text{Exposed}_i * \text{Monetary value}_i +$$

$$+ \beta_3 * \text{Exposed}_i * \text{Monetary value}_i^2 + \beta_4 * \text{Monetary Value}_i +$$

$$+ \beta_5 * \text{Monetary value}_i^2 + \epsilon_i$$

(S6)

where \( \text{Monetary value}_i \) – total expenditure for 14 months prior to the targeting campaign divided by the number of purchases made during the same period of time, thus it is the average basket amount.

**Results**

*Average effect of targeting*

In this section we present our results about the effectiveness of a conducted targeting campaign on average (average treatment effect). The two dependent variables of interest are the purchase amount of customers and the number of purchases in the period of the campaign.

We run a Zero-inflated Poisson regression to estimate the effect of targeting
on the number of purchases in two weeks (specification (2) of model (1)). Tobit regression and linear regression are used for the analysis of the same treatment effect on the post-test purchase amount (specification (4) of the model (3) and specification (7) of the model (5-6)). The control group consists of 1418 customers, whereas 8538 mobile application users received the pop-up notification.

According to our results, the probability of the number of purchases being zero is higher for customers exposed to the targeting (Table 2). Moreover, targeting leads to a significant decrease in purchase amount ($\beta_1 = -0.137$).

Table 2. Average Effect of targeting

Results reported in the articles on the effectiveness of advertising are often ambiguous. For instance, Lodish et al. (Lodish et al., 1995) report the results of 360 advertising campaigns on TV and find that only 49% of such campaigns were statistically significant at the 20% level. Lewis and Reiley (Lewis and Reiley, 2014b) discuss two reasons for small and insignificant effects that researchers find: high variance of a dependent variable and insufficient sample size. However, the only article to our knowledge where authors discuss the negative effect of advertising (Anderson et al., 2010) is devoted to the impact of deep discounts on the long-run demand of customers who had recently paid a higher price for one of the advertised products. In our opinion, the negative average effect we get is associated with the influence of messages containing information about small discounts. We will discuss this issue further in the next section.

*Heterogeneity of the effect of targeting across discount depth*

In the research by Fong et al. (Fong et al., 2015) it is assumed that purchase rate in the case of 20%-discount should be the same as purchase rate in the no-discount case. Our experimental design enables us to investigate the impact of discount depth on number of purchases and purchase amount. We form three groups of discounts: under 20%, 20%-29% and discounts greater than 29%. Discount lower than 20% was assigned to three promoted products: Sweets1,
Sweets2 and Dairy products1. The second group representing discounts from 20% to 29% consists of six advertised products: Non-food goods1, Dairy products2, Coffee1, Coffee2, Tea2 and Non-food goods2. The last group includes the following products: Juice2, Tea1, Juice1. The number of messages sent to customers with information about discounts under 20%, 20-29% and discounts greater than 29% is respectively 2121, 4292 and 2125.

We would like to note that the deepest discounts in this research (30%, 32%) are not considered as ‘deep’ in the literature. However, these were the deepest discounts in the retail chain during the promo campaign.

In order to answer the question about the effect of different discounts on number of purchases we run a Zero-inflated Poisson regression (specification (8) of model (1)). Specification (8) of the models (3) and (5-6) helps to estimate the treatment effect of three discount groups on purchase amount.

Table 3. The Effect of Discount Depth

We find a very interesting result that messages containing information about small discounts (under 20%) have a negative impact on the number of purchases made by customers and their purchase amount (Table 3). We suppose that the negative average effect of the targeting campaign we find is caused by the negative influence of small discounts. This can be the case that application users feel disappointed with promotions they are informed about because the offered discount is too small. Consequently, they make fewer purchases than those who did not receive any information.

Heterogeneity of the effect of targeting across RFM (Recency, Frequency, Monetary value) characteristics of customers.

Recency of the last purchase.

We expect that the effect of targeting can be dependent on recency of customer’s last purchase. We test a quadratic form of relationship between targeting effect and Recency variable (how much time has passed since a customer’s last purchase) by estimating the specification (10) of models (1), (3) and (5-6). Table VI reports the coefficients from estimating these models. We find
a quadratic form of relationship between the Exposed and Recency variables. This implies it is effective to target customers who did not visit a shop for a long time that is consistent with an informative view of advertising. At the same time, customers who did not make any purchase for more than half a year are unlikely to respond to offers. Figure 1 describes the relationship between treatment effect on number of purchases (Model 1, second equation for \(E(Q)\)) and log of Recency variable. Blake et al. (Blake et al., 2015) and Gonul and Shi (Gonul and Shi, 1998) receive qualitatively the same results.

Table 4. Impact of Targeting By Recency of Prior Purchase

Figure 1. Heterogeneity of treatment effect across Recency variable.

Frequency of purchases.

In this part of our work we explore interaction between Frequency and Exposed variables. We test quadratic form of relationship between frequency of purchases made by a customer within 14 months prior to the targeting campaign and treatment effect.

Table 5 reports the coefficients from estimating specification (11) of models (1), (3) and (5-6). We find that the relationship between the variables of interest is quadratic.

To better understand the interaction between treatment effect and frequency of purchases we draw a graph (Figure 2) showing relationship between log of Frequency variable and the effect of targeting on number of purchases (Zero-inflated Poisson model, second equation).

The result that targeting is effective for infrequent customers is consistent with findings by Blake et al. (Blake et al., 2015) and Gonul and Shi (Gonul and Shi, 1998). This can be explained by informative view of advertising because customers who rarely make purchases are often unfamiliar with offerings of the retail chain. However, Gonul and Shi (Gonul and Shi, 1998) highlight that it is not optimal to target those consumers who purchased many times. It is possible that we
get an opposite result because of nature of the targeting campaign we run. Only those frequent customers who would like to get information about news and discounts of a retail chain download the application. So, they are sensitive to targeting messages sent through the application.

Table 5. Impact of Targeting by Frequency

Figure 2. Heterogeneity of treatment effect across Frequency variable.

Monetary value (average basket amount).

We suppose that the effect of targeting can also depend on the third important characteristic of customers – monetary value. We test quadratic form of relationship between log of average basket amount and the treatment effect.

Table 6. Impact of Targeting by Monetary Value

Table 6 reports the coefficients from estimating specification (12) of models (1), (3) and (5-6). We find that the relationship between the monetary value of application users and the treatment effect is quadratic. Figure 3 represents this relationship. We can conclude that it is worth targeting customers with small monetary value and high monetary value.

Figure 3. Heterogeneity of treatment effect across Monetary value variable.

Conclusion & Discussion

This work contributes to the literature on promotion and personalised targeting by presenting empirical findings about the effectiveness of push-up promo campaign through a new and very special medium – branded mobile application. The main results of our research are the following:

• The impact of targeting campaign either on number of purchases or on purchase amount is slightly negative on average. This is due to small discount offers.
• The effect of targeting depends nonlinearly on RFM characteristics of consumers.

These results can help a retail chain to create some rules for individual targeting and understand who are more sensitive to the message.

An average effect of targeting we get can be underestimated. Some customers turn “push” notifications off from their smartphones. Randomization procedure ensures that equal rates of consumers do not receive a message across all groups. Nevertheless, the users who did not get a message were not influenced by variation of offers. Therefore, aggregate difference between treatment and control groups is likely to be smaller than in the situation when all consumers are affected by targeting. Thus, the effect we estimate is called “intention to treat effect”. It means that we analyse all the customers that were treated. Dividing “intention to treat effect” by the share of users who actually have seen the message leads to “treatment on treated” effect (Lewis and Reiley (Lewis and Reiley, 2014b)) that can be greater and statistically significant. However, we do not have data about a share of such customers.

We believe that the main limitation of this study is associated with generalization problem that is common to field experiments. We carried out one experiment for a single retailer that makes it uncertain to what extent the results will generalize. Moreover, to make results more persuasive it is better to replicate an experiment. However, to the moment we had not an opportunity to do an experiment again and check whether we come out with the same results.

References


### Tables.

#### Table 1. Characteristics of Promoted Products and Number of Customers in Experimental Groups

<table>
<thead>
<tr>
<th>Product Type</th>
<th>$P_0$</th>
<th>$P_1$</th>
<th>Discount</th>
<th>Number of customers treated with offer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee 1</td>
<td>1119</td>
<td>849</td>
<td>24</td>
<td>699</td>
</tr>
<tr>
<td>Coffee 2</td>
<td>124,9</td>
<td>94,9</td>
<td>24</td>
<td>734</td>
</tr>
<tr>
<td>Tea 1</td>
<td>499</td>
<td>349</td>
<td>30</td>
<td>729</td>
</tr>
<tr>
<td>Tea 2</td>
<td>79,90</td>
<td>59,90</td>
<td>25</td>
<td>714</td>
</tr>
<tr>
<td>Dairy product 1</td>
<td>78,5</td>
<td>64,9</td>
<td>17</td>
<td>697</td>
</tr>
<tr>
<td>Dairy product 2</td>
<td>59,4</td>
<td>45,9</td>
<td>23</td>
<td>727</td>
</tr>
<tr>
<td>Juice 1</td>
<td>86,9</td>
<td>58,9</td>
<td>32</td>
<td>701</td>
</tr>
<tr>
<td>Juice 2</td>
<td>61,5</td>
<td>42,9</td>
<td>30</td>
<td>695</td>
</tr>
<tr>
<td>Sweets 1</td>
<td>296,5</td>
<td>249,9</td>
<td>16</td>
<td>702</td>
</tr>
<tr>
<td>Sweets 2</td>
<td>66,9</td>
<td>59,9</td>
<td>10</td>
<td>722</td>
</tr>
<tr>
<td>Non-food 1</td>
<td>149</td>
<td>119</td>
<td>20</td>
<td>707</td>
</tr>
<tr>
<td>Non-food 2</td>
<td>139</td>
<td>100</td>
<td>28</td>
<td>711</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>8538</strong></td>
</tr>
</tbody>
</table>

Note. Table 1 summarizes information about discounts on advertised products and their prices before ($P_0$) and during ($P_1$) special offer period. Last column contains the number of customers treated with each offer.

#### Table 2. Average Effect of Targeting

<table>
<thead>
<tr>
<th></th>
<th>Number of purchases</th>
<th>Purchase amount (OLS)</th>
<th>Purchase amount (Tobit model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{Pr}(Q_i = 0)$</td>
<td>$\text{E}(Q_i)$</td>
<td>$\text{E}(Q_i)$</td>
</tr>
<tr>
<td>Exposed</td>
<td>0.041***</td>
<td>-0.024</td>
<td>-148.743</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.013)</td>
<td>(88.071)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.844***</td>
<td>1.579***</td>
<td>2204.617***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(83.333)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-25955</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td></td>
<td></td>
<td>0.0002</td>
</tr>
<tr>
<td>Sample size</td>
<td>9956</td>
<td>9956</td>
<td>9956</td>
</tr>
</tbody>
</table>

Note. Table 2 reports the coefficients from estimating equations (2), (4) and (6). Bootstrap clustered (across different groups) standard errors are in parentheses.  
***Significantly different from zero, $p<.01$  
**Significantly different from zero, $p<.05$

#### Table 3. The Effect of Discount Depth

<table>
<thead>
<tr>
<th>Discount</th>
<th>Number of purchases</th>
<th>Purchase amount (OLS)</th>
<th>Purchase amount (Tobit model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{Pr}(Q_i = 0)$</td>
<td>$\text{E}(Q_i)$</td>
<td>$\text{E}(Q_i)$</td>
</tr>
<tr>
<td>Discount1</td>
<td>0.059***</td>
<td>-0.057***</td>
<td>-227.821***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.016)</td>
<td>(110.311)</td>
</tr>
<tr>
<td>Discount2</td>
<td>0.041</td>
<td>-0.022</td>
<td>-180.670</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.021)</td>
<td>(96.112)</td>
</tr>
</tbody>
</table>
Discount%3  0.022  0.004  -5.328  -0.050  
(0.048) (0.019) (105.165) (0.108)  
Intercept -0.844***  1.579***  2204.617***  5.241***  
(0.000) (0.000) (84.328) (0.005)  
Log likelihood -25948  -21707  
$R^2_{adj}$  0.0006  
Sample size  9956  9956  9956  

Note. Table 3 reports the coefficients from estimating specification (7) of models (1), (3) and (5-6). Bootstrap clustered standard errors are in parentheses.  
***Significantly different from zero, p<.01  
**Significantly different from zero, p<.05

Table 4. Impact of Targeting by Recency of Prior Purchase

<table>
<thead>
<tr>
<th></th>
<th>Number of purchases</th>
<th>Purchase amount (OLS)</th>
<th>Purchase amount (Tobit model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr($Q_i=0$)</td>
<td>E($Q_i$)</td>
<td></td>
</tr>
<tr>
<td>Exposed</td>
<td>0.331</td>
<td>-0.339***</td>
<td>-1082.01**</td>
</tr>
<tr>
<td></td>
<td>(0.210)</td>
<td>(0.091)</td>
<td>(449.848)</td>
</tr>
<tr>
<td>Exposed*Recency</td>
<td>-0.132</td>
<td>0.295***</td>
<td>735.272***</td>
</tr>
<tr>
<td></td>
<td>(0.159)</td>
<td>(0.094)</td>
<td>(297.491)</td>
</tr>
<tr>
<td>Exposed*Recency^2</td>
<td>0.014</td>
<td>-0.044**</td>
<td>-108.453***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.020)</td>
<td>(44.091)</td>
</tr>
<tr>
<td>Recency</td>
<td>1.201***</td>
<td>-1.005***</td>
<td>-3906.675***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(286.886)</td>
</tr>
<tr>
<td>Recency^2</td>
<td>0.015***</td>
<td>0.057***</td>
<td>466.046***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(42.318)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.386***</td>
<td>3.019***</td>
<td>8097.702***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(434.648)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-21123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td></td>
<td>0.233</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>9956</td>
<td>9956</td>
<td>9956</td>
</tr>
</tbody>
</table>

Note. Table 4 reports the coefficients from estimating specification (9) of models (1), (3) and (5-6). Bootstrap clustered standard errors are in parentheses.  
***Significantly different from zero, p<.01  
**Significantly different from zero, p<.05

Table 5. Impact of Targeting by Frequency

<table>
<thead>
<tr>
<th></th>
<th>Number of purchases</th>
<th>Purchase amount (OLS)</th>
<th>Purchase amount (Tobit model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr($Q_i=0$)</td>
<td>E($Q_i$)</td>
<td></td>
</tr>
<tr>
<td>Exposed</td>
<td>0.335***</td>
<td>0.027***</td>
<td>-188.736**</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.010)</td>
<td>(85.025)</td>
</tr>
<tr>
<td>Exposed*Frequency</td>
<td>-0.149***</td>
<td>0.114***</td>
<td>2.898</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.022)</td>
<td>(95.165)</td>
</tr>
<tr>
<td>Exposed*Frequency^2</td>
<td>-0.161***</td>
<td>0.050***</td>
<td>45.567</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.014)</td>
<td>(51.374)</td>
</tr>
</tbody>
</table>

26
Table 5 reports the coefficients from estimating specification (10) of models (1), (3) and (5-6). Bootstrap clustered standard errors are in parentheses.

***Significantly different from zero, p<.01

**Significantly different from zero, p<.05

Table 6. Impact of Targeting by Monetary Value

<table>
<thead>
<tr>
<th></th>
<th>Number of purchases</th>
<th>Purchase amount (OLS)</th>
<th>Purchase amount (Tobit model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr(Qi = 0)</td>
<td>E(Qi)</td>
<td></td>
</tr>
<tr>
<td>Exposed</td>
<td>1.845</td>
<td>3.341***</td>
<td>-1767.948</td>
</tr>
<tr>
<td></td>
<td>(1.074)</td>
<td>(0.488)</td>
<td>(3529.878)</td>
</tr>
<tr>
<td>Exposed*Monetary value</td>
<td>-0.653</td>
<td>-1.088***</td>
<td>680.360</td>
</tr>
<tr>
<td></td>
<td>(0.347)</td>
<td>(0.157)</td>
<td>(1263.218)</td>
</tr>
<tr>
<td>Exposed*Monetary value^2</td>
<td>0.057**</td>
<td>0.087***</td>
<td>-65.732</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.013)</td>
<td>(111.286)</td>
</tr>
<tr>
<td>Monetary value</td>
<td>-3.787***</td>
<td>2.544***</td>
<td>-2630.840**</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(1140.839)</td>
</tr>
<tr>
<td>Monetary value^2</td>
<td>0.269***</td>
<td>-0.204***</td>
<td>366.270***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>0.000</td>
<td>(101.181)</td>
</tr>
<tr>
<td>Intercept</td>
<td>12.104***</td>
<td>-6.277***</td>
<td>4374.495</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(3165.504)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-25640</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R^2_{adj}</td>
<td></td>
<td>0.164</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>9956</td>
<td>9956</td>
<td>9956</td>
</tr>
</tbody>
</table>

Note. Table 6 reports the coefficients from estimating specification (11) of models (1), (3) and (5-6) on a sample of consumers who received only one advertising message or did not receive any message. Bootstrap clustered standard errors are in parentheses.

***Significantly different from zero, p<.01

**Significantly different from zero, p<.05
Figures.

Figure 1. Heterogeneity of treatment effect across Recency variable.


Figure 2. Heterogeneity of treatment effect across Frequency variable.


Figure 3. Heterogeneity of treatment effect across Monetary value variable.
The relationship between treatment effect and log of Monetary Value:
2nd equation of Zero-Inflated model