LOYALTY PROGRAMS AND PERSONAL DATA SHARING PREFERENCES IN THE CZECH REPUBLIC

Radek Tahal, Tomáš Formánek, Hana Mohelská

Introduction

Loyal customers are a key factor for successful retail activities. Customer loyalty can be defined as a higher probability of making new and repeated purchases, spontaneously recommending a particular retailer and spreading the positive word-of-mouth. "Loyal customers are less likely to switch to a competitor due to price inducement, and these customers make more purchases compared to less loyal customers" (Dhal, 2015; Mohelska & Sokolova, 2016) For the retailers, customer loyalty is a key factor in reaching long-term commercial success and profitability. It is much cheaper to retain and curate existing customers than to acquire new ones. Similarly, customer loyalty has been increasingly studied in health and social services over the past few years – particularly in economies where the private sector dominates this field. Many such studies focus on the topic of demand regulation within the health and social services sector, which is strongly motivated by population aging and related factors occuring in developed countries (Gavurová et al., 2014; Šoltés et al., 2014).

In modern customer-oriented marketing, loyalty programs (defined as marketing programs that reward members with purchase incentives) are perceived as the strategic instrument for creating and maintaining effective contact with customers (Bacik et al., 2015). "In the retail scene, loyalty programs involve a concentrated effort by retailers to build store traffic, increase basket size and increase frequency which creates deeper relationship ties with its customer base" (Omar, Wel, Musa, & Nazri, 2010). An important part in creating consumers' satisfaction and inducing consumer loyalty is played by the human factor (e.g. Wasan and Tripathi (2015)). It is often claimed that retail personnel should be well instructed and motivated towards increasing the number of loyal customers and creating harmony between corporate strategic aims and the customers' demand.

As loyalty programs (LPs) are implemented, managed and evaluated, retailers need appropriate data and means to identify individual participating customers. Such data allow for effective evaluation of consumers' purchasing behaviour and habits. Specifically, for the purpose of registering, uniquely identifying and properly managing customers in a LP, various identification data are required. In the Czech Republic, establishing and managing any such consumer/ LP database is bound by legal framework, implemented for consumer and personal data protection. Besides legislature, customers often fear that their personal data – once passed to LP-operator – might be misused (e.g. sold to third parties and used in an intrusive manner).

Generally speaking, consumers worry about their privacy and are afraid of losing control over their personal data – to some extent. Importantly, consumers' privacy concerns are not uniform. Individual preferences, sociodemographic and lifestyle factors play a significant role. From a good-faith LP organizer point of view, the quality of personal data and contact information collected through the LP has an enormous effect on their ability to manage and evaluate LPs properly. This study provides a structured quantitative analysis of customers´ willingness to share various types of personal and contact data with LP organizers.

1. Literature Review

The use of LP-related personal and contact data for the benefit of marketers and businessmen is a delicate topic, discussed by academic researchers and retail marketers, as well as by institutions supervising legal aspects of such activities (see e.g. Albrecht (2006) or Mariner and Cannella (2015)).

"Loyalty programmes, appropriately managed, are considered to allow structured and effective actions to manage, select, relate, and control customers' buying behaviour" (Lara & De Madariaga, 2007). In order to meet their objectives, LPs need to be beneficial to both the retailer and the customers. A new study by Nielsen (a corporate provider of information and insights into what consumers watch and buy), revealed that nearly 60 percent of global respondents reported that LPs were available to them through local retailers. From those consumers with access to LP-organizing retailers, 84 percent report being more likely to visit such retailers (Nielsen, 2013).

Lara and De Madariaga (2007) discuss the importance of finding a desirable balance between the volume of data required by the retailer and customers' willingness of sharing them. As consumers are disinclined to facilitate access to certain data types to the companies (LP organizers), this can be a factor that moderates consumer willingness to participate in such programme. "Ironically, even though loyalty program members crave a more personalized, relevant experience, they also show concern about sharing the information required to enable the retailer to deliver on this desire." (PR Newswire, 2013). Consumer incentives, along with trust and privacy assurances are the key strategic tools in LP consumer engagement.

Numerous studies point out that people are often afraid of their personal data becoming an object of trading among companies (see e.g. Wade (2010), Spiekermann, Acquisti, Böhme and Hui (2015)). "Compounding this problem is the common practice of businesses selling customer information to other businesses, marketing companies, mailing lists and so forth, thereby further increasing the amount of unwanted offers and advertising" (Marketing Weekly News, 2013).

In their corporate research, Aimia (2011) describe the young generation (generation Y or the Millennials, born approximately between 1977 and 1994) as being less concerned when sharing personal data with a retailer for shopping reasons. Of all named marketing channels in their survey, loyalty and reward programs are perceived as the most privacy-friendly by Millennials: fewer than 20% of Millennial loyalty program members are concerned about sharing personal information with loyalty program organizers.

As Koponen and Mangiaracina (2014) state, a good way to illustrate the increasing commercial value of personal data is by considering the recent growth of the online advertising sector. "Our culture places a high value on privacy. Putting control of the data – what may be collected and retained and what must be deleted (the right to be forgotten) – in each of our hands through a contractual arrangement with the collector or seller, promotes individual choice and control" (Anderson, 2015).

It is retail marketers' task to communicate personal data requirements to consumers as well as the advantages of LPs. Both LP-organizers and consumers may be viewed as being in a process of developing a new understanding of the rules of consumer engagement programs (such as LPs) that not only match consumer preferences but help consumers accomplish goals in real-time – see Document News (2013) for detailed discussion. Customers ought to be assured that their data will be utilized ethically, lawfully and for improving retail offers. "Imagine a member of a customer loyalty program who is completely informed and agrees that data about his shopping habits are used to optimize business processes and to provide advertising material and special offers" (Matzner, 2014).

2. Research Focus

This research is focused on customers´ behaviour, privacy and data-sharing preferences related to LP participation in the Czech Republic. According to our previous research, Czech customers are less inclined to sharing their personal and contact data as compared to customers in selected EU countries (see e.g. Tahal (2015)). We aim to help marketers with a particularly important task: to find the appropriate balance between the amount of personal data required from consumers participating in a given LP and the extent to which consumers are willing to share such data. In this paper, we use the primary survey data to answer two main research questions:

Research question 1: What determines consumer willingness to provide different types of contact and personal information to organizers of LPs within the FMCG – are there any significant differences that would be determined by socio-demographic and lifestyle factors?

Research question 2: Does the extent of individual participation in LPs affect consumer's willingness to provide contact and personal information?

It should be explicitly noted that our research activities (and conclusions made) aim towards fair LP schemes, i.e. LPs where customers may realistically feel that in exchange for the personal information provided they are rewarded by adequate purchase incentives and/or improved and more personalized services by the retailer.

Our research of consumer willingness to provide contact and personal information to LP organizers focuses on the following 6 types of personal data: (a) Name & Surname, (b) Email, (c) Address/residence, (d) Birthdate, (e) Phone number, (f) Personal ID number. Overall, this classification reflects common marketing practice, as LP-based personal data are used for retail segmentation, analysis, auditing, forecasting, etc. In this paper, the first data type (Name & Surname) and the last data type (Personal ID number) have a limited, benchmark-type usage: Providing one's name in order to participate in a LP is the very minimum requirement. Besides Czech language-specific gender segmentation, name bears little usable information and it is often seen as the least sensitive data type. Therefore, Name/Surname data may serve as a benchmark in measuring the basic propensity to share non-sensitive personal data. In contrast, the Personal ID number is a very sensitive type of information with a non-negligible misuse potential. Actually, LP organizers in the Czech Republic are not allowed to legally collect and use this type data under most practical circumstances (exceptions may apply). Also, Personal ID bears little usable information, once data types (b) to (e) are controlled for. Therefore, we use the information on consumer willingness to share their ID numbers as a second benchmark, for the other end of the personal data sensitivity spectrum.

3. Data and Research Methodology

Our research is based on primary survey data from the Czech Republic. A complex, anonymized and stratified (quota) sampling was performed during the period from November 2015 to April 2016, gathering data for a sample of 411 respondents from the population of FMCG consumers aged 15+. The stratified quota sampling was based on three factors: location, age and gender of the respondents. Hence, our methodology ensures consistency and relevance of the results – conclusions may be drawn with respect to the population of 15+ consumers. A combination of personal and online data collection was used to gather socio-demographic information, lifestyle preferences, attitudes towards different types of work and free time activities of the respondents. Both quantitative (interval based) and qualitative (Yes/No, Likert scale-based) questions were used in the survey.

The survey was organized and performed by a research team at the University of Economics, Prague. This team is led by university employees and teachers who supervise and coordinate the tasks performed by students specializing in marketing research. This study is part of a systematic long-term project of specialized marketing analyses (see e.g. Tahal and Stříteský (2014)). Our empirical analysis (logistic regression and related statistical inference) is adjusted to control for the quota sampling, thus ensuring proper conclusions are made towards the FMCG consumer population based on the results and their interpretation. Also, for survey data validation, the Wald-Wolfowitz "Runs" test was used to test the H_0 of order of observations being attributable to chance against the H_1 of potential data collection mishandling (Wackerly et al., 2008).

All the surveyed data (interval-based, Yes/No, Likert scale) need to be conveniently stored for subsequent quantitative analysis. Given the nature of our questionnaires, the gathered data may be conveniently recorded as logical (binary) variables. The transformation of Yes/No answers (e.g. for questions related to willingness to provide personal data) is straight-forward. For interval-based quantitative topics such as age or income, we use binaries to record respondent's appropriate interval entry. For example, the variable Age_65_plus equals 1 for all the respondents aged 65 and older and is zero otherwise. Answers to Likert scale-based questions are also recorded as binary variables. For the sake of our analysis, ordered multinomial data may be approached in a way similar to the interval-based quantitative variables. For example, respondents are asked to position themselves towards a statement "I like reading books" using a five degree Likert scale ("1" = this statement describes me very

well, …, "5" = this statement does not describe me at all). Subjectively perceived importance to individual lifestyle is addressed by this question, rather than some actual measure of reading time (or page volume count). Also, the fact that "1" is a better rating than "2" (in terms of agreeing with the statement evaluated) bears ordinal meaning only, i.e. we cannot say that the difference between "2" and "4" is somehow twice as important as the difference between "4" and "5". Therefore, the surveyed answers to this statement were used to produce two binary variables: *LS_books_yes* equals 1 for those who respond "1" on the Likert scale and zero otherwise, *LS_books_no* equals 1 for respondents who dissociate themselves from the statement by answering "5" (and zero otherwise). Although binary variable may be produced for each of the Likert scale answers, we find it empirically convenient to combine statements "2" to "4" – i.e. not a very strong position of the respondent – into a single reference category. Our approach has three advantages: First, we retrieve all cases where respondents have strong personal positions on given lifestyle activities and topics such as reading books, doing sports, eating organic food, etc. Second, the combined base answers "2" to "4" may still be included implicitly in the analysis as a reference category, necessary for interpretation of the estimated regression models. Third, using such reference base de-couples the *LS_book_yes* and *LS_book_no* binaries that are not linearly dependent and may be both used as explanatory variables in the same regression model.

Using the above described approach, we have transformed the surveyed material into a 402-row dataset (9 respondents were disregarded due to missing data issues) with 106 variables. Six of the variables describe consumer willingness to share personal data with LP organizers. The remaining one hundred variables form a pool of potential/conceivable regressors bearing socio-demographic, lifestyle and other relevant information that may be used to model consumers' personal data sharing preferences. For our dataset, an exhaustive (brute-force) search for a truly optimal parametric model specification is computationally inaccessible, as it would require an estimation and evaluation of 6×2^{100} models. Hence, in order to identify a small, informative and consistent set of explanatory variables, we combine a forward-stepwise selection method

(this is a potentially suboptimal algorithm that produces nested sequences of models) with the non-parametric random forest approach. Differences in outputs from the two methods are analysed in order to detect any potential sub-optimality in output from the stepwise method. This approach allows for efficient and computationally feasible evaluation of individual explanatory variables with respect to prediction accuracy of the model, as only 6×100^2 models and $6 \times 5{,}000$ random forests are produced and evaluated (fast and automated evaluation procedures are available in R and other software packages). Although our methodology does not guarantee a truly optimal model specification, it may be regarded as an acceptable approximation with a relatively low potential for sub optimality). For detailed discussion of model selection methodology, see James et al. (2013). Hence, in order to answer the research questions (RQs) 1 and 2, the above described model selection process was used to generate a consistent model specification as outlined by equation (1):

$$
\begin{aligned}
y_i = \beta_0 + \beta_1\, \text{Female}_i + \\
+ \beta_2\, \text{Age\_15\_24}_i + \beta_3\, \text{Age\_65\_plus}_i + \\
+ \beta_4\, \text{Moravia}_i + \beta_5\, \text{Earnings\_high}_i + \\
+ \beta_6\, \text{LS\_TV\_no}_i + \beta_7\, \text{LS\_books\_no}_i + \\
+ \beta_8\, \text{LS\_Internetuse\_no}_i + \\
+ \beta_9\, \text{LS\_Paycard\_yes}_i + \\
+ \beta_{10}\, \text{LS\_exotics\_yes}_i + \\
+ \beta_{11}\, \text{LS\_cooking\_no}_i + \\
+ \beta_{12}\, \text{LP\_Memb\_1\_2}_i + \\
+ \beta_{13}\, \text{LP\_Memb\_3\_plus}_i + u_i
\end{aligned}
\tag{1}
$$

where y_i is a binary dependent variable – six different dependent variables are used with the left hand side of the equation and therefore six different equations are estimated using the model (1) – Yes/No answers were recorded for the question "Would you provide the following type of personal information in order to become a member of a loyalty program?" for each respondent and data category (a) to (f) as defined above. Please note that such question applies identically to consumers who already are members and/or active users of a LP as well as to individuals not participating in LPs. On the right hand side of (1), β_j are the coefficients to be estimated through logistic regression (see Davidson, MacKinnon (2009, p. 454-465)). *Female$_i$* is a binary explanatory variable distinguishing between female and

male respondents, $Age\_15\_24_i$, is a binary indicating the 15-24 age group while $Age\_65\_plus_i$ depicts individuals aged 65 and older (upon variable importance evaluation as described above, age ranges 25-34, 35-49 and 50-64 are combined into a single base category). $Moravia_i$ describes the residence of respondents and Bohemia serves as its reference category. Respondents with high earnings (defined by a monthly household income over 80 thousand CZK) are discerned using $Earnings\_high_i$. Likert scale-based lifestyle variables $LS\_TV\_no_i$, $LS\_books\_no_i$ and $LS\_Internetuse\_no_i$ mark respondents who dissociate themselves from watching TV, reading books and using the internet. $LS\_Paycard\_yes_i$ indicates whether the i-th respondent uses pay cards (credit and debit) frequently and $LS\_exotics\_yes_i$ determines whether respondents are keen on spending their holidays at exotic destinations or by the sea (as the Czech Republic is a landlocked country). $LS\_cooking\_no_i$ discerns people who dissociate themselves from cooking (again, subjectively perceived importance is addressed here). $LP\_Memb\_1\_2_i$ identifies individuals (consumers) who actively participate in 1 or 2 LPs. Similarly, $LP\_Memb\_3\_plus_i$ describes respondents who actively take part in 3 or more LPs. Finally, u_i is the potentially heteroskedastic random element.

The interpretation of most of the lifestyle variables included in equation (1) is relatively straight-forward, given the intuition on Likert scale data transformation as provided above. However, the last two variables (*LP_Memb_1_2* and *LP_Memb_3_plus*) require some additional explanation: In the questionnaire, respondents were asked two questions related to the quantity of LPs. First, respondents provided the total number of LPs they are members of. The first question served mostly as a lead-in to the next question, where individuals reported their actual (active) LP participation. In this article, we focus on active LP usage rather than a simple LP membership as we find active LP participants to be a far more interesting group as far as LP organizers are concerned. Although the respondents were choosing from five options (0 LPs actively used, 1-2, 3-4, 5-9, 10+), low observed quantities in the last two categories led us to combine the last three LP categories into a single category of 3+ LPs actively used. Hence, we include *LP_Memb_1_2* and

LP_Memb_3_plus variables in equation (1) to describe the effect of active LP participation, while consumers using zero LPs form a base (reference) category. The two LP-focused variables discussed in this paragraph are indispensable for answering the research question 2. In this context, it is worth mentioning that the inclusion of both variables is based on the variable selection (regressor importance evaluation) process described above – i.e. we didn't have to "force" the two variables into equation (1) in order to allow for answering research question 2.

The logistic function used for estimation of the β_j coefficients in equation (1) may be expressed as

$$P(y_i = 1 \mid \mathbf{x}_i^T) = G(\mathbf{x}_i^T\boldsymbol{\beta}) =$$
$$= \exp(\mathbf{x}_i^T\boldsymbol{\beta})/[1 + \exp(\mathbf{x}_i^T\boldsymbol{\beta})], \qquad (2)$$

where $P(y_i = 1 \mid \mathbf{x}_i^T)$ is the probability of success (i.e. consumer willingness to provide the specific type of his/her personal data), given the observed row vector of explanatory variables $\mathbf{x}_i^T$. The expression $G(\mathbf{x}_i^T\boldsymbol{\beta})$ is a simplified notation for the logistic function: $\exp(\mathbf{x}_i^T\boldsymbol{\beta})/[1 + \exp(\mathbf{x}_i^T\boldsymbol{\beta})]$ which guarantees that all fitted values of the dependent variable lie within the $\langle 0, 1 \rangle$ interval. In model (2), the direction of the effect of change in the explanatory variable x_i on the probability of "success" in the dependent variable is always determined by the sign of the corresponding β_j coefficient. However, the magnitudes of the individual β_j coefficients are not sufficiently informative, given the nonlinear nature of the logistic function. The effect of a change in x_j on the probability of "success" for the i-th respondent must be calculated individually: as can be seen from equation (2), a change in conditional probability of success is calculated from a compound function that depends on the following arguments: β_j, all the remaining coefficients in vector $\boldsymbol{\beta}$ and all the observed values of the explanatory variables for the i-th respondent ($\mathbf{x}_i^T$). Hence, for the i-th respondent and a chosen binary explanatory variable, say x_k, the partial effect from changing x_k from 0 to 1 (while holding all other explanatory variables unchanged) may be simply calculated as

$$\Delta G(\,.\,) = G(\beta_0 + \beta_1 x_{1,i} + ... + \beta_{k-1,i} x_{k-1,i} + \beta_k)$$
$$- G(\beta_0 + \beta_1 x_{1,i} + ... + \beta_{k-1,i} x_{k-1,i}) \qquad (3)$$

In expression (3), we may note that the β_k coefficient is present when $G(.)$ is evaluated for $x_k = 1$ and omitted for $x_k = 0$. For each individual consumer, the expression (3) may be used for evaluation of changes in conditional success probabilities. However, for effective model interpretation, we need to summarize the individual information obtained from equation (3) across all individuals. This is done through the average partial effect (APE) statistics:

$$
\begin{aligned}
\text{APE}(x_k) = n^{-1} \sum_{i-1}^{n} \Big[& G(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + ... + \\
& -G(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + ... \\
& + \hat{\beta}_{k-1,i} x_{k-1,i} + \hat{\beta}_k) \Big], \\
& ... + \hat{\beta}_{k-1,i} x_{k-1,i})
\end{aligned}
\tag{4}
$$

where $\hat{\beta}_j$ are the sample estimates of β_j coefficients. In equation (4), the expected partial effect of changing a given binary regressor x_k from 0 to 1 (ceteris paribus) is obtained for each of the survey respondents and then averaged across individuals. Using this approach, $\text{APE}(x_k)$ values are usually reported along with their corresponding standard errors and significance statistics. Using expression (4), consistent APEs may be calculated for all binary regressors x_j in model (1). Although all regressors in model (1) are binary, the specification chosen provides enough control for diverse observed factors that it allows for a straightforward interpretation of individual APEs; a situation that is analogous to the Ignorability of treatment assumption (as in Wooldridge (2010, p. 908)).

4. Empirical Results

Individual willingness to disclose personal information to LP organizers is approached using diverse data evaluation methods in order answer RQs 1 and 2 and to quantify important sociodemographic and lifestyle aspects of this type of consumer behaviour. First, table 1 summarizes the overall consumer readiness to share personal data. Rows are organized in descending order and we may see that there are prominent differences in personal information sharing preferences related to the type of data. The willingness to share personal data ranges from 91.3% (Name & Surname) to 3.7% (Personal ID number). This range between the two benchmark data types (as described above) sets an interpretation framework for the remaining data types. In fact, as we test for statistically significant differences in observed means (relative "success" ratios, i.e. the average willingness to share a specific type of data), we find that all mean pairs are statistically different (6 categories make for 15 possible category-pairs), with the exception of Birthdate and Phone number – means of those two categories are not statistically different at any reasonable significance level. Such evaluation is based on the Wilcoxon signed rank test (for matched/correlated pairs). For detailed description of the test, see Wackerly et al. (2008). The row ordering and the positive outcome ratios shown in table 1 provide an interesting insight into individual privacy preferences and the general attitude of consumers to personal data disclosure. For example, we may point out the quite low level of preparedness to share phone numbers, that

Tab. 1: Observed willingness to share different types of personal information

Personal information type & consumer willingness to share it	No. of „Successes" (out of 402 respondents)	Positive outcome ratio	Variance
Name & Surname	367	0.913	0.080
Email	260	0.647	0.229
Address/residence	227	0.565	0.246
Birthdate	156	0.388	0.238
Phone number	149	0.371	0.234
Personal ID number	15	0.037	0.036

Source: own

consumers prefer to keep undisclosed more than other contact information such as Email and Address/residence. Most probably, this attitude is related to the avoidance of advertisement voice calls and text messages, which may be perceived as far more intrusive when compared to emails or paper-based leaflets. Deleting irrelevant e-mails is less bothering than dealing with unsolicited phone calls and text messages (e.g. Blackburn (2015)).

The variance information in table 1 is provided mostly for readers' convenience. Given the binomial nature of the underlying dummy variable representing willingness or unwillingness to share personal data, variance equals $p(1-p)$, where p is the Positive outcome ratio. Table 1 provides reasonable overall insight into consumers' personal data sharing preferences. Yet, in order to answer RQ1 and RQ2, we need to focus on the sociodemographic and lifestyle aspects. At the individual level, many random elements and factors play a significant role in defining the extent of personal data sharing. However, by means of logistic regression, we are often able to quantify statistically significant differences in personal data sharing habits between specific groups of respondents (either sociodemographic or lifestyle-based). Such results may be presented in an informative, easily accessible and often actionable form that may be used for diverse marketing and LP-management purposes.

Next, we turn our attention to the RQ1. Based on the explanatory variables (sociodemographic and lifestyle factors), model (1) was estimated for each of the six dependent variables (willingness to share different personal data types). All estimated models provide reasonable estimation accuracy and are statistically significant at the 5% significance level – with the exception of the model estimated for Personal ID number, which is only significant at α = 10%. In logistic regression, the individual coefficient estimates are not particularly informative, except for their signs and statistical significance. Therefore, we skip the regression output from equation (1) and focus on the APE values as defined in (4). In fact, all signs and statistical significances of the $\beta_{c,j}$ coefficients are unambiguously reflected in the corresponding $\mathrm{APE}_c(x_j)$, where x_j is the j-th explanatory variable and subscript c denotes the c-th type of personal information – a dependent variable in (1). All logistic

estimation outputs omitted from this article are available from the authors upon request, along with primary data and the R-code used. In table 2, all $\mathrm{APE}_c(x_j)$ values are reported along with their standard errors (heteroscedasticity corrected values) and p-values. Columns in table 2 are organized by the overall consumer willingness to share personal data in the same descending order as in table 1 – thus allowing for a simple comparison.

For interpretation of the sociodemographic and lifestyle factors related to RQ1 – as shown in table 2 – we shall start with the variable Female as an example: Women are roughly 10% less likely to share their email address and phone number with LP organizers when compared to the reference category (i.e. men). For the remaining four personal data types, gender plays no significant role. Such interpretation is made in a ceteris paribus context – we control for all the remaining variables explicitly included in model (1). Interestingly, our findings somewhat contradict a common stereotype that regards women as more likely to conceal their age. This conclusion is implied through the dependent variable $\mathrm{Birthdate}$, for which respondent's gender is not a significant regressor. For illustration and readers' convenience, the ceteris paribus effects on willingness to share personal data related to the regressor Female are included in figure 1 (along with corresponding bars showing 90% significance intervals). For example, the leftmost bar (within the Female group) shows that women are 3.5% less likely to share their name with LP organizers. At the same time, the corresponding 90% significance interval includes zero and therefore this particular effect is not statistically significant at α = 0.1 (significance level of 10%).

The remaining sociodemographic factors influencing consumer willingness to share personal data with LP organizers may be briefly summarized as follows: From table 2 we can see that people aged 15 to 24 are 21% more likely to share their birthdate compared to the age-reference group of people aged 25 to 64. For other personal data types, significant influence of the variable $\mathrm{Age\_15\_24}$ is not identified. In contrast, respondents aged 65 and older diverge from the age-reference group rather significantly. They are almost 28% less likely to share their email addresses (note that the non-use of the Internet is controlled for

Tab. 2: **APEs for selected types of personal data**

	Name & Surname	Email	Address	Birthdate	Phone number	Personal ID number
Female (s.e.) [p-value]	-0.0352 (0.0269) [0.1904]	-0.1032 * (0.0407) [0.0112]	-0.0014 (0.0490) [0.9779]	0.0464 (0.0483) [0.3371]	-0.1015 * (0.0489) [0.0378]	0.0047 (0.0236) [0.8428]
Age_15_24	-0.0144 (0.0600) [0.8108]	0.0841 (0.0675) [0.2130]	0.0393 (0.0743) [0.5970]	0.2106 * (0.0751) [0.0050]	0.0820 (0.0747) [0.2724]	0.0281 (0.0361) [0.4364]
Age_65_plus	-0.0582 ˙ (0.0310) [0.0606]	-0.2792 * (0.0628) [0.0000]	-0.1520 * (0.0624) [0.0148]	-0.0306 (0.0627) [0.6261]	0.1186 ˙ (0.0629) [0.0593]	0.0085 (0.0283) [0.7629]
Moravia	-0.0179 (0.0299) [0.5503]	-0.0839 ˙ (0.0446) [0.0600]	-0.0431 (0.0489) [0.3774]	-0.0865 ˙ (0.0476) [0.0693]	-0.0094 (0.0475) [0.8436]	0.0300 (0.0244) [0.2181]
Earnings_high	-0.1349 ˙ (0.0689) [0.0503]	-0.0443 (0.1065) [0.6775]	-0.2708 * (0.1085) [0.0126]	-0.0050 (0.1006) [0.9607]	-0.1118 (0.0938) [0.2333]	-0.0383 * (0.0093) [0.0000]
LS_TV_no	-0.0863 (0.0586) [0.1409]	-0.0839 (0.1038) [0.4187]	0.0423 (0.1037) [0.6833]	0.0203 (0.1255) [0.8716]	0.0019 (0.1118) [0.9866]	-0.0385 * (0.0094) [0.0000]
LS_books_no	0.0949 * (0.0131) [0.0000]	0.1755 * (0.0700) [0.0121]	0.1460 ˙ (0.0873) [0.0943]	0.3214 * (0.0836) [0.0001]	0.1028 (0.0960) [0.2844]	0.0475 (0.0406) [0.2418]
LS_Internet_use_no	-0.0828 (0.0529) [0.1174]	-0.3343 * (0.1320) [0.0113]	-0.0191 (0.0886) [0.8290]	-0.0393 (0.0915) [0.6671]	-0.2538 * (0.0782) [0.0012]	-0.0155 (0.0207) [0.4535]
LS_Paycard_yes	-0.0361 (0.0311) [0.2464]	0.1170 * (0.0439) [0.0076]	-0.0158 (0.0508) [0.7555]	-0.0069 (0.0502) [0.8908]	0.1083 * (0.0514) [0.0353]	-0.0276 (0.0224) [0.2193]
LS_exotics_yes	0.0424 (0.0292) [0.1466]	0.0337 (0.0525) [0.5205]	0.1086 ˙ (0.0583) [0.0628]	0.1023 ˙ (0.0581) [0.0786]	0.0647 (0.0575) [0.2608]	-0.0206 (0.0198) [0.2962]
LS_cooking_no	0.0103 (0.0287) [0.7187]	-0.1707 * (0.0805) [0.0339]	-0.0290 (0.0725) [0.6886]	-0.1575 * (0.0768) [0.0404]	0.0776 (0.0808) [0.3371]	0.0073 (0.0274) [0.7890]
LP_Memb_1_2	0.1254 * (0.0230) [0.0000]	0.1015 * (0.0422) [0.0162]	0.1450 * (0.0511) [0.0046]	0.0428 (0.0583) [0.4629]	0.1807 * (0.0534) [0.0007]	-0.0294 (0.0199) [0.1405]
LP_Memb_3_plus	0.1555 * (0.0237) [0.0000]	0.2412 * (0.0484) [0.0000]	0.3267 * (0.0548) [0.0000]	0.1735 * (0.0643) [0.0069]	0.3548 * (0.0603) [0.0000]	0.0577 (0.0418) [0.1680]

Source: own

Note: * – coefficient significant at α = 0.05; ˙ – coefficient significant at α = 0.1.
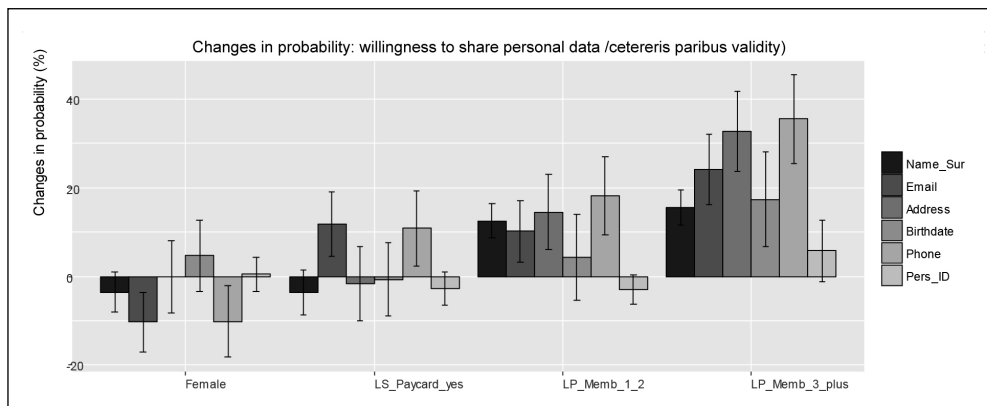
by a separate lifestyle explanatory variable $LS\_Internetuse\_no$), 15% less likely to divulge their address/residence and also 6% less likely to provide their names. On the other hand, members of the age group 65+ are about 12% more likely to hand over their phone numbers when compared to the reference. The reference group exhibits a rather uniform behavior in terms of personal data sharing preferences. Hence, it is not convenient to study the age ranges 25-34, 35-49 and 50-64 individually: we combine them into a single 25-64 reference and study how the younger and older consumers differ.

Respondent domiciled in Moravia are roughly 8.5% less likely to share their email and birthdate compared to the reference group (Bohemia combined with the separately surveyed region of Prague). Consumers with high earnings are 27% less likely to provide LP organizers with their address/residence, most probably as a security precaution. Similarly, they are 13.5% less likely to provide their name. Education (surveyed as primary/secondary/ university degree) does not have a significant effect on individual willingness to share different types of personal data, once the other factors in model (1) are controlled for. Therefore, education-related variables are omitted from equation (1) and from table 2.

Next, we briefly summarize the RQ1-related lifestyle factors that influence consumers' willingness to share data most prominently. Generally speaking, people who dissociate themselves from reading books (LS_books_no equals 1) are more likely to provide their personal data to LP organizers when compared to the reference group (active book readers combined with people without a strong position on this topic). The difference is most prominent for the $Birthdate$ dependent variable – we observe a 32% increase in probability. Respondents who dissociate themselves from using the Internet ($LS\_Internetuse\_no$) are less likely to share their personal data: we observe a decrease of 33% in willingness to share email (for obvious reasons, people who dissociate themselves from using the Internet are less likely to actually have an email address), a 25% decrease in likelihood of sharing phone number and even an 8% decrease in willingness to share name (however, this result narrowly misses statistical significance at α = 0.1). In contrast, a statistically significant effect of LS_TV_no is observed only for the Personal ID number (a decrease of 3%). Consumers who report being active paycard users (42% of the respondents strongly identify themselves with using a credit or debit card as measured by the Likert scale-based variable $LS\_Paycard\_yes$) are 12% more likely to share email address and 11% more likely to share phone number with LP organizers, while the effect of this lifestyle factor is not statistically significant for the other four types of personal data. For illustration, the variable $LS\_Paycard\_yes$ is also included in figure 1.

Fig. 1: **Illustration of selected results from table 2**

Source: own

Individuals who identify themselves with spending holidays at exotic destinations (including holidays by the sea as the Czech Republic is landlocked) are depicted by the regressor $LS\_exotics\_yes$. Such individuals are 11% more likely to share their addresses and 10% more likely to share their birthdate when compared to the reference group – perhaps the reason is that such consumers are used to share such data with different tour and travel operators. However, the effect of the $LS\_exoticsc\_yes$ is not significant for the remaining four data types considered (including Personal ID). People who dissociate themselves from cooking (subjectively perceived importance to one's lifestyle controlled for using the variable $LS\_cooking\_no$) are less inclined to share their personal data with LP organizers as follows: -17% for emails and -16% for birthdates.

The last two regressors in model (1) – $LP\_Memb\_1\_2$ and $LP\_Memb\_3\_plus$ – are used for answering our RQ2. From table 2 as well as from figure 1 we can see that active LP members are significantly more inclined to share personal data with organizers of a "new" LP they want to become members of. The interpretation of the last two rows in table 2 is somewhat different from the above factors as both $LP\_Memb\_1\_2$ and $LP\_Memb\_3\_plus$ are interpreted with respect to the same reference group – i.e. consumers who are not active members of any LP. This may be illustrated using the $Phone$ dependent variable: consumers who participate in 1 or 2 LPs are 18% more likely to share their phone number with LP organizers as compared to the reference. For consumers who participate in 3 or more LPs, this willingness increases by 35.5% as compared to the reference group of LP non-participants (i.e. 1-2 LP users are not directly compared with 3+ LPs participants). Similarly to the $Phone$ data, 1-2 LP users are 15.5% more likely to share address and 10 % more likely to share email, while for 3+ LP users the increase in probabilities (compared to base group) are 33% and 24% respectively.

For correct interpretation of the results in table 2, we need to keep in mind that the $APE_c(x_j)$ values in different columns of each row (as the influence of a selected factor is studied across different data types) are mutually independent by the nature of our regression-based methodology. This does not directly alter the interpretation of table 2, yet due to observed positive pairwise correlations among different personal data types (consumer willingness to share such data), some of the $APE_c(x_j)$ values for individual x_j regressors (i.e. values in rows of table 2) may be relatively similar in their signs, magnitudes and significance levels. We skip detailed correlation output from this article, yet the typical correlation between two personal data types (willingness to share) ranges from + 0.2 to + 0.3. The most prominent correlation was observed between the $Address$ and $Birthdate$ variables (+ 0.44), while correlation of $Personal\_ID$ with other dependent variables is somewhat lower (roughly + 0.1 to + 0.2). We do not find evidence for any mutually exclusive consumer preferences in personal data sharing – i.e. none of the dependent variables are negatively correlated.

Conclusions

In this paper, we apply various quantitative analysis tools to examine personal data-sharing preferences that are related to consumer participation in loyalty programs. In addition to sociodemographic aspects, we also focus on different lifestyle factors. Our detailed and stratified research results may be used by professional marketers and academic researchers in order to identify differences and similarities in consumer willingness to share diverse types of personal data. As far as different sociodemographic and lifestyle variables are concerned, we may conclude that the willingness to share personal data and contact information changes significantly among specific consumer groups. Most of the important aspects and answers to our research questions are explicitly interpreted in the paper. However, readers may easily draw their own conclusions upon the estimates provided in tables 1 and 2, given the sociodemographic group and/or data type of their interest.

Among other topics, our research emphasizes the fact that personal data-sharing preferences are a convoluted phenomenon which requires a complex, structured and quantified research approach. Marketing professionals and LP organizers can use our results to amend the incentives incorporated in their LPs to optimize consumer data-gathering processes. Cost-benefit assessments may lead LP managers into focusing relevant purchasing incentives and other LP-based benefits towards consumer groups that are least willing to share

their data. Alternatively, such analyses may lead towards data-gathering compromises: partial redefinitions of LP structures and evaluation processes may take place in order to require only such personal information that target customer group-members are willing to share.

To highlight some of our empirical results, respondents aged 65 and older are significantly less inclined to share their personal data as compared to other age groups – which conforms to our prior (pre-research "general expectations"). On the other hand, we do not find statistical evidence for education-based differences in data sharing preferences: i.e. individual willingness to share personal and contact information does not depend on the education level achieved. The lifestyle-based structure of data-sharing preferences is complex and may be best observed from table 2. Finally, consumers who already participate in some LP are significantly more likely to share their personal data in order to sign up for a new LP. This overall increase in personal data-sharing propensity is even more prominent for consumers who participate in three or more LPs.

References

Aimia. (2011). *How Generation Y will reshape customer loyalty*. Retrieved November 24, 2016, from https://www.aimia.com/en/media-center/news-releases/viewer.html/en/how-generation-y-will-reshape-customer-loyalty.

Albrecht, K. (2006). *Supermarket "loyalty" cards and consumer privacy education: An examination into consumer knowledge about cards' data collection function.* Cambridge, US: Harvard University.

Anderson, J. (2015). The economics of privacy: Rethinking the value of personal data collection. *Journal of Information Ethics*, *24*(2), 5-8. doi:10.5210/fm.v19i12.5008.

Apriva, llc; researchers submit patent application, "system and method for secure marketing of customer data in a loyalty program", for approval. (2013). *Marketing Weekly News*, p. 495.

Bacik, R., Gavurova, B., & Fedorko, I. (2015). The analysis of the impact of selected marketing communication factors on the online consumer behavior. *Journal of Applied Economic Sciences*, *10*(7), 999-1004.

Blackburn, C. (2015). New technology, personal data protection and implications for financial services regulation. *JASSA*, *1*(4), 59-65.

Consumers believe total data privacy no longer exists. (2013). *Document News*, *31*(1), 7-8.

Davidson, R., & MacKinnon, J. G. (2009). *Econometric Theory and Methods* (pp. 454-465). New York: Oxford University Press.

Dhal, S. R. (2015). Demographic profile and customer loyalty: A study on organized retail sector. *International Journal on Global Business Management & Research*, *4*(1), 23-32.

Gavurova, B., Stefko, R., & Bacik, R. (2014). The analysis of social services' structure in a specific region and its significance for health and social policy in Slovakia. *Polish Journal of Management Studies, 10*(2), 43-53.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with applications in R.* New York: Springer-Verlag.

Koponen, J., & Mangiaracina, A. (2014). No free lunch: Personal data and privacy in EU competition law. *Computer and Internet Lawyer*, *31*(6), 7-15.

Lara, P. R., & De Madariaga, J. G. (2007). The importance of rewards in the management of multisponsor loyalty programmes. *Journal of Database Marketing & Customer Strategy Management*, *15*(1), 37-48. doi:10.1057/palgrave.dbm.3250066.

Mariner, W. K., & Cannella, M. E. (2015). The consumer privacy bill of rights: Window dressing for data mining? *Human Rights*, *41*(4), 21-22.

Matzner, T. (2014). Why privacy is not enough privacy in the context of "ubiquitous computing" and "big data". *Journal of Information, Communication & Ethics in Society*, *12*(2), 93-106. doi:10.1108/JICES-08-2013-0030.

Mohelska, H., & Sokolova, M. (2016). Smart, connected products change a company's business strategy orientation. *Applied Economics. 48*(47), 4502-4509. doi:10. 1080/00036846.2016.1158924.

Nearly a third of consumers are worried about the safety of their personal information in loyalty programs. (2013). Mintel Group Ltd, PR Newswire. Retrieved July 4, 2016, from http://www.mintel.com/press-centre/retail-press-centre/privacy-in-loyalty-programs.

Nielsen survey: 84% of global respondents more likely to visit retailers offering a loyalty program. (2013). Entertainment Close – Up. Retrieved July 7, 2016, from http://www.nielsen.com/us/en/press-room/2013/nielsen-survey-84-percent-of-global-respondents-more-likely-to.html.

Omar, N. A., Wel, C. A. C., Musa, R., & Nazri, M. A. (2010). Program benefits, satisfaction and loyalty in retail loyalty program: Exploring the roles of program trust and program commitment. *IUP Journal of Marketing Management*, *9*(4), 6-28.

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*(2), 161-167. doi:10.1007/s12525-015-0191-0.

Soltes, V., Balloni, A. J., Gavurova, B., & Soltes, M. (2014). Economic aspects of a health system electronization. In A. Moumtzoglou, A. Kastania, & S. Archondakis (Eds.), *Laboratory Management Information Systems: Current Requirements and Future Perspectives* (pp. 32-49). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-6320-6.ch002

Tahal, R. (2015). Comparison of customer attitudes towards loyalty programs in Prague and in London. *Central European Business Review*, *4*(3), 5-12. doi:10.18267/j.cebr.128.

Tahal, R., & Stříteský, V. (2014). Věrnostní programy a jejich vnímání zákazníky a provozovateli retailových internetových obchodů. *Acta Oeconomica Pragensia*, *22*(4), 30-41.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2008). *Mathematical Statistics with Applications* (7th ed). Belmont, CA: Brooks/Cole.

Wade, A. E. (2010). A new age of privacy protection: A proposal for an international personal data privacy treaty. *The George Washington International Law Review*, *42*(3), 659-685.

Wasan, P. G., & Tripathi, G. (2015). Revisiting social marketing mix: A socio-cultural perspective. *Journal of Services Research*, *14*(2), 127-144.

Mgr. Radek Tahal, Ph.D.
University of Economics, Prague
Faculty of Business Administration
Department of Marketing
radek.tahal@vse.cz

Ing. Tomáš Formánek, Ph.D.
University of Economics, Prague
Faculty of Informatics and Statistics
Department of Econometrics
formanek@vse.cz

doc. Ing. Hana Mohelská, Ph.D.
University of Hradec Králové
Faculty of Informatics and Management
Department of Management
hana.mohelska@uhk.cz

Abstract

LOYALTY PROGRAMS AND PERSONAL DATA SHARING PREFERENCES IN THE CZECH REPUBLIC

Radek Tahal, Tomáš Formánek, Hana Mohelská

Effective loyalty program management and evaluation requires that retailers have access to relevant data. In most cases, loyalty program organizers aim to establish consumer databases for the purpose of identification of individual customers: loyalty program members. The structure and quality of customer data often has a strategic effect on retailers' decision-making accuracy and profitability. On the other hand, consumers worry about their privacy and fear their personal data may be misused. For a good-faith loyalty program organizer, it is an ongoing task to reconcile their corporate interests with the interests of consumers who are often rewarded by purchase incentives and personalized services.

Consumer's willingness to disclose personal information to loyalty program organizers is not uniform. In fact, individual preferences, sociodemographic and lifestyle factors play a very important role. This study provides a structured quantitative analysis of customers´ willingness to share selected key types of personal and contact data with loyalty program organizers in the Czech Republic. Cost-benefit assessments based on our results may help marketing managers with establishing and/or amending key LP incentives. We identify and discuss important differences in personal and contact data-sharing preferences among specific consumer groups. To highlight some of the empirical results, respondents aged 65 and older are significantly less willing to disclose personal data as compared to younger consumers. On the other hand, we do not find a statistically significant evidence for education-based differences in data sharing preference. Our results may be utilized by marketing professionals (loyalty program organizers) as well as by academic researchers in order to optimize their consumer data-gathering processes.