Introductory Seminar Econometrics

Predictive Analytics Enhancing Donor Targeting in Charitable Fundraising through Direct Mailing

Team Q5

David Chen: 559994

Twan Mulder: 572626

Konstantin Asam: 576634

Quinten van den Vijver: 559692

Supervisor:

Sam van Meer

May 22, 2023

Wordcount: 4982

Abstract

This paper presents a scientific investigation into modelling household responses to direct mailings conducted by a prominent Dutch charity. We propose two novel target selection techniques that surpass the charity's current approach employing robust econometric methods. The first technique involves ordering households based on their expected donations, while the second technique prioritizes households based on response probabilities. Logistic regressions are employed to model the response probabilities, and Ordinary Least Squares is used to estimate the donation amounts upon response. Additionally, advanced machine learning techniques, including Ridge, Lasso, and Elastic-Net regression, are considered for fitting both models. The data set consists of 5000 households and 19 direct mailings collected by the charity. Our results demonstrate the superior performance of our target selection methods compared to the charity's approach. Furthermore, we find that receiving consecutive direct mailings does not adversely affect household response probabilities, and its impact on donation amounts is statistically insignificant.

1 Introduction

In this paper, we aim to develop an optimal target selection strategy for a direct mailing campaign by a large Dutch charity. The campaign involves sending quarterly mailings to selected groups of Dutch households, so-called "targets", soliciting donations. Our goal is to predict the behaviour of the targets, specifically whether they will donate and the amount they will contribute. To achieve this, we employ response modelling techniques leveraging data on past donating behaviour of 5000 Dutch households and 19 direct mailings which were sent between 1995 and 2000. Our objective is to maximise expected proceeds, as the charity heavily relies on public fundraising for its non-profit endeavors.

We quantify response behaviour with two models: response probability and donation amount models. We employ commonly used response modelling techniques, such as logistic regressions for response probability and Ordinary Least Squares (OLS) for donation amounts. Additionally, we compare these results with forecasts generated by advanced machine learning techniques such as Lasso, Ridge, or Elastic-Net regressions. Furthermore, we evaluate two target selection strategies: ranking groups based on expected donation amount or response probability. We anticipate that more accurate targeting will significantly enhance the efficiency of the direct mailing process, keeping costs constant while increasing campaign proceeds. The additional funds raised can be allocated to charitable projects, making this research highly relevant not only for the charity but also for society as a whole.

Target selection for direct mailings is already extensively studied in academic literature. In this paper, we build mainly upon two research papers. The first paper is that of Otter et al. (1999), in which the authors jointly model the probability of donating and quantity of donations by recipients of direct mailings. The second paper that we reviewed is van Diepen et al. (2009). They studied the effects of irritation induced by charitable direct mailings on donating behaviour. Diepen et al. (2009) found that direct mailings do result in irritation, but that it does not affect the donating behaviour of households. We want to check whether these results also hold for the large Dutch charity that we consider and, therefore, incorporate the effect of irritation into our response models.

This paper is organised as follows. In section 2 we introduce the data and describe the transformations performed on the variables. In section 3 we explain the econometric methods that underpin our response models and discuss two target selection techniques. Furthermore, in section 4 we present and compare the results. Finally, we conclude this paper in section 5.

2 Data

For our research, we analyse data from a large Dutch charity concerning past donating behaviour resulting from direct mailings sent to 5,000 households between 1995 and 2000. During this period, the charity has sent out 19 direct mailings (one per quarter) where only data from mailing seven up to mailing 19 is included in our data set. Each mailing includes identification variables, RFM (recency, frequency, and monetary value) variables, and zip-code area characteristics which are explained below.

The identification variables consist of $mail_{it}$, $resp_{it}$, $amount_{it}$, $hhid_i$ and $mailid_t$. $mail_{it}$ equals one if household i received mailing t, $resp_{it}$ equals one if household i responded to mailing t and donated money, and $amount_{it}$ equals the amount donated by household i for mailing t in Dutch guilders[1]. $hhid_i$ and $mailid_t$ contain the household and mailing identification number, respectively, linking observations to households and mailings.

The RFM variables include $lastresp_{it}$, $avresp_{it}$ and $avamount_{it}$. $lastresp_{it}$ is a binary variable and equals one if household i responded to the last received mailing. Additionally, $avresp_{it}$ represents the response percentage of household i and $avamount_{it}$ denotes the average donation of household i, both over the last 12 months preceding mailing t. Note that a household always donates money when it responds.

Additionally, we consider the variables $urblvl_i$, $hhsize_i$, $highinc_i$, $lowinc_i$ and $nojob_i$, which describe characteristics of the zip-code areas of the households. First, $urblvl_i$ indicates the urbanization level, in other words, the number of addresses divided by land area in the zip-code area of household i. Second, $hhsize_i$ equals the average household size in zip-code area of household i. Finally, $highinc_i$, $lowinc_i$ and $nojob_i$ represent the percentages of households with high income, with low income, and without a job in the zip-code area of household i, respectively.

To facilitate interpretation of the variable $urblvl_i$, we introduce the dummy variable D_urblvl_i, which equals one if the urbanization level of household i exceeds the median urbanization level of 1.34 (see Table 1). Moreover, we include two dummy variables, $mail.lag1_{it}$ and $mail.lag1\&2_{it}$, where the former equals 1 if household i received the mailing preceding mailing t and the latter equals 1 if household i received both of the two mailings preceding mailing t. These variables allow us to investigate the potential effects of irritation resulting from consecutive mailings among targeted households and to relate our findings to those of van Diepen et al., 2009.

[1] 1 Dutch guilder is about 0.45 euro

Table 1 Summary statistics identification, RFM and zip-code variables

Statistic	**N**	**Mean**	**Median**	**St. Dev.**	**Min**	**Max**
resp	65000	0.271	0.00	0.444	0.000	1.000
amount	65000	4.980	0.000	15.881	0.000	1500.000
lastresp	65000	0.393	0.000	0.488	0.000	1.000
avresp	65000	41.800	33.333	38.637	0.000	100.000
avamount	65000	11.600	10.000	22.136	0.000	1500.000
urblvl	65000	1.664	1.343	1.505	0.018	11.562
hhsize	65000	2.987	3.000	0.206	2.200	4.300
*plus*65	65000	15.130	14.500	5.972	1.000	54.000
lowinc	65000	39.360	39.000	6.151	21.000	59.000
highinc	65000	21.410	21.000	7.719	5.000	61.000
nojob	65000	17.410	16.000	6.762	5.000	47.000

Table 2 Correlation of Variables

	resp	*amount*	*lastresp*	*avresp*	*avamount*	*urblvl*	*hhsize*	*plus*65	*lowinc*	*highinc*	*nojob*
resp	1										
amount	0.515	1									
lastresp	0.243	0.111	1								
avresp	0.360	0.174	0.789	1							
avamount	0.123	0.315	0.225	0.298	1						
urblvl	0.003	0.032	0.0003	0.002	0.044	1					
hhsize	−0.001	−0.021	−0.001	−0.001	−0.027	−0.404	1				
*plus*65	0.019	0.026	0.022	0.029	0.032	0.051	−0.648	1			
lowinc	−0.015	−0.033	−0.023	−0.026	−0.063	0.070	−0.198	0.202	1		
highinc	0.026	0.041	0.034	0.043	0.080	−0.171	0.171	−0.003	−0.857	1	
nojob	−0.007	−0.005	−0.012	−0.013	−0.017	0.374	−0.488	0.337	0.710	−0.627	1

3 Methodology

In this section, we describe and motivate the chosen methods. We have two main objectives: (1) modelling the responses for direct mailings of a large Dutch charity, and (2) providing at least a better target selection than the charity. First, we will discuss the response modelling in this section. For this objective, we want to model the response probability of households to direct mailings and model the amount that will be donated by them. Secondly, we will describe our target selection techniques which are based on the law of total expectation and on response probability ordering. In our modelling approaches we use econometrics and machine learning techniques, all of which will be outlined in this section.[2]

3.1 Response probability modelling with logistic models

First, we model the response probability by using logistic regressions. As was mentioned in section 2, we have data on households of mailings 7 up until 19. However, to assess the predictive quality of our models, we only use data up to mailing 18 to fit our models. More specifically, for modelling the response probability, an observation related to household i is included in the estimation sample when this household received mailing t, for $t \in \{7, 8, ..., 18\}$. We exclude observations that did not receive a mailing as they cannot respond and, hence, would otherwise introduce bias to the regression coefficients.

Next, we describe our model. Since we want to model the responses to direct mailings, the dependent variable is the response variable, $resp_{it}$. This binary variable should not be modelled by a linear regression, as this could yield predictions above one and below zero. Therefore, to bound the response probability between zero and one, we consider logistic regressions. The logistic model can be formulated as

$$\mathbb{P}[resp_{it} = 1] = \Lambda(x_{it}'\beta) = \frac{e^{x_{it}'\beta}}{1 + e^{x_{it}'\beta}} \tag{1}$$

$$\begin{aligned} x_{it}'\beta = \beta_0 &+ \beta_1 lastresp_{it} + \beta_2 avresp_{it} + \beta_3 avamount_{it} + \beta_4 D\_urblvl_i + \beta_5 hhsize_i \\ &+ \beta_6 highinc_i + \beta_7 lowinc_i + \beta_8 nojob_i + \beta_9 mail.lag1 + \beta_{10} mail.lag1\&2 + \epsilon_{it}, \end{aligned} \tag{2}$$

which is solved using Maximum Likelihood Estimation (MLE). In this model, the RFM

[2]Finally, we performed PCA to reach our objectives, but due to similar results as the other methods, it can be found in the Appendix. We will not discuss the method or the results in this paper.

variables serve as a proxy for past donating behaviour. Furthermore, the lagged mailing variables are incorporated to model potential effects of irritation as mentioned in van Diepen et al. (2009). Finally, the zip-code characteristic variables are included to control for household characteristics.

3.2 Donation amount modelling with linear models

Another decision that a recipient of a direct mailing has to make is to determine how much to donate when he/she decided to act on the direct mailing. Therefore, using Ordinary Least Squares (OLS), we model the amount that will be donated given that the household responded to the direct mailing. Our motivation for this is that, ultimately, we want to provide a target selection where we need to calculate expected donations. Therefore, by modelling the donated amount given that the household responds and the response probability, we can use the law of total expectation to calculate expected donations for each household (see section 3.4.1 for further details). So, for the amount modelling, we restrict the estimation sample to households that responded and who are related to the first 18 mailings. All in all, we fit the following model to the data

$$\begin{aligned} amount_{it} = \beta_0 + \beta_1 lastresp_{it} + \beta_2 avresp_{it} + \beta_3 avamount_{it} \\ + \beta_4 D\_urblvl_i + \beta_5 hhsize_i + \beta_6 highinc_i + \beta_7 lowinc_i \\ + \beta_8 nojob_i + \beta_9 mail.lag1 + \beta_{10} mail.lag1\&2. \end{aligned} \tag{3}$$

The variables included in this model are justified using the same reasoning as in section 3.1. Instead of only using OLS regressions, this linear model will also be used in more advanced machine learning techniques, which will be discussed in the next section.

3.3 Ridge, Lasso and Elastic-Net regressions

In sections 3.1 and 3.2, we discussed the *default* model fitting approaches: logistic regression and OLS. However, it is well known that OLS performs poorly in out-of-sample predictions (Zou & Hastie, 2005). Therefore, in this section, we look at different modelling techniques that deal with this shortcoming: the Ridge, Lasso and Elastic-Net regressions. Like OLS, all three models minimize the sum of squared residuals (SSR). However, the objective function has an additional penalty function that should shrink the estimated coefficients towards zero. The main idea behind these regressions is that by shrinking the estimated coefficients towards zero, the predicted dependent variables are less sensitive to small changes in the regressors, which will improve the bias-variance trade-off.

However, while these three regression methods have the same goal, each method is different. For instance, in the Ridge regression, the estimated model will contain all variables (i.e., no non-zero coefficients). This means that Ridge regression does not automatically select appropriate variables, which leads to a parsimonious model (Zou & Hastie, 2005). For Lasso regressions, on the other hand, the estimated coefficients can and often will be set to zero, such that it automatically constructs a parsimonious model. Especially in the situation where we have a group of regressors among which the pairwise correlations are high, Lasso will select only one of the variables from this group and does not care which one is chosen. This is also immediately a disadvantage of the Lasso regression, since it does not take into account which variable is best to choose from the group of correlated variables. Finally, Elastic-Net deals with this disadvantage and is a generalization of the Ridge and Lasso regressions.

Now, we explain the mathematical formulation and algorithm for each model following the explanation by Friedman et al. (2010). We will focus on the Elastic-Net regressions, since Ridge and Lasso are specific instances of this model. Suppose we want to fit a simple linear model to the data (the following procedure can also be related to logistic models), where $Y \in \mathbb{R}$ is the dependent variable and $X \in \mathbb{R}^k$ are the k independent variables. Furthermore, we have N observations. Then the Elastic-Net regression solves the following problem

$$\min_{(\beta_0,\beta_1)\in\mathbb{R}^{k+1}} R_\lambda(\beta_0,\beta) = \min_{(\beta_0,\beta_1)\in\mathbb{R}^{k+1}} \left[\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i'\beta)^2+\lambda P_\alpha(\beta)\right] \tag{4}$$

$$P_\alpha(\beta) = (1-\alpha)\frac{1}{2}\|\beta\|_{l_2}^2+\alpha\|\beta\|_{l_1} \tag{5}$$

$$= \sum_{j=1}^{k}\left[\frac{1}{2}(1-\alpha)\beta_j^2+\alpha|\beta_j|\right] \tag{6}$$

where $P_\alpha(\beta)$ is called the *Elastic-Net penalty.* In the Elastic-Net regressions, α should be a value between 0 and 1, and is refered to as a *scaling parameter.* Furthermore, we get a Ridge regression when we restrict α to 0 in the above equations, and a Lasso regression when α is set to 1. Moreover, the parameter λ is restricted to be non-negative for all three methods, and is referred to as the *regularization parameter.*

For Ridge and Lasso regressions we use 10-Fold Cross-Validation to determine the optimal value of the regularization parameter. The optimal value of this hyperparameter, λ, is unknown and, therefore, has to be determined from the *training data*. The training data for the Ridge, Lasso and Elastic-Net regressions are the first 18 mailings (when we model amount we restrict to households that responded and when modelling response probability we restrict the sample to those who got a mail). The testing data set, on the other hand, contains the observations that are related to the last (nineteenth) mailing. For each value of the regularization parameter, 10-Fold Cross-Validation splits the training data in ten subsets of (approximately) equal size. Iteratively, for each of the ten subsets, one *fold* or subset will be seen as a *validation* data set which will be used to evaluate the model that is fitted on the data in the remaining nine folds. We make sure that all the observations related to the same household are in the same fold, such that we make use of the correlations between these observations over the different mailings. Besides that, we will use the Mean Squared Prediction Errors (MSPE) to evaluate the performance of the models on the valuation data set for each value of λ. Ultimately, the regularization parameter which is related to the lowest MSPE will be used to fit the model for the training sample (first 18 mailings), and predict the values for the testing data (mailing 19).

For the Elastic-Net regressions, we do not only have the regularization parameter but also a scaling parameter, α. We do not use an efficient way to determine the perfect scaling parameter, but rather use an iterative procedure by looking at the values of $\alpha \in \{0.01, 0.02, ..., 0.99, 1\}$. Having set a fixed value for α, the problem is then similar as for the situations with Ridge and Lasso regressions outlined above, and of which the optimization procedure will be described below. We will store the best MSPE for all these models and use the nineteenth mailing to determine the best value of alpha, by selecting the one with the lowest MSPE.

Finally, we explain how we determine the best value of λ and estimate the regression coefficients (β_0 and β_1 in equation (4)). We consider 100 regularization parameters equally spaced on a log-scale. The question remains, however, how we determine which parameters are considered. To determine the sequence of λ parameters considered, we first find a value λ_{max} for which all regression coefficients are set to zero (Friedman et al., 2010). Furthermore, based on this value, we determine $\lambda_{min} = \epsilon\lambda_{max}$. The typical value for ϵ is 0.001. Then the 100 equally-spaced regularization parameters (on a log-scale) from λ_{max} to λ_{min} are considered by our algorithm, one-for-one in decreasing order. Furthermore, the regression coefficients are estimated by using coordinate descent for a fixed regularization parameter (and potentially scaling parameter). We will leave the details about this algorithm out, but can be found in Friedman et al. (2010).

3.4 Target selection

In this section, we will describe our target selection approach. We motive each approach and the methods used to implement them. We know the target selection technique of the large Dutch charity and our goal is to do at least better than this. More specifically, the charity ordered the households into five groups for the last (nineteenth) mailing. Group 5 contains the households who are expected to donate the most, while group 1 contains households with the lowest expected donation. Therefore, we will also group the households in five groups, while we maintain the same group sizes as those of the charity.

3.4.1 Selection method 1: Expected donations

The first target selection approach is based on expected donations. We ordered households by expected donations in decreasing order, and split them in the five groups. For this approach, we need to predict the probability that an household will respond given that they received a mail, $\mathbb{P}[resp_{it} = 1|mail_{it} = 1]$, and the expected amount donated given that the household responded, $\mathbb{E}[amount_{it}|resp_{it} = 1]$. Then, using the law of total expectation, we get

$$\begin{aligned}\mathbb{E}[amount_{it}|mail_{it} = 1] &= \sum_{k=0}^{1} \mathbb{E}[amount_{it}|resp_{it} = k]\mathbb{P}[resp_{it} = k|mail_{it} = 1] \\ &= \mathbb{E}[amount_{it}|resp_{it} = 1]\mathbb{P}[resp_{it} = 1|mail_{it} = 1]\end{aligned}$$

So, we can use the response probability and amount models that we have described in the previous sections to determine the expected amount donated (given that someone receives a mail).

3.4.2 Selection method 2: Response probability ordering

Furthermore, we also consider a second target selection approach where we order the households based on predicted response probabilities. This approach is motivated by the fact that biased amount predictions may lead to wrong target selections when we order based on expected values. For instance, one may question whether a charity is better off targeting households who are willing to donate a lot when they decide to donate, instead of mailing households who will donate with almost certainty irrespective of the donation amount. Besides that, we also expect some bias in our amount predictions since we use OLS as an estimation technique for censored data (donation amount cannot be negative) which leads to inconsistent estimators. This could have been solved by considering censored regressions, which we do not consider due to time constraints. However, for the interested reader, we performed the Tobit Type I regression in the Appendix.

Finally, for the second target selection technique, we use the response probabilities that are predicted by the response models described in the previous sections. Then, if we find that this approach performs better than the first target selection approach, we might advice the large Dutch charity to target households based on response probability, and to take the predicted donation amounts not into account.

3.5 Brier Score and Mean Squared Prediction Error

In order to evaluate the accuracy of the target selection approaches, we analyse the out-of-sample prediction accuracy. We will do so based on two very similar measures: the Brier Score and the Mean Squared Prediction Error (MSPE).

The MSPE is used to determine the predictive accuracy of models that predict continuous values. Using MSPE as an absolute accuracy measure for one model is not very informative. Together with the MSPE measures of other similar models, however, we can determine which model best predicts out-of-sample, and should be used for target selection. All in all, we consider the MSPE measures for the donation amount models and where we use the last direct (ninteenth) mailing as the testing sample. Finally, the MSPE can be formulated as

$$MSPE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{7}$$

Where, in our situation, N is the total number of observations that receive a direct mailing in the testing sample. $\hat{y}_i$ is the predicted donation amount for household i, and y_i is the actual donated amount.

The Brier Score (BS), on the other hand, is the equivalent of the MSPE for discrete (binary and categorical) variables. Therefore, this measure will be used for the response probability models. Again, we consider the last direct mailing as the testing sample, the other mailings will be used to fit the model. Finally, to be complete, the Brier Score can be formulated as follows

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \tag{8}$$

For our models, f_i, is the predicted response probability of household i for mailing 19. Besides that, o_i, refers to the true value of this variable.

4 Results

In this section, we present the results of our response models and target selections. As mentioned in section 3, we have two main objectives: (1) to model the household responses to direct mailings of a large Dutch charity, and (2) to provide at least a better target selection than the charity. We start discussing the former by presenting the estimated models, where we will mainly focus on parameter interpretation. After that, we will present the results of our target selection techniques and compare those with the charity. Finally, we will evaluate the out-of-sample performance of our models.

4.1 Response modelling

4.1.1 Response probability

We start by presenting the results of the logistic model as discussed in section 3.1. Looking at the p-values of this regression in Table 3, we find that the variables *avamount*, *D_urblvl*, *plus*65, *hhsize*, *highinc*, *lowinc* and *nojob* have an insignificant effect on the response probability. Therefore, we can conclude that the household characteristics at the zip-code level does not have an significant effect on the likelihood to donate. This observation also holds for the variable , *avamount* of the RFM variables. On the other hand, the variable *lastresp* has a statistically significant and negative effect on the response probability. This means that an household who responded to its last received mailing is now less likely to donate. This observation can be explained by budget constraints of households or periodic altruism. Furthermore, the variable *avresp* has a significant and positive effect on the likelihood to donate. This means that an higher response percentage over the last 12 months preceding mailing t increases the response probability.

Finally, the two lagged mailing variables, *mail.lag*1 and *mail.lag*1&2, are also statistically significant. From the sign of the coefficients we can conclude that an household who received the previous mailing is less likely to donate now, given that the household received the current mailing and all the other regressors stay fixed. However, an household who received the previous two mailings has a higher response probability now (i.e., the net effect of $mail.lag1$ and $mail.lag1\&2$ is positive). Based on these results, we can conclude that irritation does not seem to have a negative effect on the response probability. Besides that, the effect of irritation on response probability is significant, which is in not accordance with van Diepen et al. (2009). Note, however, that these coefficients are difficult to interpret. Whether someone received a mailing now or in the past is not exogenous, but determined by the charity. An consequence of this endogeneity might be inconsistent estimates.

Furthermore, we evaluate the results of the Ridge, Lasso and Elastic-Net regressions for the response probability modelling, which are presented in Table 4. We observe that the Lasso regression sets the effect of variables $hhsize$ and $nojob$ to zero. An explanation for this could be given by the high correlation between $nojob$ and $highinc$, and between $hhsize$ and $plus65$, see Table 2. In section 3.3, we mentioned that the Lasso regression chooses one variable from a group of highly correlated variables, and sets the parameters of the remaining variables to zero. Furthermore, an intuition for these high pairwise correlations can be given by the facts that households living in an neighbourhood with an higher percentage of people aged over 65 have a lower chance of having kids still living in their parents house. Also, a neighbourhood with an higher unemployment rate logically has a lower percentage of households with high income.

Moreover, the optimal scale parameter, α, is 1 for the Elastic-Net regression for the response probability model. As mentioned in section 3, this means that the Elastic-Net regression is the same as the Lasso regression. Therefore, we can conclude that the Lasso regression has selected the best variables from the groups of highly correlated variables, otherwise the results for the Lasso and Elastic-Net regression would have been different. For the same reason, we can conclude that the characteristic of the Ridge regression never setting coefficients to zero is not beneficial from a model-fitting perspective. Finally, the interpretation provided in the first paragraph for the logistic regression also applies to the Ridge, Lasso and Elastic-Net regression, since sign and magnitude are very similar for all coefficients.

4.1.2 Donation amount

Next, we look at the results of the OLS regression used to estimate the donation amount presented in Table 3. We observe that all zip-code characteristics are not significant, except the urbanization dummy variable, D_urblvl, and the percentage of households with high income in zip-code area, $highinc$. More specifically, given that an household

Table 3 Regression table of response probability model estimated by logistic regression and amount model fitted with OLS

Model	**Logistic**	**OLS**
Dependent Variable	*resp*	*amount*
lastresp	-0.196^{***}	1.477^{*}
	(0.036)	(0.619)
avresp	0.030^{***}	-0.123^{***}
	(0.001)	(0.009)
avamount	-0.001	0.899^{***}
	(0.001)	(0.010)
D_urblvl	0.015	1.325^{**}
	(0.029)	(0.459)
hhsize	0.067	-1.081
	(0.093)	(1.476)
*plus*65	0.004	-0.007
	(0.003)	(0.049)
highinc	0.0004	0.139^{*}
	(0.004)	(0.056)
lowinc	-0.005	0.132
	(0.005)	(0.080)
nojob	0.002	-0.055
	(0.003)	(0.050)
*mail.lag*1	-0.280^{***}	-0.792
	(0.039)	(0.658)
*mail.lag*1&2	0.415^{***}	-0.392
	(0.035)	(0.599)
Constant	-1.935^{***}	7.218
	(0.345)	(5.445)
Observations	34783	12776
R^2		0.383
Adjusted R^2		0.382
Log Likelihood	-19048.220	
Akaike Inf. Crit.	38120.440	
Residual Std. Error		22.390
F Statistic		719.276^{***}
Note:	$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$	

has decide to donate and the other regressors stay fixed, households donate more when they live in a high-urbanized area or in a high-income neighbourhood. For instance, an one percent increase in the number of households with a high income in a zip-code area, corresponds with an increase of 0.139 Dutch Guilders in amount donated.

Furthermore, while the monetairy RFM variable, *avamount*, has an insignificant effect on the response probability, it significantly and positively affects the amount that households donate to charity. Also the remaining two RFM-variables are significant. Nevertheless, these RFM variables have an opposite effect on the donation amount, compared to their effect on the response probability (see Table 3). For instance, an household that responded to its previous mailing will now donate 1.477 Dutch Guilders more. Also, an household that responded one percent more often to its mailings in the last 12 months, will now donate 0.123 Dutch Guilders less. Note that these interpretations only hold for those households that already decided to donate based on the current direct mailing and only have to decide how much they will donate, this follows from how we constructed our estimation samples.

Besides that, irritation – measured by the lagged mailing variables – does not significantly affect the amount that households donate, given that they decide to respond. This is in accordance with van Diepen et al. (2009), where the authors conclude that irritation does not affect donating behaviour.

Finally, we will interpret the results for the Ridge, Lasso and Elastic-Net regression for estimating the donation amount presented in Table 4. We observe that the Lasso regression does not exclude any regressor variables, contrary to the Lasso regression for the response probability model. This indicates that all variables are significant enough to be included in the model. Moreover, when tuning the optimal scaling parameter, α, for the Elastic-Net model, we end up with a parameter equal to 0. This means that Ridge regression seems to fit best for the donation amount modelling. Lastly, the interpretation provided above for the OLS regression also applies to the Ridge, Lasso and Elastic-Net regression, since sign and magnitude are very similar for all coefficients.

4.2 Target selection

In this section we present the results for the expected donation and response probability ordering target selection techniques. We perform both target selection techniques in order to beat the selection of the charity as presented in Table 5. Group 5 contains households who are expected to donate the most, while group 1 contains households with the lowest expected donation. Besides that, for a fair comparison, all our target selections use

Table 4 Estimated coefficients Ridge, Lasso, and Elastic-Net regressions

	Response Probability			**Donation Amount**		
Method	Ridge	Lasso	Elastic-net	Ridge	Lasso	Elastic-Net
Intercept	-1.835	-1.794	-1.795	9.461	8.568	9.461
lastresp	0.096	-0.166	-0.163	0.494	1.191	0.494
avresp	0.024	0.029	0.029	-0.102	-0.119	-0.102
avamount	0.000	-0.001	-0.001	0.842	0.896	0.842
D_urblvl	0.016	0.009	0.009	1.242	1.115	1.242
hhsize	0.043			-0.793	-0.469	-0.793
*plus*65	0.003	0.002	0.002	0.013	0.007	0.013
highinc	0.001	0.001	0.001	0.101	0.096	0.101
lowinc	-0.002	-0.002	-0.002	0.056	0.055	0.056
nojob	0.001			-0.032	-0.022	-0.32
*mail.lag*1	-0.178	-0.249	-0.247	-0.784	-0.751	-0.784
*mail.lag*1&2	0.355	0.390	0.388	-0.143	-0.355	-0.143

the same group sizes as the charity. Furthermore, The average response percentage and average donated amount for mailing 19 over all the households is, respectively, 33.4% and 6.4056 Dutch Guilders.

Table 5 Average response (in %) and donation amounts (in Dutch Guilders) for groups according to the Charity's ordering of households.

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Percentage Response	28.443	29.469	20.000	30.989	48.895
Average Donation	6.171	4.462	3.384	4.382	10.372
Group size	1227	1035	835	455	1448

In Table 6, we observe that ordering based on expected donations creates a fifth group that donates significantly more than group 5 in Table 5, for all regression methods that we consider. Especially for the Elastic-Net regressions, we observe that households in group 5 donate on average 4.523 Dutch Guilders more than the benchmark. But also for the other regression methods similar results hold. Moreover, the decrease in average donations and response percentage from group 5 to group 4, and for the same transition to other lower groups, indicates that our target selection techniques are better in targeting the households that are expected to donate the most, compared to the target selection of the charity.

Table 6 Average response (in %) and donation amounts (in Dutch Guilders) for groups selected according to expected value ordering.

		Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Logistic/OLS	Percentage Response	11.084	11.884	45.509	51.429	55.041
	Average Donation	1.898	1.966	4.201	5.809	14.857
Ridge	Percentage Response	10.595	12.271	45.150	53.626	54.696
	Average Donation	2.003	1.834	4.178	5.998	14.816
Lasso	Percentage Response	11.084	11.981	45.030	51.868	55.111
	Average Donation	1.943	1.912	4.146	5.868	14.871
Elastic-Net	Percentage Response	10.595	12.657	44.192	51.429	55.663
	Average Donation	2.015	1.834	4.095	5.870	14.895

From Table 7, it can be seen that ordering based on probability yields similar results as the ordering based on expected donation. A notable difference between these techniques, however, is that group 5 in the target selection on probability ordering is significantly more likely to donate, which holds for all regression methods. However, from group 4 onwards, this extra response probability decreases and groups 1 to 4 have lower probability of donating compared to the selection based on expected donations. Response probability targeting clusters the households that are likely to donate in group 5, hence this method is useful in situations where the charity desires to sent out a small batch of mailings to ensure more certain but smaller donations from its donors.

4.3 Comparing Forecasts

In this section, we will use the Brier score and mean squared prediction error (MSPE) to evaluate the models based on their predictions for mailing 19. The Brier score is used to evaluate the out-of-sample prediction of response probability, hence is determined over all 5000 households that received mailing 19. The MSPE, on the other hand, is calculated over the observations that have responded to the mailing, hence is only be calculated over a subset of the 5000 households.

Table 7 Average response (in %) and donation amounts (in Dutch Guilders) for groups selected according to response probability ordering.

		Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Logistic/OLS	Percentage Response	9.943	10.918	40.120	45.714	61.602
	Average Donation	1.849	2.313	8.216	9.092	11.304
Ridge	Percentage Response	10.024	12.560	40.240	41.099	61.740
	Average Donation	1.841	2.397	8.200	8.303	11.508
Lasso	Percentage Response	9.943	10.918	40.240	43.736	62.155
	Average Donation	1.817	2.342	8.182	8.919	11.385
Elastic-Net	Percentage Response	9.943	10.918	40.240	43.736	62.155
	Average Donation	1.817	2.342	8.182	8.919	11.385

Looking at the Brier score, we can see that the Elastic-Net, Lasso and logistic regression perform almost equally well and slightly outperform Ridge regression. Logistic regression just slightly outperforms Lasso and Elastic-Net. For the MSPE, it can be seen that Elastic-Net and Ridge regression perform equally well, due to the scaling hyper-parameter α having the optimal value of 0, and outperform both the linear and Lasso regression.

Based on the results in Table 8, we conclude that Elastic-Net and Ridge regression outperform the other models for donation amount modelling, and logistic regression just slightly outperforms the other models for response probability modelling.

Table 8 Performance evaluation response probability and donation amount models.

	Brier score	**MSPE**
Logistic Regression	0.1742799	
Linear Regression		231.5392
Ridge Regression	0.1751717	217.9859
Lasso Regression	0.1742824	230.6566
Elastic-Net Regression	0.1742824	217.9859

5 Conclusion

In this paper, we focus on the direct mailing of a large Dutch charity and pursue two main objectives: (1) modelling households responses to the direct mailings of the charity, and (2) providing at least a better target selection than the charity with sound econometric techniques. We approach the first objective by modelling the response probability of households and the donation amount using logistic regressions and OLS, respectively. Moreover, we explore the advanced model fitting techniques Ridge, Lasso and Elastic-Net regressions. For the second objective, we develop two selection techniques. The first approach is based on the law of total expectation and the second on response probability ordering. These approaches involve ordering households and dividing them into five groups, mirroring the group sizes maintained by the charity.

Our first main result is that all the above-mentioned models provide a better target selection than the large Dutch charity, irrespective of whether we order households based on expected donations or response probabilities. However, we would advice the charity to use the Elastic-Net regression for our target selection techniques. Besides that, the Elastic-Net regressions also provide more accurate out-of-sample predictions, which follows from the lowest Brier Score and Mean Squared Prediction Error.[3]

Besides that, we conclude that the target selection based on expected donations outperforms the response probability ordering approach across all regression methods. This is motivated by the observation that the latter approach is better able to group the households that will donate the most (in group 5).

Our final conclusion is that irritation has a statistically insignificant effect on donation amount, as measured by the two lagged mailing variables in the models. This is in accordance with the result of van Diepen et al. (2009), where the authors conclude that irritation does not affect donating behaviour. Nevertheless, irritation does have a significant effect on the response probabilities, and its effect is (mixed) positive. These results should, however, be interpreted with care. First, the models are constructed with an estimation sample consisting either of households who received the current mailing (in the response probability models) or households that did respond to the current mailing (in the amount models). Second, the charity has implemented its own target selection, which is not exogenous. Therefore, these two points together might indicate that our are inconsistent by endogeneity. For instance, one could argue that households who received the previous two mailings and the current mailing, are people that the charity believes

[3]In fact, the logistic regression attains the lowest Brier Score. The Brier Score for the Elastic-Net regression is, however, less than 0.002 percent larger.

are likely to donate and not influenced much by irritation.

Finally, we acknowledge that future research can explore additional machine learning techniques, such as Random Forest, for response modelling. While we have made progress in implementing this approach, time constraints prevented its completion. Furthermore, in light of possible endogeneity, the use of censored regression models could be further investigated.

References

Diepen, M. V., Donkers, B., & Fransens, P. (2009). Dynamic and competitive effects of direct mailings: A charitable giving application. *Jopurnal of Marketing Research*, *46*(1), 120–133. https://doi.org/10.1509/jmkr.46.1.120

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Otter, P., der Scheer, H. V., & Wansbeek, T. (1999). Direct mail selection by joint modeling of the probability and quantity of response. *IFCA Proccedings Volumes*, *32*(2), 6115–6120. https://doi.org/10.1016/S1474-6670(17)57043-3

van Diepen, M., Donkers, B., & Franses, P. H. (2009). Does irritation induced by charitable direct mailings reduce donations? *International Journal of Research in Marketing*, *26*(3), 180–188. https://doi.org/https://doi.org/10.1016/j.ijresmar.2009.03.007

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *67*(2), 301–320. Retrieved May 14, 2023, from http://www.jstor.org/stable/3647580

6 Appendix

6.1 Dimension reduction using Principal Component Analysis

Here, we discuss the implementation of PCA. First, we used the data on the first 18 mailings and the last nine variables in Table 1, to construct the correlation matrix used for eigendecompensition.[3] Then we construct the principal components such that over 90 percent of the variance is explained. In turn, these principal components are be used as independent variables in the models described in section 3.2 and section 3.1 to determine the amount model and the response probability model, respectively. Finally, these models can be used together with the estimated principal components for mailing 19 to predict the response probability and amount donated. The results are presented below.

6.2 PCA Results

Table 9 Target Selection methods PCA

		Group 1	Group 2	Group 3	Group 4	Group 5
Target Selection 1: Expected Value	Percentage Response	10.269	12.657	45.030	54.286	54.558
	Average Donation	1.852	1.996	4.269	5.927	14.798
Target Selection 2: Response Probability	Percentage Response	10.024	12.657	44.192	37.582	60.497
	Average Donation	1.784	2.322	7.933	9.327	11.442

Table 10 Principal Components and Eigenvalues

PCs	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eigenvalues	2.877	1.906	1.507	0.986	0.852	0.291	0.254	0.235	0.094
lastresp	0.0175	-0.634	0.153	-0.075	-0.307	0.018	0.036	-0.687	-0.002
avresp	0.021	-0.655	0.144	-0.054	-0.172	-0.012	-0.030	0.718	0.003
avamount	0.024	-0.370	-0.057	0.229	0.893	0.030	0.006	-0.094	0.003
urblvl	-0.250	-0.059	-0.238	0.803	-0.234	-0.110	-0.401	-0.012	-0.065
hhsize	0.374	0.097	0.531	0.033	0.074	-0.596	-0.408	-0.036	0.197
$plus65$	-0.276	-0.107	-0.524	-0.504	0.066	-0.247	-0.517	-0.034	0.224
highinc	0.454	-0.070	-0.441	-0.021	-0.048	-0.455	0.196	0.002	-0.588
lowinc	-0.484	0.048	0.378	-0.184	0.108	-0.031	-0.237	-0.012	-0.720
nojob	-0.529	-0.016	0.046	0.063	0.011	-0.601	0.556	0.012	0.207

[4]The variables are: $lastresp_{it}$, $avresp_{it}$, $avamount_{it}$, $urblvl_{it}$, $hhsize_{it}$, $plus65_{it}$, $lowinc_{it}$, $highinc_{it}$, and $nojob_{it}$.

Table 11 Logistic response probability model and linear donation amount model using regressors estimated with PCA

	Dependent variable:	
	resp	*amount*
	(Logistic)	(OLS)
PC1	0.026***	−0.098
	(0.007)	(0.096)
PC2	−0.625***	2.053***
	(0.008)	(0.121)
PC3	0.112***	3.887***
	(0.009)	(0.132)
PC4	−0.050***	−10.511***
	(0.011)	(0.161)
PC5	−0.130***	−13.145***
	(0.013)	(0.173)
Constant	−0.525***	18.315***
	(0.011)	(0.163)
Observations	40,977	15,924
R^2		0.412
Adjusted R^2		0.412
Log Likelihood	−23,954.850	
Akaike Inf. Crit.	47,921.700	
Residual Std. Error		20.579
F Statistic		2,234.288***
Note:	*p<0.05; **p<0.01; ***p<0.001	

6.3 Censored Regression: Tobit

Table 12 Regression table of amount

	Dependent Variable
	Amount (Tobit)
Intercept	1.321
lastresp	0.494-1.054
avresp	0.055***
avamount	0.331***
D_urblvl	0.707
hhsize	-0.533
*plus*65	0.006
highinc	0.036
lowinc	-0.010
nojob	-0.005
*mail.lag*1	-1.91*
*mail.lag*1&2	0.359
Observations	12776
Log Likelihood	-79324.29
Akaike Inf. Crit.	4.562
S.E. of Regression	21.831
Log Lik.	-79324.29
Note:	*p<0.05; **p<0.01; ***p<0.001