Modeling the gift in fundraising process: A parametric approach for the donations' count

Luca Barzanti[1] and Martina Nardon[2]

[1] Department of Mathematics, University of Bologna,
Piazza di Porta San Donato 5, 40126 Bologna, Italy
luca.barzanti@unibo.it
[2] Department of Economics, Ca' Foscari University of Venice,
Cannaregio 873, 30121 Venezia, Italy
mnardon@unive.it

Abstract. In fundraising management, the availability of accurate estimates of the expected gift is crucial to implement a successful fundraising campaign. To this aim, assessing the gift process is relevant. In this contribution, we first suggest modeling the gift as an individual risk which can be seen from different perspectives: occurrence, number and timing of donations, and the gift's amount. Then, we focus on one of these aspects; in particular, we model the number of donations as a Poisson random variable with an intensity parameter that depends on the individual characteristics of the Donors. The expected number of donations and the probability of gift can be estimated by performing a Poisson regression.

Keywords. Fundraising Management, Gift Probability, Poisson Regression.

M.S.C. classification. 62P20, 91B74.
J.E.L. classification. D64.

1 Introduction

The fundraising (FR) activity is focused on how to effectively contact potential Donors to raise funds for a particular purpose. In this process, the efficient use of information on Donors and past campaigns plays a central role. As regards the available data, Associations are classified according to the existence of a structured database (DB) and the presence in the DB of specific qualitative information of Donors' profiles (like personal interests and attitudes, and relationship network), in addition to the usual information on the gifts (gift history), and the typical personal profile data. Normally, this classification strictly depends on the Organization's size.

MATHEMATICAL METHODS IN ECONOMICS AND FINANCE – m²ef
Vol. 17/18, No. 1, 2022/23 – Published online on June 5, 2024.

In FR management, estimating the gift probability and the expected gift amount is a prominent issue in implementing a successful campaign. Problems regarding managing the gift in FR will provide a framework for the discussion of various models and examples. The gift is a process that can be analyzed from different perspectives: its occurrence, the number and timing of donations, and the donations's amount. Each one of these aspects can be modeled and estimated using quantitative methods.

Recent approaches to FR are characterized by a significant use of mathematical modeling and soft computing. [4] introduce the use of mathematical modeling and Decision Support Systems (DSS) to help Associations decide the kind of campaign and the features to implement, and choose the Donors to be contacted to maximize the expected return of the campaign, satisfying time and budget constraints. The approach has been specialized for different kinds of Organizations. On one hand, [5] and [9] consider large-sized Associations, with lists of millions of Donors and a powerful organizational system requiring a very sophisticated DSS. On the other hand, [6] consider small-sized Organizations and develop a DSS based only on essential information without an organized DB. The approach has been validated both in the operational world by Associations that test it (as documented in [5], [6] and [9]), and in the literature (see [23] and [19]). Medium-sized Organizations are addressed in [8] and in [7], where a targeted DSS has been developed.

Relevant features, such as the expected gift amount and the probability of donation, can be assessed using parametric and non-parametric approaches. A very recent research stream for FR is based on non-parametric Machine Learning (ML) models; see, for example, [15] and [11]. Along this line, [2] propose a Multi-Layer Perceptron (MLP) to predict the number of donations and the gift amount. [3] extend the analysis, focusing on the relative importance of the input variables in the MLP model with the aim of enhancing the effectiveness of FR campaigns.

As regards parametric approaches, [10] consider the gift as an individual risk and, in particular, the authors suggest modeling the number of gifts as a Poisson random variable with an intensity parameter that depends on Donors' characteristics. In this contribution, we extend the previous work [10], including a quantitative analysis. A Poisson regression is performed to estimate the expected number of donations, the probability of gift, and to assign a score to each Donor measuring their propensity to the donation.

The remainder of the paper is organized as follows. Section 2 discusses the importance of collecting Donors' information. In Sections 3 and 4, the gift is modeled introducing a suitable probability distribution; in particular, we adopt a Poisson model for the number of donations. In Section 5, the data set is described, and in Section 6, the result of the Poisson regression are presented. Finally, Section 7 concludes with some remarks.

2 Donors' characteristics and the giving pyramid

In the pursue of their mission, Associations adopt strategies specifically designed to reach the goal of a FR campaign (see [21]) and maximize the expected gift. In this activity, the position of the Donor is of central importance (see [14] and [18]), and particularly the search for potential Donors' profiles (Contacts) that match some specific gift propensities (see e.g. [13]) in order to support the effectiveness of the FR process. Practitioners claim that most of the success of a FR campaign is determined by choosing the appropriate target of Donors to whom the strategy is addressed, while motivations and creativity are not predominant factors.

Economists agree that information on Donors plays a crucial role to achieve the improvement of the FR strategies [20]. Studies have evidenced some main factors that influence individuals in the choice of donating. In [1] the economic and social foundations of altruism is characterized by factors such as the own community or the social network, and the so called "enlightened self-interest". These features are also studied by [14] and [22]. According to [18], an individual tends to assume a role-identity as Donor, that depends on their network of social relationship. The authors identify some variables that can affect role-identity, shape individual preferences and attitudes, and impact on the utility people get from their decision on how and to what extent donate [12].

Several factors should be considered when implementing FR strategies: information on Donors, information on past campaigns, and operational knowledge and rules of thumb of the experts in the field.

Regarding Donors characteristics to be included in the analysis, these can be further summarized in some categories: personal situation variables (gender, age, number of children, educational level, place of origin, size of residence town, etc.); financial situation variables (wage, wealth, investments, debts); risk aversion variables (the number of insurance policies subscribed by the individual is usually adopted as a proxy); other personal information (personal interests, religious involvement, social network, etc.).

All these data may not be accessible, also because in some cases they are sensitive personal information. When available, information is usually managed by an organized DB (exceptions may include small Organizations). However, classical tools for DB management have some limitations, and decisions are in practice supported by knowledge of experts in the field (see [13]). There is indeed an increasing interest both by Associations and providers of software and services in the FR sector to develop new soft-computing tools able to elaborate more efficiently the available amount of data.

Some of the above mentioned factors may have a strong influence on the gift in terms of amount and frequency. Donors can also be classified according to how often and how much they donate. To this respect, the Donors' segmentation is determined by the *giving pyramid*. An example is represented in Fig. 1, where the ground of the pyramid is constituted by the Contacts. A Donor who donates occasionally (once or very few times) small amounts is called "*Sporadic*"; a Donor who gives more frequently and/or their donations are more generous is called "*Regular*"; large and normally regular donations are associated to the so-called "*Large*" Donors; finally, on top of the pyramid there are inheritances and legacies (Donors that may be particularly involved in the Association's mission). A finer classification is possible, depending on the characteristics and size of the DB, and the type of FR activity.

Organizations manage an amount of information about past campaigns and, for each Donor in the DB, the gift history is recorded. This allows to construct a giving pyramid based on past history, and update the pyramid layers based on the behavior of the Donors along time and across different campaigns. The objective is to let more Contacts

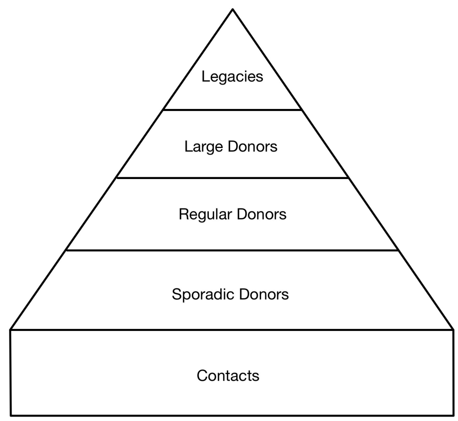

Fig. 1. The *giving pyramid* in FR management representing the segmentation of the Donors

become Donors, and stabilize sporadic Donors' to become regular ones. As getting in touch with a Donor implies some costs, one main aim of FR management is to select the most promising Donors/Contacts in order to maximize the expected return, subject to budget constraints, and, at the same time, to control the return variability.

The percentage of positive responses from Donors, for every level in the pyramid, is usually considered as a parameter estimated by the experts' knowledge. It would be useful to obtain accurate estimates of the probability of gift, based on the available information on Donors, and their gift history. To this aim, in the next section, the FR process will be formalized adopting suitable probabilistic models and specific assumptions will be made about the distribution of the quantities of interest.

3 Modeling the gift

As discussed in the previous section, the appropriate use of the information about Donors is crucial for the accuracy of the expected gift estimation and, in the end, for optimizing the resources. The task of integrating the information on Donors to find an optimal FR strategy is complex; not only does it require a clear identification of the goals, but also a rigorous definition of the variable included in the analysis.

Besides the importance of systemically collecting and updating the data, we stress the need of introducing quantitative tool to exploit such information. To this purpose, the 'gift' can be modeled as an *individual risk*, in analogy with other main domains of applications (see [16]): finance, credit risk, insurance, and marketing.

Hence, the gift can be viewed from four viewpoints:

- the *occurrence* of a donation: the outcome is either 'yes' or 'no';
- the *frequency* or *count* of the donations received in a period of time (a year or the duration of the campaign): the number of gifts is zero or any positive integer;

- the *timing* or *duration* (when a donation has occurred or the interval between donations): the outcome is an interval of time, usually measured with reference to a fixed point of origin, such as the beginning of the campaign or when the potential Donor has been contacted for the first time;
- the *gift amount*: the amount of money given by the Donor for each donation (the outcome is measured in currency units, but could be also in terms of hours).

With respect to all these perspectives, the gift is quantifiable and can be modeled using statistical methodology, determining for any aspect listed above which kind of random variable can be adopted: a dichotomous variable, a count variable, a duration variable, and a continuous positive variable, respectively.

Either dichotomous or count variables can be used to model the occurrence of the gift event. Let us simply consider a dichotomous random variable[1] Y; then the probability of gift is equal to $\mathbb{E}(Y) = p$. Let X be a continuous random variable that represents the amount of money given by the Donor for a donation, or the total gift of all donations filed in the considered period. In this case the expected gift for each Donor can be computed by the product of the gift probability and expected gift amount, $\mathbb{E}(Y)\mathbb{E}(X)$. Considering the whole campaign, both the number of gifts and the gift amount are random, hence campaign's return can be modeled as a *random sum*; in order to compute its expectation, some assumptions need to be introduced (such as independence amongst Donors, and independence of gifts count and gift amounts). All these features can be modeled in alternative ways; in Section 4 we suggest a model for the number of gifts considering a single Donor.

We make some assumptions about the mechanism that gives rise to the gift: any gift is associated with an individual i, the Donor; a Donor can be a person, a company, or other entity that can be represented by some individual characteristics which are collected in a data set; the individual characteristics of the Donor are synthesized by a *score*; the gift history (gift events, timing and gift amounts) of the Donor is recorded.

A *score* is a statistical measure of individual risk based on individual characteristics [16]. In the context of FR, it can be used to quantify the individual propensity to donate (the higher the score, the higher the propensity to the gift), to rank Donors in a population, to distinguish between (expected) "good" and "bad" Donors. This latter procedure is called *segmentation* and in FR could be used to distinguish potential Contacts or to address *ad hoc* advertising to subclasses of Donors.

Let x_i be the vector which collects selected observable, qualitative and quantitative, individual characteristics of Donor i, in a sample of n Donors. Define z_i as the vector of transformed individual characteristics (where qualitative features are properly transformed into quantitative or dummy variables). The score can be defined as a scalar function of covariates $z_i'\theta$, where θ is a vector of parameters. The score, which summarizes the information about the Donor, can be determined by more sophisticated approaches (see [16]).

In the next section, we focus on one of the aspects related to the gift process. In particular, we consider the number of donations and let depend them on Donors' individual characteristics.

[1]Formally, denoting with G the gift/donation event, we have $Y = \mathbf{1}_G(\omega)$, where $\mathbf{1}_G$ is the indicator function of G, with $\mathbb{P}[Y = 1] = p$.

4 Poisson regression in FR

The arrival of a donation to an Association, such as a new claim to an Insurer, can be viewed as the outcome of a random variable. In a very simple model, a dichotomous variable indicates whether or not a gift is received. In this contribution, we suggest to use a parametric approach for the number of gifts (in a certain period), in analogy with insurance theory where count variables are used to model, for example, the number of claims on one policy in a year. We consider a model, the Poisson distribution, that is usually adopted as a starting approach for count variables; we apply it for the number of gifts, which can then be estimated by the *Poisson regression model*.

Let Y represent the number of gifts in a unit of time; in a basic count variable model, we assume that Y has a Poisson distribution with intensity parameter λ. It is well known that $\mathbb{E}(Y) = \lambda$, which is equal to its variance $\mathbb{V}(Y) = \lambda$.

In the Poisson regression model, λ depends on the values of observable characteristics x_i of each individual or entity i. As the intensity varies across individuals, its specification for Donor i will be

$$\lambda_i = \exp(z_i'\theta), \tag{1}$$

where θ is the vector of unknown parameters and z_i is a vector of transformed individual characteristics; the exponential form ensures positivity of the intensity.

Let us consider a sample of n Donors; the gift count variables $Y_1, \ldots, Y_n$ in this model are independent, conditional on the covariates, and the conditional distribution of Y_i is a Poisson distribution with parameter λ_i as in (1). It is worth noting that

$$\mathbb{E}[Y_i|x_i] = \mathbb{V}[Y_i|x_i] = \exp(z_i'\theta); \tag{2}$$

it turns out that the model is heteroskedastic by construction. Parameters θ can be estimated by maximum likelihood; the resulting log-likelihood function is[2]:

$$\begin{aligned} l(\theta) &= \ln\left\{\prod_{i=1}^{n}\left[\exp\left(-\exp(z_i'\theta)\right)\frac{\exp(y_i z_i'\theta)}{y_i!}\right]\right\} \\ &= \sum_{i=1}^{n}\left[y_i z_i'\theta - \exp(z_i'\theta) - \ln(y_i!)\right]. \end{aligned} \tag{3}$$

$l(\theta)$ is concave with respect to θ; the maximum likelihood estimator $\hat{\theta}_n$ is obtained imposing first-order conditions:

$$\frac{\partial l(\hat{\theta}_n)}{\partial \theta} = 0 \qquad \Leftrightarrow \qquad \sum_{i=1}^{n}\left[y_i - \exp(z_i'\hat{\theta}_n)\right] z_i = 0\,. \tag{4}$$

The residuals associated with Donor i are $\hat{u}_i = y_i - \hat{\lambda}_i$, and conditions (4) are equivalent to the orthogonality conditions for residuals and variable z_i.

[2] See also [16] for details and properties of the estimators.

Once estimated, the model can be used to compute the expected number of gifts for a single Donor (or a new Contact), λ_i, and the probability of gift

$$\mathbb{P}[Y_i = y] = \exp(-\lambda_i)\frac{\lambda_i^y}{y!}, \qquad y = 0, 1, 2, \ldots \tag{5}$$

As a further result, once the model is estimated, it allows to obtain, for each Donor in the DB or for new potential Donors, a score $z_i'\theta$. Such an indicator can also be used for rating Donors with respect to their propensity to the gift; the higher the score is, then the higher the expected number of gifts λ_i, which indicates "good" Donors.

Poisson regression model is easy to interpret; a possible drawback is that it is based on some strong assumptions. Nevertheless, the model allows for various extensions. For instance, a gamma distributed *heterogeneity factor* can be introduced; as a result, one obtains a *negative-binomial* model [16].

Here the Poisson model is adopted as a first approach for FR, in a fieled where the use of more sophisticated models is not yet well developed, registering at the same time an increasing interest for quantitative approaches and artificial intelligence. This basic specification for count variables allowed us to establish a relationship between the count variable risk models and the models based on dichotomous qualitative variables linked to Donors' individual characteristics. In the next section, we describe the information available on Donors' for a medium Organization, while in Section 6 we present an application.

5 Description of the data

The numerical analysis in Section 6 is based on a simulated DB, already used in other contributions in the literature ([2] and [3]), constructed from experts' knowledge, and based on a realistic composition of a set of Donors.

Starting with about 400 000 Contacts, a set of $N = 30\,000$ Donors is obtained. These values constitute medium to high numbers for a medium-sized Organization, or high numbers for a small-sized Organization. In the set of Donors, 75 % are *Sporadic Donors* (labeled 'sd'). Among them, about 25 % made only one donations (labeled 'sd1'), and the rest made more than one donation (labeled 'sd2'). The remaining 25 % are: 19 % *Regular Donors*[3] (labeled 'rd'), and 6 % *Large Donors*. Legacies are not present in the considered sample.

Besides information about gift history of the Donor, other personal profile variables collected are: age and number of children, educational level[4] (in four categories: Master and Ph.D., Bachelor, High School, other/lower school level), wealth (measured in thousands of euro), risk aversion (measured as numbers of insurance policies signed by the Donor).

[3] A further subdivision in "stable" (labeled rd1) and "dynamic" (labeled rd2) is possible.

[4] Categorical variable transformed into values ranging from 1 to 4, assigning 4 to the highest category.

Regarding the gift history, the dataset includes for each Donor: the number of donations, the gift amount for each donation[5], and the number of gift requests (or also number of times when the Donor searched for information about the FR campaign).

Table 1. Distribution of some Donors' individual characteristics along the *giving pyramid*

Donors	low wealth	n. risks ≥ 1	min gift amount	max gift amount
col2-col5				
Sporadic (sd1)	70 %	35 %	20	50
Sporadic (sd2)	70 %	35 %	30	100
Regular (rd1)	40 %	65 %	50	400
Regular (rd2)	40 %	65 %	100	500
Large	10 %	65 %	300	1000

Table 2. Main statistics for the gift history (number and amount of donations, number of donation requests), and Donors' individual characteristics

	mean	std. dev.	min	max
n. donations	6.40	5.20	1	28
gift amount	133.65	158.20	20	1000
gift requests	15.10	8.37	1	29
age	53.43	20.86	18	89
n. children	1.50	1.12	0	3
education	2.51	1.12	1	4
wealth	398.47	310.17	10	1000
n. risks	1.07	1.67	0	5

Tables 1 and 2 report a synthesis of the data collected in the DB. In particular, Table 1 shows the composition (segmentation) of the Donors population in the giving pyramid related to some characteristics. About 70 % of the Sporadic Donors have "low wealth"; whereas, such a percentage decreases to about 40 % and 10 % for Regular Donors and Large ones, respectively. In the second column, the percentage of Donors who subscribed at least one insurance contract is reported; it can be observed that the number increases when considering higher layers of the pyramid. In the last two columns, the minimum and maximum Donation amounts are shown; in this case, results depend on

[5] The average donation for each Donor is used in the analysis.

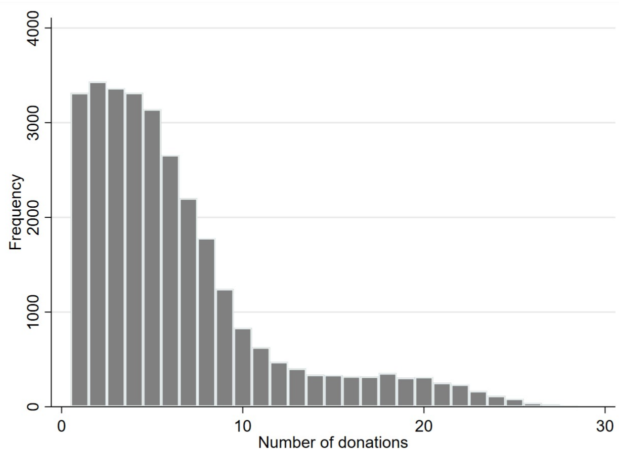

Fig. 2. Empirical distribution of the number of donations

the very definition of Sporadic (low gift amount, low frequency), Regular (low/medium gift amount, medium/high frequency), and Large (higher gift amount) Donors.

Table 2 reports the main statistics for the gift history (number of donations, amounts, number of requests), and some Donor's individual characteristics (age, number of children, educational level, wealth, and risk aversion).

The empirical distribution of the number of donations is shown in Fig. 2. It is worth noting that, as we considered a sample of Donors, the number of donations range from 1 to the maximum observed number. This choice allows us to avoid the inference issues associated with the excess of zeros that arise when considering all the Contacts in the DB.

6 Application and results

Information described in the previous section can be thought of as realization of a process that resembles those data to be modeled. The Poisson regression is the basic count variable model for individual risk. It is easy to estimate and interpret, but it relies on some strong simplifying assumptions as well.

Various problems arise when trying to apply Poisson model. For instance, the Poisson distribution assumes the possibility of zero counts, but in practice there may not be any. When considering the number of donations, we focus on the gift history of Donors' already present in the DB, excluding Contacts who have not donated yet. Hence, zero donations are not a possibility for the data being modeled, as shown in Fig. 2. On the other hand, including in the analysis the information on Contacts would lead to another problem: the excess of zeros. In these cases, the underlying distribution may need to be adjusted to take into consideration or exclude zero counts[6]. The model could be

[6] More advanced approaches include two-part hurdle models and mixtures models (see [17] for a discussion).

amended considering, for example, a truncated distribution. In Section 6.1, we start considering the standard Poisson regression, allowing for zero counts; in Section 6.4 we address this issue.

Another problem is over-dispersion. Theoretically, the mean and the variance have the same value for a Poisson distribution. In practice, one observe data with larger variance. Considering data reported in Table 2, it is evident that the variance is much larger than the observed mean. Over-dispersion occurs also when observed and predicted variances of the response differ. In Section 6.3, we will test over-dispersion and discuss how to treat this problem.

6.1 Choice of profile's variables

The Donor's individual features to be used in the regression model can be divided into: personal profile variables (age, number of children, educational level), risk aversion variable (the number of insured risks), and economic situation (wealth measured in thousand of monetary units). Besides these variables, the analysis takes into consideration information about gift history, namely, the average donation amount for each Donor, and the number of gift requests.

We first run a Poisson regression taking into consideration all the listed variables (including a constant term); predictors with $p-$values less than the generally acceptable level of 0.05 appear to significantly contribute to explain the number of donations. Two information, namely the age and number of children, turned out not to be significant and, also on the basis of information criteria (AIC and BIC), have been excluded. The results reported in Table 3 are those of the reduced form Poisson regression model.

Table 3. Results of the Poisson regression for the number of donations when gift amount, number of risks, wealth, number of gift requests and the two highest educational levels are considered as explanatory variables. Pseudo $R^2 = 0.3730$. LR $\chi^2(6) = 79381.67$; $Prob > \chi^2 = 0.0000$. Number of observations = 30000. Residual degrees of freedom df = 29993; (1/df)Deviance = 1.021245; (1/df)Pearson = 1.036365.

	coefficient	std. err.	z	$P > \|z\|$	95% conf. interval
gift amount	0.0018902	0.0000113	166.69	0.000	[0.0018680, 0.0019124]
n. risks	0.0216880	0.0013268	16.35	0.000	[0.0190875, 0.0242886]
wealth	0.0000377	0.0000077	4.92	0.000	[0.0000227, 0.0000527]
gift requests	0.0627460	0.0002960	211.99	0.000	[0.0621659, 0.0633261]
education3	0.0163401	0.0055632	2.94	0.003	[0.0054365, 0.0272437]
education4	0.0167768	0.0055933	3.00	0.003	[0.0058142, 0.0277394]
const.	0.4106476	0.0074735	54.95	0.000	[0.3959998, 0.4252954]

Regarding the educational level, this is a categorical variable (with first level as the default reference). We found that only the two highest level of education were significantly different from the level of reference.

Besides the estimated coefficients θ, the table reports the standard errors of the model parameter estimates, the confidence intervals, and statistics of the regression. In particular, $R^2 = 0.3730$.

6.2 Goodness of fit

A first test used to assess the results obtained from the Poisson regression is the deviance goodness-of-fit (gof) test. Table 4 reports the deviance statistic $D = 30630.2$, the residual degrees of freedom (df$=$ 29993), and the resulting χ^2 $p-$value. With a $p-$value less than 0.05, one can consider the model well fitted. In place of the deviance, on can also consider the Pearson χ^2 statistic (see Table 4).

When we divide the two statistics by the residual degrees of freedom, it results (1/df)Deviance$=$ 1.021245, and (1/df)Pearson$=$ 1.036365. We note that, the dispersion statistic based on Pearson gof has a value greater than 1 indicating variability in the model higher than expected. In this case, there is a moderate amount of over-dispersion. With a large number of observation, the statistic is less than 1.05; in such a case, one can try to amend the model to eliminate such excess of dispersion.

Over-dispersion is an important issue, as it may cause standard errors of the estimates to be underestimated. It can be due to several reasons[7]: positive correlations in responses, excess variation between response probabilities or counts, violations in the distributional assumptions of the data.

Table 4. Goodness-of-fit tests

Test	Statistics	$Prob > \chi^2(29993)$
Deviance gof	30630.2	0.0049
Pearson gof	31083.7	0.0000

6.3 Adjust over-dispersion

When modeling count data, the assumption of equi-dispersion (the mean and the variance are the same) is rarely satisfied. Then usually the Poisson model need some adjustments to account for under- or over-dispersion (which is more often the case when dealing with real data). More generally, the term over-dispersion can also be used when

[7] See [17] for a discussion.

Table 5. Results of the Quasi-likelihood model regression for the number of donations when gift amount, number of risks, wealth, number of gift requests and the two highest educational levels are considered as variables. (Deviance = 0.9854105, Pearson = 1, with dispersion: 1.036365). Number of observations = 30 000.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. interval
gift amount	0.0018902	0.0000111	169.69	0.000	[0.0018684, 0.0019120]
n. risks	0.0216880	0.0013034	16.64	0.000	[0.0191335, 0.0242425]
wealth	0.0000377	0.0000075	5.00	0.000	[0.0000229, 0.0000524]
gift requests	0.0627460	0.0002907	215.81	0.000	[0.0621762, 0.0633159]
education3	0.0163401	0.0054647	2.99	0.003	[0.0056295, 0.0270507]
education4	0.0167768	0.0054943	3.05	0.002	[0.0060082, 0.0275453]
const.	0.4106476	0.0073412	55.94	0.000	[0.3962591, 0.4250361]

the observed variance of the count outcomes is larger than the expected variance (the variance of the predicted or expected counts).

Considering the observed occurrences for the gift counts as shown in Fig. 2, and statistics reported in Table 2, there is evidence of over-dispersion.

A first method we used to deal with over-dispersion is Quasi-Likelihood that allows parameter estimates to be obtained without explicit specification on an underlying log-likelihood function, but based only on the mean and variance of the observations. The Pearson dispersion statistic obtained in the standard Poisson regression, (1/df)Pearson = 1.036365, is used as the variance multiplier.

The results are reported in Table 5 and the summary statistics can be compared with those of the standard Poisson regression in Table 3. The deviance statistic is lower (0.9854105), and the Pearson dispersion value is now 1.

Furthermore, we have implemented robust regression. Robust variance estimators is used to adjust standard errors for correlation in the data. The results are collected in Table 6.

Finally, we applied another method to adjust standard errors, nonparametric bootstrapping, which is not based on specific assumptions about the underlying distribution. Results are reported in Table 7.

It is worth noting that, the values of the bootstrapped and robust standard errors do not differ substantially from the ones of the standard Poisson regression (see Table 3); this is a further evidence that the model is not extradispersed.

6.4 Truncated Poisson regression

Count data relates to the number of observations that may take only nonnegative integer values, ranging from zero to infinity; but in many cases of practical interest or study design the outcomes are limited to some determined value. When considering the number of gifts from a Donor in a certain interval of time, one has to deal with data that have

Table 6. Results of the robust regression for the number of donations when gift amount, number of risks, wealth, number of gift requests and the two highest educational levels are considered as variables. (Deviance = 1.021245, Pearson = 1.036365). Pseudo $R^2 = 0.3730$. Wald $\chi^2(6) = 64026.20$; $Prob > \chi^2 = 0.0000$. Number of observations = 30000.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. interval
gift amount	0.0018902	0.0000134	140.92	0.000	[0.0018639, 0.0019165]
n. risks	0.0216880	0.0015824	13.71	0.000	[0.0185865, 0.0247896]
wealth	0.0000377	0.0000094	4.03	0.000	[0.0000193, 0.0000560]
gift requests	0.0627460	0.0003195	196.40	0.000	[0.0621198, 0.0633722]
education3	0.0163401	0.0064480	2.53	0.011	[0.0037021, 0.0289780]
education4	0.0167768	0.0065371	2.57	0.010	[0.0039643, 0.0295893]
const.	0.4106476	0.0076841	53.44	0.000	[0.3955871, 0.4257081]

Table 7. Bootstrapped regression (number of samples = 1000) for the number of donations when gift amount, number of risks, wealth, number of gift requests and the two highest educational levels are considered as variables. (Deviance = 1.021245, Pearson = 1.036365). Pseudo $R^2 = 0.3730$. Wald $\chi^2(6) = 64026.20$; $Prob > \chi^2 = 0.0000$. Number of observations = 30000.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. interval
gift amount	0.0018902	0.0000134	141.07	0.000	[0.0018639, 0.0019165]
n. risks	0.0216880	0.0015824	13.71	0.000	[0.0185874, 0.0247886]
wealth	0.0000377	0.0000099	3.82	0.000	[0.0000183, 0.0000570]
gift requests	0.0627460	0.0003143	199.66	0.000	[0.0621301, 0.0633619]
education3	0.0163401	0.0065128	2.51	0.012	[0.0035752, 0.0291050]
education4	0.0167768	0.0067262	2.49	0.013	[0.0035937, 0.0299598]
const.	0.4106476	0.0076504	53.68	0.000	[0.3956531, 0.4256421]

been truncated or censored. Censoring occurs when counts can possibly exist, but due to the study design (or other reasons) some outcomes are not present in the observed data.

Table 8. Results of the truncated Poisson regression for the number of donations when gift amount, number of risks, wealth, number of gift requests and the two highest educational levels are considered as variables. Pseudo $R^2 = 0.3890$. LR $\chi^2(6) = 82750.85$; $Prob > \chi^2 = 0.0000$. Number of observations = 30 000. Minimum number of donations is 1 and maximum number of donations considered is 28.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. interval
gift amount	0.0021528	0.0000135	159.01	0.000	[0.0021263, 0.0021794]
n. risks	0.0233569	0.0014066	16.60	0.000	[0.0205999, 0.0261138]
wealth	0.0000402	0.0000081	4.99	0.000	[0.0000244, 0.0000559]
gift requests	0.0690811	0.0003298	209.45	0.000	[0.0684347, 0.0697275]
education3	0.0135521	0.0058815	2.30	0.021	[0.0020245, 0.0250796]
education4	0.0174984	0.0059166	2.96	0.003	[0.0059020, 0.0290947]
const.	0.2391184	0.0085511	27.96	0.000	[0.2223585, 0.2558783]

As soon as a Donor is registered in the DB, the number of gifts Y is 1, as it the case with data represented in 2. That is, the range of the count variable is $Y \geq 1$. Generally, also the maximum number of donation can be modeled.

Here, we have considered a model where the number of gift is truncated. In particular, we exclude zero counts and limit the maximum number of gift, based on the gift history. The performance of the Poisson truncated regression are displayed in Table 8. One can observe that the pseudo R^2 has slightly improved.

7 Concluding remarks

In FR, the assessment of the expected gift is a crucial task. The accuracy of the estimated number of donations and gift amounts depends on the efficient use of the knowledge of Donors' individual characteristics and gift history. Information based quantitative approaches are implemented to optimize the resources by selecting the most promising Donors/Contacts from an organized DB, with the aim of maximizing the expected global gift of a particular campaign, under budget constraints.

In this contribution, we propose the use of parametric models for the prediction of Donors' behavior. In particular, the Poisson regression model is adopted for the number of gifts. This basic specification allowed us to establish a relationship between the count variable risk models and the models based on dichotomous qualitative variables linked to Donors' individual characteristics. If we can estimate the parameters of the distribution underlying the data, possibly with little bias, then we could use the resulting estimated model to make predictions and classifications. Furthermore, one can use the resulting model to assess the probability of future events.

The aim of the present work is to provide some guidelines on how to construct, interpret, and evaluate models in FR, such as for instance models for the gift count, as to their fit. Then, we performed a Poisson regression on a simulated but realistic dataset, illustrating problems that may arise when dealing with count data. In particular, we addressed two main issues: dispersion and truncation of results. We discussed how to interpret model coefficients and how predictions are produced.

Nevertheless, when dealing with real data, the underlying assumptions of the basic Poisson regression model seem quite unrealistic, due to the existence of some idiosyncratic risk related to Donors. When basic assumptions are relaxed, parametric and semi-parametric extensions to the Poisson regression model could be applied. For example, one may consider the residual heterogeneity of Donors as further unobserved random variable; when gamma heterogeneity is assumed, one obtains a negative binomial model. This issue is left for future research.

References

1. Andreoni, J.: Philantropy. In: Kolm, S.C., Ythier, J. (eds.): Handbook of the Economincs of Giving, Altruism and Reciprocity **2**, Elsevier (2006) 1201–1269
2. Barro, D., Barzanti, L., Corazza, M., Nardon, M.: Machine Learning and Fundraising: Applications of Artificial Neural Networks. University Ca' Foscari of Venice, Department of Economics Research Paper Series, 33/WP/2023 (2023)
3. Barro, D., Barzanti, L., Corazza, M., Nardon, M.: Input relevance in Multi-Layer Perceptron for fundraising. In: Corazza M., Gannon F., Legros F., Pizzi C., Touzé V. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2024, Springer Nature Switzerland (2024) *forthcoming*
4. Barzanti, L., Dragoni, N., Degli Esposti, N., Gaspari, M.: Decision making in fund raising management: A knowledge based approach. In: Ellis, R., Allen, T., Petridis, M. (eds.): Applications and Innovations in Intelligent Systems XV, Springer (2007) 189–201
5. Barzanti, L., Gaspari, M., Saletti, D.: Modelling decision making in fund raising management by a fuzzy knowledge system. Expert Systems with Applications **36** (2009) 9466–9478
6. Barzanti, L., Giove, S.: A decision support system for fund raising management based on the Choquet integral methodology. Expert Systems **29(4)** (2012) 359–373
7. Barzanti, L., Giove, S.: A decision support system for fund raising management in medium-sized Organizations. Mathematical Methods in Economics and Finance **9/10** (2018) 3–11
8. Barzanti, L., Giove, S., Pezzi, A.: A decision support system for non profit Organizations. In: Petrosino, A., Loia, V., Pedrycz, W. (eds.): Fuzzy Logic and Soft Computing Applications. Lecture Notes in Artificial Intelligence, Springer, Switzerland (2017) 270–280
9. Barzanti, L., Mastroleo, M.: An enhanced approach for developing an expert system for fund raising management. In: Segura, J.M., Reiter, A.C. (eds.): Expert System Software: Engineering, Advantages and Applications, Nova Science Publishers (2013) 131–156
10. Barzanti, L., Nardon, M.: Estimation of the gift probability in fund raising management. In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2022, Springer, Cham (2022) 70–75
11. Cagala, T., Glogowsky, U., Rincke, J., Strittmatter, A.: Optimal targeting in Fundraising: A Machine-Learning approach. CESifo Working Papers 9037 (2021)

12. Cappellari, L., Ghinetti, P. Turati, G.: On time and money donations. Journal of Socio-Economics **40(6)** (2011) 853–867
13. Duffy, J., Ochs, J., Vesterlund, L.: Giving little by little: Dynamic voluntary contribution games. Journal of Public Economics **91** (2007) 1708–1730
14. Duncan, B.: Modeling charitable contributions of time and money. Journal of Public Economics **72** (1999) 213–242
15. Farrokhvar, L., Ansari, A., Kamali, B.: Predictive models for charitable giving using machine learning techniques. PLoS ONE **13(10)** (2021) 1–14
16. Gourieroux, C., Jasiak, J.: The Econometrics of Individual Risk. Credit, Insurance, and Marketing. Princeton University Press, Princeton and Oxford (2007)
17. Hilbe, J.M.: Modeling Count Data. Cambridge University Press (2014)
18. Lee, L., Piliavin, J.A., Call, V.R.A.: Giving time, money, and blood: Similarities and Differneces. Social Psycology Quarterly **62(3)** (1999) 276–290
19. Melandri, V.: Fundraising. Civil Sector Press, Toronto, Canada (2017)
20. Nudd, S.P.: Thinking strategically about information. In: Tempel, E. (ed.): Hank Rosso's Achieving Excellence in Fund Raising, John Wiley & Sons, New York (2003) 349–365
21. Sargeant, A.: Using donor lifetime value to inform fundraising strategy. Nonprofit Managment and Leadership **12(1)** (2001) 25–38
22. Smith, W., Chang, C.: Shipping the good apples out: A note on contributions of time and money. Economic Bulletin **10(1)** (2002) 1–14
23. Verhaert, G.A., Van den Poel, D.: The role of seed money and threshold size in optimizing fundraising campaigns: Past behavior matters! Expert Systems with Applications **39** (2012) 13075–13084

MATHEMATICAL METHODS IN ECONOMICS AND FINANCE – m²ef
Vol. 17/18, No. 1, 2022/23
ISSN print edition: 1971-6419 – ISSN online edition: 1971-3878
Web page: http://www.unive.it/m2ef/ – E-mail: m2ef@unive.it