Ca' Foscari University of Venice

Department of Economics

Working Paper

Luca Barzanti
Martina Nardon

A Negative Binomial model for the donations count in Fundraising Management

ISSN: 1827-3580
No. 28/WP/2025

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 28/WP/2025
ISSN 1827-3580

A Negative Binomial model for the donations count in Fundraising Management

Luca Barzanti
University of Bologna

Martina Nardon
Ca' Foscari University of Venice

Abstract
Forecasting expected gifts is a key task in Fundraising Management. In this study, we propose modeling a gift as an individual risk that can be analyzed from multiple perspectives: the occurrence, frequency, and timing of donations, as well as their monetary amounts. We focus specifically on modeling the number of donations as a Poisson random variable whose intensity parameter depends on individual donor characteristics. By introducing a Gamma-distributed heterogeneity factor, a Negative Binomial model arises as a natural extension of the starting framework. This approach enables the estimation of both the expected number of donations and the probability of a gift through Negative Binomial regression. We illustrate the methodology with an empirical application.

Keywords
Fundraising Management; Expected Gift; Poisson Regression; Negative Binomial Regression

JEL Codes
D64, C63

Address for correspondence:
Martina Nardon
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: mnardon@unive.it

The Working Paper Series
is available only on line
(https://www.unive.it/dec-workingpapers)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

A Negative Binomial model for the donations count in Fundraising Management

Luca Barzanti[1] and Martina Nardon[2]
[1] Department of Mathematics, University of Bologna
luca.barzanti@unibo.it
[2] Department of Economics, Ca' Foscari University of Venice
mnardon@unive.it

Abstract

Forecasting expected gifts is a key task in Fundraising Management. In this study, we propose modeling a gift as an individual risk that can be analyzed from multiple perspectives: the occurrence, frequency, and timing of donations, as well as their monetary amounts. We focus specifically on modeling the number of donations as a Poisson random variable whose intensity parameter depends on individual donor characteristics. By introducing a Gamma-distributed heterogeneity factor, a Negative Binomial model arises as a natural extension of the starting framework. This approach enables the estimation of both the expected number of donations and the probability of a gift through Negative Binomial regression. We illustrate the methodology with an empirical application.

Keywords: Fundraising Management; Expected Gift; Poisson Regression; Negative Binomial Regression.

JEL Classification: D64, C63.

1 Introduction

Fundraising (FR) activities aim to effectively engage potential donors and raise funds for a specific purpose. In pursuing their mission, associations design targeted strategies to achieve the objectives of an FR campaign and maximize the expected gift[1]. Estimating both the probability of a gift and its expected amount is a central problem. To this end, the efficient use of information about donors, combined with insights gained from past campaigns (see [18]), is essential.

A key component of the FR process is identifying donor profiles (contacts) that exhibit particular propensities to give[2]. Practitioners in the field often claim that the success of a campaign depends primarily on selecting the right target group of donors, while factors such as messaging creativity or motivation play a comparatively minor role.

Economists also emphasize the critical role of donor information in improving FR strategies [15]. Prior studies identify several factors influencing individuals' decisions to donate. According to [1], the economic and social foundations of altruism are shaped by community membership, social networks, and "enlightened self-interest". Similar aspects are discussed in [8] and [17]. [13] highlight that individuals may develop a "role-identity" as donors, which depends on their social relationships. Variables that influence this role-identity (such as personal preferences, attitudes, and social connections) can affect both the motivation and the utility individuals derive from donating [6].

Regarding available data, associations can be classified according to whether they maintain a structured database (DB) and the type of qualitative information it contains. In addition to standard donor and gift data (e.g., gift history and demographic information), some DBs include richer qualitative details such as personal interests, attitudes, or social network relationships. Typically, the extent and quality of this information depend on the organization's size and resources.

Since contacting donors and collecting data are costly, a main goal of FR Management is to identify the most promising donors or contacts to maximize expected returns under time and

[1] See [16] and [14] for an overview on the subject.
[2] See [8], [13], and [7].

budget constraints. The proportion of positive donor responses is often treated as a parameter estimated through expert judgment. Therefore, developing accurate statistical estimates of gift probability – based on donor characteristics and gift history – is of great practical value.

FR-related modeling can address several aspects of the giving process: the occurrence of a gift, the number and timing of donations, and the donation amount. Each of these dimensions can be analyzed and estimated through quantitative models.

Recent approaches[3] increasingly employ mathematical and statistical methods, artificial intelligence (AI) algorithms, and decision support systems (DSS) to assist organizations in designing campaigns, selecting strategies, and managing outcome variability. Both non-parametric and parametric approaches can be used to estimate relevant quantities such as the expected gift amount and the donation probability. A very recent research stream for FR applies non-parametric Machine Learning models[4]. Along these lines, [2] use a Multi-Layer Perceptron (MLP) to predict both the number of donations and the gift amount, while [3] extend the analysis by evaluating the relative importance of input variables to improve campaign effectiveness.

The FR process can also be formalized using probabilistic models. In parametric approaches, specific assumptions are made about the distributions of quantities of interest. For example, [4] model the number of gifts as a Poisson random variable with an intensity parameter depending on donor characteristics. A Poisson regression is then used to estimate the expected number of donations, the probability of a gift, and to assign a donor score measuring their propensity to donate.

In this paper, we extend the framework proposed by [4] by introducing donor heterogeneity as an unobserved random effect. In particular, assuming a Gamma-distributed heterogeneity factor leads to a Negative Binomial (NB) distribution for the donation count. Model parameters are estimated through NB regression.

The remainder of the paper is organized as follows. Section 2 introduces the model of the gift as an individual risk. Section 3 presents the Poisson–Gamma mixture model for the number of donations. Section 4 describes the dataset containing donor characteristics and gift histories. Section 5 discusses the results of the NB regression. Section 6 concludes.

[3]See [4] and references therein.
[4]See, for example, [9] and [5].

2 Modeling the gift as a risk

As previously discussed, the appropriate use of donor information is crucial for producing accurate forecasts of expected gifts and, ultimately, for optimizing resources under given constraints. Integrating donor data to design an optimal FR strategy is a complex task: it requires a clear definition of campaign objectives and also a rigorous specification of the variables included in the analysis.

Beyond the importance of systematically collecting and updating data, it is essential to introduce quantitative tools that can effectively exploit this information, particularly in the FR context, where mathematical models are still rarely used, except within large organizations. In this regard, the 'gift' can be modeled as an *individual risk*, analogously to approaches used in other applied fields such as finance, insurance, and marketing (see [10]).

The gift can thus be analyzed from several perspectives:

Occurrence — whether a donation takes place or not; the outcome is binary ('yes' or 'no');

Count or ***Frequency*** — the number of donations received within a given period (e.g., a year or campaign duration); this is a nonnegative integer variable;

Timing or ***Recency*** — the interval between donations or the time elapsed since the most recent gift; this is typically measured from a reference point, such as the campaign start date or the donor's first contact;

Amount — the monetary value of each donation, or equivalently, the total amount donated during a given period.

From each of these perspectives, the gift can be treated as a random variable: dichotomous (for occurrence), count (for frequency), duration (for timing), or continuous positive (for amount).

Both dichotomous and count variables can describe the occurrence of the gift event. Formally, if we denote by G the donation event, we define the dichotomous random variable $Y = \mathbf{1}_G(\omega)$, where $\mathbf{1}_G$ is the indicator function of G, with $\mathbb{P}[Y = 1] = p$. The probability of a gift is therefore $\mathbb{E}(Y) = p$. Let X be a continuous random variable representing either the amount donated in a single gift or the total amount donated within a given period. In this case, the expected gift per donor can be expressed as the product $\mathbb{E}(Y)\mathbb{E}(X)$.

At the campaign level, both the number of gifts and the gift amounts are random, so the campaign's total return can be represented as a *random sum*. To compute its expectation, assumptions such as independence among donors and between donation counts and amounts are typically required. Various modeling alternatives can be developed under these assumptions.

We assume that:

- each gift is made by a donor (an individual, company, or organization) characterized by a set of measurable features stored in a dataset;
- donor characteristics are summarized by a *score*;
- each donor's gift history (events, timing, and amounts) is recorded.

In the FR context, a *score* is a statistical measure of individual risk based on donor characteristics (see, e.g., [10]). It quantifies the individual propensity to donate – the higher the score, the higher the expected likelihood of giving – and can be used to rank donors. *Segmentation* techniques can further classify donors into groups, allowing for targeted communication and customized campaign strategies.

Let x_i denote the vector of selected observable characteristics (both qualitative and quantitative) for donor i in a sample of n donors. Define z_i as the transformed vector of individual characteristics, where qualitative attributes are appropriately encoded into quantitative or dummy variables. The score is then defined as a scalar function of the covariates,

$$s_i = z_i'\theta,$$

where θ is a vector of parameters. This score summarizes the donor's relevant information and can be obtained using more sophisticated approaches (see [10]).

In the next section, we focus on one specific aspect of the gift process: the number of donations, modeled as a function of individual donor characteristics.

3 Poisson-Gamma mixture model for donations count

The arrival of a donation to an association can be regarded as the realization of a random variable. In the simplest case, a dichotomous variable indicates whether or not a gift has been received. In this study, we adopt a parametric approach to model the *number* of gifts received within a given period, drawing an analogy with insurance theory, where count variables are used to describe the frequency of loss events over time. As a starting point, we consider the Poisson distribution, which is the standard model for count data, and apply it to represent the number of gifts. The model parameters can then be estimated through Poisson regression.

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 $\mathbb{V}(Y) = \lambda$. This hypothesis is frequently violated by real data which more often exhibit over-dispersion.

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 $Y_1, \dots, Y_n$ 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). As

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

it results that the model is heteroskedastic. Parameters θ can be estimated by maximum likelihood (ML); the resulting log-likelihood function is[5]:

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

$l(\theta)$ is concave with respect to θ; the ML 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.

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, the model 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 segmenting 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 more regular donors.

3.1 A Negative Binomial Regression in FR

The Poisson model is adopted as a starting point for FR analysis, a field in which the use of advanced statistical models is still limited, despite the growing interest in quantitative and AI-based approaches. This basic specification for count variables provides a useful framework for linking risk models for count data with models based on dichotomous qualitative variables associated with donors' individual characteristics.

One of the main advantages of the Poisson regression model is its simplicity and ease of interpretation. However, it relies on assumptions that are often unrealistic in practice. In par-

[5] See also [10] for details and properties of the estimators.

ticular, individual donors may exhibit idiosyncratic risk that should be explicitly accounted for. The Poisson model, nonetheless, serves as a flexible baseline that allows for several extensions.

For instance, introducing a Gamma-distributed *heterogeneity factor* leads to the Negative Binomial (NB) model[6], which accommodates unobserved heterogeneity among individuals. In the context of automobile insurance, for example, the NB model is employed to derive closed-form updating formulas for policy premiums, forming the basis of the *bonus–malus* system.

The individual intensity μ_i is specified as a deterministic function of observable covariates. We assume that all donors follow this specification; however, some explanatory variables may be omitted because they are unobserved or deemed irrelevant. The variable u_i accounts for all characteristics of donor i that differentiate their behavior from the average predicted by the model, given the set of observable features.

In Section 4, we introduce a segmentation framework for contacts and donors based on the so-called *giving pyramid*. An individual may be misclassified within this structure due to the presence of unobserved heterogeneity captured by the latent factor u_i.

Individual intensity can then be written as

$$\mu_i = \exp(z_i'\theta + \varepsilon_i) = u_i \exp(z_i'\theta), \tag{6}$$

where u_i (or ε_i) is the latent (unobserved) variable, called the heterogeneity factor or the omitted heterogeneity.

The model is defined in two steps: first assume that the conditional distribution of Y_i, given both x_i and u_i, is Poisson distributed with parameter $\mu_i = u_i \exp(z_i'\theta)$; then specify the conditional distribution of u_i given covariates as a Gamma distribution[7], Gamma(a,a), with density

$$f(u) = \frac{a^a}{\Gamma(a)} u^{a-1} e^{-au}, \tag{7}$$

where $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u}\, du$, for $a > 0$.

Note that, the equality constraint on the two parameters of the distribution (a,a) yields $\mathbb{E}[u] = 1$. Such a constraint is not too restrictive, provided that a constant term is included among the observable covariates. The variance of the heterogeneity factor is $\mathbb{V}[u] = 1/a$.

[6]See [11] and [12] for a comprehensive treatment of the subject; see also [10] for applications in insurance.

[7]As an alternative, a semi-parametric approach can be adopted, where no specific distributional form is assumed for the heterogeneity factor (see [10], p. 70).

The first- and second-order conditional moments of Y given the covariates are:

$$\mathbb{E}[Y_i|x_i] = \mathbb{E}\left[\mathbb{E}\left(Y_i|x_i,u_i\right)|x_i\right] = \mathbb{E}\left[u_i \exp(z_i'\theta)|x_i\right] = \exp(z_i'\theta) = \lambda_i\,, \tag{8}$$

and

$$\mathbb{V}[Y_i|x_i] = \mathbb{E}\left[u_i \exp(z_i'\theta)|x_i\right] + \mathbb{V}\left[u_i \exp(z_i'\theta)|x_i\right] = \exp(z_i'\theta) + (1/a)\exp(2z_i'\theta)\,. \tag{9}$$

Hence, the conditional variance can be greater than, or equal to, the conditional mean allowing to relax the equi-dispersion assumption of the Poisson model (which is a limiting case).

The model depends on the parameters: θ, the vector of sensitivity coefficients of the observable explanatory variables, and scalar a that measures the amount of heterogeneity. When a decreases there is more heterogeneity among the donors, and the over-dispersion increases.

The conditional density is

$$f(y_i|x_i) = \frac{\Gamma(y_i+a)}{\Gamma(y_i+1)\Gamma(a)} \frac{\left[(1/a)\exp(z_i'\theta)\right]^{y_i}}{\left[1+(1/a)\exp(z_i'\theta)\right]^{y_i+a}}\,, \tag{10}$$

which is the density of an NB random variable[8].

The parameters θ and a can be estimated by ML, based on the conditional density (10), given the observable covariates. Log-likelihood can be maximized numerically in this case. In the applications, we used the `STATA` commercial statistical software where the NB is parametrized with a direct relationship between the mean and the dispersion. The variance function for such a paramterization[9] of NB is $\lambda + \alpha\lambda^2$, where λ is the intensity of the original Poisson model and $\alpha = 1/a$. The larger variability or correlation there is in the data, then the larger the dispersion parameter α will be. Parameter α is a measure of the adjustment made to the over-dispersion in the model without α (the Poisson model). With this parametrization, the NB model belongs also to the exponential family of distributions at the basis of generalized linear models (GLM)[10].

[8]Here the NB distribution arises from a mixture of a Poisson and a Gamma distribution. There are other ways to obtain a NB distribution (see, [11], p. 5 for a short history of the NB), and alternative parametrizations are used. A NB is also defined as a discrete random variable X which describes the number of failures (or successes) in a sequence of *iid* Bernoulli trials before a specified number of successes (or failures) $r > 0$ is observed.

[9]There is also an alternative parametrization with variance of the form $\lambda + \alpha\lambda$, not considered in this work.

[10]See [11], Ch. 8.

4 Modeling the information on donors

Donors' individual characteristics can strongly influence both the amount and frequency of their gifts. Several elements should be considered when developing FR strategies: donor-specific information, data from past campaigns, and the operational knowledge and experience of experts in the field.

Regarding donor characteristics to be included in the analysis, these can be grouped into the following categories:

- **Personal situation variables:** gender, age, number of children, education level, place of origin, size of residence town, etc.;
- **Financial situation variables:** income, wealth, investments, debts;
- **Risk aversion indicators:** proxies such as the number of insurance policies held by the individual;
- **Other personal information:** interests, religious involvement, social network, and similar factors.

In practice, such data may not always be accessible, particularly when they involve sensitive personal information. When available, data are typically managed within a structured DB, although smaller organizations may lack such systems. Traditional DB management tools often have limitations, and decision-making is therefore supported by the expertise and intuition of professionals in the field (see [7]). Nonetheless, there is growing interest among associations and software providers in developing new soft-computing tools capable of processing and analyzing large volumes of donor data more efficiently.

Donors segmentation is often guided by the *giving pyramid*, an instrument used to classify donors according to both the frequency and amount of their contributions. An example is shown in Figure 1, where the base of the pyramid consists of *contacts*. Donors who give occasionally and contribute small amounts are classified as *Sporadic*. Donors who contribute more frequently or with greater generosity are labeled *Regular*, while those who make large and typically consistent donations are identified as *Major Donors*. At the top of the pyramid are

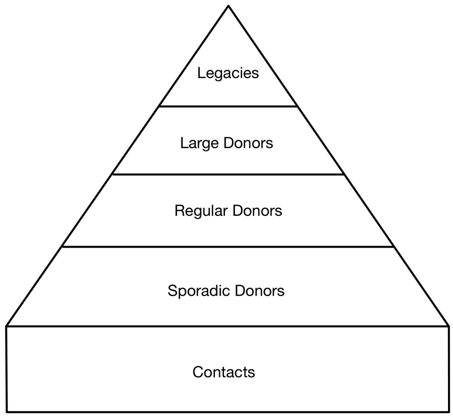

Figure 1: The *giving pyramid* in FR Management representing the segmentation of the donors.

bequests and legacies—donors who are often deeply committed to the organization's mission. Further sub-classifications can be defined depending on the characteristics and size of the DB, as well as the specific type of FR activity.

Organizations typically maintain detailed records of past campaigns, including the gift history of each donor stored in their DB. This information makes it possible to construct a giving pyramid based on historical behavior and to update its layers as donors' engagement evolves over time and across campaigns. The main objective is to convert contacts into active donors and to encourage Sporadic donors to become Regular contributors, thereby fostering long-term donor retention.

4.1 The dataset

The numerical analysis presented in Section 5 is based on a simulated DB previously used in other studies ([2], [3]). The dataset was constructed using expert knowledge and is designed to reflect the realistic composition of a donor base.

The DB comprises $n = 30000$ donors and approximately 400000 contacts. These figures correspond to medium-to-large volumes for a mid-sized organization, or high volumes for a small one. Within the donors set, 75 % are classified as *Sporadic Donors* (labeled 'sd'). Among them, roughly 25 % made only one donation (labeled 'sd1'), while the remaining 50 % made multiple donations (labeled ‘sd2’). The remaining 25 % of donors are composed of 19 % *Regular Donors*[11] (labeled ‘rd’) and 6 % *Large Donors*. Legacies are not included in the present sample.

In addition to donation history, the dataset contains several personal profile variables, including: age, number of children, educational level[12], wealth (measured in thousands of euros), and risk aversion (proxied by the number of insurance policies held by the donor).

Regarding gift history, the dataset records, for each donor: the total number of donations, the amount of each donation,[13] and the number of gift requests (i.e., the number of times a donor was contacted or sought information about the FR campaign).

Tables 1 and 2 provide a summary of the data contained in the database. In particular, Table 1 presents the composition (segmentation) of the donor population within the giving pyramid, broken down by selected characteristics. Approximately 70 % of *Sporadic Donors* are classified as having “low wealth”, compared with about 40 % among *Regular Donors* and 10 % among *Large Donors*. The second column reports the percentage of donors who hold at least one insurance policy, which increases across higher tiers of the pyramid. The last two columns display the minimum and maximum donation amounts. These figures reflect the definitions adopted for each donor group: *Sporadic Donors* (low gift amount, low frequency), *Regular Donors* (low to medium gift amount, medium to high frequency), and *Large Donors* (high gift amount).

Table 2 reports the main descriptive statistics related to gift history (number of donations, amounts, and number of requests), together with individual donor characteristics (age, number of children, education level, wealth, and risk aversion).

The empirical distribution of the number of donations is illustrated in Figure 2. Since the analysis is based on a sample of donors only, the number of donations ranges from one to the maximum observed value. This choice avoids inference issues related to the excess of zeros that would arise if all contacts in the database were included.

[11]They further may be subdivided into “stable” (labeled ‘rd1’) and “dynamic” (labeled ‘rd2’).

[12]This categorical variable is transformed into values from 1 to 4, with 4 corresponding to the highest educational level.

[13]The average donation amount per donor is used in the analysis.

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

Donors	low wealth	n. risks ≥ 1	min gift amount	max gift amount
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

5 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 recorded 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 Figure 2. On the other hand, including in the analysis the information on contacts would lead to a different problem: the excess of zeros. In these cases, the underlying distribution may need to be adjusted to take into consideration

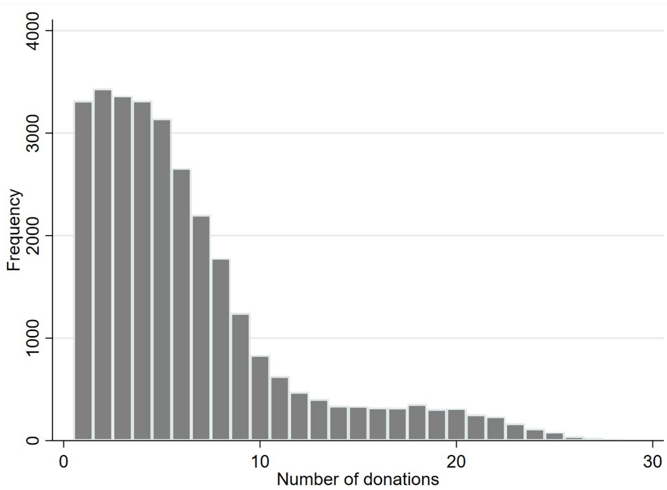

Figure 2: Empirical distribution of the number of donations.

or exclude zero counts[14]. The model could be amended considering, for example, a truncated distribution.

Another problem is over-dispersion in the data. 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 5.1, we start considering the standard Poisson regression (as in [4]). Results presented in [4] are reported here for sake of comparison; we tested over-dispersion and discussed how to treat such an issue. Finally, we address the problem of truncated data. Section 5.2 presents the results for the NB model compared to those obtained with Poisson regression. Regression outputs are obtained using `Stata` software[15].

[14]More advanced approaches include two-part hurdle models and mixtures models (see [12] for a discussion).

[15]Precisely, `Stata` functions **`poisson`** and **`nbreg`** are used for the Poisson and NB regression, respectively. In both cases, **`glm`** regression can be used as well. In particular, the dispersion parameter in the NB regression α is inserted as a constant in the GLM.

5.1 Poisson regression

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.

Regarding the choice of profile's variables, 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. int.
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.

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$=29\,993$), 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).

Table 4: Goodness-of-fit tests results.

Test	Statistics	$Prob > \chi^2(29\,993)$
Deviance gof	30 630.2	0.0049
Pearson gof	31 083.7	0.0000

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 observations, the statistic is less than 1.05; hence, one can try to amend the model to eliminate the excess of dispersion.

Over-dispersion is an important issue, as it may cause standard errors of the estimates to be underestimated. It can be caused by several reasons (see [12] for a discussion): positive correlations in responses, excess variation between response probabilities or counts, violations in the distributional assumptions of the data.

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 requires 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 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 Figure 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.

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. int.
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]

Robust variance estimators are used to adjust standard errors for correlation in the data. The results of the robust regression are reported in Table 6.

5.1.1 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 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.

Because the sample includes donors only, $Y \geq 1$ by construction (zero-truncation). Generally, also the maximum number of donation can be modeled.

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 7. One can observe that the pseudo R^2 has slightly improved.

Table 6: Results of the robust 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. (Deviance $= 1.021245$, Pearson $= 1.036365$). AIC = 4.448941, BIC = -278566.2; 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. int.
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: 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. Log likelihood $= -64992.692$; Pseudo $R^2 = 0.3890$. LR $\chi^2(6) = 82750.85$; $Prob > \chi^2 = 0.0000$. Number of observations = 30000. Minimum number of donations is 1 and maximum number of donations considered is 28.

	coefficient	std. err.	z	$P>\|z\|$	95 % conf. int.
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]

5.2 Negative Binomial regression

NB regression employs a further parameter α that directly addresses the over-dispersion in the Poisson model. When the value of α approaches zero, the model becomes Poisson.

As regards the choice of the explanatory variables, we acted as in the case of Poisson regression, obtaining a model for the number of donations as a function of the gift amount, the risk aversion proxy, wealth, gift requests, and the two highest educational levels. All these variables are significant. Results are reported in Table 8.

Table 8: Results of the NB 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. Log likelihood = -66311.109; Pseudo $R^2 = 0.2163$; LR $\chi^2(6) = 36595.46$; $Prob > \chi^2 = 0.0000$. Number of observations = 30 000. Residual degrees of freedom df = 29 993; (1/df)Deviance = 0.8240057; (1/df)Pearson = 0.8337767.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. int.
gift amount	0.0019820	0.0000142	139.47	0.000	[0.0019541, 0.0020098]
n. risks	0.0212261	0.0015082	14.07	0.000	[0.0182701, 0.0241822]
wealth	0.0000381	0.0000086	4.44	0.000	[0.0000212, 0.0000549]
gift requests	0.0636931	0.0003322	191.70	0.000	[0.0630419, 0.0643443]
education3	0.0137469	0.0062654	2.19	0.028	[0.0014669, 0.0260268]
education4	0.0154310	0.0062991	2.45	0.014	[0.0030849, 0.0277771]
const.	0.3801181	0.0082978	45.81	0.000	[0.3638546, 0.3963815]
α	0.0301555	0.0012743			[0.0277586, 0.0327594]

A Likelihood Ratio test of $\alpha = 0$ (which is the case in the Poisson model) gives $\chi^2(1) = 832.03$; the corresponding p-value is 0.000 indicating that the NB model with $\alpha = 0.0301555$ significantly differs from the Poisson one.

The AIC and BIC statistics for the Poisson and NB models are: $\text{AIC}^P = 4.448941$, $\text{BIC}^P = -278566.2$, and $\text{AIC}^{NB} = 4.421207$, $\text{BIC}^{NB} = -284482$, respectively. The values for the NB model are less than those of the Poisson one, indicating that the data are better modeled as NB. The statistics have been reduced, even though not remarkably, and estimated parameters in both models are quite similar; such a result is not unexpected, due to the estimated value for α close to zero (but still significantly different from zero). It is also worth observing that dispersion statistics in the Poisson regression where greater than 1, but not notably larger and over-dispersion appeared limited. With little excess of variability in the Poisson model, the NB model converged with estimates outputs very close to those of the Poisson regression. The resulting NB model is actually under-dispersed.

The NB model can be fitted using a robust variance estimator; empirically based standard errors did not result in a difference in p-value significance (with some minimal differences). Results are reported in Table 9. It can be observed that all these variables are still significant.

Table 9: Results of the robust NB 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= 0.8240057, Pearson= 0.8337767); Log likelihood= -66311.109; Wald $\chi^2(6) = 62539.31$; $Prob > \chi^2 = 0.0000$. Number of observations = 30 000.

	coefficient	std. err.	z	$P > \|z\|$	95 % conf. int.
gift amount	0.0019820	0.0000152	130.18	0.000	[0.0019521, 0.0020118]
n. risks	0.0212261	0.0015472	13.72	0.000	[0.0181937, 0.0242586]
wealth	0.0000381	0.0000090	4.24	0.000	[0.0000205, 0.0000556]
gift requests	0.0636931	0.0003151	202.12	0.000	[0.0630755, 0.0643107]
education3	0.0137469	0.0062636	2.19	0.028	[0.0014705, 0.0260233]
education4	0.0154310	0.0063392	2.43	0.015	[0.0030065, 0.0278556]
const.	0.3801181	0.0074273	51.18	0.000	[0.3655609, 0.3946753]
α	0.0301555	0.001062			[0.0281441, 0.0323106]

5.2.1 Truncated NB regression

When the data are truncated, censored, clustered, or there are excessive zero counts, the model might result over-dispersed (in some cases under-dispersed). Considering the shape of the giving pyramid, we should enter into the analysis all zero counts associated with contacts, leading to a problem with excess of zeros. Instead, in the modeling of the number of donations, we excluded the possibility of having zero counts and tried to use a zero-truncated NB model. Results are shown in Table 10. Outcomes of truncated NB regression are in line with those obtained in the non-truncated model, and the estimated value of α slightly increased. In this case the performance of the model registered a worsening in the Log likelihood statistic.

Table 10: Results of the truncated NB 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. Log likelihood $= -65128.075$; Pseudo $R^2 = 0.2134$. LR $\chi^2(6) = 35328.93$; $Prob > \chi^2 = 0.0000$. Number of observations $= 30000$. Minimum number of donations is 1 and maximum number of donations considered is 28.

	coefficient	std. err.	z	$P > \lvert z \rvert$	95 % conf. int.
gift amount	0.0020642	0.0000154	134.14	0.000	[0.0020340, 0.0020943]
n. risks	0.0223926	0.0016031	13.97	0.000	[0.0192505, 0.0255347]
wealth	0.0000411	0.0000091	4.51	0.000	[0.0000233, 0.0000590]
gift requests	0.0682706	0.0003710	184.03	0.000	[0.0675435, 0.0689977]
education3	0.0134903	0.0066709	2.02	0.043	[0.0004156, 0.0265649]
education4	0.0165339	0.0067043	2.47	0.014	[0.0033938, 0.0296741]
const.	0.2559655	0.0093647	27.33	0.000	[0.2376109, 0.2743200]
α	0.0409014	0.0014975			[0.0380691, 0.0439444]

6 Conclusions

In FR management, data-driven approaches are increasingly used to guide strategic decisions, optimize resource allocation, and identify the most promising donors within an organized DB. The ultimate goal is to maximize the expected total gift of a campaign under financial and operational constraints. The accuracy of forecasts for both donation frequency and gift amount depends on the effective use of donor information, including individual characteristics and historical giving behavior.

This study provides a quantitative framework for modeling donors' behavior, focusing on the prediction of gift counts. Parametric models -— particularly those based on the Poisson distribution -— offer a starting point for understanding how observable donor characteristics relate to giving frequency. This specification establishes a link between models for count data and binary models that describe the probability of giving. Once estimated, these models allow fundraisers to assess the likelihood of future donations and to classify donors according to their predicted giving propensity.

However, the simplifying assumptions of the Poisson regression model may not hold in practice, as donors often exhibit individual variability not captured by observable characteristics. To address this, the analysis incorporates extensions that allow for unobserved heterogeneity, specifically through a Gamma-distributed random effect, yielding the NB model. This model captures a wider range of behavioral differences among donors and provides a more flexible representation of real-world data.

Using a simulated yet realistic dataset, we applied both Poisson and NB regressions to illustrate the methodological and interpretive challenges of modeling count data in the FR context. Two issues in particular —overdispersion and truncation- were discussed as central to developing reliable predictive tools. Expanding the model to include richer donor information could further enhance predictive accuracy and allow for meaningful comparison across alternative specifications.

From a managerial perspective, these findings highlight the value of integrating statistical modeling into FR strategy design. Combining traditional parametric models with Machine Learning and DSSs may enable organizations to segment donors more effectively, allocate resources more efficiently, and personalize communication strategies. Future research could test these methods on real-world datasets, exploring their potential to strengthen long-term donor relationships and improve the sustainability of nonprofit FR efforts.

Acknowledgments

The authors are thankful for insights and useful comments to participants to the *Fundraising Festival*, Italy, Rimini 2022, and the *Workshop: Analisi e gestione efficiente delle informazioni per l'ottimizzazione dei processi decisionali di Fundraising*, Italy, Treviso, 2022.

References

[1] Andreoni, J. Philanthropy. In *Handbook of the Economics of Giving, Altruism and Reciprocity*; Kolm, S.C., Ythier, J., Eds.; **2**, Elsevier, 2006; pp. 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 *Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2024*; Corazza M., Gannon F., Legros F., Pizzi C., Touzé V., Eds.; Springer Nature: Switzerland, 2024; pp. 31–36.

[4] Barzanti, L., Nardon, M. Modeling the gift in fundraising process: A parametric approach for the donations count. *Math. Meth. Econ. Fin.* **2024**, 17/18, 29–44.

[5] Cagala, T., Glogowsky, U., Rincke, J., Strittmatter, A. Optimal targeting in Fundraising: A Machine-Learning approach. CESifo Working Papers 9037, 2021.

[6] Cappellari, L., Ghinetti, P. Turati, G. On time and money donations. *J. Socio-Econ.* **2011**, 40(6), 853–867.

[7] Duffy, J., Ochs, J., Vesterlund, L. Giving little by little: Dynamic voluntary contribution games. *J. Public Econ.* **2007**, 91, 1708–1730.

[8] Duncan, B. Modeling charitable contributions of time and money. *J. Public Econ.* **1999**, 72, 213–242.

[9] Farrokhvar, L., Ansari, A., Kamali, B. Predictive models for charitable giving using machine learning techniques. *PLoS ONE* **2018**, 13(10), 1–14.

[10] Gourieroux, C., Jasiak, J. *The Econometrics of Individual Risk. Credit, Insurance, and Marketing*. Princeton University Press: Princeton and Oxford, 2007.

[11] Hilbe, J.M. *Negative Binomial Regression*; Cambridge University Press: Cambridge, UK, 2011.

[12] Hilbe, J.M. *Modeling Count Data*; Cambridge University Press: Cambridge, UK, 2014.

[13] Lee, L., Piliavin, J.A., Call, V.R.A. Giving time, money, and blood: Similarities and Differneces. *Social Psychology Quarterly* **1999**, 62(3), 276–290.

[14] Melandri, V. *Fundraising*. Civil Sector Press: Toronto, Canada, 2017.

[15] Nudd, S.P. Thinking strategically about information. In *Hank Rosso's Achieving Excellence in Fund Raising*; Tempel, E., Ed.; John Wiley & Sons: New York, US, 2003; pp. 349–365.

[16] Sargeant, A. Using donor lifetime value to inform fundraising strategy. *Nonprofit Management & Leadership* **2001**, 12(1), 25–38.

[17] Smith, W., Chang, C. Shipping the good apples out: A note on contributions of time and money. *Economic Bulletin* **2002**, 10(1), 1–14.

[18] Verhaert, G.A., Van den Poel, D. The role of seed money and threshold size in optimizing fundraising campaigns: Past behavior matters! *Expert Sys. Appl.* **2012**, 39, 13075–13084.