



Optimal reimbursement limitation for a health plan

Fabio Baione¹ · Davide Biancalana¹ · Massimiliano Menzietti² 

Received: 1 December 2022 / Accepted: 24 October 2023
© The Author(s) 2023

Abstract

The goal of this paper is the assessment of an optimal reimbursement strategy for employer-based health insurance plans (HP), that cover several categories of medical services. Indeed, a health plan may offer several cost-sharing provisions for these categories, and the “Percent Expense Paid” by the plan also known as “Actuarial Value” represents a summary measure of the protection provided. Starting from a wide industrial applied model in insurance for rate-making, the generalized linear models (GLM), we estimate the expected value and variance of the health expenditure for each category. Moreover, different reimbursement rules (e.g. deductibles, co-payments, policy limits, etc.) involve a change of the “Actuarial Value” calculated as the ratio between the benefits paid by the plan (i.e. reimbursement amounts) and the expenses paid by policyholders (i.e. expenditures). The latter is the percentage of expenditure reimbursed by the plan and is sometimes defined in actuarial literature as Indicated Deductible Relativity (IDR); an IDR can be calculated for each category covered by the Health Plan or per policyholder and is a commonly used method for scoring the benefits of health insurance. Hence, we calculate the optimal IDR for each category, using the optimization problem proposed by de Finetti (Il problema dei pieni, *Giornale dell’istituto italiano degli attuari*, 1940) in the context of proportional reinsurance. The goal is the minimization of the variance of the total reimbursement of the Health Plan by fixing the total gain. Furthermore, we propose a numerical application to a real dataset, containing observed expenditures of an Italian HP.

Keywords Health plan · Generalized linear model · Optimal strategy · Policy limitation · Tweedie

F. Baione, D. Biancalana and M. Menzietti have contributed equally to this work.

✉ Massimiliano Menzietti
massimiliano.menzietti@unical.it

Fabio Baione
fabio.baione@uniroma1.it

Davide Biancalana
davide.biancalana@uniroma1.it

¹ Department of Statistics, Sapienza University of Rome, Viale Regina Elena, 295, 00161 Rome, Italy

² Department of Economics, Statistics and Finance, University of Calabria, Ponte Pietro Bucci, Cubo 3D, 87036 Arcavacata di Rende, CS, Italy

1 Introduction

In many developed countries, the ageing of the population due to lower fertility and mortality rates and the increasing cost of new medical technologies are leading to increase in health spending. This phenomenon has been accentuated by the spread of the Covid-19. In 2019, OECD countries spent, on average, around 8.8% of their GDP on health care, with an average per capita health expenditure estimated to be more than USD 4000 (OECD, 2021). Government financing schemes are the main form of financing of health care in many of OECD countries but rising costs, the economic crisis and budgetary constraints have led to financial cuts in many countries and a reduction of the overall share of public health spending (Rechel, 2019). This leads to an increased demand for enrolment in health plans and for buying health insurance policy on the market. e.g., in Italy health plans policyholders increased from less than 7 million in 2013 to more than 14 million in 2019 (Ministero della Salute, 2021).

The first pillar in the Italian health system is represented by the National Health Insurance (Servizio Sanitario Nazionale, SSN) that accounts for 74% of total health spending (OECD, 2021). The SSN does not allow people to opt out of the system, therefore there are no substitute insurance but only complementary and supplementary private health insurance. They cover several medical services not provided by SSN, offer a higher level of comfort in hospital facilities and allow for a wider choice between public and private providers (Tikkanen et al., 2020). The second pillar is mainly characterized by private group health plans that are differentiated in “full-insured” and “self-insured”.

Financial sustainability analysis of “full-insured” and “self-insured” health plans requires the adoption of the actuarial risk assessment methodology (Carroll & Mange, 2017). This is especially true for a self-insured health plan which is a true risk-taking entity as it retains the risk for paying medical claims made by the plan’s participants and operates on its own.

Academic literature has mainly focused on forecasting the annual cost of healthcare in both the public and the private sectors due to the overall impact on the economy. Projections for annual healthcare expenditure, for an individual or a group, has been obtained using several methodologies: loss models, multivariate regression models (Frees et al., 2011, 2013), phase-type model (see Govorun et al. (2015) and literature therein) and, more recently, artificial intelligence methodologies (Vimont et al., 2021; Duncan et al., 2016). For a review of models applied to healthcare in an actuarial context reader can refer to Yang et al. (2016) and Duncan (2018).

A standard methodology to estimate health plans expenditure are two-part models, where the first part models a frequency component and the second a claim amount (or expenditure) (Frees et al., 2011; Duncan et al., 2016). In actuarial literature, frequency-severity approach (Frees, 2014) has been extensively used in modeling non-life insurance claim amount. In a standard frequency-severity approach the number of claims and expenditure per claim are used. For an example of the use of a frequency-severity model in health insurance see (Keeler & Rolph, 1988).

A key element in determining the expenditure of a health plan is the introduction of reimbursement limitations (e.g. deductibles, co-payments,...). On the one hand, they reduce costs for the health plan through cost sharing between plan and policyholders and allow contribution reduction, on the other hand they motivate policyholders to reduce expenses and, more in general, counteract moral hazard (Daykin et al., 1993; van Winssen et al., 2016).

Coverage modification changes the claim frequency and severity distributions (e.g. censoring or truncation). Parametric loss models for deductible ratemaking allow to measure

the effects of coverage modification on claim frequency and severity distributions. The main advantage of these models is that they are accurate and allow to calculate theoretically correct deductible rates (Klugman et al., 2012).

Alternatively, the regression approach could be adopted, including the deductible as an explanatory variable inside a Generalized Linear Models (GLM) (Lee, 2017). A standard approach proposed in the actuarial literature for reimbursement limitation is focused on the estimate of “Indicated Deductible Relativity” (IDR) that provides an assessment of how much the insurance loss is reduced by a deductible. Considering that a health plan offers several medical services, reimbursement rules may differ among services (e.g. dental care, specialist medical visits). Therefore, in determining a reimbursement limitation strategy the level of protection offered to members should also be considered, along with the costs and risks of the health plan. Cost sharing strategies can make it difficult for members to understand the degree of coverage offered by a health plan. A possible index is given by the Actuarial Value (AV). AV is a measure of the relative percentages paid by a health benefits plan and its members. It is calculated using the expected health claims from a standard population, along with a plan’s cost-sharing provisions, to simulate the payment of claims. It ranges from 0 for a plan that pays nothing to 1 for a plan that pays all medical expenses.

The main goal of the paper is to assess the optimal IDR for each health service, by defining a mean-variance optimization problem. In this way, we calculate the IDRs by minimizing the total variance of the expenditure of the Health Plan (HP), fixing the expected gain. The optimization problem is solved both in the special case of no-correlation and the general case of correlation among health services. Such a problem adapts, in the health insurance context, the original optimization problem introduced by de Finetti De Finetti (1940) in the context of proportional reinsurance. Furthermore, our optimization problem requests as input the estimate of the mean and variance of expenditure for each health service. Hence, we propose a regression approach based on GLM. Starting from an Italian health plan data set, we compare two approaches: Two-Part and Tweedy. In general, these two approaches are used to determine the expected value of the response variable but results are not significantly different. However, for the purposes of this work is relevant to measure the variance of the expenditure in order to compare the optimization results using different methods to estimate mean and variance of the expenditure.

The paper is organized as follows. Section 2 describes the actuarial framework, the introduction of reimbursement limitation and the optimization problem. Section 3 compares Two-Part regression model and Tweedie regression model for the estimate of mean and variance of the expenditure. Section 4 presents a case study based on a real dataset and discusses the main findings. Section 5 concludes.

2 Actuarial framework and optimization problem

2.1 The frequency-severity model

We consider a health plan (HP) offering health care services to r policyholders. We group the health services in J groups (or ‘branch’). Let:

- i index the i th policyholder, $1 \leq i \leq r$;
- j index the j th branch of health expenditure, $1 \leq j \leq J$.

Following a standard approach in actuarial literature, we adopt a double stochastic aggregate expenditure model, where both the number of claims and the size of each claim are stochastic

variables (see, for instance, (Daykin et al., 1993)), the so called frequency-severity model. Therefore, we separately model the number of claims in a certain time period (frequency) and the expenditure per claim (severity). In the following we denote with:

- N the random variable (r.v.), number of claims per year;
- Y the r.v. expenditure for single claim;

In the classical collective risk model, the expenditure of the i -th policyholder given a specific branch j is:

$$Z_{i,j} = \sum_{k=1}^{N_{i,j}} Y_{i,j,k}, \quad (1)$$

where $Y_{i,j,k}$ is the r.v. expenditure for i th policyholder, due to the k th claim in the j th branch.

Under the typical assumption of independence between $N_{i,j}$ and $Y_{i,j,k}$ and identical distribution on $Y_{i,j,k} \sim Y_{i,j}, \forall k = 1, \dots, N_{i,j}$ (However, there are works that investigate the dependence between frequency and claim amount. For further details, see Garrido et al. (2016) and Clemente et al. (2022)), we get the well-known formulas for mean and variance of Z_{ij} see (Daykin et al., 1993):

$$E[Z_{i,j}] = E[N_{i,j}] \cdot E[Y_{i,j}]. \quad (2)$$

$$\text{Var}[Z_{i,j}] = E[N_{i,j}] \cdot \text{Var}[Y_{i,j}] + \text{Var}[N_{i,j}] \cdot E[Y_{i,j}]^2. \quad (3)$$

In order to reduce its expenditures, the health plan could introduce some payment limitations. The reimbursement amount $L_{i,j,k}$, after the application of a coverage modification, is obtained as a function $h(Y_{i,j,k}, \Theta)$, with Θ a set of possible coverage limitation rules. To measure the impact of the reimbursement rules on the expenditure, we focus on the r.v. proportion of expenditure reimbursed (IDR) for a single policyholder, service and claim:

$$IDR_{i,j,k} = \frac{L_{i,j,k}}{Y_{i,j,k}}. \quad (4)$$

whereas, the IDR_j at branch level is:

$$IDR_j = \frac{\sum_{i=1}^r \sum_{k=1}^{N_{i,j}} L_{i,j,k}}{\sum_{i=1}^r \sum_{k=1}^{N_{i,j}} Y_{i,j,k}}. \quad (5)$$

2.2 The effect of reimbursement limitations in the profit and loss of a health plan.

In the following we assume no administrative and operational costs and the related expense loading. We assume that the total contribution requested by the health plan for the j th service group, C_j , is determined as the sum of the expected expenditures for the service, $E[Z_j]$, and a safety loading m_j :

$$C_j = E[Z_j] + m_j \quad (6)$$

Considering the typical actuarial assumption of independence between policyholders, the expected value and variance of the expenditure for single branch are:

$$E[Z_j] = \sum_{i=1}^r E[Z_{i,j}], \quad (7)$$

$$\text{Var}[Z_j] = \sum_{i=1}^r \text{Var}[Z_{i,j}], \tag{8}$$

respectively.

In the following, we assume that a proportional reimbursement rule (coinsurance) is applied to each claim, based on a coinsurance factor $0 \leq \alpha_j \leq 1$ for each branch j . Hence the r.v. reimbursement for single claim $L_{i,j,k}$ are obtained as follows:

$$L_{i,j,k} = Y_{i,j,k} \cdot \alpha_j \tag{9}$$

where α_j is the j th element of the J -vector $\alpha = (\alpha_1, \dots, \alpha_J)$ and it represents the proportion of expenditure reimbursed for a single claim.

It is easy to show by Eq. (5) that α_j corresponds to the Indicated Deductible Relativity of the j th branch (i.e. $IDR_j = \alpha_j$).

Anyway, the introduction of a policy limitation reduces not only the reimbursement at branch level but also the contribution of the same proportion. Then, we can define a new r.v. profit (loss) for the j th branch as follows:

$$G_j(\alpha_j) = \alpha_j \cdot (C_j - Z_j) \tag{10}$$

whose expected value is $E[G_j(\alpha_j)] = \alpha_j \cdot m_j$.

Therefore, the total expected gain of the HP after the introduction of the reimbursement rule is a function of the $IDR_j = \alpha_j$ as follows:

$$E[G(\alpha)] = \sum_{j=1}^J E[G_j(\alpha_j)] = \sum_{j=1}^J \alpha_j \cdot m_j \tag{11}$$

It is important to note that the r.v. expected total profit for the HP, before the application of the reimbursement rules, i.e. if $\alpha_j = 1, j = 1, \dots, J$, is $E[G(\mathbf{1})] = \sum_{j=1}^J m_j$.

Under the assumption that the health service groups are uncorrelated, the variance of total profit is given by:

$$\text{Var}[G(\alpha)] = \sum_{j=1}^J \text{Var}[G_j(\alpha_j)] = \sum_{j=1}^J \alpha_j^2 \cdot \sigma_j^2 \tag{12}$$

whereas, for correlated service groups $\text{Var}[G(\alpha)]$ is:

$$\text{Var}[G(\alpha)] = \sum_{j=1}^J \sum_{w=1}^J \alpha_j \alpha_w \sigma_{jw} \tag{13}$$

where $\sigma_j^2 = \text{Var}[Z_j]$ is the variance of Z_j and $\sigma_{jw} = \text{cov}[Z_j, Z_w]$ is the covariance between Z_j and Z_w .

2.3 The optimization problem

Once the expected value and variance of r.v. Z_j have been defined, it is possible to introduce an optimization problem to determine the optimal $IDR_j = \alpha_j, j = 1, \dots, J$ for each branch. The optimal strategy is achieved by finding the optimal vector (α^*) that minimizes the total variance $\text{Var}[G(\alpha)]$ once the expected profit target \bar{G} has been set. Then, considering Σ as the non singular variance covariance matrix, we are able to formalize the optimization

problem:

$$\begin{cases} \min \boldsymbol{\alpha}^T \Sigma \boldsymbol{\alpha} \\ \text{s.t.} \\ \boldsymbol{\alpha}^T \mathbf{m} = \bar{G} \\ 0 < \boldsymbol{\alpha} \leq 1 \end{cases} \quad (14)$$

where \mathbf{m} is the J -vector of safety loadings, whose generic element is m_j .

Following Pressacco and Ziani (2010), it is possible to write the Lagrangian of this problem as:

$$L(\boldsymbol{\alpha}, \tau, \mathbf{u}, \mathbf{v}) = \frac{1}{2} \boldsymbol{\alpha}^T \Sigma \boldsymbol{\alpha} + \tau \cdot (\bar{G} - \boldsymbol{\alpha}^T \mathbf{m}) + \mathbf{u} \cdot (\boldsymbol{\alpha} - 1) - \mathbf{v} \cdot \boldsymbol{\alpha} \quad (15)$$

Making recourse to the Karush-Kuhn-Tucker (KKT) conditions, also known as the Kuhn-Tucker conditions, it turns out that $\boldsymbol{\alpha}^*$ is optimal if there exists a triplet $(\tau^*, \mathbf{u}^*, \mathbf{v}^*)$ non negative vector such that:

- i) $\boldsymbol{\alpha}^*$ minimizes $L(\boldsymbol{\alpha}, \tau, \mathbf{u}^*, \mathbf{v}^*)$;
- ii) $\boldsymbol{\alpha}^*$ is feasible;
- iii) either $\alpha_j^* = 0$ or $v_j^* = 0$ (or both) and either $\alpha_j^* = 1$ or $u_j^* = 0$ (or both).

The optimization problem (14) is the same proposed in De Finetti (1940), where a solution for the optimal retention problem in reinsurance market is provided.

De Finetti described an algorithm to solve this optimization problem for uncorrelated case and suggested a procedure to solve the general case of correlated risks under implicit “regularity hypothesis” (see Pressacco and Ziani (2010)). The basic idea underlying the de Finetti solution, even in the case of correlation, is that, rather than dealing with two goal variables (variance and expectation), it is convenient to work with what he called advantage functions:

$$F_j(\boldsymbol{\alpha}) = \begin{cases} \frac{\sigma_j^2}{m_j} \alpha_j, & \text{uncorrelated branches} \\ \sum_{w=1}^J \frac{\sigma_{jw}}{m_j} \alpha_w, & \text{correlated branches} \end{cases}, \quad j = 1, \dots, J \quad (16)$$

The advantage functions intuitively capture the advantage coming at α from a small reduction of the j th risk. As stated in Pressacco and Ziani (2010), “Given the form of the advantage functions, it was easily seen that this implied a movement on a segment of the cube characterized by the set of equations $F_i(x) = \lambda$ for all the current best performers. Here λ plays the role of the benefit parameter. And we should go on this segment until the advantage function of another, previously non efficient, risk matches the current value of the benefit parameter, thus becoming a member of the efficient set. Accordingly, at this point, the direction of the efficient path is changed as it is defined by a new set of equations, with the addition of the equation of the newcomer risk.” (Pressacco and Ziani, 73).

In our context it means that, starting from the case of absence of reimbursement rules, i.e. $\bar{G} = E[G(\mathbf{1})] = \mathbf{m}$, the introduction of reimbursement rules requires the reduction of the most efficient branch in terms of advantage function, i.e. $\left\{ j : \max_j F_j(\mathbf{1}) \right\}$. The reduction on branch j stops when the advantage function of another branch, say w , becomes higher than the one measured on j .

To find the path of the optimal solutions, we solve the optimal problem (14) considering several decreasing value of the expected contribution \bar{G} .

In order to solve the optimization problem, it is firstly necessary to get a robust estimate of $\hat{E}[Z_{ij}]$ and $\hat{\text{Var}}[Z_{ij}]$ and then calculate $\hat{E}[Z_j]$ and $\hat{\text{Var}}[Z_j]$ according to Eqs. (7) and (8).

To this aim we propose in the following section two actuarial approaches based on regression models.

3 One-part model versus two-part model for health expenditure estimate

In actuarial literature, to predict the cost of a claim, it is a usual approach to decompose two-part data, one part for the frequency - indicating whether a claim has occurred or, more generally, the number of claims - one part for the severity, indicating the amount of a claim. This approach is also known as Two-Part or Two-Stage model (Duan et al., 1983) and in health care expenditure analysis has been proposed in Frees (2010), Frees et al. (2013), among others.

However, the request of health-care services can be influenced by a set of demographic and/or geographic variables, among others. Hence, an univariate analysis for each homogeneous risk class is not feasible, and a multivariate approach is more appropriate. We adopt a GLM (Mc Cullagh & Nelder, 1989) as it represents the standard model in industry for ratemaking. For further details on the basic theory of GLM see Appendix 1. For a full description of the regression model in actuarial and financial application reader can refer to Frees (2010).

As the expenditure per single policyholder and category has a mass probability at zero and a positive, continuous, and often right-skewed component for positive values, Tweedie regression is a valid alternative to predict such an expenditure (Kurz, 2017) respect to Two-Part models.

3.1 A negative binomial-Gamma two-part regression model for the estimate of the health expenditure

Fixing the j th branch, the goal is the calculation of $E[N_{i,j}]$, $\text{Var}[N_{i,j}]$, $E[Y_{i,j}]$, $\text{Var}[Y_{i,j}]$ necessary for the estimate of mean and variance of $Z_{i,j}$ (see (2) and (3)). To this aim, we consider a two part model where a Negative binomial regression is used for the estimate of frequency component N_j and a Gamma regression for the estimate of severity component Y_j .

Considering the (29) with a logarithm link function it is possible to write the density of Negative Binomial r.v. as:

$$f_{N_j}(n_i, \mathbf{x}_i, \beta_{N_j}, \theta_j) = \frac{\Gamma(y + 1/\theta_j)}{\Gamma(y + 1) \cdot \Gamma(1/\theta_j)} \cdot \frac{(\theta_j e^{\mathbf{x}_i \cdot \beta_{N_j}})^y}{(1 + \theta_j e^{\mathbf{x}_i \cdot \beta_{N_j}})^{y+1/\theta_j}} \tag{17}$$

Then a likelihood function is associated to the Negative Binomial model:

$$L_{N_j}(n_i | \mathbf{x}_i, \beta_{N_j}, \theta_j) = \prod_{i=1}^r f_{N_j}(n_i, \mathbf{x}_i, \beta_{N_j}, \theta_j) \tag{18}$$

Therefore, it is possible to estimate β_{N_j} and θ_j parameters, by maximizing the (18). Exploiting (29) and (28) it is possible to calculate an estimate for $E[N_{i,j}]$ and $\text{Var}[N_{i,j}]$. It is important to note that a number J of Negative Binomial GLM must be carried on.

For the severity component, selecting the logarithm link function it is possible to write the density of Gamma r.v. as:

$$f_{Y_j}(y_i, \mathbf{x}_i, \beta_{Y_j}, \eta_j) = \frac{1}{y_i \cdot \Gamma(\eta_j)} \cdot \left(\frac{\eta_j \cdot y_i}{\exp(\mathbf{x}_i \beta_{Y_j})} \right)^{\eta_j} \cdot e^{-\eta_j \cdot y_i / (\exp(\mathbf{x}_i \beta_{Y_j}))} \quad (19)$$

Then, a likelihood function is associated to the Gamma model:

$$L_{Y_j}(y_i | \mathbf{x}_i, \beta_{Y_j}, \eta_j) = \prod_{i=1}^r f_{Y_j}(y_i, \mathbf{x}_i, \beta_{Y_j}, \eta_j) \quad (20)$$

Therefore, it is possible to estimate β_{Y_j} and η_j parameters, by maximizing the (20). Exploiting (29) and (28) it is possible to calculate an estimate for $E[Y_{i,j}]$ and $\text{Var}[Y_{i,j}]$. It is important to note that a number J of Gamma GLM must be carried on.

3.2 Tweedie regression model for the estimate of mean and variance of the expenditure

In this section, we use a Tweedie regression to get by a single model an estimate of mean and conditional variance of a response variable. Consider the collective risk model in Eq. (1), we temporarily remove for sake of simplicity the dependence on i and j . Suppose that N has a Poisson distribution with mean ψ , and each Y has a gamma distribution shape parameter ζ and scale parameter ρ . The Tweedie distribution is derived as the Poisson sum of gamma variables. By using iterated expectations, mean and variance can be expressed as:

$$E(Z) = \psi \zeta \rho \quad \text{and} \quad \text{Var}(Z) = \psi \zeta (1 + \zeta) \rho^2. \quad (21)$$

Now, define three parameters μ , ϕ , and p through the following reparametrization

$$\psi = \frac{\mu^{2-p}}{\phi(2-p)}, \quad \zeta = \frac{2-p}{p-1}, \quad \rho = \phi(p-1)\mu^{p-1}, \quad (22)$$

it is easy to show that $E(Z) = \mu$ and $\text{Var}(Z) = \phi\mu^p$ (see Kaas (2005), Tweedie (1956)). As a consequence, under these assumptions Z denotes a Tweedie random variable with mean $\mu > 0$, variance $\phi\mu^p$ ($\phi > 0$), where ϕ is the so-called dispersion parameter, and $p \in (-\infty, 0] \cup [1, \infty)$ the power parameter.

It is worth noting that Tweedie distribution is a special case of exponential dispersion models, a class of models used to describe the random component in GLMs, allowing us to use the Tweedie distribution with GLMs to model $E(Z_{i,j})$ and $\text{Var}(Z_{i,j})$.

The conditional expected expenditure and variance for the i th policyholder and j th branch are given by:

$$E[Z_j | \mathbf{x}_i] = E[Z_{i,j}] = g^{-1}(\mathbf{x}_i \cdot \beta_{Z_j}), \quad \text{Var}[Z_j | \mathbf{x}_i] = \text{Var}[Z_{i,j}] = \phi \cdot \mu_{i,j}^p \quad (23)$$

The estimate of the row-vector of regression parameters β_{Z_j} , of the dispersion parameter ϕ and the power p are provided by a Maximum Likelihood Estimation (MLE) approach (see Frees (2014)).

It is worth noting that a number J of Tweedie regressions must be carried on. In the following section, we report a numerical investigation applied to a real dataset.

Table 1 Description of branches policy limitations

j	Description	Acronym
1	Oral hygiene	OH
2	Supplementary care	SC
3	Dental care	DC
4	Rehabilitation and physical care	RPC
5	Specialist medical visits	SMV

Table 2 Number of claims and total expenditure (in EUR thousand) per branch

j	Acronym	N. of claims	Expenditure
1	OH	8786	755
2	SC	3963	625
3	DC	13,201	5742
4	RPC	4261	433
5	SMV	45,043	2745
Total		75,254	10,300

4 Numerical application

4.1 Data set

We consider a dataset from an Italian HP, where $J = 5$ branches are covered. A description of each branch is shown in Table 1.

The observation has a time frame of three years (2017, 2018 and 2019). Each year contains claims from $r = 16,206$ policyholders per year, whose age varies from 0 to 69. In the following we report some descriptive statistics of this HP between 2017 and 2019.

In Table 2 the number of claims and total expenditures (data in EUR thousand) are classified for branch; SMV ($j = 5$) is the branch with the greatest number of claims, whereas the greatest annual expenditure corresponds to DC ($j = 3$).

In Table 3, the observed frequency for the three most claimed branches (SMV, DC and OH) are reported for sex and age class.

The SMV shows greater frequency for females than males with a peak in age class [50,60). It is also important to highlight that the females have a higher frequency also for DC and OH, but differences between gender are lower than SMV case.

In Table 4, observed average expenditure per claim for the three most claimed branches (SMV, DC and OH) are reported for sex and age class.

The OH has no significant differences for sex, conversely females are in general more expensive than males for SMV and DC.

In Table 5 observed average per policyholder expenditure for the three most claimed branches (SMV, DC and OH) are reported for sex and age class. It is important to note that in SMV case females have an expenditure per policyholder greater than males, with a peak in age class [40;50). In DC and OH case, the higher riskiness of females is confirmed; furthermore, considering the whole population, the most expensive age classes are [50;60) for DC (144.85 EUR thousand) and [40;50) for OH (16.67 EUR thousand).

Table 3 Observed frequency in % of branch $j = 5, 3, 1$ for sex and age

Age class	SMV			DC			OH		
	F	M	Total	F	M	Total	F	M	Total
[10; 20)	0.00	NA	0.00	0.00	NA	0.00	0.00	NA	0.00
[20; 30)	68.47	35.79	60.03	14.76	21.58	16.52	13.20	14.74	13.60
[30; 40)	90.39	43.78	74.82	19.07	16.72	18.29	14.41	14.82	14.55
[40; 50)	109.98	63.36	95.51	31.94	25.46	29.93	20.33	17.81	19.55
[50; 60)	115.63	70.76	100.06	30.80	25.74	29.04	19.83	17.02	18.85
[60; 70)	82.46	59.57	73.34	21.57	20.31	21.07	13.32	12.86	13.14
Total	107.24	63.42	92.65	28.70	24.05	27.15	18.77	16.67	18.07

Table 4 Observed average expenditure per claim (in EUR thousand) of branch $j = 5, 3, 1$ for sex and age class

Age class	SMV			DC			OH		
	F	M	Total	F	M	Total	F	M	Total
[10; 20)	NA	NA	NA	NA	NA	NA	NA	NA	NA
[20; 30)	74.50	98.67	78.22	255.35	229.31	246.57	82.40	83.05	82.58
[30; 40)	66.79	66.08	66.65	315.20	297.36	309.75	85.65	81.46	84.23
[40; 50)	64.21	54.32	62.17	367.90	368.98	368.18	85.97	83.55	85.29
[50; 60)	59.73	53.97	58.32	504.12	486.64	498.74	87.47	85.45	86.84
[60; 70)	60.14	55.92	58.77	739.17	483.48	640.97	90.58	85.30	88.52
Total	62.43	55.87	60.93	439.24	424.89	435.01	86.71	84.30	85.97

Table 5 Observed average expenditure per policyholder (in EUR thousand) of branch $j = 5, 3, 1$ for sex and age class

Age class	SMV			DC			OH		
	F	M	Total	F	M	Total	F	M	Total
[10; 20)	NA	NA	NA	NA	NA	NA	NA	NA	NA
[20; 30)	51.01	35.31	46.95	37.68	49.48	40.73	10.88	12.24	11.23
[30; 40)	60.37	28.93	49.87	60.12	49.73	56.65	12.34	12.07	12.25
[40; 50)	70.62	34.42	59.38	117.50	93.95	110.19	17.48	14.88	16.67
[50; 60)	69.07	38.19	58.35	155.26	125.27	144.85	17.34	14.55	16.37
[60; 70)	49.59	33.31	43.10	159.46	98.21	135.06	12.07	10.97	11.63
Total	66.95	35.43	56.45	126.08	102.17	118.12	16.28	14.05	15.54

4.2 Regression models results

In the following paragraph, regression models outcomes are reported. All the regression models introduced in the previous section use sex and age as covariates. This choice is motivated by actuarial practice and literature (Frees, 2014; Duncan, 2018), where this demographic factors are considered crucial in each model for the annual expenditure. We limit the analysis to this two risk factors to facilitate the representation of results. Age variable is treated

Table 6 Output of negative binomial model

<i>j</i>	Estimate (<i>p</i> -values)				
	Intercept	Sex = Male	Age 2	Age 3	Dispersion
1	- 3.0494 (***)	- 0.1154 (***)	- 0.1154 (***)	0.0019 (***)	2580.24 (***)
2	- 3.2960 (***)	- 0.5278 (***)	- 0.5278 (***)	0.0018 (***)	0.0283 (***)
3	- 3.1650 (***)	- 0.1751 (***)	- 0.1751 (***)	0.0024 (***)	0.2946 (***)
4	- 5.4760 (***)	- 0.5721 (***)	- 0.5721 (***)	0.0036 (***)	0.0358 (***)
5	- 1.2189 (***)	- 0.5358 (***)	- 0.5358 (***)	0.0015 (***)	0.1856 (***)

Signif. codes: 0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 () 1 ' - ' Not Used

Table 7 Output of Gamma model

<i>j</i>	Estimate (<i>p</i> -values)				
	Intercept	Sex = Male	Age 2	Age 3	Dispersion
1	4.3791 (***)	- 0.0289 (*)	0.0001 '-'	0 '-'	0.1848 (***)
2	4.5845 (***)	0.1252 '-'	0.0008 '-'	0 '-'	14.7782 (***)
3	5.4231 (***)	- 0.0443 '-'	0.0001 '-'	0 '-'	3.2889 (***)
4	4.4424 (***)	0.1234 ()	0.0002 '-'	0 '-'	3.2606 (***)
5	4.5011 (***)	- 0.1028 (***)	- 0.0003 (.)	0 ()	5.0953 (***)

Signif. codes: 0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 () 1 ' - ' Not used

as quantitative and we use a polynomial function whose degree is selected by a stepwise procedure.

For the estimate of the expected number of claims we consider a negative binomial model; in Table 6 the results are largely in line with those found in the literature (Frees et al., 2011); indeed, sex and age are statistically significant and females shows greater frequencies than males for all the branches in accordance with the observed frequencies in Table 3. The stepwise regression suggests a third-degree polynomial to model age, without the coefficient of degree 1.

For the amount of expenditures (Gamma regression model), Table 7 shows that almost every age coefficient is not significant. It is important to note that the most significant age coefficients are the ones of SMV, in accordance with 4. Age is modelled by a third-degree polynomial without the coefficient of degree 1.

As previously explained, a valid alternative approach to estimate mean and variance of $Z_{i,j}$ consists on the application of Tweedie regression, whose results are reported in Table 8. As one can see, females are riskier than males, in accordance with observations reported in Table 5. The coefficients are all significant. Age is modeled by a third-degree polynomial without the coefficient of degree 1.

In order to measure the estimation accuracy, we introduce the goodness-of-fit (GoF) measure called Nash-Sutcliffe Efficiency (*NSE*). The latter is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (see (Nash & Sutcliffe, 1970)). Given a dependent variable y and its estimate \hat{y} , *NSE* can be defined as:
$$NSE = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - E(y))^2}.$$

The *NSE* can be interpreted as test statistic for the accuracy of model predictions. The *NSE* ranges from $-\infty$ to 1: if $NSE = 1$, there is a perfect match of the modeled to

Table 8 Output of Tweedie model

j	Estimate (p -values)				
	Intercept	Sex = Male	Age 2	Age 3	Dispersion
1	1.3223(***)	-0.1442 (***)	0.0019 (***)	0 (***)	79.9719 (***)
2	1.5276(***)	-0.3865 (***)	0.0022 (***)	0 (***)	502.5076 (***)
3	2.2128(***)	-0.2299 (***)	0.0026 (***)	0 (***)	152.7613 (***)
4	-1.1675(*)	-0.4614 (***)	0.0039 (***)	0 (***)	191.1936 (***)
5	3.3156(***)	-0.6336 (***)	0.0012 (***)	0 (***)	101.1635 (***)

Signif. codes: 0 (***) 0.001 (***) 0.01 (*) 0.05 (.) 0.1 () 1 '–' Not used

Table 9 NSE values for branch. Two-Part and Tweedie case

j	1	2	3	4	5
Two-part	0.9243	0.8188	0.9812	0.9536	0.9771
Tweedie	0.9242	0.7933	0.9837	0.9534	0.9787

Table 10 $E[Z_j]$ (data in EUR) and $\text{Var}[Z_j]$ (data in EUR squared) for branch, Year 2019

j	Two-part		Tweedie	
	$E[Z_j]$	$\text{Var}[Z_j]$	$E[Z_j]$	$\text{Var}[Z_j]$
1	250,879.22	25,598.871	250,863.33	24,147,372
2	202,379.74	190,682,291	201,987.42	343,698,039
3	1,964,390.73	3,060,383,134	1,962,698.06	3,546,897,983
4	145,696.05	76,122,543	146,204.47	75,822,123
5	913,986.92	406,617,622	913,105.17	368,588,641

the observed data; if $NSE = 0$, the model predictions are as accurate as the mean of the observed data, if $-\infty < NSE < 0$, the observed mean is a better predictor than the model. It is important to note that NSE is a GoF measure that allows to accept or reject the model using 0 as a critical value.

In Table 9: we report results for NSE :

where all the values are compliant with acceptance.

Finally, exploiting Eqs. (2) and (3) it is possible to calculate $E[Z_{i,j}]$ and $\text{Var}[Z_{i,j}]$ in the case of Two-Part model. whereas the Tweedie model gives a direct estimate of this values. By (7) and (8) we can get an estimate of mean and variance of the expenditure for each branch, whose results for 2019 are reported in Table 10

The branch DC has the greatest variance of expenditure. If we consider the ratio between variance and mean representing the variance per unit of mean, the DC value is the greatest (1558 and 1807 in two-part and Tweedie estimate respectively) and the OH value is the lowest (102 and 96 in two-part and Tweedie estimate respectively) confirming the ranking observed in Table 11

4.3 The calculation of the optimal IDR

We show an example of the calculation of optimal IDRs in the case of no-correlation and correlation among branches. Before proceeding, we need to establish the general rule to assess the risk loading m_j for each branch j . To this aim, standard actuarial pricing principles (see Gerber (1979)) use a functional that assigns a loaded premium, called pure premium, to any distribution of the expenditure per policyholder. Traditional “loaded” net premium principles correct the equivalence principle to generate a premium that is bigger than the expected loss. In the following context we use the simplest method, the so-called *Equivalence Premium*, that starting from the expected loss assigns a positive “safety loading” to avoid ruin with a certainty proportional to the expected loss, i.e. $m_j = \beta \cdot E[Z_j]$, $\beta > 0$. Therefore, in the following we consider $\beta = 10\%$ for each branch and the risk loading m_j ($j = 1, \dots, 5$) are obtained according to the values $E[Z_j]$ show in Table 10. To find the path of the optimal solutions, we solve the optimal problem proposed in Eq. (14) considering several decreasing values of the expected contribution (\bar{G}) according to a proportion $0 \leq \delta \leq 1$ of the maximum contribution $E[G(\mathbf{1})]$, i.e. $\bar{G} = \delta \cdot E[G(\mathbf{1})]$. These solutions are equivalent to consider an overall IDR (or AV) equal to δ .

4.3.1 The no-correlation case

Assuming the absence of correlation among branches, the optimization problem in Eq. (14) can be solved numerically using Eq. (15) where the objective function is easily expressed as the weighted sum of the $\text{Var}[Z_j]$, with weights α_j .

As an alternative, identical solutions can be achieved by following the De Finetti “algorithm” (for further details see Pressacco and Ziani (2010)) proposed in his famous “problema de pieni” (De Finetti (1940)). The de Finetti “algorithm”, in the uncorrelated case, can be summarized as follows:

- i) order branches by the terms $F_j(\mathbf{1}) = \frac{\sigma_j^2}{m_j}$, $j = 1, \dots, J$, which is the expression of the advantage function in Eq. (16) in the no-correlation case;
- ii) by reducing the expected gain to a level \bar{G} , the IDRs must be reduced consequently, following the order defined in step i : the first IDR to be reduced is the one for which $\frac{\sigma_j^2}{m_j}$ is maximum and the last to be reduced is the one for which $\frac{\sigma_j^2}{m_j}$ is minimum;
- iii) For any $0 \leq \lambda \leq \max_j \frac{\sigma_j^2}{m_j}$, an optimal retention is given by: $\alpha_j^* = \min \left\{ \frac{\lambda}{\frac{\sigma_j^2}{m_j}}, 1 \right\}$

As previously stated λ plays the role of the benefit parameter. Table 11 shows the values of $F_j(\mathbf{1})$ obtained using values in Table 10.

As observable, according to the de Finetti algorithm we expect that the first branch to move is *DC* and the last *OH*. This is confirmed by value reported in Table 12 where the optimal IDR respect to a decreasing contribution δ , according to the Two-Part regression model. Similar values are obtained with the Tweedie regression model.

Fig. 1 gives a straight representation of the concept of the functioning of the de Finetti algorithm. Indeed, the first branch reduced is *DC* as it has the maximum $F_3(\mathbf{1})$, whereas the last one is *OH* where is observed the minimum $F_1(\mathbf{1})$. It is worth noting that, that the second highest value of the ratio is $F_2(\mathbf{1})$ (*SC*) where the Tweedie model shows a higher

Table 11 de Finetti ordered $F_j(\mathbf{1})$ values (data in EUR). No-Correlation case

Acronym	Two-Part		Tweedie	
	$F_j(\mathbf{1})$	Ranking	$F_j(\mathbf{1})$	Ranking
OH	1020	5	963	5
SC	9422	2	17,016	2
DC	15,579	1	18,072	1
RPC	5224	3	5186	3
SMV	4449	4	4037	4

Table 12 No-correlation case—optimal IDRs for different values of expected contribution reduction δ . Two-Part regression model. Mean and standard deviation in EUR

Acronym	δ				
	100%	80%	60%	40%	20%
OH	100.00%	100.00%	100.00%	100.00%	100.00%
SC	100.00%	100.00%	55.80%	31.77%	12.39%
DC	100.00%	64.60%	33.75%	19.21%	7.49%
RPC	100.00%	100.00%	100.00%	57.29%	22.34%
SMV	100.00%	100.00%	100.00%	67.28%	26.24%
$E[G(\alpha^*)]$	347,733.27	278,186.61	208,639.96	139,093.31	69,546.65
$\sigma[G(\alpha^*)]$	61,313.90	44,452.37	30,269.15	19,152.56	8,802.89

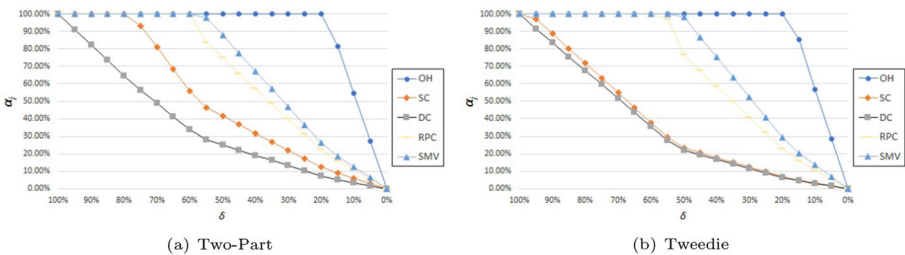


Fig. 1 No-correlation case—optimal IDR paths for different values of overall IDR δ

value compared to the Two-Part model. This involves a quicker movement from 1 to 0 of SC IDR concerning higher values of δ , in the Tweedie model compared to Two-Part one.

4.3.2 The general case of correlated branches

To provide an analysis on correlated branches, it is necessary to introduce the correlation matrix Ω and, consequently, calculate the variance/covariance matrix Σ , concerning the two regression models provided.

$$\Omega = \begin{bmatrix} 1.0000 & 0.0540 & 0.1048 & 0.0959 & 0.2521 \\ 0.0540 & 1.0000 & 0.0122 & 0.1537 & 0.0877 \\ 0.1048 & 0.0122 & 1.0000 & 0.0294 & 0.0926 \\ 0.0959 & 0.1537 & 0.0294 & 1.0000 & 0.1495 \\ 0.2521 & 0.0877 & 0.0926 & 0.1495 & 1.0000 \end{bmatrix} \tag{24}$$

Table 13 de Finetti ordered $F_j(\mathbf{1})$ values (data in EUR). Correlation case

Acronym	Two-Part		Tweedie	
	$F_j(\mathbf{1})$	Ranking	$F_j(\mathbf{1})$	Ranking
OH	3,534	5	3,493	5
SC	12,191	2	20,700	1
DC	16,374	1	18,914	2
RPC	9,565	3	9,917	3
SMV	6,415	4	6,072	4

Table 14 Correlation case - Optimal IDR for different values of expected contribution reduction δ . Two-Part regression model. Mean and standard deviation in EUR

Acronym	δ				
	100%	80%	60%	40%	20%
OH	100.00%	100%	100%	100%	100%
SC	100.00%	89%	46%	25%	10%
DC	100.00%	66%	37%	21%	8%
RPC	100.00%	100%	67%	33%	12%
SMV	100.00%	100%	100%	70%	26%
$E[G(\alpha^*)]$	347,733.27	278,186.61	208,639.96	139,093.31	69,546.65
$\sigma[G(\alpha^*)]$	65,403.81	48,797.18	34,190.49	21,949.99	10,364.45

$$\Sigma_{Two-Part} = \begin{bmatrix} 25,598,871 & 3,775,780 & 29,334,013 & 4,234,214 & 25,721,650 \\ 3,775,780 & 190,682,291 & 9,329,362 & 18,511,149 & 24,418,098 \\ 29,334,013 & 9,329,362 & 3,060,383,134 & 14,206,070 & 103,304,284 \\ 4,234,214 & 18,511,149 & 14,206,070 & 76,112,543 & 26,297,286 \\ 25,721,650 & 24,418,098 & 103,304,284 & 26,297,286 & 406,617,622 \end{bmatrix} \tag{25}$$

$$\Sigma_{Tweedie} = \begin{bmatrix} 24,147,372 & 4,923,393 & 30,671,320 & 4,104,565 & 23,784,895 \\ 4,923,393 & 343,698,039 & 13,484,093 & 24,804,849 & 31,212,104 \\ 30,671,320 & 13,484,093 & 3,546,897,983 & 15,264,418 & 105,884,565 \\ 4,104,565 & 24,804,849 & 15,264,418 & 75,822,123 & 24,989,563 \\ 23,784,895 & 31,212,104 & 105,884,565 & 24,989,563 & 368,588,641 \end{bmatrix} \tag{26}$$

Once again, the optimization problem in Eq. (14) can be solved numerically and, as an alternative, de Finetti “algorithm” can be improved. Anyway, the latter, is based on the advantage function in Eq. (16) where the $F_j(\alpha) = \lambda$ depends on variance and covariance between branches. Hence, a more complex algorithm compared to the one introduced in the no-correlation case is required. For further details, the reader can refer to Pressacco and Serafini (2007). However, the basic rule of the maximum advantage function to identify starting point for the first declining risk is respected. Table 13 shows the ranking obtained in the correlated branches cases.

As one can see in Table 14, the first branch to move is *DC* and the last is *OH* as in the no-correlation case. Anyway, this time, the Tweedie model does not show similar values as

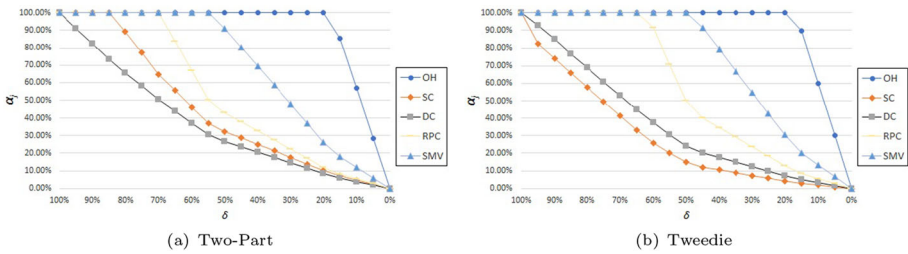


Fig. 2 Correlation case - Optimal IDR paths for different values of overall IDR δ

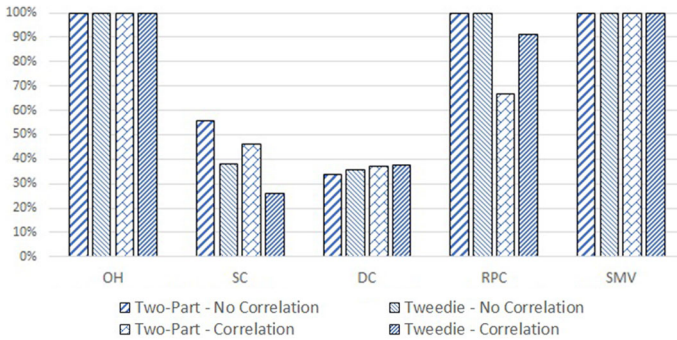


Fig. 3 A comparison of optimal IDR solutions with $\delta = 0.6$ with no correlated and correlated branches

shown in Fig. 2. Indeed, the ranking in this case changes, as the branch with maximum value of $F_i(\mathbf{1})$ is *SC* instead of *DC* (see Table 13) as in the Two-Part model.

5 Conclusion

This study investigates a methodological framework to assess optimal reimbursement strategy for employer-based health insurance plans. The reimbursement strategy is measured in terms of the percentage of expenditure reimbursed by the health plan defined in actuarial literature as Indicated Deductible Relativity and in the U.S. Actuarial Value. The IDR may be different for each category of services or branch provided by HP and the optimal strategy may be found assuming as objective the minimization of the risk, measured through the variance, or maximizing the expected gain. To this aim, we propose an optimization problem similar to the one proposed and solved by (De Finetti (1940)) in the context of proportional reinsurance. Firstly, we adopt a basic actuarial regression model based on a two-part approach (Negative Binominal - Gamma model) and we compare it with a Tweedie regression model. The latter has the main advantage to analyze the distribution of the loss, representing the total expenditure per policyholder, as a whole. This distribution is particularly indicated considering the large number of zero claims we observe in the data for each group of services. The regression models allow for the calculation of the mean and variance of the expenditure r.v. for each policyholder profile considered. Then, assuming independence among policyholders, we assess the overall expected expenditure and variance for each branch. These values are considered as the inputs of the optimization problem whose aim is to identify the optimal IDR for each branch considering the absence or the presence of correlation among branches.

We calibrate our model to a real database that offers five types of services. The findings provided confirm that the advantage function, as defined by de Finetti, represents the main quantity to be controlled to define an optimal strategy. In the case of the HP, the positive correlation observed among branches represents another relevant element to be considered. As observable, in the following Fig. 3 we compare the set of optimal solutions obtained with $\delta = 0.60$.

The quantitative analysis shows that to obtain the same gain, the positive correlation cases require a greater reduction of IDR. In such cases, the IDR is reduced over 50% but this condition may be not feasible as the IDR or AV is a commonly used method by policyholders for scoring a health insurance plan.

In the US HP context, it is relevant that the IDR is not lower than 60% as stated in Obamacare (2010) to ensure that policyholder can have a proper insurance cover. This means that further researches can be addressed to use 60% as lower limit for α in the optimization problem (14). Such new constraint will give a sub-optimal result but is more compliant with indications in Obamacare (2010).

Funding Open access funding provided by Università degli Studi di Salerno within the CRUI-CARE Agreement.

Declarations

Conflict of interest Fabio Baione declares he has no conflict of interests, Davide Biancalana declares he has no conflict of interests, Massimiliano Menzietti declares he has no conflict of interests. This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Basic elements of GLMs

This section deals with the introduction of the essential elements of the Generalized Linear Models, which are regression techniques that are useful in the insurance context, amongst others, and were introduced by Nelder and Mc Cullagh (Mc Cullagh and Nelder (1989)). GLMs are based on the following three building blocks:

1. The dependent variable Y belongs to the exponential family with the probability density function $f(y, \kappa, \varphi) = \exp \left\{ \frac{y\kappa - b(\kappa)}{a(\varphi)} + c(y, \varphi) \right\}$, where a , b and c are known functions; κ and φ are unknown canonical and dispersion parameters, respectively. The function $a(\varphi)$ is commonly in the form $a(\varphi) = \varphi/\omega$, where φ is constant over observation, and ω is a known prior weight that can vary from observation to observation.
2. A linear combination of independent variables is considered as $\vartheta = \sum_h x_h \beta_h$, where β_h are unknown parameters and x_h are given values of predictors.
3. The dependence of the conditional mean is described by a link function g , which is strictly monotonous and twice differentiable $E(Y | x) = E(Y) = \mu = g^{-1}(\vartheta)$.

Table 15 Characteristics of some common univariate distributions in the exponential family

Name	Notation	Density function $f(y, \kappa, \varphi)$	Dispersion parameter φ	Canonical parameter κ	Mean Value $\mu(\kappa)$	Variance Function $\text{Var}(\mu)$
Poisson	$P(\psi)$	$\frac{\psi^y \cdot e^{-\psi}}{y!}$	1	$\log(\psi)$	e^κ	ψ
Neg binomial	$NBin(\mu, \theta)$	$\frac{\Gamma(y+1/\theta)}{\Gamma(y+1)\Gamma(1/\theta)} \cdot \frac{(\theta\mu)^y}{(1+\theta\mu)^{y+1/\theta}}$	θ	$\log(\mu)$	e^κ	$\mu + \theta\mu^2$
Gamma	$G(v, \eta)$	$\frac{1}{\Gamma(v)y} \left(\frac{y}{\eta}\right)^v e^{-\frac{y}{\eta}}$	$\frac{1}{v}$	$-\frac{1}{\eta v}$	$-\frac{1}{\kappa}$	μ^2

The expectation and variance can be obtained under the assumption that b is twice continuously differentiable, as follows:

$$E[Y] = b'(\kappa), \quad \text{Var}[Y] = a(\varphi) b''(\kappa). \quad (27)$$

where the latter expression can be rewritten using the variance function which is defined as:

$$\begin{aligned} \text{Var}(\mu) &= b'' \left[\left(b' \right)^{-1}(\mu) \right] \\ \text{Var}(Y) &= a(\varphi) \cdot V(\mu) \end{aligned} \quad (28)$$

It means that the variance only depends on the mean value. The maximum likelihood method is used to estimate the parameters of GLMs.

In the following Table 15, the members of the exponential family we refer to herein are shown.

The gamma distribution is positive right-skewed, with a sharp peak and a long tail to the right. Gamma is the most adopted distribution in practice for modeling severity.

Let (\mathbf{x}_i, y_i) be a member of the set of observations ($i = 1, \dots, r$) where y_i is a dependent variable in the regression equation, and $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$ is a row-vector of independent variables (covariates). The conditional mean for the i th policyholder is given by:

$$E(Y|\mathbf{x}_i) = E(Y^{(i)}) = g^{-1}(\mathbf{x}_i \cdot \beta) \quad (29)$$

where β is the column-vector of regression coefficients, introduced in point 2 of this paragraph. In some cases, the choice of the link function g is bound with the features of statistical procedures.

In applications, the choice is often made to simplify practical considerations. In the insurance ratemaking context, a logarithm function is usually selected as it allows to combine two models by means of a factorization of their conditional expectation. Then a log-link function eq. (29) could be re-written as:

$$E\left(Y^{(i)}\right) = e^{\mathbf{x}_i \cdot \beta} \quad (30)$$

Moreover, GLM provide also an estimate of conditioned variance and of the whole conditioned probability distribution by estimating the dispersion parameter φ . Overdispersion is a phenomenon that is often observed in practice where the variance need not to be equal to the expected value and φ scales the relationship between the variance and the mean as in (28). An estimate of the dispersion parameter could be provided in several ways. The most robust model error estimation is based on Pearson's chi-square statistic X^2 (see Mc Cullagh and

Nelder (1989)). An approximately unbiased estimator of φ is:

$$\hat{\varphi}_X = \frac{\varphi \cdot X^2}{r - k} = \frac{1}{r - k} \sum_i \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (31)$$

The solution follows statistical theory where X^2 is approximately χ_{r-k}^2 -distributed, with k standing for the number of regression coefficients, so that $E[X^2] \sim r - k$.

The knowledge of conditional mean and variance for each ($i = 1, \dots, r$) allows to well-define the parameters for the exponential family distributions used in GLMs.

References

- Carroll H., Mange J. (2017). The actuarial role in self-insurance: A health section strategic initiative. In Health Watch Issue 84, SOA.
- Clemente, G. P., Savelli, N., Spedicato, G. A., & Zappa, D. (2022). Modeling general practitioners' total drug costs through GAMLSS and collective risk models. *North American Actuarial Journal*, 26(4), 610–625.
- Daykin, C. D., Pentikainen, T., & Pesonen, M. (1993). *Practical risk theory for actuaries*. Chapman and Hall/CRC.
- De Finetti, B., (1940). Il problema dei pieni. *Giornale dell'istituto italiano degli attuari*
- Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *J of Business and Economic Statistics*, 1, 115–126.
- Duncan I., (2018). Healthcare risk adjustment and predictive modeling, (2nd edn), ACTEX Learning.
- Duncan, I., Loginov, M., & Ludkovski, M. (2016). Testing alternative regression frameworks for predictive modeling of health care costs. *North American Actuarial Journal*, 20(1), 65–87.
- Frees, E. W. (2010). *Regression modeling with actuarial and financial applications*. Cambridge University Press.
- Frees, E. W. (2014). Frequency and severity models. In E. W. Frees, G. Meyers, & R. A. Derrig (Eds.), *Predictive modeling applications in actuarial science*. Cambridge University Press.
- Frees, E. W., Gao, J., & Rosenberg, M. A. (2011). Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal*, 15(3), 377–392.
- Frees, E. W., Jin, X., & Lin, X. (2013). Actuarial applications of multivariate two-part regression models. *Annals of Actuarial Science*, 7, 258–287.
- Garrido J., Genest C., Schulz J., (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, Elsevier, vol. 70, (C), pp. 205–15.
- Gerber, H., (1979). An introduction to mathematical risk theory. Huebner Foundation.
- Govorun, M., Latouche, G., & Loisel, S. (2015). Phase-type aging modeling for health dependent costs. *Insurance Mathematics and Economics*, 62, 173–183.
- Kaas, R., (2005). Compound poisson distribution and GLM's-Tweedie's distribution in proceedings of the Contact Forum "3rd actuarial and financial mathematics day", pp. 3–2. Brussels: Royal Flemish Academy of Belgium for Science and the Arts.
- Keeler, E. B., & Rolph, J. E. (1988). The demand for episodes of treatment in the health insurance experiment. *Journal of Health Economics*, 7, 337–367.
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions* (4th ed.). Wiley.
- Kurz, C. F. (2017). Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Medical Research Methodology*, 17, 171.
- Lee, G. Y. (2016). General insurance deductible ratemaking. *North American Actuarial Journal*, 21(4), 620–638.
- Mc Cullagh, P. A., & Nelder, J. A. (1989). *Generalized linear models*. Taylor & Francis.
- Ministero della Salute, Le attività dell'Anagrafe fondi sanitari e i dati del Sistema Informativo Anagrafe Fondi, Rome (2021).
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-a discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- OECD.: Health at a glance 2021: OECD indicators, OECD Publishing, (2021). <https://doi.org/10.1787/ac3016b9-en>
- Pressacco, F., Ziani, L., (2010). Bruno de Finetti forerunner of modern finance. Convegno di studi su Economia e Incertezza, Trieste, 23 ottobre 2009, Trieste, EUT Edizioni Università di Trieste, pp. 65–84.

- Pressacco, F., & Serafini, P. (2007). The origins of the mean-variance approach in finance: Revisiting de Finetti 65 years later. *Decisions in Economics and Finance*, 30, 19–49.
- Rechel, B. (2019). Funding for public health in Europe in decline? *Health Policy*, 123(1), 21–26. <https://doi.org/10.1016/j.healthpol.2018.11.014>
- The patient protection and affordable care act of 2010, Pub. L. No. 111-148, 110 Stat. 2033 (2010). <https://www.govinfo.gov/content/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- Tikkanen, R., Osborn, R., Mossialos, E., Djordjevic, A. and Wharton, G., (2020). International profiles of health care systems [online] Commonwealthfund.org. Available at: https://www.commonwealthfund.org/sites/default/files/2020-12/International_Profiles_of_Health_Care_Systems_Dec2020.pdf
- Tweedie, M. C. K. (1956). Some statistical properties of inverse Gaussian distributions. *Virginia Journal of Science New Series*, 7, 160–165.
- van Winssen, K. P., van Kleef, R., & C., van de Ven, W.P. (2016). Potential determinants of deductible uptake in health insurance: How to increase uptake in The Netherlands? *The European Journal of Health Economics*, 17(9), 1059–1072.
- Vimont, A., Leleu, H., & Durand-Zaleski, I. (2021). Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France. *The European Journal of Health Economics*, 23(2), 211–223.
- Yang, S.-Y., Wang, C.-W., & Huang, H.-C. (2016). The valuation of lifetime health insurance policies with limited coverage. *Journal Risk and Insurance*, 83, 777–800.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.