



ORIGINAL RESEARCH PAPER

# Optimizing insurance risk assessment: a regression model based on a risk-loaded approach

Zinoviy Landsman<sup>1,2</sup>  and Tomer Shushi<sup>3</sup>

<sup>1</sup>Actuarial Research Center, Department of Statistics, University of Haifa, Haifa, Israel; <sup>2</sup>Faculty of Sciences, Holon Institute of Technology, Holon, Israel; and <sup>3</sup>Department of Business Administration, Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel

**Corresponding author:** Zinoviy Landsman; Email: [landsman@stat.haifa.ac.il](mailto:landsman@stat.haifa.ac.il)

(Received 20 November 2023; revised 20 April 2024; accepted 24 April 2024; first published online 31 May 2024)

## Abstract

Risk measurement and econometrics are the two pillars of actuarial science. Unlike econometrics, risk measurement allows taking into account decision-makers' risk aversion when analyzing the risks. We propose a hybrid model that captures decision-makers' regression-based approach to study risks, focusing on explanatory variables while paying attention to risk severity. Our model considers different loss functions that quantify the severity of the losses that are provided by the risk manager or the actuary. We present an explicit formula for the regression estimators for the proposed risk-based regression problem and study the proposed results. Finally, we provide a numerical study of the results using data from the insurance industry.

**Keywords:** Loss functions; optimization; regression theory; penalty loss function; trade-off coefficient

## 1. Introduction

Regression theory plays an essential role in actuarial science. Using regression models, one obtains predictions for the losses based on a priori knowledge of the explanatory variables considered by the risk manager or actuary. Several papers developed and studied regression theory in the context of risks and risk measures (He *et al.*, 2020; Keilbar & Wang, 2022; Gaglianone *et al.*, 2011; Xiao *et al.*, 2015; Barkai *et al.*, 2021; Rennie & Srebro, 2005; Buccini *et al.*, 2020; Ki Kang *et al.*, 2020; Pitselis, 2020; Jeong & Valdez, 2020; Daouia *et al.*, 2021). Unlike standard regression analysis, when dealing with risks one commonly wishes to focus on risk severity levels, which can be quantified by expected losses. In this paper, we develop a regression model that both minimizes the standard OLS error function and the loss function of the severity of the studied risk. We prove that the minimized solutions, that is, the regressors  $\tilde{\beta}$  of the model, can be explicitly solved, which allows us to capture the fundamental behavior of the slopes for the given econometric problem. We then study prototypical examples of such a general approach and test it with empirical analysis.

Let  $\mathbf{Y}$  be a risk with historical data  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ ,  $X$  the  $n \times p$  designed matrix that defines the dependent variables  $(x_1, x_2, \dots, x_n)$ , and  $\beta \in \mathbb{R}^p$  be the vector of slopes. Then, the linear regression takes the form  $\mathbf{Y} = X\beta + \epsilon$ , where  $\epsilon$  is the error term and the total error is given by  $\varepsilon_{Total}(\beta) = \|\mathbf{Y} - X\beta\|^2$ , whose minimization with respect to  $\beta$  gives the minimum least square estimator:

$$\hat{\beta} = \arg \inf_{\beta \in \mathbb{R}^p} \varepsilon_{Total}(\beta) = (X^T X)^{-1} (X^T \mathbf{Y}). \quad (1)$$

If

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n), \quad (2)$$

where  $\sigma > 0$  is a dispersion parameter,  $I_n$  is an  $n \times n$  identical matrix that satisfies all of the BLUE assumptions, and  $\hat{\boldsymbol{\beta}}$  is the BLUE estimator and also an MLE estimator of multivariate parameter  $\boldsymbol{\beta}$ . We consider the case in which matrix  $X^T X$  is not singular, that is,

$$\text{Rank}(X) = p. \quad (3)$$

We define the loss intensity of the problem by  $I_Y = E(\|\mathbf{Y}\|^2) = \sum_{i=1}^n EY_i^2 = \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} + n\sigma^2$  quantified by the total sum of expected squares of the losses. Given a loss function  $\mathcal{L} : u \in \mathbb{R}_{\geq 0} \Rightarrow \mathcal{L}(u) \geq 0$ , we consider  $\mathcal{L}(I_Y)$  as the severity level of the loss intensity  $I_{Y_n}$ . Following a standard linear regression  $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , our multi-objective problem is given by:

$$\min \|\mathbf{Y} - X\boldsymbol{\beta}\|^2 \text{ AND } \min \mathcal{L}(I_Y). \quad (4)$$

In practical terms, loss intensity measures  $\mathcal{L}(I_Y)$  can take many forms depending on the specific risks in the system. For example, in insurance, it might involve calculating the average claim payout per policyholder over a certain period. It can also be a way to assess the potential financial impact of different types of risks, such as natural disasters, cybersecurity breaches, or credit risks. The measure  $\mathcal{L}(I_Y)$  has a special role in risk measurement. The square root of the loss intensity  $I_Y^{1/2} = \sqrt{E(\|\mathbf{Y}\|^2)}$  is an objective measure that provides us an index of the magnitude of loss in the system in the units of the loss, for example, in US dollars, while the function  $\mathcal{L}$  is subjective and depends on the decision-maker. The function  $\mathcal{L}$  quantifies how the decision-maker is risk-averse toward the loss intensity, and  $\lambda$  as the trade-off parameter quantifies the weight we assign to the consideration of  $\mathcal{L}$ , that is, how important it is to minimize  $\mathcal{L}$  compared to the minimization of the regression error  $\|\mathbf{Y} - X\boldsymbol{\beta}\|^2$ .

As we wish to find  $\boldsymbol{\beta}$  that minimizes both the error term  $\|\mathbf{Y} - X\boldsymbol{\beta}\|^2$  and the loss function  $\mathcal{L}(I_Y)$ , the severity is quantified by the sum of expected squares of the losses. We set a trade-off goal function between them in the following standard trade-off form:

$$F_{\mathcal{L},\lambda}(\boldsymbol{\beta}) = \|\mathbf{Y} - X\boldsymbol{\beta}\|^2 + \lambda \mathcal{L}(I_Y), \quad (5)$$

with a trade-off parameter  $\lambda > 0$ .  $F_{\mathcal{L},\lambda}(\boldsymbol{\beta})$  considers the balance sum of two functionals: the error term and the penalty term of the minimized goal function.

This paper explores the extension of the regression theory that captures the magnitude of loss in the system of risks. Taking the proposed goal function (5) for obtaining the regression parameters  $\boldsymbol{\beta}^*$  provides a more general framework than the classical regression model, in which one captures both the minimization of the error provided by  $\|\mathbf{Y} - X\boldsymbol{\beta}\|^2$  while also taking into account the minimization of  $\mathcal{L}(I_Y)$ .

In this paper, we explore the problem of minimization of (5) and prove that it has explicit closed-form solutions in both an unrestricted (Section 2) and a restricted (Section 3) model. The proposed findings hold crucial implications within actuarial analysis. Specifically, we demonstrate that the ratio of intensities, represented as:

$$r = \frac{I_{Y^*}}{I_{\hat{Y}}} \leq 1,$$

where  $I_{Y^*}$  is the loss intensity obtained by the results of Sections 2 and 3, and  $I_{\hat{Y}}$  is the intensity obtained by the classical least square method. Notice that our main theorems do not require the normality assumption of the residuals  $\boldsymbol{\varepsilon}$ , as we highlighted in Section 2, Remark 1. This might be particularly useful in insurance data, where right-skewed distributions are often observed.

Regression analysis in actuarial science plays a vital role by modeling relationships between variables to assess risk and predict future outcomes. Actuaries often employ regression to analyze historical data, such as insurance claims or mortality rates, to develop models that estimate

the impact of various factors on future events. By utilizing regression techniques, actuaries can make informed decisions regarding pricing, risk assessment, and policy development within the insurance and financial sectors. In that follows Section 4 is devoted to a numerical illustration of the risk-loaded approach using real data. In our numerical study demonstrating the risk-loaded approach in regression analysis, we focus on the claims experience of a prominent property and casualty insurer in the Midwestern United States. Specifically, we delve into the realm of private passenger automobile insurance. Within this context, the variable we're analyzing as the dependent factor is the monetary value paid on closed claims, measured in US dollars.

## 2. Unrestricted $F_{\mathcal{L},\lambda}(\beta)$ model

In this section, we consider problem of minimization of functional (5) and for while we do not apply any restrictions on the choice of regression coefficients  $\beta = (\beta_1, \dots, \beta_p)^T$ . The following theorem represents just such a case:

**Theorem 1.** Assume functions  $\mathcal{L}'(y) \geq 0$ ,  $\mathcal{L}''(y) \geq 0$ , and define  $a = \mathbf{Y}^T \mathbf{X} \hat{\beta}$ . If the univariate equation

$$w = \frac{1}{1 + \lambda \mathcal{L}'(aw^2)}, \quad (6)$$

has the solution  $w^*$ , this solution is unique, and the solution of the problem of minimization of functional (5) has the following explicit form:

$$\beta^* = \arg \inf_{\beta \in \mathbb{R}^p} F_{\mathcal{L},\lambda}(\beta) = w^* \hat{\beta}, \quad (7)$$

where the coefficient  $w^*$  is depending on  $\mathbf{Y}^T \mathbf{X} \hat{\beta}$ .

*Proof.* For the proof of the theorem, we will need the following Lemma, the proof of which we have included in the appendix.

**Lemma 1.** Let  $\alpha \in \mathbb{R}^n$  be some vector and  $A > 0$  be some positive definite matrix. Consider the following optimization problem:

$$\mathcal{F}(\mathbf{x}) = F(\alpha^T \mathbf{x}, \mathbf{x}^T \mathbf{A} \mathbf{x}) \rightarrow \inf.$$

Suppose bivariate function  $F(x, y)$  is twice continuously differentiable in the feasible space  $\mathcal{X} \subset \mathbb{R} \times \mathbb{R}_+$  and functional  $\mathcal{F}(\mathbf{x})$  is convex. Then, the analytic solution to the optimization problem (5) is as follows:

$$\mathbf{x}^* = \frac{x_1^*}{A_{1*}^{-1} \alpha} A^{-1} \alpha,$$

where  $A_{1*}^{-1}$  is the first row of matrix  $A^{-1}$ , that is,  $A_{1*}^{-1} \alpha = \sum_{j=1}^n A_{1j}^{-1} \alpha_j$ , and  $x_1^*$  is the root of the univariate equation:

$$x_1 = -\frac{1}{2} G(ax_1, bx_1^2) A_{1*}^{-1} \alpha, \quad (8)$$

where

$$G(x, y) = \frac{F'_x(x, y)}{F'_y(x, y)},$$

$$a = \frac{1}{A_{1*}^{-1} \alpha} \alpha^T A^{-1} \alpha,$$

and

$$b = \frac{1}{(A_{1*}^{-1}\alpha)^2} \alpha^T A^{-1} \alpha.$$

Using this Lemma, we now prove the problem of minimization of functional (5), which we represent as the following problem:

$$\mathcal{F}_1(\beta) = F(\alpha^T \beta, \beta^T A \beta) \rightarrow \inf, \quad (9)$$

where

$$F(x, y) = b - 2x + y + \lambda \mathcal{L}(y), \quad (10)$$

and  $\alpha = X^T Y$ ;  $A = X^T X$ ;  $b = Y^T Y$ . Now we show that the functional (9) is convex. We first observe that

$$F'_x(x, y) = -2 \quad (11)$$

$$F'_y(x, y) = 1 + \lambda \mathcal{L}'(y).$$

$$F''_{xx}(x, y) = F''_{xy}(x, y) = 0,$$

$$F''_{yy}(x, y) = \lambda \mathcal{L}''(y),$$

and we have

$$\Delta = F''_{xx}(x, y)F''_{yy}(x, y) - F''_{xy}(x, y)^2 = 0. \quad (12)$$

As the first derivatives of function  $\mathcal{L}(y)$  is also nonnegative, we use Lemma 3.1 of Landsman *et al.* (2020) and conclude that the considered functional (9) is convex. Then, we use Lemma 1 and find an analytic solution of problem (5). Since

$$F'_x(x, y) = -2 \quad (13)$$

$$F'_y(x, y) = 1 + \lambda \mathcal{L}'(y),$$

we obtain  $G(x, y) = -\frac{2}{1 + \lambda \mathcal{L}'(y)}$ , and the solution of the optimization problem (5) is given by the following explicit form:

$$\begin{aligned} \beta^* &= \frac{x_1^*}{(X^T X)_{1*}^{-1} (X^T Y)} (X^T X)^{-1} (X^T Y) \\ &= \frac{x_1^*}{\hat{\beta}_1} \hat{\beta}, \end{aligned} \quad (14)$$

where  $x_1^*$  is a solution of the following equation:

$$x_1 = \frac{1}{1 + \lambda \mathcal{L}'\left(\frac{a}{\hat{\beta}_1^2} x_1^2\right)} \hat{\beta}_1. \quad (15)$$

Recall and verify that  $\hat{\beta}_1 = (X^T X)_{1*}^{-1} (X^T Y)$ . After designating  $w = x_1 / \hat{\beta}_1$ , we reduce equations (15) to (6). Since the functional (9) is convex, the solution of equation (6) is unique and vector  $\beta^*$  presented in expression (14) is a minimum point of functional (9). It is not superfluous to notice that

$$a = \mathbf{Y}^T X \hat{\boldsymbol{\beta}} = \mathbf{Y}^T K \mathbf{Y} \geq 0, \quad (16)$$

where  $K = X(X^T X)^{-1} X^T$  is an idempotent matrix and is, consequently, a nonnegative definite. Therefore,  $a w^2 \in R_+$ .  $\square$

**Corollary 1.** As  $\mathcal{L}'(y) \geq 0$  from equations (6) and (16) immediately follows that  $0 \leq w^* \leq 1$ .

This corollary has an important value for understanding the actuarial and economic sense of risk-loaded regression result. Define  $\hat{\mathbf{Y}} = X \hat{\boldsymbol{\beta}}$  and  $\mathbf{Y}^* = X \boldsymbol{\beta}^*$ . Then for empirical intensities  $I_{\hat{\mathbf{Y}}} = \hat{\boldsymbol{\beta}}^T X^T X \hat{\boldsymbol{\beta}}$ ,  $I_{\mathbf{Y}^*} = \boldsymbol{\beta}^{*T} X^T X \boldsymbol{\beta}^*$ , it follows that the ratio of intensities:

$$r = \frac{I_{\mathbf{Y}^*}}{I_{\hat{\mathbf{Y}}}} = w^{*2} \leq 1. \quad (17)$$

**Remark 1.** The conditions of Theorem 1 do not include the assumption of normal distribution of residuals  $\boldsymbol{\varepsilon}$  (cf. equation (2)).

**Remark 2.** As can be seen from the proposed model, the desired coefficients  $\boldsymbol{\beta}^* = w^* \hat{\boldsymbol{\beta}}$  are biased estimators stemming from the risk-loaded term  $\lambda \mathcal{L}(E(\mathbf{Y}^T \mathbf{Y}))$ , in such a way that the rate of the risk-loaded  $\boldsymbol{\beta}^*$  is  $w^*$ , since the biased estimator is achieved by  $\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}$  which is the solution when  $\lambda = 0$ , or  $\mathcal{L}(E(\mathbf{Y}^T \mathbf{Y})) = 0$ .

**Remark 3.** The proposed risk-loaded approach for regression analysis of risks modifies the standard concept of regression theory by adding a penalty term for the uncertainty of the risk and allows the decision-maker to choose different functionals to both the error term  $\|\mathbf{Y} - X\boldsymbol{\beta}\|^2$  and  $E(\mathbf{Y}^T \mathbf{Y})$ . The simplest form of  $\mathcal{L}$ ,  $\mathcal{L}(u) = u$ , is the celebrated Ridge regression.

### 3. The equality constrained $F_{\mathcal{L},\lambda}(\boldsymbol{\beta})$ model

In certain situations, it is possible to possess non-sample information (a priori information on the  $\boldsymbol{\beta}$  parameters), which can vary in nature. We specifically focus on precise a priori information regarding coefficients. Let us now scrutinize the proposed risk-based regression model under equality constraints. Generally, this prior information on the coefficients can be represented as follows: we address the issue (5) with linear constraints:

$$R\boldsymbol{\beta} = \mathbf{r}, \quad (18)$$

where  $R$  is matrix,  $m \leq p$ , and  $\mathbf{r}$  is  $m \times 1$  vector. A notable and significant example of such constraints is

$$\beta_1 + \beta_2 + \dots + \beta_p = 1, \quad (19)$$

where matrix  $R = \mathbf{1}^T$ ,  $\mathbf{1}$  is merely a vector-column of ones, and  $\mathbf{r} = 1$ . This constraint allows us to regard the coefficients  $\beta_i$ s as weights of the factors involved in the regression model. To solve the optimization problem (5) under constraints (18), we use Theorem 3.1 of Landsman *et al.* (2020), where such a problem was solved under even more general conditions. Denote the following vectors

$$\boldsymbol{\beta}_0 = (X^T X)^{-1} R^T (R(X^T X)^{-1} R^T)^{-1} \mathbf{r} \quad (20)$$

$$\boldsymbol{\beta}_1 = (X^T X)^{-1} X^T \mathbf{Y} - (X^T X)^{-1} R^T (R(X^T X)^{-1} R^T)^{-1} R(X^T X)^{-1} X^T \mathbf{Y}$$

and the following values

$$\alpha_1 = \mathbf{Y}^T X \boldsymbol{\beta}_1, \quad \alpha_2 = \boldsymbol{\beta}_0^T X^T X \boldsymbol{\beta}_0. \quad (21)$$

**Theorem 2.** Under the conditions of Theorem 1, if the following univariate equation

$$w = \frac{1}{1 + \lambda \mathcal{L}'(\alpha_2 + \alpha_1 w^2)}, \quad (22)$$

has the solution  $w^*$ , then this solution is unique, and the solution of the problem of minimization of functional (5) under linear constraints (18) has the following explicit form:

$$\boldsymbol{\beta}_R^* = \arg \inf_{R\boldsymbol{\beta}=\mathbf{r}} F_{\mathcal{L},\lambda}(\boldsymbol{\beta}) = \boldsymbol{\beta}_0 + w^* \boldsymbol{\beta}_1. \quad (23)$$

The vectors  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\beta}_1$  are orthogonal with respect to  $X^T X$  in the sense that

$$\boldsymbol{\beta}_0^T X^T X \boldsymbol{\beta}_1 = 0. \quad (24)$$

*Proof.* The proof immediately follows from Theorem 3.1 of Landsman *et al.* (2020), and we also refer to the main results given in Landsman *et al.* (2018). In fact, the problem (5), subject to the system of linear constraints (18), is a special case of Theorem 3.1, where  $F(x, y)$  has a form (10), equation (18) of Theorem 3.1 becomes (22), and the solution of the optimization problem (18) of Theorem 3.1 becomes (23). Vectors  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\beta}_1$  are calculated by vectors  $\mathbf{x}^0$  and  $\mathbf{z}$  given in Theorem 3.1, respectively, where matrices  $\Sigma = X^T X$ ,  $B = R$ , and vectors  $\mathbf{c} = \mathbf{r}$  and  $\boldsymbol{\mu} = X^T \mathbf{Y}$ . The orthogonal property (24) immediately follows from the orthogonal property of vectors  $\mathbf{x}^0$  and  $\mathbf{z}$ , see equation (19) of Theorem 3.1.  $\square$

Note that the vectors  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\beta}_1$  are not dependent on the loss function  $\mathcal{L}(u)$  but are dependent only on matrix  $X$ , vector  $\mathbf{Y}$ , restriction matrix  $R$ , and vector  $\mathbf{r}$ . When  $w^* = 1$ , we have the classical minimum least squared estimator under restriction (19):

$$\hat{\boldsymbol{\beta}}_R = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1. \quad (25)$$

Function  $\mathcal{L}(u)$ , in turn, determines equations (6) and (22).

As in the unrestricted case considered in Section 2, we have the following

**Corollary 2.** As  $\mathcal{L}'(y) \geq 0$  from equations (22) and (16) immediately follows that  $0 \leq w^* \leq 1$ .

We define the empirical intensities of  $\hat{\mathbf{Y}}_R = X \hat{\boldsymbol{\beta}}_R$  and  $\mathbf{Y}_R^* = X \boldsymbol{\beta}_R^*$ , by:

$$I_{\hat{\mathbf{Y}}_R} = \hat{\boldsymbol{\beta}}_R^T X^T X \hat{\boldsymbol{\beta}}_R \quad (26)$$

and

$$I_{\mathbf{Y}_R^*} = \boldsymbol{\beta}_R^{*T} X^T X \boldsymbol{\beta}_R^*, \quad (27)$$

respectively. Then, Corollary 2 and the orthogonal property of  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\beta}_1$  (see, (24)) immediately lead to the following intensity ratio:

$$\begin{aligned} r &= \frac{I_{\mathbf{Y}_R^*}}{I_{\hat{\mathbf{Y}}_R}} = \frac{\boldsymbol{\beta}_R^{*T} X^T X \boldsymbol{\beta}_R^*}{\hat{\boldsymbol{\beta}}_R^T X^T X \hat{\boldsymbol{\beta}}_R} = \frac{(\boldsymbol{\beta}_0 + w^* \boldsymbol{\beta}_1)^T X^T X (\boldsymbol{\beta}_0 + w^* \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1)^T X^T X (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1)} \\ &= \frac{\boldsymbol{\beta}_0^T X^T X \boldsymbol{\beta}_0 + w^{*2} \boldsymbol{\beta}_1^T X^T X \boldsymbol{\beta}_1}{\boldsymbol{\beta}_0^T X^T X \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^T X^T X \boldsymbol{\beta}_1} \leq 1. \end{aligned} \quad (28)$$

The above bound supports the preferability of a risk-loaded approach.

In the following, we examine special cases for our proposed actuarial and econometric measure, for different functions  $\mathcal{L}$ .

### 3.1 Classical case

The classical case is obtained when  $\mathcal{L} \Leftarrow u \Rightarrow u$ . Then, the solution of equations (6) and (22) reduces to

$$w^* = \frac{1}{1 + \lambda}, \quad (29)$$

and the solution to the optimization unrestricted optimization problem is  $\beta^* = \frac{1}{1+\lambda} \hat{\beta}$ . For the case  $\lambda = 0$ , which is the classical minimum least squared case, we have

$$\beta^* = \hat{\beta}, \quad (30)$$

which conforms well with the solution of the classical problem. Notice that as  $\hat{\beta}$  is unbiased and a consistent estimator of  $\beta$ ,  $\beta^*$  is unbiased and consistent estimator of  $\frac{1}{1+\lambda} \beta$ .

In the context of the equality constrained model from Theorem 2 and (29), it immediately follows that

$$\beta_R^* = \beta_0 + \frac{1}{1 + \lambda} \beta_1. \quad (31)$$

When  $\lambda = 0$ , which is the classical minimum least squared case under linear constraints (18), we have

$$\beta_R^* = \beta_0 + \beta_1 = \hat{\beta}_R, \quad (32)$$

which conforms well with the solution of the classical problem, cf. Özkale & Selahattin (2007), eq (1.16).

### 3.2 Powered penalty function

Assume  $\mathcal{L}(u) = u^\delta, \delta > 0$ . Then equation (6) has the following form:

$$w = \frac{1}{1 + \lambda \delta a^{\delta-1} w^{2\delta-2}}, \quad (33)$$

which can be reduced to the power equation:

$$w + \lambda \delta a^{\delta-1} w^{2\delta-1} = 1. \quad (34)$$

A special case of this example is the case in which  $\delta = 1/2$ . Then, equation (34) has an analytic solution:

$$w^* = 1 - \frac{\lambda}{2\sqrt{a}}, \quad (35)$$

and, the explicit solution of the minimization problem becomes

$$\beta^* = \left( 1 - \frac{\lambda}{2\sqrt{\mathbf{Y}^T X \hat{\beta}}} \right) \hat{\beta}. \quad (36)$$

It is well known that under natural condition

$$\frac{1}{n} X^T X \rightarrow Q \text{ as } n \rightarrow \infty, \quad (37)$$

where  $Q$  is  $p \times p$  matrix and  $\hat{\beta}$  is a consistent estimator of  $\beta$ . We show that in the case that residuals are normally distributed,  $\epsilon \sim N_n(\mathbf{0}, \sigma^2 I_n)$ ,  $\beta^*$  is also consistent estimator of  $\beta$ . Recall  $\beta^* = w^* \hat{\beta}$ , where  $w^* = 1 - \frac{\lambda}{2\sqrt{a}}$ , and  $a = \mathbf{Y}^T X \hat{\beta} = \mathbf{Y}^T K \mathbf{Y}$ , where  $K = X(X^T X)^{-1} X^T$ . As  $K^2 = X(X^T X)^{-1} X^T = K$ , matrix  $K$  is idempotent. Recall that matrix  $X$  has maximal rank equal

to  $p$  and vector  $\mathbf{Y}$  is distributed  $N_n(X\boldsymbol{\beta}, \sigma^2 I_n)$ . Then  $\mathbf{Y}/\sigma \sim N_n(X\boldsymbol{\beta}/\sigma, I_n)$  and  $a/\sigma^2 \sim \chi_n(\Delta)$ , where  $\chi_n(\Delta)$  is a non-central chi-squared distribution with  $p$  degrees of freedom, and  $\Delta = \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T X \boldsymbol{\beta}$  is a non-centrality parameter. Then,  $E(a/n) = (n + \Delta) = 1 + \frac{1}{n} \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \rightarrow 1$ , and  $\text{Var}(a) = 2(n + 2\Delta) = 2n + 4\boldsymbol{\beta}^T X^T X \boldsymbol{\beta}$ . From Chebyshev's inequality follows that

$$P\left(\left|\frac{a}{n} - E\left(\frac{a}{n}\right)\right| > \varepsilon\right) \leq \frac{2n + 4\boldsymbol{\beta}^T X^T X \boldsymbol{\beta}}{n^2 \varepsilon^2} = \frac{2}{n\varepsilon^2} + \frac{4}{n\varepsilon^2} \boldsymbol{\beta}^T \frac{X^T X}{n} \boldsymbol{\beta} \rightarrow 0. \quad (38)$$

This implies that

$$\frac{a}{n} \xrightarrow{P} 1$$

and then

$$w^* = 1 - \frac{\lambda}{2\sqrt{a}} \xrightarrow{P} 1,$$

that is,  $\boldsymbol{\beta}^*$  is a consistent estimator of  $\boldsymbol{\beta}$  for this case.

The same happens even for the more general case,  $1/2 \leq \delta < 1$ . In fact, it follows from equation (33) that:

$$w^* = \frac{1}{1 + \lambda\delta[(a/n)^{1-\delta} n^{1-\delta} w^{*2(1-\delta)}]^{-1}} \xrightarrow{P} 1, n \rightarrow \infty. \quad (39)$$

Considering the second special case  $\delta = 1$ , we obtain from the power equation (34) that  $w^* = \frac{1}{1+\lambda}$ , which conforms well with Subsection 3.1, but in this case,  $\boldsymbol{\beta}^*$  is not a consistent estimator of  $\boldsymbol{\beta}$ , but a consistent estimator of  $\frac{1}{1+\lambda} \boldsymbol{\beta}$ .

#### 4. Numerical analysis

We illustrate the risk-loaded approach for regression analysis by conducting a numerical study. We consider claims experience from a large Midwestern (US) property and casualty insurer for private passenger automobile insurance. The dependent variable is the amount paid on a closed claim, in (US) dollars (claims that were not closed by year end are handled separately). The independent variables are State code, vehicle Class code, Gender, and Age. We obtained the data from Frees (2010, Table 4). To have a design matrix  $X$  be only numerically valued, we denoted the different class vehicle codes by numbers from 1 to 18 and Gender variable by 1 or 0. In Table 1, we present the first 10 lines (of  $n = 6773$  lines) that, in fact, are the first 10 rows of matrix  $X$  and vector  $\mathbf{Y}$ :

Consider, first, the unrestricted case and powered penalty function. Then the solution of equation (34) is, in fact, the zero of function:

$$F(w) = w + \lambda\delta a^{\delta-1} w^{2\delta-1} - 1. \quad (40)$$

In Fig. 1, we show the graph of function  $F(w)$  for  $\lambda = 0.2$  and  $\delta = 1.1$ , and we can see that this function has the unique root  $w^* = 0.170$ .

In Fig. 2, we provide the graph of the ratio of empirical intensities. One can see that the ratio of intensities decreases from 1 to 0.

Now assume that we have the equality constraint (19). Then equation (22) in Theorem 2 takes of the form:

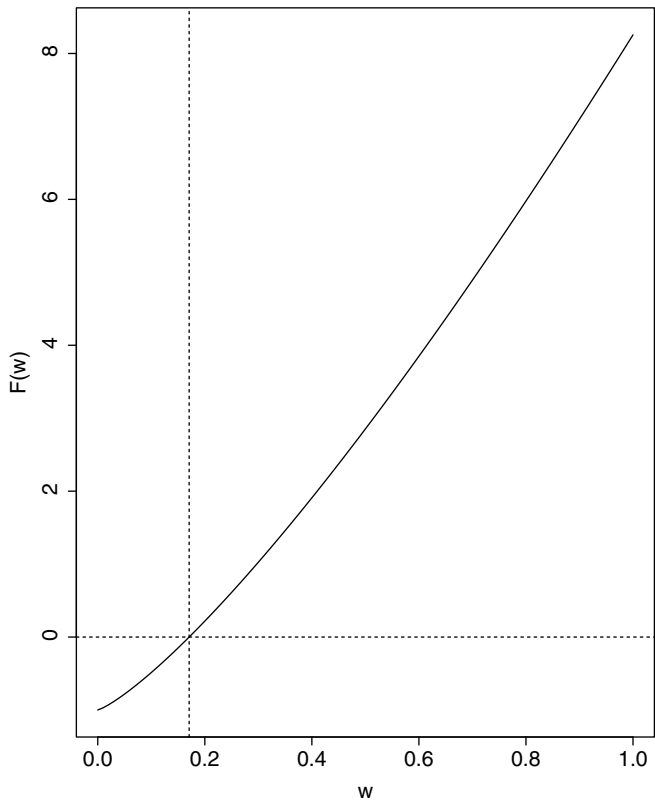
$$F_1(w) = w + \lambda\delta w (\alpha_2 + \alpha_1 w^2)^{\delta-1} - 1 = 0. \quad (41)$$

The solution of this equation is a zero of function  $F_1(w)$ . In Fig. 3, we show the graph of function  $F_1(w)$  for  $\lambda = 0.2$  and  $\delta = 1.1$ . We can see that this function has the unique root  $w^* = 0.341$ .



**Table 1.** First 10 lines of matrix **X** and vector **Y**

Obs.	Intercept	State code From(1–18)	Class code From (1–17)	Gender Mail= 1, Female= 0	Age From (70–97)	Paid Y
1	1	14	1	0	97	1134.44
2	1	15	1	0	96	3761.24
3	1	15	2	0	95	7842.31
4	1	15	3	1	95	2384.67
5	1	15	3	0	95	650
6	1	15	3	0	95	391.12
7	1	15	2	0	94	3775.83
8	1	10	1	0	94	415.35
9	1	14	2	0	93	2283.56
10	1	3	2	0	93	665.48



**Figure 1.** Graph of function  $F(w)$  for  $\lambda = 0.2$  and  $\delta = 1.1$ ;  $w^* = 0.1701$ .

Using obtained  $w^*$ , we can provide the solution of the problem (5) under linear constraints (18), that is, coefficients  $\beta_R^*$  as follows:

In the following table, we provide the classical estimators for comparison with the risk-loaded estimators.

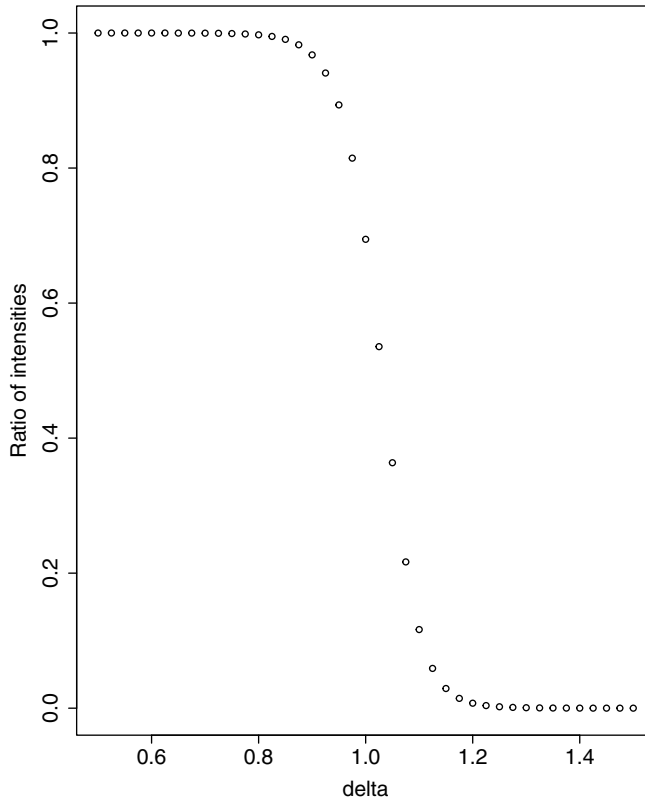


Figure 2. Changing the ratio of empirical intensities  $r = I_{Y^*} / I_{\hat{Y}}$  when power parameter  $\delta$  increases from 0.5 to 1.5.

In order to collate these two solutions for the minimum least squared problem, we calculate the ratio of empirical intensities:

$$r = \frac{I_{Y_R^*}}{I_{\hat{Y}_R}} \quad (42)$$

using the expression (28). In the considered numerical case  $r = 0.116$ , this implies that the intensity is significant lower using risk-loaded approach. In addition, we can say that as the  $\beta_R$ -coefficients satisfy the restriction (19), they indicate the weights of the factors such as Intercept, State, Class Vehicle, Gender, and Age in the considered regression model. Comparing the risk-loaded estimators reported in Table 2 with the classical estimators given in Table 3, we can conclude that the amplitude of changing weights for the risk-loaded estimators is much less than for the classical regression estimators. Regarding the difference between non-restricted and restricted models, we can observe that for the considered numerical data, the intensity ratio for the unrestricted approach, that is,  $r = 0.029$ , which is lower than for the restricted case. This result is quite natural because in the non-restricted case, we have an absolute minimum. However, in some sense, the unrestricted case can be considered as useless because we cannot interpret the meaning of each  $\beta$ -coefficient.

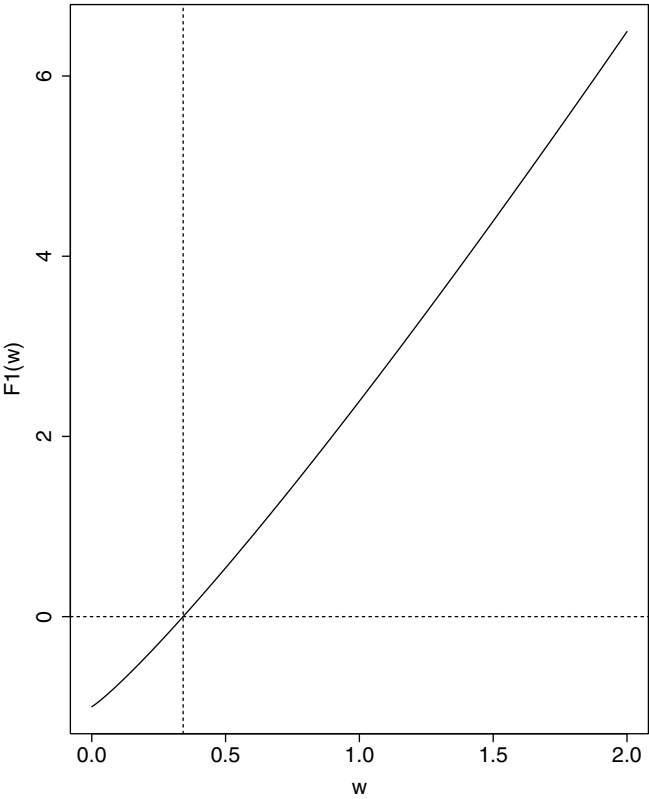
In Fig. 4, we show how the ratio of intensities changes when power parameter  $\delta$  increases from 0.5 to 1.5. In this figure, we see that the ratio of intensities decreases when  $\delta$  increases from 0.5 to 1.5 and that its graph is very similar to the graph given in Fig. 2.

**Table 2.** Solution of Theorem 2 under power risk-loaded function

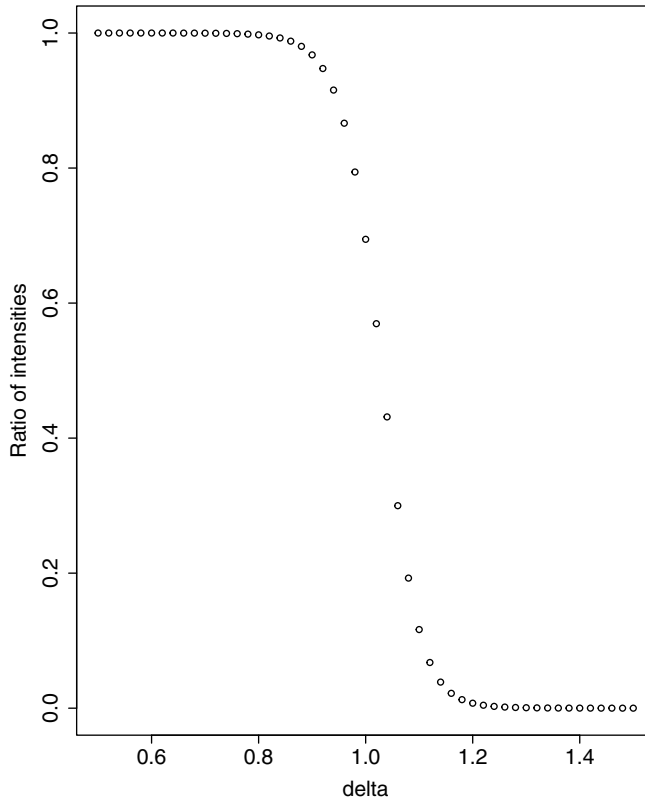
Variables	Coefficients ( $\beta_R^*$ )
Intercept	−16.028
State	2.909
Class Vehicle	14.272
Gender	−8.266
Age	8.112

**Table 3.** The classical minimum least squared estimator under restriction (19)

Variables	Coefficients ( $\hat{\beta}_R$ )
Intercept	−48.894
State	8.527
Class Vehicle	41.842
Gender	−24.260
Age	23.784



**Figure 3.** Graph of function  $F_1(w)$  for  $\lambda = 0.2$  and  $\delta = 1.1$ ;  $w^* = 0.34146$ .



**Figure 4.** Changing of the ratio of empirical intensities when power parameter  $\delta$  increases from 0.5 to 1.5: Linear constraint is present.

Notice that sometimes the data need to be log-transformed. In our case, it is unnecessary because our factors (State, Vehicle Class, Gender, and Age) are integers, and the relatively low ratios ( $r$ ) indicate a reasonably good level of adequacy.

## 5. Conclusion

In this paper, we proposed a hybrid risk-loaded regression model that simultaneously minimizes the classical least squared loss function and penalty loss function of suggested loss intensity of the problem, being the sum of expected squared historical data of responses  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ . Imposing rather general conditions on the model and the loss function, we find that the explicit solution of the minimization problem takes a proportional form of the least squared estimator  $\hat{\beta}$ , where the coefficient of proportionality depends on response vector  $\mathbf{Y}$  and design matrix  $X$ . Special attention was given to the powered penalty loss function, which essentially generalized the classical case of identical loss function. In addition to the unconditional problem, we also considered a situation in which the solution satisfies the linear equality constraints. In that case, we also found the explicit closed-form solution. In both the unconditional and conditional cases, we demonstrated that the ratio of empirical intensities for classical least squared estimator and for hybrid risk loaded estimator proposed in the paper is always less than 1, which speaks in favor of the proposed method. We also provided a numerical illustration of the proposed model using real data.

**Acknowledgment.** The authors would like to thank the reviewer for the comments, which significantly improved the paper.

**Data availability statement.** Replication materials are available on request from the authors. The data and code that support the findings of this study are available from the corresponding author [Zinoviy Landsman] upon reasonable request.

**Funding statement.** This work received no specific grant from any funding agency, commercial, or not-for-profit sectors.

**Competing interests.** The author(s) declare none.

## References

- Barkai, I., Shushi, T., & Yosef, R. (2021). A cryptocurrency risk–return analysis for bull and bear regimes. *The Journal of Alternative Investments*, **24**(1), 95–118.
- Buccini, A., De la Cruz Cabrera, O., Donatelli, M., Martinelli, A., & Reichel, L. (2020). Large-scale regression with non-convex loss and penalty. *Applied Numerical Mathematics*, **157**, 590–601.
- Daouia, A., Gijbels, I., & Stupfler, G. (2021). Extremile regression. *Journal of the American Statistical Association*, **117**, 1579–1586.
- Frees, E. W. (2010). Instructors' Manual for Regression Modeling with Actuarial and Financial Applications. Available at: <https://instruction.bus.wisc.edu/jfrees/jfreesbooks/regression%20modeling/bookwebdec2010/DataDescriptions.pdf>. Accessed 24 May 2024.
- Gaglianone, W. P., Lima, L. R., Linton, O., & Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, **29**(1), 150–160.
- He, Y., Hou, Y., Peng, L., & Shen, H. (2020). Inference for conditional value-at-risk of a predictive regression. *The Annals of Statistics*, **48**(6), 3442–3464.
- Jeong, H., & Valdez, E. A. (2020). Predictive compound risk models with dependence. *Insurance: Mathematics and Economics*, **94**, 182–195.
- Keilbar, G., & Wang, W. (2022). Modelling systemic risk using neural network quantile regression. *Empirical Economics*, **62**(1), 93–118.
- Ki Kang, S., Peng, L., & Golub, A. (2020). Two-step risk analysis in insurance ratemaking. *Scandinavian Actuarial Journal*, **2021**, 532–542.
- Landsman, Z., Makov, U., & Shushi, T. (2018). A generalized measure for the optimal portfolio selection problem and its explicit solution. *Risks*, **6**(1), 19.
- Landsman, Z., Makov, U., & Shushi, T. (2020). Portfolio optimization by a bivariate functional of the mean and variance. *Journal of Optimization Theory and Applications*, **185**(2), 622–651.
- Pitselis, G. (2020). Multi-stage nested classification credibility quantile regression model. *Insurance: Mathematics and Economics*, **92**, 162–176.
- Rennie, J. D., & Srebro, N. (2005). Loss functions for preference levels: Regression with discrete ordered labels. In Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling, Menlo Park, CA. AAAI Press, vol. 1.
- Xiao, Z., Guo, H. & Lam, M. S. (2015) Quantile regression and value at risk. In *Handbook of financial econometrics and statistics* (pp. 1143–1167). Springer.
- Özkale, M. R., & Selahattin, K. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics—Theory and Methods*, **36**(15), 2707–2725. doi: [10.1080/03610920701386877](https://doi.org/10.1080/03610920701386877).

## Appendix

### Proof of Lemma 1:

To find the stationary points of the optimization problem (5), we obtain the following system of equations:

$$\frac{d}{d\mathbf{x}} F(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) = F'_x(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) \boldsymbol{\alpha} + 2F'_y(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A \mathbf{x} = \mathbf{0}.$$

After algebraic calculations, we obtain  $\mathbf{x} = -\frac{1}{2} G(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A^{-1} \boldsymbol{\alpha}$ . Thus, we have the following system of equations:

$$\begin{cases} x_1 = -\frac{1}{2} G(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A_{1*}^{-1} \boldsymbol{\alpha} \\ x_2 = -\frac{1}{2} G(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A_{2*}^{-1} \boldsymbol{\alpha} \\ \dots \\ x_n = -\frac{1}{2} G(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A_{n*}^{-1} \boldsymbol{\alpha} \end{cases}, \quad (\text{A1})$$

where  $A_{i*}^{-1}$  is the  $i$ -th row of matrix  $A^{-1}$ . Dividing the  $i$ -th equation,  $i = 2, \dots, n$ , into the first equation, we obtain

$$\frac{x_i}{x_1} = \frac{A_{i*}^{-1} \boldsymbol{\alpha}}{A_{1*}^{-1} \boldsymbol{\alpha}}, \quad i = 2, \dots, n,$$

where, without loss of generality, we assume that  $A_{1*}^{-1} \boldsymbol{\alpha} \neq 0$ . Then  $x_i = \frac{A_{i*}^{-1} \boldsymbol{\alpha}}{A_{1*}^{-1} \boldsymbol{\alpha}} x_1$ ,  $i = 2, \dots, n$ , and hence  $\mathbf{x}$  can be represented by the first variable  $x_1$ , in the following manner:

$$\mathbf{x} = \frac{x_1}{A_{1*}^{-1} \boldsymbol{\alpha}} A^{-1} \boldsymbol{\alpha}, \quad (\text{A2})$$

where  $x_1$  is the solution of the first equation of system (A1). Substituting (A2) into this equation, we obtain equation:

$$\begin{aligned} x_1 &= -\frac{1}{2} G(\boldsymbol{\alpha}^T \mathbf{x}, \mathbf{x}^T A \mathbf{x}) A_{1*}^{-1} \boldsymbol{\alpha} = -\frac{1}{2} G\left(\frac{x_1}{A_{1*}^{-1} \boldsymbol{\alpha}} \boldsymbol{\alpha}^T A^{-1} \boldsymbol{\alpha}, \frac{x_1^2}{(A_{1*}^{-1} \boldsymbol{\alpha})^2} \boldsymbol{\alpha}^T A^{-1} A A^{-1} \boldsymbol{\alpha}\right) A_{1*}^{-1} \boldsymbol{\alpha} \\ &= -\frac{1}{2} G\left(\frac{x_1}{A_{1*}^{-1} \boldsymbol{\alpha}} \boldsymbol{\alpha}^T A^{-1} \boldsymbol{\alpha}, \frac{x_1^2}{(A_{1*}^{-1} \boldsymbol{\alpha})^2} \boldsymbol{\alpha}^T A^{-1} \boldsymbol{\alpha}\right) A_{1*}^{-1} \boldsymbol{\alpha}, \end{aligned}$$

If this equation has a solution, this solution is unique due to the convexity of function  $F$ , and the obtained stationary point given in (A2) is a minimum point. The Lemma is proved