Journal of Applied Statistics
Vol. 00, No. 00, 00 2012, 1–17

RESEARCH ARTICLE

MODELLING SME LOAN DEFAULTS AS RARE EVENTS: THE GENERALIZED EXTREME VALUE REGRESSION MODEL

Raffaella Calabrese[a]* and Silvia Angela Osmetti[b]

[a] *University of Milano-Bicocca, Italy*; [b] *University Cattolica del Sacro Cuore, Italy*

(*v4.4 released XXXXXX*)

A pivotal characteristic of credit defaults that is ignored by most credit scoring models is the rarity of the event. The most widely used model to estimate the Probability of Default (PD) is the logistic regression model. Since the dependent variable represents a rare event, the logistic regression model shows relevant drawbacks, for example underestimation of the default probability, which could be very risky for banks. In order to overcome these drawbacks we propose the Generalized Extreme Value (GEV) regression model. In particular, in a Generalized Linear Model (GLM) with binary dependent variable we suggest the quantile function of the GEV distribution as link function, so our attention is focused on the tail of the response curve for values close to one. The estimation procedure used is the maximum likelihood method. This model accommodates skewness and it presents a generalisation of GLMs with complementary log-log link function. We analyse its performance by simulation studies. Finally, we apply the proposed model to empirical data on Italian Small and Medium Enterprises (SMEs).

Keywords: credit defaults; Small and Medium Enterprises; Generalized Linear Model; Generalized Extreme Value distribution; rare events; binary data.

1. INTRODUCTION

Credit risk forecasting is one of the leading topics in modern finance, as the bank regulation has made increasing use of external and internal credit ratings (Basel Committee on Banking Supervision, 2004). Statistical credit scoring models try to predict the probability that a loan applicant or existing borrower will default over a given time-horizon, usually of one year. According to the Basel Committee on Banking Supervision (2004), banks are required to measure the one year default probability for the calculation of the equity exposure of loans. In this framework, banks adopting the Internal-Rating-Based (IRB) approach are allowed to use their own PD estimates. Moreover, Basel II requires these banks to build a rating system and provides a formula for the calculation of minimum capital requirements where the PD is the main input. For that reason, in many credit risk models such as CreditMetrics (Gupton et al., 1997), CreditRisk+ (Credit Suisse Financial Products, 1997) or CreditPortfolioView (Wilson, 1998), default probabilities are essential input parameters.

Altman (1968) was the first to use a statistical model to predict the default probabilities of firms, calculating his well-known Z-Score using a standard discriminant model. Almost a decade later Altman et al. (1977) modified the Z-Score by extending the dataset to larger-sized and distressed firms. Even if different methodologies

*Corresponding author. Email: raffaella.calabrese1@unimib.it

ISSN: 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664763.YYYY.XXXXXX
http://www.tandfonline.com

are now widely used for credit risk assessment (Hand and Henley 1997), most of the academic literature (Altman and Sabato, 2006; Aziz et al., 1998; Becchetti and Sierra, 2002; Charitou and Trigeorgis, 2002; Gentry et al., 1985; Keasey and Watson, 1987; Lizal, 2002; Mossman et al., 1998; Ooghe et al., 1995; Platt and Platt, 1990; Zavgren, 1983) uses logistic regression to predict default.

Only a few authors (e.g. Kiefer, 2010) regard default as a rare event; this means that the number of defaults in a sample is very small. The logistic regression shows important drawbacks in rare events studies: the probability of rare event is underestimated and the logit link is a symmetric function, so the response curve approaches zero at the same rate it approaches one. Moreover, commonly used data collection strategies are inefficient for rare event data (King and Zeng, 2001). The bias of the maximum likelihood estimators of logistic regression parameters in small sample sizes, which has been well analysed in the literature (McCullagh and Nelder, 1989; Manski and Lerman 1977; Hsieh Manski and McFadden, 1985), is amplified in the rare event study. Most of these problems are relatively unexplored in the literature (King and Zeng, 2001).

Since for banks the underestimation of the PD is very risky, the main aim of this paper is to overcome the drawbacks of logistic regression for estimating the PD. We propose a new model for binary dependent data with an asymmetric link function given by the quantile function of the Generalized Extreme Value (GEV) random variable. In the extreme value theory, the GEV distribution is used to model the tail of a distribution (Kotz Nadarajah, 2000; Coles, 2004). Since we focus our attention on the tail of the response curve for values close to 1, we have chosen GEV distribution. In GLMs (Agresti, 2002), the log-log and the complementary log-log link functions are used, since they are asymmetric functions. In particular, the complementary log-log link function is the quantile function of the Gumbel random variable. The inverse function of the log-log is one minus the cumulative distribution function of the Gumbel random variable. These models represent particular cases of the GEV regression model.

We apply the model proposed here to data on Italian SMEs. SMEs play a very important role in the economic system of many countries and particularly in Italy (about 90% of Italian firms are SMEs, see Vozzella and Gabbi, 2010). Furthermore, a large part of the literature (Altman and Sabato, 2006; Ansell et al., 2009; Berger, 2004; Ciampi and Gordini 2008; Figini and Fantazzini, 2009; Figini and Giudici, 2011; Jacobson et al., 2005; Saurina and Trucharte, 2004; Vozzella and Gabbi, 2010) has focused on the special character of small business lending and the importance of relationship banking for solving information asymmetries. The information asymmetries puzzle particularly affects SMEs due to the difficulty of estimating and making known their fair value. Therefore, lending to SMEs is riskier than lending to large corporates (Altman and Sabato, 2006; Dietsch and Petey, 2004; Saurina and Trucharte, 2004). As a consequence, Basel II (Basel Committee on Banking Supervision, 2004) establishes that banks should develop credit risk models specifically addressed to SMEs. Only a few studies consider SMEs (Altman and Sabato, 2006; Altman et al. 2010; Figini and Fantazzini, 2009; Figini and Giudici, 2011), since it is relatively difficult to gather data on SMEs.

The present paper is organised as follows. In the next section we explain the main drawbacks of the logistic regression model for estimating the PD. In Section 3 we propose the GEV model for credit defaults. In Subsection 3.1 we present the Weibull regression model as a particular case of the GEV model. In Section 4 we analyse the performance of the proposed model through Monte Carlo simulations. Finally, in Section 5 we apply our proposal to empirical data on Italian SMEs. In particular, the first subsection describes the dataset and the second subsection shows the estimation results. In the following subsection, the predictive accuracies of the logistic regression model and the GEV model are compared for different percentages of the defaults in the sample. Finally, the last section is devoted to conclusions. In the appendix, we report the score functions and the Fisher information matrix of the parameters of the GEV model.

2. THE MAIN DRAWBACKS OF THE LOGISTIC REGRESSION FOR CREDIT DEFAULTS AS RARE EVENTS

We consider n borrowers and the event default or non-default of each borrower is represented by a Bernoulli random variable Y_i with parameter π_i that represents the PD of the i-th borrower, for $i = 1, 2, ..., n$. To estimate π_i a Generalized Linear Model (GLM) (McCullag and Nelder, 1989; Dobson and Barnett, 2008) for Y_i considers a monotonic and twice differentiable function $g(\cdot)$, called *link function*, and a covariate vector $\mathbf{x}_i$ such that

$$g(\pi_i) = \boldsymbol{\beta}'\boldsymbol{x}_i.$$

In the logistic regression model the PD π_i is a logistic cumulative distribution function

$$\pi(\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}'\boldsymbol{x}_i)} \tag{1}$$

with

$$\boldsymbol{\beta}' = [\beta_0, \beta_1, ..., \beta_k] \qquad \boldsymbol{x}' = [1, x_1, ..., x_k].$$

The maximum likelihood method is usually used to estimate the parameters vector $\boldsymbol{\beta}$ (McCullag and Nelder, 1989).

The sample number of defaults is very small (Kiefer, 2010), and for this reason default is a rare event. When the logistic regression model is applied to predict defaults and, in general, to rare events, it shows pivotal drawbacks. Firstly, the logistic regression could underestimate the PD. Secondly, commonly used data collection strategies are inefficient for rare event data (King and Zeng, 2001). In order to overcome this drawback choice-based or endogenous stratified sampling (case-control design) is used. The strategy is to select on Y by collecting observations for which $Y = 1$ and a random selection of observations for which $Y = 0$. This sampling method is usually supplemented with a prior correction of the bias of MLE estimators. An alternative procedure is weighting the data to compensate the differences in the sample and population fractions of ones induced by choice-based sampling by the weighted exogenous sampling maximum estimator. Manski and Lerman (1977) and McCullagh and Nelder (1989) show an analytical approximation for the bias in the MLE estimates to account for finite sample. Such bias increases when the logistic regression is applied to rare events. Thirdly, the logit link is symmetric about 0.5

$$logit(\pi(\boldsymbol{x}_i)) = ln\left(\frac{\pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)}\right) = -logit(\pi(\boldsymbol{x}_i)) = -ln\left(\frac{\pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)}\right).$$

This means that the response curve for $\pi(\boldsymbol{x}_i)$ approaches zero at the same rate that it approaches one. If the dependent variable represents a rare event, a symmetric link function is not appropriated. Since a counting rare event is usually modelled by a Poisson distribution (Falk et al, 2010), which has positive skewness, it is

coherent to choose an asymmetric link function in order to obtain a response curve that approaches zero at a different rate than it approaches one.
From the variance matrix

$$V(\hat{\boldsymbol{\beta}}) = \left[\sum_{i=1}^{n} \pi_i(1-\pi_i)\boldsymbol{x}_i'\boldsymbol{x}_i\right]^{-1} \tag{2}$$

we can deduce that default characteristics (covariates of the values one of the dependent variable) are more informative than non-default ones. The part of the matrix (2) affected by the rarity of default event is the factor $\pi_i(1-\pi_i)$. Most credit scoring models yield small estimates of PD $P\{Y_i = 1|\boldsymbol{x}_i\} = \pi_i$ for all borrowers. However, if the logistic regression has some explanatory power, the estimate of π_i for defaults should be larger, and closer to 0.5, because PD studies are normally very small, than among observations for non-defaults. The result is that $\pi_i(1-\pi_i)$ should be larger for ones than zeros and so the variance will be smaller. In this situation, additional defaults will cause the variance to drop more and hence are more informative than additional non-defaults.
For this reason, in this paper, we focus our attention on the tail of the response curve for the values close to one.

3. THE GENERALIZED EXTREME VALUE (GEV) REGRESSION MODEL

Extreme value theory is a robust framework to analyse the tail behaviour of distributions. Embrechts (1999, 2000) considers the potential and limitations of extreme value theory for risk management. Without being exhaustive, De Haan et al. (1994) and Danielsson and de Vries (1997) study quantile estimation. Bali (2003) uses the GEV distribution to model the empirical distribution of returns. McNeil (1999) and Dowd (2002) give an extensive overview of extreme value theory for risk management.
The class of GEV distributions is very flexible with the tail shape parameter τ controlling the shape and the size of the tails of the three different families of distributions subsumed under it. The three families of extreme value distributions can be nested into a single parametric representation, as shown by Jenkinson (1955) and von Mises (1936). This representation is known as the Generalized Extreme Value (GEV) distribution and its cumulative distribution function is given by

$$F_X(x) = \exp\left\{-\left[1+\tau\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\tau}}\right\} \quad -\infty < \tau < \infty \quad -\infty < \mu < +\infty \quad \sigma > 0 \tag{3}$$

defined on $S_X = \{x : 1 + \tau(x-\mu)/\sigma > 0\}$. The parameter τ is a shape parameter, while μ and $\sigma(>0)$ are location and scale parameters respectively.
The Type II (Fréchet-type distribution) and the Type III (Weibull-type distribution) classes of the extreme value distribution correspond respectively to the case $\tau > 0$ and $\tau < 0$, while the Type I class (Gumbel-type distribution) arises in the limit as $\tau \to 0$. The corresponding distributions of $(-X)$ are also called extreme value distributions. We underline that Fréchet and Weibull distributions are related by a change of sign.
In this paper we propose a new GLM to overcome the drawbacks of the logistic model to forecast defaults. We propose a generalisation of the complementary log-log model by using the quantile function of the GEV distribution as link function. We define the proposed model as Generalized Extreme Value (GEV) regression model.

To estimate the PD $\pi(\boldsymbol{x}_i) = P\{Y_i = 1|\boldsymbol{x}_i\}$, we suggest the GEV cumulative distribution function as the response curve

$$\pi(\boldsymbol{x}_i) = \exp\{-[1 + \tau(\boldsymbol{\beta}'\boldsymbol{x}_i)]^{-1/\tau}\}. \tag{4}$$

in a GLM. For $\tau \to 0$ the expression (4) becomes the response curve of the complementary log-log model and for $\tau < 0$ the Weibull response curve, particular cases of the GEV model.
The link function of the GEV model is given by

$$\frac{\{-ln\,[\pi(\boldsymbol{x}_i)]\}^{-\tau} - 1}{\tau} = \boldsymbol{\beta}'\boldsymbol{x}_i \tag{5}$$

that represents a non-canonical link function.
For the interpretation of the parameters $\boldsymbol{\beta}$ and τ, we suppose that the value of the j-th regressor (with $j = 1, 2, ..k$) is increased by one unit and all the other independent variables remain unchanged. Let $\boldsymbol{x}^*$ denote the new covariate values, whereas $\boldsymbol{x}$ denotes the original covariate values. From the equation (5) we deduce that $\beta_j = g(\pi(\boldsymbol{x}^*)) - g(\pi(\boldsymbol{x}))$ with $j = 1, 2, .., k$. This means that if the parameter β_j (with $j = 1, 2, ..k$) is positive and all the other parameters are fixed, by increasing the j-th regressor the estimate of $\pi(\boldsymbol{x})$ decreases. Otherwise, if β_j is negative, by increasing the j-th regressor the estimate of $\pi(\boldsymbol{x})$ also increases.
Moreover, we analyse the parameter β_0: for all fixed values of τ and for null values of the independent variables, β_0 has a positive monotonic relationship with the estimate of $\pi(\boldsymbol{x})$. Finally, we analyse the influence of the τ parameter on $\pi(\boldsymbol{x})$. We find that for $\beta_0 = 0$ and by considering null values for all the covariates, from the GEV model we obtain an estimate of $\pi(\boldsymbol{x})$ that is approximately equal to e^{-1} for all the values of τ. This means that the variations of $\pi(\boldsymbol{x})$ depend on the covariate variations and not on τ variations.
We propose to estimate the parameters of this model using the maximum likelihood method. Let $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ a simple random sample of size n from Y, the log-likelihood function is

$$l(\boldsymbol{\beta}, \tau) = \sum_{i=1}^{n} \left\{-y_i[1 + \tau(\boldsymbol{\beta}'\boldsymbol{x}_i)]^{-1/\tau} + (1 - y_i)ln[1 - \exp[-[1 + \tau(\boldsymbol{\beta}'\boldsymbol{x}_i)]^{-1/\tau}]]\right\}. \tag{6}$$

Some simulation studies are developed to verify the existence of the maximum of the likelihood function, considered as a function of only one parameter for fixed values of the other parameters (likelihood profile function). Since the inverse of the link function (4) is a cumulative distribution function only for the values $\{\boldsymbol{x}_i : 1 + \tau\boldsymbol{x}_i > 0\}$, the (6) exists only for $\{\boldsymbol{x}_i : 1 + \tau\boldsymbol{\beta}'\boldsymbol{x}_i > 0\}$.
The score functions, obtained by differentiating the log-likelihood function with respect to the known parameters $\boldsymbol{\beta}$ and τ (see Appendix) are given by

$$\frac{\partial l(\boldsymbol{\beta}, \tau)}{\partial \beta_j} = -\sum_{i=1}^{n} x_{ij} \frac{\ln[\pi(\boldsymbol{x}_i)]}{1 + \tau\boldsymbol{\beta}'\boldsymbol{x}_i} \frac{y_i - \pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)} \quad j = 0, 1, ..., k; \tag{7}$$

$$\frac{\partial l(\boldsymbol{\beta}, \tau)}{\partial \tau} = \sum_{i=1}^{n} \left[\frac{1}{\tau^2} ln(1 + \tau\boldsymbol{\beta}'\boldsymbol{x}_i) - \frac{\boldsymbol{\beta}'\boldsymbol{x}_i}{\tau(1 + \tau\boldsymbol{\beta}'\boldsymbol{x}_i)}\right] \frac{y_i - \pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)} ln[\pi(\boldsymbol{x}_i)]. \tag{8}$$

The asymptotic standard errors of the maximum likelihood estimators of the parameters in the models are given by the Fisher information matrix (see Appendix).

Since the Fisher information matrix is not a diagonal matrix (see Appendix), the maximum likelihood estimators of the parameters $\boldsymbol{\beta}$ and τ are dependent and they cannot be computed separately.

Since the score functions do not have closed form, the maximum likelihood estimators need to be obtained by numerically maximising the log-likelihood function using iterative optimisation algorithms. Optimisation algorithms require the specification of initial values to be used in an iterative scheme.

Our suggestion is to use as initial point estimate for τ a value close to zero. For this value the GEV model becomes the complementary log-log model. Hence, in order to obtain the initial point estimate for $\boldsymbol{\beta}$, we analyse the complementary log-log or Gumbell regression model (see Agresti, 2002) with the response curve

$$\pi(\boldsymbol{x}_i) = \exp(-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)). \tag{9}$$

We compute the log-likelihood function of the Gumbel regression

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\{y_i \ln[\pi(\boldsymbol{x}_i)] + (1-y_i)\ln[1-\pi(\boldsymbol{x}_i)]\} \qquad (10) \\ &= \sum_{i=1}^{n}\{y_i \ln[\exp[-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)]] + (1-y_i)\ln[1-\exp[-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)]]\} \\ &= \sum_{i=1}^{n}\{y_i[-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)] + (1-y_i)\ln[1-\exp[-\exp(\boldsymbol{\beta}'\mathbf{x}_i)]]\}. \end{aligned}$$

The score functions are given by

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij}\ln[\pi(\boldsymbol{x}_i)]\frac{y_i - \pi(\boldsymbol{x}_i)}{1-\pi(\boldsymbol{x}_i)} \qquad j = 0, 1, ..., k. \tag{11}$$

To identify the initial values for $\boldsymbol{\beta}$, we choose $\beta_j^* = 0$ for $j = 1, ..., k$. By substituting $\beta_j^* = 0$ for $j = 1, ..., k$ in equation (10) we obtain

$$\beta_0^* = \ln[-\ln(\overline{y})].$$

We use the initial values proposed for the complementary log-log model in order to identify the initial values $\boldsymbol{\beta}^*$ for the GEV regression model. In particular, we propose to use $\tau^* \simeq 0$, $\beta_j^* = 0$ for $j = 1, ..., k$ and $\beta_0^* = \ln\left[-ln(\overline{y})\right]$.

Afterwards, by substituting the initial values for the parameter $\boldsymbol{\beta}$ in the equation (8) we obtain the estimate of τ for the first step of the iterative procedure. By using this estimate of τ in the equation (7), we obtain the estimates of β_j with $j = 0, 1, ..., k$ for the first step in the GEV regression.

3.1 WEIBULL REGRESSION MODEL FOR PREDICTING DEFAULTS

A particular case of the GEV cumulative distribution function (3) for $\tau < 0$ is the Weibull-type cumulative distribution function

$$F(x) = \begin{cases} \exp\left\{-\left[-\frac{x-\mu}{\sigma}\right]^k\right\} & x < \mu \qquad -\infty < \mu < +\infty \qquad \sigma > 0 \qquad k > 0; \\ 1 & x > \mu. \end{cases} \tag{12}$$

where μ and $\sigma (> 0)$ are, respectively, a location and a scale parameters and $k = \left|\frac{1}{\tau}\right|$

is a shape parameter.
By considering the cumulative distribution function (12) in the GLM for binary dependent variable, the response curve becomes

$$\pi(\boldsymbol{x}_i) = \exp[-(\boldsymbol{\beta}'\boldsymbol{x}_i)^k], \tag{13}$$

where $k > 0$. We call this model Weibull regression model for binary data. The response curve of the Weibull regression model (13) is a particular case of the GEV response curve (4) for $\tau < 0$. On the one hand, the Weibull response curve is an asymmetric function, analogously to the response curve (9) of the Gumbel regression model. On the other hand, unlike the Gumbel response curve (9), the $\pi(\boldsymbol{x}_i)$ in the Weibull model (13) approaches 1 sharply and approaches 0 slowly. In particular, the behaviour of the Weibull response curve depends on k: if k increases $\pi(\boldsymbol{x}_i)$ approaches both 0 and 1 sharply. If the value of the j-th regressor (with $j = 1, 2, ..k$) is increased and all the other independent variables remain unchanged, the Weibull response curve (13) decreases when $\beta_j > 0$ and increases when $\beta_j < 0$. The link function of the Weibull regression model is

$$\left[\ln\left(\frac{1}{\pi(\boldsymbol{x}_i)}\right)\right]^{1/k} = \boldsymbol{\beta}'\boldsymbol{x}_i. \tag{14}$$

We compute the log-likelihood function of the Weibull regression

$$\begin{aligned} l(\boldsymbol{\beta}, k) &= \sum_{i=1}^{n}\{y_i \ln[\pi(\boldsymbol{x}_i)] + (1 - y_i)\ln[1 - \pi(\boldsymbol{x}_i)]\} \qquad (15)\\ &= \sum_{i=1}^{n}\{-y_i(\boldsymbol{\beta}'\boldsymbol{x}_i)^k + (1 - y_i)\ln[1 - \exp(-(\boldsymbol{\beta}'\boldsymbol{x}_i))^k]\}. \end{aligned}$$

The score functions are given by

$$\frac{\partial l(\boldsymbol{\beta}, k)}{\partial \beta_j} = -k\sum_{i=1}^{n} x_{ij}\frac{\ln[\pi(\boldsymbol{x}_i)]}{\boldsymbol{\beta}'\boldsymbol{x}_i}\frac{y_i - \pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)} \quad j = 0, 1, ..., k; \tag{16}$$

$$\frac{\partial l(\boldsymbol{\beta}, k, \boldsymbol{y})}{\partial k} = -k\sum_{i=1}^{n}\ln[\pi(\boldsymbol{x}_i)]\ln[\boldsymbol{\beta}'\boldsymbol{x}_i]\frac{y_i - \pi(\boldsymbol{x}_i)}{1 - \pi(\boldsymbol{x}_i)}. \tag{17}$$

In order to apply an iterative algorithm, we need to identify the initial values $\boldsymbol{\beta}^*$ and k^* for the parameters. Our suggestion is to use $k^* = 1$, $\beta_j^* = 0$ for $j = 1, ..., k$ and

$$\beta_0^* = \ln\left[1 - \frac{n}{y}\right]. \tag{18}$$

We obtain the initial value (18) by substituting $\beta_j^* = 0$ for $j = 1, ..., k$ and $k^* = 1$ in (16). We highlight that the Weibull regression with $k = 1$ is a log-linear model whose response curve is the cumulative distribution function of an exponential random variable (McCullagh and Nelder, 1989).

4. SIMULATION STUDIES

In this section we compare the performance of the GEV regression model with the probit and the logistic models using Monte Carlo simulations. In our Monte Carlo analysis we generate $1,000$ replications. In order to evaluate how the performance of the models varies according to the number of observations, we consider two different sample sizes: $n = 500$ and $n = 1,500$. We consider three datasets generated from a probit model, a logistic model and a complementary log-log model, respectively. For each model we consider one covariate whose values are drawn from a normal distribution N(0,1) with mean equal to zero and variance equal to one and the parameter vector $\boldsymbol{\beta}' = [0,1]$. The residuals are generated from a normal distribution N(0,1), a logistic distribution with location parameter equal to zero and scale parameter equal to one and a Gumbel distribution with location parameter equal to zero and scale parameter equal to one, respectively.

In order to evaluate how the properties of the three models vary according to a change of the sample percentage of defaults, we consider two different percentages of defaults ($Y = 1$) in the samples, 1% and 5%.

For both sample sizes and for both percentages of one in the samples we evaluate the predictive accuracy of the models using two performance measures: the Mean Square Error (MSE) and the Mean Absolute Error (MAE), defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{19}$$

where y_i and $\hat{y}_i$ are the observed and the predicted dependent variables, respectively. Models with lower MSE and MAE forecast the dependent variable more accurately. The main aim of this subsection is to show that the GEV model overcomes the drawback of the probit[1] and the logistic regression models in the underestimation of PDs. For this reason, we focus our attention on the tail of the response curve for values of the dependent variable equal to one that represents the default. Therefore, we compare the models by computing MAE and MSE only for the defaulters (the values one). This means that in the equations (19) we consider only the positive errors $(y_i - \hat{y}_i) > 0$ and n is the number of defaulters. We denote these errors by MAE$^+$ and MSE$^+$, respectively. In particular, Table 1 reports the average of the MAE$^+$ and the MSE$^+$ computed on 1,000 Monte Carlo samples, for the sample percentage of defaults 5%.

TABLE 1 AROUND HERE

From the results reported in Table 1, our proposal model shows both the MAE$^+$ and the MSE$^+$ lower than the respective errors of the probit and the logistic regression models for both the sample sizes and for the different datasets drawn from the probit, logistic and complementary log-log models. This means that the GEV model overcomes the drawback of the underestimation of the PD by applying the logistic and the probit models.

By comparing the errors for different sample sizes, unlike the probit and logistic models, our model improves its accuracy by increasing the sample size from $n = 500$ to $n = 1,500$.

Table 2 reports the results of the averages of the MAE$^+$ and the MSE$^+$ for the sample percentage of defaults of 1%.

TABLE 2 AROUND HERE

[1] The drawbacks explained in Section 2 are due to a symmetric link function. This means that not only the logistic model but also the probit model show the above-mentioned disadvantages.

Table 2 confirms the previous results: the values of the means of the MAE$^+$ and the MSE$^+$ are lower than those obtained with the probit and the logistic regression models.

Moreover, by comparing the errors in Table 1 and Table 2 for different sample percentages of defaults (rare events) the logistic and the probit models show a worse performance for lower sample percentage of defaults. On the contrary, the GEV model improves its accuracy. From these results the GEV model can be considered a suitable regression model for predicting credit defaults.

5. DEFAULT PREDICTION OF ITALIAN SMALL AND MEDIUM ENTERPRISES

Data used in our analysis comes from AIDA-Bureau van Dijk, a large Italian financial and balance sheet information provider. We consider defaulted and non defaulted SMEs over the years $2005-2011$. This time horizon is particularly important, since it includes the financial crisis of 2008. This event is so relevant that the periods of time after and before the crisis show totally different characteristics of credit risk. The database contains accounting data of around 210,000 Italian firms with total assets below 10 million Euros (Vozzella and Gabbi, 2010). We exclude firms without the necessary information on the covariates from the sample.

In accordance with Basel II, the PD is one year forecasted. Therefore, let Y_t be a binary r.v. such that

$$Y_t = \begin{cases} 1, \text{ if a SME is defaulted at time } t; \\ 0, \text{ otherwise.} \end{cases}$$

and let $\boldsymbol{x}_{t-1}$ be the covariate vector at time $t-1$. We aim at estimating the conditional PD $\pi(\boldsymbol{x}_{t-1}) = P(Y_t = 1|\boldsymbol{x}_{t-1})$ by applying and comparing the GEV and the logistic regression models.

Often default definitions for credit risk models concern single loan defaults of a company versus a bank, as also emerges from the Basel II instructions. This is the case for banks building models based on their portfolio data, i.e. relying on data on individual loans which are reserved (e.g., Altman and Sabato (2006) develop a logit model for Italian SMEs based on the portfolio of a large Italian bank). However, traditional structural models (i.e. Merton, 1974) refer to a firm-based definition of default: a firm defaults when the value of the assets is lower than the value of the liabilities, i.e. when equity is negative. In this work default is intended as the end of the SMEs activity, i.e. the status in which the SME needs to liquidate its assets for the benefit of its creditors. In practice, we consider a default to have occurred when a specific SME enters a bankruptcy procedure as defined by the Italian law. The reason for this choice lies in the availability of data.

In accordance with Altman and Sabato (2006), we apply a choice-based or endogenous stratified sampling to this dataset. In this sampling scheme data are stratified by the values of the response variable. We draw the observations randomly within each stratum defined by the two categories of the dependent variable (1=default, 0=non-default) and we consider all the defaulted firms. Then, we select a random sample of non-defaulted SMEs over the same year of defaults in order to obtain a percentage of defaults in our sample as close as possible to the default percentage (5 %) for Italian SMEs (Cerved Group, 2011), in order to analyse the properties of our model for two percentages of defaults (1% and 5%).

By applying choice-based sampling, the observations are dependent. Since the sample sizes of this application are high, according to the superpopulation theory (Prentice, 1986) we can consider the sample as a simple random sample.

5.1 ESTIMATION RESULTS

We apply the GEV regression model proposed in this work to the AIDA database. This application is interesting since it concerns SMEs, on which the availability of data is very difficult in the Italian credit market, which might differ in other countries.

In order to model the default event, we choose the independent variables that represent financial and economic characteristics of firms according to the recent literature (Vozzella and Gabbi, 2010; Ciampi and Gordini, 2008; Altman et al., 2006). These covariates cover the most relevant aspects of the firm's operations: leverage, liquidity and profitability.

Firstly, we consider 16 covariates: liquidity ratio, current ratio, leverage, solvency ratio, debt/EBITDA, return on equity, return on investment, turnover per employee, added value per employee, cash flow, banks/turnover, debt/equity ratio, return on solvency, EBITDA/turnover, total personnel costs/added value, cash flow/turnover.

Secondly, we examine the multicollinearity and remove the variables with a Variance Inflation Factor higher than 5 (Greene, 2000, p.257-258). Thirdly, by applying the GEV model 7 variables are significant at the level of 5% for the PD forecast:

- *Solvency ratio*: the ratio of a company's income over the firm's total debt obligations;
- *Return on equity*: the amount of net income returned as a percentage of shareholders equity;
- *Turnover per employee*: the ratio of sales divided by the number of employees;
- *Added value per employee*: the enhancement added to a product or service by a company divided by the number of employees;
- *Cash flow*: the amount of cash generated and used by a company in a given period;
- *Bank loans over turnover*: short and long term debts with banks over sales volume net of all discounts and sales taxes;
- *Total personnel costs over added value*: the ratio of a company's labour costs divided by the enhancement added to a product or service by a company.

In order to avoid overfitting, data are randomly divided into two parts: a sample on which the regression models are estimated and a control sample on which we evaluate the predictive accuracy of the models. Table 1 reports the parameter estimates obtained by applying the GEV model to the sample of 1,485 defaulters and 29,700 non-defaulters over the years $2005 - 2008$.

TABLE 3 AROUND HERE

In section 3 we explain the interpretation of the parameters of the GEV model. According to these interpretations we can analyse the influence of each variable in Table 1 on the PD estimate.

At first, Ansell et al. (2009) explain that the solvency ratio should have an inverse relationship with the PD estimate, coherent with our result but in contrast with the result obtained by Ansell et al. (2009). The return on equity and the added value per employee show the same kind of relationship with the PD: the first result coincides but the second one is in contrast with Ciampi and Gordini (2008). We highlight that our result for the added value per employee coincides with the expectations.

On the contrary, the turnover per employee and the cash flow show a direct relationship with PD, coherent with Altman and Sabato (2006) and Ciampi and

Gordini (2008). The last two results in Table 1 are in contrast with the expectations: bank loans divided by turnover and total personnel costs divided by added value show an inverse relationship with the PD estimate. For this reason we analyse the results obtained in the literature for these two variables. Ciampi and Gordini (2008) obtain a direct relationship of bank loans divided by turnover with the PD estimate. Alternatively, Altman and Sabato (2006) consider the short term debt over equity book value to model the PD and they show that this variable has an inverse relationship with the PD estimate, analogously to our result. On the contrary, Fantazzini and Figini (2009) show that the short term debt has a direct influence on PD, coherent with the expectation. Ciampi and Gordini (2008) also consider total personnel costs divided by added value in their regression model but their analysis shows a result coherent with the expectations and in contrast with the one shown in Table 1. Fantazzini and Figini (2009) also consider labour costs in their model. In particular, they introduce the personnel expenses over sales in the regression model and this variable shows a direct influence with the PD estimate.

5.2 PREDICTIVE ACCURACY

Since for banks the underestimation of the PD could be very risky, the main aim of this subsection is to show that the GEV model overcomes the drawback of the logistic regression in the underestimation of rare events.

The performance of models can be highly sensitive to the data sample used for validation. To avoid embedding unwanted sample dependency, models should be validated on observations that are not included in the sample used to estimate the model. Hence, we run out-of-sample and out-of-time validations to compare the GEV and the logistic regression models. For this aim we consider the confusion matrix (Giudici, 2003), the Area Under the Curve (AUC) index (Hand, 2001), the H measure and the MAE and MSE defined in equations (19). We consider the H measure since it overcomes the drawbacks of the AUC when the class sizes and the classification error costs are extremely unbalanced (Hand, 2009 and 2010); both these characteristics are satisfied by credit scoring models.

It is much more costly to classify an SME as non-defaulter when it is a defaulter than to classify an SME as non-defaulter when it is a defaulter. In particular, when a defaulted firm is classified as non-defaulter by the scoring model, banks will give it a loan. If the borrower becomes defaulter, the bank may lose the whole or a part of the credit exposure. On the contrary, when a non-defaulter is classified as defaulter, the bank only loses interest on loans. For this reason, the identification of defaulters is a pivotal aim for banks' internal models. For all these reasons, we compute the MAE and the MSE only for defaulters, denoted MAE$^+$ and MSE$^+$.

Models are validated by using out-of-sample and out-of-time tests. In the out-of-time approach, we estimate the model on the observations from the years $2005 - 2008$. We test the model on three default horizons: one year (observations of 2009), two years (observations of 2009-2010) and three years (observations of 2009-2010-2011). The out-of-sample and out-of-time sample sizes are reported in Table 4 and Table 5.

TABLE 4 AROUND HERE

TABLE 5 AROUND HERE

As mentioned above, we consider PDs 0.05 and 0.01. Analogously to Figini and Giudici (2011), we apply a bootstrap method (Efron and Tibshirani, 1993): for each 1,000 bootstrap samples, the H measure, AUC, MAE$^+$ and MSE$^+$ are

calculated. A test on the H measure is performed to determine if a model performs significantly better than another, at a confidence level of 95%.
TABLE 6 AROUND HERE

For simplicity, we report the confusion matrices of the two models only for the out-of-time sample 2009-2010-2011 in Table 6.
TABLE 7 AROUND HERE

TABLE 8 AROUND HERE

Table 7 and Table 8 report the means of [1]the H measure, AUC, MAE$^+$ and MSE$^+$ on the sample (subscript "c") and on the control sample (subscript "cs"). In order to evaluate the weights of the errors MSE$^+$ and MAE$^+$ on the respective (total) errors MSE and MAE, we compute the mean of the ratios (MSE^+/MSE) and (MAE^+/MAE) weighted by the PDs 0.05 and 0.01. We report their values between round brackets in Table 7 and Table 8.
From Table 7 and Table 8, our proposal shows the means of MAE$^+$ and the MSE$^+$ lower than the respective errors of the logistic regression model for both the sample and the control sample and for both the PDs. By comparing the percentages in round brackets, we deduce that for the logistic regression model the weights of misclassification of defaults are relevant. On the contrary, for our proposal the weights of these errors are negligible in both the out-of sample and the out-of-time samples. This means that the GEV model overcomes the drawback of the logistic regression in the underestimation of PD.
Since the errors for the sample and the control sample are similar, the covariates are significant for the default discrimination of both the regression models. This means that both the models are well explained. In Table 8 the means of MSE$^+$ and MAE$^+$ show that our model is more accurate in forecasting defaults than the logistic model when the time horizon changes and by considering a substantial change in credit risk after 2008 due to the financial crisis. Our model improves its accuracy by reducing the sample percentage of defaulters, while the logistic regression model shows worse performance.
From Table 7 and Table 8, the average H measure for the GEV model is always higher than the average H measure for the logistic regression model; this relationship is inverted by the AUC. To understand this result we should consider that the H measure is equivalent to averaging the misclassification loss over a cost ratio distribution which enables us to represent the highly unbalanced misclassification costs. Instead, this weight function in the AUC depends on the score distributions (Hand, 2009), so different classifiers are incoherently evaluated using different metrics.
We compare the performances of the two models by testing whether the difference between the H measures is significantly different from zero at the confidence level of 95% level. The approximate p-value, obtained by a nonparametric bootstrap procedure (Davison and Hinkley, 1997), allows us to reject the hypothesis that the models show the same performance.
Finally, in order to analyse the robustness of the GEV model we estimate the coefficients of both the regression models on a sample with a given sample percentage of defaults and we evaluate the accuracy on a sample with a different

[1] To compute the H measure and AUC, we use the H measure package of R-program. For the H measure we consider a severity ratio of 0.01.

sample percentage of defaults. By computing the average MAE$^+$ and the average MSE$^+$ on all the SMEs, the errors of our model do not change in comparison with the respective errors on a sample with the same percentage of defaulters. This means that our model is robust for different sample percentages of defaulters. On the contrary, for the logistic regression models the same errors are significantly different.

5. CONCLUSIONS REMARKS

In this work we aim to propose a new GLM regression model to overcome the drawbacks of the logistic regression model in underestimating the PD in credit risk analysis. As is well known, the GEV distribution is a suitable function for modelling extreme values and rare events data. For this reason we propose the quantile function of the GEV distribution as link function in a GLM for binary dependent variables. The GEV model depends on the regression parameters and on the shape parameter of the GEV distribution. Since the score functions do not have closed form, we obtain the maximum likelihood estimators by maximising the log-likelihood function using an iterative algorithm. We specify initial values of parameters for this iterative algorithm and the Fisher information matrix. The main advantage of the GEV model is its excellent performance to identify defaults. Thanks to this characteristic, the drawback of the logistic regression model of underestimating the PD is overcome. This result is highlighted by simulation and empirical studies in this work.

We analyse empirical data on Italian Small and Medium Enterprises (SMEs) and we model their PD over the years 2005-2011 by considering financial and economic covariates. In the validation of the GEV and the logistic regression models, we show the substantial underestimation of the PD by applying the logistic regression model. By reducing the sample percentage of defaults, the predictive performance of the logistic regression model to identify defaults becomes worse. On the contrary, the accuracy of the GEV model to identify defaults improves by reducing the sample percentage of defaults. Finally, we show that, unlike the logistic regression model, the GEV model is a robust model. A possible future development of this work could be the use of the GEV link function in a Generalized Additive Model.

References

[1] A. Agresti, *Categorical Data Analysis*, Wiley, New York, 2002.

[2] E. Altman, *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, Journal of Finance 23 (4) (1968), pp. 589–609.

[3] E. Altman, R. Haldeman, and P. Narayanan, *ZETA analysis: a new model to identify bankruptcy risk of corporations*, Journal of Banking & Finance 1 (1977), pp. 29–54.

[4] E. Altman, G. Sabato, *Modeling Credit Risk for SMEs: Evidence from the US Market*, ABACUS 19(6) (2006), pp. 716–723.

[5] E. Altman, G. Sabato, N. Wilson, *The value of non-financial information in small and medium-sized enterprise risk management*, Journal of Credit Risk 6(2) (2010), pp. 95–127.

[6] J. Ansell, S. Lin, Y. Ma, G. Andreeva, *Experimenting with Modeling Default of Small and Medium Sized Enterprises (SMEs)* in *Credit Scoring and Credit Control XI Conference*, August 2009.

[7] A. Aziz, D. Emanuel, G. Lawson, *Bankruptcy prediction An investigation of cash flow based models*, Journal of Management Studies 25 (1998), pp. 419-437.

[8] T.G. Bali, *An extreme value approach to estimating volatility and value at risk*, Journal of Business 76 (1) (2003), pp. 83-108.

[9] Basel Committee on Banking Supervision, *International convergence of capital measurement and capital standards: A revised framework*, Bank for International Settlements, Basel, June 2004.

[10] L. Becchetti; J. Sierra, *Bankruptcy risk and productive efficiency in manufacturing firms*, Journal of Banking and Finance, 27 (2002), pp. 2099–2120.

[11] A. N. Berger, *Potential Competitive Effects of Basel II on Banks in SME Credit Markets in the United States* in Finance and Economics Discussion Series paper 2004-12 (2004).

[12] A. Charitou, L. Trigeorgis, *Option-based bankruptcy prediction* 6th Annual Real Options Conference, Paphos, Cyprus, 4-6 July 2002.

[13] Cerved Group, *Caratteristiche delle imprese, governance e probabilitá di insolvenza*, Report. Milan, February 2011.

[14] F. Ciampi, N. Gordini, *Using Economic-Financial Ratios for Small Enterprize Default Prediction Modeling: an Empirical Analysis*, in Oxford Business & Economics Conference, Oxford 2008.

[15] S.G. Coles *An Introduction to Statistical Modelling of Extreme Values*, Springer-Verlag, London 2004.

[16] Credit Suisse Financial Products, *CreditRisk+: A Credit Risk Management Framework*, Credit Suisse First Boston 1997.

[17] J. Danielsson, C.G. de Vries, *Tail index estimation with very high frequency data*, Journal of Empirical Finance 4 (1997), pp. 241–257.

[18] L. De Haan, D.W. Cansen,K. Koedijk, C.G. de Vries, *Safety first portfolio selection, extreme value theory and long run asset risks*, in *Extreme value theory and applications* J. Galambos et al., eds., Kluwer, Dordrecht, 1994, pp. 471–487.

[19] A. C. Davison, D. V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, 1997.

[20] M. Dietsch, J. Petey, *Should SME Exposure be treated as Retail or as Corporate Exposures? A Comparative Analysis of Default Probabilities and Asset Correlation in French and German SMEs*, Journal of Banking and Finance 28 (2004), pp. 773–788.

[21] A. J. Dobson, A.G. Barnett, *Introduction to Generalized Linear Models* 3rd ed., Chapman and Hall/CRC, Boca Raton 2008.

[22] K. Dowd, *Measuring Market Risk*, John Wiley and Sons, Chichester and New York, 2004.

[23] B. Efron, R.J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, New York, 1993.

[24] P. Embrechts, S. Resnick, G. Samorodnitsky, *Extreme value theory as a risk management tool*, North American Actuarial Journal 3 (1999), pp. 30–41.

[25] P. Embrechts, *Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool*, Derivatives Use, Trading & Regulation 6 (2000), pp. 449–456.

[26] M. Falk, J. Haler, R. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, 3rd ed., Springer, Basel, 2010.

[27] S. Figini, D. Fantazzini, *Random Survival Forest models for SME Credit Risk Measurement*, Methodology and computing in applied probability 11 (2009), pp. 29–45.

[28] S. Figini, P. Giudici, *Statistical Merging of Rating Models*, Journal of the Operational Research Society, 62 (6) (2011), pp. 1067–1074.

[29] J.A. Gentry, P. Newbold, D.T. Whitford, *Classifying bankrupt firms with funds*

flow components, Journal of Accounting Research 23 (1) (1985).

[30] P. Giudici, *Applied data mining: statistical methods for business and industry*, Wiley, London, 2003.

[31] W.H. Greene, *Econometric Analysis*, Prentice Hall, New York, 2000.

[32] G. M. Gupton, C.C. Finger, M. Bhatia, *CreditMetrics* Technical document. J. P. Morgan, (1997).

[33] D.J. Hand, W.E. Henley, *Some developments in statistical credit scoring* in *Machine learning and statistics: the interface*, N Nakhaeizadeh, C. Taylor, eds., Wiley, New York, 1997, pp. 221–237.

[34] D.J. Hand, A.M. Niall, *Defining attributes for scorecard construction in credit scoring*, Journal of Applied Statistics 27 (5) (2000), pp. 527–540.

[35] D.J. Hand, H. Mannila, P. Smyth, *Principles of Data Mining*, MIT Press, 2001.

[36] D. J. Hand, *Measuring classifier performance: a coherent alternative to the area under the ROC curve*, Machine Learning 77 (2009), pp. 103–123.

[37] D.J. Hand, *Evaluating diagnostic tests: the area under the ROC curve and the balance of errors*, Statistics in Medicine 29 (2010), pp. 1502–1510.

[38] D.A. Hsieh, C.F. Manski, D. McFadden, *Estimation of Response Probabilities from Augmented Retrospective Observations*, Journal of the American Statistical Association 80 (391) (1985), pp. 651–662.

[39] T. Jacobson, J. Lindè, K. Roszbach, *Credit risk versus capital requirements un-der Basel II: are SME loans and retail credit really different?*, Journal of Financial Services Research 28 (2005), pp. 43–75.

[40] A.F. Jenkinson, *The frequency distribution of the annual maximum (or minimum) values of meteorological elements*, Quarterly Journal of the Royal Meteorological Society 87 (1955), pp. 158–171.

[41] N. M. Kiefer, Default estimation and expert information. Journal of Business and Economic Statistics, 28(2) (2010), pp.320–328.

[42] G. King G., L. Zeng, *Logistic Regression in Rare Events Data*, Political Analysis 9 (2001), pp. 137–163.

[43] S. Kotz, S. Nadarajah, *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London 2000.

[44] C.F. Manski, S.R. Lerman, *The Estimation of Choice Probabilities from Choice-based Samples*, Econometrica 45 (8) (1977).

[45] A.J. McNeil, *Extreme value theory for risk managers. Internal Modelling and CAD II*, RISK Books, 1999, pp. 93–113.

[46] P. McCullagh, J.A. Nelder, *Generalized Linear Model*, Chapman Hall, New York 1989.

[47] R. Merton, *On the pricing of corporate debt: The risk structure of interest rates*, Journal of Finance 29 (1974), pp. 449–470.

[48] H. D. Platt, M.B. Platt, *Development of a class of stable predictive variables: the case of bankruptcy prediction*, Journal of Business Finance & Accounting 17 (1) (1990).

[49] R.L. Prentice, *A case-cohort design for epidemiologic cohort studies and disease prevention trials*, Biometrika 66 (1986), pp. 403–411.

[50] J. Saurina, C. Trucharte, *The Impact of Basel II on Lending to Small-and Medium-Sized Firms: A Regulatory Policy Assessment Based on Spanish Credit Register Data*, Journal of Finance Service Research 26 (2004), pp. 121–144.

[51] R. von Mises, *La Distribution de la Plus Grande de n Valeurs* in Selected Papers of Richard von Mises, Providence, RI: American Mathematical Society 2 (1936), pp. 271–294.

[52] P. Vozzella, G. Gabbi, *Default and Asset Correlation: An Empirical Study for Italian SMEs*, Working Paper.

[53] T.C. Wilson, *Portfolio credit risk*, Economic Policy Review 4 (1998), pp. 71–82.

[54] C. Zavgren, *The prediction of corporate failure: the state of the art*, Journal of Accounting Literature 2 (1983), pp. 1–37.

APPENDIX

In this appendix we obtain the score functions and the Fisher information matrix for $\boldsymbol{\beta}$ and τ of the GEV regression model. The notation used here is defined in Section 3. Firstly, in order to compute the score functions we consider the following equations

$$\frac{\partial l_i(\boldsymbol{\beta},\tau,y_i)}{\partial\beta_j}=\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)}\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\beta_j} \qquad \frac{\partial l_i(\boldsymbol{\beta},\tau,y_i)}{\partial\tau}=\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)}\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\tau} \tag{20}$$

with $j=1,2,...,k$ and $i=1,2,...,n$. From equations (4) and (6) we obtain that

$$\begin{aligned}
\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)} &= \frac{y_i}{\pi(\boldsymbol{x}_i)}-\frac{1-y_i}{1-\pi(\boldsymbol{x}_i)}\\
\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\beta_j} &= -x_{ij}(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i)^{-(\frac{1}{\tau}+1)}\exp\left[-(1+\tau\beta'\boldsymbol{x}_i)^{-\frac{1}{\tau}}\right]\\
\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\tau} &= -(1+\tau\beta'\boldsymbol{x}_i)^{-\frac{1}{\tau}}\left[\frac{1}{\tau^2}ln(1+\tau\beta'\boldsymbol{x}_i)-\frac{\beta'\boldsymbol{x}_i}{\tau(1+\tau\beta'\boldsymbol{x}_i)}\right]\exp\left[-(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i)^{-\frac{1}{\tau}}\right]
\end{aligned}$$

Substituting the former results in the equations (20), the score functions (7) and (8) are obtained.

The second order partial derivatives of the log-likelihood function with respect to parameters $\boldsymbol{\beta},\tau$ are

$$\begin{aligned}
\frac{\partial^2 l_i(\boldsymbol{\beta},\tau,y_i)}{\partial^2\beta_j} &= \frac{\partial^2 l_i(\pi(\boldsymbol{x}_i))}{\partial^2\pi(\boldsymbol{x}_i)}\left[\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\beta_j}\right]^2+\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)}\frac{\partial^2(\pi(\boldsymbol{x}_i))}{\partial^2\beta_j}\\
\frac{\partial^2 l_i(\boldsymbol{\beta},\tau,y_i)}{\partial^2\tau} &= \frac{\partial^2 l_i(\pi(\boldsymbol{x}_i))}{\partial^2\pi(\boldsymbol{x}_i)}\left[\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\tau}\right]^2+\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)}\frac{\partial^2(\pi(\boldsymbol{x}_i))}{\partial^2\tau}\\
\frac{\partial^2 l_i(\boldsymbol{\beta},\tau,y_i)}{\partial\beta_j\beta_k} &= \frac{\partial^2 l_i(\pi(\boldsymbol{x}_i))}{\partial^2\pi(\boldsymbol{x}_i)}\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\beta_j}\frac{\partial\pi(\boldsymbol{x}_i)}{\partial\beta_k}+\frac{\partial l_i(\pi(\boldsymbol{x}_i))}{\partial\pi(\boldsymbol{x}_i)}\frac{\partial^2(\pi(\boldsymbol{x}_i))}{\partial\beta_j\partial\beta_k}\\
\frac{\partial^2 l_i(\beta,\tau,y_i)}{\partial\beta_j\tau} &= \frac{\partial}{\partial\beta_j}\left[\frac{\partial l_i(\boldsymbol{\beta},\tau,y_i)}{\partial\tau}\right]
\end{aligned}$$

where

$$
\begin{aligned}
\frac{\partial^2 l_i(\pi(\boldsymbol{x}_i))}{\partial^2 \pi(\boldsymbol{x}_i)} &= -\frac{y_i}{[\pi(\boldsymbol{x}_i)]^2} - \frac{1-y_i}{[1-\pi(\boldsymbol{x}_i)]^2} \\
\frac{\partial^2(\pi(\boldsymbol{x}_i))}{\partial^2 \tau} &= \pi(\boldsymbol{x}_i) ln[\pi(\boldsymbol{x}_i)] \left\{ \left[\frac{1}{\tau^2} ln(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i) - \frac{\boldsymbol{\beta}'\boldsymbol{x}_i}{\tau(1+\tau\boldsymbol{\beta}'\boldsymbol{x})} \right]^2 [ln[\pi(\boldsymbol{x}_i)]+1] \right. \\
&+ \left. \left[-\frac{2}{\tau^3} ln(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i) + \frac{\boldsymbol{\beta}'\boldsymbol{x}_i + 2\tau(\boldsymbol{\beta}'\boldsymbol{x})^2}{\tau^2(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i)^2} \right] \right\} \\
\frac{\partial^2(\pi(\boldsymbol{x}_i))}{\partial\beta_j \partial\beta_k} &= x_{ij}x_{ik}\pi(\boldsymbol{x}_i)(1+\tau\boldsymbol{\beta}'\boldsymbol{x}_i)^{-1/\tau-2} \left\{1+\tau+ln[\pi(\boldsymbol{x}_i)]\right\}
\end{aligned}
$$

The Fisher information is the negative of the expectation of the second derivatives of the log-likelihood with respect to the parameters $\boldsymbol{\beta}$ and τ

$$
\begin{aligned}
-E\left(\frac{\partial^2 l_i(\beta,\tau,y_i)}{\partial\beta_j\partial\beta_k}\right) &= -\frac{\partial\pi(\mathbf{x}_i)}{\partial\beta_j}\frac{\partial\pi(\mathbf{x}_i)}{\partial\beta_k} \\
-E\left(\frac{\partial^2 l_i(\beta,\tau,y_i)}{\partial^2\pi(\mathbf{x}_i)}\right) &= -\frac{\partial\pi(\mathbf{x}_i)}{\partial\beta_j}\frac{\partial\pi(\mathbf{x}_i)}{\partial\beta_k}\frac{1}{\pi(\mathbf{x}_i)[1-\pi(\mathbf{x}_i)]} \\
-E\left(\frac{\partial^2 l_i(\beta,\tau,y_i)}{\partial\beta_j\partial\tau}\right) &= -x_{ij}\frac{ln^2[\pi(\mathbf{x}_i)]\pi(\mathbf{x}_i)}{(1+\tau\boldsymbol{\beta}'\mathbf{x})[1+\pi(\mathbf{x}_i)]}\left[\frac{1}{\tau^2}ln(1+\tau\boldsymbol{\beta}'\mathbf{x}_i) - \frac{\boldsymbol{\beta}'\mathbf{x}_i}{\tau(1+\tau\boldsymbol{\beta}'\mathbf{x}_i)}\right] \\
-E\left(\frac{\partial^2 l_i(\beta,\tau,y_i)}{\partial^2\tau}\right) &= -\frac{\partial^2\pi(\mathbf{x}_i)}{\partial^2\tau}\frac{1}{\pi(\mathbf{x}_i)[1-\pi(\mathbf{x}_i)]} \qquad (21)
\end{aligned}
$$

since $E\left(\frac{\partial l_i(\pi(\mathbf{x}_i))}{\partial\pi(\mathbf{x}_i)}\right) = 0.$
By substituting the previous results in the equations (21) the Fisher information matrix is obtained.