

# Bayesian Semi-Parametric Logistic Regression Model with Application to Credit Scoring Data

Haitham M. Yousof <sup>1</sup>, Ahmed M. Gad<sup>2</sup>

<sup>1</sup>*Department of Statistics, Mathematics and Insurance, Benha University, Egypt*  
<sup>2</sup>*Statistics Department, Faculty of Economics and Political Science, Cairo University, Egypt*

*Abstract:* In this article a new Bayesian regression model, called the Bayesian semi-parametric logistic regression model, is introduced. This model generalizes the semi-parametric logistic regression model (SLoRM) and improves its estimation process. The paper considers Bayesian and non-Bayesian estimation and inference for the parametric and semi-parametric logistic regression model with application to credit scoring data under the square error loss function. The paper introduces a new algorithm for estimating the SLoRM parameters using Bayesian theorem in more detail. Finally, the parametric logistic regression model (PLoRM), the SLoRM and the Bayesian SLoRM are used and compared using a real data set.

*Key words:* Generalized partial linear model, semi-parametric logistic regression model, parametric logistic regression model, Profile likelihood method, Bayesian estimation, Square error loss function.

## 1. Introduction

Semi-parametric regression models include regression models that combine parametric and nonparametric components. Semi-parametric regression models are often used in situations where the fully nonparametric model may not perform well, or when the functional form of a subset of the regressors or the density of the errors is not known. Semi-parametric regression models are a particular type of semi-parametric models containing a parametric component. So, they rely on parametric assumptions and may be misspecified and inconsistent as in a fully parametric model.

Semi-parametric models combine the flexibility of a nonparametric model with the advantages of a parametric model. A fully nonparametric model will be more robust than semi-parametric and parametric models since it does not suffer from the risk of misspecification. However, nonparametric estimators suffer from low convergence rates, which deteriorate when considering higher order derivatives and multidimensional random variables. In contrast, the parametric models have the risk of misspecification, but if correctly specified they will normally enjoy  $\sqrt{n}$ -consistency with no deterioration caused by derivatives and multivariate data. The basic idea of a semi-parametric model is to take the best of both models.

The semi-parametric generalized linear model, known as the generalized partial linear model (GPLM), is one of the semi-parametric regression models, see Powell (1994), Ruppert et al. (2003), Sperlich et al. (2006), and Yousof and Gad (2015).

Many authors have tried to introduce algorithms to estimate the parameters of the semi-parametric regression models. Majumdar and Eubank (2009) have studied the Bayesian semi-parametric sales projections for the Texas lottery. Meyer et al. (2011) have introduced Bayesian estimation and inference for generalized partial linear models using shape-restricted splines. Zhang et al. (2014) have studied estimation and variable selection in partial linear single index models with error-prone linear covariates. Guo et al. (2015) have studied the empirical likelihood for the single index model with missing covariates at random. Bouaziz et al. (2015) have studied semi-parametric inference for the recurrent events process by means of a single-index model. Yousof and Gad (2015) have introduced Bayesian estimation and inference for generalized partial linear model using some multivariate conjugate prior distributions. Finucane et al. (2015) have introduced a semi-parametric Bayesian density estimation with disparate data sources.

The aim of this paper is to propose a new method for estimating semi-parametric logistic regression model. This is Bayesian type algorithm of estimation. The rest of the paper is organized as follows. In Section 2, the generalized partial linear model (GPLM) is presented. Section 3, introduces the semi-parametric logistic regression model (SLoRM). In Section 4, the Bayesian estimation and inference for the SLoRM is presented. In Section 5, the proposed method is applied to credit scoring data. Finally, some concluding remarks are presented in Section 6.

## 2. The Generalized Partial Linear Model (GPLM)

Consider the generalized linear model (GLM),

$$E[Y|X] = G[X^T\beta], \quad (1)$$

where  $G(\cdot)$  is a known and monotone link function, and  $\beta$  is an unknown finite dimensional parameter. A semi-parametric generalized linear model known as generalized partial linear model is an extension of a generalized linear model defined in Eq. (1). The GPLM has the form

$$E[Y|X, W] = G[X\beta + m(W)] \quad (2)$$

where  $G(\cdot)$  is a known link function. It is a semi-parametric model since it contains both parametric and nonparametric components. For a unit link linear function, the model in Eq. (2) can be reduced to the model  $E[Y|X, W] = X\beta + m(W)$ . This model is called partial linear model (PLM).

In the generalized linear model (GLM) the nonlinear link function  $G(\cdot)$  is fixed and monotone. In the generalized partial linear model (GPLM), a more complex relationship between the response and the regressors is defined by a non-monotone link function.

The generalized partial linear model (GPLM) is used by Burda et al. (1998), Chen (1995), Muller (2001), Peng (2003), Peng and Wang (2004), Severini and Staniswalis (1994), Yousof and Gad (2015).

The estimation methods for the GPLM, in Eq. (2), are based on the idea that an estimator  $\hat{\beta}$  is obtained for a known  $m(\cdot)$ , and an estimator  $\hat{m}(\cdot)$  is obtained for known  $\beta$ . The estimation method that will be considered are based on kernel smoothing methods in the estimation of the nonparametric component of the model. This method is known as *the profile-likelihood method*.

The *profile likelihood method* has been introduced by Severini and Wong (1992). It is based on assuming a parametric model for the conditional distribution of Y given X and W. The idea of this method is as follows:

1. Assume the parametric component of the model, i.e. the parameters vector  $\beta, \beta^*$  say.
2. Estimate the nonparametric component of the model depending on  $\beta^*$ , i.e.  $m_{\beta}(\cdot)$  using a smoothing method to obtain the estimator  $\hat{m}_{\beta}(\cdot)$ .
3. Use the estimator  $\hat{m}_{\beta}(\cdot)$  to construct profile likelihood for the parametric component using either a true likelihood or quasi-likelihood function.
4. The profile likelihood function is then used to obtain an estimator of the parametric component of the model using a maximum likelihood method.

Thus the profile likelihood method aims to separate the estimation process into two parts; the parametric part which is estimated by a parametric method, and the nonparametric part which is estimated by a nonparametric method.

Murphy and Vaart (2000) show that the full likelihood method is not a good choice for semi-parametric models. In semi-parametric models the observed information, if it exists, would be an infinite-dimensional operator. They use profile likelihood rather than a full likelihood to overcome this problem. The algorithm for profile likelihood method is derived as follows.

## 2.1 Derivation of the likelihood functions

For the parametric component of the model, the objective function is the parametric profile likelihood function which is maximized to obtain an estimator of  $\beta$ . This function is given as

$$\ell(\beta) = \sum_{i=1}^n \ell(\mu_{i,\beta}, y_i) = \sum_{i=1}^n \ell_i, \quad (3)$$

where  $\ell(\cdot)$  denotes the log-likelihood or quasi-likelihood function,  $\mu_{i,\beta} = G[X_i^T \beta + m_{\beta}(W_i)]$ , and  $\ell_i = \ell(\mu_{i,\beta}, y_i)$ .

For the nonparametric component of the model, the objective function is a smoothed or a local likelihood function which is given as

$$\ell_H[m_\beta(w)] = \sum_{i=1}^n \ell(\mu_{i,\beta(w)}, y_i) K_H(w - W_i), \quad (4)$$

where  $\mu_{i,\beta(w)} = G[X_i^T \beta + m_\beta(w)] = G(\eta)$ , and the local weight  $K_H(w - W_i)$  is the kernel weight with  $\mathcal{K}$  denoting a multidimensional kernel function and  $H$  is a bandwidth matrix. The function in Eq. (4) is maximized to obtain an estimator for the smooth function  $m_\beta(w)$  at a point  $w$ .

## 2.2 Maximization of the likelihood functions

The maximization of the local likelihood in Eq. (4) requires solving the equations

$$0 = \sum_{i=1}^n \dot{\ell}(G(\eta), y_i) K_H(w - W_i) \quad (5)$$

with respect to  $m_\beta(t)$  and  $\dot{\ell}(\cdot)$  denotes the first derivative of  $\ell(\cdot)$ . The maximization of the profile likelihood in Eq. (3) requires solving the equations

$$0 = \sum_{i=1}^n \dot{\ell}(G(\eta), y_i) \left( U_i + \nabla_{m_\beta}(W_i) \right) \quad (6)$$

with respect to the coefficient vector  $\beta$ . The vector  $\nabla_{m_\beta}(W_i)$  denotes the vector of all partial derivatives of  $m_\beta$  with respect to  $\beta$ . A further differentiation of Eq. (5) with respect to  $\beta$  leads to an explicit expression for  $\nabla_{m_\beta}$  as follows:

$$\nabla_{m_\beta}(w) = - \frac{\sum_{i=1}^n \ddot{\ell}(G(\eta), y_i) K_H(w - W_i) U_i}{\sum_{i=1}^n \dot{\ell}(G(\eta), y_i) K_H(w - W_i)} \quad (7)$$

where  $\ddot{\ell}$  denotes the second derivative of  $\ell(G(\eta), y_i)$

The Eq. (5) and Eq. (6) can only be solved iteratively. Severini and Saitniswalis (1994) presented a Newton-Raphson type algorithm for this maximization as follows.

Let  $m_j = m_\beta(W_j)$ ,  $\eta_i = (X_i^T \beta + m_i)$ ,  $\eta_{ij} = (X_i^T \beta + m_j)$ ,  $\ell_i = \ell(G(\eta_i), y_i)$  and  $\ell_{ij} = \ell(G(\eta_{ij}), y_i)$ . Also, let  $\dot{\ell}_i$ ,  $\ddot{\ell}_i$ ,  $\dot{\ell}_{ij}$  and  $\ddot{\ell}_{ij}$  be the first and second derivatives of  $\ell_i$  and  $\ell_{ij}$  with respect to their first argument, respectively. All the above values are calculated at the observations  $W_i$  instead of the free parameter  $w$ . Then Eq.(5) and Eq. (6) are transformed to

$$0 = \sum_{i=1}^n \dot{\ell}_{ij} \mathcal{K}_H(W_i - W_j), \quad (8)$$

and

$$0 = \sum_{i=1}^n \dot{\ell}_i (X_i - \nabla_{m_i}), \quad (9)$$

respectively. The estimator of  $\nabla_{m_j} = \nabla_{m_\beta}(W_j)$  based on Eq. (7) is necessary to estimate  $\beta$ , where

$$\nabla_{m_j} = - \frac{\sum_{i=1}^n \ddot{\ell}_{ij} \mathcal{K}_H(W_i - W_j) U_i}{\sum_{i=1}^n \ddot{\ell}_{ij} \mathcal{K}_H(W_i - W_j)}.$$

Eq. (8) and Eq. (9) imply the following iterative Newton-Raphson Type algorithm.

**First:** Initial values

Different strategies are presented to obtain starting values of the estimators. These include;

- Using  $\hat{\beta}^{(o)}$  and  $\hat{m}_j^{(o)}$  obtained from fitting a parametric generalized linear model (GLM).
- Using  $\beta = 0$  and  $m_j = G^{-1}[(Y_j + \bar{Y})/2]$  with the adjustment for Binomial responses as  $m_j = G^{-1}[(Y_j + 0.5)/2]$ .
- Using  $\beta = 0$  and  $m_j = G^{-1}[Y_j]$  with the adjustment for binomial responses as  $m_j = G^{-1}[(Y_j + 0.5)/2]$ . (See Severini and Staniswalis, 1994).

**Second:** The updating step for  $\beta$

The  $\beta$  can be updated as

$$\beta^{\text{new}} = \beta - B^{-1} \sum_{i=1}^n \ell_i \tilde{X}_i,$$

where B is a Hessian type matrix defined as  $B = \sum_{i=1}^n \ell_i \tilde{X}_i \tilde{X}_i^T$  and

$$\tilde{X}_j = X_j + \nabla_{m_j} = X_j - \frac{\sum_{i=1}^n \ell_{ij} \mathcal{K}_H(W_i - W_j) X_i}{\sum_{i=1}^n \ell_{ij} \mathcal{K}_H(W_i - W_j)}, j = 1, \dots, n.$$

The updating step for  $\beta$  can be summarized in a closed matrix form as follows:

$\beta^{\text{new}} = (\tilde{X}^T D \tilde{X})^{-1} \tilde{X}^T D \tilde{Z}$ , where  $\tilde{X} = (1 - S^p)U$ ,  $\tilde{Z} = \tilde{X}\beta - D^{-1}v$ ,  $X$  is the design matrix with rows  $X_i^T$ ,  $I$  is an  $(n \times n)$  identity matrix,  $v = (\ell_1, \dots, \ell_n)$ ,  $D = \text{diag}(\ell_1, \dots, \ell_n)$ , and  $S^p$  is a smoother matrix with elements  $(S^p)_{ij} = \frac{\ell_{ij} \mathcal{K}_H(W_i - W_j)}{\sum_{i=1}^n \ell_{ij} \mathcal{K}_H(W_i - W_j)}$ .

**Third:** The updating step for  $m_j = m_\beta(w_j)$

The function  $m_j = m_\beta(w_j)$  is updated by  $\hat{m}_j^{k+1} = \hat{m}_j^k \frac{\sum_{i=1}^n \ell_{ij} \mathcal{K}_H(W_i - W_j)}{\sum_{i=1}^n \ell_{ij} \mathcal{K}_H(W_i - W_j)}$ , where  $k=0, 1, 2, \dots$  is the number of iteration. It is noted that the function  $\ell_{ij}(\cdot)$  can be replaced by its expectation with respect to  $Y$  to obtain a Fisher scoring type algorithm, Severini and Staniswalis (1994).

For the above procedure we can note the following:

1. The variable  $\tilde{Z}$  is a set of adjusted dependent variable.
2. The parameter  $\beta$  is updated by a parametric method with a nonparametrically modified design matrix  $\tilde{X}$ .
3. The function  $\ell_i$  can be replaced by its expectation, with respect to  $y$ , to obtain a Fisher scoring type procedure.
4. The updating step for  $m_j$  is of quite complex structure and can be simplified in some models for identity and exponential link functions  $G$ .

### 3. The semi-parametric logistic regression model (SLoRM)

Consider the updating steps of estimating GPLM, using the profile-likelihood method which have considered in Section (2). The updated values of  $m_j$  and  $\beta$  are:

$$\hat{m}_j^{k+1} = \hat{m}_j^k - \frac{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_h(W_i - W_j)}{\sum_{i=1}^n \cdot \ell_{ij} \mathcal{K}_h(W_i - W_j)} \quad (13)$$

and

$$\beta^{\text{new}} = \beta - B^{-1} \sum_{i=1}^n \cdot \ell_i \tilde{U}_i. \quad (14)$$

The estimators of the parametric component  $\beta$  and the nonparametric component  $m(w)$  will be developed when the response variable  $Y$  belongs to a binomial distribution. This will be shown in the univariate case model using the profile likelihood method. The iterative procedure that will be used is a Newton-Raphson type algorithm.

### 3.1 Updates of the nonparametric component

The Eq. (13) can be written as

$$\hat{m}_j^{(k+1)} = \hat{m}_j^{(k)} - \frac{\sum_{i=1}^n \cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right) \mathcal{K}_h(W_i - W_j)}{\sum_{i=1}^n \cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right) \mathcal{K}_h(W_i - W_j)},$$

where

$$\cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right) = \left[ Y_i - \frac{\exp \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right)}{1 + \exp \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right)} \right]$$

and

$$\cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right) = \left[ Y_i - \frac{\exp \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right)}{\left[ 1 + \exp \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k)} \right) \right]^2} \right].$$

### 3.2 Updates of the parametric component

Eq. (14) can be written as

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} - \frac{1}{B} \sum_{i=1}^n \cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)} \right) \tilde{U}_i^{(k)},$$

where

$$\cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)} \right) = \left[ Y_i - \frac{\exp \left( U_i \hat{\beta} + \hat{m}_i^{(k+1)} \right)}{1 + \exp \left( U_i \hat{\beta} + \hat{m}_i^{(k+1)} \right)} \right],$$

$$\tilde{U}_i^{(k)} = U_i - \frac{\sum_{i=1}^n \cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)} \right) \mathcal{K}_h(W_i - W_j) U_i}{\sum_{i=1}^n \cdot \ell_i \left( U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)} \right) \mathcal{K}_h(W_i - W_j)},$$

$$\cdot \ell_i(U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)}) = \left[ Y_i - \frac{\exp(U_i \hat{\beta} + \hat{m}_i^{(k+1)})}{[1 + \exp(U_i \hat{\beta} + \hat{m}_i^{(k+1)})]^2} \right],$$

$B = \sum_{i=1}^n \cdot \ell_i(U_i \hat{\beta}^{(k)} + \hat{m}_j^{(k+1)}) \tilde{U}_i^{(k)2}$ ,  $j=1, 2, \dots, n$  is the number of observations, and  $k=0, 1, 2$ , is the iterations. The procedure will be iterated until convergence.

For  $k=0$ ,  $\hat{m}_j^{(1)} = \hat{m}_j^{(0)} - \frac{\sum_{i=1}^n \cdot \ell_i(U_i \hat{\beta}^{(0)} + \hat{m}_j^{(0)}) \mathcal{K}_h(W_i - W_j)}{\sum_{i=1}^n \cdot \ell_i(U_i \hat{\beta}^{(0)} + \hat{m}_j^{(0)}) \mathcal{K}_h(W_i - W_j)}$  and

$\hat{\beta}^{(0)} = \hat{\beta}^{(0)} - \frac{1}{B} \sum_{i=1}^n \cdot \ell_i(U_i \hat{\beta}^{(0)} + \hat{m}_j^{(1)}) \tilde{U}_i^{(0)}$ , where  $m_j^{(0)}$  and  $\beta^{(0)}$  are the initial values.

#### 4. Bayesian estimation and inference for the SLoRM

Bayesian inference derives the posterior distribution as a consequence of two antecedents; a prior probability and a likelihood function derived from a probability model for the data to be observed. In Bayesian inference the posterior probability can be obtained according to the Bayes theorem as

$$P(\beta|y) \propto \pi(\beta)L(y|\beta), \quad (15)$$

where  $P(\beta|y)$  is the posterior distribution,  $\pi(\beta)$  is the prior distribution, and  $L(y|\beta)$  is the likelihood function.

The proposed algorithm for estimating the SLoRM parameters as follows:

1. Obtain the probability distribution of response variable  $\mathbf{Y}$ .
2. Obtain the likelihood function of the probability distribution of response variable  $\mathbf{Y}$ .
3. Choose a suitable prior distribution of  $\beta$ .
4. Use Eq. (15) to obtain the posterior distribution.
5. Obtain the Bayesian estimator under the square error loss function.
6. Replace the initial value of  $\beta$  by the Bayesian estimator.
7. Use Profile likelihood method and the Newton-Raphson algorithm with the new initial value of  $\beta$  to estimate the SLoRM.

Consider the SLoRM in Section (3) and suppose that  $y_i \sim \text{Binomial}(m, \beta)$ ,  $\Rightarrow P(Y_i|m, \beta) \propto \beta^y (1 - \beta)^{m-y}$ . We have four different cases, depending on the assumed prior distribution. These cases are described below.

##### Case 1: The general case

Suppose that the conjugate prior distribution of  $\beta$  is as  $\pi(\beta) \sim \text{Beta}(\xi, \tau)$ . Using Eq. (15), the posterior distribution is  $\text{Beta}(\sum_{i=1}^n y_i + \xi, m - \sum_{i=1}^n y_i + \tau)$  with expected

value  $\frac{\sum_{i=1}^n y_i + \xi}{m + \xi + \tau}$  and variance  $\frac{(\sum_{i=1}^n y_i + \xi)(m - \sum_{i=1}^n y_i + \tau)}{[m + \xi + \tau]^2 [m + \xi + \tau + 1]}$ . Then, the Bayesian estimator under the square error loss function is

$$\hat{\beta}_{\text{Bayes}} = \frac{\sum_{i=1}^n y_i + \xi}{m + \xi + \tau}.$$

**Case 2:**

Suppose that the conjugate prior distribution of  $\beta$  is as  $\pi(\beta) \sim \text{Beta}(1, 1)$ . Then the posterior distribution is  $\text{Beta}(\sum_{i=1}^n y_i + 1, m - \sum_{i=1}^n y_i + 1)$  with expected value  $\frac{\sum_{i=1}^n y_i + 1}{m + 2}$  and variance  $\frac{(\sum_{i=1}^n y_i + 1)(m - \sum_{i=1}^n y_i + 1)}{[m + 2]^2 [m + 3]}$ . Then, the Bayesian estimator under the square error loss function is

$$\hat{\beta}_{\text{Bayes}} = \frac{\sum_{i=1}^n y_i + 1}{m + 2}.$$

**Case 3:**

Suppose that we have no information about  $\beta$ , then c Ignorance (Noninformative) prior distribution of  $\beta$  will be as follows

$$\pi(\beta) = 1$$

Then the posterior distribution is  $\text{Beta}(\sum_{i=1}^n y_i + 1, m - \sum_{i=1}^n y_i + 1)$  with expected value

$$\frac{\sum_{i=1}^n y_i + 1}{m + 2}$$

and variance

$$\frac{(\sum_{i=1}^n y_i + 1)(m - \sum_{i=1}^n y_i + 1)}{[m + 2]^2 [m + 3]}.$$

Then, the Bayesian estimator under the square error loss function

$$\hat{\beta}_{\text{Bayes}} = \frac{\sum_{i=1}^n y_i + 1}{m + 2}.$$

**Case 4:**

Using the Jeffrey's Prior for the Binomial distribution, that the conjugate prior distribution of  $\beta$  is as  $\pi(\beta) \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$ . Then the posterior distribution is  $\text{Beta}(\sum_{i=1}^n y_i + \frac{1}{2}, m - \sum_{i=1}^n y_i + \frac{1}{2})$  with expected value  $\frac{\sum_{i=1}^n y_i + \frac{1}{2}}{m + 1}$  and variance  $\frac{(\sum_{i=1}^n y_i + \frac{1}{2})(m - \sum_{i=1}^n y_i + \frac{1}{2})}{[m + 1]^2 [m + 2]}$ . Then, the Bayesian estimator under the square error loss function is  $\hat{\beta}_{\text{Bayes}} = \frac{\sum_{i=1}^n y_i + 0.5}{m + 1}$ .

The proposed algorithm can be generalized for different types of semi-parametric regression models. For instance, Yousof and Gad (2015) used a similar algorithm for estimating the generalized partial linear model using a Bayesian approach. We consider



the Wald statistic to show the significance of the parameters  $\hat{\beta}_i$ . Also, we will use the odds ratio for facilitating the selection of the significant variables.

In case of the SLoRM, the proposed algorithm is appropriate if the profile likelihood method and the Newton-Raphson algorithm are applied. The SLoRM can be used if the response variable has only two values and if the explanatory variables are divided into two groups: the first group, the non-metric (categorical) covariates and the second group, the metric (quantitative) covariates.

## 5. Application: Credit Scoring Data

The credit cards industry has been growing rapidly recently. Thus, consumer credit data are collected by the credit department of banks. The credit scoring manager often evaluates the consumer's credit with intuitive experience. However, with the support of the credit classification model, the manager can accurately evaluate the applicant's credit score. Credit scoring represents a set of common techniques to decide whether a bank should grant a loan to an applicant (borrower) or not. Many authors have tried to use Credit scoring data for example West (2000), Baesens et al. (2003), Huang et al. (2007), Haggag (2008), Brown, I. and Mues, C. (2011), Tu et al. (2014) and Sohn et al. (2016).

Credit scoring data have metric and non-metric covariates; hence the SLoRM algorithm is very convenient to this data type. Also, all the parametric and nonparametric models cannot be used for the credit scoring data for the above reason.

The current data include 400 cases taken from the National Bank of Egypt in the period 2005 – 2010. These data consist of 15 variables; 14 independent variables  $X$ 's (covariates) and one dependent variable ( $Y$ ). Each independent variable represents a question to the applicant. The first three covariates  $X_1$ ,  $X_2$  and  $X_3$  are quantitative variables whereas the remaining variables are categorical variables. The dependent variable  $Y$  is the credibility which takes two values; 1 means "credit-worthy" and 0 means "not credit-worthy". The variables and their values are presented in Table (1).

Table (1): The description of the study covariates

Covariates	Symbol	Description
Age in years	$X_1$	Quantitative variable
Amount of loan	$X_2$	Quantitative variable
Duration in months	$X_3$	Quantitative variable
Savings	$X_4$	Categorical variable: Yes (1) / No (0)
Credit card	$X_5$	Categorical variable: Yes (1) / No (0)
Reputation	$X_6$	Categorical variable: Yes (1) / No (0)
Repayments of earlier credits	$X_7$	Categorical variable: Yes (1) / No (0) / having no previous credits (2)

Experience	X <sub>8</sub>	Categorical variable: Yes (1) / No (0) / having no previous experience (2)
Relationship with partner	X <sub>9</sub>	Categorical variable: Good (1) / bad (0) / having no partner (2)
Guarantees	X <sub>10</sub>	Categorical variable: Person (1) / real estate (2) / goods (3) /nothing (0)
Marital Status	X <sub>11</sub>	Categorical variable: Married (1) / divorced (2) / widow (3) / Single (0)
Educational level	X <sub>12</sub>	Categorical variable: PhD (1) / Master (2) / Bachelor (3) / Medium (4)/ not educated (0)
Project Type	X <sub>13</sub>	Categorical variable: Manufacturing (1) / commercial (2) / professional (3)/ agricultural (4) / others (0)
Purpose of credit	X <sub>14</sub>	Categorical variable: Project construction (1) / Project developing (2) / Buying goods (3) / Buying a land (4) / Others (0)

The PLoRM, the SLoRM and the Bayesian SLoRM are applied to the data. The PLoRM model is as follows

$$\Pr(Y_i = 1|X_i) = G(X_i^T \beta) = \frac{\exp(X_i^T \beta)}{1 + \exp(X_i^T \beta)}$$

The model gives the probability that the response variable takes the value 1. The parameter estimates of PLoRM using maximum likelihood are presented in Table (2). To test hypothesis about the parameters are H<sub>0</sub>: β<sub>i</sub> = 0 vs. H<sub>1</sub>: β<sub>i</sub> ≠ 0. The Wald test statistic is  $Wald = \left\{ \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right\}^2$ .

The null hypothesis is rejected if the P-value (Significance) of  $\hat{\beta}_i < (\alpha = 0.05)$

Table (2): The results of the PLoRM model

	β	SE	P-value	Odds Ratio	Wald	df
X <sub>1</sub>	-0.04	0.0232	0.06	0.96	3.07	1
X <sub>2</sub>	-5.24	2.2017	0.10	0.01	5.66	1
X <sub>3</sub>	-0.44	0.2935	0.10	0.64	2.30	1
X <sub>4</sub>	0.98	0.1292	<b>0.01</b>	2.67	57.92	1
X <sub>5</sub>	3.01	0.2989	<b>0.01</b>	20.36	101.6	1
X <sub>6</sub>	2.83	2.0120	0.34	16.95	1.98	1
X <sub>7</sub>	0.27	0.8767	0.22	1.31	0.09	1
X <sub>8</sub>	0.39	2.3204	0.43	1.48	0.03	1
X <sub>9</sub>	0.76	0.7643	0.52	2.15	1.00	1
X <sub>10</sub>	0.87	0.1731	<b>0.03</b>	2.39	25.34	1
X <sub>11</sub>	3.99	0.9891	0.63	54.36	16.32	1
X <sub>12</sub>	-0.01	0.3665	0.93	0.98	0.00	1
X <sub>13</sub>	0.12	0.3476	0.07	1.13	0.16	1
X <sub>14</sub>	-0.61	0.1301	<b>0.03</b>	0.54	22.21	1
Const.	3.10	1.8712	0.44	22.25	2.75	1

Depending on these results there are four significant independent variables. These variables are saving ( $X_4$ ); whether the applicant has saving, credit card ( $X_5$ ); whether the applicant has a credit card, Guarantees ( $X_{10}$ ); the guarantees to the applicant, and Purpose of credit ( $X_{14}$ ). These variables influence the credit worthiness, according to our data. The remaining variables are insignificant. The three metric variables, age of the borrower, amount, and duration of the loan, are insignificant in all models.

The goodness of fit for the whole model is tested using the loglikelihood ratio test; the p-value is 0.046 which means that the model is significant at 5% significance level.

The second fitted model is the SLoRM which assumes that the probability of a good loan is given by

$$\Pr(Y_i = 1|X_i, W_i) = G(X_i^T \beta + m(W_i)) = \frac{\exp(X_i^T \beta + m(W_i))}{1 + \exp(X_i^T \beta + m(W_i))}$$

where  $G(\cdot)$  is the cumulative logistic function,  $X_i$  is a vector of the variables  $X_4, X_5, \dots, X_{14}$  and  $W_i$  is a vector of the variables  $X_1, X_2, X_3$ . The parameter estimates and their standard errors are displayed in Table (2).

Table (3): The results of the SLoRM Model

	$\hat{\beta}$	S.E	P-value	Odds Ratio	Wald
$X_4$	0.88	0.0994	<b>0.01</b>	2.42	78.72
$X_5$	2.01	0.1084	<b>0.01</b>	7.49	344.25
$X_6$	1.83	1.1129	0.71	6.29	2.72
$X_7$	0.43	0.7061	0.36	1.54	0.37
$X_8$	0.46	2.8779	0.74	1.58	0.02
$X_9$	0.50	0.6639	0.82	1.65	0.57
$X_{10}$	0.96	0.1293	<b>0.02</b>	2.61	55.26
$X_{11}$	2.73	0.7895	0.69	15.37	11.98
$X_{12}$	-0.11	0.2615	0.66	0.89	0.19
$X_{13}$	0.98	0.2439	<b>0.05</b>	2.68	16.32
$X_{14}$	-0.74	0.1418	<b>0.03</b>	0.47	27.63

From the results we can conclude that 5 variables are significant. These variables are  $X_5, X_4, X_{10}, X_{14}$ , and  $X_{13}$ . The three metric variables, age of the borrower, amount, and duration of the loan, are insignificant in all models.

The third fitted model is SLoRM using Bayesian approach. As we presented in Section 4 there are 4 cases. Here we use the second case where we use the conjugate prior distribution and  $Y \sim Binomial(n, \Pi), \pi(\beta) \sim Beta(1,1)$ . The other cases have been tried but we display results of case (2) in Table (4) for the sake of parsimony.

Table (4): The results of the Bayesian SLoRM; Case (2)

Conjugate Prior Distribution $\pi(\beta) \sim Beta(1,1)$	
--	--

	$\hat{\beta}$	S.E	P-value	Odd Ratio	Wald
$X_4$	0.90	0.0942	<b>0.02</b>	2.45	90.57
$X_5$	2.13	0.1059	<b>0.01</b>	8.44	405.69
$X_6$	1.85	1.1292	0.44	6.39	2.70
$X_7$	0.44	0.5231	0.13	1.59	0.70
$X_8$	0.45	2.2129	0.36	1.58	0.04
$X_9$	0.51	0.5734	0.31	1.67	0.80
$X_{10}$	0.96	0.1166	<b>0.02</b>	2.60	67.71
$X_{11}$	2.74	0.6929	0.59	15.54	15.68
$X_{12}$	-0.10	0.1967	0.52	0.90	0.29
$X_{13}$	0.10	0.1847	<b>0.04</b>	2.70	29.17
$X_{14}$	-0.89	0.1314	<b>0.03</b>	0.41	46.28

It is found that the five categorical variables  $X_5, X_4, X_{10}, X_{14}$  and  $X_{13}$  are the only significant variables, which influence the credit worthiness, according to our data. Three metric variables, age of the borrower, amount, and duration of the loan, are insignificant in all models.

The three models are compared and the results are displayed in Table (5). In general, SLoRM estimators are efficient than PLoRM estimators with small standard errors and deviance. The Bayesian SLoRM estimators are efficient than the PLoRM and the SLoRM estimators with small standard errors and deviance. Estimating credit worthiness using Bayesian SLoRM gives estimators with less deviance.

Table (5): The -2loglikelihood of the three models

	PLoRM	SLRM	Bayesian (SLRM)		
			Case(2)	Case(3)	Case(4)
Deviance	140.25	130.12	126.48	126.48	128.87

## 6. Conclusions

According to the PLoRM, it is found that the five categorical variables  $X_5, X_4, X_{10}$  and  $X_{14}$  are the only significant variables, which influence the credit worthiness, according to the current data. According to the SLoRM and the Bayesian SLoRM, it is found that the five categorical variables  $X_5, X_4, X_{10}, X_{14}$  and  $X_{13}$  are the only significant variables, which influence the credit worthiness, according to the current data. The three metric variables; the age of the borrower, the amount of the loan, and the duration of the loan are insignificant in all models.

The SLoRM estimators are efficient than the PLoRM estimators in sense they have smaller standard errors, deviance and smallest number of iterations. The Bayesian SLoRM estimators are efficient comparable to the PLoRM and the SLoRM estimators with smaller standard errors, deviance and mallest number of iterations. Estimating

credit worthiness using Bayesian SLoRM gives estimations with less deviance and with smallest number of iterations.

### Acknowledgements

The authors would like to thank the Editor in Chief, Associated Editors and anonymous referees for their help, very careful reading and valuable comments which greatly improved the paper.

### References

- [1] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. A. K. and Vanthienen, J. (2003). Benchmarking state of the art classification algorithms for credit scoring. *Journal of the Operational Research Society*, **54**, 627–635.
- [2] Bouaziz, O., Geffray, S. and Lopez, O. (2015). Semi-parametric inference for the recurrent events process by means of a single-index model. *Statistics: A Journal of Theoretical and Applied Statistics*, **49**, 361–385.
- [3] Brown, I. and Mues, C. (2011). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, **39**, 3446–3453.
- [4] Burda, M., Hardel, W., Müller, M. and Weratz, A. (1998). Semiparametric analysis of German east-west migration: facts and theory. *Journal of Applied Econometrics*, **13**, 525–542.
- [5] Chen, H. (1995). Asymptotically Efficient Estimation in Semiparametric Generalized Linear Models. *The Annals of Statistics*, **23**, 1102–1129.
- [6] Collet, D. (1991). *Modelling Binary Data*. Chapman and Hall, New York.
- [7] Finucane, M.M., Paciorek, C.J., Stevens, G.A. and Ezzati, M. (2015). Semiparametric Bayesian density estimation with disparate data sources: a meta-analysis of global childhood undernutrition. *Journal of the American Statistical Association*, **110**, 889–901.
- [8] Guo, X., Niu, G., Yang, Y. and Xu, W. (2015). Empirical likelihood for single index model with missing covariates at random. *Statistics: A Journal of Theoretical and Applied Statistics*, **49**, 588–601.

- [9] Haggag.M.M, (2008),"Improving the Efficiency of Semi Parametric Regression Estimation" Unpublished Dissertation submitted to ISSR, Cairo University.
- [10] Hagan, C.L. Chen, M. C. and Wang C.J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, **33**, 847–856.
- [11] Majumdar, A. and Eubank, R. (2009). Bayesian Semiparametric Sales Projections for the Texas Lottery. *Journal of Data Science*, **7**, 73-87 .
- [12] Meyer, M.C, Hackstadt, A. J. and Hoeting, J. A. (2011). Bayesian estimation and inference for generalized partial linear models using shape-restricted splines. *Journal of Nonparametric Statistics*, **23**, 867-884.
- [13] Müller, M. and Ronz, B. (1990). Credit scoring using semi-parametric methods, in measuring risk in complex stochastic systems, by Frank. J., hardel, W. and Stahel G. Springer Verlage.
- [14] Müller, M, (2001). Estimation and Testing in Generalized Partial Linear Models-A comparative Study. *Statistics and Computing*, **11**, 299-309.
- [15] Murphy, S.A. and Van Der Vaart, A.W. (2000). On Profile Likelihood. *Journal of the American Statistical Association*, **95**, 449-485.
- [16] Peng, H. and Wang, X. (2004). Moment Estimation in Semi-parametric Generalized Linear Models. *Journal of Nonparametric Statistics*, **16**, 1-22.
- [17] Powel, L. (1994). Estimation of Semi-parametric Regression Models. *Handbook of Econometrics*, **4**, Ch. 41.
- [18] Ruppert, D., Wang, M.P. and Carroll, R.J. (2003). *Semi-Parametric Regression*. Cambridge University Press, UK.
- [19] Severini, T.A. and Staniswalis, J.G. (1994). Quasi-likelihood Estimation in Semi-parametric Models. *Journal of the American Statistical Association: Theory and Methods*, **89**, **426**, 501-511.
- [20] Severini, T.A. and Wong, W.H. (1992). Profile Likelihood and Conditionally Parametric Models. *The Annals of Statistics*, **20**, **4**, 1768-1802.
- [21] Sohn, S. Y., Kim, D. H. and Yoon, J. H. (2016). Technology credit Scoring model with fuzzy logistic regression. *Applied Soft Computing*, **43**, 150-158.

- 
- [22] Sperlich, S., Hardle, W. and Aydinli, G. (2006). *The Art of Semi-parametrics*. Physical-Verlag, Germany.
- [23] Tu, X., Yang, Q., Song P. and Yang, X. (2014). A credit scoring model based on logistic regression and RONSA. *Journal of Sichuan University*, **46**, 19-24.
- [24] West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, **27**, 1131-1152.
- [25] XploRe (2000), "XploRe-the interactive statistical computing environment" WWW: <http://www.xploRe-stat.de>.
- [26] Yousof, H.M. (2015). Semi-parametric regression using Bayesian technique with application, Unpublished Ph.D. Thesis, Department of Statistics, Mathematics and Insurance, Benha University, Egypt.
- [27] Yousof, H.M. and Gad, A.M. (2015). Bayesian estimation and inference for the generalized partial linear model. *International Journal of Probability and Statistics*, **4**, 51–64.
- [28] Zhang, J., Wang,X., Yu,Y. and Gai ,Y. (2014). Estimation and variable selection in partial linear single index models with error-prone linear covariates. *Journal of Theoretical and Applied Statistics*, **48**, 1048–1070.

Haitham M. Yousof  
Department of Statistics, Mathematics and Insurance  
Benha University  
Benha, Egypt.  
[haitham.yousof@fcom.bu.edu.eg](mailto:haitham.yousof@fcom.bu.edu.eg)

Ahmed M. Gad  
Statistics Department,  
Faculty of Economics and Political Science,  
Cairo University  
Cairo, Egypt.  
[Ahmed.gad@feps.edu.eg](mailto:Ahmed.gad@feps.edu.eg)

