Shen, Xiangjin; Karibzhanov, Iskander; Tsurumi, Hiroki; Li, Shiliang

Working Paper

Comparison of Bayesian and sample theory parametric and semiparametric binary response models

Bank of Canada Staff Working Paper, No. 2022-31

Provided in Cooperation with:
Bank of Canada, Ottawa

Suggested Citation: Shen, Xiangjin; Karibzhanov, Iskander; Tsurumi, Hiroki; Li, Shiliang (2022) : Comparison of Bayesian and sample theory parametric and semiparametric binary response models, Bank of Canada Staff Working Paper, No. 2022-31, Bank of Canada, Ottawa, https://doi.org/10.34989/swp-2022-31

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2022-31

Last updated: July 4, 2022

Comparison of Bayesian and Sample Theory Parametric and Semiparametric Binary Response Models

by Xiangjin Shen,[1] Iskander Karibzhanov,[2] Hiroki Tsurumi[3] and Shiliang Li[4]

[1] Financial Stability Department
Bank of Canada
xshen@bankofcanada.ca

[2] International Economic Analysis Department
Bank of Canada
ikaribzhanov@bankofcanada.ca

[3] Department of Economics
Rutgers University
tsurumi@rci.rutgers.edu

[4] The Depository Trust & Clearing Corporation
shiliangli1@gmail.com

DOI: https://doi.org/10.34989/swp-2022-31 | **ISSN** 1701-9397

Acknowledgements

We thank participants at the 10th International Conference on Computational and Financial Econometrics and the 2018 Economic Research in High Performance Computing Environments Workshop at the Federal Reserve Bank of Kansas City for their comments. Colleagues in the Financial Stability Department at the Bank of Canada also provided helpful comments and suggestions. We are grateful to Zhentong Lu and Ken Chow for comments and to Yang Xu and Gias Uddin for data knowledge support. We also thank Professor Roger Klein at Rutgers University for a wonderful lecture. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

Abstract

This study proposes a Bayesian semiparametric binary response model using Markov chain Monte Carlo algorithms since this Bayesian algorithm works when the maximum likelihood estimation fails. Implementing graphic processing unit computing improves the computation time because of its efficiency in estimating the optimal bandwidth of the kernel density. The study employs simulated data and Monte Carlo experiments to compare the performances of the parametric and semiparametric models. We use mean squared errors, receiver operating characteristic curves and marginal effects as model assessment criteria. Finally, we present an application to evaluate the consumer bankruptcy rates based on Canadian TransUnion data.

Topics: Econometric and statistical methods; Credit risk management
JEL codes: C1, C14, C35, C51, C63, D1

1 Introduction

Since Nelder and Wedderburn published their seminal paper in 1972 on the generalized linear model (GLM), applications of the GLM have increased appreciably, especially in the past few years. From regime switching to machine learning and from microeconomic survey data to large panel data analysis, binary response models have been applied broadly as the foundation for purposes such as network mapping analysis, housing and labour market analysis and financial risk analysis.

Amemiya (1981) and Aldrich and Nelson (1984) present comprehensive discussions of the GLM binary response models. Since the distributions of binary responses are in fact unknown and are often not estimable, researchers have applied logit or probit models under specific distribution assumptions. But a persistent question is whether semiparametric models are better than the traditional logit or probit parametric models.[1]

In this paper, we propose a Bayesian semiparametric binary response model using the quasi-likelihood function as the likelihood part of the posterior distribution. We compare the performances of the Bayesian semiparametric model with the sample theory semiparametric model. We also compare the semiparametric models with probit and logit parametric models. The comparisons are based on simulated data and Monte Carlo experiments. As the criteria of comparison, we use the marginal effect, mean squared error (MSE) and receiver operating characteristic (ROC) curve. We find: (i) when the data are balanced, the performances of the semiparametric models are indistinguishable from the performances of the parametric models; (ii) however, when the data are extremely unbalanced (for example, the "Y = yes" response rate is less than 3%), the maximum likelihood estimation of the semiparametric and parametric models may not converge, whereas the Bayesian estimation converges.

We also introduce a computationally optimum bandwidth, then compare the Bayesian estimates using the regular bandwidth and compare the other estimates using the computationally optimum bandwidth. To do so, we apply GPU (graphics processing unit) computing with C/C^{++} as well as

[1] Few papers provide an answer to this persistent question. A paper published in the *Journal of Applied Econometrics* states "the results of this paper indicate that more more work is necessary." See Gerfin (1996).

MATLAB. The computed speed of the GPU computing is significantly (by hundreds or thousands[2] of times) faster than the regular computing process. Most importantly, GPU computing is an efficient tool in the Markov chain process, which is difficult to realize using the regular parallel computing.

After the simulated data and Monte Carlo experiments, we illustrate an example by estimating the consumer bankruptcy rates based on the TransUnion[3] data, and we test the robustness of the Bayesian semiparametric and other binary response models.

The organization of the paper is as follows. In Section 2, we present the Bayesian binary response model and estimation method as well as GPU computing. In Section 3, we compare different estimators using simulated data. In Section 4, we present Monte Carlo experiments. An empirical application is presented in Section 5. And Section 6 provides concluding remarks.

2 Bayesian Semiparametric Binary Response Model

2.1 Model and estimation algorithm

We have a sample of binary responses, $y_1, \ldots, y_n$, where the binary response model is to use the following latent variable regression:

$$y_i = \begin{cases} 1 & \text{if yes with probability } p_i \text{ or if } M(y_i) > \varepsilon_i \\ 0 & \text{otherwise with probability } 1 - p_i \text{ or if } M(y_i) \leq \varepsilon_i \end{cases},$$

and p_i is given by $p_i = \mathrm{F}(x_i, \beta)$, where $x_i = (x_{i1}, \cdots, x_{ik})$ and $\beta = (\beta_0, \beta_1, \cdots, \beta_k)'$.
Then, we will have a general GLM form of the binary response model:

$$M(y_i) = F(x_i, \beta).$$

[2]The speed depends on the GPU hardware setup.

[3]To protect the privacy of Canadians, no personal information was provided by TransUnion. The TransUnion dataset was "anonymized," meaning that it does not include information that identifies individual Canadians, such as names, social insurance numbers or addresses. In addition, the dataset has a panel structure, which uses fictitious account and consumer numbers assigned by TransUnion.

If we know the cumulative density $F(\cdot)$ for ε_i in latent variable regression above, we may choose a parametric estimation procedure. The most frequently used distributions are

	Cumulative Density	Probability Density
Logistic	$F(z_i) = \frac{1}{1+e^{-z_i}}$	$f(z_i) = \frac{e^{-z_i}}{(1+e^{-z_i})^2}$
Probit	$\Phi(z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{-\frac{t}{2}} dt$	$f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i}{2}}$
Linear Probability	$F(z_i) = \int_0^{z_i} dt = z_i = x_i\beta$	$U(0, 1)$

If $F(\cdot)$ is not known, we can use the quasi-likelihood function:

$$\ell(\beta \mid y_i, x_i = 1, \ldots, n) = \prod_{i=1}^{n} \hat{p}_i^{y_i} \times \prod_{i=1}^{n} (1 - \hat{p}_i)^{1-y_i}. \tag{1}$$

As one of the most efficient semiparametric methods for binary response models, the single index-parametric model proposed by Klein and Spady (1993) and Klein and Vella (2009), among others, has been broadly cited.[4] Following Klein and Spady (1993), we obtain $\hat{p}_i$ by

$$\begin{aligned} \hat{p}_i &= Pr[Y_i = 1 \mid V_i(\beta)] \\ &= \frac{p(Y=1)\hat{g}(V \mid Y=1)}{\hat{g}(V)} = \frac{\frac{n_1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K(\frac{t-v_i}{h_n})(\frac{y_i}{n_1})}{\sum_{i=1}^{n} \frac{1}{h_n} \frac{K(\frac{t-v_i}{h_n})}{n}}. \end{aligned} \tag{2}$$

$$V_i(\beta) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}. \tag{3}$$

$$\hat{g}(t) = \sum_{i=1}^{n} \frac{1}{h_n} \frac{K(\frac{t-v_i}{h_n})}{n} \tag{4}$$

is a non-parametric kernel density estimation function, where

$$K(\frac{t-v_i}{h_n})$$

is the kernel function satisfying $\int K(x)dx = 1$, $K(x) \geq 0$, n_1 is the sum of $'y_i = 1'$, and h_n is the kernel density window size or bandwidth.

[4] As of July 2021, the Jstor.org citation record for Klein and Spady (1993) is 991 times.

In equation (3), $V_i(\beta)$ is the single index. Given the linearity of V_i in equation (3), we may write:

$$X_i\beta = \beta_1(X_{i1} + \theta_2 X_{i2} + \cdots + \theta_k X_{ik}) + \beta_0, \tag{5}$$

where $\theta_i = \beta_i/\beta_1$, $\beta_1 \neq 0$. From equation (5) we get the new single index as

$$V_i(\theta) = X_{i1} + \theta_2 X_{i2} + \cdots + \theta_k X_{ik}. \tag{6}$$

Because the probability of the linear transformation of the index is the same as the probability of the original index, equation (2) will have the following property:

$$Pr(Y = 1 \mid V = v(\beta)) = Pr(Y = 1 \mid V = v(\theta)). \tag{7}$$

Rather than maximizing the quasi-likelihood function (1), we propose a Bayesian semiparametric estimation algorithm by using the quasi-likelihood function (1) as the likelihood to obtain the posterior distribution of $\theta = (\theta_2, \ldots, \theta_k)$:

$$p(\theta) \propto \pi(\theta)\ell(\theta \mid \text{data}), \tag{8}$$

where $\pi(\theta)$ is the prior and $\ell(\cdot)$ is the quasi-likelihood. We use MCMC algorithms with the Metropolis-Hastings criterion.

The MCMC algorithms are carried out as follows: let $\theta^{(i)}$ be the i-th draw of θ.

Step 1 Choose an initial value $\theta^{(0)}$. We use the ordinary least square (OLS) estimates of the standardized transformed model of equation (6):[5]

$$y_i = x_{i1} + \theta_2 x_{i2} + \cdots + \theta_k x_{ik}.$$

Step 2 We use a random walk draw:

$$\theta^{(i)} = \theta^{(i-1)} + \varepsilon_i,$$

where ε_i is normal with mean 0 and variance $c(X'X)^{-1}$. We set $c = 1$.

[5] In the maximum likelihood estimation of the semiparametric model, the covariate x_{ij} is standardized as x_{ij}/s_j, where s_j is the standard deviation of x_{ij}'s. This standardization of the covariates is done to make the convergence of the MLE procedure easier and to get rid of the large variances among different types of variables.

Step 3 Set $\theta^{(i)} = \theta^{(i)}$ if $u < \alpha$. Otherwise set $\theta^{(i)} = \theta^{(i-1)}$, where u is drawn from Uniform(0, 1) and α is given by:

$$\alpha = \min\left\{1, \frac{p(\theta^{(i)} \mid \text{data})}{p(\theta^{(i-1)} \mid \text{data})}\right\}.$$

$p(\cdot \mid \text{data})$ is the posterior pdf of θ.

Step 4 Repeat **Step 2** and **Step 3** for $i = 1, 2, \ldots, M$.

In estimating the semiparametric model, we use the kernel density of equation (4) for both the MLE and Bayesian estimation. In the case of the Bayesian Markov chain Monte Carlo (MCMC) algorithm, the kernel density is estimated for each draw of $\theta^{(i)}$. In the case of the MLE, the kernel density is estimated for each iteration until convergence is attained.

The kernel density depends on the choice of the kernel, $K(\cdot)$ and the bandwidth, h. Li (2001) shows that the choice of the bandwidth is more important than the choice of the kernel. Keeping the normal distribution as the kernel, we use two bandwidths to see if the choice of the bandwidth makes the difference in the MLE as well as in the Bayesian estimation. The first bandwidth we use is Silverman's (1986) estimation:

$$h = \left(\frac{4}{3n}\right)^{.2} \sigma. \tag{9}$$

We call this bandwidth the regular bandwidth. The second bandwidth we use is the optimal bandwidth given by

$$h^*_{optimal} = \left(\frac{R(K)}{\left(\int x^2 K(x)dx\right)^2 R(\hat{g}''(x; p(h)))}\right)^{.2}. \tag{10}$$

The optimal bandwidth $h^*_{optimal}$ is explained in the appendix. The optimal bandwidth tends to trace sharp modes of a density better than the regular bandwidth does. This is illustrated in Figures 1 and 2, where 15 Gaussian mixture densities are presented.

Figures 1 and 2 Here.

In Figures 1 and 2, the solid black lines are the true Gaussian mixture densities, whereas the lines in red are kernel densities. In Figure 1 the kernel densities are obtained using the regular

bandwidth, while in Figure 2 they are obtained using the optimal bandwidth. We see that the regular bandwidth in Figure 1 misses the sharp modes of the true densities, but the optimal bandwidth in Figure 1 traces the sharp modes fairly accurately as illustrated vividly by the multimodal claw distribution in the center of Figures 1 and 2.

2.2 Benefit of using GPU in MATLAB to improve computing efficiency

The computation of the optimal bandwidth is time-consuming, and thus we use a graphic processing unit (GPU). GPU computation has been used more and more in Bayesian estimation of many models. Equation (9) is the most popular simple bandwidth and it is only optimal for Gaussian data. If data are not Gaussian, we could use equation (10), which must need multiple complex computations.[6] This is one of the reasons that researchers, such as Klein and Shen (2010), use different estimation methods to estimate bandwidth within the single index model to build different smooth factors with specific bounds to minimize the bias in estimating bandwidth. The computation of equation (10) can be implemented efficiently by using GPU computing with C/C++ in MATLAB (Li (2011)), and its speed is at least 600 times faster[7] than the regular computing method, such as in Gauss or MATLAB itself. Since Bayesian computing needs the conditional Markov chain process with complicated matrix order setup, it is impossible to apply the parallel computing simultaneously. Therefore, when we try to estimate much more accurate optimal bandwidth, GPU computing is more efficient and applicable because we will consider all real numbers without any arbitrary lower or upper bound. Most importantly, we can realize Monte Carlo simulations effectively.

A GPU is a specialized processor dedicated to optimizing graphical computations, i.e., rendering 2D/3D scenes. Nvidia[8] introduced the term GPU in 1999 with their GeForce line of products. CUDA, short for "Compute Unified Device Architecture," is a proprietary platform that Nvidia

[6] To compute equation (10), studies such as Chen (2015) and Guidoum (2015) have tried to resolve the integration we propose by using multiple computing options, and the computing is time consuming if the data are large.

[7] This is based on a USD 200 GPU installed on a regular desktop. A high efficiency GPU will generate results thousands of times faster than the regular computing.

[8] Nvidia Corporation is an American multinational technology company that designs GPUs to accelerate computing and solve important challenges beyond the reach of normal computers.

has implemented on their GPUs. The original version was introduced in 2007. CUDA has now matured considerably and is supported on every major Nvidia GPU, including high-end products like the Tesla cards on which we run our codes.

GPU support in MATLAB is available inside the Parallel Computing Toolbox and is built on top of the CUDA driver. On the GPU, a primitive operation, called a "kernel," is executed in parallel by (possibly thousands) of CUDA threads. Kernels can be launched as 1D, 2D or 3D blocks, which can themselves be organized into 1D or 2D grids. Kernels are defined as special C/C++ functions.

After profiling our MATLAB code, we found that most of its runtime is spent calculating the optimal bandwidth in equation (10) and the probability estimates in equations (2) and (4). The algorithm for solving the fixed point in equation (10) is explained in section of 4.1 of Li (2011) and described in the appendix. Its most computationally intensive steps involve estimating higher order Gaussian density derivatives for the optimized pilot bandwidths, which we implement using 4^{th}- and 6^{th}-degree Hermite polynomials.

Fortunately, computing higher order Gaussian density derivatives is an "embarrassingly parallel" problem in the sense of Moler (2013). This means that the problem can be separated into a number of parallel tasks involving basic math operations, with no need for communication between these tasks. We have therefore efficiently hand-written three CUDA/C++ kernel functions that carry out the computations alluded to above, *DensityDerivative4.cu* and *DensityDerivative6.cu* for equation (10) and *ProbabilityEstimates.cu* for equation (2). These can be compiled using the Nvidia NVCC compiler. The resulting assembly-level PTX files can then be loaded into MATLAB as GPU kernel objects using *parallel.gpu.CUDAKernel* as an interface. At this point, the kernels can easily be run via MATLAB's *feval* function. The process of transferring the input data from host memory to GPU memory is done automatically by MATLAB via *feval*, but the results need to be loaded back into the main workspace using the *gather* function.

The three CUDA kernel functions mentioned above use a one-dimensional grid of one-dimensional blocks. In our case, one GPU thread is required for each of the N=218,213 data points in our TransUnion sample within section 5. Since the maximum number of threads possible to run simultaneously in one block is 1,024, we divide the grid into N/1,024=213 blocks. Using this approach, we were able to speed up the computation of density derivatives described above by a factor of

about 160 on a Tesla K80 card and a factor of about 600 on a Tesla P100 card.

3 Comparing the Performances of Parametric and Semiparametric Binary Response Models

Let us compare the performances of the parametric and semiparametric binary response models using the Bayesian and maximum likelihood estimators (MLE). We choose the probit and logit models as the parametric models. For the semiparametric model, we use two bandwidths: the regular bandwidth of equation (9) and the optimal bandwith of equation (10). In summary, the estimators and models we compare are:

$$\text{Bayesian} \begin{cases} \text{Probit} \\ \text{Logit} \\ \text{Semiparametric with the regular bandwidth} \\ \text{Semiparametric with the optimal bandwidth} \end{cases}$$

$$\text{MLE} \begin{cases} \text{Probit} \\ \text{Logit} \\ \text{Semiparametric with the regular bandwidth} \\ \text{Semiparametric with the optimal bandwidth} \end{cases}$$

As given in equation (5), in the semiparametric model the regression coefficients, β_i's, are transformed into θ_i's. This makes it difficult to compare the regression coefficient estimates from a parametric model with those from the semiparametric model. Therefore, let us use three model selection criteria: the marginal effects, MSEs and the ROC curve.

The marginal effect is a popular statistic for the binary response model. When the distribution is known or the model is parametric, the generalized form of the true marginal effect of X_k for models with known density distribution is:

$$\frac{\partial F(x_i\beta)}{\partial x_k} = \frac{\partial F(x_i\beta)}{\partial x_i\beta}\frac{\partial x_i\beta}{\partial x_k} = F'(x_i\beta)\beta_k = f(x_i\beta)\beta_k.$$

Within the semiparametric model, the marginal effect needs to be defined differently: we use the predicted probability, $\widehat{p}$, and define the estimated marginal effect as $\hat{p}(x+\Delta x) - \hat{p}(x)$, in which

$\hat{p}(.)$ is given in equation (2) and Δx is an increment of the x. In order to capture the entire distribution of the X, we will consider $\Delta x = \{std(x), 2 \times std(x), 3 \times std(x)\}$.[9]

In regressions, one way to select a model is to choose the model with the smallest unweighted MSE or the normal MSE, [10] which is calculated by

$$\sum_{i=1}^{n} \frac{(y_i - \hat{P}_i)^2}{n-k},$$

where $y_i = 1$ or 0. $\hat{P}_i$ is the computed probability $F(x_i\hat{\beta})$ for the case of a parametric model or equation (2) for the case of the semiparametric model.

The ROC curve is one of the best choices (McNeil and Hanley (1982, 1984), Swets et al. (2000), Fawcett (2006), etc.) to select the binary response model. We compare the area under the ROC curve (Alonzo (2002), Agresti (2007)). The bigger the area, the better the predictive power of the binary response model. We will use the algorithm from Fawcett (2006) to plot the ROC curve and calculate the area under the ROC curve.

Let us compare the performances of the different estimation methods and models using simulated data. We specify the binary choice model to be

$$Y_i^* = \beta_0 + \beta_1 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \tag{11}$$

where X_{i3} is a zero-one dummy variable to represent a discrete covariate and X_{i2}, is drawn from a uniform distribution, $U(0,a)$. The continuous regressor, X_{i2}, is included since the large sample properties of the semiparametric estimator require that at least one regressor is a continuous variable. The values of the parameters $(\beta_0, \beta_1, \beta_2, a)$ are chosen to control the percentage of $Y_i = 1$ to represent balanced or unbalanced data. The observed binary values, Y_i, are set as

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}.$$

The sample size n is set at 2,048 ($n = 2,048$).

[9] We may also use the quantile of X as Δx.

[10] Using the weighted mean squared errors is also an option. Amemiya (1981) argues for the use of the weighted MSE, but as shown in Chen and Tsurumi (2010), the unweighted MSE is a better model selection criterion than the weighted MSE.

Before we compare the performances of the sample theory and Bayesian estimates of the binary response models, let us see how well the discretized marginal effect

$$p(x+\Delta x) - p(x)$$

is estimated by the sample theory and Bayesian semiparametric models. We generated a sample of size 1,000 ($n = 1,000$) from the logistic distribution setting $\beta = \{\beta_0, \beta_1, \beta_2\}$. The percentage of $Y = 1$ is 9.25%. Δx is set at one, two, and three standard deviations.

ΔX	true marginal effect	semiparametric model	
		MLE	Bayesian
one std(x_2)	.0109	.0159	.0149
two std(x_2)	.0233	.0273	.0266
three std(x_2)	.0370	.0302	.0307

Notes: std=standard deviation; Bayesian=posterior mean. The regular bandwidth is used.

From the table above we see that the discretized marginal effects are reasonably well estimated by the sample theory and Bayesian models.

The error term in equation (11), ε_i , can also be drawn from 16 distributions that are given in Table 1. Distributions #1–15 are Gaussian mixture densities from Marron and Wand (1992) and distribution #16 is a skew logistic distribution with the distribution function given by:

$$\Pr(y_i = 1) = \frac{1}{(1+e^{-x_i\beta})^\theta}.$$

The first 15 distributions are presented in Figures 1 and 2. Some of these distributions, especially trimodal, claw, and comb distributions, may seldom occur in real data, but these distributions are different from the probit or logit distributions, and thus the semiparametric models may perform better than the parametric models.

Table 2 and Table 3 present the ROC areas and MSEs of different estimators based on simulated data. Although we have obtained results for all of the 16 distributions, the results are quite similar to those given in Tables 2 and 3. Table 2 shows the balanced cases in which the percentage of $Y = 1$ ranges from 28.8% to 36.5%, while Table 3 shows extremely unbalanced cases in which the percentage of $Y = 1$ ranges from 0.9% to 3.47%.

Table 1: Distributions of the error terms of the binary response models

	Distribution		Distribution
1	Gaussian	9	Trimodal
2	Skewed unimodal	10	Claw
3	Strongly skewed	11	Double claw
4	Kurtotic unimodal	12	Aymmetric claw
5	Outlier	13	Asymmetric double claw
6	Bimodal	14	Smooth comb
7	Separated bimodal	15	Discrete comb
8	Skewed bimodal	16	Skew logistic

Tables 2 and 3 Here.

From Tables 2 and 3 we conclude that judged by the ROC area and MSE, we cannot discriminate among the different models and estimation procedures except in the cases of extremely unbalanced data as given in Table 3: all the MLE estimation procedures failed to converge, whereas all the Bayesian MCMC algorithms attain convergence. Hence we conclude that when the data are extremely unbalanced, the Bayesian MCMC algorithms may be preferred to the MLE algorithms. Comparing the bandwidths, we see that the use of the optimal bandwidth does not have a visible advantage over the standard bandwidth.

4 Monte Carlo Experiments

In the previous section based on one sample draw, we compared the performances of the different models and estimation procedures. In this section, we conduct Monte Carlo (MC) experiments to compare the performances of the different models and estimation procedures.

In the literature, MC results using the optimal bandwidth are few because of the heavy computational burden in searching for the optimal bandwidth. The most difficult part of MC simulation is to estimate the optimal bandwidth efficiently without smoothing techniques and specific bounds,

and to run Bayesian MCMC simulations quickly. By using GPU computing with C/C++ in MATLAB (Li(2011)), which is more than 600 times faster than the regular computing method, we are able to effectively run the MC simulations.

The number of Monte Carlo replications is 500. The Monte Carlo simulation results are consistent with results obtained in the previous section for the simulated data. Therefore, we only present two examples for balanced cases and unbalanced cases.

In Table 4, the mean and standard deviation of the 500 iterations are displayed. The first part is for the balanced case with claw distribution and the second part is for the unbalanced case of the model with skewed log distribution; only Bayesian MC results can be presented for the unbalanced case because not all 500 replications yield convergence when the MLE is used. Clearly, the results from ROC and MSE are very close among different models: either parametric or semiparametric models by either MLE or Bayesian methods. However, the marginal effects from both MLE or Bayesian semiparametric methods are smaller than those from parametric methods. In addition, the optimal bandwidth estimates from both Bayesian and MLE semiparametric methods are very close, although the standard deviation of bandwidth derived from the MLE method is larger.

5 Analysis of Consumer Bankruptcy Rates using TransUnion Data

Bankruptcy rate has become one of the most important risk assessment factors in predicting financial stress. Many popular studies using binary models to predict various types of bankruptcy rates have circulated since Ohlson (1980) proposed a logistic model for predicting bankruptcy. Among these studies, Gross (2002) initiated an empirical analysis of personal bankruptcy by using probit and logistic models, and Alaminos et al. (2016) chose a logistic type model to predict bankruptcy of business globally after a thorough literature review of model options for bankruptcy. All of these indicate that a binary response model is one of the best options for predicting bankruptcy rates. In this section, we present an application to evaluate consumer bankruptcy rates using Canadian TransUnion (TU) data based on a binary response model.

5.1 TransUnion consumer data

TransUnion consumer data provide loan information about Canadian consumers, and they provide first-hand bankruptcy information for individual consumers. Each consumer may have different loan accounts, such as mortgages, credit cards, lines of credit. Our analysis will be based on consumer instead of independent loan accounts.

Several interesting topics could be answered through a binary response model using TU data. Is the older or younger generation more prone to bankruptcy? Does credit risk score really matter for credit quality? Are people with higher current credit balances more likely to have a loan problem? Which province in Canada has consumers who are relatively more vulnerable?

We focus on the major adult consumers between 18 and 82 years of age.[11] After data cleaning, we randomly choose 218,112 observations (around 1% of the total data[12]) from the February 2020[13] data. We set:

$$Y = \begin{cases} 1 & \text{if the consumer went bankrupt once in history} \\ 0 & \text{otherwise.} \end{cases}$$

We refer to the proportion of (Y=1) as the rate, which represents the total bankruptcy rate in TU data history. This bankruptcy rate (P(Y=1)) is 3.09%,[14] and we may say the data are unbalanced. Five covariates or regressors are continuous variables: they are age in months, TU risk scores, total balance of mortgages, total balance of bankcards, and total balance for lines of credit. Three covariates characterizing the location of consumers are categorical variables: Ontario (yes or no, 1 or 0), British Columbia (yes or no, 1 or 0), and Quebec (yes or no, 1 or 0).

Summary statistics of all variables are presented in Table 5. The mean and median age of consumers are around 48 years, with a standard deviation of 16 years. The average and median

[11] According to the census, the average lifespan of Canadians is around 82 years old, so we choose this as the upper bound of age.

[12] Our data are selected based on a cleaned TU sample after generic data validation from the Bank of Canada Data Statistics Office. We also filtered out data with missing values and made sure the total loan balances $>=$ sum of balances from all loan accounts.

[13] TU data are reported by month; our sample represents the February 2020 data report.

[14] The aggregate bankruptcy rate is 3.09%. The bankruptcy rates are different across geographies: 2.08% in British Columbia, 2.58% in Ontario, and 4.44% in Quebec.

consumer mortgage credit balances are 81,000 versus zero with a standard deviation of 188,000. Judged by the quantiles, most variables were distributed quite asymmetrically, except for age and TU risk score. Since not all consumers carry all types of loan accounts or declare bankruptcy, these summary statistics represent the characteristics of our variables: for example, only around 25% of consumers have mortgage accounts with a mortgage balance value; and bankcard accounts have the lowest mean balance (4,340) compared with other loan account balances. The consumer proportions of three provinces (ON: 39%, BC: 14%, QC: 22%) are consistent with their population size ratio in Canada.

Table 5 Here.

A deeper dive into consumer age is presented in Figure 3 and Figure 4. The distribution mode of the age at which consumers declare bankruptcy is generally around 42 years of age, and the bankruptcy age distributions are a little skewed to the right, indicating that more people declare bankruptcy during their younger age than during their old age. The overall age distribution of consumers is fairly close to a uniform distribution but with fewer people in each age group above the 65–70 group. In addition, the age distributions across geographies show no significant differences.

5.2 Analysis results

The parameter estimates and standard deviations are given in Table 6. The MLE probit model does not converge, so we gray out its output column. The majority of estimated parameters of the models yield the same signs except for the parameters of Ontario in the Bayes probit model. The MLE and Bayes estimates of the logistic model are similar. The signs of the estimated parameters of the semiparametric models are generally consistent with the signs of the estimated parameters of the logit and probit models. This is because the parameters of the semiparametric model are normalized by the parameter of age with positive estimates, β_1: $\theta_i = \dfrac{\beta_i - 1}{\beta_1}$.

Table 6 here.

The model parameter estimates show that consumers of the older generation are more prone to declare bankruptcy in their lifetime. This might be a result of the limited financial instruments

available in the previous century: the peak bankruptcy age is around 48, the age that consumers who were 80 years old in 2020 reached in the 1990s. The older generation might have had fewer debt channels or may have had to claim bankruptcy during several past periods of financial stress (for example, the 1990 or 2007 crisis). A deeper analysis will provide more evidence, but this is outside the topic of this paper.

Table 6 also illustrates that consumers with high mortgage balances might be less likely to claim bankruptcy, indicating that mortgage balance is a safer collateralized (through housing property) debt than high credit card balances, which normally lead to a higher likelihood of bankruptcy. Since a line of credit is generally linked with mortgage credit in Canada, its parameter is positive, showing a higher line of credit balance potentially linked to a relatively lower bankruptcy rate.

An interesting finding is the geographic heterogeneities across different provinces. Since Ontario, Quebec and British Columbia are the three major provinces, with 75% of the Canadian population, we create three dummy variables to try to find the geographic differences. The results indicate that consumers in Quebec have slightly higher bankruptcy rates than those in British Columbia and Ontario.

The MSE and ROC of the parametric and semiparametric models are fairly close, but a careful examination shows that the ROC curves of the semiparametric models are higher than those of the parametric models, as shown in Figure 5. Although the differences among models are small, generally, the semiparametric model estimated by either the MLE or Bayes MCMC yields a better ROC. Based on ROC, we may say that the semiparametric model is a better model for our TU data sample.

Figure 5 Here.

The marginal effects among different models are displayed in Table 7. The marginal effects are more informative measures than the parameters in a binary response model in assessing variable impact. For continuous variables, the marginal effects are computed by adding $\Delta(x)$ or the standard deviation to the variable, and the marginal effect of the dummy variables is computed by the model output difference between the aggregate mean projection and when the corresponding dummy equals zero. When consumer age increases by 16 years, the bankruptcy rate will increase by 133

to 206 basis points. Since we are not looking at time series data, this just reflects that the older generation in the data sample has a higher bankruptcy rate. On the other hand, the bankruptcy rate will be lowered by more than 200 basis points if the TU risk score increases by 124, which is consistent with the credit rating design that the more risk-resilient consumer should have a higher credit score. After increasing the bankcard balance by one standard deviation ($8,800), the bankruptcy rate will be increased by up to 62 basis points, showing potentially high risk in credit card utilization. Although the geographic marginal impacts are not that strong, overall consumers in Quebec seem to contribute more than 25 basis points above the national mean bankruptcy rate.

Table 7 Here.

Since TU consumer data do not include many important demographic factors, such as income, renter's information, education, and ethnic groups, more granular analysis of structural modelling is necessary to draw conclusions about the characteristics of bankruptcy rates for Canadian consumers. This is out of the scope of this research, given the purpose of this session is to demonstrate that the semiparametric model framework has the advantage of using a semiparametric scheme instead of prescribed distributions and provides better model fitting. Meanwhile, we also show how to use the marginal effect to make an assessment.

6 Concluding Remarks

We first presented a Bayesian semiparametric binary response model based on the quasi-likelihood function that is based on the kernel density estimate. The major difference between our Bayesian semiparametric binary response model and the sample theoretic semiparametric binary response model of Klein and Spady (1993) is that we use the MCMC algorithm with Metropolis-Hastings criterion rather than the traditional maximum likelihood estimator. We used the normal kernel and employed two bandwidths: the regular bandwidth and the optimal bandwidth.

Using simulated data, we compared the performances of the semiparametric models with those of the logit and probit models. We used the MLE and MCMC algorithms. The error term of the regression model is generated from 16 different distributions. The comparison of performances is

based on the MSEs, the ROC curve and the marginal effect. We find that the performances of the parametric and semiparametric models are virtually indistinguishable if they are estimated using the MLE or MCMC procedures except when the data are extremely unbalanced (for example, when % of $'Y = 1' < 3\%$). In the extremely unbalanced cases, the MCMC procedure works but the MLE sometimes does not converge.

Although the optimal bandwidth traces sharp modes better than the regular bandwidth, as shown in Figures 1 and 2, the quasi-likelihood function produced by the kernel density with the optimal bandwidth is not very different than the one produced by the regular bandwidth. Consequently, the semiparametric models based on the optimal bandwidth yield virtually the same results as the semiparametric models based on the regular bandwidth.

As an application, we estimated the binary response model using the TU consumer data. We set the consumer with a history of bankruptcy as 1 and the consumer without a bankruptcy history as 0. Judged by the ROC curves, the semiparametric models are better than the parametric models.

There are other types of Bayesian semiparametric qualitative choice models. One model is based on B-splines to approximate the link function using the Laplace transform of the normal distribution (Fahrmeir and Lang (2001), Antoniadis et al. (2004), Fahrmeir and Raach (2007)). The second type of model uses the binary response version of the median regression model (Newton and Chappell (1996), Kottas and Gelfand (2001)). Both of these methods need link functions subject to identifiability. The other type uses a complex Dirichlet process mixture for priors with multiple stages (Kleinman and Ibrahim (1998)). Lu et al. (2019) also propose a two-stage semi-nonparametric estimation of logit models. None of these models applied the single index technique. It will be interesting to compare our Bayesian semiparametric model with the models by these authors.

Appendix: Optimal Bandwidth

Wand and Jones (1995) and Silverman (1986) show that we can obtain the optimal bandwidth h by minimizing the Mean Integrated Squared Error (MISE):

$$MISE\left\{\hat{g}(xlh), g(x)\right\} = E\left[\int (\hat{g}(x;h) - g(x))^2 dx\right],$$

where $g(\cdot)$ is the non-parametric kernel density estimation function. It is clear that integration needs to be made on the whole real line, $x \in (-\infty, \infty)$, instead of a finite discrete set. Li (2011) shows that the choice of the kernel function $K(x)$ is not as important as the choice of the bandwidth. Hence, we will use the standard normal distribution for $K(\cdot) = \Phi(\cdot)$ and will find the optimal bandwidth.

By applying the Central Limit Theorem (CLT), we get an approximation of MISE called the Asymptotic Mean Integrated Squared Error (AMISE):

$$AMISE\left\{\hat{g}(x;h), g(x)\right\} = (Nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(g),$$

where $R(K) = \int K(x)^2 dx$ and $\mu_2(K) = \int x^2 K(x) dx$. The AMISE is a monotonic function of the optimal bandwidth h, and the optimal h is generally defined as:

$$h_{optimal} = \left[\frac{R(K)}{\left(\int x^2 K\right)^2 R(g'')N}\right]^{\frac{1}{5}}.$$

This optimal bandwidth cannot be calculated directly because $R(g'')$ is a function of the second order derivative of the true density function g, which is unknown.

When the data set is Gaussian or asymptotically Gaussian with standard deviation, we will get the regular optimal bandwidth in the literature:

$$h_{optimal} = \left[\frac{4}{3N}\right]^{\frac{1}{5}}\sigma. \tag{12}$$

When data are not normal, this optimal bandwidth may not fit into the real data, and we may use the most popular solve-the-equation plug-in approach and get the optimal bandwidth as

$$h^*_{optimal} = \left[\frac{R(K)}{\left(\int x^2 K(x)dx\right)^2 R(\hat{g}''(x;p(h)))}\right]^{\frac{1}{5}}. \tag{13}$$

Here, $p(h) = \left[\frac{-2K^{(4)}(0)\mu_2(K)\hat{\psi}_4}{R(K)\hat{\psi}_6}\right]^{\frac{1}{7}} h^{\frac{5}{7}}$ is the optimal pilot bandwidth and $\hat{\psi}_r = \frac{1}{N}\sum_{i=1}^{N} \hat{g}^{(r)}(x_i; p^{(r)})$, where $p^{(r)}$ is the pilot bandwidth to estimate the rth derivative of the density $g^{(r)}$.

Equation (12) is the most popular simple optimal bandwidth, and it is only optimal for Gaussian data. If data are not Gaussian, we should use equation (13), which requires multiple complex computations and is extremely time consuming. This is one of the reasons that many people use different estimation methods to estimate bandwidth, such as Shen and Klein (2010) with specific bounds to minimize the bias in estimating bandwidth. The computation in equation (13) can be realized efficiently by using graphic processing unit (GPU) computing with C/C++ in MATLAB (Li (2011)), and its speed is about 600 time faster than the regular computing methods such as in Gauss or MATLAB itself. Therefore, we can estimate much more accurate optimal bandwidth because we consider all real numbers without any arbitrary lower or upper bound.

References

Agresti, A. (2007). *An introduction to categorical data analysis*. 2nd edition, Wiley. chp5.

Alaminos, D., A.D. Castillo and M.A. Fernadez (2016). "A global model for bankruptcy prediction". *Public Library of Science*, 11(11), e0166693.

Aldrich, J.H. and F.D. Nelson (1984). "Linear probability, logit, and probit models". Beverly Hills: Sage Publications, Inc.

Alonzo, T. (2002). "Distribution-free ROC analysis using binary regression techniques". *Biostatistics*, 3, 421-432.

Amemiya, T. (1981). "Qualitative response models. A survey". *Journal of Economic Literature*, 19, 1483-1536

Antoniadis, G., G. Gerard, and I. Ian (2004). "Bayesian estimation in single-index models". *Statistica Sinica*, 14, 1147-1164.

Bester, H. and E. Petrakis (2003). "Wages and productivity growth in a competitive industry". *Journal of Economic Theory*, 109(1), 52-69.

Chen, S. (2015). "Optimal bandwidth selection for Kernel density functionals estimation". *Journal of Probability and Statistics*, 2015, Article ID 242683, 21 pages.

Chen, G. and H. Tsurumi (2010). "Probit and logit model selection". *Communications in Statistics - Theory and Methods*, 40, 159-175.

David, A.J. and H.S. Ann (1999). "Is job stability in the United States falling?" *Journal of Labor Economics*, Vol. 17(4), S1-28.

Fahrmeir, L. and S. Lang (2001). "Bayesian semiparametric regression analysis of multicategorical time-space data". *Annals of the Institute of Statistical Mathematics*, 53, 10-30.

Fahrmeir, L. and A. Raach (2007). "A Bayesian semiparametric latent variable model for mixed responses". *Psychometrika*, 72(3), 327-346.

Fawcett, T. (2006). “An introduction to ROC analysis”. *Pattern Recognition Letters*, Special issue: ROC analysis in pattern recognition archive, 27(8).

Feldstein, M. (2008). “Did wages reflect growth in productivity?” *Journal of Policy Modeling*, 30 (4), 591-594.

Gerfin, M. (1996). “Parametric and semi-parametric estimation of the binary response model of labour market participation”. *Journal of Applied Econometrics*, 11, 321-339.

Gross, B.D. (2002). “An empirical analysis of personal bankruptcy and delinquency”. *The Review of Financial Studies*, 15 (1), 319-347.

Horowitz, J.L. and N.E. Savin (2001). “Binary response models: logits, probits and semiparametrics”. *Journal of Economic Perspectives*, 15(4), 43-56.

James N.B. and L. Audrey (1992). “Interpreting panel data on job tenure”. *Journal of Labor Economics*, 10(3), 219-257.

Klein, R. and R. Spady (1993). “An efficient semi-parametric estimator for the binary response model”. *Econometrica*, 61, 387-421.

Klein, R. and F. Vella (2009). “A semi-parametric model for binary response and continuous outcomes under index heterogeneity”. *Journal of Applied Econometrics*, 24, 735-762.

Klein, R. and C. Shen (2010). “Basic corrections in testing and estimating semiparametric, single index models”. *Econometric Theory*, 26, 1593-1718.

Kleinman, P.K. and J.G. Ibrahim (1998). “A semi-parametric Bayesian approach to generalized linear mixed models”. *Statistics in Medicine*, 17, 2579 -2596.

Kottas, A. and A.E. Gelfand (2001). “Bayesian semiparametric median regression modeling”. *Journal of the American Statistical Association*, 96(456), 1458-1468.

Lawrence M., B. Jared and A. Sylvia (2006). *The state of working America*, 10th edition, ILR Press.

Li, S. (2011). "Combining MATLAB with C/C++ and graphic processing unit: an example of matrix multiplication and kernel density estimation". Chapter 2 of the dissertation, Rutgers University.

Lu, Z., S. Xiaoxia and T. Jing (2019). "Semi-nonparametric estimation of random coefficient Logit model for aggregate demand". Working paper.

Marron, J.S. and M.P. Wand (1992). "Exact mean integrated square errors". *Annals of Statistics*, 20 (2), 712-736.

McNeil, B. and J. Hanley (1984). "Statistical approaches to the analysis of receiver operating characteristic (ROC) curves". *Medical Decision Making*, 4(2), 137-150.

Moler, C. (2013). "The Intel Hypercube, part 2". MATLAB blog retrieved 30 November 2020.

Nelder, J. and R. Wedderburn (1972). "Generalized linear models". *Journal of the Royal Statistical Society*, Series A, 135(3), 370-384.

Newton, C. and M. Chappell (1996). "Bayesian inference for semiparametric binary regression". *Journal of the American Statistical Association*, 91 (433), 142- 153.

Ohlson, J. (1980). "Financial ratios and the probabilistic prediction of bankruptcy". *Journal of Accounting Research*, 18(1), 109-131.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, New York.

Swets, J.A., R.M. Dawes and J. Monahan (2000). "Better decisions through science". *Scientific American*, 283, 82-87.

Wand, M.P. and M.C. Jones (1995). *Kernel smoothing*. Chapman & Hall, New York.

Table 2: ROC areas and MSE: Balanced cases

	Strongly skewed		**Sparated bimodal**		**Claw**	
	Y=1 is 36.5%		**Y=1 is 29.1%**		**Y=1 is 28.8%**	
	ROC area	**MSE**	**ROC area**	**MSE**	**ROC area**	**MSE**
	Bayes		**Bayes**		**Bayes**	
Probit	0.696	0.202	0.974	0.060	0.993	0.029
Logit	0.696	0.201	0.974	0.061	0.993	0.029
Semi	0.697	0.196	0.974	0.061	0.992	0.035
Semi-opt	0.697	0.195	0.974	0.060	0.992	0.034
	MLE		**MLE**		**MLE**	
Probit	0.696	0.202	0.974	0.060	0.993	0.029
Logit	0.696	0.201	0.974	0.061	0.993	0.029
Semi	0.698	0.195	0.974	0.060	0.993	0.032
Semi-opt	0.699	0.196	0.974	0.061	0.993	0.033

Notes: semi = semi parametric with the bandwidth h in equation (9)

semi-opt = semi prametric with the optimal $h^*_{optimal}$ in equation (10)

Table 3: ROC areas and MSE: Extremely unbalanced cases

	Skewed logitic (θ = .1) Y=1 is 0.9%		**Outlier** Y=1 is 3.47%		**Kurtotic unimodal** Y=1 is 1.5%	
	ROC area	**MSE**	**ROC area**	**MSE**	**ROC area**	**MSE**
	Bayes		**Bayes**		**Bayes**	
Probit	.968	.007	.991	.007	.893	.311
Logit	.967	.007	.991	.007	.878	.108
Semi	.967	.006	.991	.007	.999	.006
Semi-opt	.967	.006	.991	.007	.999	.006
	MLE		**MLE**		**MLE**	
Probit	NC		NC		NC	
Logit	NC		NC		NC	
Semi	NC		NC		NC	
Semi-opt	NC		NC		NC	

Notes: semi = semi parametric with the bandwidth h in equation (9)

semi-opt = semi prametric with the optimal $h^*_{optimal}$ in equation (10)

NC = not converged

Table 4: Monte Carlo experiment result

Claw distribution, around 20% of 'Y=1' Replications = 500							
Binary model evaluation criteria						Marginal effect	
ROC area	**MEAN**	**STD.**	**MSE**	**MEAN**	**STD.**	**MEAN**	**STD.**
Bayes Probit	0.8357	0.0283	Bayes Probit	0.0370	0.0042	0.0702	0.0121
Bayes Logit	0.8356	0.0284	Bayes Logit	0.0370	0.0042	0.0739	0.0122
Bayes Semi	0.8387	0.0270	Bayes Semi	0.0365	0.0042	0.0519	0.0084
Bayes Semiopt	0.8384	0.0273	Bayes Semiopt	0.0365	0.0042	0.0516	0.0086
MLE Probit	0.8357	0.0283	MLE Probit	0.0370	0.0042	0.0704	0.0121
MLE Logit	0.8355	0.0283	MLE Logit	0.0370	0.0042	0.0739	0.0122
MLE Semi	0.8428	0.0261	MLE Semi	0.0363	0.0042	0.0576	0.0103
MLE Semiopt	0.8427	0.0258	MLE Semiopt	0.0364	0.0042	0.0552	0.0113
Bayes Semi optimal bandwidth	0.3579	0.0198	**MLE Semi optimal bandwidth**	0.3323	0.0480		
Skewed log alpha=0.1, around 0.9% of 'Y=1' Replications = 500							
Binary model evaluation criteria						Marginal effect	
ROC area	**MEAN**	**STD.**	**MSE**	**MEAN**	**STD.**	**MEAN**	**STD.**
Bayes Probit	0.9750	0.0098	Bayes Probit	0.0066	0.0015	0.0481	0.0112
Bayes Logit	0.9746	0.0099	Bayes Logit	0.0067	0.0015	0.0482	0.0120
Bayes Semi	0.9760	0.0088	Bayes Semi	0.0063	0.0014	0.0395	0.0098
Bayes Semiopt	0.9763	0.0087	Bayes Semiopt	0.0062	0.0014	0.0411	0.0102
Bayes Semi optimal bandwidth	0.2578	0.0076					

Notes: For the skewed log data sample, not all 500 MC iterations are converged for the MLE method;

only Bayes results are presented for skewed log case.

Table 5: Summary statistics of TransUnion consumer data

Variable	Mean	StDev	Min	Q1	Median	Q3	Max
Age in months of consumer	580	194	217	417	579	734	983
TU RISK 2009 Score	747	124	302	676	794	846	884
Total balance for mortgage trades	81	188	0	0	0	84	6,603
Total balance for bankcard trades	4.34	8.81	0	0.05	1.1	4.50	333
Total balance for line of credit trades	15	71	0	0	0	3.50	6,158
Ontario	0.39		0	0	0	1	1
British Columbia	0.14		0	0	0	0	1
Quebec	0.22		0	0	0	0	1

Notes: Q1 and Q3 are 25 and 75 percentiles, respectively.

Table 6: Estimations for the TransUnion consumer data sample (**Y=1 is 3.12%**)

variable		**Bayes**[1]		**MLE**[2]		**Semi** (θ)	
		Probit	**Logistic**	**Probit**	**Logistic**	**Bayes**	**MLE**
c	Constant	0.5493 0.4668	1.9331* 0.0638		1.9323* 0.0502	/	/
β_1	Age	0.0012* 0.0006	0.0031* 0.0001		0.0031* 0.0001	/	/
β_3	TU risk score	-0.0044* 0.0020	-0.0114* 0.0001		-0.0114* 0.0001	-4.8769* 0.7436	-29.0191* 0.1256
β_3	Mortgage balance	-0.0007 0.0005	-0.0023* 0.0002		-0.0022* 0.0002	-0.2916* 0.0492	-0.2368* 0.0244
β_4	Bankcard balance	0.0089* 0.0034	0.0242* 0.0011		0.0240* 0.0016	10.0038* 1.0479	-6.0898* 0.0165
β_5	Line of credit	-0.0019* 0.0010	-0.0055* 0.0007		-0.0053* 0.0006	-0.7731* 0.1309	-2.0174* 0.0633
β_7	Ontario	0.0278* 0.0250	0.0586* 0.0351		0.0611* 0.0324	-5.2680* 0.9754	10.6488* 0.0237
β_7	British Columbia	-0.1211* 0.0609	-0.3121* 0.0577		-0.3100* 0.0356	-97.9473* 1.3101	-13.2886* 0.0366
β_7	Quebec	0.3182* 0.1149	0.8179* 0.0394		0.8198* 0.0344	209.2817* 1.6437	119.2758* 0.0476
	MSE	0.0257	0.0255		0.0255	0.0235	0.0241
	ROC area	0.9175	0.9175		0.9175	0.9470	0.9402

Notes: 1. For MLE: first row is the coefficient estimates, second row is standard error. Probit MLE does not converge.

2. For Bayes: first row is posterior mean and second row is posterior standard error, respectively.

3. * means statistically significant.

Table 7: Marginal effects of the TransUnion consumer data sample

Explanatory variables	**Marginal effects**	**Bayes**		**MLE**		**Bayes-Semi**	**MLE-Semi**
		Probit	**Logistic**	**Probit**	**Logistic**		
Age	$\Delta x = 1std$	0.0206	0.0183		0.0182	0.0133	0.0205
TU risk score	$\Delta x = 1std$	-0.0282	-0.0216		-0.0216	-0.0209	-0.0207
Mortgage balance	$\Delta x = 1std$	-0.0086	-0.0090		-0.0089	-0.0030	-0.0046
Bankcard balance	$\Delta x = 1std$	0.0062	0.0056		0.0056	0.0055	0.0056
Line of credit	dummy	-0.0090	-0.0081		-0.0081	-0.0030	-0.0015
Ontario	dummy	0.0007	0.0005		0.0005	0.0001	-0.0001
British Columbia	dummy	-0.0010	-0.0008		-0.0008	-0.0007	-0.0002
Quebec	dummy	0.0055	0.0046		0.0046	0.0028	0.0025

Notes: *std* = standard deviation.

Figure 1: Various distributions and kernel densities using the regular bandwidth

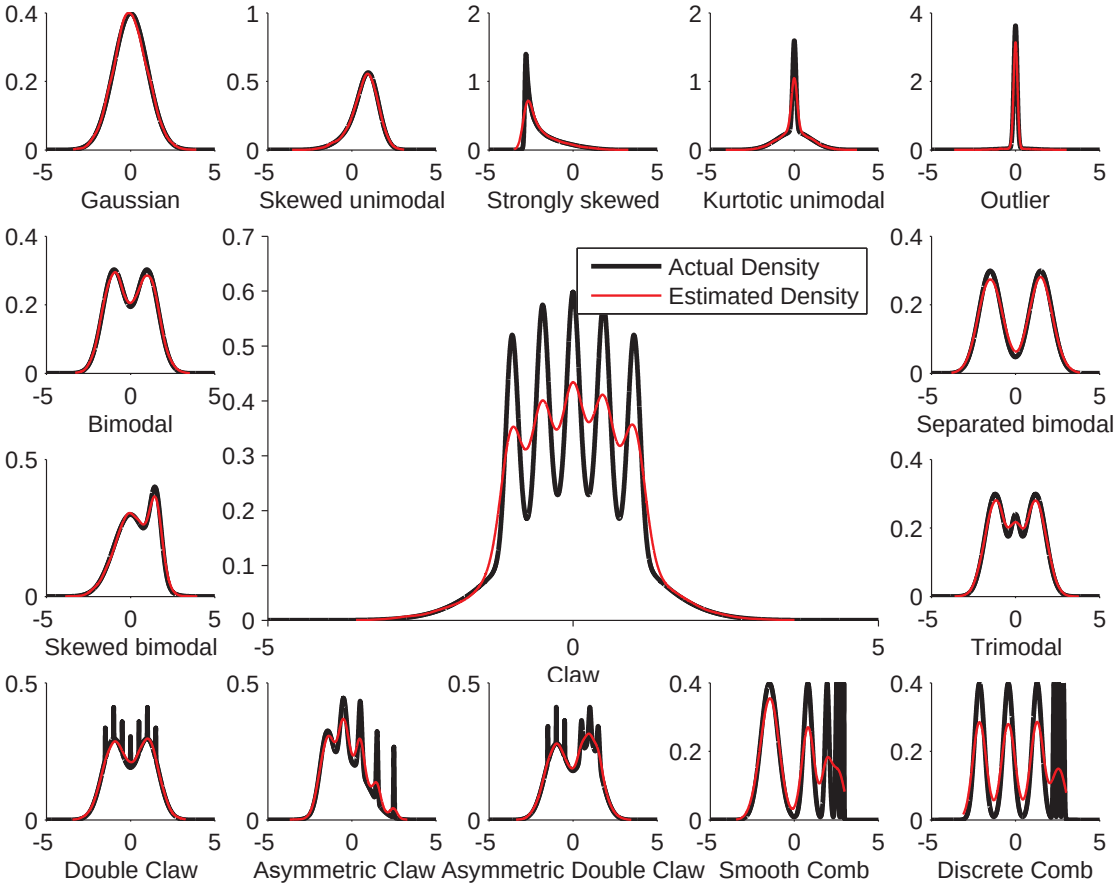

Figure 2: Various distributions and kernel densities using the optimal bandwidth

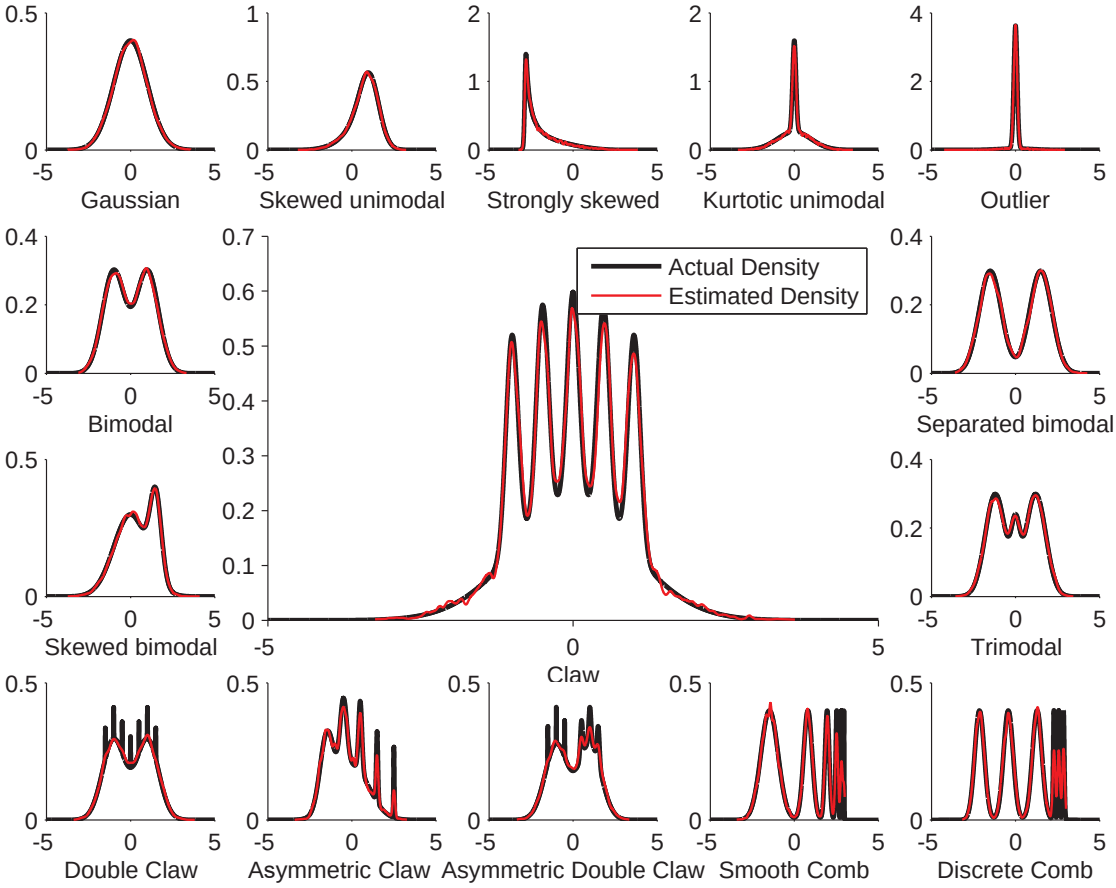

Figure 3: Age density distributions of all consumers

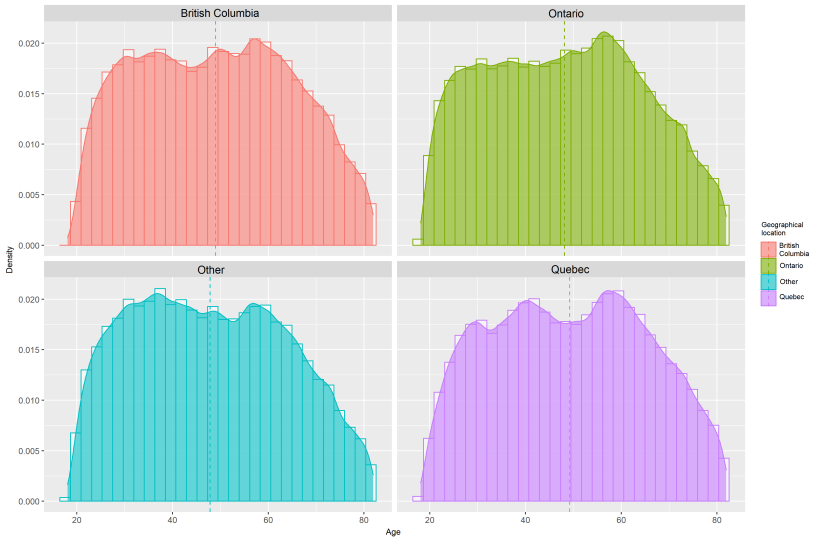

Figure 4: Bankrupt age density distributions

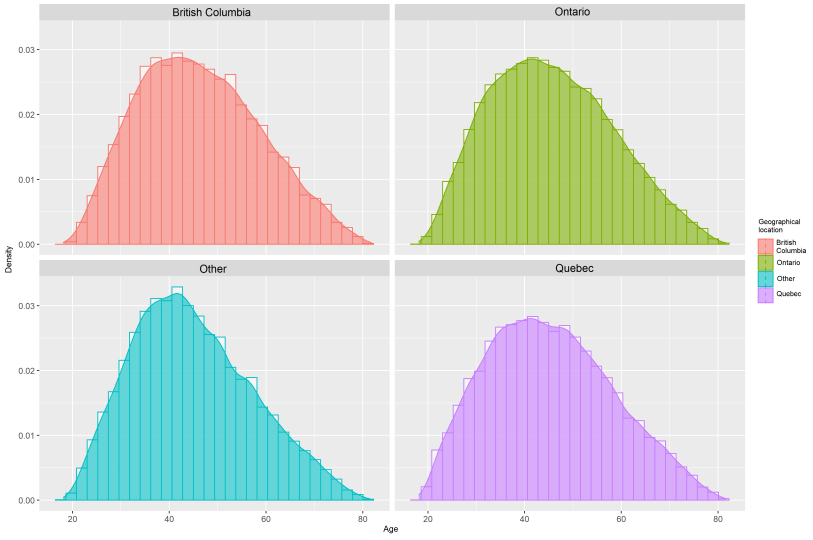

Figure 5: ROC curve analysis for the TransUnion consumer data: Semiparametric models generate larger ROC areas