Bayesian evaluation of a semi-parametric binary response model

Eliana SCHEIHING[1] **and Michel MOUCHART**[2]

March 17, 1998

Abstract

In this paper, we develop a Bayesian analysis of a semi-parametric binary choice model. The prior specification of the functional parameter, namely the distribution function of a latent variable, is of the Dirichlet process type and the prior specification of the Euclidean parameter, namely the coefficients of a linear combination of exogenous variables, is left arbitrary. The model identification is ensured by fixing the prior expectation of the functional parameter (see Mouchart et al. (1997)). Approximations for the posterior predictive distributions are obtained from two different sampling methods. Several questions are studied through an exploratory numerical analysis, such as the numerical convergence of the algorithms and of the methods and the general problem of contrasting semi-parametric and purely parametric specification.

Keywords: Discrete choice model, semi-parametric model, Dirichlet process, Gibbs sampling, Bayesian specification

AMS-MOS : Primary : 62G05, Secondary: 62P20

Acknowledgements : Research supports from the contract "Projet d'Actions de Recherche Concertées" No. 93/98-164, from the Belgian Program on Interuniversity Poles of Attraction (No. 26), from the Belgian OSTC Transport and Mobility (contract TRB3/007), from the Departmental fellowship of UCL and from DID-UACh grant S-96-30 are gratefully acknowledged. This version has benefited from comments by J.P. Florens, M.-P. Kestemont, F. Quintana and J.P. Raoult and more particularly from R. Cano and M. Steel. The help from J.M. Rolin went far beyond the border of friendly, vivid and fruitful discussions; the only limit is that he should in no way be taken as responsible for any remaining error.

[1]Instituto de Informática, Universidad Austral de Chile, Valdivia, Chile

[2]CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

1 Introduction

In this paper we consider a Bayesian analysis of a binary choice model in a semi-parametric setup. More specifically, the link function is taken as a functional parameter and the index function is specified as a linear function, the coefficients of which conform a Euclidean parameter.

The prior specification follows standard lines: prior independence between the functional and the Euclidean parameter, a Dirichlet process for the functional parameter and an arbitrary unspecified distribution for the Euclidean parameter. In a Bayesian framework, the model identification of such a model only requires to fix an (arbitrary) prior expectation for the functional parameter as shown in Mouchart et al(1997).

The main theme of this paper is to display alternative strategies for obtaining numerical evaluations of the posterior distributions, based on a careful use of the structural properties of the model. In next section, we give a precise specification of the relevant Bayesian model, viz. its sampling component and its prior component. The sampling model is specified by means of two observationally equivalent models, one of which is based on latent variables, the usefulness of which will be crucial for the simulations. Taking advantage of the structural properties of this Bayesian model, as dispelled also in next section, Section 3 presents alternative strategies for evaluating posterior distributions. Two Gibbs samplers of the latent variables are introduced. One is based on an a.s. representation of the trajectory of a Dirichlet process (as developed in Rolin (1993) and Sethuraman(1994)). The other one is based on a Polya urn representation of the Dirichlet process (as developed in Blackwell and McQueen(1973), see also Escobar(1994)). In Section 4 a numerical experiment compares results obtained from this semi-parametric model and those obtained from a purely parametric model. Some difficulties raised by the simulation of semi-parametric models are also discussed. The last section presents some concluding remarks and sketch some perspectives for future work. An appendix first summarizes, for the sake of easy reference, the definitions and properties related to conditional independence and the trajectories of the Dirichlet process and next presents the proof of most statements given in the main text.

2 Specification of the Bayesian model

2.1 Model hypotheses

Let us consider a conditionally independent Bernoulli sampling. The data are in the form $X = (Y, Z)$ where Y is a binary response variable and Z is a vector of explanatory variables. We shall use the notation $Z_1^\infty = (Z_1, Z_2, \cdots)$, respectively $Z_1^n = (Z_1, \cdots, Z_n)$, for an infinite, respectively finite of size n, trajectory of the $Z-$process. The sampling model may accordingly be characterized by:

$$(Y_i \mid Z_1^\infty, \theta) \sim Ind.Bernoulli(\psi(\beta' Z_i)) \tag{1}$$

where $\theta = (\psi, \beta)$ is the model parameter: ψ is a functional parameter, namely a (unknown) cumulative distribution function on $I\!R$, which is generally called the *link function* and $\beta \in I\!R^p$ is a *Euclidean parameter*; $\beta' Z_i$ is often called an *index function*. This model is also known as a single-index semi-parametric model. An observationally equivalent model is presented in terms of *latent variables* λ_i with the following hypotheses:

$$\begin{aligned} \perp\!\!\!\perp \lambda_i \quad &\mid \quad \theta, Z_1^\infty && (2)\\ \lambda_1^\infty \perp\!\!\!\perp (\beta, Z_1^\infty) \quad &\mid \quad \psi && (3)\\ \lambda_i \quad &\mid \quad \psi \quad \sim \quad i.i.d.\, \psi, && (4) \end{aligned}$$

the response variable Y_i being represented as:

$$Y_i = 1\!\!1_{\{\lambda_i \leq \beta' Z_i\}} \tag{5}$$

This second presentation is of interest at the level of structural modelling where de λ_i's may be interpreted as "levels of tolerance", for instance but not only in biometrics, or as "difference of residual utilities" among two alternatives, for instance in discrete choice models. As we shall see later on, this second presentation is also useful for the development of simulation based methods, particularly when simulating the posterior or predictive distributions.

For the prior specification, we assume a general independence, as usual in semi-parametric models, namely:

$$\psi \perp\!\!\!\perp \beta \perp\!\!\!\perp Z_1^\infty \tag{6}$$

For ψ we assume a Dirichlet process:

$$\psi \sim \mathcal{D}i(n_0, P_0) \tag{7}$$

where $n_0 \in I\!R_+$ represents the weight of the prior information and P_0 is a distribution function on $I\!R$ representing the prior expectation of ψ: $E(\psi) = P_0$. For β, we assume an arbitrary and unspecified distribution Q_0 on $I\!R^p$:

$$\beta \sim Q_0; \tag{8}$$

sometimes we shall also write $q_0(\beta)$ for its density.

For identification purposes (see Mouchart et al(1997)), we also assume that:

(i) The exogenous process is such that $\{Z_i : i \in I\!N\}$ is a.s. dense in $I\!R^p$.

(ii) Q_0 is not degenerate in the sense that there exists $j_0 > 1$ such that: $Q_0(\{\beta_{j_0} = 0\}) = 0$.

(iii) P_0 is not a mass point.

2.2 Some properties

The simulations to be presented make a systematic use of the properties of this model, some of which are properties of the sampling process whereas other properties depend on the prior specification. In this section we list the main properties for easier reference and delay in the appendix the proofs of these properties.

Let η_i denote *the linear predictor function* (or linear index function):

$$\eta_i = \beta' Z_i \tag{9}$$

As before we shall also use the notation $\eta_1^n = (\eta_1, \cdots, \eta_n) = (\beta' Z_1, \cdots, \beta' Z_n)$. The first property asserts that η_i, as a function of Z_i and β, operates a *split* (in the sense of definition 5.4.16 in Florens at al. (1990), see also Appendix) between (Z_i, β) and (ψ, Y) for any sample size n. More precisely:

Property 2.1 *:*

$$(Z_1^n, \beta) \perp\!\!\!\perp (\psi, Y_1^n) \mid \eta_1^n$$

The second property states that if the latent variables λ_i's were actually observable, the posterior distribution of ψ would only depend on λ_1^n. More precisely:

Property 2.2 *:*

$$\psi \perp\!\!\!\perp (X_1^n, \beta) \mid \lambda_1^n$$

More specifically:

$$\psi \mid \lambda_1^n \quad \sim \quad \mathcal{D}i((n+n_0), \frac{n_0 P_0 + n\hat{F}_n}{n+n_0})) \tag{10}$$

where $\hat{F}_n$ *is the empirical c.d.f. on* λ_1^n*:*

$$\hat{F}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{\{\lambda_j\}}$$

($\delta_{\{\lambda_j\}}$ *is the degenerate probability measure giving a unitary mass to the point* $\{\lambda_j\}$*). Consequently, therefore*

$$\beta \perp\!\!\!\perp \psi \mid X_1^n, \lambda_1^n$$

More specifically:

$$m(\beta \mid X_1^n, \lambda_1^n) \propto q_0(\beta)\mathbb{1}_{\{\beta \mid \forall i (Y_i = 1 \Leftrightarrow \lambda_i \leq Z_i'\beta)\}} \tag{11}$$

The next two properties are crucial for making feasible the Gibbs sampling on the space of the latent variables $\lambda_i' s$. The following property is based on the fact that, given ψ, λ_1^n is *sifted* (in the sense of definition 7.6.10 in Florens et al. (1990), see also appendix) by (η_1^n, Y_1^n). More precisely:

Property 2.3 *:*

(a) $\perp\!\!\!\perp_{1 \leq i \leq n} \lambda_i \mid \eta_1^n, Y_1^n, \psi$

(b) $\lambda_i \perp\!\!\!\perp (\eta_1^n, Y_1^n) \mid \eta_i, Y_i, \psi$

More specifically:

(c)

$$P(\lambda_i \leq l \mid \eta_i, Y_i, \psi) = \begin{cases} \frac{\psi(l)}{\psi(\eta_i)} & if \quad Y_i = 1, l \leq \eta_i \\ 1 & if \quad Y_i = 1, l > \eta_i \\ 0 & if \quad Y_i = 0, l \leq \eta_i \\ \frac{\psi(l)-\psi(\eta_i)}{1-\psi(\eta_i)} & if \quad Y_i = 0, l > \eta_i \end{cases} \tag{12}$$

Combining these previous properties allows one to obtain a similar property after integrating out ψ. More precisely:

Property 2.4 *:*
(a) $\lambda_i \perp\!\!\!\perp (\eta_1^n, Y_1^n) \mid (\lambda_j, j \neq i), \eta_i, Y_i$

More specifically, we obtain:
(b)

$$(\lambda_i \mid (\lambda_j, j \neq i), \eta_i, Y_i) \sim \frac{\alpha_{i,1}}{\alpha_{i,1} + \alpha_{i,2}} F_{i,1} + \frac{\alpha_{i,2}}{\alpha_{i,1} + \alpha_{i,2}} F_{i,2} \tag{13}$$

where

$$\begin{aligned}
\alpha_{i,1} &= \sum_{j \neq i} q_{ij} \\
F_{i,1}((-\infty, l]) &= \frac{1}{\alpha_{i,1}} \sum_{j \neq i} q_{ij} \mathbb{1}_{\{\lambda_j \leq l\}} \\
q_{ij} &= \begin{cases} \mathbb{1}_{\{\lambda_j \leq \eta_i\}} & if \quad Y_i = 1 \\ \mathbb{1}_{\{\lambda_j > \eta_i\}} & if \quad Y_i = 0 \end{cases} \\
\alpha_{i,2} &= \begin{cases} n_0 P_0(\eta_i) & if \quad Y_i = 1 \\ n_0(1 - P_0(\eta_i)) & if \quad Y_i = 0 \end{cases} \\
F_{i,2}((-\infty, l]) &= \begin{cases} \frac{P_0(l)}{P_0(\eta_i)} & if \quad Y_i = 1, l \leq \eta_i \\ 1 & if \quad Y_i = 1, l > \eta_i \\ 0 & if \quad Y_i = 0, l \leq \eta_i \\ \frac{P_0(l) - P_0(\eta_i)}{1 - P_0(\eta_i)} & if \quad Y_i = 0, l > \eta_i \end{cases}
\end{aligned}$$

At the end of the Section 2.1, we gave necessary conditions for the Bayesian identification in the asymptotic model. Clearly, for a finite sample, there is no hope to identify an infinity of characteristics of ψ. Next property describes which part of ψ is actually identified for a given sample size n $(< \infty)$.

Property 2.5 *Consider the index permutation such that:*

$$\eta_{(1)} < \eta_{(2)} < \cdots < \eta_{(n_\beta)}$$

where n_β is the number of distinct values in η_1^n. To simplify the presentation, without substantial loss of generality, we consider in the following $n_\beta = n$. For given β, consider the following partition of $\mathbb{R}$:

$$\begin{aligned} B_1 &= (-\infty, \eta_{(1)}] \\ B_2 &= (\eta_{(1)}, \eta_{(2)}] \\ &\cdots \\ B_n &= (\eta_{(n-1)}, \eta_{(n)}] \\ B_{n+1} &= (\eta_{(n)}, +\infty) \end{aligned}$$

Then:

(a) $Y_1^n \perp\!\!\!\perp \psi \mid \psi(B_1), \cdots, \psi(B_{n+1})$

(b) $(Y_1^n, \psi(B_1), \cdots, \psi(B_{n+1})) \perp\!\!\!\perp (\psi^{B_1}, \cdots, \psi^{B_{n+1}})$

where ψ^{B_j} is the probability measure ψ conditional on B_j.

Corollary:

$$Y_1^n \perp\!\!\!\perp (Z, \beta, \psi) \mid \psi(B_1), \cdots, \psi(B_{n+1})$$

Thus, for any value of Z, the class of observationally equivalent parameters on the (β, ψ)-space is characterized by $\psi(B_1), \cdots, \psi(B_{n+1})$; clearly this partition of the parameter space depends on Z and β. The fact that only $\psi(B_1)$, $\cdots$, $\psi(B_{n+1})$ is identified by Y_1^n also means that, conditionally on $\psi(B_1), \cdots, \psi(B_{n+1})$, the distribution of $\psi^{B_1}, \cdots, \psi^{B_{n+1}}$ is not revised by a sample of size n.

Next property is important for ensuring the convergence of the Gibbs sampler on the λ_i's (see Roberts and Smith (1994)).

Property 2.6 *The support D of the joint distribution of the vector λ_1^n conditionally on the sample and β has the following property:*

$$D = D_1 \times D_2 \times \cdots \times D_n$$

The last property ensures the convergence of the Bayesian semi-parametric model in a specific sense:

Property 2.7 *The Bayesian semi-parametric model defined in Section 2.1 verifies:*

$$\lim_{n\to\infty} E(Y_0 \mid Z_0, X_1^n) = E(\psi(Z_0'\beta) \mid Z_0, X_1^\infty) \qquad a.s.$$

3 Simulation of the posterior distribution

3.1 The posterior distribution of ψ conditionally on β

In this section we consider two methods for evaluating the posterior distribution of ψ conditionally on β. Both methods are based on Property 2.2 of previous section which ensures that, if λ_1^n were observable, the conditional posterior distribution of ψ given β would depend on λ_1^n only, and would be a standard Dirichlet process. The sample from λ_1^n is obtained in both cases through Gibbs sampling techniques.

Method 1:
The first method performs a Gibbs sampling jointly on ψ and λ_1^n, by drawing successively:

(1) $\psi^{(k+1)} \mid Y_1^n, \eta_1^n, \lambda_1^{n\,(k)}$

(2) $\lambda_1^{n\,(k+1)} \mid Y_1^n, \eta_1^n, \psi^{(k+1)}$

where drawings for the first step are performed according to the posterior distribution in equation (10) of Property 2.2 and making use of the representation of the Dirichlet process given by Rolin(1993) (see Appendix, Section A.2). For the generation of λ_1^n in the second step, each component λ_i is drawn independently from the distribution conditional on ψ as specified in equation (12) of Property 2.3.

This Gibbs sampler allows us to approximate the joint distribution of

$$\psi, \lambda_1^n \mid Y_1^n, \eta_1^n$$

and eventually the conditional expectation of any measurable function of (ψ, λ_1^n). Note that Property 2.1 implies that

$$\psi \mid X_1^n, \beta \sim \psi \mid Y_1^n, \eta_1^n$$

because η_1^n is a function of (Z_1^n, β); thus this Gibbs sampler also provides one with the posterior distribution of ψ given β. This Gibbs sampler also allows us to approximate the predictive posterior expectations. Indeed:

$$\begin{aligned}
E(Y_{n+1} \mid X_1^n, \beta, z_{n+1}) &= E[\,E(Y_{n+1} \mid X_1^n, \beta, z_{n+1}, \psi) \mid X_1^n, \beta, z_{n+1}] \\
&= E(\psi(\eta_{n+1}) \mid Y_1^n, \eta_1^{n+1}) \\
&= E[\,E(\psi(\eta_{n+1}) \mid Y_1^n, \eta_1^{n+1}, \lambda_1^n) \mid Y_1^n, \eta_1^{n+1}] \\
&= E(E(\psi \mid \lambda_1^n)(\eta_{n+1}) \mid Y_1^n, \eta_1^{n+1}) \qquad \text{by Property 2.2} \\
&= E(f(\lambda_1^n)(\eta_{n+1}) \mid Y_1^n, \eta_1^{n+1})
\end{aligned}$$

where

$$f(\lambda_1^n) \quad = \quad E(\psi \mid \lambda_1^n) = \left[\frac{1}{n+n_0}(n_0 P_0 + n\hat{F}_n(\lambda_1^n))\right]$$

Thus, the predictive posterior expectation can be approximated by:

$$\hat{E}[Y_{n+1} \mid X_1^n, \beta, z_{n+1}] \quad = \quad \frac{1}{K}\sum_{k=1}^{K} f(l_1^{n\,(k)})$$

where $l_1^{n\,(k)}$, $k = 1, \cdots, K$ are K drawings from $\lambda_1^n \mid Y_1^n, \eta_1^n$ provided by this Gibbs sampler.

Method 2:
The second method performs a Gibbs sampling on the full conditional distribution of $(\lambda_i \mid \lambda_j, j \neq i)$, integrated with respect to ψ, by drawing succesively:

(1) $\lambda_1^{(k+1)} \mid \lambda_2^{(k)}, \cdots, \lambda_n^{(k)}, Y_1^n, \eta_1^n$

(2) $\lambda_2^{(k+1)} \mid \lambda_1^{(k+1)}, \lambda_3^{(k)}, \cdots, \lambda_n^{(k)}, Y_1^n, \eta_1^n$

$\cdots$

(n) $\lambda_n^{(k+1)} \mid \lambda_1^{(k+1)}, \cdots, \lambda_{n-1}^{(k+1)}, Y_1^n, \eta_1^n$

For each step, $(\lambda_i \mid \lambda_j, j \neq i)$ is drawn following the conditional distribution in equation (13) of Property 2.4.
This Gibbs sampler allows us to approximate the joint distribution of

$$\lambda_1^n \mid Y_1^n, \eta_1^n,$$

and the expectation of any measurable function of λ_1^n. In particular, as in *Method 1*, we can approximate the predictive posterior expectation by:

$$\hat{E}[Y_{n+1} \mid X_1^n, \beta, z_{n+1}] = \frac{1}{K} \sum_{k=1}^{K} f(l_1^{n\,(k)})(\eta_{n+1})$$

where $l_1^{n\,(k)}$, $k = 1, \cdots, K$ are K drawings from $\lambda_1^n \mid Y_1^n, \eta_1^n$ provided by this Gibbs sampler.

3.2 Estimation of the joint posterior distribution

In this section we present a method to evaluate simultaneously the posterior distribution of the two parameters. The idea is to incorporate β into the iterative scheme of a Gibbs sampler.
From Property 2.2 of previous section we have an explicit expression of the posterior distribution of β conditionally to the other parameters of the two Gibbs sampling schemes of previous section. Thus we can incorporate β into both Gibbs samplings.

Method 1:
The first Gibbs sampling becomes a three-steps sampling scheme:

(1) $\psi^{(k+1)} \mid X_1^n, \lambda_1^{n\,(k)}, \beta^{(k)}$

(2) $\lambda_1^{n\,(k+1)} \mid X_1^n, \psi^{(k+1)}, \beta^{(k)}$

(3) $\beta^{(k+1)} \mid X_1^n, \psi^{(k+1)}, \lambda_1^{n\,(k+1)}$

where ψ and λ_1^n are drawn according to the posterior distribution defined in previous section by equations (10) and (12) respectively (where η_i is still

defined by $\eta_i = Z_i'\beta)$ and β is drawn according to the posterior distribution (11) defined above.
This Gibbs sampler allows us to approximate the distribution of:

$$\psi, \lambda_1^n, \beta \mid X_1^n$$

and the expectation of any measurable function of ψ, λ_1^n, β. In particular, let

$$f(\lambda_1^n, \beta, z_{n+1}) = E(\psi \mid \lambda_1^n)(\beta' z_{n+1}) = \left[\frac{1}{n+n_0}(n_0 P_0 + n\hat{F}_n(\lambda_1^n))\right](\beta' z_{n+1})$$

then, we can approximate the predictive posterior expectation by:

$$\hat{E}[Y_{n+1} \mid X_1^n, z_{n+1}] = \frac{1}{K}\sum_{k=1}^{K} f(l_1^{n\,(k)}, b^{(k)}, z_{n+1})$$

where $l_1^{n\,(k)}, b^{(k)}$, $k = 1, \cdots, K$ are K drawings from $\lambda_1^n, \beta \mid X_1^n$ provided by this Gibbs sampler.

Method 2:
The second Gibbs sampling becomes a $(n+1)$-steps sampling scheme:

(1) $\lambda_1^{(k+1)} \mid \lambda_2^{(k)}, \cdots, \lambda_n^{(k)}, \beta^{(k)}, X_1^n$

(2) $\lambda_2^{(k+1)} \mid \lambda_1^{(k+1)}, \lambda_3^{(k)}, \cdots, \lambda_n^{(k)}, \beta^{(k)}, X_1^n$

$\cdots$

(n) $\lambda_n^{(k+1)} \mid \lambda_1^{(k+1)}, \cdots, \lambda_{n-1}^{(k+1)}, \beta^{(k)}, X_1^n$

$(n+1)$ $\beta^{(n+1)} \mid \lambda_1^{n\,(k+1)}, X_1^n$

where λ_i are drawn from the posterior distribution of λ_i conditionally on all others $\lambda_{j,j\neq i}$ and β according to (13) and β is drawn according to (11).
This Gibbs sampler allows us to approximate the distribution of:

$$\lambda_1^n, \beta \mid X_1^n$$

i.e. ψ is integrated out analytically. In particular , as in *Method 1*, we can approximate the predictive posterior expectation by:

$$\hat{E}[Y_{n+1} \mid X_1^n, z_{n+1}] = \frac{1}{K} \sum_{k=1}^{K} f(l_1^{n\,(k)}, b^{(k)}, z_{n+1})$$

where $l_1^{n\,(k)}, b^{(k)}$, $k = 1, \cdots, K$ are K drawings from $\lambda_1^n, \beta \mid X_1^n$ provided by this Gibbs sampler.

4 A numerical experiment

4.1 Introduction

In this section we present some numerical experiments in order to evaluate the proposed methods. The general objective of this analysis is to gain a better understanding about the numerical behavior of the Bayesian semi-parametric models. We want to analyze topics related to the numerical and statistical stability of the methods.
From a preliminary study, we had learned that both methods in previous section, behave in a roughly similar way, at least as far as speed of convergence is concerned. Nevertheless, each iteration of the first method is more demanding from a computational point of view. This is likely due to the fact that the second method integrates analytically the functional parameter (ψ) whereas the first one requires a numerical generation ψ. So that we concentrated in this experiment on the second method.
On one hand, we analyse the numerical behaviour of the second method concerning the following aspects:

(i) The convergence of the Markov Chain Monte Carlo method towards its stationary regime

(ii) The sampling convergence of the posterior distributions (pointwise on the parameter space)

(iii) The sensitivity to the data

On the other hand we sketch a comparison of parametric and semi-parametric models, in presence of more or less specification error of the parametric model.
The structure of the simulation is given in the following section while the results are presented in the section 4.3.

4.2 Simulation design

The data have been generated according to the following scheme:

(i) First, the values of the parameters, the sample size and the vector of explanatory variables are fixed. More specifically, let

$b^* \in I\!R^p$ be a given value of the Euclidean parameter

F^* be a distribution function on $I\!R$, the given value of the functional parameter

$Z_1^n = \{Z_1, \cdots, Z_n\}$ be a given set of values of the vector of explanatory variables, $Z_i \in \{1\} \times I\!R^{p-1}, i = 1, \cdots, n$.

(ii) Next, a random sample of the latent variables is drawn according to F^*:

$$\lambda_1, \cdots, \lambda_n \sim i.i.d\, F^*$$

(iii) Finally, the response variables are generated as follows:

$$Y_i = 1\!\!1_{\{\lambda_i \leq Z_i' b^*\}}$$

Thus, we have a sample $X_1^n = (Y_1^n, Z_1^n)$, from which we perform our estimations regardless of the generated values of the latent variable.

In this paper, we present an example where $p = 2$, that is with only one explanatory variable, $b^* = (-3, 1)$, F^* is a Student distribution of parameter $(\nu, 0, 1)$, where $\nu = 1$ or 10 represent more or less specification error of the parametric model (defined below). Moreover we consider samples of size 10,20,50 and 100.

Parametric estimation:
The parametric model is defined by the following hypotheses:

(i) $Y_i \mid Z_i, \tilde{\beta} \sim I.Bernoulli(\Phi(Z_i'\tilde{\beta}))$, where Φ is the cumulative distribution function of a $\mathcal{N}(\prime, \infty)$ a standard Normal distribution, or equivalently:

$$Y_i = 1\!\!1_{\{\lambda_i \leq Z_i' \tilde{\beta}\}} \text{ and } \lambda_i \sim i.i.d.\, \mathcal{N}(0, 1)$$

.

(ii) $Z \perp\!\!\!\perp \tilde{\beta}$ and a priori $\tilde{\beta}$ is uniformally distributed on a evenly spread grid of points on a large but finite rectangle $B \subset I\!R^p$, i.e. a poorly informative prior.

Thus, for each $\tilde{\beta}^{(m)} \in B$ we compute the posterior probability as:

$$P(\tilde{\beta} = \tilde{\beta}^{(m)} \mid X_1^n) \quad \propto \quad P(Y_1^n \mid Z_1^n, \tilde{\beta}^{(m)}) P(\tilde{\beta} = \tilde{\beta}^{(m)})$$

by (ii) we also have

$$P(\tilde{\beta} = \tilde{\beta}^{(m)} \mid X_1^n) \quad \propto \quad \prod_{i=1}^{n} \Phi(Z_i' \tilde{\beta}^{(m)})^{y_i} (1 - \Phi(Z_i' \tilde{\beta}^{(m)}))^{1-y_i}$$

The predictive posterior expectation may be evaluated as:

$$\begin{aligned} E(Y_{n+1} \mid X_1^n, Z_{n+1}) &= P(Y_{n+1} = 1 \mid X_1^n, Z_{n+1}) \\ &= E[P(Y_{n+1} = 1 \mid X_1^n, Z_{n+1}, \tilde{\beta}) \mid X_1^n, Z_{n+1}] \end{aligned}$$

by (ii), the posterior predictive expectation can be approximated as:

$$\begin{aligned} \hat{E}(Y_{n+1} \mid X_1^n, Z_{n+1}) &= \sum_{\tilde{\beta}^{(m)} \in B} P(Y_{n+1} = 1 \mid X_1^n, Z_{n+1}, \tilde{\beta}^{(m)}) m(\tilde{\beta}^{(m)} \mid X_1^n) \\ &= \sum_{\tilde{\beta}^{(m)} \in B} \Phi(Z_{n+1}' \tilde{\beta}^{(m)}) m(\tilde{\beta}^{(m)} \mid X_1^n) \qquad (14) \end{aligned}$$

where Φ denotes the c.d.f. of the $\mathcal{N}(0,1)$.

Semi-parametric estimation:
We assume the following prior distribution:

$\psi \sim \mathcal{D}i(n_0, P_0), \qquad P_0 \sim Cauchy \qquad n_0 = 1$
where, we choose P_0 to be a Cauchy distribution in order to represent reasonably poor but proper prior information.

$\beta \sim$ Uniform distribution on a discrete rectangle $B \subset I\!R^p$, to represent a poorly informative prior distribution on the Euclidean parameter.

In this case, the predictive posterior expectation is approximated as follows:

$$\hat{E}(Y_{n+1} \mid X_1^n, Z_{n+1}) =$$

$$\frac{1}{K} \sum_{k=1}^{K} \frac{1}{n + n_0} \left[n_0 P_0 + n \hat{F}_{n, l_1^{n}, (k)} \right] (b_0^{(k)} + b_1^{(k)} Z_{n+1}) \qquad (15)$$

where

$$\hat{F}_{n,l_1^{n,(k)}} = \frac{1}{n}\sum_{j=1}^{n}\delta_{\{l_j^{(k)}\}},$$

and $(l_1^{n\,(k)}, b_0^{(k)}, b_1^{(k)})$, $k = 1, \cdots, K$ are K drawings from $(\lambda_1^n, \beta_0, \beta_1) \mid X_1^n$ provided by the Gibbs sampler of *Method 2* of Section 3.2 .

For each sample X_1^n, generated according to the simulation design, we compare the estimated predictive posterior expectations of the semi-parametric model (15) with those of the parametric model (14) and with the expectation from the actual model, i.e.

$$E(Y_{n+1} \mid Z_{n+1}) = F^*(Z'_{n+1}b^*)$$

For the Euclidean parameter we define, in both parametric and semi-parametric estimation, the set $B = B_1 \times B_2 \subset I\!R^2$, where B_i is a grid of $n_B = 50$ equidistant points on the interval $[-10, 10]$. For the semi-parametric estimation we initially consider 2000 iterations of the Gibbs sampling to reach the stationary phase and 4000 iterations to approximate the predictive posterior expectations. We repeat the above procedure for three different sets of initial values of parameters. If the Gibbs sampling converged in the fixed number of iterations, then the three estimated predictive posterior expectations would have to be similar. Thus, we repeat three times the procedure to test the convergence rather than to improve the approximation. For these experiments we fixed the value of the weight of prior information n_0 equal to 1, independently of the sample size.

All the simulations and estimation methods have been performed with programs written in Splus, except some subroutines written in Fortran language. The figures were also made in Splus.

4.3 Results

Figures 1 to 14 show the results of the numerical examples. Figures 1 to 5 and 10 (figures (a)) represents the estimated predictive posterior expectation from the semi-parametric model computed with the Method 2 (Section 3.2) compared with that of the parametric model and with the expectation from the actual model. For the semi-parametric model, we plotted three estimations of the predictive posterior expectation corresponding to three independent drawings from the Markov process associated to the Gibbs sampler in Method 2. Each figure corresponds to a different sample which is represented in the second plot (figures (b)), where for each value of the grid on Z we plotted the value of Y in the sample.

Given that for each scenario we simulate only one sample of the observable variables, the results clearly depend on the particularities of the generated sample. As to be expected, Figures 1 to 14 show a great variability in this respect; for instance Figures 1 or 2 show quite "typical" sample result whereas Figures 3 or 5 display a much more "atypical" result. This particularity of our experiment allows one to gain some feeling about the data sensivity of the different models under comparison.

These examples also allows one to compare the behaviour of the parametric model with that of the semi-parametric one, in presence of more or less specification error of the parametric model. Indeed we fix ν, the parameter of the Student distribution, equal to 1 to represent more specification error and equal to 10 to represent less specification error. At the same time, when $\nu = 1$ we favor the semi-parametric model, because in this case the link function has the same value as the prior mean of the Dirichlet process in the semi-parametric model.

Figures 1 to 4 suggest that for small sample size the stationary regime of the Gibbs sampler is likely to be approximately obtained after 2000 drawings, given that in each trial the 3 evaluation of the posterior predictive expectations are close together. However, the stationary regime after 2000 drawings is much more questionable once the sample size increases, see e.g. Figure 10 with sample size $n = 10$.

For the statistical convergence of the predictive posterior expectation of the semi-parametric model, these results suggest that even for small sample sizes as $n = 10$ or 20, the closeness to the true (sampling) conditional expectation is reasonably good. The parametric model, display a high degree of sensivity no only to the specification error ($\nu = 1$ or 10 whereas the parametric model corresponds to $\nu \to \infty$) but also to the data structure. For instance, the parametric results in Figures 3 or 5 are really poor.

Note that we fixed the weight of prior information equal to 1, therefore the role of the sample is more important than that of the prior information. But the prior specification clearly smoothes the empirical process which is expected to be rather erratic for small sample sizes. In parametric model, the smoothing of the empirical process is "brutal" at the cost of more sensitivity to the specification error. We observe, in particular, that the semi-parametric estimation is often far from the actual model when a point of the sample falls on a less probable value (following the actual model). See for instance Figure 4 and the sample point $x_i = (y_i, z_i) = (1, 1.5)$. Thus for small n, this suggests that the sample bias explains the discrepancy between the semi-parametric estimation and the actual model.

Figures 6 to 9 and 11 to 14 present the results of the semi-parametric model concerning the Euclidean parameter $\beta = (\beta_0, \beta_1)$ in the case of the samples of size 50 and 100 respectively.

Figures 6,8,11,13 present the estimated posterior probabilities of $\beta = (\beta_0, \beta_1)$ and cumulative probabilities from each one of the three drawings of the Markov process, in the form of histograms. We observe in the two cases (Figures 8 and 13) that the estimated posterior distribution of β_1 is virtually restricted to the interval(0,10]. For β_0 the results presented in Figures 6 and 11 are less clear and one may be suspectious about the convergence of the Gibbs sampler.
Another concerns regards the numerical stability of the Gibbs sampler. Figures 7,9,12 and 14 present the time series of drawing values of β_0 and β_1 in the Gibbs sampling associated to the semi-parametric estimations. For almost all the cases, the time series display a correlation structure that decreases with the lag length. However, the correlation structures are rather different between repetitions of the Gibbs sampling, which can be viewed as an instability of the Gibbs sampling. On the other hand we also notice in another experiment that less autocorrelated time series are obtained for smaller sample sizes, which is coherent with the fact that the proposed Gibbs sampling converges more slowly when the sample size increases.

5 Conclusions

It seems fitting to make explicit some few things to be learned from this work. Two types of issues will be sketched: some general comments, more methodologically oriented and some empirical findings on the particular model under scrutiny. These comments also allude to findings obtained in a more systematic work, of which this one is mostly a summary, namely Scheihing(1995).

At a general level, we made use of two important features of the Bayesian model in the semi-parametric binary response case. The first one is that Bayesian identification is obtained (for a proof and some comments see Mouchart et al(1997)), by fixing P_0 the prior expectation of ψ without restricting the support of the prior distribution to satisfy such conditions as zero mean and unit variances or some other similar conditions (see e.g. Newton at al. (1996)). Previous work (Scheihing (1995)) had shown that such constraints, when unduly incorporated, could possibly be rejected by the data, in the sense of producing unreasonable posterior results. The second one is that, from a non-parametric point of view, the prior distribution on ψ acts as an adaptive smoother of the empirical process. Thus the posterior predictive expectations were almost as smooth as in the parametric models but without being liable to specification error and being less sensitive to atypical data.

In a Bayesian semi-parametric model, the modalities for building a Gibbs sampler are multiple. Discussing how suitable they are crucially depends on a careful analysis of the structural properties of the model: in retrospect, the analysis of Section 2 has been particularly rewarding.

A typical problem of the Bayesian semi-parametric models is the numerical generation of infinite dimensional parameters. The necessity of truncating *each* generation adds to the complexity of the convergence problems. Indeed, several problems of convergence are raised and should be carefully distinguished. For given data, there is a first (standard) problem of the convergence of a numerical algorithm toward the (numerically) exact posterior distribution. In this case, the conditions for convergence bear on two different aspects of the numerical procedure: the size and the structure of the Gibbs sampler and the truncation in the simulation of each trajectory of the infinite dimensional functional parameter. A second issue is the sampling convergence of the posterior distribution (i.e. a convergence pointwise on the parameter space) and a third (and different) issue is the convergence *within* the Bayesian model, i.e. within the unique probability measure on the (doubly infinite dimensional) product space "parameter $\times$ observations". The last two questions are roughly equivalent in a purely parametric set-up (under a condition of correct specification of the sampling process and of regular prior specification) but they become quite different in a semi-parametric setup because of the infinite dimensionality of both the parameter and the sample space.

In this paper, we addressed, among others, the question of obtaining the stationary regime of the Markov chain associated to the Gibbs samplers after 2000 drawings. Although a careful analysis of this issue is out of the scope of this paper, we nevertheless checked that when simulating three such trajectories they gave reasonably close results and we also controlled the autocorrelation of the time series of the simulations. Vastly different results would lead to question whether the stationary regime has been actually obtained (see for references Gelfand and Smith (1990), Tierney(1994) , and Athreya et al (1996)).

Let us consider some factual findings. A more complete analysis (see Scheihing(1995)) suggested that the two methods of simulation introduced in Section 3 were roughly equivalent as far as the convergence of the Gibbs sampler toward its stationary regime. Thus in this paper we report results on the second method only. One interesting feature of the semi-parametric model is that the convergence of the Gibbs sampler towards its stationary regime happens to be slower as the sample size increase. This is opposite to the sampling convergence, i.e., for larger sample size, the exact finite sample results will be closer to the asymptotic result but require more initial iterations of the Gibbs sampler.

When comparing the semi-parametric model with the parametric models specified with various degrees of specification error it was to be noticed that even for sample size as small as $n = 10$ or $n = 20$, the semi-parametric model gave quite sensible results whereas the parametric model not only was quite sensitive to the size of the specification error (as was to be expected) but was also sensitive to the atypical data.

Finally, the hypothesis of prior independence between the Euclidean and the functional parameter is usually justified by reasons of convenience only. Its interpretation may however raise problems. But the computational cost of recognizing a prior dependence between the functional and the Euclidean parameter may be formidable.

A Appendix

A.1 Basic concepts on conditional and mutual independence

For the sake of easy reference, we summarize some definitions and properties of conditional independence extracted from Florens et al(1990), henceforth to be refered to as FMR. These properties are the basic tools in the following demonstrations. Let $(M, \mathcal{M}, P)$ be an abstract probability space, $\mathcal{M}_i, i = 1, \cdots, 5$ be sub-σ-fields of $\mathcal{M}$.

A.1.1 (Theorem 2.2.10 in FMR):

The following properties are equivalent:

(i) $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$ and $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_4 \mid \mathcal{M}_2 \vee \mathcal{M}_3$,

(ii) $\mathcal{M}_1 \perp\!\!\!\perp (\mathcal{M}_2 \vee \mathcal{M}_4) \mid \mathcal{M}_3$,

(iii) $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_4 \mid \mathcal{M}_3$ and $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_4 \vee \mathcal{M}_3$,

A.1.2 (Corollary 2.2.11 in FMR):

If $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$, $\mathcal{M}_5 \subset \mathcal{M}_1 \vee \mathcal{M}_3$, $M_4 \subset \mathcal{M}_2 \vee \mathcal{M}_3$ then:

(i) $(\mathcal{M}_1 \vee \mathcal{M}_5) \perp\!\!\!\perp (\mathcal{M}_2 \vee \mathcal{M}_4) \mid \mathcal{M}_3$,

(ii) $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3 \vee \mathcal{M}_4 \vee \mathcal{M}_5$

A.1.3 (Definition 5.4.16 in Florens et al(1990)):

Let $\mathcal{M}_i \, (i = 1, 2, 3)$ be three sub-σ-fields of $\mathcal{M}$.

(i) $\mathcal{M}_3$ splits $(\mathcal{M}_1, \mathcal{M}_2)$ if $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$;

(ii) $\mathcal{M}_3$ minimally splits $(\mathcal{M}_1, \mathcal{M}_2)$ if

- (a) $\mathcal{M}_3$ splits $(\mathcal{M}_1, \mathcal{M}_2)$
- (b) $\mathcal{M}_4 \subset \mathcal{M}_3$ splits $(\mathcal{M}_1, \mathcal{M}_2) \Rightarrow \overline{\mathcal{M}_3} = \overline{\mathcal{M}_4}$

(*iii*) $\mathcal{M}_3$ minimally and internally splits $(\mathcal{M}_1, \mathcal{M}_2)$ if

(*a*) $\mathcal{M}_3$ splits $(\mathcal{M}_1, \mathcal{M}_2)$

(*b*) $\mathcal{M}_3 \subset \mathcal{M}_1 \vee \mathcal{M}_2$

A.1.4 (Definition 7.6.1 in FMR):

Let I be an arbitrary index set. The family of sub-σ-fields of $\mathcal{M}$, $\{\mathcal{F}_i : i \in I\}$ is a family of *mutually conditionally independent* σ-fields, where the conditioning is with respect to a sub-σ-field $\mathcal{M}_1$ of $\mathcal{M}$ and we write

$$\perp\!\!\!\perp_{i \in I} \mathcal{F}_i \mid \mathcal{M}_1$$

if, for every finite subset J of I and for every $f_i \in [\mathcal{F}_i]^+$, $\forall i \in J$,

$$\mathcal{M}_1(\prod_{i \in J} f_i) = \prod_{i \in J} \mathcal{M}_1(f_i)$$

A.1.5 (Theorem 7.6.3 in FMR):

Let $\{\mathcal{F}_n : n \in I\!N\}$ be a sequence of sub-σ-fields of $\mathcal{M}$. The following conditions are equivalent:

(*i*) $\perp\!\!\!\perp_{n \in I\!N} \mathcal{F}_n \mid \mathcal{M}_1$;

(*ii*) $\perp\!\!\!\perp_{0 \leq k \leq n} \mathcal{F}_k \mid \mathcal{M}_1, \forall n \in I\!N$;

(*iii*) $\mathcal{F}_0^n \perp\!\!\!\perp \mathcal{F}_{n+1}^\infty \mid \mathcal{M}_1, \forall n \in I\!N$;

(*iv*) $\mathcal{F}_n \perp\!\!\!\perp \mathcal{F}_{]n[} \mid \mathcal{M}_1, \forall n \in I\!N$, where $\mathcal{F}_{]n[} = \mathcal{F}_{I\!N - \{n\}}$;

(*v*) $\mathcal{F}_{n+1} \perp\!\!\!\perp \mathcal{F}_0^n \mid \mathcal{M}_1, \forall n \in I\!N$.

A.1.6 (Theorem 7.6.10 and definition 7.6.11 in FMR):

Let $\mathcal{M}_1$ be a sub-σ-field of $\mathcal{M}$ and let $(\mathcal{F}_n)$ and $(\mathcal{G}_n)$ be two sequences of sub-σ-fields of $\mathcal{M}$. Then the following three properties are equivalent and define that $(\mathcal{F}_n)$ is $\mathcal{M}_1$-sifted by $(\mathcal{G}_n)$.

I (i) $\perp\!\!\!\perp_{n\in\mathbb{N}}\mathcal{F}_n \mid \mathcal{M}_1 \vee \mathcal{G}_0^\infty$;

(ii) $\mathcal{F}_n \perp\!\!\!\perp \mathcal{G}_0^\infty \mid \mathcal{M}_1 \vee \mathcal{G}_n, \forall n \in \mathbb{N}$.

II (iii) $\perp\!\!\!\perp_{0\leq k\leq n}\mathcal{F}_k \mid \mathcal{M}_1 \vee \mathcal{G}_0^n, \forall n \in \mathbb{N}$;

(iv) $\mathcal{F}_k \perp\!\!\!\perp \mathcal{G}_0^n \mid \mathcal{M}_1 \vee \mathcal{G}_k, \forall n, k \in \mathbb{N}, k \leq n$.

III (v) $\mathcal{F}_0^n \perp\!\!\!\perp \mathcal{G}_{n+1} \mid \mathcal{M}_1 \vee \mathcal{G}_0^n, \forall n \in \mathbb{N}$;

(vi) $\mathcal{F}_{n+1} \perp\!\!\!\perp \mathcal{G}_0^n \mid \mathcal{M}_1 \vee \mathcal{G}_{n+1}, \forall n \in \mathbb{N}$;

(vii) $\mathcal{F}_{n+1} \perp\!\!\!\perp \mathcal{F}_0^n \mid \mathcal{M}_1 \vee \mathcal{G}_0^{n+1}, \forall n \in \mathbb{N}$.

A.2 Representation of trajectories of a Dirichlet process

In this section we summarize a representation of the Dirichlet process developed by Rolin (1993) that allows to generate Monte Carlo samples of the Dirichlet random measure. We shall consider both prior and posterior Dirichlet processes.

A.2.1 The prior Dirichlet process

Let $P \sim \mathcal{D}\rangle(n_0, P_0)$, $n_0 > 0$, P_0 diffuse. Then P admits the a.s. representation:

$$P = \sum_{k=1}^{\infty} \alpha_k \delta_{\xi_k} \qquad \text{a.s.} \tag{16}$$

where

- $(\alpha_k)_{k\in\mathbb{N}} \perp\!\!\!\perp (\xi_k)_{k\in\mathbb{N}}$
- $(\xi_k)_{k\in\mathbb{N}} \sim \ iid\ P_0$
- the distribution of α_k may be described by the following representation:

$$\alpha_k = v_k \prod_{1 \leq h \leq k-1} (1 - v_h)$$

where $(v_h)_{h \in \mathbb{N}} \sim iid\ Beta(1, n_0)$

- δ_{ξ_k} is the unit mass at ξ_k.

For simulation purposes we shall consider the following approximation of (16):

$$\bar{P}^K = \frac{1}{\sum_{k=1}^K \alpha_k} \sum_{k=1}^K \alpha_k \delta_{\xi_k}, \qquad for\ some\ K \in \mathbb{N}, \tag{17}$$

which satisfies

$$\varepsilon_K = \sup_{B \in \mathcal{B}_p} |P(B) - \bar{P}^K(B)| = \sum_{k > K} \alpha_k. \tag{18}$$

The moments of ε_K are thus given by

$$E(\varepsilon_K^p) = \left(\frac{n_0}{n_0 + p} \right)^K. \tag{19}$$

Therefore, as $K \to \infty$, $\varepsilon_K \to 0$ a.s. and in L^p, $\forall p \in (0, \infty)$.

A.2.2 The posterior Dirichlet process

Let $P \sim \mathcal{D}i(n_0, P_0)$, $n_0 > 0$, P_0 diffuse and $x = \{x_1, \cdots, x_n\}$ with $(x_i \mid P) \sim\ iid\ P$, then a posteriori P is also distributed as a Dirichlet process:

$$(P \mid x) \sim \mathcal{D}i(n_*, P_*)$$

where

$$\begin{aligned} n_* &= n_0 + n \\ P_* &= \frac{n_0}{n_*} P_0 + \frac{n}{n_*} \hat{F}_n(x) \end{aligned}$$

and

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n \delta_{x_j} = \frac{1}{n} \sum_{j=1}^m n_j \delta_{\tilde{x}_j}$$

where $\{\tilde{x}_1, \cdots, \tilde{x}_m\}$ is the set of distinct values of the x and n_j the number of occurrences of $\tilde{x}_j$.
The posterior Dirichlet process admits the a.s. representation :

$$P = (1-\gamma)P_c + \gamma P_d \tag{20}$$

(P_c and P_d are the diffuse ("continuous") and the discrete part of P respectively) where

$$\begin{array}{rcl} \gamma \perp\!\!\!\perp & P_c & \perp\!\!\!\perp P_d \\ \gamma & \sim & Beta(n, n_0) \\ P_c & \sim & \mathcal{D}i(n_0, P_0) \\ P_d & \sim & \mathcal{D}i(n, \hat{F}_n(x)) \end{array}$$

where P_c may be represented as in (16) and P_d may be represented as

$$P_d = \sum_{j=1}^{m} \beta_j \delta_{\tilde{x}_j} \qquad \text{a.s.} \tag{21}$$

where

$$(\beta_1, \cdots, \beta_m) \sim Di(n_1, \cdots, n_m)$$

For simulation purposes, the approximation for P given in the previous section, can be used to approximate the posterior process, through the following expression:

$$\bar{P}^K = (1-\gamma)\bar{P}_c^K + \gamma P_d$$

The posterior approximation errors are equal to the prior approximation errors multiplied by $(1-\gamma)$, implying that the quality of the posterior approximation is better than the quality of the prior approximation.

For a detailed exposition of these results see Rolin (1992, 1993) and Florens and Rolin (1994).

A.3 Proof of Properties

Preliminary remarks: When there is no danger of confusion, we freely identify a random variable and the $\sigma-$field generated by the random variable. For instance, we write $Y_i \subset \sigma(\lambda_i, \eta_i)$ instead of $\sigma(Y_i) \subset \sigma(\lambda_i, \eta_i)$. In general, implications (denoted by $\Rightarrow$) without arguments are simple consequences of general properties of conditional independence as recalled in A.1.1 and A.1.2.

Proof Property 2.1:

$$\left.\begin{array}{rcl}(6) & \Rightarrow & (Z_1^n,\beta) \perp\!\!\!\perp \psi \\ & & \eta_1^n \subset \sigma(Z_1^n,\beta)\end{array}\right\} \Rightarrow \quad (Z_1^n,\beta) \perp\!\!\!\perp \psi \mid \eta_1^n \tag{22}$$

$$(1) \Rightarrow (Z_1^n,\theta) \perp\!\!\!\perp Y_1^n \mid \psi, \eta_1^n \quad \Rightarrow \quad (Z_1^n,\beta) \perp\!\!\!\perp Y_1^n \mid \psi, \eta_1^n \tag{23}$$

$$Then\ (22)\ and\ (23) \quad \Leftrightarrow \quad (Z_1^n,\beta) \perp\!\!\!\perp (\psi, Y_1^n) \mid \eta_1^n$$

Proof Property 2.2:

$$\left.\begin{array}{rcl}(3) & \Rightarrow & \lambda_1^n \perp\!\!\!\perp (Z_1^n,\beta) \mid \psi \\ (6) & \Rightarrow & \psi \perp\!\!\!\perp (Z_1^n,\beta)\end{array}\right\} \Leftrightarrow \quad (Z_1^n,\beta) \perp\!\!\!\perp (\psi, \lambda_1^n)$$

$$\Rightarrow \quad (Z_1^n,\beta) \perp\!\!\!\perp \psi \mid \lambda_1^n \tag{24}$$

Furthermore $Y_i \subset \sigma(Z_1^n, \beta, \lambda_1^n)$

$$\Rightarrow \quad Y_1^n \perp\!\!\!\perp \psi \mid Z_1^n, \beta, \lambda_1^n \tag{25}$$

$$Then\,(24)+(25) \quad \Leftrightarrow \quad \psi \perp\!\!\!\perp (X_1^n,\beta) \mid \lambda_1^n$$

Finally, from Bayes theorem:

$$\begin{array}{rcl} m(\beta \mid X_1^n, \lambda_1^n) & \propto & P(Y_1^n \mid Z_1^n, \lambda_1^n, \beta) m(\beta \mid Z_1^n, \lambda_1^n) \\ (1)+(3) \Rightarrow & \propto & P(Y_1^n \mid Z_1^n, \lambda_1^n, \beta) q_0(\beta) \\ & \propto & q_0(\beta) \mathbb{1}_{\{\beta \mid \forall i, Y_i = 1 \Leftrightarrow \lambda_i \leq Z_i' \beta\}} \end{array}$$

Proof Proposition 2.3:
We have from property (A.1.5) above that (a) is equivalent to

$$\lambda_i \perp\!\!\!\perp (\lambda_j, j \neq i) \mid \eta_1^n, Y_1^n, \psi \ \ \forall i = 1, \cdots, n \tag{26}$$

From (3) we have

$$\lambda_1^n \perp\!\!\!\perp \eta_1^n \mid \psi$$

and therefore

$$\lambda_i \perp\!\!\!\perp \eta_1^n \mid \psi, (\lambda_j, j \neq i) \tag{27}$$

Furthermore, (4) implies

$$\underline{\|}_{1\le i\le n}\lambda_i \mid \psi$$

and therefore

$$\lambda_i \underline{\|} (\lambda_j, j\neq i) \mid \psi \tag{28}$$

Thus (27) + (28) is equivalent to:

$$\lambda_i \underline{\|} \eta_1^n, (\lambda_j, j\neq i) \mid \psi$$

which implies:

$$\lambda_i \underline{\|} (\lambda_j, j\neq i) \mid \psi, \eta_1^n \tag{29}$$

From the definition of the latent variables we also have

$$\left.\begin{array}{ccc} Y_i & \subset & \sigma(\lambda_i,\eta_i) \\ Y_j, j\neq i & \subset & \sigma(\lambda_j,\eta_j, j\neq i) \end{array}\right\} \tag{30}$$

From Property (A.1.2) above and (29) + (30) we have

$$(\lambda_i, Y_i) \quad \underline{\|} \quad (\lambda_j, j\neq i), (Y_j, j\neq i) \mid \psi, \eta_1^n \tag{31}$$

Clearly (31) implies (26): This demonstrates part (a) of the property. On the other hand, from (31) we also have

$$\lambda_i \underline{\|} (Y_j, j\neq i) \mid \psi, \eta_1^n, Y_i \tag{32}$$

Furthermore, from (3)

$$\lambda_i \underline{\|} \eta_1^n \mid \psi \quad \Rightarrow \quad \lambda_i \underline{\|} (\eta_j, j\neq i) \mid \psi, \eta_i$$

which implies

$$\lambda_i \underline{\|} (\eta_j, j\neq i) \mid \psi, \eta_i, Y_i \tag{33}$$

because $Y_i \subset \sigma(\lambda_i, \eta_i)$. Note that (32) + (33) is equivalent to

$$\lambda_i \underline{\|} (\eta_1^n, Y_1^n) \mid \psi, \eta_i, Y_i$$

which demonstrates part (b) of the Property.

(c) From Bayes theorem we have:

$$
\begin{aligned}
P(\lambda_i \leq l \mid Y_i, \psi, \eta_i) &\propto P(\lambda_i \leq l, Y_i = y_i \mid \psi, \eta_i) \\
&\propto \begin{cases} P(\lambda_i \leq l, \lambda_i \leq \eta_i \mid \psi, \eta_i) & if \quad Y_i = 1 \\ \\ P(\lambda_i \leq l, \lambda_i > \eta_i \mid \psi, \eta_i) & if \quad Y_i = 0 \end{cases} \\
&\propto \begin{cases} P(\lambda_i \leq \min(l, \eta_i) \mid \psi, \eta_i) & if \quad Y_i = 1 \\ \\ P(\eta_i < \lambda_i \leq l \mid \psi, \eta_i) & if \quad Y_i = 0 \end{cases}
\end{aligned}
$$

From (3) we conclude:

$$
P(\lambda_i \leq l \mid Y_i, \psi, \eta_i) \propto \begin{cases} \psi(l) & if \quad Y_i = 1, l \leq \eta_i \\ \psi(\eta_i) & if \quad Y_i = 1, l > \eta_i \\ 0 & if \quad Y_i = 0, l \leq \eta_i \\ \psi(l) - \psi(\eta_i) & if \quad Y_i = 0, l > \eta_i \end{cases}
$$

Note that, obviously enough, $\psi(\eta_i) = P(Y_i = 1 \mid \eta_i, \psi) \in]0, 1[$ once the support of ψ is all the real line. Thus,

$$
P(\lambda_i \leq l \mid Y_i, \eta_i, \psi) = \begin{cases} \frac{\psi(l)}{\psi(\eta_i)} & if \quad Y_i = 1, l \leq \eta_i \\ 1 & if \quad Y_i = 1, l > \eta_i \\ 0 & if \quad Y_i = 0, l \leq \eta_i \\ \frac{\psi(l) - \psi(\eta_i)}{1 - \psi(\eta_i)} & if \quad Y_i = 0, l > \eta_i \end{cases}
$$

Proof Proposition 2.4:
(a) Because clearly $\eta_1^n \subset \sigma(Z_1^n, \beta)$, (3) implies

$$
\eta_1^n \perp\!\!\!\perp \lambda_1^n \mid \psi \tag{34}
$$

and (6) implies:

$$
\eta_1^n \perp\!\!\!\perp \psi \tag{35}
$$

Now, (34) and (35) are equivalent to

$$
\eta_1^n \perp\!\!\!\perp (\psi, \lambda_1^n)
$$

which implies:

$$
\lambda_i \perp\!\!\!\perp (\eta_j, j \neq i) \mid (\lambda_j, j \neq i), \eta_i
$$

and, because $Y_i \subset \sigma(\lambda_i, \eta_i)\ \forall i$, this succesively implies:

$$\begin{aligned}(\lambda_i, Y_i) &\perp\!\!\!\perp (\eta_j, Y_j, j \neq i) \mid (\lambda_j, j \neq i), \eta_i \\ \lambda_i &\perp\!\!\!\perp (\eta_j, Y_j, j \neq i) \mid (\lambda_j, j \neq i), \eta_i, Y_i\end{aligned}$$

and this is equivalent to (a).

(b) We need to demonstrate:

$$P(\lambda_i \leq l \mid (\lambda_j, j \neq i), Y_i, \eta_i) \propto \alpha_{i,1} F_{i,1}((-\infty, l]) + \alpha_{i,2} F_{i,2}((-\infty, l])$$

In fact

$$\begin{aligned}P(\lambda_i \leq l \mid (\lambda_j, j \neq i), Y_i, \eta_i) &\propto P(\lambda_i \leq l, Y_i \mid (\lambda_j, j \neq i), \eta_i) \\ &\propto \begin{cases} P(\lambda_i \leq l, \lambda_i \leq \eta_i \mid (\lambda_j, j \neq i), \eta_i) & if \quad Y_i = 1 \\ \\ P(\lambda_i \leq l, \lambda_i > \eta_i \mid (\lambda_j, j \neq i), \eta_i) & if \quad Y_i = 0 \end{cases} \\ &\propto \begin{cases} P(\lambda_i \leq \min(l, \eta_i) \mid (\lambda_j, j \neq i), \eta_i) & if \quad Y_i = 1 \\ \\ P(\eta_i < \lambda_i \leq l \mid (\lambda_j, j \neq i), \eta_i) & if \quad Y_i = 0 \end{cases}\end{aligned}$$

On the other hand, from the Polya urn representation of the Dirichlet process (Blackwell and MacQueen (1973)), we have:

$$P(\lambda_i \in B \mid (\lambda_j, j \neq i)) \propto \sum_{j \neq i} \mathbb{1}_{\{\lambda_j \in B\}} + n_0 P_0(B)$$

thus
$P(\lambda_i \leq l \mid (\lambda_j, j \neq i), Y_i, \eta_i) \propto$

$$\begin{cases} \sum_{j \neq i} \mathbb{1}_{\{\lambda_j \leq l, \lambda_j \leq \eta_i\}} + n_0 P_0((-\infty, \min(l, \eta_i)]) & if \quad Y_i = 1 \\ \\ \sum_{j \neq i} \mathbb{1}_{\{\lambda_j \leq l, \lambda_j > \eta_i\}} + n_0 P_0((\eta_i, l]) & if \quad Y_i = 0 \end{cases}$$

As $\mathbb{1}_{AB} = \mathbb{1}_A \mathbb{1}_B$ we get

$$q_{ij} \mathbb{1}_{\{\lambda_j \leq l\}} = \begin{cases} \mathbb{1}_{\{\lambda_j \leq l, \lambda_j \leq \eta_i\}} & if \quad Y_i = 1 \\ \mathbb{1}_{\{\lambda_j \leq l, \lambda_j > \eta_i\}} & if \quad Y_i = 0 \end{cases}$$

where

$$q_{ij} = \begin{cases} \mathbb{1}_{\{\lambda_j \leq \eta_i\}} & if \quad Y_i = 1 \\ \mathbb{1}_{\{\lambda_j > \eta_i\}} & if \quad Y_i = 0 \end{cases}$$

Therefore
$P(\lambda_i \leq l \mid (\lambda_j, j \neq i), Y_i, \eta_i) \propto$

$$\begin{cases} \alpha_{i,1} F_{i,1}((-\infty, l]) + n_0 P_0(\eta_i) F_{i,2}((-\infty, l]) & if \quad Y_i = 1 \\ \\ \alpha_{i,1} F_{i,1}((-\infty, l]) + n_0 (1 - P_0(\eta_i)) F_{i,2}((-\infty, l]) & if \quad Y_i = 0 \end{cases}$$

Thus, the results is obtained.

Proof of Property 2.5:

(a) Let us denote $y_{(i)}$ for the observation of Y corresponding to $\eta_{(i)}$. We note that:

$$\begin{aligned} P(Y_1^n = y_1^n \mid \eta_1^n, \psi) &= \prod_{i=1}^{n} \psi(\eta_{(i)})^{y_{(i)}} (1 - \psi(\eta_{(i)}))^{1-y_{(i)}} \\ &= \prod_{i=1}^{n} \left(\sum_{j=1}^{i} \psi(B_j) \right)^{y_{(i)}} \left(1 - \sum_{j=1}^{i} \psi(B_j) \right)^{1-y_{(i)}} \end{aligned}$$

which implies

$$Y_1^n \perp\!\!\!\perp \eta_1^n, \psi \mid \psi(B_1), \cdots, \psi(B_{n+1})$$

(b) From Rolin (1992), Corollary (5.5)(i) we have that

$$\{\psi^{B_j} : 1 \leq j \leq n+1\} \perp\!\!\!\perp \{\psi(B_j) : 1 \leq j \leq n+1\} \tag{36}$$

Thus (36) and (a) imply (b).

Proof of Property 2.6:
From Property 2.4 we have that:

$P(\lambda_i \leq l \mid \lambda_j, j \neq i, Y_i, \eta_i)$

$$\propto \begin{cases} \sum_{j \neq i} \mathbb{1}_{\{\lambda_j \leq \eta_i\}} \mathbb{1}_{\{\lambda_j \leq l\}} + n_0 P_0(l) & if \quad Y_i = 1, l \leq \eta_i \\ \\ \sum_{j \neq i} \mathbb{1}_{\{\lambda_j \leq \eta_i\}} + n_0 & if \quad Y_i = 1, l > \eta_i \\ \\ 0 & if \quad Y_i = 0, l \leq \eta_i \\ \\ \sum_{j \neq i} \mathbb{1}_{\{\eta_i < \lambda_j \leq l\}} + n_0 (P_0(l) - P_0(\eta_i)) & if \quad Y_i = 0, l > \eta_i \end{cases}$$

Therefore, the support D_i of the distribution of λ_i conditionally to $\lambda_j, j \neq i$, Y_i and η_i verifies:

$$D_i = \begin{cases} (-\infty, \eta_i] & if \quad Y_i = 1 \\ \\ (\eta_i, \infty) & if \quad Y_i = 0 \end{cases}$$

which does not depend of the values of $\lambda_j, j \neq i$ and this is for all $i = 1, \cdots, n$, from which we obtain the result.

Proof Property 2.7:
Note that:

$$\begin{aligned} E(Y_0 \mid Z_0, X_1^n) &= E(E(Y_0 \mid Z_0, X_1^n, \beta, \psi) \mid Z_0, X_1^n) \\ Y_0 \perp\!\!\!\perp X_1^n \mid \psi, \beta, Z_0 \Rightarrow &= E(E(Y_0 \mid Z_0, \beta, \psi) \mid Z_0, X_1^n) \\ &= E(\psi(Z_0'\beta) \mid Z_0, X_1^n) \end{aligned}$$

Thus, from the Doob's Martingale Convergence Theorem (see for instance, Chow and Teicher(1988), Section 7.4 Corollary 1), we have that

$$\lim_{n \to \infty} E(\psi(Z_0'\beta) \mid Z_0, X_1^n) = E(\psi(Z_0'\beta) \mid Z_0, X_1^\infty) \qquad a.s.$$

References

Antoniak, C.E.(1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2, 1152-1174.

Athreya, K.B., H. Doss and J. Sethuraman (1996) On the Convergence of the Markov Chain Simulation Method. *Annals of Statistics*, 24, 69-100.

Blackwell, D. and J.B. MacQueen (1973) Ferguson distribution via Polya urn scheme. *Annals of Statistics*, 1, 353-355.

Chow, Y.S. and H. Teicher (1988). *Probability Theory. Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

Cosslett, S.R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, 51, 765-782.

Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of American Statistical Association*, 89, 268-277.

Ferguson, T.S. (1973). A Bayesian analysis of some non-parametric problems. *Annals of Statistics*, 1, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics*, 2, 615-629.

Fisher R.A. (1935). *The Design of Experiments* (8th ed. 1966). Oliver and Boyd, Edinburgh.

Florens, J.P. and J.M. Rolin (1994). Bayes, bootstrap, moments. Institut de Statistique DP9413, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Florens, J.-P., Mouchart, M., and J.-M. Rolin (1990). *Elements of Bayesian Statistics.* Marcel Dekker, inc., New York.

Florens, J.-P., Mouchart, M., and J.-M. Rolin (1992). Bayesian analysis of mixtures: some results on exact estimability and identification. *Bayesian Statistics 4*. Proceedings of the Fourth Valencia International Meeting, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M Smith. Oxford Univeristy Press, 127-145.

Geman,D. and S. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *I.E.E.E. transactions on pattern analysis and machine intelligence*, 6, 721-741.

Horowitz, J (1991). A smoothed Maximum score estimator for binary response model. *Econometrica*, 60, 505-531.

Manski, C.F. (1985). Semi-parametric analysis of discrete response: asymptotic properties of the Maximum score estimator. *Journal of Econometrics*, 27, 313-333.

Matzkin, R. (1991). Semi-parametric estimation of monotonic and concave utility functions: The discrete choice case. *Econometrica*, 59, 1319-1327.

McFadden, D. (1984). Econometric Analysis of Qualitative Response Models. *Handbook of Econometrics, Vol. II, Z. Griliches and M.D. Intriligator ed.*, Amsterdam, North-Holland, 1395-1457.

Mouchart M., J.-M. Rolin and E. Scheihing (1997). Bayesian Identification of Semi-Parametric Binary Response Models. Institut de Statistique DP9713, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Newton, M. (1994). Computing with priors that support identifiable semiparametric models. Technical Report 924 (revised), Department of Statistics, University of Wisconsin, Madison.

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

Roberts, G.O. and A.F.M Smith (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms.

Stochastic Processes and Applications, 49, 207-216.

Rolin, J.M. (1992). Some useful properties of the Dirichlet process. Institut de Statistique DP9202, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Rolin, J.M. (1993). On the distribution of jumps of the Dirichlet process. Institut de Statistique DP9302, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Scheihing E. (1995), Some problems on the Bayesian analysis of discrete data, Doctoral thesis in Statistics. Institut de Statistique, Université Catholique de Louvain, Belgium.

Sethuraman, J. (1994), A constructive definition of Dirichlet Priors. *Statistica Sinica*, 4, 639-650.

Smith A.F. and A.E. Gelfand (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Smith A.F. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B*, 55, 3-24.

Tanner M.A. (1993). *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions.* Springer Series in Statistics, New York.

Tierney L. (1991). Exploring posterior distributions using Markov chains. In Keramidas, E.M. ed., *Computer Science and Statistics: 23rd Symposium on the Interface.* 563-570.

Tierney L. (1994). Markov chain for exploring posterior distributions. *The Annals of Statistics*, 22, 1701-1728 (with discussion).

Wilks, S.S (1962) *Mathematical Statistics.* Wiley, New York.

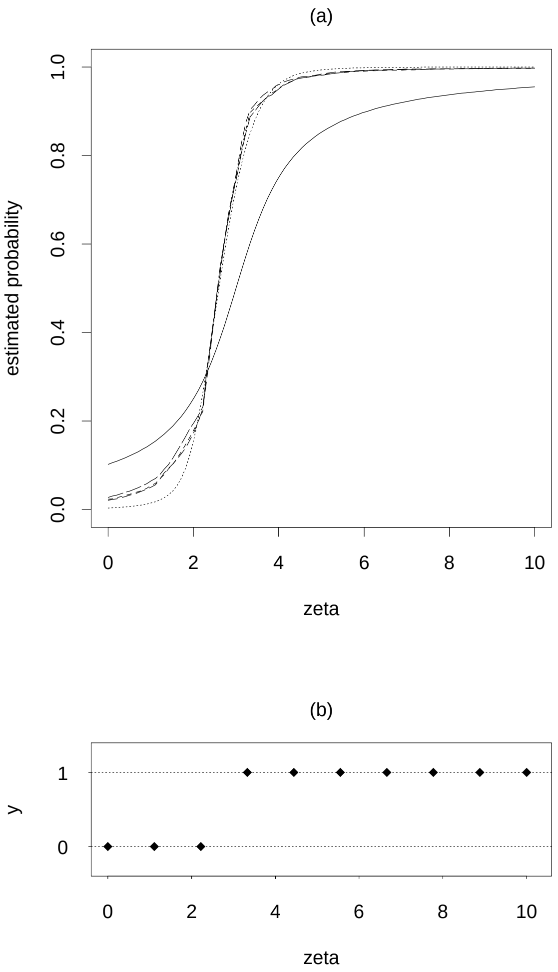

Figure 1: Predictive posterior expectations ($n = 10, \nu = 1$)

Predictive posterior expectations computed with *Method 2* (figure (a)) for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

.... $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K} \sum E(\psi(Z) \mid \lambda_1^{n\,(k)})(\beta^{(k)'} Z)$, semi-parametric model (3 drawings).

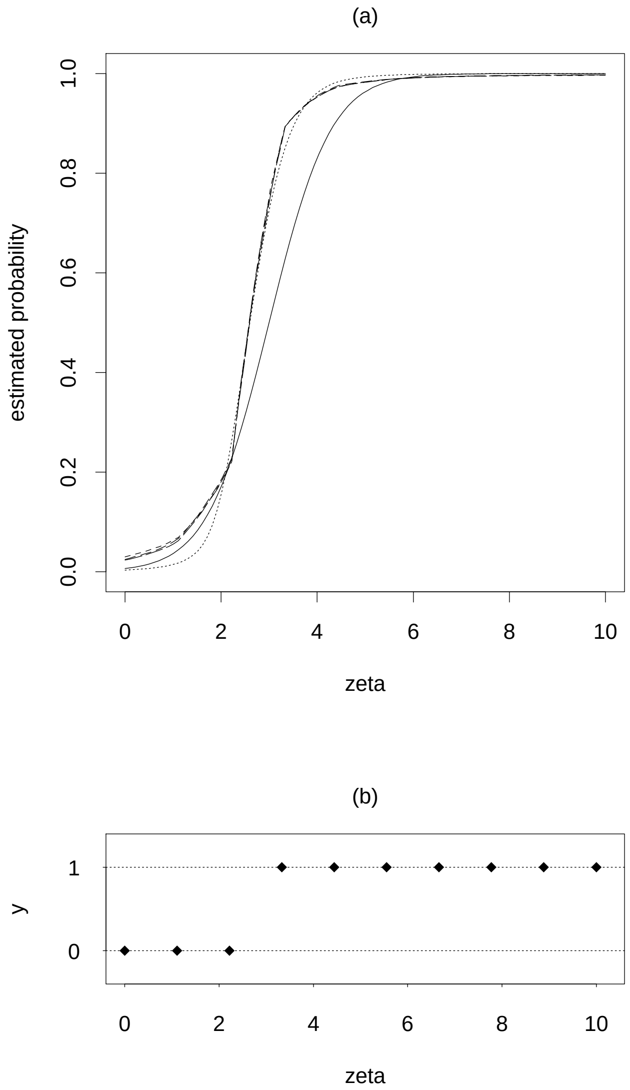

Figure 2: Predictive posterior expectations ($n = 10, \nu = 10$)

Predictive posterior expectations computed with *Method 2* (figure (a)) for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

……… $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K}\sum E(\psi(Z) \mid \lambda_1^{n\,(k)})(\beta^{(k)'}Z)$, semi-parametric model (3 drawings).

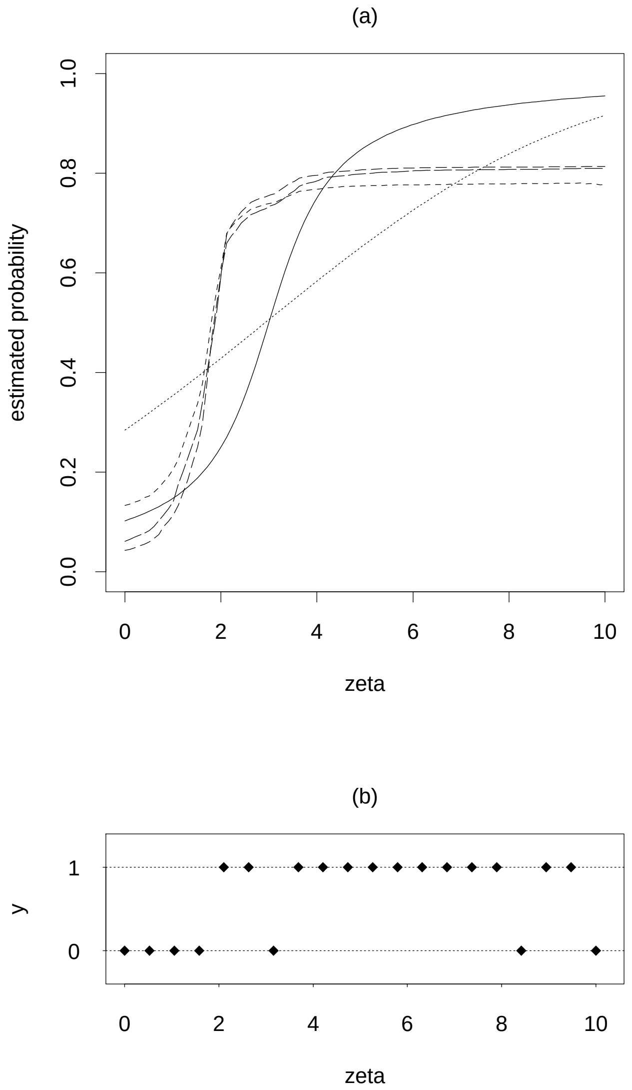

Figure 3: Predictive posterior expectations ($n = 20, \nu = 1$)

Predictive posterior expectations computed with *Method 2* (figure (a)), for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

……… $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K} \sum E(\psi(Z) \mid \lambda_1^{n\,(k)})(\beta^{(k)'} Z)$, semi-parametric model (3 drawings).

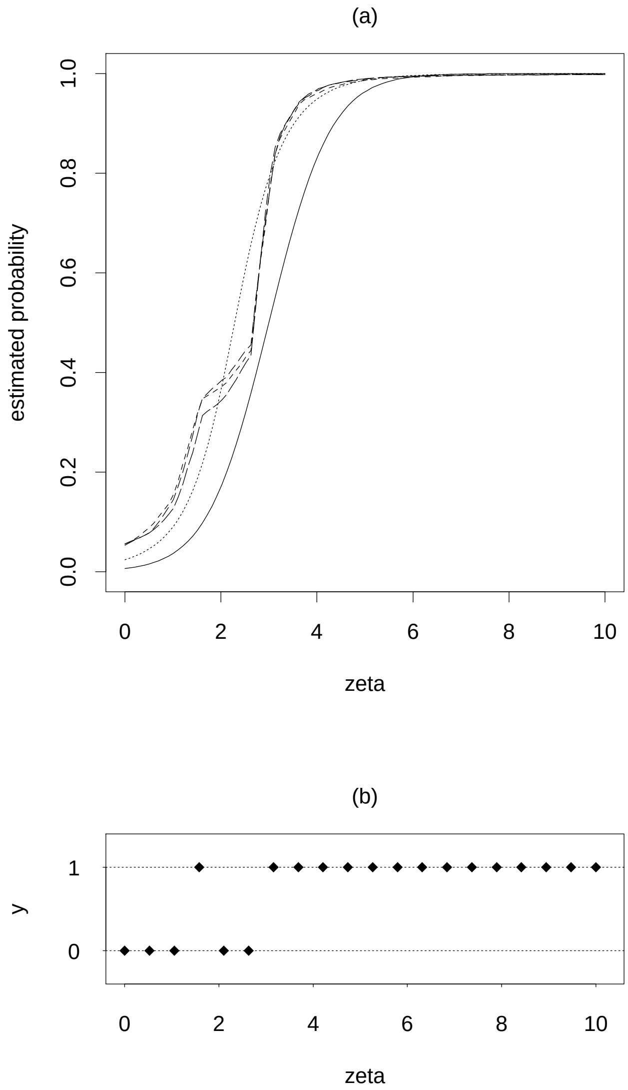

Figure 4: Predictive posterior expectations ($n = 20, \nu = 10$)

Predictive posterior expectations computed with *Method 2* (figure (a)), for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

……… $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K} \sum E(\psi(Z) \mid \lambda 1^{n\,(k)})(\beta^{(k)'} Z)$, semi-parametric model(3 drawings).

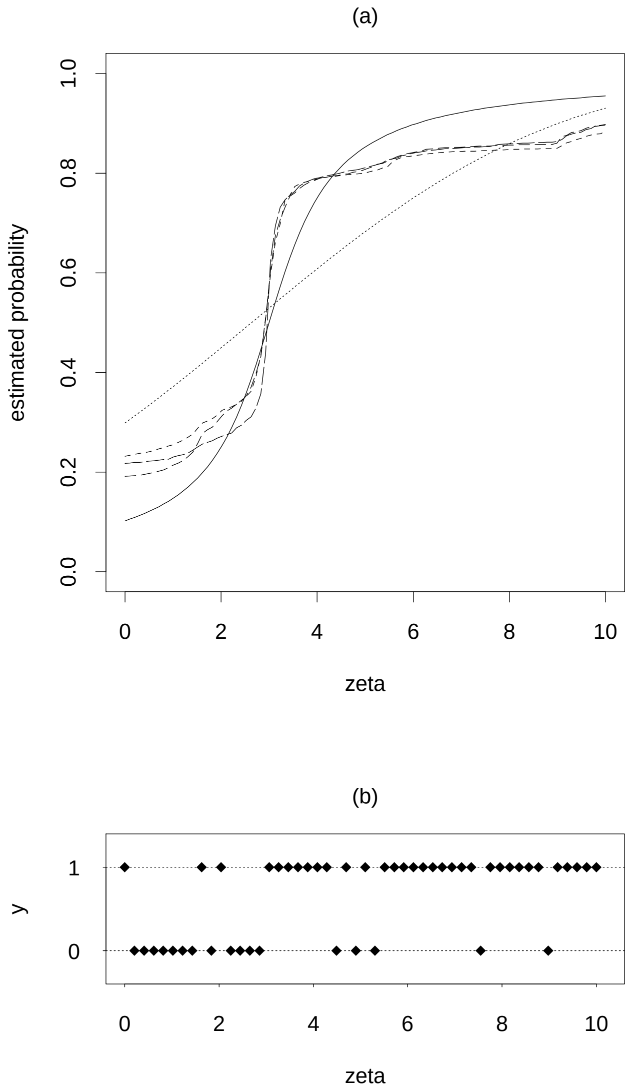

Figure 5: Predictive posterior expectations ($n = 50, \nu = 1$)

Predictive posterior expectations computed with *Method 2* (figure (a)) for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

.... $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K}\sum E(\psi(Z) \mid \lambda_1^{n\,(k)})(\beta^{(k)'}Z$, semi-parametric model(3 drawings).

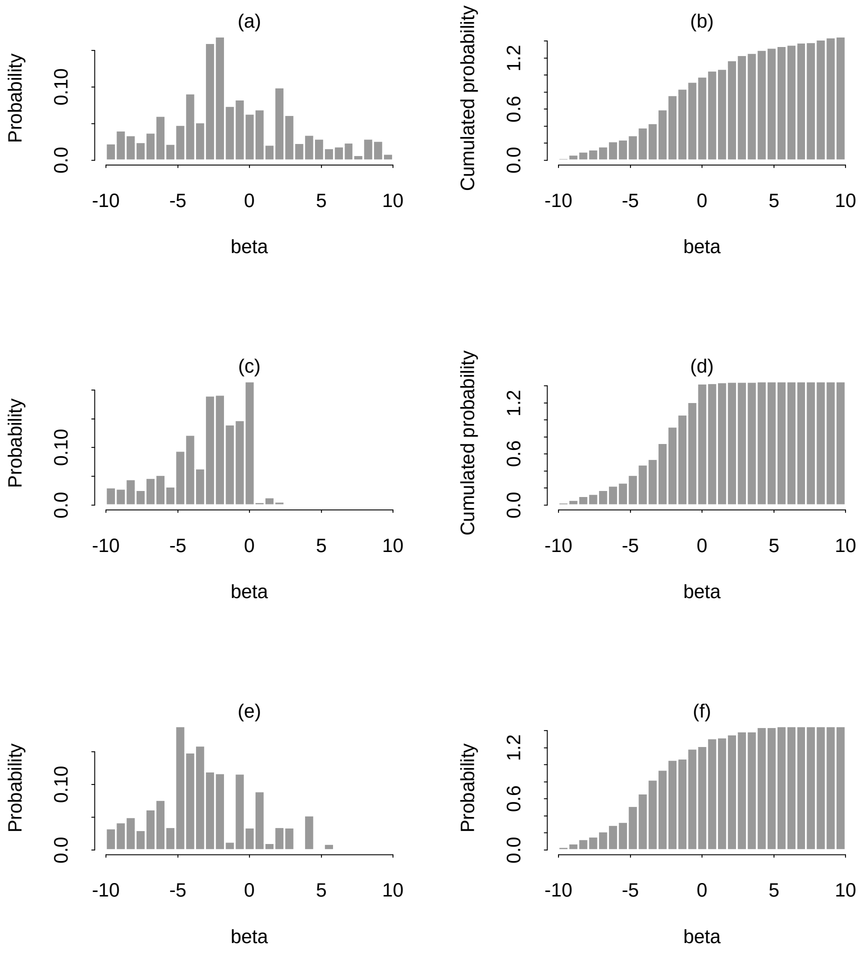

Figure 6: Posterior distributions of β_0 in the semi-parametric model $(n = 50, \nu = 1)$. Estimated posterior densities (figures (a), (c) and (e)) and cumulated distributions (figures (b), (d) and (f)) of β_0. For three drawings of the Markov process associated to the semi-parametric estimation.

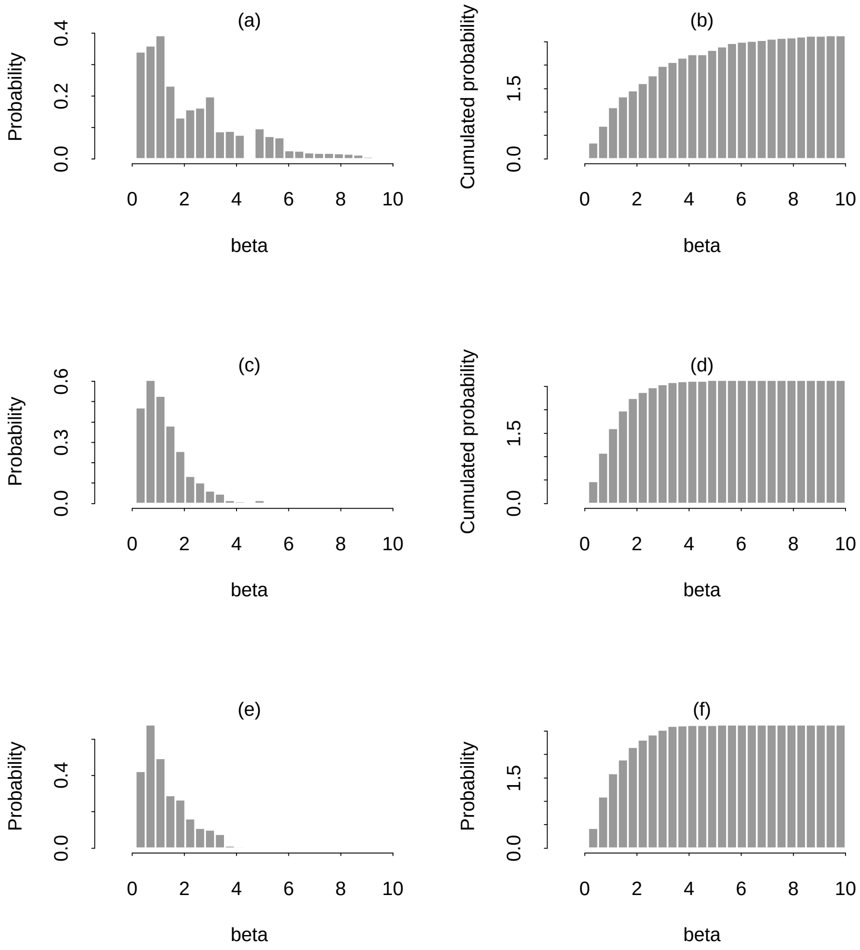

Figure 7: Time series of β_0 in semi-parametric estimation $(n = 50, \nu = 1)$
Time series of β_0 (figures (a), (c) and (e)) and their correlation structures (figures (b), (d) and (f)). For three drawings of the Markov process associated to the semi-parametric estimation.

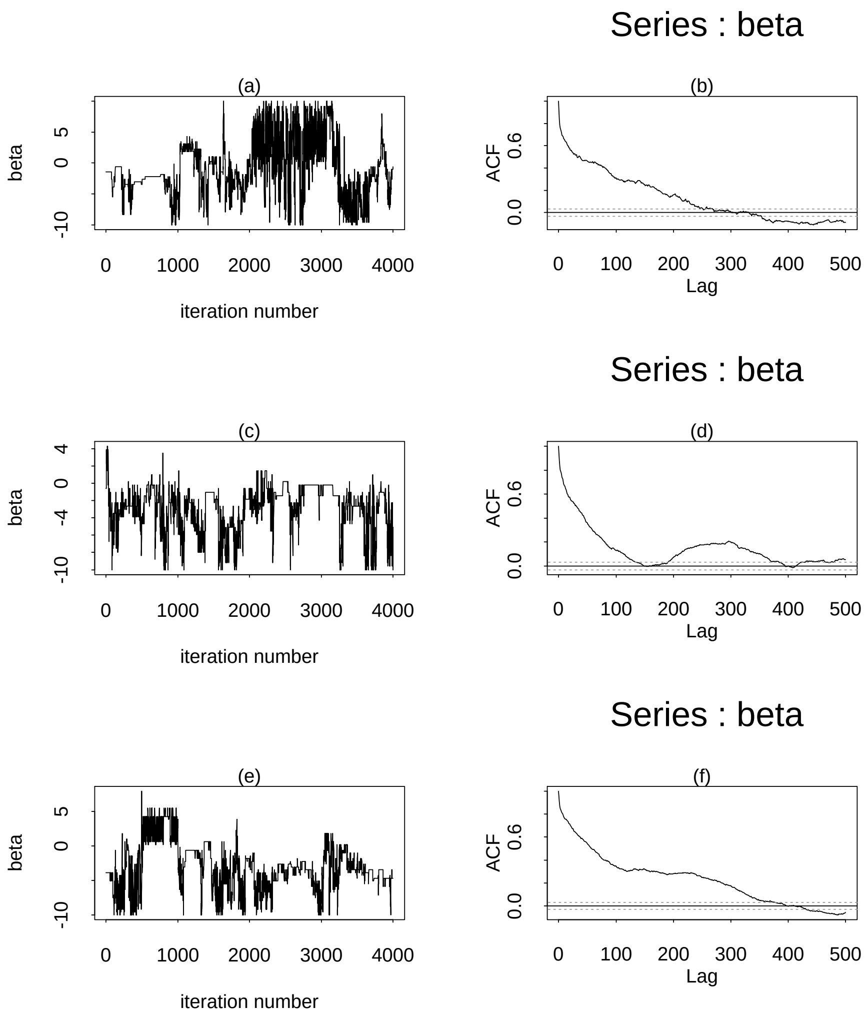

Figure 8: Posterior distributions of β_1 in the semi-parametric model $(n = 50, \nu = 1)$. Estimated posterior densities (figures (a), (c) and (e)) and cumulated distributions (figures (b), (d) and (f)) of β_0. For three drawings of the Markov process associated to the semi-parametric estimation.

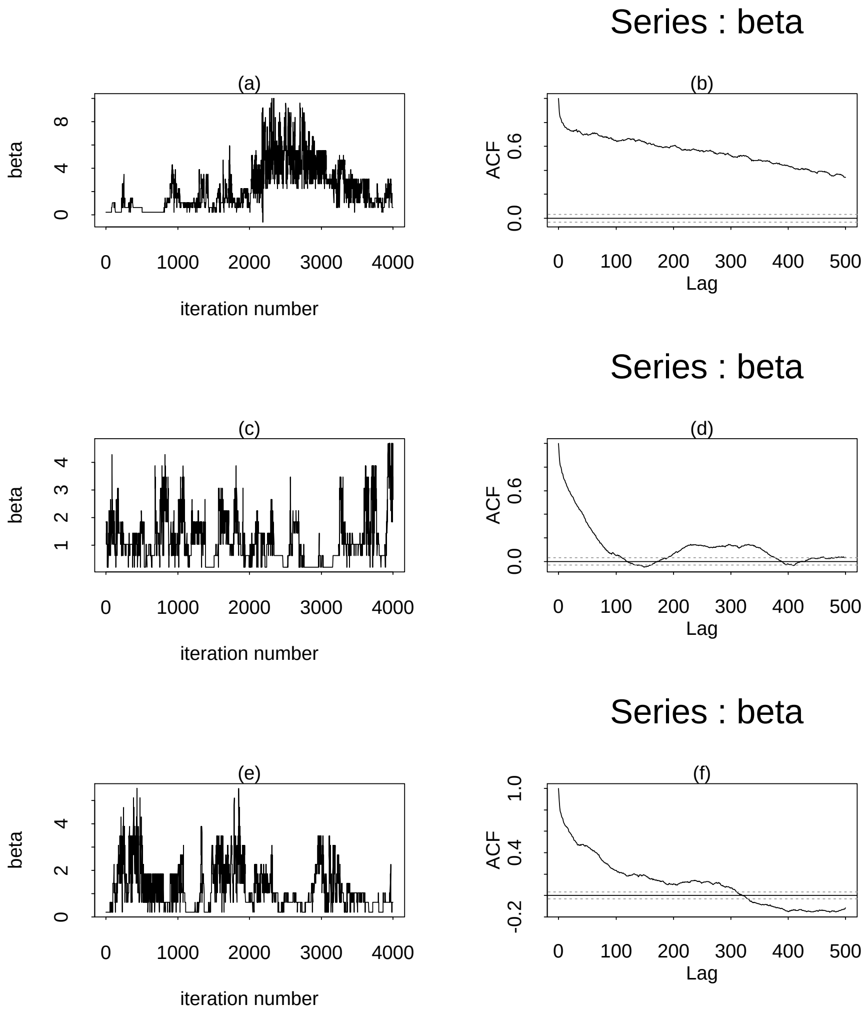

Figure 9: Time series of β_1 in semi-parametric estimation ($n = 50, \nu = 1$)
Time series of β_1 (figures (a), (c) and (e)) and their correlation structures (figures (b), (d) and (f)). For three drawings of the Markov process associated to the semi-parametric estimation.

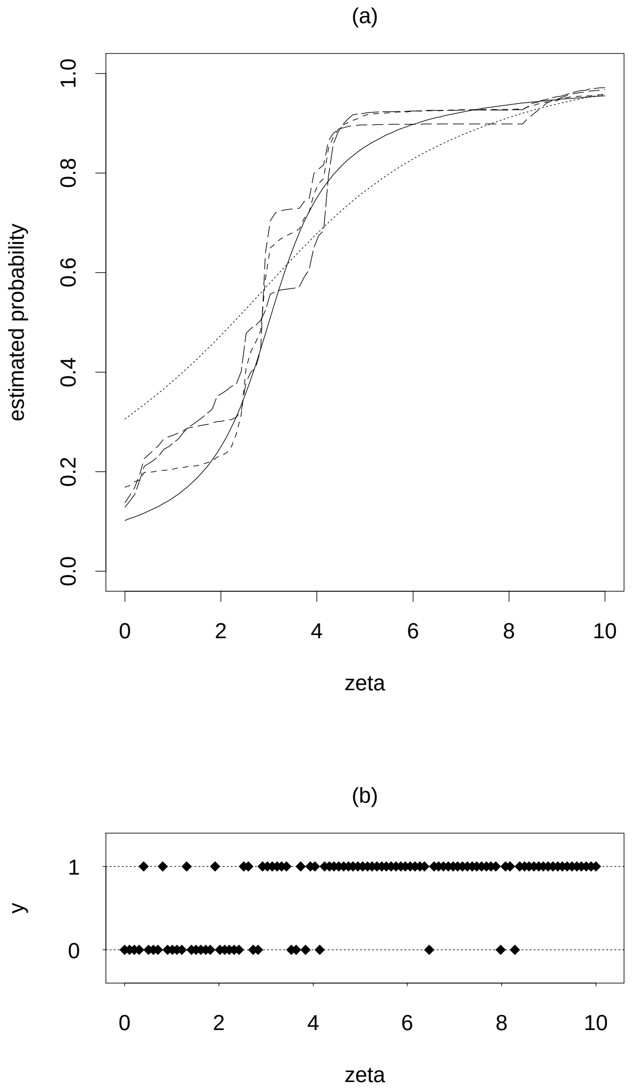

Figure 10: Predictive posterior expectations $(n = 100, \nu = 1)$

Predictive posterior expectations computed with *Method 2* (figure (a)), for the sample reproduced in figure (b) and generated from $F^* = T(\nu, 0, 1)$ and $b^* = (-3, 1)$

——— $E(Y \mid Z) = F^*(Z'b^*)$, actual model.

.... $\hat{E}(Y \mid Z, X_1^n) = \sum_{\tilde{\beta}^{(m)}} \Phi(Z'\tilde{\beta}^{(m)})\hat{m}(\tilde{\beta}^{(m)} \mid X_1^n)$, parametric model.

- - - - $\hat{E}(Y \mid Z, X_1^n) = \frac{1}{K} \sum E(\psi(Z) \mid \lambda 1^{n\,(k)})(\beta^{(k)'} Z)$, semi-parametric model (3 drawings) .

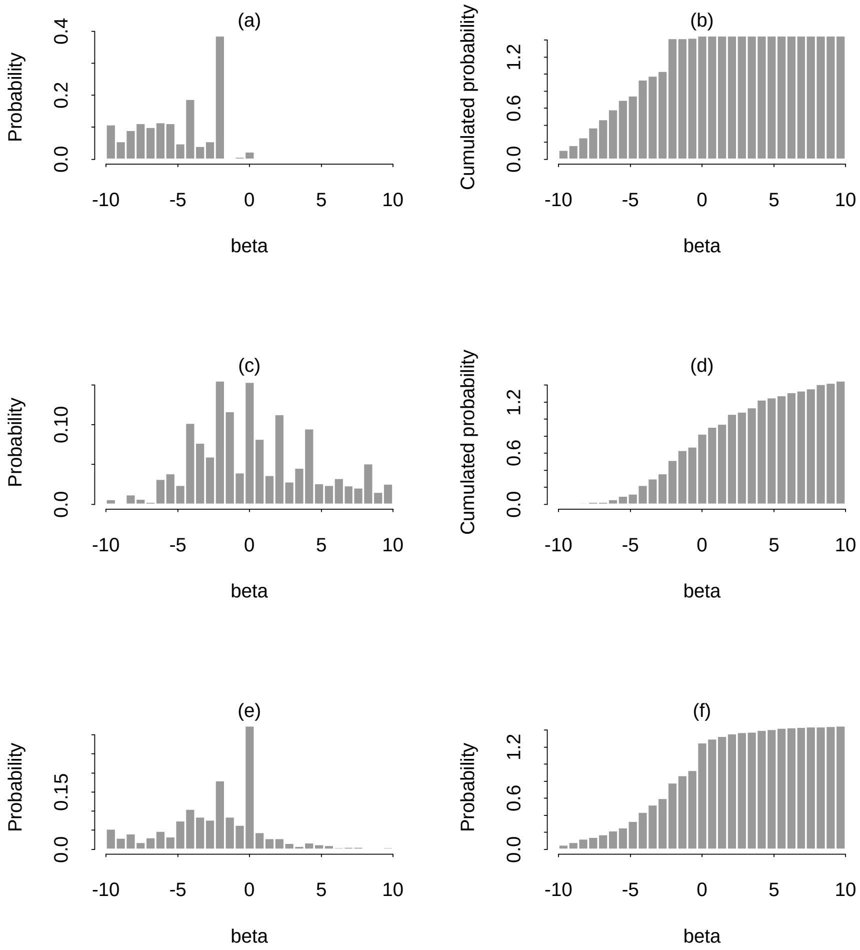

Figure 11: Posterior distributions of β_0 in the semi-parametric model $(n = 100, \nu = 1)$. Estimated posterior densities (figures (a), (c) and (e)) and cumulated distributions (figures (b), (d) and (f)) of β_0. For three drawings of the Markov process associated to the semi-parametric estimation

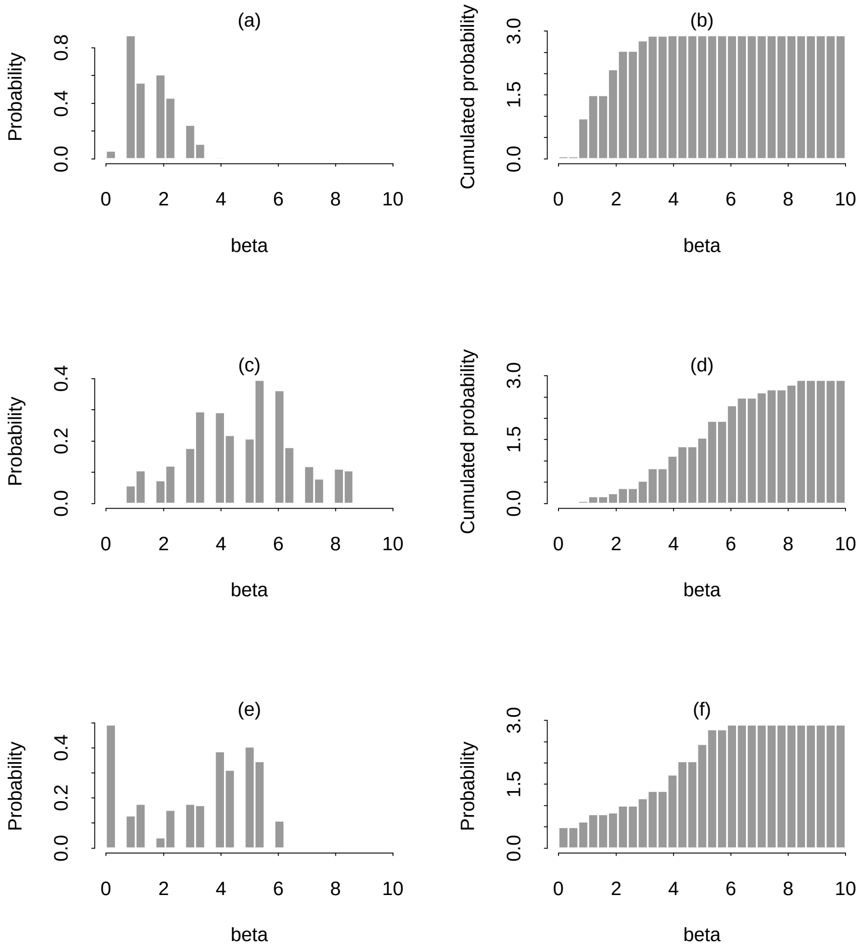

Figure 12: Time series of β_0 in semi-parametric estimation $(n = 100, \nu = 1)$
Time series of β_0 (figures (a), (c) and (e)) and their correlation structures (figures (b), (d) and (f)). For three drawings of the Markov process associated to the semi-parametric estimation

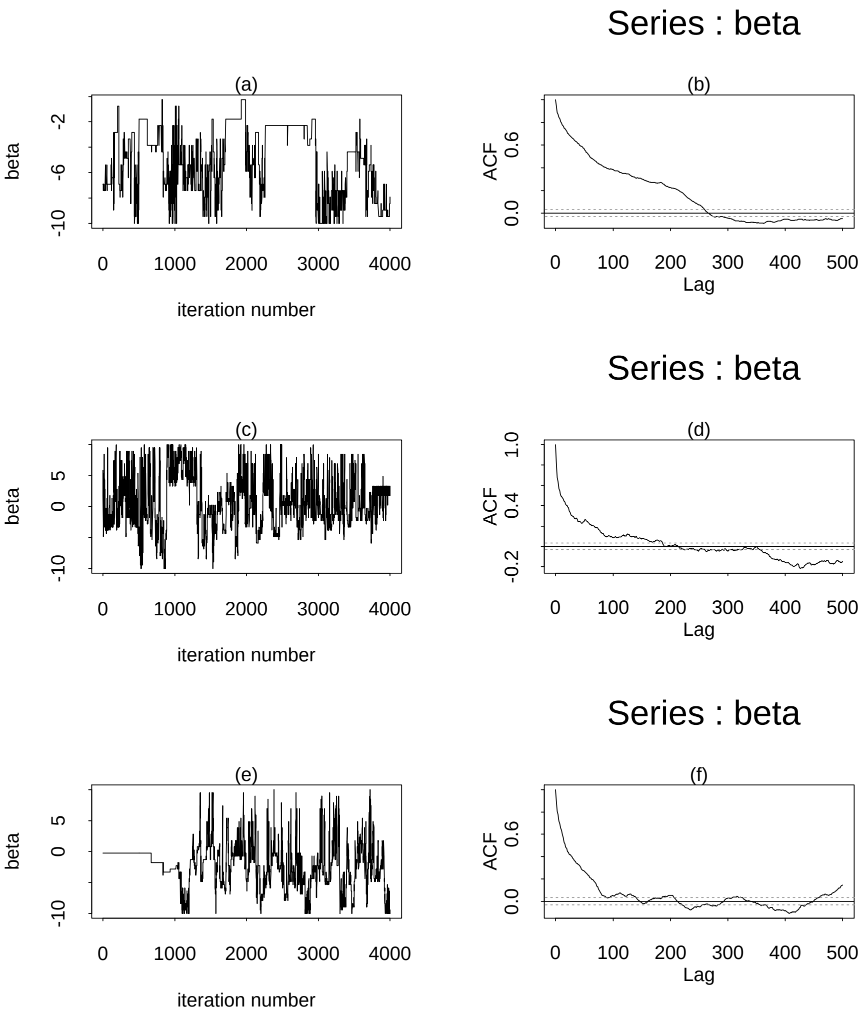

Figure 13: Posterior distributions of β_1 in the semi-parametric model($n = 100, \nu = 1$). Estimated posterior densities (figures (a), (c) and (e)) and cumulated distributions (figures (b), (d) and (f)) of β_0. For three drawings of the Markov process associated to the semi-parametric estimation

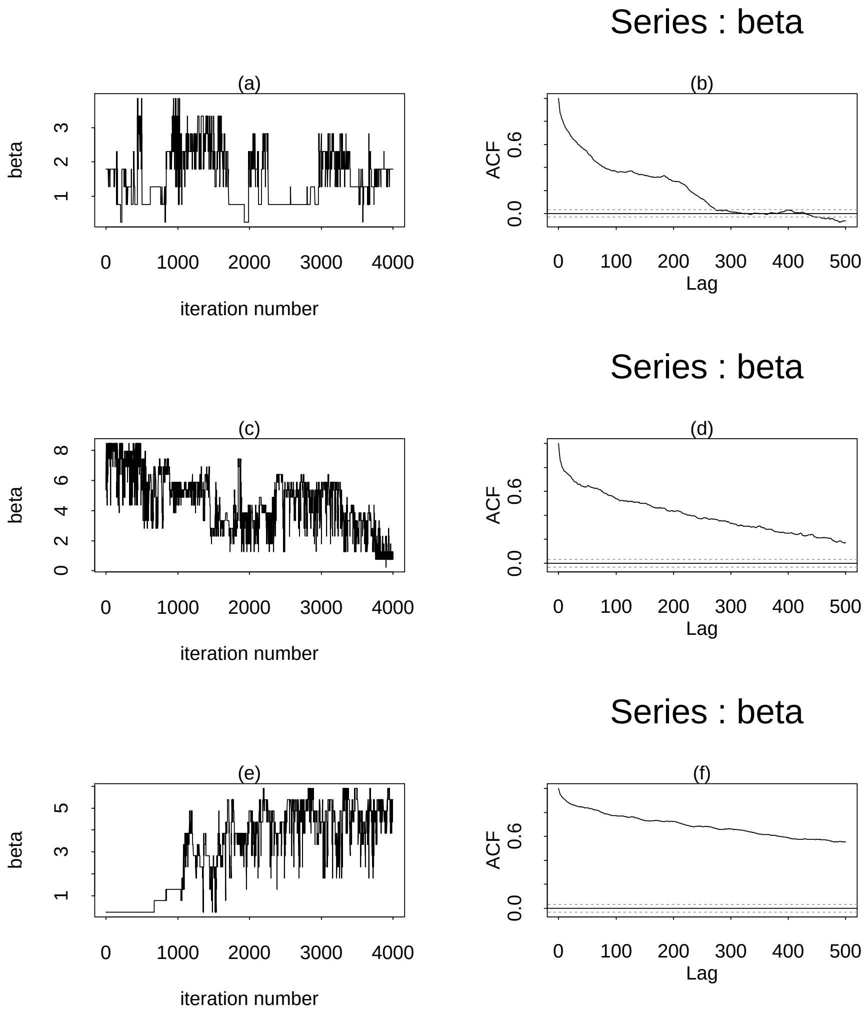

Figure 14: Time series of β_1 in semi-parametric estimation($n = 100, \nu = 1$)
Time series of β_1 (figures (a), (c) and (e)) and their correlation structures (figures (b), (d) and (f)). For three drawings of the Markov process associated to the semi-parametric estimation