



**HAL**  
open science

## Bayes factor consistency in regression problems

Judith Rousseau, Choi Taeryon

► **To cite this version:**

| Judith Rousseau, Choi Taeryon. Bayes factor consistency in regression problems. 2012. hal-00767469

**HAL Id: hal-00767469**

**<https://hal.science/hal-00767469>**

Preprint submitted on 20 Dec 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayes factor consistency in regression problems

Judith Rousseau and Taeryon Choi\*  
*Université Paris Dauphine and Korea University*

May 30, 2012

## Abstract

We investigate the asymptotic behavior of the Bayes factor for regression problems in which observations are not required to be independent and identically distributed and provide general results about consistency of the Bayes factor. Then we specialize our results to the model selection problem in the context of partially linear regression model in which the regression function is assumed to be the additive form of the linear component and the nonparametric component. Specifically, sufficient conditions to ensure Bayes factor consistency are given for choosing between the parametric model and the semiparametric alternative in the partially linear regression model.

*Keywords* : Bayes factor, Hellinger distance, Kullback-Leibler neighborhoods, Partially linear models, Rate of contraction

## 1 Introduction

Suppose we have two candidate models  $\mathcal{M}_0$  and  $\mathcal{M}_1$  for  $y^n \in \mathcal{Y}^n$ , a set of  $n$  observations from an arbitrary distribution  $P^n$  which is absolutely continuous with respect to a common measure  $\mu^n$  on  $\mathcal{Y}^n$ . We also assume two candidate models to have respective parameters and prior distributions,  $\theta$ ,  $\pi_0(\theta)$ ,  $\lambda$  and  $\pi_1(\lambda)$ ,

$$\mathcal{M}_0 = \{p_\theta^n(y^n), \theta \in \Theta, \pi_0(\theta)\}, \quad \mathcal{M}_1 = \{p_\lambda^n(y^n), \lambda \in \Lambda, \pi_1(\lambda)\}, \quad (1.1)$$

where  $p_\theta^n(y^n)$  and  $p_\lambda^n(y^n)$  denote the densities of  $y^n$  with respect to  $\mu^n$  under  $\mathcal{M}_0$  and  $\mathcal{M}_1$  respectively.

Based on this set of observations, a common Bayesian procedure to measure the evidence in favor of  $\mathcal{M}_0$  over  $\mathcal{M}_1$  is to assess the Bayes factor (Jeffreys, 1961), the ratio of the respective marginal densities or prior predictive densities

---

\*Corresponding author. Email:trchoi@gmail.com

of the data for the two competing models. Given two candidate models  $\mathcal{M}_0$  and  $\mathcal{M}_1$ , the marginal densities of  $y^n$  are computed by

$$m_0(y^n) = p(y^n|\mathcal{M}_0) = \int p(y^n|\mathcal{M}_0, \theta)\pi(\theta|\mathcal{M}_0)d\theta = \int p_\theta^n(y^n)\pi_0(\theta)d\theta,$$

$$m_1(y^n) = p(y^n|\mathcal{M}_1) = \int p(y^n|\mathcal{M}_1, \lambda)\pi(\lambda|\mathcal{M}_1)d\lambda = \int p_\lambda^n(y^n)\pi_1(\lambda)d\lambda.$$

Assuming the prior model probabilities  $\Pr(\mathcal{M}_j)$ ,  $j = 0, 1$  with  $\sum_{j=0}^1 \Pr(\mathcal{M}_j) = 1$ , the Bayes factor, i.e. the ratio of posterior odds and prior odds, is equivalent to the ratio of two marginal densities (Kass and Raftery, 1995), given by

$$B_{01} = \frac{\Pr(\mathcal{M}_0|y^n)}{\Pr(\mathcal{M}_1|y^n)} \Big/ \frac{\Pr(\mathcal{M}_0)}{\Pr(\mathcal{M}_1)} = \frac{p(y^n|\mathcal{M}_0)}{p(y^n|\mathcal{M}_1)} = \frac{m_0(y^n)}{m_1(y^n)}.$$

Alternatively, the posterior probability of  $\mathcal{M}_0$  is represented by

$$\Pr(\mathcal{M}_0|y^n) = \frac{\Pr(\mathcal{M}_0) \cdot B_{01}}{\Pr(\mathcal{M}_0) \cdot B_{01} + \Pr(\mathcal{M}_1)}.$$

Note that the large value of  $B_{01}$  indicates the strong evidence in support of model  $\mathcal{M}_0$  (Jeffreys (1961) and Kass and Raftery (1995)). Accordingly,  $B_{01}$  is expected to converge to infinity as the sample size increases when  $\mathcal{M}_0$  is the true model, and this concept can be formulated as,

$$\lim_{n \rightarrow \infty} B_{01} = \infty, \text{ equivalently } \lim_{n \rightarrow \infty} \Pr(\mathcal{M}_0|y^n) = 1, \quad (1.2)$$

when  $\mathcal{M}_0$  is the true model. The convergence in (1.2) denotes *in-probability* convergence under the true sampling distribution of  $y^n$ , and that the former is called *Bayes factor consistency* or consistency of the Bayes factor, and the latter is often called *posterior model consistency* or posterior consistency for model choice. Note that consistency of Bayesian model selection procedure is the fundamental issue to be secured, whereas the model selection using the classical tools such as  $C_p$  and AIC generally do not guarantee model selection consistency (e.g. see Berger and Pericchi (1996) and Yang (2005)).

Recent results in the Bayes factor consistency include works by Verdinelli and Wasserman (1998), Dass and Lee (2004), Ghosal et al. (2008), and McVinish et al. (2009), which focus on the density estimation in the Bayesian goodness of fit testing problems. Their results are based on verifying sufficient conditions related to posterior consistency and posterior convergence rates. In addition, those conditions are mainly designed for the case of independent and identically distributed (i.i.d.) observations, and thus it is expected to generalize them to the case of non i.i.d observations as in the case of regression problems. The consistency of the Bayes model selection in regression problems has been largely studied in Gaussian linear regression models, particularly in the context of variable selection procedures, for example, Liang et al. (2008), Casella et al. (2009), Moreno et al. (2010) and Shang and Clayton (2011). On the other hand, a recent work by Choi et al. (2009) investigated the Bayes factor consistency in the

partially linear regression model with a specific trigonometric representation of the nonparametric component, in which the analytic form of the Bayes factor was directly evaluated for its asymptotic behavior under suitable conditions.

Alternatively, this paper investigates the asymptotic behavior of the Bayes factor for regression problems in which observations are not required to be independent and identically distributed. In particular, we consider a uniform version of consistency of the Bayes factor and discuss general results on the Bayes factor consistency in Section 2. Then we specialize our results to the model selection problem in the context of partially linear regression model, in which the regression function is assumed to be the additive form of the linear component and the nonparametric component. Specifically, sufficient conditions to ensure Bayes factor consistency are given in Section 3 for choosing between the parametric model and the semiparametric alternative in the partially linear regression model. sufficient conditions to ensure Bayes factor consistency are given for the partially linear regression model. Section 4 makes a concluding remark and discusses further extension of the consistency of the Bayes factor based on the general theorem we propose in the paper.

## 2 General Theorem

In this section, we give the general theorem to obtain consistency of the Bayes factor where observations are not required to be independent and identically distributed (IID) and in a framework where  $\Theta \subset \mathbb{R}^k$  for some  $k \geq 1$ , where  $\Lambda$  is typically infinite dimensional. For this purpose, we make use of asymptotic results established in Ghosal and van der Vaart (2007) and McVinish et al. (2009) and extend general results of the Bayes factor consistency to the non-IID observations.

Then, the Bayes factor is given by

$$B_{01} = \frac{\int_{\Theta} p_{\theta}^n(y^n) d\pi_0(\theta)}{\int_{\Lambda} p_{\lambda}^n(y^n) d\pi_1(\lambda)}. \quad (2.1)$$

Consistency of the Bayes factor is usually formulated as follows :

$$\lim_{n \rightarrow \infty} B_{01} = \begin{cases} \infty, & \text{in } P_{\theta_0}^n \text{ probability, if } p_{\theta_0}^n \in \mathcal{M}_0 \\ 0, & \text{in } P_{\lambda_0}^n \text{ probability, if } p_{\lambda_0}^n \in \mathcal{M}_1, \end{cases}$$

where  $P_{\theta_0}$  represents the true probability measure belonging to the null model,  $P_{\lambda_0}$  represents the true probability measure belong to the alternative model. One drawback of the above formulation is that it is pointwise and not uniform. We therefore consider in this paper a uniform version of consistency of the Bayes factor written as : let  $\Lambda_0 \subset \Lambda$ , for all  $K$  compact subset of  $\Theta$  and for all  $\epsilon > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\theta_0 \in K} P_{\theta_0}^n [B_{01}^{-1} > \epsilon] &= 0 \\ \lim_{n \rightarrow \infty} \sup_{\lambda_0 \in \Lambda_0} P_{\lambda_0}^n [B_{01} > \epsilon] &= 0 \end{aligned} \quad (2.2)$$

In the above formulation  $\Lambda_0$  is to be understood as some functional class whose *distance* to the null hypothesis is bounded from below. We shall make this notion more precise in Assumptions **A1** and **A2**.

In (2.1) and (2.2), we have typically in mind that  $\Lambda$  is much bigger than  $\Theta$ , and  $\Theta$  is often nested to  $\Lambda$ . In such cases, the difficulty comes from the fact that if  $p_{\theta_0}^n \in \mathcal{M}_0$ , it can also be approximated by densities in  $\mathcal{M}_1$ . Specifically, when  $p_{\theta_0}^n \in \mathcal{M}_0$ , the Kullback-Leibler property (Walker et al., 2004), a basic condition to be satisfied for posterior consistency and Bayes factor consistency, holds for both prior distributions under  $\mathcal{M}_0$  and  $\mathcal{M}_1$ . Since the Bayes factor is known to asymptotically support the model with the prior that satisfies the Kullback-Leibler property (Walker et al., 2004), it is often the case that the Bayes factor based on prior distributions only with the Kullback-Leiber property may not be enough, and additional conditions are required for consistent model selection between two competitive models. Ghosal et al. (2008) and McVinish et al. (2009) investigated this issue for IID observations, and we adapt it to non-IID observations.

In this respect, our investigation on Bayes factor consistency begins with the case that the observations  $y^n$  is actually generated by  $\mathcal{M}_1$ . We consider first a set of assumptions to obtain consistency under  $\mathcal{M}_1$  in (2.2), i.e. when the true model for  $y^n$  is assumed to be  $p_{\lambda_0}^n$ . For this purpose, we write the Bayes factor  $B_{01}$  as

$$B_{01} = \frac{J_0^{\lambda_0}(y^n)}{J_1^{\lambda_0}(y^n)}, \quad (2.3)$$

where

$$J_0^{\lambda_0}(y^n) = \int_{\Theta} \frac{p_{\theta}^n(y^n)}{p_{\lambda_0}^n(y^n)} d\pi_0(\theta), \text{ and } J_1^{\lambda_0}(y^n) = \int_{\Lambda} \frac{p_{\lambda}^n(y^n)}{p_{\lambda_0}^n(y^n)} d\pi_1(\lambda).$$

Moreover let  $d_n$  be a semimetric on  $\Lambda \cup \Theta$  and define

$$h(\lambda_0) = \liminf_n \inf_{\theta \in \Theta} d_n(p_{\lambda_0}^n, p_{\theta}^n), \quad \lambda_0 \in \Lambda$$

**Assumption A1.** Let  $\Lambda_0 \subset \Lambda$  satisfies : **For**  $\Lambda_0 \subset \Lambda$ , there exists  $\epsilon_n$  converging to 0 such that

$$\sup_{\lambda_0 \in \Lambda_0} P_{\lambda_0}^n \left[ J_1^{\lambda_0}(y^n) < e^{-n\epsilon_n^2} \right] = o(1)$$

**Assumption A2.**

**A2-1.**

$$\inf_{\lambda_0 \in \Lambda_0} h(\lambda_0) > 0$$

**A2-2.** There exists  $\epsilon_0 > 0$  such that for all  $\lambda_0 \in \Lambda_0$  and all  $\epsilon_0 > \epsilon > 0$ , there exists  $\Theta_n(\lambda_0) \subset \Theta$ , such that

$$\pi_0(\Theta_n(\lambda_0)^c) \leq e^{-2n\epsilon},$$

**A2-3.** For all  $\epsilon > 0$  there exists  $a > 0$  such that for all  $\lambda_0 \in \Lambda_0$ , there exists a sequence of tests  $\phi_n(\lambda_0)$  satisfying

$$\sup_{\lambda_0 \in \Lambda_0} E_{\lambda_0}^n [\phi_n(\lambda_0)] = o(1), \quad \sup_{\lambda_0 \in \Lambda_0} \sup_{\theta \in \Theta_n(\lambda_0)} E_{\theta}^n [1 - \phi_n(\lambda_0)] \leq e^{-an}.$$

**Theorem 1.** *Suppose that assumptions **A1** and **A2** hold. Then for all  $\epsilon > 0$  there exists  $\delta > 0$  such that*

$$\sup_{\lambda_0 \in \Lambda_0} P_{\lambda_0}^n [B_{01} e^{\delta n} > \epsilon] = o(1)$$

*That is, the Bayes factor is exponentially decreasing under  $\mathcal{M}_1$  (uniformly over  $\Lambda_0$ ).*

*Proof.* Under **A1**,  $J_1^{\lambda_0}(y^n) \geq e^{-n\epsilon_n^2}$ , with probability going to 1, under  $P_{\lambda_0}^n$ , uniformly over  $\Lambda_0$ .

Let  $\epsilon > 0$  and  $\delta > 0$ , then assumption **A2** implies that uniformly over  $\Lambda_0$

$$\begin{aligned} P_{\lambda_0}^n [B_{01} e^{\delta n} \geq \epsilon] &\leq E_{\lambda_0}^n (\phi_n(\lambda_0)) + E_{\lambda_0}^n [(1 - \phi_n(\lambda_0)) \mathbb{1}_{B_{01} e^{\delta n} \geq \epsilon}] \\ &\leq E_{\lambda_0}^n (\phi_n(\lambda_0)) + P_{\lambda_0}^n [J_1^{\lambda_0}(y^n) < e^{-n\epsilon_n^2}] \\ &\quad + \frac{e^{n\epsilon_n^2 + \delta n}}{\epsilon} \left[ \int_{\Theta_n(\lambda_0)} E_{\theta}^n (1 - \phi_n(\lambda_0)) d\pi_0(\theta) + \pi_0(\Theta_n(\lambda_0)^c) \right] \\ &\leq o(1) + e^{-n\epsilon/2 - na/2}, \end{aligned}$$

as soon as  $\delta < (\epsilon \wedge a)/2$ . □

Note that requiring these uniform assumptions is often not a difficulty in the Bayesian setting, where we typically control the above terms uniformly on balls with a given radius, such as Hölder or Besov balls.

We next consider a set of assumptions to obtain consistency under  $\mathcal{M}_0$  in (2.2), i.e. when the true model for  $y^n$  is assumed to be  $p_{\theta_0}^n$ . We write the Bayes factor  $B_{01}$  as

$$B_{01} = \frac{J_0^{\theta_0}(y^n)}{J_1^{\theta_0}(y^n)}, \tag{2.4}$$

where

$$J_0^{\theta_0}(y^n) = \int_{\Theta} \frac{p_{\theta}^n(y^n)}{p_{\theta_0}^n(y^n)} d\pi_0(\theta), \quad \text{and} \quad J_1^{\theta_0}(y^n) = \int_{\Lambda} \frac{p_{\lambda}^n(y^n)}{p_{\theta_0}^n(y^n)} d\pi_1(\lambda)$$

Let  $KL(f, g)$  denote the Kullback-Leibler divergence between  $f$  and  $g$  and  $V(f, g) = \int f(\log f / \log g)^2$ , and let  $d_n$  be a semimetric on  $\Lambda \cup \Theta$ .

**Assumption B1.** For all  $K \subset \Theta$  compact and all  $\theta_0 \in K$ , there exists  $k_0 > 0$  such that

$$\inf_{\theta_0 \in K} n^{k_0/2} \pi_0 [\{\theta : KL(p_{\theta_0}^n, p_{\theta}^n) \leq 1, \quad V(p_{\theta_0}^n, p_{\theta}^n) \leq 1\}] \geq C$$

for some positive constants  $C$ .

**Assumption B2.** For all  $K \subset \Theta_0$ , for all  $\theta_0 \in K$ , there exists  $\epsilon_n > 0$  going to 0, with  $A_{\epsilon_n}(\theta_0) = \{p_\lambda : d_n(p_\lambda^n, p_{\theta_0}^n) < \epsilon_n\}$  such that

**B2-1.**

$$\sup_{\theta_0 \in K} P_{\theta_0}^n [\pi_1 [A_{\epsilon_n}^c(\theta_0) | y^n]] = o(1)$$

and such that

**B2-2.**

$$\sup_{\theta_0 \in K} \pi_1 [A_{\epsilon_n}(\theta_0)] = o(n^{-k_0/2}).$$

where  $k_0$  is the same positive constant as in Assumption **B1**.

In other words, the posterior probability of  $A_{\epsilon_n}$  from  $\pi_1$  is converging to 1 in  $p_{\theta_0}^n$ -probability with rate  $\epsilon_n$ , and the prior probability of  $A_{\epsilon_n}$  from  $\pi_1$  has a positive probability but converging to 0. In the above assumptions  $k_0$  and  $\epsilon_n$  depend on  $\theta_0$ . Note that the same condition was also discussed in (McVinish et al., 2009, Assumption A3) and a similar but stronger condition was given in (Ghosal et al., 2008, (4.1)) for nonparametric Bayesian density estimation.

**Lemma 1.** Define  $S_n = \{\theta : KL(p_{\theta_0}^n, p_\theta^n) \leq 1, V(p_{\theta_0}^n, p_\theta^n) \leq 1\}$ , for some  $\theta_0 \in \Theta$ . Then,

$$\lim_{C \rightarrow \infty} \sup_n P_{\theta_0}^n [J_0^{\theta_0}(y^n) < e^{-C} \pi_0(S_n)/2] = 0.$$

*Proof.* The proof is similar to the proof of Theorem 1 in McVinish et al. (2009) for instance, given as follows :

First, note that

$$J_0^{\theta_0}(y^n) \geq \int_{\Theta} \frac{p_{\theta_0}^n(y^n)}{p_{\theta_0}^n(y^n)} \mathbb{1}_{\Omega_n}(y^n, \theta) d\pi_0(\theta) \geq e^{-C} \int_{S_n} \mathbb{1}_{\Omega_n}(y^n, \theta) d\pi_0(\theta),$$

where

$$\Omega_n = \{(y^n, \theta) : \ell_n(\theta) - \ell_n(\theta_0) \geq -C\}, \quad \ell_n(\theta) = \log p_\theta^n(y^n).$$

Then for sufficiently large  $C > 0$ , we have

$$\begin{aligned} P_{\theta_0}^n [J_0^{\theta_0}(y^n) < e^{-C} \pi_0(S_n)/2] &\leq P_{\theta_0}^n [\pi_0(S_n \cap \Omega_n^c) > \pi_0(S_n)/2] \\ &\leq 2 \frac{\int_{S_n} P_{\theta_0}^n [\Omega_n^c(\theta)] d\pi_0(\theta)}{\pi_0(S_n)} \\ &\leq \frac{8}{C^2}. \end{aligned}$$

where the latter inequality comes from Chebyshev inequality on  $\ell_n(\theta_0) - \ell_n(\theta) - KL(p_{\theta_0}^n, p_\theta^n)$ . Therefore, it follows that

$$P_{\theta_0}^n [J_0^{\theta_0}(y^n) < e^{-C} \pi_0(S_n)/2] \rightarrow 0 \text{ as } C \rightarrow \infty,$$

uniformly in  $\theta_0$ . □

**Theorem 2.** Suppose that assumptions **B1** and **B2** hold. Let  $P_{\theta_0}^n$  denote the joint distribution of  $y^n$ . If  $\theta_0 \in \Theta_0$ , then

$$B_{01} \rightarrow \infty \quad \text{in } P_{\theta_0}^n\text{-probability.}$$

That is, the Bayes factor is increasing to infinity under  $\mathcal{M}_0$ .

*Proof.* Let  $B_{10} = B_{01}^{-1}$ ,  $\epsilon > 0$  and  $\theta_0 \in K$  for some compact set  $K$ , define

$$A_n = \{y^n : J_0^{\theta_0}(y^n) > e^{-C} \pi_0(S_n)/2\} \cap \{y^n : \pi_1(A_{\epsilon_n}|y^n) > 1 - \epsilon\},$$

where  $S_n$  is defined in the proof of Lemma 1 and  $\epsilon_n$  and  $A_{\epsilon_n}$  are defined in Assumption **A2**. Suppose that  $y^n \in A_n$ . Then, by assumption **B1**,

$$B_{10} \leq e^C \frac{2}{C} n^{d/2} J_1^{\theta_0}(y^n) = e^C \frac{2}{C} n^{d/2} \frac{\int_{A_{\epsilon_n}} \frac{p_{\lambda}^n(y^n)}{p_{\theta_0}^n(y^n)} d\pi_1(\lambda)}{\pi_1(A_{\epsilon_n}|y^n)}.$$

Thus, by Lemma 1 and assumption **B2-1**,

$$\begin{aligned} P_{\theta_0}^n \left[ B_{10} > e^{2C} n^{\mathbf{ko}/2} \pi_1(A_{\epsilon_n}) \right] &\leq P_{\theta_0}^n \left[ A_n \cap \left\{ B_{10} > e^{2C} n^{\mathbf{ko}/2} \pi_1(A_{\epsilon_n}) \right\} \right] + P_{\theta_0}^n [A_n^c] \\ &\leq \frac{C e^{-C}}{2(1-\epsilon)} + P_{\theta_0}^n [A_n^c] \\ &\leq o(1) + 2C e^{-C} + 8/C^2 \xrightarrow{C \rightarrow \infty} 0. \end{aligned}$$

Note the above bounds are uniform over  $K$ . Therefore, under  $\mathcal{M}_0$ , the Bayes factor goes to infinity at a rate bounded by  $O(n^{d/2} \pi(A_{\epsilon_n}))$ .

Hence, by assumption **B2-2**,  $B_{01}^{-1} \rightarrow 0$  with  $P_{\theta_0}^n$  probability tending to 1, which implies  $B_{01}$  converges to infinity with  $P_{\theta_0}^n$  probability tending to 1 when the true model is from  $\mathcal{M}_0$ .  $\square$

### 3 Application to the partially linear model

In this section we apply the general theorems in the previous section to the model selection problem for partially linear models in choosing between the linear regression model, and the semiparametric alternative. Bayesian methods in partially linear models have been developed in for example Lenk (1999), Koop and Poirier (2004), and Ko et al. (2009), whereas theoretical validation of these Bayesian methods has little been investigated, in particular for consistency of Bayes factor except for the recent result by Choi et al. (2009).

Accordingly, we attempt to investigate Bayes factor consistency in the partially linear regression based on the general theorems, Theorem 1 and Theorem 2 which we established in the previous section. Specifically, we adapt assumptions **A1** and **A2** of Theorem 1 and assumptions **B1** and **B2** of Theorem 2 to the partially linear regression models, and provide sufficient conditions to ensure consistency of the Bayes factor.



For this purpose, we consider the following partially linear regression model,

$$y_i = \alpha + \beta^t d_i + f(x_i) + \sigma \epsilon_i, \quad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1), \quad (3.1)$$

where the mean function of the regression model in (3.1) has two parts: a  $p$ -dimensional parametric part with  $\beta^t d_i$ ,  $\{d_i\}_{i=1}^n \in [-1, 1]^p$  and a nonparametric part with an unknown function  $f(x_i)$ ,  $\{x_i\}_{i=1}^n \in [0, 1]^q$  in the infinite dimensional parameter space, with  $p, q \geq 1$ . We consider here the case of random design, i.e.  $(d_i, x_i) \sim \mu$  independently, with  $\mu$  a probability measure on  $[-1, 1]^p \times [0, 1]^q$ , we assume that  $E[d] = 0$  under this measure and that

$$\int dd^t d\mu_p(d) > 0,$$

where the latter inequality means that the covariance matrix of  $d$  is positive definite.

To begin with, we introduce additional notations and assumptions necessary for the technical details in the remainder of the paper. For all function  $g \in L^2([0, 1]^q)$ , we denote  $\|g\| = \left(\int_0^1 g^2(x) dx\right)^{1/2}$ , and  $g^n = (g(x_1), \dots, g(x_n))^t$ . Also for all  $n$ -dimensional vector  $\eta^n = (\eta_1, \dots, \eta_n)^t \in \mathbb{R}^n$ , we denote  $\|\eta^n\|_n^2 = n^{-1} \sum_{i=1}^n \eta_i^2$ . Let  $Z$  be the matrix whose  $i$ -th row is given by  $Z_i = (1, d_i^t)$ ,  $i = 1, \dots, n$ . Let  $\gamma_0 \in \mathbb{R}^{p+1}$  and  $f_0 \in L^2([0, 1]^q)$  be the true values of unknown parameters in (3.1).

Bayesian inference for the partially linear regression model in (3.1) begins with the specification of prior distributions for  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^p$ ,  $f(\cdot)$  on a given class of measurable functions and  $\sigma \in \mathbb{R}^+$ . Based on the model structure in (3.1) with suitable prior distributions for unknown parameters, we build the posterior distribution and estimate the regression function  $\eta_{\alpha, \beta, f}(d, x) = \alpha + \beta^t d + f(x)$ . After the model estimation, we perform the Bayesian model checking procedure and see the adequacy of the model structure we assumed. Specifically, under the partially linear regression model in (3.1), we would like to choose between a parametric component and its semiparametric counter part,  $\mathcal{M}_0$  and  $\mathcal{M}_1$ , given by

$$\mathcal{M}_0 : y_i = \alpha + \beta^t d_i + \sigma \epsilon_i, \text{ vs. } \mathcal{M}_1 : y_i = \alpha + \beta^t d_i + f(x_i) + \sigma \epsilon_i, \quad (3.2)$$

and model selection is made by computing the Bayes factor in (2.1),

$$B_{01} = \frac{\int_{\Theta} p_{\theta}^n(y^n) d\pi_0(\theta)}{\int_{\Lambda} p_{\lambda}^n(y^n) d\pi_1(\lambda)},$$

where  $\theta = (\gamma, \sigma)$ ,  $\gamma = (\alpha, \beta)$  and  $\lambda = (\gamma, f, \sigma)$ . This is equivalent to testing

$$\begin{aligned} H_0 : & \inf_{\gamma \in \mathbb{R}^{p+1}} \int_{[0, 1]^{p+q}} \|(1, d^t)\gamma - f(x)\|^2 d\mu(d, x) = 0 \quad \text{vs} \\ H_1 : & \inf_{\gamma \in \mathbb{R}^{p+1}} \int_{[0, 1]^{p+q}} \|(1, d^t)\gamma - f(x)\|^2 d\mu(d, x) > 0 \end{aligned}$$

If  $d$  and  $x$  are independent under  $\mu$  then this is equivalent to testing

$$H_0 : f = \text{constant} \quad \text{vs} \quad H_1 : f \neq \text{constant}$$

For each  $f \in L^2[0, 1]^q$ , we write

$$H(f) = \inf_{\gamma \in \mathbb{R}^{p+1}} \int_{[0,1]^{p+q}} \|(1, d^t)\gamma - f(x)\|^2 d\mu(d, x)$$

$H(f)$  acts as a distance to the null hypothesis.

We consider the following general families of prior distributions under  $\mathcal{M}_0$  and  $\mathcal{M}_1$  :

- Prior distribution  $\pi_0$  on  $\mathcal{M}_0$ : The prior  $\pi_0$  on the parametric model is assumed to be absolutely continuous with respect to the Lebesgue measure with positive, continuous and bounded density on  $\mathbb{R}^{p+1} \times \mathbb{R}^+$ .
- Prior distribution  $\pi_1$  on  $\mathcal{M}_1$ : The prior  $\pi_1$  on  $\lambda$  is assumed to be

$$d\pi_1(\lambda) = \pi_p(\gamma, \sigma) d\pi_f(f) d\gamma d\sigma,$$

with  $\pi_p(\gamma, \sigma)$  continuous and positive on  $\mathbb{R}^{p+1} \times \mathbb{R}^+$  and  $\pi_f$  is assumed to have support in  $L^2([0, 1]^q)$ .

We assume the following condition on  $\pi_0$  :

CONDITION (P0) (Parametric prior  $\pi_0$ ): for all  $\epsilon > 0$  there exists  $a_\epsilon > 0$  and  $N_\epsilon$  such that  $\forall n \geq N_\epsilon$ ,

$$\pi_0 \left[ e^{-a_\epsilon n} \leq \sigma \leq e^{a_\epsilon n}; \|\gamma\|_{p+1} \leq e^{a_\epsilon n} \right] \geq 1 - e^{-\epsilon n}$$

This is a very weak assumption on the prior  $\pi_0$ . In particular if  $\sigma$  follows a either a *Gamma*( $a, b$ ) with  $a, b > 0$ , or an *inverse Gamma*( $a, b$ ) or a truncated Gaussian as in Gelman (2006) and if the prior on  $\alpha$  and  $\beta$  has at least polynomial tails then condition (P0) is satisfied.

We first study the consistency of the Bayes factor under the alternative hypothesis.

### 3.1 Consistency under $\mathcal{M}_1$

CONDITION (P0) and basic assumptions on  $\pi_0$  and  $\pi_1$  above are sufficient to ensure the following consistency result under  $\mathcal{M}_1$ :

**Theorem 3.** *Consider the above framework with  $\pi_p$  and  $\pi_0$  satisfying CONDITION (P0), then for all  $L > 0$  and all all functional classes  $\mathcal{C} \subset \{(\gamma, \sigma, f) \in \mathbb{R}^{p+1} \times \mathbb{R}^+ \times L^2([0, 1]^q); \alpha^2 + \|\beta\|_p^2 + \|f\|^2 \leq L\}$  such that there exists  $\epsilon_n \downarrow 0$ , with  $n\epsilon_n^2 \rightarrow +\infty$ , for which*

$$\inf_{f_0 \in \mathcal{C}} \pi(\|f - f_0\|_2 \leq \epsilon_n) \geq e^{-cn\epsilon_n^2} \quad (3.3)$$

for some  $c > 0$ , we have for all  $\epsilon > 0$  there exists  $\delta > 0$  such that

$$\sup_{f_0 \in \mathcal{C}: H(f_0) > \epsilon} P_0 [B_{01} e^{\delta n} > \epsilon] = o(1).$$

Before we prove Theorem 3, we consider the following Lemma which will be useful for the proof of Theorem 3.

**Lemma 2.** *There exist  $0 < c_1 \leq C_1 < +\infty$  such that*

$$\mu [c_1 n \leq \|Z^t Z\| \leq C_1 n] = 1 + o(1) \quad (3.4)$$

We now prove Theorem 3.

*Proof.* Let  $\epsilon > 0$  be fixed and  $K$  be a compact subset of  $\mathbb{R}^{p+1} \times \mathbb{R}^+$  and define  $\Lambda_0 = \{(\alpha, \beta, \sigma, f); (\alpha, \beta, \sigma) \in K, f \in \mathcal{C} \cap \{H(f) > \epsilon\}\}$ . To prove Theorem 3 we verify assumptions **A1** and **A2**. We write  $\lambda = (\alpha, \beta, \sigma, f)$ . Under the Gaussian noise assumption in (3.1), it follows that

$$\begin{aligned} KL(p_{\lambda_0}^n, p_{\lambda}^n) &= \mathbb{E}_{p_{\lambda_0}^n} \left( \frac{\log p_{\lambda_0}^n}{\log p_{\lambda}^n} \right) = \frac{n}{2} \left( \frac{\sigma_0^2}{\sigma^2} - 1 - \log(\sigma_0^2/\sigma^2) \right) + \frac{\|\eta^n - \eta_0^n\|_n^2}{2\sigma^2}, \\ V(p_{\lambda_0}^n, p_{\lambda}^n) &= \text{Var}_{p_{\lambda_0}^n} \left( \frac{\log p_{\lambda_0}^n}{\log p_{\lambda}^n} \right) = n \left( \frac{\sigma_0^2}{\sigma^2} - 1 \right)^2 + \frac{\sigma_0^4}{\sigma^4} \|\eta^n - \eta_0^n\|_n^2, \end{aligned} \quad (3.5)$$

where  $\eta^n = (\alpha + \beta^t d_1 + f(x_1), \dots, \alpha + \beta^t d_n + f(x_n))^t$ , and  $\eta_0^n = (\alpha_0 + \beta_0^t d_1 + f_0(x_1), \dots, \alpha_0 + \beta_0^t d_n + f_0(x_n))^t$ .

Since  $\|\eta^n - \eta_0^n\|_n^2 \leq 2 [\|f^n - f_0^n\|_n^2 + (\gamma - \gamma_0)^t Z^t Z (\gamma - \gamma_0)]$ , we have using Lemma 2, on  $Z$ , that if

$$\|\gamma - \gamma_0\|_{p+1}^2 \leq \epsilon_n^2, \quad \|f^n - f_0^n\|_n^2 \leq n\epsilon_n^2, \quad |\sigma - \sigma_0| \leq \epsilon_n,$$

then there exists  $\tau > 0$  such that  $KL(p_{\lambda_0}^n, p_{\lambda}^n) \leq \tau n\epsilon_n^2$  and  $V(p_{\lambda_0}^n, p_{\lambda}^n) \leq \tau n\epsilon_n^2$ .

Then, similarly to Lemma 1,

$$P_{\lambda_0}^n \left[ J_1^{\lambda_0} < \frac{e^{-2\tau n\epsilon_n^2} \pi_1 (\{KL(p_{\lambda_0}^n, p_{\lambda}^n) \leq \tau n\epsilon_n^2; V(p_{\lambda_0}^n, p_{\lambda}^n) \leq \tau n\epsilon_n^2\})}{2} \right] \leq \frac{2\tau}{n\epsilon_n^2}$$

so that under assumption (3.3),

$$\sup_{\lambda_0 \in \Lambda_0} P_{\lambda_0}^n \left[ J_1^{\lambda_0} < \frac{C e^{-(c+2)n\epsilon_n^2}}{2} \right] \leq \frac{2}{n\epsilon_n^2}$$

and assumption **A1** is satisfied. We now consider assumption **A2**. By definition of  $\Lambda_0$ , **A2-1** is satisfied with  $d_n$  defined through the functional  $H(\lambda)$ . However the metric which is used to construct tests in regression models is often the average of the squares of the Hellinger distances for the  $n$  observations given by

$$d_n^2(p_{\lambda_1}^n, p_{\lambda_2}^n) = \frac{1}{n} \sum_{i=1}^n \int \left( \sqrt{\phi_{\lambda_1}(y_i|w_i)} - \sqrt{\phi_{\lambda_2}(y_i|w_i)} \right)^2 d\mu,$$

where  $\phi_\lambda(y|w)$  is the Gaussian density of  $y$  given  $w = (d, x)$  with mean  $\eta(d, x) = \alpha + \beta^t d + f(x)$  and variance  $\sigma^2$ . Note that obvious calculations lead to

$$d_n^2(p_{\lambda_1}^n, p_{\lambda_2}^n) = 2 - 2 \left(1 - \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1^2 + \sigma_2^2}\right)^{1/2} \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{(\eta_{i1} - \eta_{i2})^2}{4(\sigma_1^2 + \sigma_2^2)}\right), \quad (3.6)$$

where  $\eta_{ij} = \alpha_j + \beta_j^t d_i + f_j(x_i)$ ,  $i = 1, \dots, n$  and  $j = 1, 2$ . We thus have

$$\begin{aligned} d_n^2(p_{\lambda_1}^n, p_{\lambda_2}^n) &\geq 2 - 2 \left(1 - \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1^2 + \sigma_2^2}\right)^{1/2} \exp\left(-\frac{\|\eta_1^n - \eta_2^n\|_n^2}{4n(\sigma_1^2 + \sigma_2^2)}\right) \\ &\leq 2 - 2 \left(1 - \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1^2 + \sigma_2^2}\right)^{1/2} \left(1 - \frac{\|\eta_1^n - \eta_2^n\|_n^2}{4n(\sigma_1^2 + \sigma_2^2)}\right), \end{aligned} \quad (3.7)$$

where  $\eta_j^n = (\eta_{1j}, \dots, \eta_{nj})^t$ ,  $j = 1, 2$ . Let  $\lambda_0 \in \Lambda$  satisfy  $H(\lambda_0) > \epsilon$ ,  $\gamma_n^* = \operatorname{argmin}_\gamma \|Z(\gamma - \gamma_0) - f_0^n\|_n^2$  and  $\gamma^* = \operatorname{argmin}_{\alpha, \beta} \int_{d, x} (\alpha - \alpha_0 - (\beta - \beta_0)d - f_0(x))^2 d\mu(d, x)$ . We can represent

$$\gamma_n^* = (Z^t Z)^{-1} Z^t f_0^n = V_d^{-1} \int (1, d^t)^t f_0(x) d\mu(d, x) + o_p(1) = \gamma^* + o_p(1),$$

where  $V_d$  is the  $(p+1) \times (p+1)$  symmetric matrix whose components are given by  $V_d(1, 1) = 1$ ,  $V_d(1, j) = 0$  for  $j = 2, \dots, p+1$  and  $V_d(i, j) = E(d_{i-1} d_{j-1})$  for  $i, j = 2, \dots, p+1$ .

Let  $\epsilon > 0$  and  $\lambda_0 \in \Lambda_0 = \mathcal{C} \cap \{H(\lambda) > \epsilon\}$ , then

$$\begin{aligned} P_{\lambda_0}^n \left[ \frac{1}{n} \|Z\gamma_n^* - f_0^n\|_n^2 < \epsilon/2 \right] &= P_{\lambda_0}^n \left[ \frac{1}{n} \|Z\gamma^* - f_0^n\|_n^2 < \epsilon/3 \right] + P_{\lambda_0}^n \left[ \frac{1}{n} \|Z \cdot (\gamma^* - \gamma_n^*)\|_n^2 < \epsilon^2 \right] \\ &= P_{\lambda_0}^n \left[ \frac{1}{n} \|Z\gamma^* - f_0^n\|_n^2 < \epsilon/3 \right] + P_{\lambda_0}^n \left[ \frac{1}{n} \|\gamma^* - \gamma_n^*\|_n^2 < C\epsilon^2 \right] \\ &= P_{\lambda_0}^n \left[ \left| \frac{1}{n} \|Z\gamma^* - f_0^n\|_n^2 - H(\lambda_0) \right| > 2\epsilon/3 \right] + o(1) = o(1) \end{aligned} \quad (3.8)$$

Uniformly over  $\Lambda_0$ . Hence assumption **A2-1** is satisfied with  $P_{\lambda_0}^n$  probability going to 1, uniformly in  $\lambda_0 \in \Lambda_0$ .

Let  $\epsilon > 0$  and consider  $a_\epsilon > 0$  defined by CONDITION (P1) and

$$\Theta_n = \{(\alpha, \beta, \sigma); |\alpha| \leq e^{a_\epsilon n}, \|\beta\| \leq e^{a_\epsilon n}, e^{-a_\epsilon n} \leq \sigma \leq e^{a_\epsilon n}\}$$

this set satisfies assumption **A2-2**. We now verify assumption **A2-3**, using results in Birgé (1983), Le Cam (1986), or Lemma 2 of Ghosal and van der Vaart (2007). It follows that there exists a sequence of tests  $\phi_1^n$  such that for all  $\lambda_0 \in \Lambda_0$  and all  $\theta_1 \in \Theta$ ,

$$\begin{aligned} E_{\lambda_0}^n [\phi_1^n] &\leq e^{-nd_n^2(\lambda_0, \theta_1)/2} \\ \sup_{d_n(\theta_1, \theta) \leq d_n(\theta_1, \lambda_0)/18} E_\theta^n [1 - \phi_1^n] &\leq e^{-nd_n^2(\lambda_0, \theta_1)/2}, \end{aligned}$$

where  $d_n$  is the average Hellinger entropy. Then, combining (3.7) with (3.8) we have for  $n$  large enough  $\inf_{\alpha, \beta} \|\eta_1^n - \eta_0^n\|_n^2 \geq \epsilon n/2$  and if  $|\sigma_1 - \sigma_0| \leq 2\sigma_0$  then there exists a constant  $a > 0$  such that  $d_n^2(\lambda_0, \theta_1) > a$ , with  $P_{\lambda_0}^n$  probability going to 1. If  $|\sigma_1 - \sigma_0| > 2\sigma_0$  then direct computations imply that  $d_n^2(\lambda_0, \theta_1) \geq (\sqrt{8} - \sqrt{7})/\sqrt{2}$  and choosing  $a \leq (\sqrt{8} - \sqrt{7})/\sqrt{2}$  we have that  $d_n^2(\lambda_0, \theta_1) > a$ . This leads to

$$E_{\lambda_0}^n [\phi_1^n] \leq e^{-na^2/2}, \quad \sup_{d_n(\theta_1, \theta) \leq a/18} E_{\theta}^n [1 - \phi_1^n] \leq e^{-na^2/2}, \quad (3.9)$$

Let  $t > 0$  and let  $N_n(t, \Theta_n, d_n^2)$  be the  $t$  covering number of  $\Theta_n$  in  $d_n$  norm. Then, the second inequality of (3.7) implies that for all  $\theta_j = (\alpha_j, \beta_j, \sigma_j) \in \Theta_n$  and  $\theta = (\alpha, \beta, \sigma) \in \Theta_n$ , if  $|\alpha - \alpha_j| \leq \tau\sigma_j$ ,  $|\beta - \beta_j| \leq \tau\sigma_j$ ,  $\|\eta^n - \eta_j^n\|_n^2 \leq na\sigma_j^2/(18)^2$ , and  $|\sigma_j - \sigma|^2 \leq a\sigma_j^2/[4(18)^2]$ , by choosing  $\tau$  small enough,

$$d_n^2(p_{\theta}^n, p_{\theta_j}^n) \leq 2 - 2 \left(1 - \frac{(\sigma - \sigma_j)^2}{\sigma^2 + \sigma_j^2}\right)^{1/2} \left(1 - \frac{\|\eta^n - \eta_j^n\|_n^2}{4n(\sigma^2 + \sigma_j^2)}\right),$$

so that  $d_n^2(p_{\theta}^n, p_{\theta_j}^n) \leq a/(18)^2$ . For each subinterval  $(\sigma_j, \sigma_{j+1})$  with  $\sigma_j = e^{-a_\epsilon n}(1 + \sqrt{a_\epsilon}/36)^j$ ,  $j = 0 \dots J_n$  where  $J_n = \lfloor 36[\exp(a_\epsilon n) + a_\epsilon n]/\sqrt{a_\epsilon} \rfloor + 1$ .

we bound the number of intervals in  $\alpha, \beta$  satisfying the above constraint by

$$N_{n,j} \leq 4e^{2a_\epsilon n} \tau^{-2} \sigma_j^{-2} \vee 1$$

Moreover, when let  $j$  be such that  $\sigma_j \geq 2e^{a_\epsilon n}/\tau$ , uniformly in  $\sigma \in (\sigma_j, \sigma_{j+1})$  and  $\alpha^2 + \|\beta\|^2, (\alpha')^2 + \|\beta'\|^2 \leq e^{2a_\epsilon n}$ , with  $\theta = (\alpha, \beta, \sigma_j)$  and  $\theta' = (\alpha', \beta', \sigma)$ ,  $d_n^2(p_{\theta}^n, p_{\theta'}^n) \leq a/(18)^2$ , so that the global covering number of  $\Theta_n$  is bounded by

$$N_n(a/(18)^2, \Theta_n, d_n^2) \leq \frac{e^{4a_\epsilon n}}{\tau^2} \left[ \sum_{j=0}^{\infty} (1 + \sqrt{a_\epsilon}/36)^{-2j} + J_n \right] \leq e^{an/4}, \quad (3.10)$$

if the constant  $a_\epsilon$  in the definition of  $\Theta_n$  is small enough. Finally, combining (3.10) with (3.9), we prove assumption **A2-3** and Theorem 3 is proved.  $\square$

We now turn to studying the consistency of the Bayes factor under the null hypothesis.

### 3.2 Under $\mathcal{M}_0$

Suppose that the true model is  $\mathcal{M}_0$ . That is, the true regression model is assumed to be the parametric linear model,  $y = \alpha_0 + \beta_0^t d + \sigma_0 \epsilon_i$ ,  $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$ ,  $p_{\theta_0}^n \in \mathcal{M}_0$ . Thus, assuming that  $H(f) = 0$ , the parametric vector  $\eta_0^n$  has components equal to  $\eta_{0i} = \alpha_0 + \beta_0^t d_i$ , and we verify Assumptions **B1-B2**. Verification of assumption **B1** is relatively straightforward using (3.5), since if

$$|\sigma - \sigma_0| \leq n^{-1/2}, |\alpha - \alpha_0| \leq n^{-1/2}, \|\beta - \beta_0\|_{\mathbf{p}} \leq n^{-1/2}, \quad \theta = (\alpha, \beta, \sigma),$$

then  $KL(p_{\theta_0}^n, p_{\theta}^n) + V(p_{\theta_0}^n, p_{\theta}^n) \leq C$  for some positive constant  $C$ . The conditions on the prior density  $\pi_0$  implies that assumption **B1** is verified with  $k_0 = p + 2$ . In order to verify assumption **B2** we need to verify that the semiparametric posterior probability  $\pi_1(\cdot | y^n)$  of the  $\epsilon_n$ -shrinkage ball of  $p_{\theta_0}^n$  based on  $d_n$  metric converges to 1, while the semiparametric prior  $\pi_1$  assigns negligible probability on the  $\epsilon_n$ -shrinkage ball. Define

$$A_{u_n} = \{\lambda \in \Lambda : d_n(p_{\lambda}^n, p_{\theta_0}^n) < u_n\}.$$

We now prove that there exists  $u_n$  going to 0, such that

$$\pi_1[A_{u_n} | y^n] = 1 + o_p(1), \quad \pi_1[A_{u_n}] = o(n^{-(p+2)/2}).$$

Recall that

$$d_n^2(p_{\lambda}^n, p_{\theta_0}^n) = 2 - 2 \left(1 - \frac{(\sigma - \sigma_0)^2}{\sigma^2 + \sigma_0^2}\right)^{1/2} \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{(\eta_i - \eta_{i0})^2}{4(\sigma^2 + \sigma_0^2)}\right),$$

where  $\eta_i = \alpha + \beta^t d_i + f(x_i)$ , and  $\eta_{i0} = \alpha_0 + \beta_0^t d_i$ , and that  $d_n^2(p_{\lambda}^n, p_{\theta_0}^n) \leq u_n^2$  implies that  $(\sigma - \sigma_0)^2 \leq C u_n^2$  for some positive  $C$ , regardless of  $\eta$ . Obviously **B2-1** and **B2-2** depends on the chosen prior on  $f$ . We consider the following assumptions on  $\pi_f$  and  $\pi_p$  :

**CONDITION (P1)** (Semiparametric prior  $\pi_1$ ) :

- (i) There exists  $u_n \downarrow 0$ ,  $C_1, c_1, \tau > 0$  and  $\mathcal{F}_{n,1} \subset L^2([0, 1]^q)$  s.t.

$$\pi_f(\|f^n\|_n \leq u_n) \geq C_1 e^{-c_1 n u_n^2}, \quad \pi_f(\mathcal{F}_{n,1}^c) = o(e^{-(2+c_1) n u_n^2} u_n^{p+1}),$$

and if  $\bar{\mathcal{F}}_{n,1} = \{f^n = (f(x_1), \dots, f(x_n))^t, f \in \mathcal{F}_{n,1}\}$

$$\mu[1 \log N(u_n, \bar{\mathcal{F}}_{n,1}, \|\cdot\|_n) > \tau n u_n^2] = o(1)$$

- (ii) For all  $a_1, b > 0$  there exists  $a_2 > 0$  such that for all  $n$  large enough,

$$\pi_p \left[ a_1 u_n \leq \sigma \leq e^{a_2 n u_n^2}; \|\gamma\| \leq e^{a_2 n u_n^2} \right] \geq 1 - e^{b n u_n^2} \quad (3.11)$$

Moreover for all  $\epsilon > 0$ , there exists  $C > 0$  such that if  $w_n = \sqrt{\log(n u_n^2)}$ ,

$$\pi_f [\|f\|_{\infty} > C \sqrt{n} u_n / w_n] + \pi_f [H(f) \leq C u_n^2] < \epsilon \left( \frac{1}{u_n \sqrt{n}} \right)^{p+2}, \quad (3.12)$$

- (iii) For all  $\epsilon > 0$ , there exists  $C > 0$  such that

$$\mu \left[ \pi_f (f^n - P_z f^n \|_n \leq C u_n) > \epsilon \left( \frac{1}{u_n \sqrt{n}} \right)^{p+2} \right] = o(1) \quad (3.13)$$

We then have the following result :

**Theorem 4.** *Under condition (P1), (i) and either (ii) or (iii), the Bayes factor  $B_{01}$  is uniformly increasing to infinity under  $P_{\theta_0}$  uniformly over any compact subset of  $\Theta$ .*

The proof is postponed to the Appendix (Section 5).

**Remark 1.** *Set  $z(d) = (1, d^t)$  and  $\langle z(d), f \rangle = \int_{[-1,1]^p \times [0,1]^q} (1, d^t) f(x) d\mu(d, x)$ , then we can write  $H(f) = \|f - (1, d^t) V_d^{-1} \langle z(d), f \rangle\|^2$ , and note that if  $d$  and  $x$  are independent under  $\mu$ , writing  $\tilde{f} = f - \int_{[0,1]^q} f(x) d\mu(x)$  we obtain that  $H(f) = \|\tilde{f}\|_2^2$ . Here,  $\tilde{f}(x)$  is a centered random quantity of  $f(x)$ , and  $H(f) = 0$  is equivalent to  $\tilde{f}(x) = 0$ , i.e.  $f(x) = \text{constant}$ . Therefore,  $H(f)$  can be regarded as a distance to the null hypothesis as mentioned before.*

**Remark 2.** *Note that condition (ii) implies condition (iii) (see Section 5, Appendix). However, it is quite possible that depending on the families of priors considered, (iii) might be easier to prove than (ii). We consider two examples with Gaussian process priors or hierarchical Gaussian process priors on  $f$  in the following section, in which condition (ii) is easier to prove.*

### 3.3 The case of Gaussian process priors

In this section we assume that CONDITION (P0) is satisfied by the parametric prior  $\pi_0$  under model  $\mathcal{M}_0$  and that CONDITION (P1) is satisfied by the parametric prior  $\pi_p$  under model  $\mathcal{M}_1$ .

In this section we consider two specific examples of the nonparametric prior,  $\pi_f$  for studying the validity of condition (3.3) of Theorem 3 and CONDITION (P1) of Theorem 4. For this purpose, we deal with a family of Gaussian process priors for the nonparametric component  $f$  which is a common family of priors for such models. Throughout this section, we assume that CONDITION (P0) is satisfied by the parametric prior  $\pi_0$  under model  $\mathcal{M}_0$  and that equation 3.11 in CONDITION (P1) is satisfied by the parametric prior  $\pi_p$  under model  $\mathcal{M}_1$ . Consequently, we focus on the nonparametric prior,  $\pi_f$  and investigate the Bayes factor consistency. To be specific, we assume that  $f$  is distributed as a zero-mean Gaussian process prior on  $\mathbb{B} = \mathcal{C}([0, 1])$  the Banach space of continuous functions over  $[0, 1]$  associate with  $\|\cdot\|_\infty$ , under  $\pi_f$ , with reproducing kernel Hilber space  $\mathbb{H}$  (RKHS) and concentration function  $\phi_f$  defined by : for all  $f \in \mathbb{B}$ ,

$$\phi_f(\epsilon) = \inf_{h \in \mathbb{H} : \|h - f_0\|_\infty < \epsilon} \|h\|_{\mathbb{H}}^2 - \log \pi_f[\|f\|_\infty < \epsilon],$$

see van der Vaart and van Zanten (2008) for a more complete discussion of the notion of RKHS and concentration function. Recall that the support  $\mathcal{S}$  of  $\pi_f$  is the closure of  $\mathbb{H} \in \mathbb{B}$ . Hence, for all  $f_0 \in \mathcal{S}$ , there exists  $\epsilon_n$  going to 0 such that  $\phi_{f_0}(\epsilon_n) \leq n\epsilon_n^2$  and using Theorem 2.4 of van der Vaart and van Zanten (2008),

$$\pi_f[\|f - f_0\|_\infty \leq \epsilon_n] \geq e^{-cn\epsilon_n^2}$$

for some  $c > 0$  and condition (3.3) is satisfied. Thus, Theorem 3 is valid and the Bayes factor is consistent under  $\mathcal{M}_1$ .

To verify the consistency under model  $\mathcal{M}_0$  we study the validity of CONDITION (P1) (i) and (ii) equation (3.12) and apply Theorem 4. To do so we need to consider some assumptions on the Gaussian process prior. First note that the constant function equal to 0 is in the support of all zero-mean Gaussian process, so that for all zero mean Gaussian process on  $\mathbb{B}$ , there exists  $u_n$  such that

$$\varphi_0(u_n) = -\log \pi_f(\|f\|_\infty \leq u_n) \leq nu_n^2.$$

Thus using Theorem 2.4 of van der Vaart and van Zanten (2008) CONDITION (P1) (i) of Theorem 4 is verified. Next, we check CONDITION (P1) (ii) of Theorem 4 as follows.

Denote  $m_j(x) = E[d_j|X = x]$ ,  $\Sigma$  to be the expectation of the conditional covariance matrix of  $d$  given  $X$ ,  $\beta(f) = (\beta_1(f), \dots, \beta_p(f))$  with  $\beta_j(f) = \langle m_j, f \rangle$  and assume that  $\Sigma$  is positive definite. Denote also  $\tilde{f} = f - \int_0^1 f(x)d\mu(x)$ .

A Markov inequality implies that for all  $s > p + 2$

$$\pi_f [\|f\|_\infty > \sqrt{n}u_n/w_n] \leq \frac{E[\|f\|_\infty^s]w_n^s}{(\sqrt{n}u_n)^s} = o((\sqrt{n}u_n)^{-(p+2)}).$$

Thus equation (3.12) of CONDITION (P1) (ii) is satisfied if

$$\pi_f [H(f) \leq Cu_n^2] \leq o((\sqrt{n}u_n)^{-(p+2)}), \quad H(f) = \int (f(x) - (1, d^t)V_d^{-1} \langle z(d), f \rangle)^2 dx.$$

Without loss of generality we can assume that  $V_d$  is the identity matrix, then

$$\begin{aligned} H(f) &= \int (\tilde{f} - \sum_{j=1}^p m_j \beta_j(f))^2(x) d\mu(x) + \beta(f)^t \Sigma \beta(f) \\ &\geq \int (\tilde{f} - \sum_{j=1}^p m_j \beta_j(f))^2(x) d\mu(x) + c_0 \|\beta(f)\|^2 \end{aligned}$$

for some positive  $c_0$ . Hence  $H(f) \leq Cu_n^2$  implies that  $\|\beta(f)\|^2 \leq Cc_0^{-1}u_n^2 \lesssim u_n^2$ , which in turns implies that  $\|\tilde{f}\|_2 \lesssim u_n$  and that

$$\pi_f [H(f) \leq Cu_n] \leq \pi_f [\|\tilde{f}\|_2 \leq C'u_n]$$

for some positive  $C'$ .

Therefore, if  $\pi_f$  is the distribution of a zero mean Gaussian process on  $\mathcal{C}([0, 1])$ , (3.12) CONDITION (P1) is verified if and only if

$$\pi_f [\|\tilde{f}\|_2 \leq C'u_n] = o(u_n \sqrt{n})^{-(p+2)}.$$

We now illustrate on how to verify such a condition on two types of Gaussian (or conditionally) Gaussian priors. Note first that the above computations remain valid for conditionally Gaussian process priors.



- Purely Gaussian prior :  $f$  can be represented as an infinite series

$$f = \sum_k \sigma_k Z_k \xi_k, \quad \sigma_k > 0 \quad Z_k \sim \mathcal{N}(0, 1), \quad i.i.d, \quad (3.14)$$

where  $(\xi_k, k \geq 0)$  is an orthonormal basis of  $L^2([0, 1])$ , and each  $\xi_k$  is bounded. If  $\sigma_k \propto k^{-\tau-1/2}$ , following van der Vaart and van Zanten (2008) or Castillo (2008) it can be proved that  $u_n \lesssim n^{-\tau/(2\tau+1)} \log n$ . To simplify the presentation we assume that  $1 = \xi_0$ .

- Hierarchical Gaussian prior, that is  $f$  can be represented as a truncated Gaussian with random truncation :

$$f = \sum_{k=0}^K \sigma_k Z_k \xi_k, \quad \sigma_k > 0 \quad Z_k \sim \mathcal{N}(0, 1), \quad i.i.d \quad (3.15)$$

and  $K$  is random and is distributed according to some probability on  $\mathbb{N}^*$ , which we assume to satisfy

$$e^{-a_1 k L(k)} \lesssim P(K = k) \lesssim e^{-a_2 k L(k)}$$

where  $L(k)$  is either equal to 1 (as in the Hypergeometric distribution ) or  $\log k$  (as in the Poisson distribution). We also assume that  $\sigma_k \propto k^{-\tau-1/2}$ .

In the case of the Purely Gaussian prior distribution, then  $\tilde{f} = \sum_{k=1}^{\infty} \sigma_k Z_k \xi_k$

$$\begin{aligned} \pi_f \left[ \|\tilde{f}\|_2 \leq C' u_n \right] &= P \left[ \sum_{k=1}^{\infty} \sigma_k^2 Z_k^2 \leq C u_n^2 \right] \\ &\leq e^{-A n u_n^2} = o \left( (\sqrt{n} u_n)^{-(p+2)} \right). \end{aligned}$$

for some  $A > 0$ .

In the case of the hierarchical Gaussian prior, then following Arbel (2012) it can be proved that  $u_n \lesssim \sqrt{\log n/n}$  and

$$\begin{aligned} \pi_f \left[ \|\tilde{f}\|_2 \leq C' u_n \right] &= \sum_{k=1}^{\infty} P[K = \kappa] P_r \left[ \sum_{k=\kappa}^{\infty} \sigma_k^2 Z_k^2 \leq C u_n^2 \right] \\ &\lesssim u_n = o \left( (\sqrt{n} u_n)^{p+2} \right). \end{aligned}$$

Hence, the Bayes factor is consistent under  $\mathcal{M}_0$  when the Gaussian process prior  $\pi_f$  is constructed with both Gaussian-type priors in (3.14) and (3.15).

## 4 Discussion

In this paper, we investigated the consistency of the Bayes factor for independent but non identically distributed observations. In particular, we considered a

uniform version of consistency of the Bayes factor and discussed general results on the Bayes factor consistency. Then we specialized our results to the model selection problem in the context of partially linear regression model, in which the regression function is assumed to be the additive form of the linear component and the nonparametric component. Specifically, sufficient conditions to ensure Bayes factor consistency were given for choosing between the parametric model and the semiparametric alternative in the partially linear regression model. These results extend the work of McVinish et al. (2009) to the non-IID observations and complement their work in the context of Bayesian lack of fit testing for partially linear models. The main challenge was to deal with the prior probabilities when the true model is a parametric regression model and in particular to lower bound prior probability mass of neighbourhoods of the true model. Here two commonly used family of Gaussian type priors for the nonparametric component in the partially linear model were shown to satisfy the required conditions and thus illustrated validity of sufficient conditions we presented. Our computations generalize very easily to other families of priors up to the condition

$$\pi_f \left[ \|\tilde{f}\|_2 \leq C' u_n \right] = o(u_n \sqrt{n})^{-(p+2)},$$

which can be checked on a case by case basis. The computations we proposed for the priors 3.15 should be relatively easy to extend to other families of priors based basis expansions such as orthogonal priors with wavelets or Legendre polynomials. The investigation of this inequality nonorthogonal priors with spline bases or mixture priors discussed in de Jonge and van Zanten (2010) would be of interest.

## acknowledgements

This work is partially supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy 75012 Paris) through the 2010–2013 project *Bandhits*.

## 5 Appendix : Proof of Theorem 4

First we show that assumption (i) implies that  $\pi_1 [A_{Mu_n} | y^n] = 1 + o_p(1)$ , for some  $M > 0$ . Using (3.5), there exists  $M_0$  such that if  $\lambda = (\sigma, \gamma, f)$  satisfies

$$\|f^n\|_n \leq \sigma_0 u_n / 2, \quad (\gamma - \gamma_0)^t Z^t Z (\gamma - \gamma_0) \leq u_n \sigma_0 / 2 \quad 0 < \sigma - \sigma_0 < \epsilon u_n$$

for some  $\epsilon > 0$ , then

$$\lambda \in S_n = \{\lambda : KL(p_{\lambda_0}^n, p_\lambda^n) \leq nu_n^2, \quad V(p_{\lambda_0}^n, p_\lambda^n) \leq M_0 nu_n^2\}.$$

Assumption (i) implies that with probability going to 1,  $\pi_1(S_n) \gtrsim C_1 e^{-c_1 nu_n^2} u_n^{p+1}$ , so that

$$J_1^{\lambda_0}(y^n) \gtrsim e^{-(2+c_1)nu_n^2} u_n^{p+1} \tag{5.1}$$

with probability going to 1. Moreover define

$$\mathcal{F}_n(a_1, a_2) = \{(\gamma, \sigma, f); \|\gamma\| \leq e^{a_2 n u_n^2}, a_1 u_n \leq \sigma \leq e^{a_2 n u_n^2}, f \in \mathcal{F}_{n,1}\}$$

then under (i), we have that for all  $a > 0$  there exists  $c_0 > 0$  with  $\pi(\mathcal{F}_n^c(a_1, a_2)) = o(e^{-(2+c_1) n u_n^2 u_n^{p+1}})$ . Let  $\epsilon_0 > 0$  and define  $S_{n,1} = \{\lambda; d_n^2(p_\lambda^n, p_{\theta_0}^n) \leq \epsilon_0^2\}$  and  $S_{n,2} = \{\lambda; d_n^2(p_\lambda^n, p_{\theta_0}^n) > \epsilon_0^2\}$ . Then  $\lambda \in S_{n,1}$  implies that there exists  $C > 0$  such that

$$|\sigma - \sigma_0| \leq \sigma_0/2, \quad \text{and} \quad \|\eta^n - \eta_0^n\|_n \leq C d_n(p_\lambda^n, p_{\theta_0}^n).$$

Thus if  $\lambda, \lambda' \in S_{n,1}$ , with  $\lambda = (\sigma, \gamma, f)$  and  $\lambda' = (\sigma', \gamma', f')$  and

$$|\sigma - \sigma'| \leq u_n, \quad \|\gamma - \gamma'\| \leq u_n, \quad \|f^n - f'^n\|_n \leq u_n$$

then there exists  $\rho > 0$  such that  $d_n(p_\lambda^n, p_{\lambda'}^n) \leq \rho u_n$ . Therefore, it follows that there exists  $\rho$  such that

$$\begin{aligned} \log N(\rho u_n, \mathcal{F}_n \cap S_{n,1}, d_n) &\lesssim n u_n^2 + \log N(u_n, \mathcal{F}_{n,1}, \|\cdot\|_n) \\ &\lesssim n u_n^2, \end{aligned}$$

with probability going to 1. Next, suppose that  $\lambda \in S_{n,2}$  and  $\sigma \in (\sigma_0/2, 3\sigma_0/2)$ . Then using the same computations as before,

$$\begin{aligned} &\log N(\epsilon_0/18, \mathcal{F}_n \cap S_{n,2} \cap \{\sigma \in (\sigma_0/2, 3\sigma_0/2)\}, d_n) \\ &\lesssim n u_n^2 + \log N(u_n, \mathcal{F}_{n,1}, \|\cdot\|_n) = o(n) \end{aligned}$$

Set  $\sigma_n = a_1 u_n$  and  $\bar{\sigma}_n = e^{a_2 n u_n^2}$ . We consider separately the cases  $\sigma \in (\sigma_n, \sigma_0/2)$  or  $\sigma \in (3\sigma_0/2, \bar{\sigma}_n)$ . In the former case, we define  $\sigma_j = \sigma_n(1 + a_1)^j$  with  $a_1 > 0$  chosen as small as needed be, and  $j = 1, \dots, J_{n,1}$  where  $J_{n,1} = \lfloor a_1^{-1} \log(\sigma_0/(2\sigma_n)) \rfloor + 1$ . For each  $j \leq J_{n,1}$ , for all  $\sigma, \sigma' \in (\sigma_j, \sigma_{j+1})$ , all  $\gamma, \gamma'$  such that  $\|\gamma - \gamma'\| \leq \rho \sigma_j$  and all  $f, f'$  such that  $\|f^n - (f')^n\|_n \leq a_1^{-1} \sigma_j$  with  $\rho$  small enough and  $a_1$  large enough,  $d_n(p_\lambda^n, p_{\lambda'}^n) \leq \epsilon_0/18$ . Thus, based on the similar derivation of covering number to the case of (3.10), we have that

$$\log N(\epsilon_0/18, \mathcal{F}_n \cap \{\sigma \in (\sigma_n, \sigma_0/2)\}, d_n) = o(n).$$

For the latter, we have the covering number in the same manner as before,

$$\log N(\epsilon_0/36, \mathcal{F}_n \cap \{\sigma \in (3\sigma_0/2, \bar{\sigma}_n)\}, d_n) = o(n).$$

Finally this implies that  $\pi[A_{M u_n} | y^n] = 1 + o_{p_{\theta_0}^n}(1)$ , uniformly over all compact  $K \subset \mathbb{R}^{p+1} \times \mathbb{R}^{+*}$ , for some  $M > 0$ .

Now we bound from above  $\pi_1[A_{M u_n}]$ . We first note that  $\lambda \in A_{M u_n}$  implies that there exist  $\tau_1, \tau_2 > 0$  such that  $|\sigma - \sigma_0| \leq \tau_1 u_n$  and  $\|\eta^n - \eta_0^n\|_n^2 \leq \tau_2 u_n^2$ . Thus, it follows that

$$\begin{aligned} \pi_1(A_{M u_n}) &\lesssim \pi_1[|\sigma - \sigma_0| \leq \tau_1 u_n] \cap \{\|\eta^n - \eta_0^n\|_n^2 \leq \tau_2 u_n^2\} \\ &\lesssim u_n \pi_1[\|\eta^n - \eta_0^n\|_n^2 \leq \tau_2 n u_n^2]. \end{aligned}$$

Let  $P_z$  be the projection operator from  $\mathbb{R}^n$  onto the vector space spanned by  $Z$ , we then obtain

$$\begin{aligned}\|\eta^n - \eta_0^n\|_n^2 &= \|Z(\gamma - \gamma_0) - P_z f^n\|_n^2 + \|f^n - P_z f^n\|_n^2 \\ &\gtrsim \|\gamma - \gamma_0 - (Z^t Z)^{-1} Z^t f^n\|^2 + \|f^n - P_z f^n\|_n^2.\end{aligned}\quad (5.2)$$

So that

$$\pi_1 [\|\eta^n - \eta_0^n\|_n^2 \leq \tau_2 u_n^2] \lesssim u_n^{p+1} \pi_f [\|f^n - P_z f^n\|_n^2 \leq \tau_3 u_n^2]$$

for some  $\tau_3 > 0$ .

Thus, to prove that for all  $\epsilon > 0$ ,  $\pi_1 [A_{Mu_n}] < \epsilon n^{-(p+2)/2}$  with probability going to 1, it is enough to prove that

$$\mu \left( \pi_f [\|f^n - P_z f^n\|_n^2 \leq \tau_3 u_n^2] > \epsilon (nu_n^2)^{-(p+2)/2} \right) = o(1)$$

Condition (iii) implies the above result which terminates the proof. We now prove that condition (ii) implies condition (iii).

Condition (ii) (3.12) implies that we can restrict ourselves to the set  $\{\|f^n - P_z f^n\|_n^2 \leq \tau_3 u_n^2\} \cap \{f; \|f\|_\infty \leq C\sqrt{nu_n}\} \cap \{f; \|f\|_2 \leq C\sqrt{nu_n}/w_n\}$  and for the sake of simplicity we still call this set  $A_n$ . Moreover, let  $\Omega_{n,1}(B) = \{Z; \|Z^t Z/n - V_d\| \leq B/\sqrt{n}\}$ , then for all  $\epsilon > 0$  there exists  $B > 0$  such that  $\mu [\Omega_{n,1}(B)^c] \leq \epsilon$ . Since

$$\|f^n - P_z f^n\|_n^2 \geq (\|f^n - n^{-1} Z V_d^{-1} Z^t f^n\|_n - \|(P_z - n^{-1} Z V_d^{-1} Z^t) f^n\|_n)^2. \quad (5.3)$$

and when  $Z \in \Omega_{n,1}(B)$ ,

$$\begin{aligned}\|(P_z - n^{-1} Z V_d^{-1} Z^t) f^n\|_n^2 &\lesssim n^{-2} \|Z[(Z^t Z)/n - V_d] Z^t f^n\|_n^2 \\ &\leq \frac{B^2}{n} \left( \frac{\sum_{i=1}^n \|Z_i\|^2}{n} \right)^2 \|f^n\|_n^2 \\ &\lesssim \frac{\|f^n\|_n^2}{n} \lesssim \frac{\|f\|_\infty^2}{n} = o(u_n^2)\end{aligned}$$

if  $f \in A_n$ . Hence, if  $Z \in \Omega_{n,1}(B)$ , there exists  $C > 0$  such that for all  $f \in A_n$ ,  $\|f^n - n^{-1} Z V_d^{-1} Z^t f^n\|_n^2 \leq C u_n^2$ . We also have for all  $A > 0$ , using a Markov inequality,

$$\begin{aligned}&\mu \left[ \pi_f \left( A_n \cap \left\{ \left\| Z V_d^{-1} \left( \frac{Z^t f^n}{n} - \langle z(d), f \rangle \right) \right\|_n^2 > A u_n^2 \right\} \right) > \epsilon (\sqrt{nu_n})^{-(p+2)} \right] \\ &\leq \frac{(\sqrt{nu_n})^{(p+2)}}{\epsilon} \int_{A_n} \mu \left[ \left\| Z V_d^{-1} \left( \frac{Z^t f^n}{n} - \langle z(d), f \rangle \right) \right\|_n^2 > A u_n^2 \right] d\pi_f(f)\end{aligned}$$

Note that for  $Z \in \Omega_n(B)$ ,

$$\left\| Z V_d^{-1} \left( \frac{Z^t f^n}{n} - \langle z(d), f \rangle \right) \right\|_n^2 \lesssim \sum_{j=1}^{p+1} \left( \sum_{i=1}^n \frac{Z_{ij} f(x_i)}{n} - \langle z_j(d), f \rangle \right)^2$$

where  $z_j(d) = 1$  if  $j = 1$  and  $z_j(d) = d_{j-1}$  if  $j > 1$ . Set  $\sigma_{j,f}^2 = \int d_{j-1}^2 f(x)^2 d\mu(d_j, x)$  if  $j \geq 2$  and  $\sigma_{1,f}^2 = \int f(x)^2 d\mu(x)$ , then we can write

$$\mu \left[ \sum_{i=1}^n (Z_{ij} f(x_i) - \langle z_j(d), f \rangle) > \sqrt{A} n u_n \right] = \mu \left[ \sum_{i=1}^n \frac{(Z_{ij} f(x_i) - \langle z_j(d), f \rangle)}{\sqrt{n} \sigma_{j,f}} > \frac{\sqrt{A} n u_n}{\sigma_{j,f}} \right].$$

Since for all  $f \in A_n$ ,  $Z_{ij} f(x_i)$  is bounded by  $\|f\|_\infty$ , Kolmogorov's exponential inequality see Shorack and A. (1986) p 855, implies that

$$\begin{aligned} \mu \left[ \sum_{i=1}^n (Z_{ij} f(x_i) - \langle z_j(d), f \rangle) > \sqrt{A} n u_n \right] &\leq e^{-\frac{A n u_n^2}{4 \|f\|_\infty^2}} \quad \text{if } \sigma_{j,f}^2 \geq \sqrt{A} u_n \|f\|_\infty \\ &\leq e^{-\frac{\sqrt{A} n u_n}{4 \|f\|_\infty}} \quad \text{if } \sigma_{j,f}^2 < \sqrt{A} u_n \|f\|_\infty \end{aligned}$$

In both cases the right hand side is  $o(e^{-A w_n^2/4C}) = o((\sqrt{n} u_n)^{-(p+2)})$  on  $A_n$  by choosing  $A$  large enough. We finally obtain that there exists  $c > 0$ , such that

$$\pi_f [A_n] \leq \pi_f [\|f^n - ZV_d^{-1} \langle z(d), f \rangle\|_n^2 \leq c u_n^2] + o_p((\sqrt{n} u_n)^{-(p+2)/2}).$$

Consider  $f \in A_n$  with  $H(f) > 2c u_n^2$ , then

$$\begin{aligned} &\mu \left[ \pi_f [\{\|f^n - ZV_d^{-1} \langle z(d), f \rangle\|_n^2 \leq c u_n^2\} \cap \{H(f) > 2c u_n^2\}] > \epsilon (\sqrt{n} u_n)^{-(p+2)} \right] \\ &\leq \frac{(\sqrt{n} u_n)^{(p+2)}}{\epsilon} \int_{A_n} \mu (\{-\|f^n - ZV_d^{-1} \langle z(d), f \rangle\|_n^2 + H(f) > (H(f)/2 + c u_n^2)/2\}) d\pi_f(f) \end{aligned}$$

Using Kolmogorov inequality, if  $H(f) \geq 2c u_n^2$ , and since  $(f(x_i) - Z_i V_d^{-1} \langle z(d), f \rangle)^2 \leq c_0 \|f\|_\infty^2$ , for some positive  $c_0$ ,

$$\begin{aligned} \mu (\{-\|f^n - ZV_d^{-1} \langle z(d), f \rangle\|_n^2 + H(f) > H(f)/4\}) &\leq \exp \left( -\frac{c C n H(f)^2}{w^2(f)} \right) \quad \text{if } H(f) \leq \frac{16 w^2(f)}{c_0 \|f\|_\infty^2} \\ &\leq \exp \left( -\frac{c C n H(f)^2}{\|f\|_\infty^2} \right) \quad \text{if } H(f) > \frac{16 w^2(f)}{\|f\|_\infty^2} \end{aligned}$$

for some positive  $C$ , where  $w^2(f) = \int (f(x) - z(d) V_d^{-1} \langle z(d), f \rangle)^4 d\mu(d, x) - H(f)^2$ . In both cases the right hand term is bounded by  $e^{-c C w_n^2} = o((\sqrt{n} u_n)^{-(p+2)})$  by choosing  $c$  large enough.

## References

- Arbel, J. (2012). Bayesian optimal adaptive estimation using a sieve prior.
- Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *J. Amer. Statist. Assoc.*, 91(433):109–122.

- Birgé, L. (1983). Robust testing for independent nonidentically distributed variables and Markov chains. In *Specifying statistical models (Louvain-la-Neuve, 1981)*, volume 16 of *Lecture Notes in Statist.*, pages 134–162. Springer, New York.
- Casella, G., Girón, F. J., Martínez, M. L., and Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *Ann. Statist.*, 37(3):1207–1228.
- Castillo, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299.
- Choi, T., Lee, J., and Roy, A. (2009). A note on the Bayes factor in a semi-parametric regression model. *J. Multivariate Anal.*, 100(6):1316–1327.
- Dass, S. C. and Lee, J. (2004). A note on the consistency of Bayes factors for testing point null versus non-parametric alternatives. *J. Statist. Plann. Inference*, 119(1):143–152.
- de Jonge, R. and van Zanten, J. H. (2010). Adaptive nonparametric bayesian inference using location-scale mixture priors. *Ann. Statist.*, 38(6):3300–3320.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533.
- Ghosal, S., Lember, J., and van der Vaart, A. (2008). Nonparametric Bayesian model selection and averaging. *Electron. J. Stat.*, 2:63–89.
- Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.*, 35(1):192–223.
- Jeffreys, H. (1961). *Theory of probability*. Third edition. Clarendon Press, Oxford.
- Kass, R. E. and Raftery, A. (1995). Bayes factors. *J. Amer. Statist. Assoc.*, 90:773–795.
- Ko, K., Qu, L., and Vannucci, M. (2009). Wavelet-based Bayesian estimation of partially linear regression models with long memory errors. *Statist. Sinica*, 19(4):1463–1478.
- Koop, G. and Poirier, D. J. (2004). Bayesian variants of some classical semi-parametric regression techniques. *J. Econometrics*, 123(2):259–282.
- Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York.
- Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a Fourier representation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(4):863–879.

- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of  $g$  priors for Bayesian variable selection. *J. Amer. Statist. Assoc.*, 103(481):410–423.
- McVinish, R., Rousseau, J., and Mengersen, K. (2009). Bayesian goodness of fit testing with mixtures of triangular distributions. *Scand. J. Stat.*, 36(2):337–354.
- Moreno, E., Girón, F. J., and Casella, G. (2010). Consistency of objective Bayes factors as the model dimension grows. *Ann. Statist.*, 38(4):1937–1952.
- Shang, Z. and Clayton, M. K. (2011). Consistency of Bayesian linear model selection with a growing number of parameters. *J. Statist. Plann. Inference*, 141(11):3463–3474.
- Shorack, Galen, R. and A., W. J. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
- van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, 36(3):1435–1463.
- Verdinelli, I. and Wasserman, L. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *Ann. Statist.*, 26(4):1215–1241.
- Walker, S., Damien, P., and Lenk, P. (2004). On priors with a Kullback-Leibler property. *J. Amer. Statist. Assoc.*, 99(466):404–408.
- Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*, 92(4):937–950.