NIH Public Access
Author Manuscript
*J Am Stat Assoc*. Author manuscript; available in PMC 2015 March 01.

Published in final edited form as:
*J Am Stat Assoc*. 2014 March 1; 109(505): 437–447. doi:10.1080/01621459.2014.881153.

# Bayes variable selection in semiparametric linear models

**Suprateek Kundu**[1] and **David B. Dunson**[2]

Suprateek Kundu: sk@stat.tamu.edu; David B. Dunson: dunson@stat.duke.edu

[1]Postdoctoral Research Associate in the Dept. of Statistics, Texas A&M University, College Station, TX 77843, USA

[2]Arts & Sciences Distinguished Professor in Dept. Statistical Science, Duke University, Durham, NC 27708, USA

## Abstract

There is a rich literature on Bayesian variable selection for parametric models. Our focus is on generalizing methods and asymptotic theory established for mixtures of *g*-priors to semiparametric linear regression models having unknown residual densities. Using a Dirichlet process location mixture for the residual density, we propose a semiparametric *g*-prior which incorporates an unknown matrix of cluster allocation indicators. For this class of priors, posterior computation can proceed via a straightforward stochastic search variable selection algorithm. In addition, Bayes factor and variable selection consistency is shown to result under a class of proper priors on *g* even when the number of candidate predictors *p* is allowed to increase much faster than sample size *n*, while making sparsity assumptions on the true model size.

### Keywords

Asymptotic theory; Bayes factor; *g*-prior; Large *p*; small *n*; Model selection; Posterior consistency; Subset selection; Stochastic search variable selection

## 1. INTRODUCTION

Bayesian variable selection is widely applied, with O'Hara and Sillanpää providing a recent review (2009). There is a rich literature proposing variable selection methods and studying asymptotic properties for parametric models, while our focus is variable selection in semiparametric linear regression models of the form:

$$\mathbf{Y}^n = \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}, \quad \varepsilon_i \sim f, \quad (1)$$

where $\mathbf{Y}^n$ is $n \times 1$, $\boldsymbol{\gamma} = \{\gamma_j, j = 1, \ldots, p\} \in \Gamma$, $\gamma_j = 1$ if the $j$th candidate predictor is included in the model with $\gamma_j = 0$ otherwise, $\Gamma$ is the set of all possible subsets that are given non-zero prior probability, $p_\gamma = \sum_{j=1}^{p} \gamma_j$ is the size of model $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_\gamma$ is the $p_{\boldsymbol{\gamma}} \times 1$ vector of regression coefficients, $\mathbf{X}_{\boldsymbol{\gamma}}$ is the $n \times p_{\boldsymbol{\gamma}}$ design matrix containing the predictors in model $\boldsymbol{\gamma}$, and $f$ is an unknown residual density. Our focus is on avoiding parametric assumptions on $f$, while accommodating high-dimensional settings in which the number of candidate predictors $p$ can be much larger than the sample size $n$ but $\Gamma$ is restricted to sparse models having $p_{\boldsymbol{\gamma}} < n$.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

There has been limited consideration of variable selection in semiparametric Bayesian models, with essentially no results on asymptotic properties. In particular, it would be appealing to provide a computationally efficient procedure for Bayesian variable selection based on (1) for which it can be shown that the posterior probability on the true model converges to one as $n \to \infty$ even when the number of candidate predictors increases much faster than $n$. In order for the asymptotic analysis to reflect the high dimensionality, it is important to allow $p$ to grow with $n$. There has been some consideration of increasing $p$ asymptotics in Bayesian parametric models. Castillo and van der Vaart (2012) study concentration of the posterior distribution in the normal means problem. Armagan et al. (2013) provide conditions for consistency in high-dimensional normal linear regression with shrinkage priors on the coefficients. Jiang (2007) studies convergence rates of the predictive distribution resulting from Bayesian model averaging in generalized linear models with high-dimensional predictors. These approaches do not consider consistency of model selection or semiparametric settings.

This article proposes a practical, useful and general methodology for Bayesian variable selection in semiparametric linear models (1), while providing basic theoretical support by showing Bayes factor and variable selection consistency. We also extend our approach and theory to increasing model dimensions involving $p \gg n$ candidate predictors while making sparsity assumptions on the true model. Our approach relies on placing a Dirichlet process (DP, Ferguson, 1972) location mixture of Gaussians (Lo, 1984) prior on the residual density $f$, inducing clustering of subjects. We introduce a prior on the coefficients $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ specific to each model $\boldsymbol{\gamma}$, which generalizes mixtures of $g$-priors (Zellner and Siow, 1980; Liang et al., 2008) to include cluster allocation indices induced through the Dirichlet process. The formulation leads to a straightforward implementation via a stochastic search variable selection (SSVS) algorithm (George and McCulloch, 1997).

Section 2 develops the proposed framework. Section 3 considers asymptotic properties. Section 4 contains simulation results. Section 5 applies the approach to a type 2 diabetes data example, and the proofs of Theorems are contained in the Appendix.

## 2. MIXTURES OF SEMIPARAMETRIC *g*-PRIORS

### 2.1 Model Formulation

In this section, we propose a new class of priors for Bayesian variable selection in linear regression models with an unknown residual density characterized via a Dirichlet process (DP) location mixture of Gaussians. In particular, let

$$\begin{aligned} y_i &= \mathbf{x}_{\boldsymbol{\gamma},i}'\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \varepsilon_i, \quad \varepsilon_i \sim f, \quad i = 1, \ldots, n, \\ f(\cdot) &= \int N(\cdot; \alpha, \tau^{-1}) dP(\alpha), \quad P \sim DP(mP_0), \quad P_0 = N(0, \tau^{-1}), \end{aligned} \tag{2}$$

where $\mathbf{x}_{\boldsymbol{\gamma},i}$ is the $i$th row of $\mathbf{X}_{\boldsymbol{\gamma}}$ and does not include an intercept as we do not restrict $f$ to have zero mean, and $f$ is a density with respect to Lebesgue measure on $\Re$. We address uncertainty in subset selection by placing a prior on $\boldsymbol{\gamma}$, while the prior on $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ characterizes prior knowledge of the size of the coefficients for the selected predictors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The DP mixture prior on the density *f* induces clustering of the *n* subjects into *k* groups/subclusters, where *k* is random and each group has a distinct intercept in the linear regression model. Let **A** denote an $n \times k$ allocation matrix, with $\mathbf{A}_{ij} = 1$ if the *i*th subject is allocated to the *j*th cluster and 0 otherwise. The *j*th column of **A** then sums to $n_j$, the number of subjects allocated to subcluster *j*, with $\sum_{j=1}^{k} n_j = n$. Following Kyung, Gill and Casella (2009), conditionally on the allocation matrix **A**, (2) can be represented as a linear model with random intercepts

$$\mathbf{Y}^n = \mathbf{A}\boldsymbol{\eta} + \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}, \quad \boldsymbol{\eta} \sim N(0, \tau^{-1}\mathbf{I}_k), \quad \boldsymbol{\varepsilon} \sim N(0, \tau^{-1}\mathbf{I}_n), \quad (3)$$

where **A** is random with a certain prior probability given by the coefficients in the summation of the likelihood expression (8) and the response and predictors are centered prior to analysis. In the special case in which $\mathbf{A} = \mathbf{1}_n$, the model reduces to a linear regression model with a common intercept $\eta$ and Gaussian residuals. In this case, the conditional posterior for $\eta$ given $\mathbf{A} = \mathbf{1}_n$ is $N\left(\frac{1}{1+n}\sum_{i=1}^{n}(y_i - \mathbf{x}_{\gamma,i}\boldsymbol{\beta}_\gamma), \sqrt{\frac{1}{\tau(1+n)}}\right)$, which has realizations increasingly concentrated at zero as *n* increases.

We would like the prior on the regression coefficients to retain the essential elements of Zellner's *g*-prior (Zellner, 1986), while being suitably adapted to the semiparametric case. To this effect, we propose a mixture of semi-parametric *g*-priors constructed to scale the covariance matrix in Zellner's *g*-prior to reflect the clustering phenomenon as follows:

$$\pi(\boldsymbol{\beta}_\gamma) = N(0, g\tau^{-1}(\mathbf{X}_\gamma' \textstyle\sum_A^{-1} \mathbf{X}_\gamma)^{-1}), \quad \textstyle\sum_A = \mathbf{I}_n + \mathbf{A}\mathbf{A}', \quad g \sim \pi(g). \quad (4)$$

Prior (4) inherits advantages of previous mixtures of *g*-priors including computational efficiency in computing marginal likelihoods (conditional on **A**) and robustness to mis-specification of *g*. The prior can be interpreted as having arisen from the analysis of a conceptual sample generated using a scaled design matrix $\sum_A^{-1/2}\mathbf{X}_\gamma$, reflecting the clustering phenomenon due to the DP kernel mixture prior. Moreover, the proposed prior leads to Bayes factor and variable selection consistency in semi-parametric linear models (2) as we will show.

Note that $(\mathbf{X}_\gamma' \sum_A^{-1} \mathbf{X}_\gamma) = (\mathbf{X}_\gamma'\mathbf{X}_\gamma) - \mathbf{X}_\gamma'\mathbf{A}(\mathbf{I}_k + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}_\gamma < (\mathbf{X}_\gamma'\mathbf{X}_\gamma)$, so $(\mathbf{X}_\gamma' \sum_A^{-1} \mathbf{X}_\gamma)^{-1} > (\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}$, implying that the prior variance of $\boldsymbol{\beta}_\gamma$ conditional on $(g, \tau)$ is higher for the semi-parametric *g*–prior as compared to the traditional *g*–prior for any allocation matrix **A**. To assess the influence of **A** on the prior for $\boldsymbol{\beta}_\gamma$, we did simulations which revealed that for fixed $(n, p)$, $\text{var}(\beta_{\gamma l})$ increases but the $\text{cov}(\beta_{\gamma l}, \beta_{\gamma l'})$ decreases as the number of underlying subclusters in the data increase ($l'$, $l = 1, \ldots, p$, $l' \neq l$). This suggests that as the number of groups in **A** increase, the components of $\boldsymbol{\beta}_\gamma$ are likely to be more dispersed with decreasing association between each other.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

### 2.2 Bayes Factor in Semiparametric Linear Models

In studying asymptotic properties of our proposed approach, we follow standard practice in Bayesian model selection, and assume that the data $\mathbf{Y}^n = (y_1, \ldots, y_n)'$ arise from one of the models in the list under consideration. This true model is denoted $\mathcal{M}_1$ as defined in equation (5). For pairwise comparison, we evaluate the evidence in favor of $\mathcal{M}_1$ compared to an alternative model $\mathcal{M}_2$ using the Bayes factor, where

$$\begin{gathered}\mathscr{M}_1: \boldsymbol{Y}^n = \mathbf{X}_{\gamma_1}\boldsymbol{\beta}_{\gamma_1} + \boldsymbol{\varepsilon}_1, \quad \varepsilon_{1i} \sim f \\ \mathscr{M}_2: \boldsymbol{Y}^n = \mathbf{X}_{\gamma_2}\boldsymbol{\beta}_{\gamma_2} + \boldsymbol{\varepsilon}_2, \quad \varepsilon_{2i} \sim f \\ f(\cdot) = \int N(\cdot; \alpha, \tau^{-1}) dP(\alpha), \quad P \sim DP(mP_0), \quad P_0 = N(0, \tau^{-1}) \\ \boldsymbol{\beta}_{\gamma_j} \sim \pi(\boldsymbol{\beta}_{\gamma_j}), j = 1, 2, \quad \pi(\tau^{-1}) \propto 1/\tau^{-1}, \quad g \sim \pi(g),\end{gathered} \tag{5}$$

where $\gamma_j \in \Gamma$ indexes models of dimension $p_j$ and $\pi(\boldsymbol{\beta}_{\gamma j})$ is defined in (4), $j = 1, 2$. Our prior specification philosophy is similar to the one adopted by Guo and Speckman (2009) for normal linear models, in that we assign proper priors on all elements of both $\boldsymbol{\beta}_{\gamma 1}, \boldsymbol{\beta}_{\gamma 2}$ conditional on $(g, \tau^{-1})$, and an improper prior on $\tau^{-1}$ for a more objective assessment. However unlike Guo and Speckman (2009), our focus is on Bayesian variable selection in semi-parametric linear models.

Note that the conditional likelihood of the response after marginalizing out $\boldsymbol{\eta}$ in (3) is $L(\mathbf{Y}^n | \mathbf{A}, \boldsymbol{\beta}_\gamma, \tau^{-1}) = N(\mathbf{X}_\gamma \boldsymbol{\beta}_\gamma, \tau^{-1}\boldsymbol{\Sigma}_A)$ (Kyung et. al., 2009). Thus conditional on $\mathbf{A}$ and under the DP mixture of Gaussians prior on $f$, $\mathcal{M}_j$ in (5) reduces to the normal linear model:

$$\sum\nolimits_A^{-1/2} \boldsymbol{Y}^n = \boldsymbol{Z}_A = \tilde{\mathbf{X}}_{A,\gamma_j} \boldsymbol{\beta}_{\gamma_j} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \tau^{-1}\mathbf{I}_n), \quad \pi(\boldsymbol{\beta}_{\gamma_j}) = N(0, g\tau^{-1}(\tilde{\mathbf{X}}'_{A,\gamma_j}\tilde{\mathbf{X}}_{A,\gamma_j})^{-1}), \tag{6}$$

where $\tilde{\mathbf{X}}_{A,\gamma_j} = \sum\nolimits_A^{-1/2} \mathbf{X}_{\gamma_j}$. Under a mixture of semi-parametric $g$-priors, we can directly use expression (17) in Guo and Speckman (2009) to obtain (conditional on $\mathbf{A}$) for $j = 1, 2$

$$L(\boldsymbol{Z}_A | \mathscr{M}_j) \equiv L(\boldsymbol{Y}^n | \mathbf{A}, \mathscr{M}_j) \propto (\boldsymbol{Z}'_A \boldsymbol{Z}_A)^{-n/2} \int_0^\infty (1+g)^{-p_j/2} \left[1 - \frac{g}{1+g} \frac{\boldsymbol{Z}'_A \tilde{\mathbf{H}}_{A,j} \boldsymbol{Z}_A}{\boldsymbol{Z}'_A \boldsymbol{Z}_A}\right]^{-n/2} \pi(dg), \tag{7}$$

where $\tilde{\mathbf{H}}_{A,j} = \tilde{\mathbf{X}}_{A,\gamma_j}(\tilde{\mathbf{X}}'_{A,\gamma_j}\tilde{\mathbf{X}}_{A,\gamma_j})^{-1}\tilde{\mathbf{X}}'_{A,\gamma_j}$.

Also, marginalizing over all possible subcluster allocations for a given sample size $n$, the following marginal likelihood can be obtained under a DP prior on $f$ (Kyung et. al., 2009):

$$L(\boldsymbol{Y}^n | \mathscr{M}_j) = \frac{\Gamma(m)}{\Gamma(m+n)} \sum_{k=1}^n m^k \sum_{\mathbf{A} \in \mathscr{A}_k} \prod_{i=1}^k \Gamma(n_i) L(\boldsymbol{Y}^n | \mathbf{A}, \mathscr{M}_j) = \sum_{\mathbf{A}_l \in \mathscr{C}_n} w_l L(\boldsymbol{Y}^n | \mathbf{A}_l, \mathscr{M}_j), \tag{8}$$

where $\mathcal{A}_k$ is the collection of all possible $n \times k$ matrices corresponding to different allocations of $n$ subjects into $k$ subclusters, and $\mathcal{C}_n$ is the collection of all possible allocation matrices for a sample size $n$ with $\sum_{\mathbf{A}_l \in \mathcal{C}_n} w_l = 1$. In the limiting case as $n \to \infty$, we have $\mathcal{C}_\infty$ as the class of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

limiting allocation matrices. Further using (7), the Bayes factor in favor of $\mathcal{M}_2$ conditional on the allocation matrix **A** is given by

$$BF^n_{21,A}=\frac{L(\boldsymbol{Z}_A|\mathcal{M}_2)}{L(\boldsymbol{Z}_A|\mathcal{M}_1)}=\frac{\int_0^\infty(1+g)^{-p_2/2}\left[1-\frac{g}{1-g}\tilde{R}^2_{A,2}\right]^{-n/2}\pi(dg)}{\int_0^\infty(1+g)^{-p_1/2}\left[1-\frac{g}{1+g}\tilde{R}^2_{A,1}\right]^{-n/2}\pi(dg)}, \quad (9)$$

where $\tilde{R}^2_{A,j}=\boldsymbol{Z}'_A\tilde{\mathbf{H}}_{A,j}\boldsymbol{Z}_A/\boldsymbol{Z}'_A\boldsymbol{Z}_A$, ($j$ = 1, 2). Finally using (8), the unconditional Bayes factor in favor of $\mathcal{M}_2$ marginalizing out **A** is

$$BF^n_{21}=\frac{L(\boldsymbol{Y}^n|\mathcal{M}_2)}{L(\boldsymbol{Y}^n|\mathcal{M}_1)}=\frac{\sum_{\mathbf{A}_l\in\mathcal{C}_n}w_lL(\boldsymbol{Z}_{\mathbf{A}_l}|\mathcal{M}_2)}{\sum_{\mathbf{A}_l\in\mathcal{C}_n}w_lL(\boldsymbol{Z}_{\mathbf{A}_l}|\mathcal{M}_1)}. \quad (10)$$

### 2.3 Posterior Computation

We propose a MCMC algorithm for posterior computation for model (2), which combines a stochastic search variable selection algorithm or SSVS (George and McCulloch, 1997) with recently proposed methods for efficient computation in DP mixture models. In particular, we utilize the slice sampler of Walker (2007) incorporating the modification of Yau et al. (2011). Using Sethuraman's (1994) stick-breaking representation, let

$$P=\sum_{j=1}^{\infty}w_j\eta_j, \quad \eta_j\sim N(0,\tau^{-1}), \quad w_j=\nu_j\prod_{l<j}(1-\nu_l), \quad \nu_l\sim Beta(1,m). \quad (11)$$

The slice sampler of Walker (2007) relies on augmentation with uniform latent variables, which allows us to move from an infinite summation for $P$ in (11) to a finite sum given the uniform latent variable. In particular,

$$f_{w,\eta}(y|u)\propto\sum_{j\in B_w(u)}N(y|\eta_j), \quad B_w(u)=\{j:w_j>u\} \text{ is a finite set}, \quad u\sim U(0,1).$$

For the DP precision parameter, we specify the hyperprior $m \sim Ga(a_m, b_m)$ for greater flexibility. We specify a $Ga(a_\tau, b_\tau)$ prior on $\tau$ and $Be(a_1, b_1)$ prior on $\Pr(\gamma_l = 1)$ for implementing SSVS, $l = 1, \ldots, p$. We choose $\pi(g)$ as the hyper-$g$ prior with $a = 4$ and use the fact that $\frac{g}{1+g}\sim Be(1,1)$ to sample $g$ using a griddy Gibbs approach employing equally spaced quantiles. Inverting the $n \times n$ matrix $\boldsymbol{\Sigma}_A$ in the mixtures of semiparametric $g$-prior in (4) does not add much to the computational burden even for large $n$, as we can use the closed form expression $\sum_A^{-1}=\mathbf{I}_n-\mathbf{A}(\mathbf{I}_k+\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, where $k$ grows at a rate $\log(n)$ (Antoniak, 1974) and is small to moderate in most practical applications. We outline the posterior computation steps in Appendix I.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## 3. ASYMPTOTIC PROPERTIES

In this section we establish asymptotic properties for the proposed approach using $\gamma_1$ to index the true model $\mathcal{M}_1$ defined in (5) and $\gamma_2$ to index an arbitrary model $\mathcal{M}_2$ being compared to $\mathcal{M}_1$, with $\mathcal{M}_1 \subset \mathcal{M}_2$ denoting nesting of $\mathcal{M}_1$ in $\mathcal{M}_2$. Before proceeding, we introduce some regularity conditions essential for the development of asymptotic theory.

**(A1′)** $\lim_{n\to\infty} \frac{\beta_{\gamma_1}'(\mathbf{X}_{\gamma_1}'\mathbf{X}_{\gamma_1})\beta_{\gamma_1}}{n} \to b_1 > 0$.

**(A2′)** For $\mathcal{M}_1 \not\subseteq \mathcal{M}_2$, $\lim_{n\to\infty} \frac{\beta_{\gamma_1}'\mathbf{X}_{\gamma_1}'\mathbf{H}_2\mathbf{X}_{\gamma_1}\beta_{\gamma_1}}{n} \to b_2 \in [0, b_1]$, with $\mathbf{H}_2 = \mathbf{X}_{\gamma_2}(\mathbf{X}_{\gamma_2}'\mathbf{X}_{\gamma_2})^{-1}\mathbf{X}_{\gamma_2}'$.

**(A1)** For $p_1 = O(n^{a_1})$, $0 \le a_1 < 1$, $\lim_{n\to\infty} \frac{\beta_{\gamma_1}'(\mathbf{X}_{\gamma_1}'\sum_A^{-1}\mathbf{X}_{\gamma_1})\beta_{\gamma_1}}{n} \to b_{A,1} > 0$.

**(A2)** For $\mathcal{M}_1 \not\subseteq \mathcal{M}_2$, $\lim_{n\to\infty} \frac{\beta_{\gamma_1}'\tilde{\mathbf{X}}_{A,\gamma_1}'\tilde{\mathbf{H}}_{A,2}\tilde{\mathbf{X}}_{A,\gamma_1}\beta_{\gamma_1}}{n} \to b_{A,2}$, where $b_{A,2} \in [0, b_{A,1})$ for fixed $p_1$, $p_2$, and $b_{A,2} \in (0, b_{A,1})$ for $p_j = O(n^{a_j})$ ($j = 1, 2$, $0 \le a_1 < a_2 < 1$).

*(A1)*, *(A2)* depend on the allocation matrix $\mathbf{A}$, which is a $n \times k$ binary matrix that for large $n$ tends to have $k \ll n$, and be very sparse containing mostly zeros with sparsity increasing with column index. We also assume the following for the class of proper priors $\pi(g)$ on $g$:

**(A3)** There exists a constant $k \ge 0$ such that $\int_{a_n}^{c_0 a_n} \pi(dg) \approx n^{-k}$ for any constant $c_0 > 1$ and any sequence $a_n \approx n$. Here $a_n \approx b_n$ implies that $\lim_{n\to\infty} a_n/b_n > 0$.

**(A4)** There exists a constant $k_u$ such that $k - (p_2 - p_1)/2 < k_u \le k$ and $\int_0^\infty (1+g)^{k_u}\pi(dg) \approx 1$.

We state *(A1′)*, *(A2′)* as the standard assumptions for establishing Bayes factor consistency in normal linear models, on which our assumptions *(A1)*, *(A2)* are based. We develop asymptotic theory for semiparametric linear models (5) based on assumptions *(A1)–(A4)*. We note that *(A1)* is stronger compared to *(A1′)*, since *(A1)* implies *(A1′)* as $\sum_A^{-1} = \mathbf{I}_n - \mathbf{A}(\mathbf{I}_k + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. Further, in the extreme case when $\mathbf{A} = \mathbf{I}_n$, we have $\frac{\mathbf{x}_{\gamma_1}'\sum_A^{-1}\mathbf{x}_{\gamma_1}}{n} = \frac{1}{2}\frac{\mathbf{x}_{\gamma_1}'\mathbf{x}_{\gamma_1}}{n}$, so that *(A1′)* implies *(A1)*. Again when $\mathbf{A} = \mathbf{1}_n$, $\mathbf{X}_{\gamma_1}'\sum_A^{-1}\mathbf{X}_{\gamma_1} \approx \mathbf{X}_{\gamma_1}'\mathbf{X}_{\gamma_1} - n\overline{\mathbf{X}}_{\gamma_1}'\overline{\mathbf{X}}_{\gamma_1}$ for large $n$, for $\overline{\mathbf{X}}_{\gamma_1} = \mathbf{1}_n'\mathbf{X}_{\gamma_1}/n$. Hence $\frac{\beta_{\gamma_1}'(\mathbf{X}_{\gamma_1}'\sum_A^{-1}\mathbf{X}_{\gamma_1}')\beta_{\gamma_1}}{n} \approx \frac{\beta_{\gamma_1}'(\mathbf{X}_{\gamma_1}^{c\prime}\mathbf{X}_{\gamma_1}^c)\beta_{\gamma_1}}{n}$, where $\mathbf{X}_{\gamma_1}^c$ is the centered design matrix. When $\lim_{n\to\infty} \frac{\beta_{\gamma_1}'(\mathbf{X}_{\gamma_1}^{c\prime}\mathbf{X}_{\gamma_1}^c)\beta_{\gamma_1}}{n} > 0$, *(A1′)* implies *(A1)*.

Assumption *(A2)* can be interpreted as a positive 'limiting distance' between the two models corresponding to design matrices $\mathbf{X}_{\gamma 1}$ and $\mathbf{X}_{\gamma 2}$ in (3) conditional on $\mathbf{A}$, after marginalizing out $\eta$, i.e. $\Delta_{21,A} = \lim_{n\to\infty} \frac{\beta_{\gamma_1}'\tilde{\mathbf{X}}_{A,\gamma_1}'(\mathbf{I}_n - \tilde{\mathbf{H}}_{A,2})\tilde{\mathbf{X}}_{A,\gamma_1}\beta_{\gamma_1}}{n\tau^{-1}} = \frac{b_{A,1} - b_{A,2}}{\tau^{-1}} \in (0, \infty)$. Such a 'limiting distance' ($\Delta_{21,A}$) can be considered as a natural extension of the definition of distance

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between two normal linear models in Casella et. al. (2009) and Moreno et. al. (2010) to models with random intercept as in (3).

Assumptions *(A3)*, *(A4)* define a class of proper priors for *g* described in Guo and Speckman (2009). This class includes $\text{hyper}-g\ (\frac{a-2}{2}(1+g)^{-a/2})$ and $\text{hyper}-g/n\ (\frac{a-2}{2n}(1+g/n)^{-a/2})$ priors with $2 < a \leq 4$ (Liang et. al. 2008), Zellner-Siow priors (Zellner and Siow, 1980) as well as beta-prime priors (Maruyama and George, 2008). It is clear that these assumptions on $\pi(g)$ are satisfied by quite a few standard priors are hence are quite reasonable.

The following lemma gives the limits of quantities such as $\tilde{R}^2_{A,j} = \boldsymbol{Z}'_A \check{\mathbf{H}}_{A,j} \boldsymbol{Z}_A / \boldsymbol{Z}'_A \boldsymbol{Z}_A$, which would be useful for establishing asymptotic properties. The proof follows directly using Lemmas 1, 2 of Guo and Speckman (2009) and from (6) which essentially states that under the DP mixture of Gaussians prior on *f* for $\mathcal{M}_j$ in (5) and conditional on allocation matrix **A**, $\boldsymbol{Z}_A = \sum_A^{-1/2} \boldsymbol{Y}^n \sim N(\tilde{\mathbf{X}}_{A,\gamma_j} \boldsymbol{\beta}_{\gamma_j}, \tau^{-1}\mathbf{I}_n)$, $j = 1, 2$.

**Lemma 1—**Let assumptions (A1), (A2) hold.

**i.** *If* $\mathcal{M}_1 \subset \mathcal{M}_2$, *conditional on* **A**, $\tilde{R}^2_{A,1} \xrightarrow{a.s.} \frac{b_{A,1}}{\tau^{-1}+b_{A,1}}$, $\tilde{R}^2_{A,2} \xrightarrow{a.s.} \frac{b_{A,1}}{\tau^{-1}+b_{A,1}}$, *under* $\mathcal{M}_1$

**ii.** *If* $\mathcal{M}_1 \not\subseteq \mathcal{M}_2$, *conditional on* **A**, $\tilde{R}^2_{A,1} \xrightarrow{a.s.} \frac{b_{A,1}}{\tau^{-1}+b_{A,1}}$, $\tilde{R}^2_{A,2} \xrightarrow{a.s.} \frac{b_{A,2}}{\tau^{-1}+b_{A,1}}$, *under* $\mathcal{M}_1$

As shown by the following result, the proposed approach leads to Bayes factor consistency when comparing fixed dimensional models as well as models growing at the rate $O(n^t)$, $0 < t < 1$, when the truth is sparse.

**Theorem I—**Let assumptions (A1), (A2) hold.

**I.** Suppose $p_1$ and $p_2$ are fixed. If $\mathcal{M}_1 \subset \mathcal{M}_2$, then under $\mathcal{M}_1$ and assumptions (A3), (A4), $BF^n_{21} \xrightarrow{P} 0$ *as* $n \to \infty$ *and if* $p_2 - p_1 > 2 + 2(k - k_u)$, $BF^n_{21} \xrightarrow{a.s.} 0$ *as* $n \to \infty$. *Further, if* $\mathcal{M}_1 \not\subseteq \mathcal{M}_2$, *then under* $\mathcal{M}_1$ *and assumption (A3),* $BF^n_{21} \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

**II.** Suppose $p_j$ is growing at the rate $O(n^{aj})$, j=1,2, with $0 \leq a_1 < a_2 < 1$. Then under $\mathcal{M}_1$ and assumption (A3), $BF^n_{21} \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

**REMARK 1:** Although we omit the proof here, Theorem I can be modified to accommodate the case of improper priors on *g* (i.e. $\pi(g) \propto \frac{1}{1+g}$). In such a case, assumptions *(A3)*, *(A4)* are excluded and we require $p_2 - p_1 \geq 3$ for a.s. convergence in *(I)* for $\mathcal{M}_1 \subset \mathcal{M}_2$.

The next result establishes model selection consistency for the proposed approach, even in cases when the cardinality of the model space increases with *n*. In particular, we consider cases when the number of candidate predictors $p_n$ is growing at the rate $O(n^a)$, $a > 0$, but the prior on the model space assigns zero probability to models growing at a rate equal to or faster than *n*. When $a \geq 1$, the prior support consists of models constructed using $O(n^t)$ ($0 \leq t < 1$) sized subsets of $p_n = O(n^a)$ candidate predictors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To elaborate, let the support of the prior on the model space be $\mathcal{M} = \mathcal{M}_F \cup \mathcal{M}_I$, where $\mathcal{M}_F$ is the set of all (non-null) models $\boldsymbol{\gamma}$ such that there exists a sample size $n_0 < \in$ for which $\gamma_j = 0$ for all $j > p_{n0}$, and $\mathcal{M}_I$ is the set of all models with dimensions growing at a rate strictly less than $n$, $\mathcal{M}_I = \{\boldsymbol{\gamma}: \sum_{j=1}^{p_n} \gamma_j = O(n^t), 0 < t < 1\}$. Letting $p_0 = \max\{j: \boldsymbol{\gamma} \in \mathcal{M}_F, \gamma_j = 1\}$, we can discard predictors having a higher index than $p_0$ for all $\boldsymbol{\gamma} \in \mathcal{M}_F$ and treat $\mathcal{M}_F$ as finite dimensional having $2^{p0} - 1$ elements (excluding the null model). Let $\boldsymbol{\gamma}_{jl}$ denote the $l$th model having dimension $p_j$. Consider the following sequence of priors which penalizes models with increasing dimensions, thus encouraging sparsity:

$$\pi^n(\boldsymbol{\gamma}_{jl}) \propto \pi^{p_j}(1-\pi)^{p_0-p_j} I[\boldsymbol{\gamma}_{jl} \in \mathcal{M}_F] + \begin{pmatrix} p_n \\ p_j \end{pmatrix}^{-1} I[\boldsymbol{\gamma}_{jl} \in \mathcal{M}_I], \quad \pi \sim Be(a_1, b_1). \quad (12)$$

When the truth is sparse such that $\mathcal{M}_I \in \mathcal{M}_F$, we have the following result.

**Theorem II—**Suppose assumptions (A1)–(A4) hold. For fixed $p$ and under $P(\mathcal{M}_1|\boldsymbol{Y}^n) \xrightarrow{P} 1$ for any prior on $\Gamma$ with $\pi(\mathcal{M}_1) > 0$. When $p_n = O(n^a)$ $(a > 0)$ and $P(\mathcal{M}_1|\boldsymbol{Y}^n) \xrightarrow{P} 1$ under $\mathcal{M}_1$, for $\pi^n(\boldsymbol{\gamma}_{jl})$ defined as in (12).

## 4. SIMULATION STUDY

We present the results of two simulation studies comparing our method (SLM) with the normal linear model (NLM) having $\boldsymbol{\beta}_\gamma \sim N(0, g\tau^{-1}(\mathbf{X}_\gamma' \sum_{A=1_n}^{1} \mathbf{X}_\gamma)^{-1})$ (designed to assign comparable prior information when the residual is Gaussian), the lasso (Tibshirani, 1996) and elastic net (Zou and Hastie, 2005), as well as robust variable selection methods including an MM-type regression estimator (Yohai, 1987; Koller and Stahel, 2011), and a median regression model with SSVS for variable selection (Yu et al., 2013). The data is generated as follows:

$$\begin{aligned} \textbf{Case I}: &\quad y_i = \mathbf{x}_i \boldsymbol{\beta}_T + \varepsilon_i, \quad \varepsilon_i \sim 0.5N(2.5, 1) + 0.5N(-2.5, 1), \\ \textbf{Case II}: &\quad y_i = 1 + \mathbf{x}_i \boldsymbol{\beta}_T + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1), \quad i - 1, \ldots, n, \end{aligned}$$

where $\mathbf{x}_i$ is a ten dimensional predictor (p=10), with $x_{ij}$, $j$ = 1, …, 10 generated independently from U(−1,1), and $\boldsymbol{\beta}_T$ = (3, 2, −1, 0, 1.5, 1, 0, −4, −1.5, 0).

We used $Ga(0.1, 1)$ prior on the DP precision parameter and $Be(0.1, 1)$ prior on $P(\gamma_j = 1)$, j=1,…,p, which corresponds to a weakly informative prior favoring parsimony. We update $g$ using the griddy Gibbs approach having 1000 equally spaced quantiles for $\frac{g}{1+g} \sim Be(1, 1)$ corresponding to $a$ = 4 in the hyper-$g$ prior. For both SLM and NLM, we ran 50,000 iterations with a burn in of 5,000. We implemented the lasso (L1) and elastic net (EL) using the GLMNET package in R with default settings, while the MM-type estimator (LMR) was implemented using 'lmrob' function in 'robustbase' package in R and the median regression with SSVS (QR) was implemented using function 'SSVSquantreg' in 'MCMCpack' package in R, with a $Be(0.1, 1)$ prior on the prior inclusion probability for predictors. All

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

results are summarized across 20 replicates. The computation time for SLM per iteration was marginally slower than NLM. The mixing for the fixed effects was good under both the methods. The results for SLM do not appear to be sensitive to the hyper-parameters in $\pi(m)$, but are mildly sensitive to hyper-parameters in $\pi(g)$ for $n = 100$.

We study the marginal inclusion probabilities (MIP) under SLM and NLM over varying sample sizes in Figures 1 and 2. These plots suggest a faster rate of increase of the MIP for the important predictors under SLM as compared to NLM when the true residuals are non-Gaussian, and a very similar rate of increase under both methods when the true residuals are Gaussian (thus justifying the prior choice for NLM). In contrast, the exclusion probabilities for the unimportant predictors converge to one slowly under both the methods, reflecting the well known tendency for slower accumulation of evidence in favor of the true null.

Tables 1 and 2 present some summaries for $n = 100$ for Case I. The MIPs in Table I suggests correct variable selection decision by SLM, but poor performance by NLM which fails to exclude any of the unimportant predictors under median probability model. Further, L1, EL and QR seem to favor an overly complex model by choosing a superset of important predictors. In terms of estimation of the fixed effects, SLM has the highest degree of accuracy as reflected by the smallest mean square error ( $\frac{\|\hat{\beta}-\beta_T\|_2}{p}$) in Table 2, where $\boldsymbol{\beta}_T$ is the vector of true regression coefficients. In addition, the replicate average mean square error for out of sample prediction for a test sample size of 25 (Table 2) is smallest under the SLM, followed by lasso and elastic net. NLM is seen to be clearly inadequate for prediction purposes as indicated by the extremely high out of sample predictive MSE. Thus in conclusion, when the true residual is non-Gaussian, the SLM has the best performance compared to competitors, whereas NLM performs poorly in general.

## 5. APPLICATION TO DIABETES DATA

The prevalence of diabetes in the United States is expected to more than double to 48 million people by 2050 (Mokdad et. al., 2001). Previous medical studies have suggested that Diabetes Mellitus type II (DM II) or adult onset diabetes could be associated with high levels of total cholesterol (Brunham et. al., 2007) and obesity (often characterized by BMI and waist to hip ratio) (Schmidt et. al., 1992), as well as hypertension (indicated by a high systolic or diastolic blood pressure or both) which is twice as prevalent in diabetics compared to non-diabetic individuals (Epstein and Sowers, 1992).

We develop a comprehensive variable selection strategy for indicators of DM II in African-Americans based on data obtained from Department of Biostatistics, Vanderbilt University website. Our primary focus is to discover important indicators of DM II by modeling the continuous outcome glycosylated hemoglobin ($> 7mg/dL$ indicates a positive diagnosis of diabetes) based on predictors such as total cholesterol (TC), stabilized glucose (SG), high density lipoprotein (HDL), age, gender, body mass index (BMI) indicator (overweight and obese with normal as baseline), systolic and diastolic blood pressure (SBP and DBP), waist to hip ratio (WHR) and postprandial time indicator (PPT) (1 if blood was drawn within 2 hours of a meal, 0 otherwise). We note that lower levels of HDL have been known to be associated with insulin resistance syndrome, often considered a precursor of DM II with a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

conversion rate around 30%. We also expect PPT to be a significant indicator as blood sugar levels are high up to 2 hours after a meal.

After excluding the records containing missing values, the data consisted of 365 subjects which was split into multiple training and test samples of sizes 330 and 35 respectively. The replicate averaged fixed effects estimates (multiplied by 100) for the SLM, NLM, L1, EL, LMR and QR are presented in Table 3, and the marginal inclusion probabilities (MIP) for the SLM, NLM and QR are summarized in Table 4. We also evaluate the out of sample predictive performance for each training-test split using predictive MSE in Table 5, and additionally provide the mean coverage (COV) and width (CIW) of 95% pointwise credible intervals for the predicted responses under SLM and NLM. The same values of hyper-parameters were used as in section 5. For each replicate, we randomized the initial starting points and made 100,000 runs for SLM (burn in = 20,000) and 50,000 runs for NLM (burn in = 5,000).

It is interesting to note from Table 4 that the variable selection decisions under SLM (using median probability model) are quite different compared to the NLM. In particular, while both the models successfully identify total cholesterol, stabilized glucose and postprandial time as important predictors, it is only the SLM which identifies systolic blood pressure (MIP = 0.72), HDL (MIP = 0.64) and waist to hip ratio (MIP = 0.93) as important indicators, compared to NLM which assigns MIP = 0.14, 0.39 and 0.13 to these three predictors respectively. Age is identified as an important predictor under NLM (MIP = 0.67), but not under SLM (MIP=0.43). For both the methods, the MIPs for BMI (overweight and obese) were low, which could potentially be attributed to adjusting for the other obesity factors such as waist to hip ratio. From Tables 3 and 4, we also see that the lasso, elastic net and the MM-type estimator select an overly complex model by excluding minimal number of predictors, while the quantile regression with SSVS fails to include several important predictors and selects a highly parsimonious and inadequate model.

Variable selection in this application is clearly influenced by the assumptions on the residual density, with the nonparametric residual density providing a more realistic characterization that should lead to a more accurate selection of the important predictors. Figure 3 shows an estimate of the residual density obtained from the SLM analysis, suggesting a uni-modal right skewed density with a heavy right tail. The SLM results suggest that a mixture of two Gaussians provides an adequate characterization of this density. The computation time for SLM is only marginally slower than NLM, and in addition SLM exhibits good mixing for most of the fixed effects (Table 6). These results are robust to SSVS starting points, and consistency in the results across training-test splits also indirectly suggests adequate computational efficiency of SSVS.

In terms of out of sample predictive MSE (Table 5), the relative performance between SLM, NLM, L1 and EL vary across training-test splits so that none of the models can be said to dominate the others, while LMR and QR produce relatively inferior prediction results. Overall, the NLM has narrower 95% pointwise credible intervals compared to SLM, often resulting in poorer coverage for out of sample predictions. In conclusion, SLM succeeds in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

choosing the most reasonable model for DM II, consistent with previous medical evidence, and compares favorably with other competitors for prediction purposes.

## Acknowledgments

This work was supported by Award Number R01ES017240 from the National Institute of Environmental Health Sciences. The authors thank the referees and the associate editor for their valuable comments.

## References

1. Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Annals of Statistics. 1974; 2:11521174.
2. Armagan A, Dunson DB, Lee J, Bajwa WU, Strawn N. Posterior consistency in high-dimensional linear models. Biometrika. 2013 To appear.
3. Brunham LR, Kruit JK, Pape TD, Timmins JM, Reuwer AQ, Vasanji Z, Marsh BJ, Rodrigues B, Johnson JD, Parks JS, Verchere CB, Hayden MR. *β*-cell ABCA1 influences insulin secretion, glucose homeostasis and response to thiazolidinedione treatment. Nature Medicine. 2007; 13:340–347.
4. Casella G, Girón FJ, Martínez ML, Moreno E. Consistency of Bayesian procedures for variable selection. Annals of Statistics. 2009; 37:1207–1228.
5. Castillo I, van der Vaart A. Needles and straws in a haystack: Posterior concentration for possibly sparse sequences. Annals of Statistics. 2012; 40:2069–2101.
6. Epstein M, Sowers JR. Diabetes mellitus and hypertension. Hypertension. 1992; 19:403–418. [PubMed: 1568757]
7. Ferguson TS. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1972; 1:209–230.
8. George EI, McCulloch RE. Approaches for Bayesian Variable Selection. Statistica Sinica. 1997; 7(2):339–74.
9. Guo, R.; Speckman, P. Bayes factor consistency in linear models. In the 2009 International Workshop on Objective Bayes Methodology; Philadelphia. 2009; 2009.
10. Jiang W. Bayesian variable selection for high dimensional generalized linear models: Convergence rates of the fitted densities. Annals of Statistics. 2007; 35:1487–1511.
11. Koller M, Stahel WA. Sharpening Wald-type inference in robust regression for small samples. Computational Statistics and Data Analysis. 2011; 55:25042515.
12. Kyung M, Gill J, Casella G. Characterizing the variance improvement in linear Dirichlet random effects models. Statistics and Probability Letters. 2009; 79:2343–2350.
13. Liang F, Paulo R, Molina G, Clyde MA, Berger JO. Mixtures of g-priors for Bayesian Variable Selection. Journal of the American Statistical Association. 2008; 103:410–423.
14. Lo AY. On a class of Bayesian nonparametric estimates: I. density estimates. Annals of Statistics. 1984; 12:351–357.
15. Maruyama Y, George E. A *g*–prior extension for $p > n$. 2008 arxiv:0801.4410v1 [stat.ME].
16. Mokdad AH, Bowman BA, Ford ES, Vinicor F, Marks JS, Koplan JP. The continuing epidemics of obesity and diabetes in the United States. Journal of the American Medical Association. 2001; 286:1195–1200. [PubMed: 11559264]
17. O'Hara RB, Sillanpää MJ. Review of Bayesian variable selection methods: What, how and which. Bayesian Analysis. 2009; 4:85–118.
18. Yu K, Chen CWS, Reed C, Dunson D. Bayesian Variable Selection in Quantile Regression. Statistics and its Interface. 2013; 6:261–274.
19. Ritter C, Tanner MA. Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. Journal of the American Statistical Association. 1992; 87:861–868.
20. Schmidt MI, Duncan BB, Canani LH, Karohl C, Chambless L. Association of waist-hip ratio with diabetes mellitus. Strength and possible modifiers. Diabetes Care. 1992; 15:912–4. [PubMed: 1516514]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


21. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B. 1996; 58(1):267–288.
22. Walker S. Sampling the dirichlet mixture model with slices. Communication in Statistics - Simulation and Computation. 2007; 36:45–54.
23. Yau C, Papaspiliopoulos O, Roberts G, Holmes C. Bayesian non-parametric hidden Markov models with applications in genomics. Journal of the Royal Statical Society, Series B. 2011; 73(Part 1):33–57.
24. Yohai VJ. High breakdown-point and high efficiency estimates for regression. Annals of Statistics. 1987; 15:642–65.
25. Zellner, A.; Siow, A. Bayesian Statistics: Proceedings of the First International Meeting. Valencia: University of Valencia Press; 1980. Posterior odds ratios for selected regression hypotheses; p. 585-603.
26. Zellner A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. 1986:233–243.
27. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B. 2005; 67:301–320.


# APPENDIX A: PROOF OF RESULTS

## Proof of Theorem I

Using similar methods as in the proof of Theorem 2 in Guo and Speckman (2009), it can be shown that conditional on **A** and assumptions (*A3*) and (*A4*), the upper and lower bounds of $L(\boldsymbol{Y}^n|\mathbf{A},M_1)=\int_0^\infty(1+g)^{-p_1/2}\left[1-\frac{g}{1+g}\tilde{R}^2_{A,1}\right]^{-n/2}\pi(dg)$ are

$$L(\boldsymbol{Y}^n|\mathbf{A},M_1)\le\left(\frac{p_1+2k_u}{n-p_1-2k_u}\right)^{p_1/2+k_u}\left(\frac{1-\tilde{R}^2_{A,1}}{\tilde{R}^2_{A,1}}\right)^{p_1/2+k_u}\left(\frac{n}{n-p_1-2k_u}\right)^{-n/2}\left(1-\tilde{R}^2_{A,1}\right)^{-n/2}$$
$$\approx\left(\frac{p_1+2k_u}{n-p_1-2k_u}\right)^{p_1/2+k_u}\left(\frac{1-\tilde{R}^2_{A,1}}{\tilde{R}^2_{A,1}}\right)^{p_1/2+k_u}\left(1-\tilde{R}^2_{A,1}\right)^{-n/2}=U_{A,1}(n),$$

and $L(\boldsymbol{Y}^n|\mathbf{A},M_1)\ge n^{-p_1/2-k}\left(1-\tilde{R}^2_{A,1}\right)^{-n/2}=L_{\mathbf{A},1}(n)$. Similarly, $L_{\mathbf{A},2}(n)\le L(\boldsymbol{Y}^n|\mathbf{A},M_2)=\int_0^\infty(1+g)^{-p_2/2}\left[1-\frac{g}{1+g}\tilde{R}^2_{A,2}\right]^{-n/2}\pi(dg)\le U_{A,2}(n)$. Therefore, $\mathrm{BF}^n_{21,A}\le\frac{U_{A,2}(n)}{L_{\mathbf{A},1}(n)}$

$$=\left(\frac{p_2+2k_u}{n-p_2-2k_u}\right)^{p_2/2+k_u}\left(\frac{1-\tilde{R}^2_{A,2}}{\tilde{R}^2_{A,2}}\right)^{p_2/2+k_u}\left(1-\tilde{R}^2_{A,2}\right)^{-n/2}/\left(n^{-p_1/2-k}(1-\tilde{R}^2_{A,1})^{-n/2}\right). \quad (13)$$

Case (I): For fixed $p_j$ ($j$ = 1, 2) and large $n$, $\mathrm{BF}^n_{21,A}\le\zeta(\mathbf{A},n)=n^{\frac{p_1-p_2}{2}+k-k_u}\left(\frac{1-\tilde{R}^2_{A,2}}{1-\tilde{R}^2_{A,1}}\right)^{-n/2}$, ignoring terms independent of $n$. Using the results in proof of Theorems 2, 3 in Guo and Speckman (2009), we can show that $n\zeta(\mathbf{A},n)\xrightarrow{a.s.}0$ when $\mathcal{M}_1\not\subseteq\mathcal{M}_2$. Again for $\mathcal{M}_1\subset\mathcal{M}_2$,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

using results in the aforementioned proofs, we have $\left(\frac{1-\tilde{R}^2_{A,2}}{1-\tilde{R}^2_{A,1}}\right)^{-n/2}=O_P(1)$. Further for $\mathcal{M}_1 \subset \mathcal{M}_2$ when $p_2 - p_1 > 2 + 2(k - k_u)$, we have $n^{\delta}\zeta(\mathbf{A}, n) \xrightarrow{a.s.} 0$, where $\delta > 0$ is such that $i \leq 2(k - k_u) + 2\delta < i + 1$ when $i \leq 2(k - k_u) < i + 1$. This implies that for large enough $n$,

$$\begin{gathered}\text{for } \mathcal{M}_1 \subset \mathcal{M}_2{:}\zeta(\mathbf{A}, n)=n^{-\frac{p_2-p_1}{2}+(k-k_u)} \text{ in probability,}\\ \text{for } \mathcal{M}_1 \subset \mathcal{M}_2{:}\zeta(\mathbf{A}, n) \leq n^{-\delta} \text{ almost surely when } p_2-p_1>2+2(k-k_u),\\ \text{for } \mathcal{M}_1 \not\subseteq \mathcal{M}_2{:}\zeta(\mathbf{A}, n) \leq n^{-1} \text{ almost surely.}\end{gathered} \quad (14)$$

Then for large enough $n$, we have,

$$\begin{gathered}\mathrm{BF}^n_{21,A} \leq \zeta(\mathbf{A}, n) \iff L(\mathbf{Y}^n|\mathbf{A}, \mathcal{M}_2) \leq \zeta(\mathbf{A}, n)L(\mathbf{Y}^n|\mathbf{A}, \mathcal{M}_1)\\ \Rightarrow L(\mathbf{Y}^n|\mathcal{M}_2) \leq \sum_{\mathbf{A}_l \in \mathcal{C}_n} w_l \zeta(\mathbf{A}_l, n)L(\mathbf{Y}^n|\mathbf{A}_l, \mathcal{M}_1) \leq \zeta^*(n)L(\mathbf{Y}^n|\mathcal{M}_1),\end{gathered} \quad (15)$$

where $\zeta^*(n)$ is the LHS in equations (14) which is independent of $\mathbf{A}$, and $\zeta^*(n) \to 0$ as $n \to \infty$ (using (*A4*)). Dividing both sides of (15) by $L(\mathbf{Y}^n|\ \mathcal{M}_1)$, we have $\mathrm{BF}^n_{21} \leq \zeta^*(n) \to 0$ as $n \to \infty$.

Case (II): For increasing dimensions $p_1 = O(n^{a_1})$, $p_2 = O(n^{a_2})$ with $0 \leq a_1 < a_2 < 1$, we will only assume *(A3)* for $g \sim \pi(g)$ so that $k_u = 0$. We have using (13)

$$BF^n_{21,A} \leq \zeta(\mathbf{A}, n)n^{p_1/2-(1-a_2)p_2/2+k}\left(\frac{1-\tilde{R}^2_{A,2}}{\tilde{R}^2_{A,2}}\right)^{p_2/2}\left(\frac{1-\tilde{R}^2_{A,2}}{1-\tilde{R}^2_{A,1}}\right)^{-n/2}. \quad (16)$$

Let us consider the following cases under $0 \leq a_1 < a_2 < 1$.

Case C1: $\mathcal{M}_1 \subset \mathcal{M}_2$. We have $Q_j=\tau(\boldsymbol{Z}'_A\boldsymbol{Z}_A-\boldsymbol{Z}'_A\tilde{\mathbf{H}}_{A,j}\boldsymbol{Z}_A)\sim\chi^2_{n-p_j}(0)$, j=1,2, and $Q_1-Q_2=\tau\left(\boldsymbol{Z}'_A(\tilde{\mathbf{H}}_{A,2}-\tilde{\mathbf{H}}_{A,1})\boldsymbol{Z}_A\right)\sim\chi^2_{p_2-p_1}(0)$. Using Lemma 1 of Guo et. al. (2009),

$$\frac{1-\tilde{R}^2_{A,1}}{1-\tilde{R}^2_{A,2}}=\frac{\boldsymbol{Z}'_A\boldsymbol{Z}_A-\boldsymbol{Z}'_A\tilde{\mathbf{H}}_{A,1}\boldsymbol{Z}_A}{\boldsymbol{Z}'_A\boldsymbol{Z}_A-\boldsymbol{Z}'_A\tilde{\mathbf{H}}_{A,2}\boldsymbol{Z}_A}=\frac{Q_1}{Q_2}=1+\frac{(Q_1-Q_2)/(p_2-p_1)}{Q_2/(n-p_2)}\frac{p_2-p_1}{n-p_2} \xrightarrow{a.s.} 1.$$

Moreover $\left(\frac{1-\tilde{R}^2_{A,2}}{\tilde{R}^2_{A,2}}\right) \xrightarrow{a.s.} \left(\frac{\tau^{-1}}{b_{A,1}}\right)$ under $\mathcal{M}_1$. Then for large $n$,

$$\zeta(\mathbf{A}, n) \approx n^{p_1/2-(1-a_2)p_2/2+k}\left(\frac{\tau^{-1}}{b_{A,1}}\right)^{p_2/2}=n^{p_1/2-(1-a^*-a_2)p_2/2+k}\left(\frac{\tau^{-1}}{n^{a^*}b_{A,1}}\right)^{p_2/2} \text{a.s.},$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $a^* > 0$ is such that $0 < 1 - a^* - a_2 < 1$. This implies that $n^{K^*}\zeta(\mathbf{A}, n) \xrightarrow{a.s.} 0$ under $\mathcal{M}_1$ for any constant $K^* > 0$. Case C2: $\mathcal{M}_1 \not\subset \mathcal{M}_2$. Using Lemma 1,

$$\frac{1-\tilde{R}^2_{A,2}}{\tilde{R}^2_{A,2}} \xrightarrow{a.s.} \frac{\tau^{-1}+b_{A,1}-b_{A,2}}{b_{A,2}} > 1, \quad \frac{1-\tilde{R}^2_{A,1}}{1-\tilde{R}^2_{A,2}} \xrightarrow{a.s.} \frac{\tau^{-1}}{\tau^{-1}+b_{A,1}-b_{A,2}} < 1, \text{ under } \mathcal{M}_1.$$

For fixed $\tau^{-1}$ and $b_{A,2} > 0$ (under *(A2)*), $\left(\frac{1-\tilde{R}^2_{A,2}}{\tilde{R}^2_{A,2}}\right)^{p_2/2}\left(\frac{1-\tilde{R}^2_{A,2}}{1-\tilde{R}^2_{A,1}}\right)^{-n/2} \xrightarrow{a.s.} 0$. This implies that in the limiting case when $n \to \infty$, we have

$$\begin{aligned} &\text{for } \mathcal{M}_1 \subset \mathcal{M}_2 : \zeta(\mathbf{A}, n) \le n^{-K^*} \text{ almost surely,} \\ &\text{for } \mathcal{M}_1 \not\subset \mathcal{M}_2 : \zeta(\mathbf{A}, n) \le n^{p_1/2-(1-a_2)p_2/2+k} \le n^{-K^*} \text{ almost surely,} \end{aligned} \tag{17}$$

where $K^* > 0$ is a constant. Denoting the upper bounds as $\zeta^*(n)$, it is clear that $\zeta^*(n)$ is independent of $\mathbf{A}$ and $\zeta^*(n) \to 0$ as $n \to \infty$ when $0 \le a_1 < a_2 < 1$. Using similar arguments as in equation (15) of Case (I), we have $\mathrm{BF}^n_{21} \le n^{-K^*}$ and consistency follows.

## Proof of Theorem II

Given the assumptions *(A1)–(A4)*, Bayes factor consistency holds under the different cases elaborated in Theorem I. For fixed $p$, the proof follows trivially using Bayes factor consistency. For increasing $p_n = O(n^a)$ $(a > 0)$, our prior is $\pi^n(\gamma_{jl}) \propto \pi^{p_j}(1-\pi)^{p_0-p_j} I[\gamma_{j_l} \in \mathcal{M}_F] + N_j^{-1} I[\gamma_{j_l} \in \mathcal{M}_I]$, where $\pi \sim Be(a_1, b_1)$ and $N_j = \begin{pmatrix} p_n \\ p_j \end{pmatrix}$. Let $W_\gamma$ denote the prior weight for $\gamma \in \mathcal{M}_F$ after marginalizing out $\pi$ under the $Be(a_1, b_1)$ prior ($W_1$ being the weight for $\mathcal{M}_1$). Let $\mathrm{BF}^n_{\gamma 1}$ = Bayes factor between models $\gamma$ and $\mathcal{M}_1$, let $D = \{p_\gamma : \gamma \in \mathcal{M}_I\}$ and denote $\mathcal{H}_j = \{\gamma \in \mathcal{M}_I : \dim(\gamma) = p_j\}$. Note that under *(A1)–(A4)* and $\mathcal{M}_1 \in \mathcal{M}_F$, $\mathrm{BF}^n_{\gamma 1} \xrightarrow{P} 0$ for all $\gamma \in (\mathcal{M}_F \cap \mathcal{M}_1^c) \cup \mathcal{M}_I$, using Theorem I. Also,

$$\begin{aligned} P(\mathcal{M}_1|\mathbf{Y}^n) &= [1 + W_1^{-1} \sum_{\gamma \in \mathcal{M}_F \cap \mathcal{M}_1^c} W_\gamma BF^n_{\gamma 1} + W_1^{-1} \sum_{\gamma \in \mathcal{M}_I} N_j^{-1} BF^n_{\gamma 1}]^{-1} \\ &= [1 + W_1^{-1} \sum_{\gamma \in \mathcal{M}_F \cap \mathcal{M}_1^c} W_\gamma BF^n_{\gamma 1} + W_1^{-1} \sum_{p_j \in D} \sum_{\gamma_{jl} \in \mathcal{H}_j} N_j^{-1} BF^n_{\gamma_{jl} 1}]^{-1} \\ &\ge [1 + \varepsilon_0 + W_1^{-1} \sum_{p_j \in D} \sum_{\gamma_{jl} \in \mathcal{H}_j} N_j^{-1} BF^n_{\gamma_{jl} 1}]^{-1}, \end{aligned}$$

where $W_1^{-1} \sum_{\gamma \in \mathcal{M}_F \cap \mathcal{M}_1^c} W_\gamma BF^n_{\gamma 1} \le \varepsilon_0$ for large enough $n$, and $\varepsilon_0 \to 0$ as $n \to \infty$ since all the individual terms in the finite summation $\to 0$ using Theorem I. Further using (17), the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

upper bound of $BF^n_{\gamma_{jl}1}$ for any $\gamma_{jl} \in \mathcal{H}_j$ is given by $\zeta^*(n) = n^{-K^*}$ when $n$ is large, where $K^* > 0$ is a constant. Noting that the cardinality of $\mathcal{H}_j \leq N_j = \begin{pmatrix} p_n \\ p_j \end{pmatrix}$, we have for large $n$,

$$\sum_{\gamma_{jl} \in \mathcal{H}_j} N_j^{-1} BF^n_{\gamma_{jl}1} \leq n^{-K^*} \Rightarrow P(\mathcal{M}_1|\mathbf{Y}^n) \geq [1+\varepsilon_0+W_1^{-1}\sum_{p_j \in D} n^{-K^*}]^{-1}, \text{ under } \mathcal{M}_1.$$

Now note that $W_1$ is fixed and the cardinality of $D < \kappa_0 n$ for some constant $\kappa_0 > 0$. Thus it is clear that $W_1^{-1}\sum_{p_j \in D} n^{-K^*} \to 0$ as $n \to \infty$ for large $K^*$. The rest is straightforward.

## APPENDIX B: COMPUTATIONAL STEPS FOR MCMC

The posterior computation steps are:

**Step 1.1** Update the $\nu$'s after marginalizing out the augmented uniform variable using $\pi(\nu_h|-) = Be(1 + n_h, \Sigma_{j>h}\, n_j + m)$, h=1, …, M, where M is the total number of clusters satisfying $\sum_{h=1}^{M} w_h > 1 - \min(u_1, \ldots, u_n)$, with $w_h = \nu_h \Pi_{l<h}(1 - \nu_l)$.

**Step 1.2** Update $u_i$, i=1, …, n, from its full conditional as described in Walker (2007).

**Step 2** Update the cluster membership of different subjects using $f(y_i|u_i, \mathbf{A}_{ih} = 1) \propto N(y_i|\eta_h, \mathbf{x}_{\gamma,i}, \boldsymbol{\beta}_\gamma, \tau^{-1})I(h \in B_w(u_i))$, h=1, …, M, with $B_w(u_i)$ defined as in section 2.3.

**Step 3** Update the Dirichlet process atom $\eta_l$ for the l-th cluster using $\pi(\eta_l|-) = N\left(\frac{\sum_{i:A_{il}=1}(y_i = \mathbf{x}_{\gamma,i}\beta_\gamma)}{1+n_l}, \sqrt{\frac{1}{\tau(1+n_l)}}\right)$, where $n_l = \sum_{i=1}^{n} \mathbf{A}_{il}$ is the cardinality of the l-th cluster, l=1, …, M.

**Step 4** Update the DP precision using $\pi(m|-) = Ga(a_m + M, b_m - \sum_{l=1}^{M} \log(1-\nu_l))$.

**Step 5** Letting $p_\gamma = \sum_{j=1}^{p} \gamma_j$, update precision $\tau$ using $\pi(\tau|-) = Ga\left(a_\tau + \frac{n+p_\gamma}{2}, b_\tau + \frac{1}{2}\left\{(\mathbf{Y}^n - \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma)'\sum_A^{-1}(\mathbf{Y}^n - \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma) + \frac{1}{g}\boldsymbol{\beta}_\gamma'(\mathbf{X}_\gamma'\sum_A^{-1}\mathbf{X}_\gamma)\boldsymbol{\beta}_\gamma\right\}\right)$.

**Step 6** Using the hyper-$g$ prior and the fact that $\frac{g}{1+g} \sim Be(1, 1)$ for $a = 4$, we can adopt the griddy Gibbs approach (Ritter and Tanner, 1992) to update $g$.

**Step 7** Update the prior inclusion probability $\pi = \Pr(\gamma_j = 1)$ using $f(\pi|-) = Be(a_1 + p_\gamma, b_1 + p - p_\gamma)$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**Step 8** Update $\gamma_j$'s one at a time by computing their posterior inclusion probabilities after marginalizing out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and conditional on inclusion indicators for the remaining predictors as well as $g$, $\tau$ and $A$. Denote $\boldsymbol{\gamma}(j)$ and $\boldsymbol{\gamma}(-j)$ as the vector of current variable inclusion indicators with $\gamma_j$ fixed at 1 and 0 respectively, and let $p_{\boldsymbol{\gamma}(j)}$ and $p_{\boldsymbol{\gamma}(-j)}$ denote the corresponding vector sums. We can sample $\gamma_j$ from the Bernoulli conditional posterior distribution with probabilities $\Pr(\gamma_j = 1|-) = p_{j1}/(p_{j1} + p_{j0})$ and $\Pr(\gamma_j = 0|-) = p_{j0}/(p_{j1} + p_{j0})$, where

$$p_{j1}=\pi(1+g)^{-p_{\gamma(j)}/2}\exp\left\{\frac{\tau}{2}\left(\frac{g}{1+g}\right)\left((\boldsymbol{Y}^n)^T\sum_A^{-1}\mathbf{X}_{\gamma(j)}(\mathbf{X}'_{\gamma(j)}\sum_A^{-1}\mathbf{X}_{\gamma(j)})^{-1}\mathbf{X}'_{\gamma(j)}\sum_A^{-1}\boldsymbol{Y}^n\right)\right\},$$
$$p_{j0}=(1-\pi)\,(1+g)^{-p_{\gamma(-j)}/2}\exp\left\{\frac{\tau}{2}\left(\frac{g}{1+g}\right)\left((\boldsymbol{Y}^n)^T\sum_A^{-1}\mathbf{X}_{\gamma(-j)}(\mathbf{X}'_{\gamma(-j)}\sum_A^{-1}\mathbf{X}_{\gamma(-j)})^{-1}\mathbf{X}'_{\gamma(-j)}\sum_A^{-1}\boldsymbol{Y}^n\right)\right\}.$$

**Step 9** Set $\{\beta_j : \gamma_j = 0\} = 0$ and update $\boldsymbol{\beta}_{\boldsymbol{\gamma}} = \{\beta_j : \gamma_j = 1\}$ using $\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|-) = N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}; \mathbf{E}, \mathbf{V})$, where $\mathbf{V}=\left(\frac{\tau}{g}(\mathbf{X}'_\gamma\sum_A^{-1}\mathbf{X}_\gamma)+\tau(\mathbf{X}'_\gamma\mathbf{X}_\gamma)\right)^{-1}$ and $\mathbf{E}=\mathbf{V}\left(\tau\mathbf{X}'_\gamma(\boldsymbol{Y}^n-\mathbf{A}\boldsymbol{\eta})\right)$.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



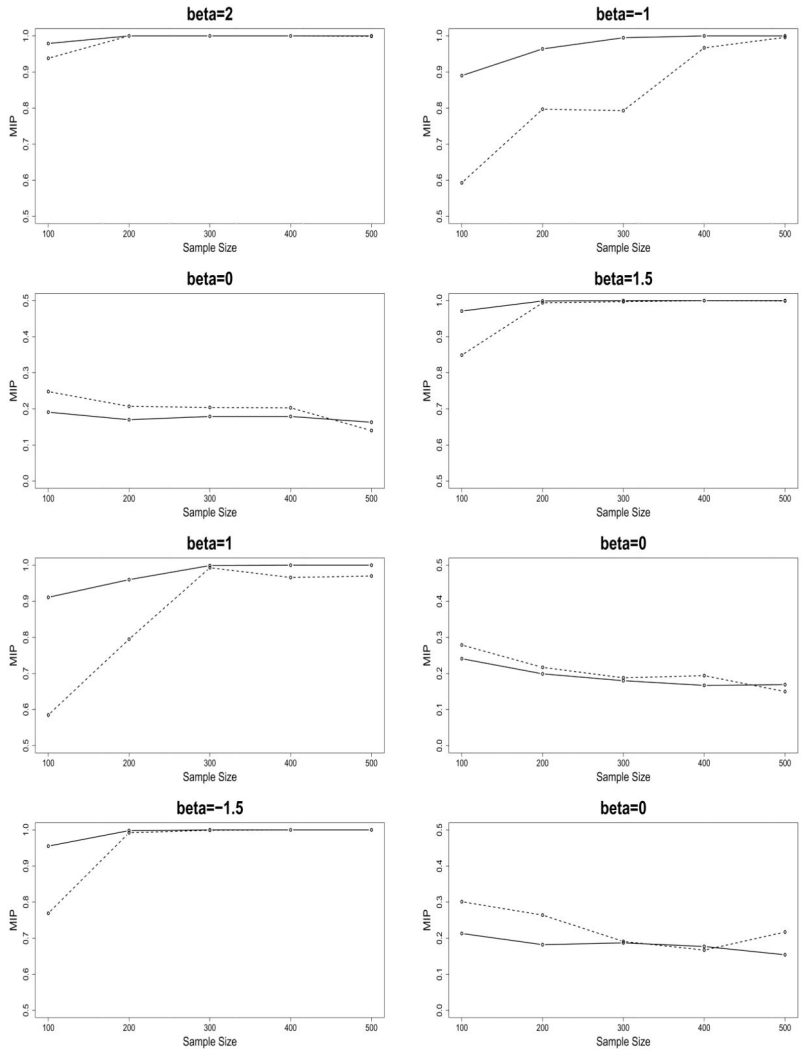


**Figure 1.**

Marginal Inclusion Probabilities (MIP) over varying sample sizes: Truth generated from bimodal residual. Solid lines - Semi-parametric Linear Model, dashed lines - Normal Linear Model.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

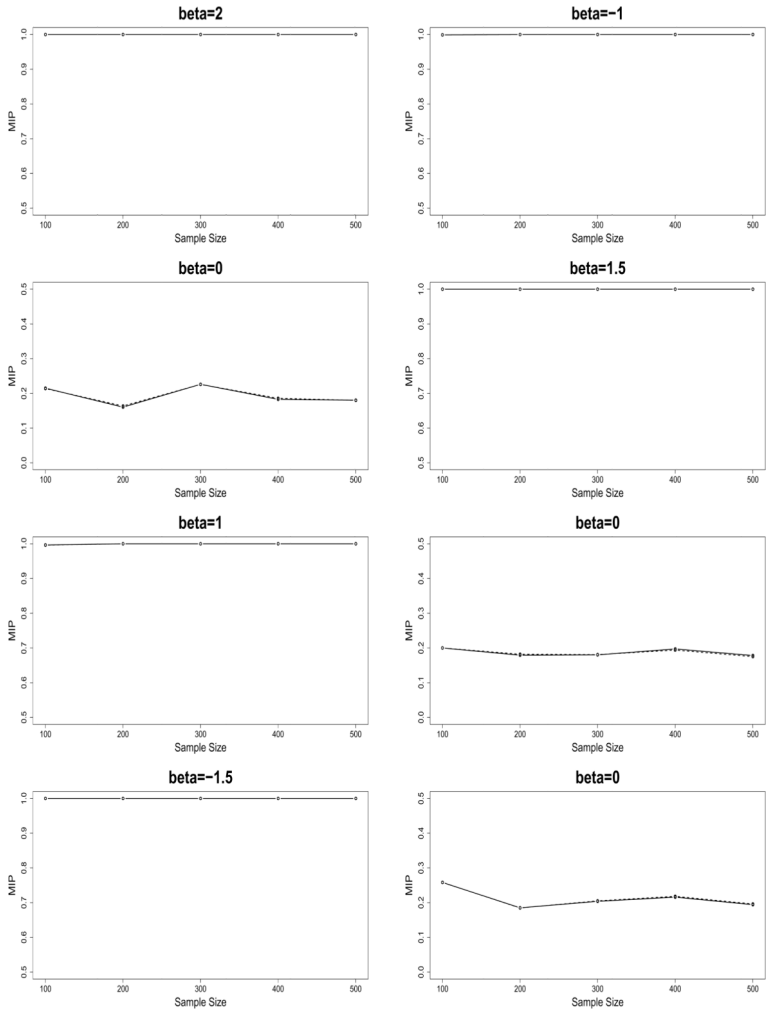


**Figure 2.**

Marginal Inclusion Probabilities (MIP) over varying sample sizes: Truth generated from Gaussian residual. Solid lines - Semi-parametric Linear Model, dashed lines - Normal Linear Model.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

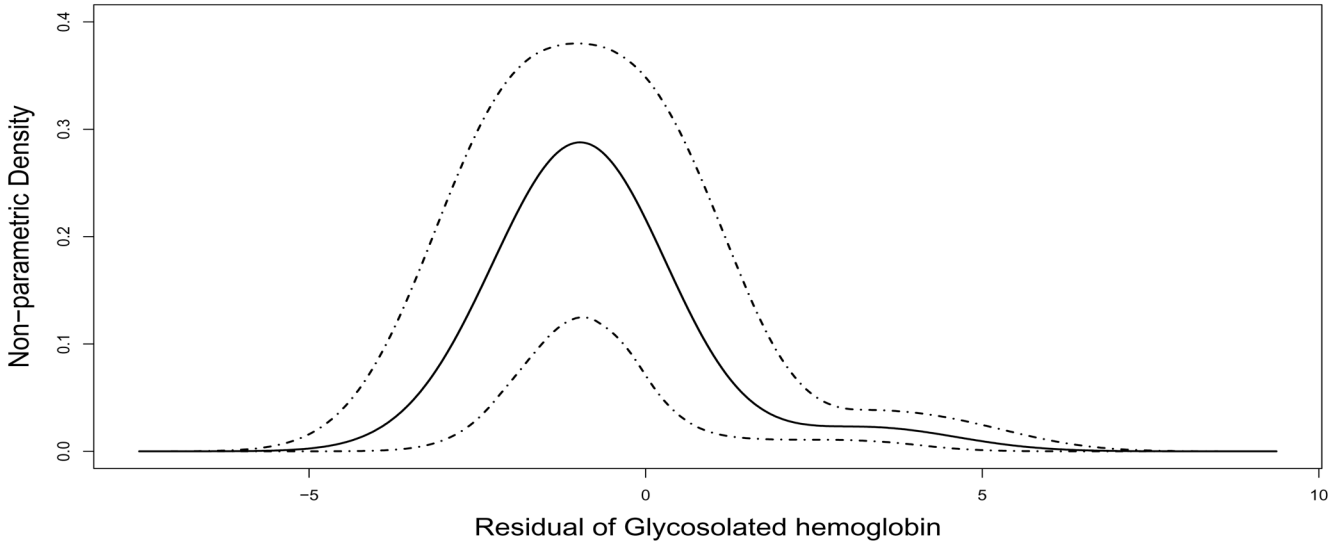


**Figure 3.**
Residual density for Type II Diabetes study under Semi-parametric Linear Model.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



*J Am Stat Assoc*. Author manuscript; available in PMC 2015 March 01.

**Table 1**

Fixed effects estimates and marginal inclusion probabilities (MIP) for fixed effects for Case I when n=100.

| $\beta_T$ | $\text{MIP}_{SLM}$ | $\beta_{SLM}$ | $\text{MIP}_{NLM}$ | $\beta_{SLM}$ | $\beta_{L1}$ | $\beta_{EL}$ | $\beta_{LMR}$ | $\beta_{QR}$ |
|---|---|---|---|---|---|---|---|---|
| 3 | 1.00 | 2.88(2.34, 3.41) | 1.00 | 2.83(1.86, 3.81) | 3.08 | 3.08 | 3.15 | 2.92 |
| 2 | 0.99 | 1.89(1.34, 2.44) | 0.98 | 1.95(0.96, 2.91) | 2.06 | 2.06 | 2.11 | 1.84 |
| −1 | 0.93 | −0.91(−1.46, −0.36) | 0.75 | −0.78(−1.75,0.03) | −0.98 | −0.98 | −0.87 | −0.78 |
| 0 | 0.45 | −0.01(−0.44,0.44) | 0.53 | 0.006(−0.82, 0.81) | 0.01 | 0.009 | −0.003 | −0.02 |
| 1.5 | 0.98 | 1.43(0.89,1.98) | 0.90 | 1.35(0.35, 2.35) | 1.54 | 1.54 | 1.57 | 1.29 |
| 1 | 0.90 | 0.79(0.28, 1.35) | 0.68 | 0.54(−0.26, 1.48) | 0.74 | 0.74 | 0.66 | 0.42 |
| 0 | 0.43 | −0.005(−0.44, 0.42) | 0.53 | −0.05(−0.85, 0.73) | −0.04 | −0.04 | −0.09 | −0.06 |
| −4 | 1.00 | −3.89(−4.43, −3.33) | 1.00 | −3.75(−4.74, −2.74) | −4.05 | −4.04 | −4.14 | −3.95 |
| −1.5 | 0.99 | −1.54(−2.08, −0.98) | 0.92 | −1.43(−2.41, −0.41) | −1.57 | −1.57 | −1.54 | −1.30 |
| 0 | 0.42 | 0.008(−0.43, 0.43) | 0.54 | −0.12(−0.93, 0.64) | −0.12 | −0.12 | −0.06 | −0.14 |

SLM: Semi-parametric linear model, NLM: Normal linear model, L1: Lasso, EL: Elastic Net, LMR: MM-type estimator, QR: Median regression with SSVS.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 2**

Summaries for Case I when n=100.

| Measure | SLM | NLM | L1 | EL | LMR | QR |
|---|---|---|---|---|---|---|
| MSE around $\boldsymbol{\beta}_T$ | 0.07 | 0.21 | 0.24 | 0.24 | 0.40 | 0.50 |
| MSE for out of sample prediction | 7.70 | 16.44 | 8.33 | 8.32 | 8.83 | 9.11 |

SLM: Semi-parametric linear model, NLM: Normal linear model, L1: Lasso, EL: Elastic Net, LMR: MM-type estimator, QR: Median regression with SSVS. MSE: mean square error and $\boldsymbol{\beta}_T$ is the vector of true regression coefficients.

*J Am Stat Assoc*. Author manuscript; available in PMC 2015 March 01.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript





*J Am Stat Assoc*. Author manuscript; available in PMC 2015 March 01.

**Table 3**

Fixed effects (times 100) for type-II diabetes example.

| **Predictor** | $\hat{\beta}_{SLM}$ | $\hat{\beta}_{NLM}$ | $\hat{\beta}_{L1}$ | $\hat{\beta}_{EL}$ | $\hat{\beta}_{LMR}$ | $\hat{\beta}_{QR}$ |
|---|---|---|---|---|---|---|
| TC | 0.55(0.11,0.73) | 0.74(0.25,1.20) | 0.75 | 0.75 | 0.29 | 0.01 |
| SG | 2.11(1.75,2.48) | 2.82(2.5,3.15) | 2.83 | 2.82 | 2.99 | 3.23 |
| HDL | −0.50(−1.4,0.015) | −0.36(−1.61,0) | −1.02 | −1.02 | −0.42 | 0 |
| Age | 0.34(−0.06,1.3) | 0.98(0,2.35) | 1.19 | 1.19 | 0.57 | 0.04 |
| Gender | −3.72(−30.12,4.39) | −1.53(−25.46,3.22) | −19.66 | −19.81 | −7.87 | −0.86 |
| BMI(overwt) | 1.55(−9.43,24.03) | 2.04(−3.33,29.53) | 4.33 | 4.27 | 15.12 | 1.84 |
| BMI(obese) | −0.74(−20.33,13.44) | −0.91(−21.93,6.14) | −14.88 | −15.03 | 8.16 | 0.62 |
| SBP | 0.53(0,1.35) | 0.03(−0.13,0.65) | 0.25 | 0.25 | 0.56 | 0.009 |
| DBP | −0.03(−0.99,0.69) | 0(−0.45,0.45) | 0.018 | 0.017 | −0.55 | 0.002 |
| WHR | 224.27(67.72,381.88) | 3.16(−44.74,91.4) | 90.47 | 91.53 | 90.79 | 129.23 |
| PPT | 21.42(1.89,57.49) | 33.04(0,80.39) | 47.31 | 47.32 | 37.55 | 18.99 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 4**

Marginal Inclusion Probabilities for SLM, NLM, QR in type-II diabetes data.

| Predictor | TC | SG | HDL | Age | Gender | BMI(overwt) | BMI(obese) | SBP | DBP | WHR | PPT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $MIP_{SLM}$ | 0.97 | 1.00 | 0.64 | 0.43 | 0.17 | 0.15 | 0.22 | 0.72 | 0.23 | 0.93 | 0.64 |
| $MIP_{NLM}$ | 0.98 | 1.00 | 0.39 | 0.67 | 0.12 | 0.13 | 0.11 | 0.14 | 0.10 | 0.13 | 0.68 |
| $MIP_{QR}$ | 0.02 | 1.00 | 0.002 | 0.03 | 0.08 | 0.10 | 0.08 | 0.01 | 0.004 | 0.71 | 0.42 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 5**

Out of Sample Prediction.

| Replicate | S 1 | S 2 | S 3 | S 4 | S 5 | S 6 | S 7 | S 8 |
|---|---|---|---|---|---|---|---|---|
| $MSE_{SLM}$ | 1.25 | 1.24 | 1.55 | 1.21 | 1.45 | 1.47 | 3.44 | 1.23 |
| $MSE_{NLM}$ | 1.23 | 1.33 | 1.74 | 1.29 | 1.14 | 1.46 | 3.43 | 1.52 |
| $MSE_{L1}$ | 1.28 | 1.45 | 2.49 | 2.34 | 1.13 | 1.45 | 3.47 | 1.75 |
| $MSE_{EL}$ | 1.29 | 1.47 | 2.51 | 2.36 | 1.14 | 1.45 | 3.48 | 1.75 |
| $MSE_{LMR}$ | 2.23 | 1.21 | 2.15 | 1.02 | 1.09 | 1.36 | 4.06 | 1.69 |
| $MSE_{QR}$ | 1.82 | 1.91 | 2.64 | 1.15 | 1.64 | 2.68 | 3.98 | 2.44 |
| $Cov_{SLM}$ | 100.00 | 97.14 | 100.00 | 97.14 | 100.00 | 100.00 | 91.42 | 100.00 |
| $Cov_{NLM}$ | 97.12 | 97.14 | 94.28 | 97.14 | 100.00 | 97.14 | 91.42 | 100.00 |
| $CIW_{NLM}$ | 5.92 | 5.41 | 5.84 | 5.94 | 5.93 | 5.91 | 5.59 | 5.90 |
| $CIW_{SLM}$ | 6.93 | 6.16 | 6.80 | 6.81 | 6.84 | 6.86 | 6.13 | 6.77 |

MSE: out of sample predictive mean square error, Cov: 95% credible interval coverage, CIW: 95% credible interval width.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 6**

Auto-correlations across lags for fixed effects in type-II diabetes data.

| Predictor | Lag 1 | | Lag 5 | | Lag 10 | | Lag 25 | | Lag 50 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SLM | NLM | SLM | NLM | SLM | NLM | SLM | NLM | SLM | NLM |
| TC | 0.22 | 0.18 | 0.113 | 0.194 | 0.073 | 0.159 | 0.032 | 0.111 | 0.013 | 0.059 |
| SG | 0.59 | 0.06 | 0.386 | 0.038 | 0.285 | 0.022 | 0.14 | 0.009 | 0.06 | 0.016 |
| HDL | 0.19 | 0.02 | 0.081 | 0.012 | 0.041 | 0.013 | 0.01 | 0.021 | 0.0005 | −0.006 |
| Age | 0.21 | 0.04 | 0.072 | 0.009 | 0.053 | −0.0001 | 0.025 | 0.006 | 0.007 | −0.014 |
| Gender | 0.06 | −0.007 | 0.030 | 0.0003 | 0.013 | −0.006 | 0.009 | −0.014 | 0.005 | 0.019 |
| BMI(overwt) | 0.02 | −0.002 | 0.01 | −0.006 | 0.006 | 0.013 | −0.006 | 0.009 | 0.0014 | 0.018 |
| BMI(obese) | 0.02 | 0.002 | 0.017 | 0.004 | 0.004 | 0.018 | 0.007 | −0.003 | 0.000 | 0.000 |
| SBP | 0.29 | 0.0711 | 0.137 | 0.019 | 0.096 | 0.007 | 0.047 | 0.03 | 0.014 | 0.022 |
| DBP | 0.07 | 0.0239 | 0.021 | 0.019 | 0.019 | 0.031 | 0.009 | −0.003 | 0.004 | −0.012 |
| WHR | 0.44 | 0.0642 | 0.353 | 0.043 | 0.321 | 0.061 | 0.251 | 0.06 | 0.186 | −0.003 |
| PPT | 0.22 | 0.0600 | 0.118 | 0.047 | 0.068 | 0.045 | 0.015 | 0.004 | −0.002 | 0.019 |