Anna Freni Sterrantino

Varying coefficient models as Mixed Models: reparametrization methods and bayesian estimation
Quaderni di Dipartimento

Serie Richerche 2013, n. 5
ISSN

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche
"Paolo Fortunati"

Abstract

Non-linear relationships are accommodated in a regression model using smoothing functions. Interaction may occurs between continuous variable, in this case interaction between non-linear and linear covariate leads to varying coefficient model (VCM), a subclass of generalized additive model.
Additive models can be estimated as generalized linear mixed models, after being reparametrized.
In this article we show three different type of matrix design for mixed model for VCM, by applying b-spline smoothing functions. An application on real data is provided and model estimates are computed with a Bayesian approach.

Introduction

Non-linear relations between covariates and response variable are integrated in a regression model by smooth functions. In presence of interaction between a smooth function and a variable, we deal with Varying coefficients models (VCM)(Hastie and Tibshirani, 1993), a subclass of Generalized additive models (GAM).
Both classes share estimation methods, like Penalized Iteratively reweighted least squares (P-IRLS) and software procedures.
Alternative way to treat generalized additive models is to reparametrize them into a Generalized Linear Mixed Models (GLMM) and then estimate with maximum likelihood or penalized quasi-likelihood methods, (Durban, 2009; Brumback and Rice, 1998; Verbyla et al., 1999) or by Bayesian methods (Crainiceanu et al., 2005; Fahrmeir and Lang, 2001).
Wrapping GAM, as a generalized linear mixed model, relays on a smoothing basis anchored on knots and on a penalty parameter to ensure fitting roughness. This processes is performed by a varied of basis, like: thin plate splines (Wood, 2006), p-splines (Marx, 2010) or truncated polynomial.
We present a reparametrization of a varying coefficient model - a GAM model with an interaction term - in a GLMM form with p-splines basis (i); three ways on how to express non-linear effect in fixed and random effect matrices (ii) and provide code for matrices set up and for Bayesian estimation (iii).
The paper is structured as follows: section (1) introduces a non-linear regression and linear mixed model frame. Section (2) presents varying coefficient models with a generic spline basis. Generalized linear mixed model representation is described in section (3). B-spline basis and the three ways to reparametrize a VCM are introduced in section (4) and (5). Finally, an example is provided on real data in section (6). Discussion and appendix follow respectively.

1 Non-linear regression

A non-linear regression holds for a variable y that depends non-linearly on a variable x, for the i-th observation equation model is:

$$y_i = f_1(x_i) + \epsilon_i \tag{1}$$

where $\epsilon \sim N(0, \sigma^2 I)$, $f_1(\cdot)$ a smoothing function, and y a Gaussian distribution. If $\widetilde{X}$ is a generic basis for the smoothing function $f_1(\cdot)$, then

$$y_i = \sum_j^{q_j} \beta_j b_j(x_i) + \epsilon_i$$

with β_j and $b_j(x_i)$ respectively smoothing coefficients and basis function.
A spline basis is constructed on knots q_j - a selection of observed x_i values - to limit smoothness and allow for wiggliness. In matrix form $f_1(x_1) = \widetilde{\mathbf{X}}\beta$, wiggliness is translated as a penalization of smoothing coefficients controlled by a penalty parameter λ:

$$y = \widetilde{\mathbf{X}}\beta + \lambda\beta^T \mathbf{P}\beta.$$

To find model estimates, we minimize

$$\mathcal{S}(\beta) = (y - \widetilde{\mathbf{X}}\beta)^T (y - \widetilde{\mathbf{X}}\beta) + \lambda\beta^T \mathbf{P}\beta$$

with being P a penalty matrix.
For a given λ, the solution to the optimization problem satisfies:

$$(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}} + \lambda\mathbf{P})\hat{\beta} = \widetilde{\mathbf{X}}^T y$$

then $\hat{y} = \widetilde{\mathbf{X}}\hat{\beta} = \mathbf{H}y$ with hat matrix $\mathbf{H}$:

$$\mathbf{H} = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}} + \lambda\mathbf{P})^{-1} \widetilde{\mathbf{X}}^T.$$

1.1 Adding linear covariates

Suppose we would like to include other linear covariates, indicated in matrix form X^*, then:

$$y = \mathbf{X}^*\theta + \widetilde{\mathbf{X}}\beta + \epsilon \tag{2}$$

we minimize for (θ, β) :

$$\mathcal{S}(\theta, \beta) = (y - \mathbf{X}^*\theta - \widetilde{\mathbf{X}}\beta)^T (y - \mathbf{X}^*\theta - \widetilde{X}\beta) + \lambda\beta^T \mathbf{P}\beta.$$

that satisfies:

$$\begin{bmatrix} \mathbf{X^{*T}X} & \mathbf{X^{*T}}\widetilde{\mathbf{X}} \\ \widetilde{\mathbf{X}}^{\mathbf{T}}\mathbf{X} & \widetilde{\mathbf{X}}^{\mathbf{T}}\widetilde{\mathbf{X}} + \lambda\mathbf{P} \end{bmatrix} \begin{bmatrix} \hat{\theta} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{X^{*T}} \\ \widetilde{\mathbf{X}}^{\mathbf{T}} \end{bmatrix} y$$

Estimation method is penalized iteratively re-weighted least squares (P-IRLS) or maximum likelihood, while optimal penalty coefficient value is found by generalized cross validation (GCV), see Wood (2006) for more details.
But, both models in equation (1) and (2) could be re-parametrize into a linear mixed model form:

$$y = \mathbf{X}\beta + \mathbf{Z}u + \epsilon \tag{3}$$

with $\epsilon \sim N(0, \sigma_\epsilon^2 I)$ and $u \sim N(0, \sigma_u^2 I)$, and

$$cov(y) = \begin{pmatrix} \sigma_\epsilon^2 I & 0 \\ 0 & \sigma_u^2 I \end{pmatrix}$$

The internal structure of $\mathbf{X}$ and $\mathbf{Z}$ matrices depends by the smoothing basis, with fixed matrix containing observed predictors and random matrix incorporating penalties.

2 Varying coefficient models: definition

A varying coefficient model is defined as a generalized additive model with an interaction effect between a smoothed function that represents a non-linear relationship and a continuous variable (binary or categorical variables are included too). Given the following model:

$$g(\mu_i) = \mathbf{X_i^*}\theta + f_1(x_{1i}) + f_2(x_{2i})s_i + ... \tag{4}$$

$\mu_i = E(y_i)$ and y_i belonging to the exponential family, g a known monotonic link function, $\mathbf{X_i^*}$ i-th row ($i = 1, \ldots, n$-observations) of model matrix for any parametric model components, θ parameter vector, f_j's smooth function for x_j covariates and s a linear variable that interacts with x_2.
Given a generic basis b_{ji} for each function (more in section 4), the j-th smoothing function is:

$$f_j(x_j) = \sum_{j}^{q_j} \beta_j b_j(x_i)$$

A model matrix is defined as $\widetilde{\mathbf{X}}_\mathbf{j}$, for each j. Therefore $\mathbf{f_j}$ is a vector, such that $\mathbf{f_j}_i = f_j(x_{ji})$ and $\widetilde{\beta}_\mathbf{j} = [\beta_{j1}, \ldots, \beta_{jq_j}]^T$, and

$$\mathbf{f_j} = \widetilde{\mathbf{X}}_\mathbf{j,ik} = \mathbf{b_{jk}}(\mathbf{x_i})\widetilde{\beta}_\mathbf{j} = \widetilde{\mathbf{X}}_\mathbf{j}\widetilde{\beta}_\mathbf{j}$$

Substituting in equation (4)

$$g(\mu_i) = \mathbf{X_i^*}\theta + \widetilde{X}_1\widetilde{\beta}_1 + \widetilde{X}_2\widetilde{\beta}_2 s_i + ... \tag{5}$$

Because model equation in (5) is not identifiable, each smooth basis has to be centered.
A suitable constraints is that the sum of the elements in model matrix equals zero,

$$\mathbf{1^T}\widetilde{\mathbf{X}}_\mathbf{j}\widetilde{\beta}_\mathbf{j} = 0.$$

An efficient solution use QR decomposition on matrix $\widetilde{\mathbf{X_j}}^\mathbf{T}$[1]. Constraints imply that we redefine $\mathbf{X_j} = \widetilde{\mathbf{X_j}}\mathbf{Z}$, such that $\mathbf{f_j}$ equals $\mathbf{X_j}\beta_j$. The final form for a varying coefficient model is:

$$g(\mu) = \mathbf{X}^* + \mathbf{X_1}\beta_1 + (\mathbf{X_2}\beta_2)s + ... \tag{6}$$

[1]See Appendix B

To include the effect of a linear covariate is sufficient to multiple $diag(s_i)$ - a diagonal matrix with all element zero except on the principal diagonal - to model matrix $\mathbf{X_2}$. A compact view of model (6) gives:

$$g(\mu_i) = \mathbf{X_i}\beta$$

with $\mathbf{X} = [\mathbf{X}^* : \mathbf{X_1} : \mathbf{X_2}diag(s) : \ldots]$.
A penalization term to control overfitting is represented by a quadratic form of equation coefficients:

$$\widetilde{\beta}_j^T \tilde{\mathbf{S}}_j \tilde{\beta}_j$$

matrix $\widetilde{\mathbf{S}}_j$ is composed by known coefficients that control j-th wiggliness. Penalty term is subjected to re-parametrization by QR procedure, for identifiably issues, and it becomes

$$\beta_j^T \overline{\mathbf{S}}_j \beta_j$$

with

$$\overline{\mathbf{S}}_j = \mathbf{Z}^T \tilde{\mathbf{S}}_j \mathbf{Z}.$$

For notational practice, penalty is noted as $\beta^T \mathbf{S}_j \beta$, where $\mathbf{S_j}$ equals $\overline{\mathbf{S_j}}$, but has a sparse structure, that

$$\beta^T \mathbf{S}_j \beta = \beta_j^T \overline{\mathbf{S}}_j \beta_j.$$

Varying coefficient model is reduced to:

$$g(\mu) = \mathbf{X}\beta + \lambda \beta^T \mathbf{S}_j \beta$$

3 Varying coefficient model in mixed model formulation

Given the non-parametric model:

$$g(\mu) = \mathbf{X}\beta + \lambda \beta^T \mathbf{S}_j \beta \tag{7}$$

the aim is to re-parametrize in a parametric form of equation (3) For each $f_j(x_j)$ there is a model matrix and a wiggliness function, respectively $b_j(x_i) = \mathbf{X}^f$, $J(f) = \beta^T \mathbf{S}_j \beta$.
Because we want to compress equation (7) into fixed and random matrices, smoothing coefficients βs are partitioned in β_F for fixed effect and b_R for random effects.
A prior distribution for smooth function is

$$\mathbf{f}_\beta(\beta) \propto exp(-\lambda \beta^T \mathbf{S}^T \beta / 2). \tag{8}$$

But, it is an improper prior for β, given $\mathbf{S}$ rank deficiency.
Re-parametrization brings two groups of parameters: one with proper prior for random effects and one with improper prior for fixed effect. Penalty matrix $\mathbf{S}$ has to be strictly positive and given the singular value decomposition (svd) $\mathbf{S} = \mathbf{U^T D U}$, we extract $\mathbf{D}$ a diagonal matrix, with eigenvalues on its principal diagonal arranged in descending

order. From $\mathbf{D}$ we recover to submatrices: $\mathbf{D}_+$the smallest submatrix that contains only positive eigenvalues and $\mathbf{D}_*$ that contains on its diagonal the remaining ordered eigenvalues.

$$D = \begin{bmatrix} \mathbf{D}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_* \end{bmatrix} \tag{9}$$

Penalty acts only on random effect coefficients, and parameter vector is partitioned:

$$(b_R^T, \beta_F^T)^{=}\mathbf{U}^{\mathbf{T}}\beta$$

substituting in penalty term of equation (7)

$$\beta^T \mathbf{S} \beta = \beta^T \mathbf{U}\mathbf{D}\mathbf{U}^{\mathbf{T}}\beta = (b_R, \beta_F^T)\mathbf{D}(b_R^T, \beta_F^T)^T = \\ (b_R, \beta_F^T)[\mathbf{D}_+ : \mathbf{D}_*](b_R^T, \beta_F^T)^{=} b_R^T \mathbf{D}_+ b_R$$

is clear that matrix $\mathbf{D}_+$ penalizes only β_R. Smooth distribution for β_R, with proper prior for random effects is:

$$\mathbf{f}_{\mathbf{b_R}}(\mathbf{b_R}) \propto exp(-\lambda b_R^T \mathbf{D}_+{}^T b_R/2)$$

Whereas, random coefficients distribution is:

$$b_R \sim N(0, \mathbf{D}_+{}^{-1})/\lambda.$$

Orthogonal matrix, resulting from svd penalty matrix, is split in two block matrices $\mathbf{U} = [\mathbf{U_R} : \mathbf{U_F}]$. Two matrices composed by eigenvectors associated with null eigenvalues $\mathbf{U_F}$ and eigenvectors associated with positive eigenvalues $\mathbf{U_R}$.
Fixed and random matrices are:

$$\mathbf{X_F} = \mathbf{X^f U_F} \\ \mathbf{X_R} = \mathbf{X^f U_R}$$

Additional transformation on random coefficients

$$b = \sqrt{\mathbf{D}_+{}^{-1}} b_R$$

and

$$\mathbf{Z} = \mathbf{X_R}\sqrt{\mathbf{D}_+}$$

lead to $\mathbf{X_F}\beta_{\mathbf{F}} + \mathbf{Zb}$, and

$$b \sim N(0, \mathbf{I}/\lambda).$$

obtaining desired distribution for random coefficients, stated in equation (3). No particular spline basis for $\widetilde{\mathbf{X}}$ and or any penalty matrices $\mathbf{S}$, have been selected.

4 B-spline basis

Non-linear relationship is arranged in a regression model by interpolating data or by fitting a truncated power function:

$$1, x, \ldots, x^p, (x-k_1)_+^p, \ldots, (x-k_k)_+^p$$

with k indicating data observations called knots ($k < n$) selected to anchor the smoothing function. The number of knots or its location is not so relevant for the fitted model, as stated by (Ruppert, Wand, and Carroll, 2003). A general rule of thumb is the following:

knots locations: $k_k = (\frac{k+1}{K+2})th$ sample quantile of unique x_i

knots number: $K = \min(1/4 \times \text{number of unique} x_i, 35)$

A generic B-spline is computed as differences of truncated power functions, cubic b-spline basis is generated by four polynomial of degree three and is non-zero within the range off knots.
A bspline is defined recursively as

$$\begin{array}{cc} b_1(x) = 1 & b_2(x) = x_i \\ b_3(x) = R(x_i, k_1) & b_4(x) = R(x_i, k_2) \\ b_k(x) = R(x_i, k_k) & \end{array}$$

And function $R(x, k)$ assures that cubic polynomial segments of data are connected smoothly between them

$$R(x,k) = \frac{1}{4} \times \left(\left(k - \frac{1}{2}\right)^2 - \frac{1}{12} \right) \times \left(\left(x - \frac{1}{2}\right)^2 - \frac{1}{12} \right)$$
$$-\frac{1}{24} \times \left(\left(|x-k| - \frac{1}{2}\right)^4 - \frac{1}{2} \times \left(|x-k| - \frac{1}{2}\right)^2 + \frac{7}{240} \right)$$

In Figure 1 we plot b-splines for different degree value, (Wood, 2006).

Penalties matrices are basis specific and consent for smoothing flexibility while avoiding over fitting. For b-splines, penalty matrix is a based on a difference operator of order u, (Eilers and Marx, 1996; Eilers and Marx, 2004).
For cubic b-splines, a second order difference penalty is sufficient. First order differences penalize too big jumps between successive parameters and second order differences penalize deviations from linear trends. A bspline accompanied by a penalty matrix is called p-spline (ps).

5 Varying coefficient model reparametrization with B-spline

Three different fixed and random effect matrices design are presented. All methods are adapted for a p-spline basis function and for varying coefficient models, given the starting model form :

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) s_i + \epsilon$$

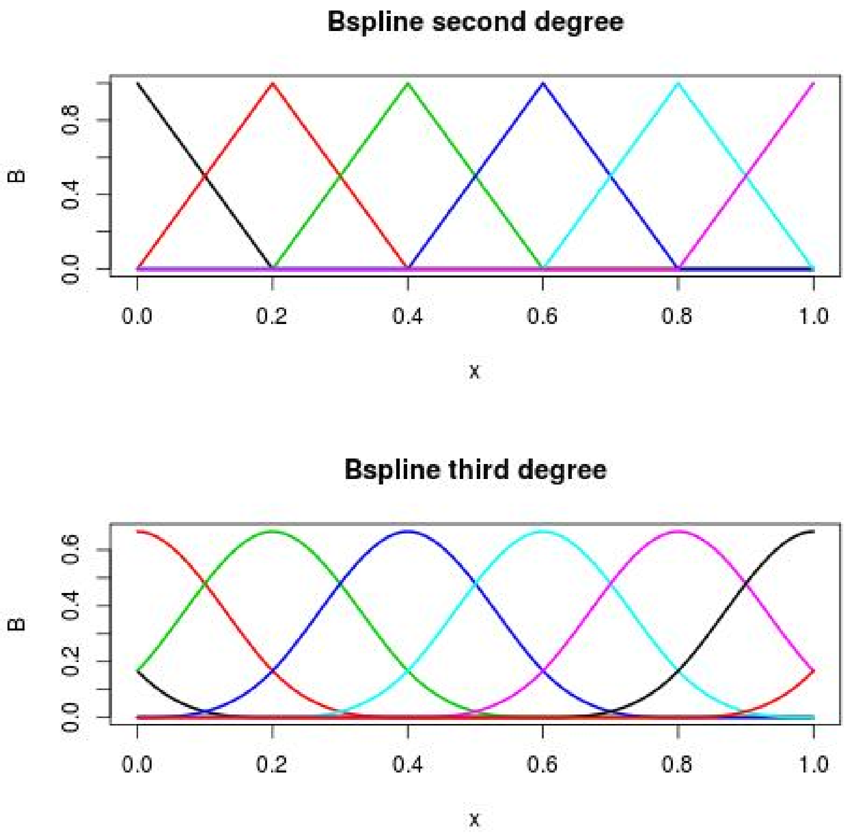


Figure 1: Bspline for different degree and 10 internal knots

we want to specify:

$$g(\mu_i) = \mathbf{X}\beta + \mathbf{Z}u + \epsilon.$$

For each non-linear covariate, a b-spline matrix has to be computed. Given a j-th non-linear predictor, (subscript j omitted) $\mathbf{x} = [x_1, \ldots, x_n]$ a B cubic b-spline matrix is computed and has dimension $n \times (k + deg)$, deg is polynomial degree -for cubic b-spline equals 3 - and k the number of internal knots.
A penalty matrix $\mathbf{P}$ is derived by assuming a difference diagonal matrix of order 2,

$$\mathbf{P} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots \\ 0 & 1 & -2 & 1 & 0 & \ldots \\ 0 & 0 & 1 & -2 & 1 & \ldots \\ \ldots & & & & \ldots & \ldots \end{bmatrix}$$

of dimension $(k + deg) \times (k + deg) - 2$.
Penalty matrix is defined, as $\mathbf{S} = \mathbf{P^tP}$. And singular value decomposition

$$\mathbf{PP^T_{(k+deg)\times(k+deg)}} = \mathbf{UDU^T}$$

The block matrix

$$\mathbf{U} = [\mathbf{U_F}, \mathbf{U_R}]$$

has $\mathbf{U}_{\mathbf{F}(k+deg)\times 2}$ and $\mathbf{U}_{\mathbf{R}(k+deg)-2\times 2}$ In table 1 we reported the three matrix design, for a GAM and for a varying coefficient model. For the three designs, fixed and random matrices are orthogonal and full column rank.

5.1 Matrix design: I

This first parametrization has been seen in Breitner (2007) to fit a time varying coefficient model, and in Currie, Durban, and Eilers (2006).
In this case, fixed effect matrix is defined:

$$\mathbf{X_1} = [1\,, x_1]$$

and random matrix is

$$\mathbf{Z_1} = \mathbf{B_1 U_R D_+^{-1/2}}$$

with $\mathbf{B_1}$ b-spline matrix associated with the first predictor x_1, and $\mathbf{D}_+$ equal to definition in been in equation(9).
For our second additive member of equation: $f_2(x_2)s$, matrix b-spline is equal to

$$diag(s_1, s_2, \ldots, s_n)\mathbf{B_{2s}}.$$

Without loss of generality, a unique penalty matrix is defined, for both smooth function as long knots number is the same for both. Random matrix is

$$\mathbf{Z_2} = \mathbf{B_{2s} U_R D_+^{-1/2}}$$

for fixed matrix, given the two zeros eigenvalues, it may be composed as follow

$$\mathbf{X_2} = [x_2, x_2 s].$$

Random and fixed matrices are combined in block matrices:

$$\mathbf{X} = [1, x_1, x_2, x_2 s, X^*]$$

$$\mathbf{Z} = [\mathbf{B_1 U_R D_+^{-1/2}} | \mathbf{B_{2s} U_R D_+^{-1/2}}].$$

If this parametrization is used, it is necessary to ensure that block matrix is $[\mathbf{X} : \mathbf{Z}]$ is full column rank and $\mathbf{X^T Z} = 0$, i.e. orthogonal. Both are full column rank,

$$rank(\mathbf{Z} = [\mathbf{B_1 U_R D_+^{-1/2}} | \mathbf{B_{2s} U_R D_+^{-1/2}}]) = (k + deg) - 2$$

and

$$rank(\mathbf{X} = [1, x_1, x_2, x_2 s]) = 4$$

Whereas, orthogonality is ensured by transformation applied to obtain this reparametrization, for details (Lee, 2010).

5.2 Matrix design: II

Second type of parametrization was adapted to p-splines by Durban (2009) in generalized linear models context, following the approach of Brumback and Rice (1998) that have reformulate the two matrices in equation (as a solution to an optimization problem).
For the term with no interaction, fixed matrix:

$$\mathbf{X} = [1, x_1, x_2, x_2 s, X^*]$$

while random matrices

$$\mathbf{Z_1} = \mathbf{B_1 U D^{-1/2}}$$

With $\mathbf{D}$ is defined in equation(9). Analogously for interaction term

$$\mathbf{Z_2} = \mathbf{B_{2s} U D^{-1/2}}$$

Compact block matrix:

$$\mathbf{Z} = [\mathbf{B_1 U D^{-1/2}} | \mathbf{B_{2S} U D^{-1/2}}] \qquad (10)$$

If the number of knots, b-spline degree and order of penalty matrix are same, $\mathbf{U}$ and $\mathbf{D}$ are identical for both smooth functions in the model.

5.3 Matrix design: III

Third parametrization is found in Verbyla et al. (1999), following indication from Green and Silverman (1994) who apply a cubic smoothing spline. Whereas, Currie and Durban (2002) adapts this parametrization procedure to p-splines, defining first smooth term

$$\mathbf{X_1} = \mathbf{B_1 G}$$

and

$$G = \begin{pmatrix} 1 & k_1 & k_1^2 & & k_1^{(deg-1)} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & k_k & k_k^2 & & k_k^{(deg-1)} \end{pmatrix}$$

With *deg* being the b-spline degree is cubic, matrix $\mathbf{G}$ is reduced to $\mathbf{G} = [1\,, k]$, where q are the selected knots. Random effect matrix, for term is defined as:

$$\mathbf{Z_1} = \mathbf{B_1 D^T (D D^T)^{-1}}$$

For interaction term:

$$\mathbf{X_2} = \mathbf{B_{2s} G}$$

and

$$\mathbf{Z_1} = \mathbf{B_{2s} D^T (D D^T)^{-1}}$$

Fixed and random effect matrices are:

$$\mathbf{X}_{=}[\mathbf{B_1 G} | \mathbf{B_{2s} G} | \mathbf{X}^*]$$

$$\mathbf{Z} = [\mathbf{B_1 D^T (D D^T)^{-1}} | \mathbf{B_{2s} D^T (D D^T)^{-1}}]$$

Model	Fixed Effect Matrix	Random Effect Matrix
$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + \epsilon$	$\mathbf{X} = [1, x_1 \vert \mathbf{X}^*]$	$\mathbf{Z} = \mathbf{B U_R D_+^{-1/2}}$
	$\mathbf{X} = [1, x_1 \vert \mathbf{X}^*]$	$\mathbf{Z} = \mathbf{B U D^{-1/2}}$
	$\mathbf{X} = [\mathbf{B} \vert 1, k \vert \mathbf{X}^*]$	$\mathbf{Z} = \mathbf{D^T (D D^T)^{-1}}$
$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) s_i + \epsilon$	$\mathbf{X} = [1, x_1, x_2, x_2 s, X^*]$	$\mathbf{Z} = [\mathbf{B_1 U D^{-1/2}} \vert \mathbf{B_{2s} U D^{-1/2}}]$
	$\mathbf{X} = [1, x_1 x_2, x_2 s, X^*]$	$\mathbf{Z} = [\mathbf{B_1 U_R} \mathbf{D}_+^{-1/2} \vert \mathbf{B_{2s} U_R D_+^{1/2}}]$
	$\mathbf{X} = [\mathbf{B_1 G} \vert \mathbf{B_{2s} G}, \mathbf{X}^*]$	$\mathbf{Z} = [\mathbf{B_1 D^T (D D^T)}^{-1} \vert \mathbf{B_{2s} D^T (D D^T)^{-1}}]$

Table 1: Reparametrization scheme into a GLMM, for a generalized additive model and a varying coefficient model, with a penalty matrix of second order and a cubic degree.

6 Illustration on real data

We show the application of two matrix designs with data from Cleveland, Grosse, and Shyu (1991). Ethanol dataset has 88 observations for two predictors: the equivalence ratio (E), a measure of the fuel-air mixture, the compression ratio (C) of the engine; and a response variable NOx, the concentration of nitric oxide and nitrogen dioxide in engine exhaust, normalized by the work done by the engine.
NOx presents a quadratic effect with predictor E and an interaction with compression ratio. As deductible in Figure 2, each scatter-plot depicts the concentration of NOx against engine compression ratio, by equivalence ratio quartile.
Despite linear relation in the four panels between NOx and compression ratio, equivalence ratio modifies this relationships: both slope and intercept vary non parametrically.
Equation model is:

$$NOx = \beta_1(E) + \beta_2(E)C + \epsilon$$

Reparametrization for fixed and random effect

$$NOx = X\beta + Zu$$

with $X = [1,\, C,\, E,\, CE]$ and $Z = [Z_1,\, Z_2]$.

For Bayesian estimates, priors are:

$$\begin{cases} \beta_0, \beta_1 \sim Norm(0, 0.001) \\ \sigma_\epsilon \sim Gamma(10^{-4}, 10^{-4}) \end{cases}$$

We estimated two models, one with a matrix design as in section (5.1) with graphs in Figure 3 and a second model with matrix design in section (5.2), and graphs in Figure 4.
The two figures, similar to those in Hastie and Tibshirani (1993) and Ruppert, Wand, and Carroll (2003), plot slope and intercept - the smoothed terms- versus equivalence ratio.
In detail, plot in Figure 3a shows a quadratic effect between response variable and equivalence ratio predictor. There is a positive linear effect up to 0.85 of equivalence ratio, then declines to zero for higher values.
Plots in Figure 4 are obtained by estimating a model with a random effect equal to equation (10), and conclusion draw for the previous plots are identical, despite some variability in pointwise credible sets.
In appendix A, there are all the steps to perform the computations.

7 Discussion

We show how to reparametrize a varying coefficient model as a generalized linear mixed model, with b-spline basis, and provided three different matrix design for fixed and random matrices, followed by an application on real data and Bayesian estimation.
Crainiceanu, Ruppert, and Wand (2005) published a similar work,

with application of Bayesian estimates for generalized additive models reparametrized by thin plate splines in a mixed model form, but no interaction term.
One motivation behind this work is that p-splines are default basis in BayesX - a software that estimates GAM and VCM - but is not possible for users to decide knots selection or a priori parameters value. WinBUGS software, on the contrary permits to control all estimation aspects, at the price of some consuming computational time.
Bayesian estimation is supported by the fact that frequentist methods for mixed models estimation and non Gaussian data present non-optimal approximation for small dataset, either using WinBUGS or R package like MCMCglmm by Jarrod (2010).
For a pure Bayesian approach of GAM-VCM, we suggest to refer to Lang and Brezger, 2004.
B-splines, as thin plate splines, are natural cubic splines, characterized for numerical stability and computational efficiency and a change of basis do not compromise the model fit.
One limitation of this reparametrization is for complex model, an interaction with tensor products, renders complicate design matrices and GLMM computationally inefficient and less interpretable, but for simple model, GLMM is an elegant solution and well-understand for non statistical public.
Aim of this paper was to show how express a Varying Coefficient Model in a generalized linear mixed model framework. We provided three matrix designs methods on a reparametrization of a VCM in a mixed model, using b-splines smoothing functions.

References

Breitner, S. (2007). "Time-varying coefficient models and measurement error". PhD thesis. LMU Mnchen: Faculty of Mathematics, Computer Science and Statistics.

Brumback, B. A. and J. A. Rice (1998). "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves". English. In: *Journal of the American Statistical Association* 93.443, pp. 961–976. ISSN: 01621459. URL: http://www.jstor.org/stable/2669837.

Cleveland, W. S., E. Grosse, and W. M. Shyu (1991). "Local regression model". In: *Statistical Models in S*. Ed. by J. M. Chambers and T. Hastie. Pacific Grove: Wadsworth and Brooks/Cole.

Crainiceanu, C. M., D. Ruppert, and M.P. Wand (Sept. 2005). "Bayesian Analysis for Penalized Spline Regression Using WinBUGS". In: *Journal of Statistical Software* 14.14, pp. 1–24. ISSN: 1548-7660. URL: http://www.jstatsoft.org/v14/i14.

Currie, D. and M. Durban (2002). "Flexible smoothing with p-splines: a unified approach". In: *Statistical Modeling* 4, pp. 333–349.

Currie, I. D., M. Durban, and P. H. C. Eilers (2006). "Generalized Linear Array Models with Applications to Multidimensional Smoothing". English. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 68.2, pp. 259–280. ISSN: 13697412. URL: http://www.jstor.org/stable/3647569.

Durban, M. (2009). “An introduction to smoothing splines with penalties: P-spline”. In: *Boletin de Estadistica e Investigacion Operativa* vol.25, No.3.vol.25, No.3.
Eilers, P. H. C. and B. D. Marx (1996). “Flexible Smoothing with B-splines and Penalties”. English. In: *Statistical Science* 11.2, pp. 89–102. ISSN: 08834237. URL: http://www.jstor.org/stable/2246049.
Eilers, P.H.C and B.D. Marx (2004). *Splines, Knots, and Penalties.*
Green, P.J.A. and B.W.A. Silverman (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approch.* Monographs on Statistics & Applied Probability. Chapman & Hall. ISBN: 9780412300400. URL: http://books.google.it/books?id=-AIVXozvpLUC.
Hastie, T. and R. Tibshirani (1993). “Varying-Coefficient Models”. English. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 55.4, pp. 757–796. ISSN: 00359246. URL: http://www.jstor.org/stable/2345993.
Jarrod, D. H. (2010). “MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package”. In: *Journal of Statistical Software* 33.2, pp. 1–22. URL: http://www.jstatsoft.org/v33/i02/.
Lang, S. and A. Brezger (2004). “Bayesian P-splines”. In: *J. Computnl Graph. Statist.* 13, pp. 183–212.
Lee, D. (2010). “Smoothing mixed models for spatial and spatio-temporal data”. PhD thesis. Universidad Carlos III De Madrid.
Marx, B. D. (2010). “P-spline Varying Coefficient Models for Complex Data”. English. In: *Statistical Modelling and Regression Structures.* Ed. by Thomas Kneib and Gerhard Tutz. Physica-Verlag HD, pp. 19–43. ISBN: 978-3-7908-2412-4. DOI: 10.1007/978-3-7908-2413-1_2. URL: http://dx.doi.org/10.1007/978-3-7908-2413-1_2.
Ruppert, D., M.P. Wand, and R.J. Carroll (2003). *Semiparametric Regression.* Cambridge University Press.
Verbyla, A. P. et al. (1999). “The Analysis of Designed Experiments and Longitudinal Data by Using Smoothing Splines”. English. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 48.3, pp. 269–311. ISSN: 00359254. URL: http://www.jstor.org/stable/2680826.
Wood, S. (2006). *Generalized Additive Models: An introduction with R.* Chapman & Hall/CRC.

Appendix A: Bugs/R code

This is the code used for the data example. Ethanol data are found in *lattice* R pacakage.

- Compute b-spline matrix for the smoothed variable, with b-spline function from R *splines* package, or see in Eilers and Marx (1996) R code for function *bbase*

```
tpower <- function(x, t, p)
```

```
# Truncated p-th power function
    (x - t) ^ p * (x > t)
bbase <- function(x, xl = min(x), xr = max(x), nseg = 10,deg = 3){
# Construct B-spline basis
    dx <- (xr - xl) / nseg
    knots <- seq(xl - deg * dx, xr + deg * dx, by = dx)
    P <- outer(x, knots, tpower, deg)

    n <- dim(P)[2]
    D <- diff(diag(n), diff = deg + 1) / (gamma(deg + 1) * dx ^ deg)
    B <- (-1) ^ (deg + 1) * P %*% t(D)
    return(B)
}
```

- Covariates are standardized.
- Define X matrix and compute Z by one of the possible designs.

```
X <- cbind(rep(1,nrow(response)), x1, x2,x1*x2)
```

- Random effect matrix with first design

```
B1 <- bbase(x,nseg=5)
sdia <- diag(s)
B2s <-sdia%*%B1
D <- diff(diag(ncol(B1)),diff=2)
P = t(D) %*% D
decomp <- svd(P)
sigma <- diag(decomp$d)
un <- decomp$u[,(ncol(B1)-1):ncol(B1)]
us <-decomp$u[,1:(ncol(B1)-2)]
Z1 <- B1%*%us%*%diag(1/sqrt(decomp$d[1:ncol(B1)-2]))
Z2 <- B2s%*%us%*%diag(1/sqrt(decomp$d[1::ncol(B1)-2]))
Z <- cbind(Z1,Z2)
```

- Random effect matrix with second design

```
B1 <- bbase(x,nseg=5)
sdia <- diag(s)
B2s <-sdia%*%B1
D <- diff(diag(ncol(B1)),diff=2)
P = t(D) %*% D
decomp <- svd(P)
Z1 <- B1%*%decomp$u%*%diag((1/sqrt(decomp$d)))
Z2 <- B2s%*%decomp$u%*%diag((1/sqrt(decomp$d)))
Z <- cbind(Z1,Z2)
```

- OpenBUGS model for the Cleveland data is:

```
 model{

for(j in 1:n){
```

```
response[j]~dnorm(mu[j],taub)
mu[j]<-inprod(beta0[],X[j,])+inprod(beta1[],Z[j,])
}

taub~dgamma(0.0001,0.0001)
for(i in 1:num.knots){
beta1[i]~dnorm(0,0.001)}

for( k in 1:P){
beta0[k]~dnorm(0,0.001)
}
sigmab<-1/sqrt(taub)
}
```

- R Library BRugs, runs OpenBUGS from R

```
model.fit<-bugs(dati.bugs,inits.bugs, c("beta0","taub",
"beta1","sigmab"),"modelethanol.txt",
n.chains=1,n.iter=n.iter,n.thin=n.thin,
n.burnin=n.burnin, codaPkg=FALSE)
```

Appendix B: QR Decomposition

If $\widetilde{\mathbf{X_j}}^{\mathbf{T}}$ has dimension $n \times q_j$ and $(q_j \leq n)$ then:

$$\widetilde{\mathbf{X_j}}^{\mathbf{T}} = \underbrace{[\mathbf{A} : \mathbf{Z}]}_{\text{Orthogonal matrix}} \underbrace{\begin{bmatrix} \mathbf{P} \\ \mathbf{0} \end{bmatrix}}_{\text{Upper triangular matrix}} \tag{11}$$

Where orthogonal part is represented by a block matrix by $\mathbf{D}_{n\times(n-q_j)}$ and $\mathbf{Z}_{n\times q_j-(n-q_j)}$. Therefore β are defined as $\beta = \mathbf{Z}\beta_z$ and it is proved that

$$\widetilde{\mathbf{X_j}}\beta = [\mathbf{P}^T\mathbf{0}] \begin{bmatrix} \mathbf{A^T} \\ \mathbf{Z^T} \end{bmatrix} \mathbf{Z}\beta_z = [\mathbf{P}^T\mathbf{0}] \begin{bmatrix} \mathbf{0} \\ \mathbf{I_{p-m}} \end{bmatrix} \beta_z = 0$$

For more details refer to (Wood, 2006).

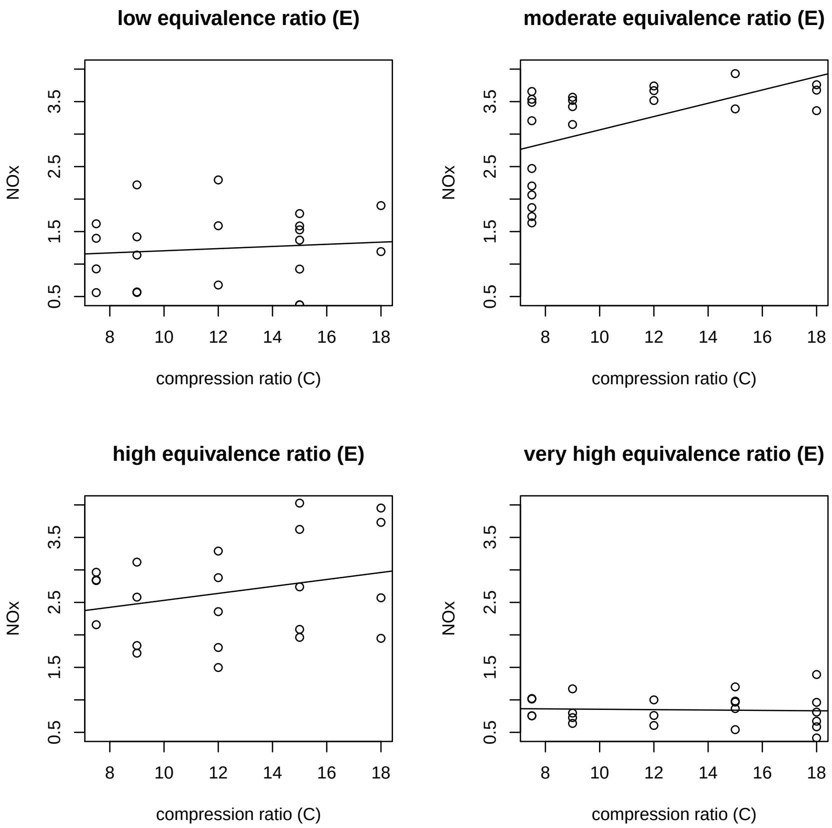
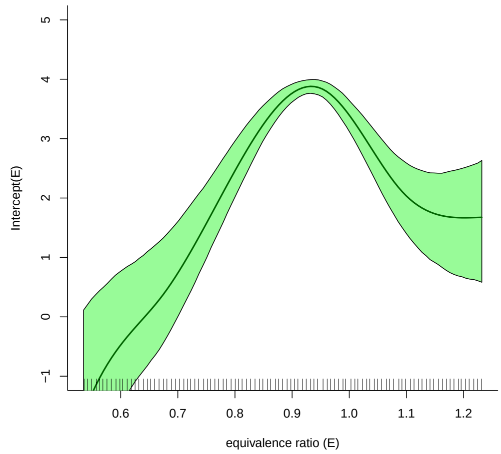
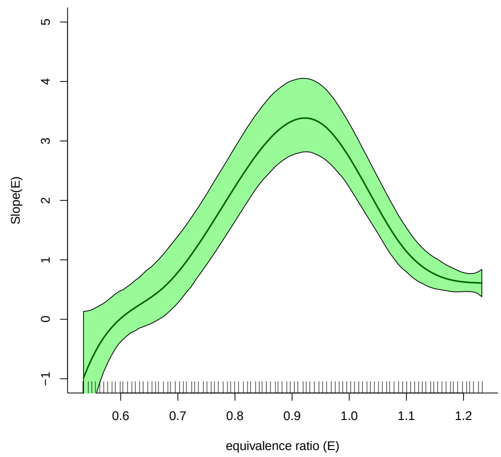


Figure 2: Plots of NOx versus compression ratio (C), for low, moderate, high and very high values of equivalence ratio (E).

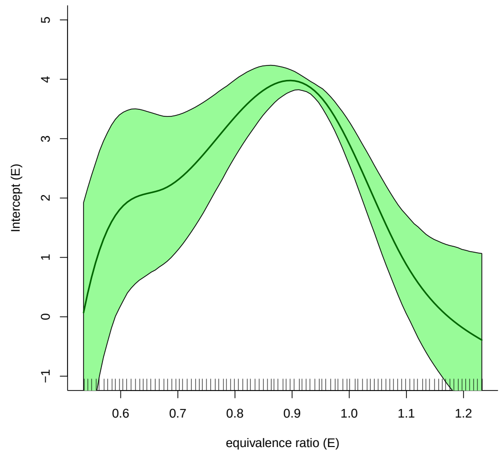


(a)

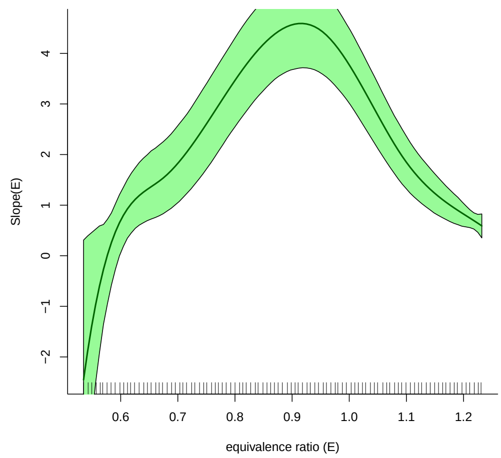


(b)

Figure 3: Plots obtained from estimates, with a random matrix by first design in section 5.1. In figures (a) and (b) plot of the posterior mean of NOx with all the other covariates set on their average values. Shaded area is the corresponding 95% pointwise credible sets.



(a)



(b)

Figure 4: Plots obtained from estimates, with a random matrix by second design in section 5.2. In figures (a) and (b) plot of the posterior mean of NOx with all the other covariates set on their average values. Shaded area is the corresponding 95% pointwise credible sets.