RandomCoefficients: Adaptive estimation in the linear random coefficients model

Christophe Gaillac
CREST and TSE

Eric Gautier
TSE

Abstract

This vignette presents the R package **RandomCoefficients** associated to Gaillac and Gautier (2019). This package implements the adaptive estimation of the joint density linear model where the coefficients - intercept and slopes - are random and independent from regressors which support is a proper subset. The estimator proposed in Gaillac and Gautier (2019) is based on Prolate Spheroidal Wave functions which are computed efficiently in **RandomCoefficients** based on Osipov, Rokhlin, and Xiao (2013). This package also provides a parallel implementation of the estimator.

Keywords: Adaptation, Ill-posed Inverse Problem, Random Coefficients, R.

1. How to get started

The R package **RandomCoefficients** can be downloaded from `https://cran.r-project.org`. To install the **RandomCoefficients** package from R use

```
install.packages("RandomCoefficients")
```

The installation of the package should proceed automatically. Once the **RandomCoefficients** package is installed, it can be loaded to the current R session by the command

```
library(RandomCoefficients)
```

Online help is available in two ways: either help(package="RandomCoefficients") or ?rc_estim. The first returns all available commands in the package. The second gives detailed information about a specific command. A valuable feature of R help files is that the examples used to illustrate commands are executable, so they can be pasted into a session or run as a group with a command like example(rc_estim).

The R package **RandomCoefficients** can be also downloaded from Github `https://github.com/cgaillac/RationalExp`. To install the **RandomCoefficients** package from Github, the devtools library is required. Then, use the command

```
library("devtools")
install_github('RandomCoefficients','cgaillac')
```

2. Theory

2.1. Random coefficients density estimation in a linear random coefficients model

For $\boldsymbol{\beta} \in \mathbb{C}^d$, $(f_m)_{m\in\mathbb{N}_0}$ functions with values in $\mathbb{C}$, and $\boldsymbol{m} \in \mathbb{N}_0^d$, denote by $\boldsymbol{\beta}^{\boldsymbol{m}} = \prod_{k=1}^d \boldsymbol{\beta}_k^{\boldsymbol{m}_k}$, $|\boldsymbol{\beta}|^{\boldsymbol{m}} = \prod_{k=1}^d |\boldsymbol{\beta}_k|^{\boldsymbol{m}_k}$, and $f_{\boldsymbol{m}} = \prod_{k=1}^d f_{\boldsymbol{m}_k}$. $|\cdot|_\infty$ stands for the ℓ_∞ norm of a vector. The Fourier transform of $f \in L^1\left(\mathbb{R}^d\right)$ is $\mathcal{F}[f](\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{b}^\top \boldsymbol{x}} f(\boldsymbol{b})d\boldsymbol{b}$ and $\mathcal{F}[f]$ is also the Fourier transform in $L^2\left(\mathbb{R}^d\right)$. Denote by $\mathcal{F}_{1\text{st}}[f]$ the partial Fourier transform of f with respect to the first variable. For a random vector $\boldsymbol{X}$, $\mathbb{P}_{\boldsymbol{X}}$ is its law, $f_{\boldsymbol{X}}$ its density and $f_{\boldsymbol{X}|\mathcal{X}}$ the truncated density of $\boldsymbol{X}$ given $\boldsymbol{X} \in \mathcal{X}$ when they exist, and $\mathbb{S}_{\boldsymbol{X}}$ its support. The inverse of a mapping f, when it exists, is denoted by f^I. Finally denote by $W_{[-R,R]} = \mathbb{1}\{[-R,R]\} + \infty\mathbb{1}\{[-R,R]^c\}$, where $R > 0$.

We first explain the estimator of the joint density $f_{\alpha,\beta}$ in the linear random coefficients model

$$Y = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X}, \tag{1}$$

$$(\alpha, \boldsymbol{\beta}^\top) \text{ and } \boldsymbol{X} \text{ are independent.} \tag{2}$$

The researcher has at her disposal n i.i.d observations $(Y_i, \boldsymbol{X}_i^\top)_{i=1}^n$ of $(Y, \boldsymbol{X}^\top)$ but does not observe the realizations $(\alpha_i, \boldsymbol{\beta}_i^\top)_{i=1}^n$ of $(\alpha, \boldsymbol{\beta}^\top)$. In this version of the package the number of regressors is limited to $p = 1$. Moreover, we assume here that

Assumption 1 *(H1.1) $f_{\boldsymbol{X}}$ and $f_{\alpha,\beta}$ exist;*

(H1.2) The support of $\boldsymbol{\beta}$ is a subset of $[-R,R]^p$, where $R > 0$ is known by the researcher;

(H1.3) For some $x_0 \in (0,\infty)$ and $\mathcal{X} = [-x_0, x_0]^p$ a subset of the support of $\boldsymbol{X}$, we have at our disposal an i.i.d sample $(Y_i, \boldsymbol{X}_i)_{i=1}^n$ and an estimator $\widehat{f}_{\boldsymbol{X}|\mathcal{X}}$ of the truncated density $f_{\boldsymbol{X}|\mathcal{X}}$ based on a sample of size n_0 independent of $(Y_i, \boldsymbol{X}_i)_{i=1}^n$;

(H1.4) The set $\mathcal{X}$ is such $\left\|f_{\boldsymbol{X}|\mathcal{X}}\right\|_{L^\infty(\mathcal{X})} \le C_{\boldsymbol{X}}$ and $\left\|1/f_{\boldsymbol{X}|\mathcal{X}}\right\|_{L^\infty(\mathcal{X})} \le c_{\boldsymbol{X}}$, where $c_{\boldsymbol{X}}, C_{\boldsymbol{X}} \in (0,\infty)$.

Assumption (H1.3) is not restrictive because, for all $\underline{\boldsymbol{x}}$ in the interior of $\mathbb{S}_{\boldsymbol{X}}$, we can rewrite (1) as $Y = \alpha + \boldsymbol{\beta}^\top \underline{\boldsymbol{x}} + \boldsymbol{\beta}^\top(\boldsymbol{X} - \underline{\boldsymbol{x}})$, take $\underline{\boldsymbol{x}} \in \mathbb{R}^p$ and x_0 such that $\mathcal{X} \subseteq \mathbb{S}_{\boldsymbol{X}-\underline{\boldsymbol{x}}}$, and there is a one-to-one mapping between $f_{\alpha+\boldsymbol{\beta}^\top\underline{\boldsymbol{x}},\boldsymbol{\beta}}$ and $f_{\alpha,\boldsymbol{\beta}}$. This mapping is not yet directly implemented in this version of the package even if there is the option to estimate $f_{\alpha+\boldsymbol{\beta}^\top\underline{\boldsymbol{x}},\boldsymbol{\beta}}$ (see the parameter `center` in Section 3). The constant $C_{\boldsymbol{X}}$ in (H1.4) is not needed in the estimation, whereas the constant $c_{\boldsymbol{X}}$ will be estimated using the estimator $\widehat{f}_{\boldsymbol{X}|\mathcal{X}}$ and the cross-entropy method (using the **RCEIM** package). There is a trade-off in the choice of $\mathcal{X}$ between the sample size used and the impact of a small $c_{\boldsymbol{X}}$ on the convergence rates (see the parameter `trunc` in Section 3 for a practical solution). Gaillac and Gautier (2019) relax (H1.2) and treat the more general case $f_{\alpha,\boldsymbol{\beta}} \in L^2\left(w \otimes W^{\otimes p}\right)$, where $W = \cosh(\cdot/R)$ and $w \ge 1$.

For $c \in \mathbb{R}$, let us introduce the operator $h \in L^2(W^{\otimes p}) \to \mathcal{F}[h](c\ \cdot) \in L^2([-1,1]^p)$, where $W = W_{[-R,R]}$, which

1. has SVD $\left(\sigma_m^{W,c}, \varphi_m^{W,c}, g_m^{W,c}\right)_{m\in\mathbb{N}_0}$ when $p=1$;

2. else its SVD is the product $\left(\sigma_{\boldsymbol{m}}^{W,c}, \varphi_{\boldsymbol{m}}^{W,c}, g_{\boldsymbol{m}}^{W,c}\right)_{\boldsymbol{m}\in\mathbb{N}_0^p}$.

The estimator in Gaillac and Gautier (2019) aims at minimizing the Mean Integrated Squared Error (MISE) conditional on the sample used to estimate $f_{\boldsymbol{X}|\mathcal{X}}$. The estimation of $f_{\alpha,\boldsymbol{\beta}}$ is an inverse problem, as detailed below, and the estimation strategy is 1) to estimate $\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\left(\cdot,\star\right)\right](t)$ for all $t\neq 0$, then 2) to estimate $f_{\alpha,\boldsymbol{\beta}}$ taking the Fourier inverse of the estimator in 1). Assume that the researcher knows a superset $[-R,R]^p$ containing the support of $\boldsymbol{\beta}$. Then, consider the following three steps estimator, for $0<\epsilon<1<T$ and $N:\mathbb{R}\to\mathbb{N}_0$:

(S.1) for all $t\neq 0$, obtain a preliminary approximation of $F_1(t,\cdot)=\mathcal{F}_{1\mathrm{st}}\left[f_{\alpha,\boldsymbol{\beta}}\right](t,\cdot)$

$$\widehat{F}_1^{N,T}(t,\cdot)=\mathbb{1}\{|t|\le T\}\sum_{\boldsymbol{m}\in\mathbb{N}_0^p,|\boldsymbol{m}|_\infty\le N(t)}\frac{\widehat{c}_{\boldsymbol{m}}(t)}{\sigma_{\boldsymbol{m}}^{W,tx_0}}\varphi_{\boldsymbol{m}}^{W,tx_0}(\cdot),$$

where

$$\widehat{c}_{\boldsymbol{m}}(t)=\frac{1}{n}\sum_{j=1}^n\frac{e^{itY_j}}{x_0^p\widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j)}\overline{g_{\boldsymbol{m}}^{W,tx_0}}\left(\frac{\boldsymbol{X}_j}{x_0}\right)\mathbb{1}\left\{\boldsymbol{X}_j\in\mathcal{X}\right\}, \tag{3}$$

is an estimator of $c_{\boldsymbol{m}}(t)=\langle\mathcal{F}\left[f_{Y|\boldsymbol{X}=x_0\cdot}\right](t),g_{\boldsymbol{m}}^{W,tx_0}(\cdot)\rangle_{L^2([-1,1]^p)}$, $\widehat{f}^{\delta}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j)=\widehat{f}_{\boldsymbol{X}|\mathcal{X}}(\boldsymbol{X}_j)\vee\sqrt{\delta(n_0)}$ and $\delta(n_0)$ is a trimming factor converging to zero with n_0;

(S.2) refine for $|t|\le\epsilon$

$$\widehat{F}_1^{N,T,\epsilon}(t,\cdot)=\widehat{F}_1^{N,T}(t,\cdot)\mathbb{1}\{|t|\ge\epsilon\}+\mathcal{I}_{\underline{a},\epsilon}\left[\widehat{F}_1^{N,T}(\star,\cdot)\right](t)\mathbb{1}\{|t|<\epsilon\},$$

where, for $\underline{a},\epsilon>0, f\in L^2(\mathbb{R})$, $\rho_m^{W,c}=\left(\sigma_m^{W,c}\right)^2|c|/(2\pi)$, $\mathcal{I}_{\underline{a},\epsilon}$ defined as

$$\mathcal{I}_{\underline{a},\epsilon}\left[f\right](\cdot)=\sum_{m\in\mathbb{N}_0}\frac{\rho_m^{W_{[-1,1]},\underline{a}\epsilon}}{\left(1-\rho_m^{W_{[-1,1]},\underline{a}\epsilon}\right)\epsilon}\left\langle f(\star),g_m^{W_{[-1,1]},\underline{a}\epsilon}\left(\frac{\star}{\epsilon}\right)\right\rangle_{L^2(\mathbb{R}\setminus[-\epsilon,\epsilon])}g_m^{W_{[-1,1]},\underline{a}\epsilon}\left(\frac{\cdot}{\epsilon}\right).$$

performs interpolation (see Gaillac and Gautier (2019));

(S.3) take $\widehat{f}_{\alpha,\boldsymbol{\beta}}^{N,T,\epsilon}(\cdot_1,\cdot_r)=\max\left(\mathcal{F}_{1\mathrm{st}}^I\left[\widehat{F}_1^{N,T,\epsilon}(\star,\cdot_r)\right](\cdot_1),0\right)$.

Let $\epsilon>0$, $K_{\max}=\lfloor\log(n)/(6p\log(2))\rfloor$, and $T_{\max}=2^{K_{\max}}$. Then, choose the parameters (N,T) in a data-driven way following an adaptation of the Goldenshluger-Lepski method (see Goldenshluger and Lepski (2014) and the references therein). First, obtain $\widehat{N}$ by solving univariate minimisation problems

$$\forall t\in\mathbb{R}\setminus(-\epsilon,\epsilon),\quad \widehat{N}(t)\in\underset{0\le N\le N_{\max}(W,t)}{\mathrm{argmin}}\left(B_1(t,N)+c_1\Sigma(t,N)\right), \tag{4}$$

where $c_1 \geq 31/30$ is greater than 1 to handle the estimation of $f_{\boldsymbol{X}|\mathcal{X}}$ and

$$B_1\left(t,N\right) = \max_{N_{\max,q}(W,t)\geq N'\geq N}\left(\sum_{N\leq|\boldsymbol{m}|_\infty\leq N'}\frac{\left|\widehat{c}_{\boldsymbol{m}}(t)\right|^2}{\left(\sigma_{\boldsymbol{m}}^{W,tx_0}\right)^2}-\Sigma\left(t,N'\right)\right)_+,$$

$$\Sigma(t,N) = \frac{84(1+2((2\log(n))\vee 3))c_{\boldsymbol{X}}}{n}\left(\frac{|t|}{2\pi}\right)^p\nu(W,N,tx_0).$$

Second, define $\widehat{T}$ as

$$\widehat{T}\in\underset{T\in\mathcal{T}_n}{\operatorname{argmin}}\left(B_2\left(T,\widehat{N}\right)+\int_{\epsilon\leq|t|\leq T}\Sigma\left(t,\widehat{N}(t)\right)dt\right), \tag{5}$$

where

$$B_2\left(T,\widehat{N}\right) = \max_{T'\in\mathcal{T}_n,T'\geq T}\left(\int_{T\leq|t|\leq T'}\sum_{|\boldsymbol{m}|_\infty\leq\widehat{N}(t)}\frac{\left|\widehat{c}_{\boldsymbol{m}}(t)\right|^2}{\left(\sigma_{\boldsymbol{m}}^{W,tx_0}\right)^2}-\Sigma\left(t,\widehat{N}(t)\right)dt\right)_+,$$

$$\mathcal{T}_n = \left\{2^k:\ k=1,\dots,K_{\max}\right\}.$$

The functions ν and $N_{\max}$ are defined, for $t\neq 0$ by

$$\nu(W,N,t) = \left(\frac{2(1\vee N)}{(R\,|t|)\vee 1}\right)^p\exp\left(2Np\ln\left(\left(\frac{7\pi(N+1)}{R\,|t|}\right)\vee 1\right)\right),$$

$$N_{\max}(W,t) = \left\lfloor\frac{\ln(n)}{2p}\left(\mathcal{W}\left(\frac{7\pi}{R\,|t|}\frac{\ln(n)}{2p}\right)\vee 1\right)^{-1}\right\rfloor. \tag{6}$$

where $\mathcal{W}$ is the inverse of $x\in[0,\infty)\mapsto xe^x$. Finally, this package uses a Gaussian kernel density estimator to estimate $f_{\boldsymbol{X}|\mathcal{X}}$ through kde in the package **ks**.

2.2. Computation of the SVD

The estimator requires the SVD of $\mathcal{F}_c$ for $c\neq 0$. When $W=W_{[-1,1]}$, by Proposition A.1 in Gaillac and Gautier (2019), we have $g_m^{W(\cdot/R),c} = g_m^{W,Rc}$ for all $m\in\mathbb{N}$. When $W=W_{[-1,1]}$, the first coefficients of the decomposition of the eigenfunctions on the Legendre polynomials can be obtained by solving for the eigenvectors of two tridiagonal symmetric Toeplitz matrices (for even and odd values of m, see Section 2.6 in Osipov *et al.* (2013)). We use that $\mathcal{F}_c^*\left(g_m^{W,Rc}\right)=\sigma_m^{W,Rc}\varphi_m^{W,Rc}$ and that $\varphi_m^{W,Rc}$ has norm 1 to obtain the remaining of the SVD.

3. The main function in the RandomCoefficients package

The function rc_estim() implements the adaptive estimation of the joint density of random coefficients model. The function takes as inputs data (Y,X) then estimates the density and return its evaluation on a grid b_grid times a_grid. By setting nbCores greater than 1 computations are done in parallel.

```
rc_estim<-function(X,Y,b_grid,a_grid,nbCores,M_T,N_u,epsilon,n_0,trunc,center)
```

X	Vector of size N, N being the number of observation and the number of regressors limited to 1 in this version of the package.
Y	Outcome vector of size N.
b_grid	Vector grid on which the estimator of the density of the random slope will we evaluated. No default.
a_grid	Vector grid on which the estimator of the density of the random intercept will we evaluated. Default is beta_grid.
nbCores	Number of cores for the parallel implementation. Default is 1, no parallel computation.
M_T	Number of discretisation points for the estimated partial Fourier transform. Default is 60.
N_u	Maximal number of singular functions used in the estimation. Default is the maximum of 10 and (6).
epsilon	Parameter for the interpolation. Default is $(\log(N)/\log(\log(N)))^{-\sigma_0}$ as in (T5.1) in Gaillac and Gautier (2019) with $\sigma_0 = 4$.
n_0	Parameter for the sample splitting. If n_0$= 0$ then no sample splitting is done and we use the same sample of size N to perform the estimation of $f_{\boldsymbol{X}\|\mathcal{X}}$. If n_0$> 0$, then this is the size of the sample used to perform the estimation of $f_{\boldsymbol{X}\|\mathcal{X}}$. Default is $n_0 = 0$.
trunc	Dummy for the truncation of the density of the regressors to an hypercube $\mathcal{X}$. If trunc$= 1$, then truncation is performed and $\mathcal{X}$ is defined using the argmin of the ratio of the estimated constant $c_{\boldsymbol{X}}$ over the percentage of observation in $\mathcal{X}$. Default is 0, no truncation.
center	Dummy to trigger the use of $X - \underline{x}$ instead of X as regressor. If center$= 1$, then use $X - \underline{x}$ where $\underline{x}$ is the vector of the medians coordinates by coordinates for X. Default is center$= 0$, where regressors are left unchanged.

We refer to the reference manual or help file for additional details.

4. Example

We give the following example of a linear random coefficients model when regressors have limited variation. We take $p = 1$ and $(\alpha, \beta)^\top = \xi_1 \mathbb{1}\{\theta \geq 0\} + \xi_2 \mathbb{1}\{\theta < 0\}$ with $\mathbb{P}(\theta \geq 0) = 0.5$. The law of X is a truncated normal based on a normal of mean 0 and variance 2.5 and truncated to $\mathcal{X}$ with $x_0 = 1.5$. The laws of ξ_1 and ξ_2 are truncated normals based on normals with means $\mu_1 = (-2 \ \ -3)^\top$ and $\mu_2 = (3 \ 0)^\top$, same covariance $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, and truncated to $[-6, 6]^{p+1}$. We refer to Gaillac and Gautier (2019) for an analysis of the performances of the estimator using Monte Carlo simulations.

Start by loading useful packages, defining the output grids, and the apriori on the support of the regressors.

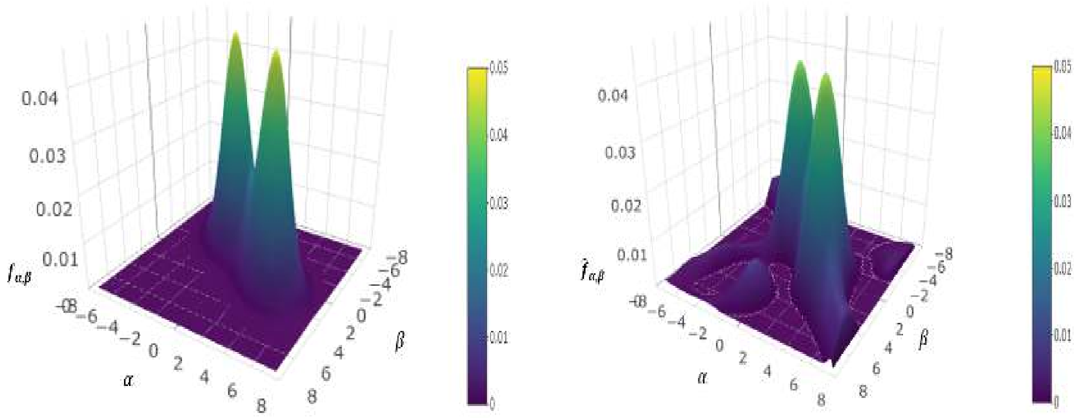
```
library("orthopolynom")
library("polynom")
```

```
library(tmvtnorm)
library(ks)
library(snowfall)
library(sfsmisc)
library(fourierin)
library(rdetools)
library(statmod)
library(RCEIM)
library(robustbase)
library(VGAM)
library(RandomCoefficients)

# beta (output) Grid
M=100
# Apriori on the support of the random slope.
limit =7.5
b_grid <- seq(-limit ,limit ,length.out=M)
a = limit
# Support apriori limits (taken symmetric)
up =1.5
down = -up
und_beta <- a
x.grid <- as.matrix(expand.grid(b_grid,b_grid))
```

Then, simulate the data:

```
# sample size
N <- 1000
#number of regressors
d = 1
Mean_mu1 = c(-2,- 3)
Mean_mu2= c(3,  0)
Sigma= diag(2, 2)
Sigma[1,2] = 1
Sigma[2,1] = 1
vect <- as.matrix(expand.grid(b_grid,b_grid ))
x.grid <- vect
beta_model <-  1/2*matrix(dmvnorm( vect ,Mean_mu1  , Sigma), M ,M)
             + 1/2*matrix(dmvnorm( vect ,Mean_mu2  , Sigma), M ,M)
Sigma= diag(2, 2)
Sigma[1,2] = 1
Sigma[2,1] = 1
# Generate truncated normals, for the regressors and the random coefficients
lim2 = 6
xi1<-rtmvnorm(N,Mean_mu1,Sigma,lower=        c(-lim2,-lim2),upper=c(lim2,lim2))
xi2<-rtmvnorm(N,Mean_mu2,Sigma,lower=c(-lim2,-lim2),upper=c(lim2,lim2))
theta = runif(N, -1 , 1)
```



(a) True density

(b) Estimator

```
beta <- 1*(theta >=0) * xi1  + 1*(theta <0) * xi2

X <- rtmvnorm(N, mean = c(0), sigma=2.5, lower=c( down), upper=c(up))
X_t <- cbind(matrix(1, N,1),X)
Y <-rowSums(beta*X_t)
```

Finally, perform estimation using rc_estim and parallel computation. Then, the estimator is plotted using the code below. Figure 4 is obtained using the plotly package.

```
out <- rc_estim( X,Y,b_grid,b_grid,nbCores = 4, M_T = 60)
# The output matrix
mat <- out[[1]]
# The evaluation grid, random slope then intercept.
b_grid <- out[[2]]
alpha_grid <- out[[3]]
# To plot the output estimator
x11()
filled.contour(alpha_grid ,b_grid, mat)
```

References

Gaillac C, Gautier E (2019). "The linear random coefficients model when regressors have limited variation." Preprint arXiv:1905.06584.

Goldenshluger A, Lepski O (2014). "On adaptive minimax density estimation on $\mathbb{R}^d$." *Probab. Theory Related Fields*, **159**, 479–543.

Osipov A, Rokhlin V, Xiao H (2013). *Prolate spheroidal wave functions of order zero.* Springer.

Affiliation:

Christophe Gaillac
CREST
5, avenue Henry Le Chatelier
91 120 Palaiseau, France
and Toulouse School of Economics
E-mail: christophe.gaillac@ensae.fr

Eric Gautier
Toulouse School of Economics, Toulouse Capitole University
21 allée de Brienne
31000 Toulouse, France
E-mail: eric.gautier@tse-fr.eu