

*A calibration function built from change points using linear mixed models: A solution through Evolutionary Algorithms*

Ehidy Karime García Cruz - Ph.D. (C) in Statistics  
Juan Carlos Correa Morales Ph.D in Statistics - Director  
Juan Carlos Salazar Uribe Ph.D in Statistics - Codirector

Institutional Seminar School of Statistics  
National University of Colombia- Campus Medellín



# OUTLINE

## ① INTRODUCTION

## ② PROBLEM STATEMENT

## ③ OBJECTIVES

## ④ BACKGROUND

## ⑤ PARAMETRIC APPROACH

## ⑥ CONCLUSIONS & COMMITMENTS

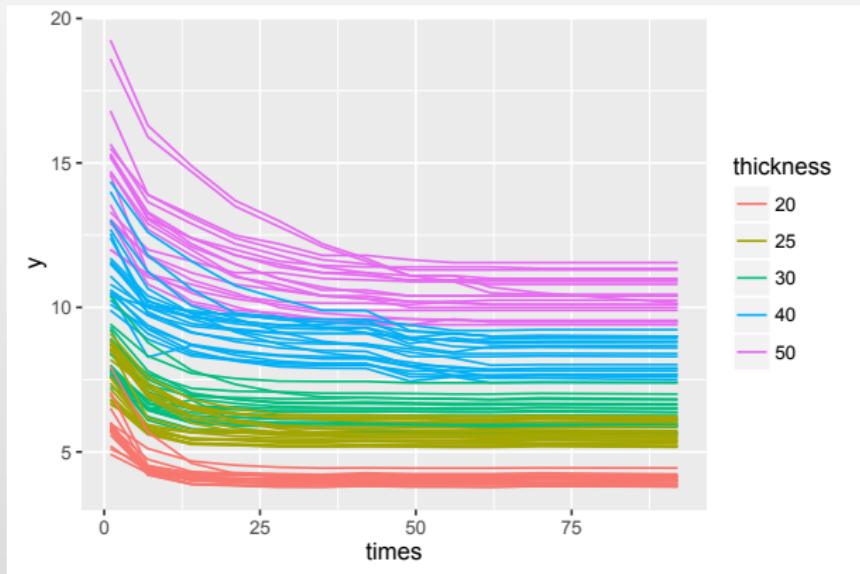
## ⑦ BIBLIOGRAPHY



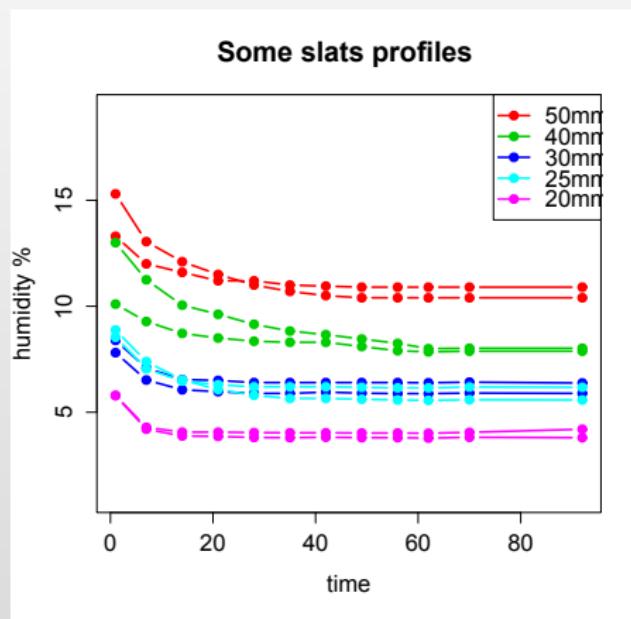
# THE ORIGIN OF THE PROBLEM



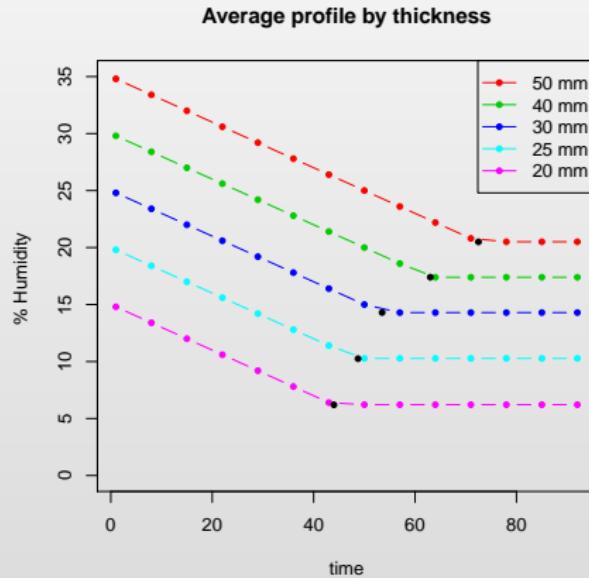
# LONGITUDINAL DATA SET PLOT



# SOME SUBJECTS



# PROBLEM



# OBJECTIVES

## GENERAL OBJECTIVE

To build a complete calibration function from subject-specific change points that were estimated using a linear mixed model approach.



# LINEAR REGRESSION MODEL

## SIMPLE LINEAR REGRESSION MODEL

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad (1)$$

with  $\varepsilon_i \sim N(0, \sigma^2)$ .

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (2)$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad (3)$$

It is well known that the  $\hat{\alpha}$  and  $\hat{\beta}$  are Gauss-Markov minimum variance unbiased estimators (Rao(1973), Berkson(1969)).



# CALIBRATION PROBLEM

## INVERSE SIMPLE REGRESSION MODEL

$$X_i = \frac{Y_i - \alpha}{\beta} + \eta_i, \quad (4)$$

where  $\eta_i \sim N\left(0, \frac{\sigma^2}{\beta^2}\right)$

For example, if time is the explanatory variable of interest then  $\hat{t}$  must be estimated, and we want to predict it as precise as possible BUT...

Under a longitudinal setting.



# CALIBRATION PROBLEM

The suggested estimators are:

Classic	Inverse <sup>1</sup>
$\hat{X}_c = \frac{Y - \hat{\alpha}}{\hat{\beta}}$	$\hat{X}_I = \hat{\gamma} + \hat{\delta} Y$
IMSE <sup>2</sup>	Practically Unbiased
$\hat{X}_M = \hat{\lambda}_0 + \hat{\lambda} Y$ $\hat{\lambda}_0$ and $\hat{\lambda}$ minimize the IMSE ( $X_M$ )	$\hat{X}_u = \bar{X} + \frac{Y - \bar{Y}}{\left[ \hat{\beta} + \frac{\hat{\sigma}^2}{\hat{\beta}^2 + \sum(x_i - \bar{x})^2} \right]}$

<sup>1</sup>Krutchkoff, R.(1969)

<sup>2</sup>Brown, G.(1979), Naszódi, L. J.(1978)



## CP: USING LINEAR MODELS

For a single change point and using SLRM<sup>3</sup>, the model is expressed as:

$$y_i = \begin{cases} \beta_{10} + \beta_{11}x_i + \varepsilon_{1i}, & i = 1, \dots, s \\ \beta_{20} + \beta_{21}x_i + \varepsilon_{2i}, & i = s + 1, \dots, n \end{cases}, \quad \varepsilon_{1i} \sim N(0, \sigma_1^2), \quad \varepsilon_{2i} \sim N(0, \sigma_2^2) \quad . \quad (5)$$



<sup>3</sup>Hofrichter(2007), Hinich & Farley (1970)

BUT...

- Estimates can only be fitted with the standard IWLS method if the change point is known, Hofrichter (2007).
- If the change point is unknown there is not an analytical solution available to estimate  $x_s = \tau$ .
- To avoid a suboptimal value for  $\tau$  an iterative procedure must be executed.

A more extensive literature review on this problem can be seen at  
Garcia,E. Correa,J. & Salazar,J. (2017)



# LINEAR MIXED MODELS

The model given in (1), can be extended under some conditions as follow:

$$\begin{aligned} \mathbf{Y}_i = & \underbrace{\mathbf{X}_i \boldsymbol{\beta}}_{\text{Fixed}} + \underbrace{\mathbf{Z}_i \mathbf{b}_i}_{\text{Random}} + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i & \sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\varepsilon}_i & \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \\ \boldsymbol{\varepsilon}_i \text{ and } \mathbf{b}_i & \text{ are independent.} \end{aligned} \tag{6}$$



## CP: USING LINEAR MIXED MODELS

Lai & Albert (2014) suggested the change point for fixed effects models can be found by partitioning the information about the subjects in two blocks, so that  $B_1 = \{t_{ij} : t_{ij} \leq \tau\}$  and  $B_2 = \{t_{ij} : t_{ij} > \tau\}$  as

$$\mathbf{Y}_{ij} = \sum_{k=1}^2 \mathbf{X}_{ij}\boldsymbol{\beta}_{B_k} I(t_{ij} \in B_k) + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{e}_{ij} \quad (7)$$

and in case of a fixed effects and random effects model we have,

$$\mathbf{Y}_{ij} = \sum_{k=1}^2 \mathbf{X}_{ij}\boldsymbol{\beta}_{B_k} I(t_{ij} \in B_k) + b_{i,0} + \sum_{k=1}^2 b_{i,B_k} I(t_{ij} \in B_k) + \mathbf{e}_{ij} \quad (8)$$

They proposed a test the fixed effect given the blocks built.



# THE MODEL TO CONSIDER

$$\begin{aligned}y_{ij} &= (\beta_1 x_{ij} + \beta_2 t_{ij}) 1(t_{ij} \leq \tau_i) + (\beta_1 x_{ij} + \beta_2 \tau_i) 1(t_{ij} > \tau_i) \\&\quad + b_{00i} + b_{01i} 1(t_{ij} \leq \tau_i) + b_{02i} 1(t_{ij} > \tau_i) + e_{ij} \\&= \beta_1 x_{ij} + \beta_2 \{t_{ij} 1(t_{ij} \leq \tau_i) + \tau_i 1(t_{ij} > \tau_i)\} \\&\quad + b_{00i} + b_{01i} 1(t_{ij} \leq \tau_i) + b_{02i} 1(t_{ij} > \tau_i) + e_{ij}\end{aligned}$$



Without loss of generality, but magnifying the importance of this we will take:

$$\mathbf{X}_i(\tau_i) = \mathbf{X}_i$$

And our objective function is to maximize the ML, which should be written as:

$$\ell(\boldsymbol{\theta})_{ML} = -\frac{1}{2} \sum_{i=1}^N \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \left\{ \sum_{i=1}^N \frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\} \quad (9)$$

Or REML

$$\begin{aligned} \ell(\boldsymbol{\theta})_{REML} = & -\frac{1}{2} \sum_{i=1}^N \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \left\{ \sum_{i=1}^N \frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\} \\ & - \frac{1}{2} \log \left| \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right| \end{aligned} \quad (10)$$



# SIMULATION RESULTS

Change points using DEA for a LMM

$n = 25, n_i = 14$

Fixed effects values $\mathbf{x}_i$					
id	50mm	40mm	30mm	25mm	20mm
1	65.171	56.210	49.342	47.046	40.448
2	65.359	57.573	48.738	44.213	42.088
3	67.417	60.824	47.427	47.488	37.969
4	65.148	58.623	48.109	48.139	41.413
5	65.223	60.328	47.076	48.597	42.235
$\tau$	65	57	49	45	41
$\bar{\tau}_i$	65.6636	58.7116	48.1384	47.0966	40.83
$S_{\bar{\tau}_i}$	0.9836	1.9129	0.9279	1.7181	1.7484



## CHANGE POINT FOR THE REAL DATA SET

Change points using EA for a LMM  $n = 100, n_i = 14$

id	Fixed effects values Thickness				
	50mm	40mm	30mm	25mm	20mm
1	46.493	39.055	26.763	21.515	16.699
2	41.979	35.315	27.271	20.365	16.382
3	43.871	32.047	27.904	28.515	20.715
4	45.210	35.042	28.578	25.647	15.512
5	40.508	32.608	27.016	22.924	21.096
6	41.826	33.479	27.835	25.464	23.310
7	45.576	31.917	26.739	21.836	18.351
8	44.247	37.215	31.769	20.526	15.010
9	40.318	30.310	30.394	22.584	24.067
10	41.946	31.141	27.243	23.763	24.099
11	48.875	31.757	32.623	20.091	19.911
12	42.569	32.937	29.566	21.613	18.726
13	44.158	33.631	34.190	27.331	15.682
14	44.637	36.064	29.115	23.851	19.684
15	41.280	35.524	25.407	21.632	15.633
16	44.049	35.755	26.688	28.232	24.328
17	47.245	39.618	28.011	20.029	20.721
18	42.327	30.459	30.115	29.076	17.599
19	41.159	36.307	28.222	25.708	18.730
20	45.634	35.842	32.707	23.460	22.258
$\hat{\gamma}_i$	43.695	34.3012	28.903	23.7081	19.4257
$S_{\hat{\gamma}_i}$	2.3495	2.693	2.3793	2.9313	3.098



# A CLASSICAL BAYESIAN APPROACH

## THE PROBLEM TO SOLVE

$$\begin{aligned} y_{ij} \mid \beta, \mathbf{b}_i, \Sigma_i &\sim N \left( \mathbf{x}_{ij}^T(\tau_i) \beta + \mathbf{z}_{ij}^T(\tau_i) \mathbf{b}_i, \Sigma_i \right) && \text{Level 1} \\ \mathbf{b}_i \mid \mathbf{D} &\sim N_q(\mathbf{0}, \mathbf{D}) && \text{Level 2} \\ \tau_i &\sim \pi(\gamma) && \text{Level 3} \\ \Sigma_i &\sim \pi(\sigma^2, \rho), \quad \beta \sim \pi(\beta) \text{ and } \mathbf{D} \sim \pi(\mathbf{D}), && \text{Priors} \\ \gamma &\sim \pi(\gamma_0, \gamma_1), \quad \psi^2 \sim \pi(\psi^2) \end{aligned} \tag{11}$$



# CONCLUSIONS

- The adapted algorithm allows to estimate a global maximum more than a local maximum which is one the main advantages of Evolutionary Algorithms.
- The change points associated to the real data set are quite precise according to the subject-specific profile.



- The results from the real data set generate new questions about the utility of this change points and its behaviour on large samples.
- In the illustration with real data, we observed that this approach could permit to predict, in a plausible and precise way, the change point given a specific value for the thickness. From a practical point of view, this prediction process allows to reduces both storage time and storage expenses.



## REMARKS

- This proposal has not been developed by any author before, according to the literature review.
- Computational problems to optimize the objective function have arisen almost all time during this researching, we had to check the convergence to the original values in the simulation process and parallelizing most of the routines executed on R.
- We have used three different servers (OpenSuse, Apolo, RStudio Server) to execute faster the routines and the optimization process.



## REFERENCES

-  Garcia, E., Correa, J. & Salazar, J. (2017) , A calibration function built from change points: a review', *Comunicaciones en Estadística* **10**(1), 113–128.
-  Lai Y. & Albert P.(2014). 'Identifying multiple change-points in a linear mixed effects model'. *Statist. Med.*33:1015 - 1028.  
Accessed:20/05/2015.Available Online doi: 10.1002/sim.5996
-  Rosenfield,D., Zhou,E., Wilhem F., Conrad A., Roth, W. & Meuret, A. (2010). 'Change point analysis for longitudinal Physiological data: detection of cardio-respiratory changes preceding panic attacks '. *Biological psychology* **84** (1), 112 - 120.
-  Krutchkoff (1967). 'Classical and Inverse regression methods of calibration'. *Technometrics* **9** (3), 425 - 439.



## REFERENCES

- Rao, C. R. (1973). *Linear Statistical inference and its applications*. 2nd Ed. Jhon Wiley & Sons: U.S.
- Fitzmaurice, G. Davidian, M. Verbeke, G. & Molenberghs, G.(2008). *Handbook of Modern Statistical Methods: Longitudinal data analysis*. Chapman& Hall: U.S.
- McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models* (2nd ed.).London: Chapman and Hall.



# Thank you



# Questions

