



HAL
open science

Bootstrap test for variance components in nonlinear mixed effects models

Tom Guédon, Charlotte Baey, Estelle Kuhn

► **To cite this version:**

Tom Guédon, Charlotte Baey, Estelle Kuhn. Bootstrap test for variance components in nonlinear mixed effects models. european meeting of statisticians, Jul 2023, Varsovie, Poland. . hal-04350696

HAL Id: hal-04350696

<https://hal.inrae.fr/hal-04350696>

Submitted on 18 Dec 2023

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

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



Bootstrap test for variance components in nonlinear mixed effects models

Tom Guédon¹ Charlotte Baey² Estelle Kuhn¹

¹INRAE, France.

²Univ. Lille, France.

Introduction

Context:

- hypothesis testing under nonstandard conditions
- usual MLE asymptotic theory not applicable

Motivations:

- build a procedure
- non asymptotic
- applicable to any type of mixed-effect models (linear, nonlinear)

Mixed effects models

We consider the following nonlinear mixed effects model

$$\begin{cases} y_{ij} = g(x_{ij}, \beta, \Lambda \xi_i) + \varepsilon_{ij} & \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \xi_i \sim \mathcal{N}(0, I_p) \end{cases}, \quad (1)$$

- y_{ij} : the j th response of the i th individual ($i = 1, \dots, N; j = 1, \dots, J_i$)
- β : vector of fixed effects
- x_{ij} : known covariates
- Λ : lower triangular matrix with nonnegative diagonal coefficients
- $(\xi_i)_i$ and $(\varepsilon_{ij})_{ij}$ are mutually independent
- $y_i = (y_{ij})_{j=1 \dots J_i} \sim f_i(y_i; \theta)$, $y_i | \xi_i \sim f_i(y_i; \xi_i, \theta)$, $\xi_i \sim \pi(\xi_i)$

Variance components testing

Objective: Test the nullity of the last r variances of the scaled random effects $(\Lambda \xi_i)_i$.

$$H_0: \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \quad H_1: \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ \Lambda_{12} & \Lambda_2 \end{pmatrix}$$

Likelihood ratio test statistic (lrt):

$$\text{lrt}(y_{1:N}) = -2 \left(\sup_{\theta \in \Theta} l(\theta; y_{1:N}) - \sup_{\theta \in \Theta_0} l(\theta; y_{1:N}) \right).$$

- Θ : unrestricted parameter space
- Θ_0 : restricted parameter space (under H_0)
- $l(\theta; y_{1:N})$: log marginal likelihood

$$l(\theta; y_{1:N}) = \sum_{i=1}^N \log \{ f_i(y_i; \theta) \} = \sum_{i=1}^N \log \int f_i(y_i; \xi_i, \theta) \pi(\xi_i) d\xi_i$$

Proposed test procedure

Shrunked parametric bootstrap for variance components testing

- Input:** $c_N > 0$, $B \in \mathbb{N}^*$, $0 < \alpha < 1$
- Set:** $\beta_N^* = \hat{\beta}_N$, $\Lambda_N^* = \hat{\Lambda}_N$, and $\sigma_N^{*2} = \hat{\sigma}_N^2$
- Set:** $\Lambda_{2,N}^* = \Lambda_{12,N}^* = 0$
- Set:** $[\Lambda_{1,N}^*]_{mn} = [\hat{\Lambda}_{1,N}]_{mn} \mathbb{1}([\hat{\Lambda}_{1,N}]_{mn} > c_N)$
- for** $b = 1, \dots, B$ **do**
- for** $i = 1, \dots, N$ **do**
- draw independently $\varepsilon_i^{*,b} \sim \mathcal{N}(0, \sigma_N^{*2} I_{J_i})$ and $\xi_i^{*,b} \sim \mathcal{N}(0, I_p)$
- build the i th value of the b th bootstrap sample $y_i^{*,b} = g(x_i, \beta_N^*, \Lambda_N^* \xi_i^{*,b}) + \varepsilon_i^{*,b}$
- compute the likelihood ratio statistic $\text{lrt}(y_{1:N}^{*,b})$
- end for**
- end for**
- Compute the bootstrap p -value as $p_{boot} = \frac{1}{B} \sum_{b=1}^B \mathbb{1}(\text{lrt}(y_{1:N}^{*,b}) > \text{lrt}(y_{1:N}))$
- Reject H_0 if $p_{boot} < \alpha$

Theoretical issues

Boundary issue: $\Lambda_2 = 0 \notin \overset{\circ}{\Theta}$

Singularity issue: vanishing score

Example of a linear model with one random effect:

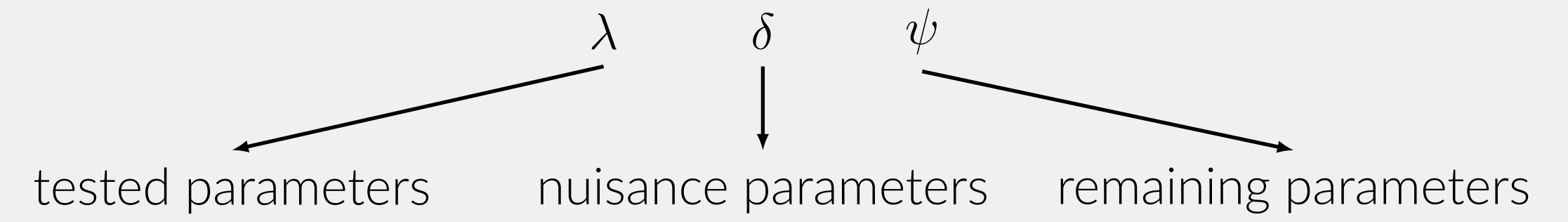
$$\begin{cases} y_{ij} = \lambda \xi_i + \varepsilon_{ij} & \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \xi_i \sim \mathcal{N}(0, 1) \end{cases},$$

$$\frac{\partial l(\theta; y_{1:N})}{\partial \lambda} \Big|_{\lambda=0} = \sum_{i=1}^N f_i(y_i; \theta)^{-1} \sum_{j=1}^{J_i} \frac{y_{ij}}{\sigma^2} \int \xi_i \pi(\xi_i) d\xi_i = 0$$

- singularity and boundary issues:** sources of inconsistency of the bootstrap procedure
- nuisance parameters:** issues cited above occurring at unknown locations

Theoretical results

Notations: $\theta = (\lambda, \delta, \psi)$



True value: $\theta_0 = (0, 0, \psi_0)$

Theorem: Under regularity conditions, if θ_N^* is chosen such that $\theta_N^* \in \Theta_0$, $\theta_N^* = \theta_0 + o_p(1)$ and $N^{1/4} \delta_N^* = o_p(1)$ then as $N \rightarrow +\infty$, it holds in probability that

$$pr^* \{ \text{lrt}(y_{1:N}^*) \leq t \} \rightarrow pr \{ \text{lrt}_\infty \leq t \}.$$

How to choose the bootstrap parameter $\theta_N^* = (\lambda_N^*, \delta_N^*, \psi_N^*)$?

Proposition: Let $(c_N)_{N \in \mathbb{N}}$ be a sequence such that $\lim_{N \rightarrow +\infty} c_N = 0$ and $\lim_{N \rightarrow +\infty} N^{1/4} c_N = +\infty$. Let $\hat{\theta}_N = (\hat{\psi}_N, \hat{\delta}_N, \hat{\lambda}_N)$ be the MLE of θ_0 .

Under regularity conditions if

- $\forall k = 1, \dots, d_\psi$, $d_\psi \psi_{N,k}^* = \hat{\psi}_{N,k} \mathbb{1}(\hat{\psi}_{N,k} > c_N)$
- $\forall k = 1, \dots, d_\delta$, $d_\delta \delta_{N,k}^* = \hat{\delta}_{N,k} \mathbb{1}(\hat{\delta}_{N,k} > c_N)$
- $\lambda_N^* = 0_{d_\lambda}$

then θ_N^* verifies the hypothesis of the main theorem.

Simulations study

Small sample properties:

Test that one variance is null in a linear model with two random effects

Level α	$N = 10$		$N = 20$		$N = 30$		$N = 40$		$N = 100$		max sd
	boot	asym	boot	asym	boot	asym	boot	asym	boot	asym	
1%	1.14	0.68	0.98	0.68	1.20	0.94	0.74	0.70	0.86	0.72	0.15
5%	5.20	3.64	5.22	3.82	5.74	4.30	4.86	3.94	5.26	4.50	0.33
10%	10.72	7.16	10.80	7.98	10.30	8.40	10.80	8.44	10.34	8.86	0.44

Table 1. Empirical levels (in %) for $K = 5000$ simulated datasets and $B = 500$

Test that one variance is null in a nonlinear model with three random effects (logistic model)

Level α	boot	asym	max sd
1%	0.80	0.80	0.28
5%	5.10	3.60	0.70
10%	10.30	7.00	0.96

Table 2. Empirical levels (in %) for $K = 1000$ datasets of size $N = 40$ and $B = 300$.

Effect of the shrinkage in presence of nuisance parameters:

Test that one variance is null in a linear model with 8 random effects

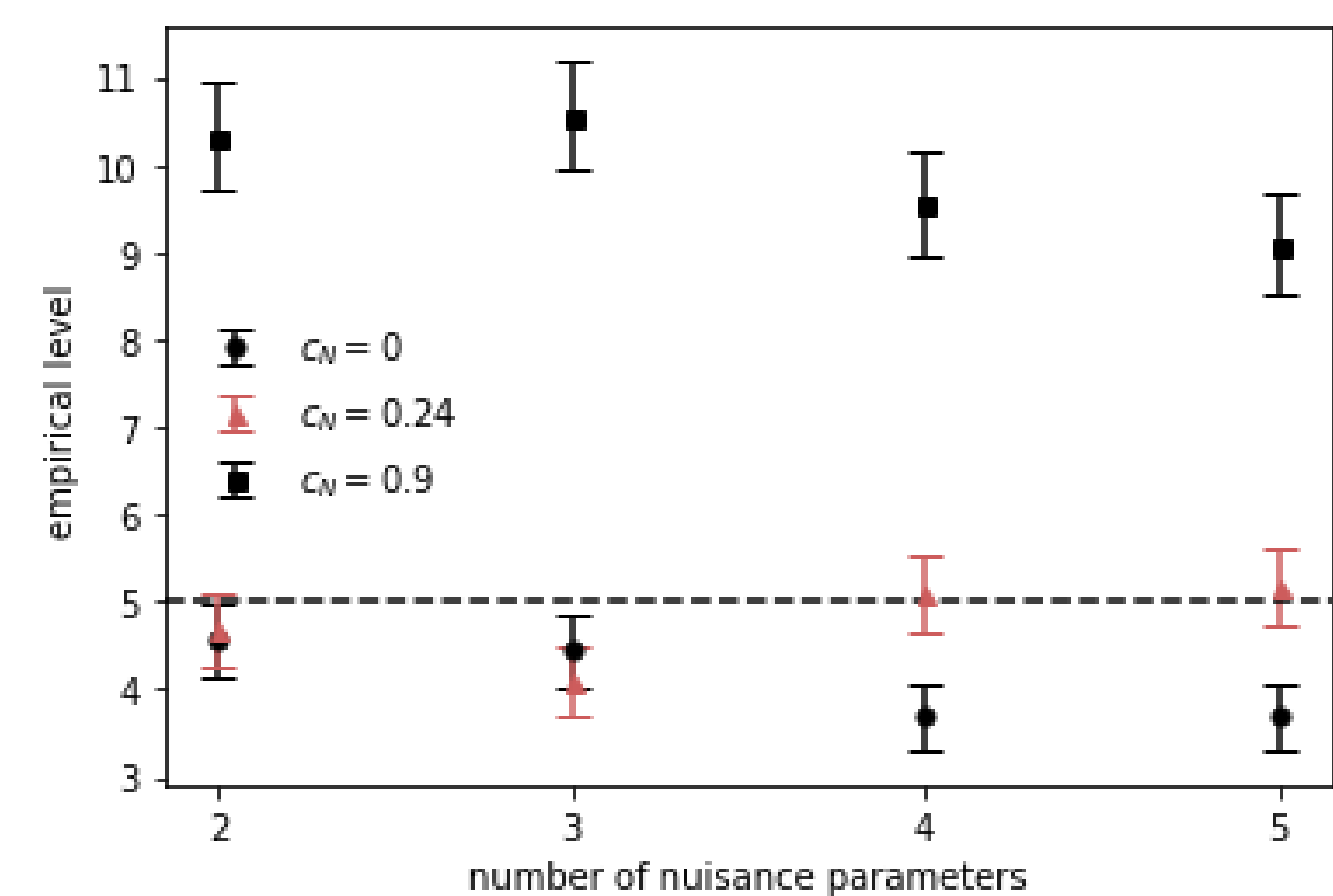


Figure 1. Empirical levels (in %) for $K = 2500$ datasets of size $N = 30$ and $B = 300$

Conclusion

Contributions: a test procedure

- with good small samples properties
- robust to nuisance parameters
- applicable to any type of mixed effects models

Perspectives:

- choice of c_N
- efficient computation of $\text{lrt}(y_{1:N})$

References

- D. W. Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, pages 399–405, 2000.
- C. Baey, P.-H. Cournède, and E. Kuhn. Asymptotic distribution of lrt statistics for variance components in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 135:107–122, 2019.
- G. Cavaliere, H. B. Nielsen, and A. Rahbek. On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *Journal of Time Series Analysis*, 38(4):513–534, 2017.