Les modèles linéaires généralisés à effets mixtes: Les prévisions sont-elles fiables?

Mathieu Fortin
UMR LERFoB
AgroParisTech/INRA

Journées CAQ-CAPSIS 8-10 avril 2013

Un modèle linéaire généralisé ?

- La grande famille des modèles dont la variable réponse est discrète
  - Par exemple
    - Régression logistique (oui/non)
    - Régression de Poisson (fréquence)

Un modèle linéaire, mais une forme non linéaire

$$\Pr(y_{ij} = 1|u_i) = \frac{e^{\mathbf{x}_{ij}\boldsymbol{\beta} + u_i}}{1 + e^{\mathbf{x}_{ij}\boldsymbol{\beta} + u_i}}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta} = \beta_0 + \beta_{1,s} + (\beta_{2,s} + \beta_{3,s}\mathrm{merch}_{ij})\Delta\mathrm{dbh}_{ij}$$

Modèle linéaire généralisé à effets mixtes (fixes et aléatoires)

- Une extension des modèles linéaires généralisés pour tenir compte de l'autocorrélation

- Une technique statistique relativement récente <10 ans

Un modèle linéaire, mais une forme non linéaire

$$\Pr(y_{ij} = 1|u_i) = \frac{e^{x_{ij}\beta + u_i}}{1 + e^{x_{ij}\beta + u_i}} \qquad u_i \sim N(0, \sigma_u^2)$$

$$\mathbf{x}_{ij}\boldsymbol{\beta} = \beta_0 + \beta_{1,s} + (\beta_{2,s} + \beta_{3,s}\mathrm{merch}_{ij})\Delta\mathrm{dbh}_{ij}$$

Des packages statistiques

- SAS
  - Procédure NLMIXED
  - Procédure GLIMMIX

- R
  - Fonction glmer – package lme4

En pratique,

- On ne connaît que très rarement l'effet aléatoire

- Que faire?

$$u_i \sim N(0, \sigma_u^2)$$

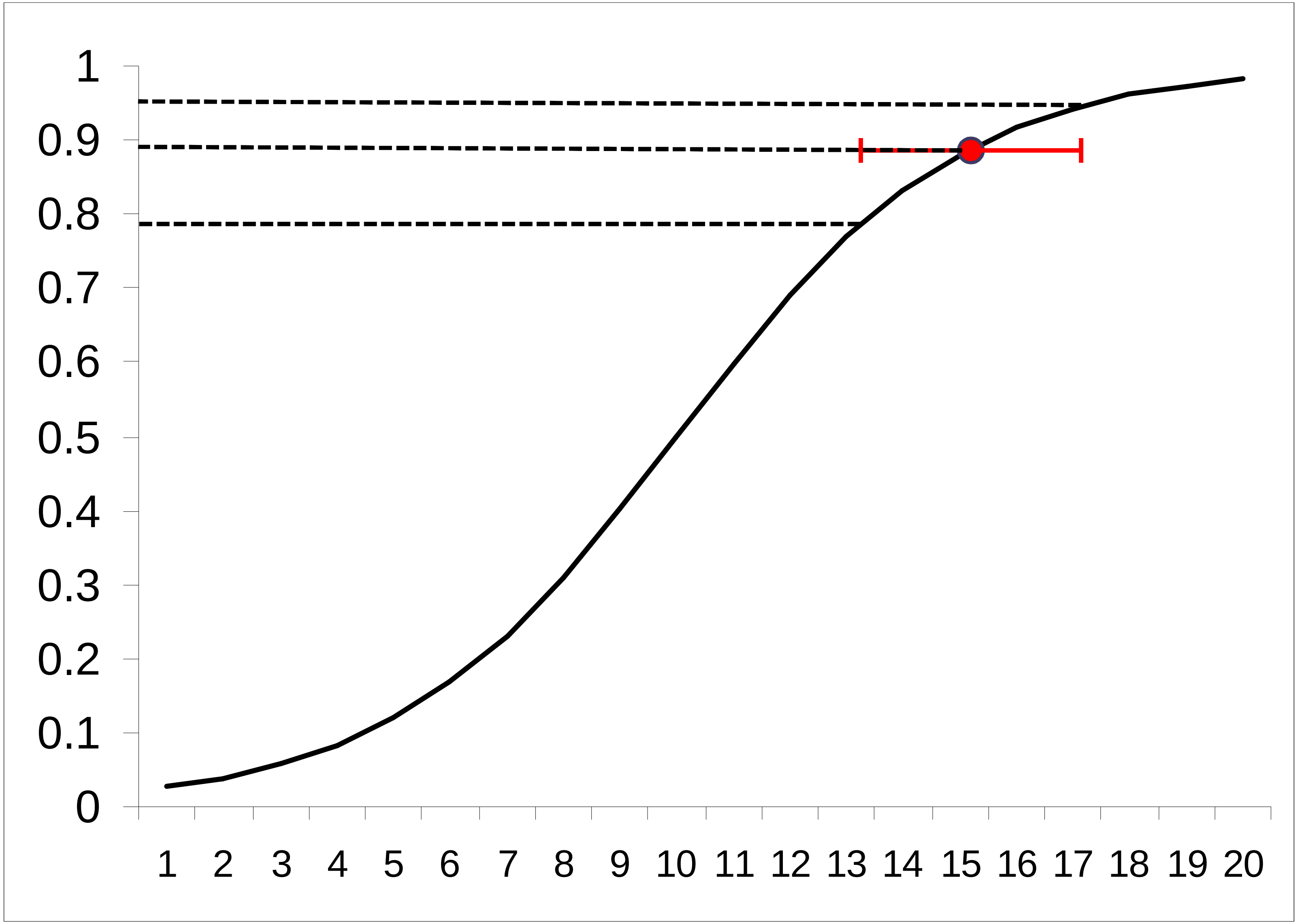
$$\Pr(y_{ij} = 1|u_i) = \frac{e^{x_{ij}\beta + u_i}}{1 + e^{x_{ij}\beta + u_i}}$$

$$\mathbf{x}_{ij}\boldsymbol{\beta} = \beta_0 + \beta_{1,s} + (\beta_{2,s} + \beta_{3,s}\text{merch}_{ij})\Delta\text{dbh}_{ij}$$

En pratique,

- On remplace l'effet aléatoire par la valeur la plus probable, soit zéro!

- Ca cause des biais (Groom et al. 2012)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

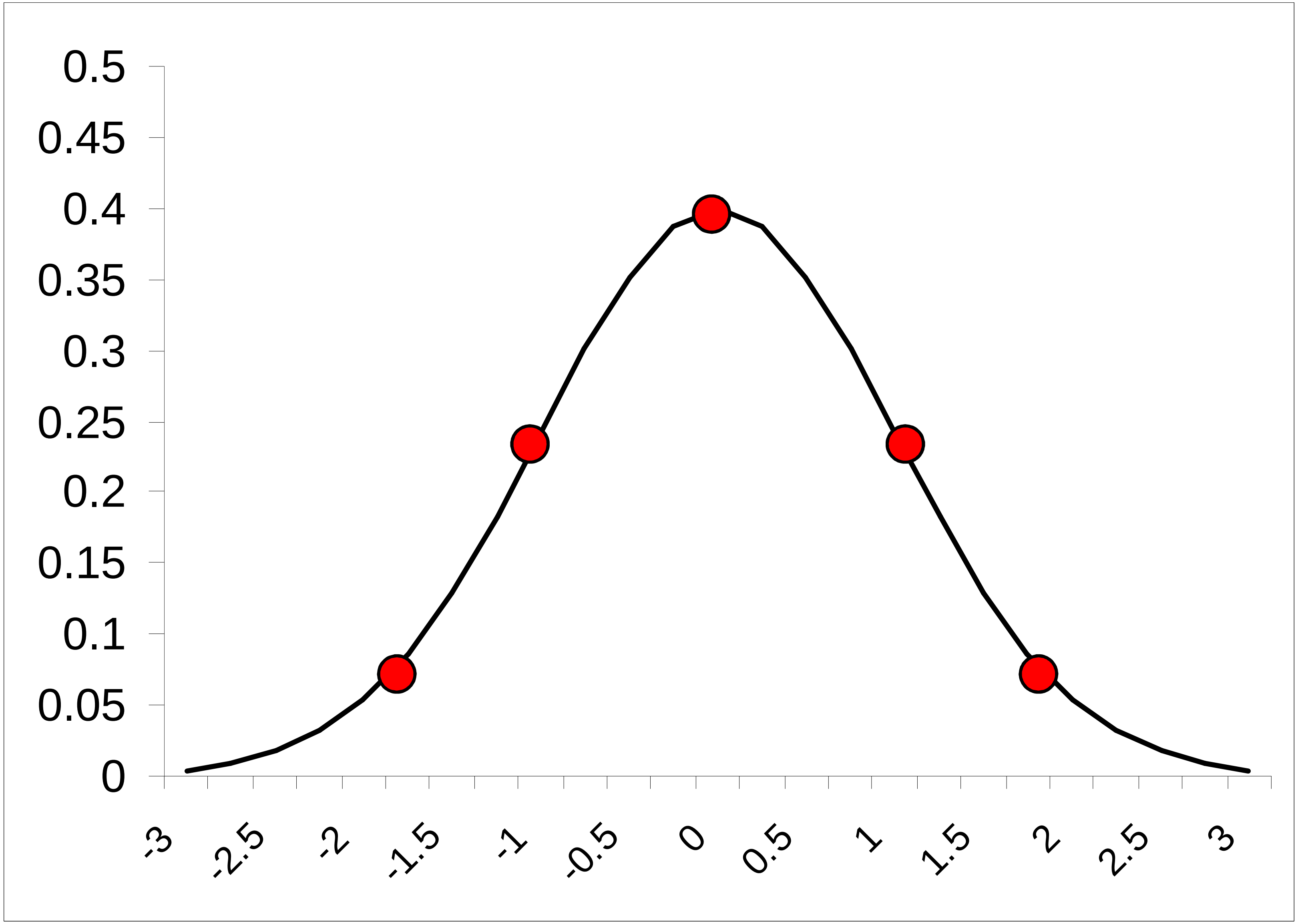
D’un point de vue statistique,

$$
\begin{aligned}
E[y_{ij}|\mathbf{x}_{ij}] &= \int E[y_{ij}|\mathbf{x}_{ij}, u_i]\mathrm{pdf}(u_i;\sigma_u^2)\mathrm{d}u_i \\
&= \int f(\mathbf{x}_{ij}\boldsymbol{\beta} + z_{ij}u_i)\mathrm{pdf}(u_i;\sigma_u^2)\mathrm{d}u_i
\end{aligned}
$$

Quadratures de Gauss

$$f(v_i) = w(v_i)h(v_i)$$

$$\int f(v_i)\mathrm{d}v_i \approx \sum_{k=1}^{r} w'_k h(v_{i,k})$$


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3

Quadrature de Gauss-Hermite

- On peut corriger l'estimation
- On peut obtenir un intervalle de confiance sans biais

Etude par simulation

- On tire 1000 échantillons d'une population dont on connaît les paramètres

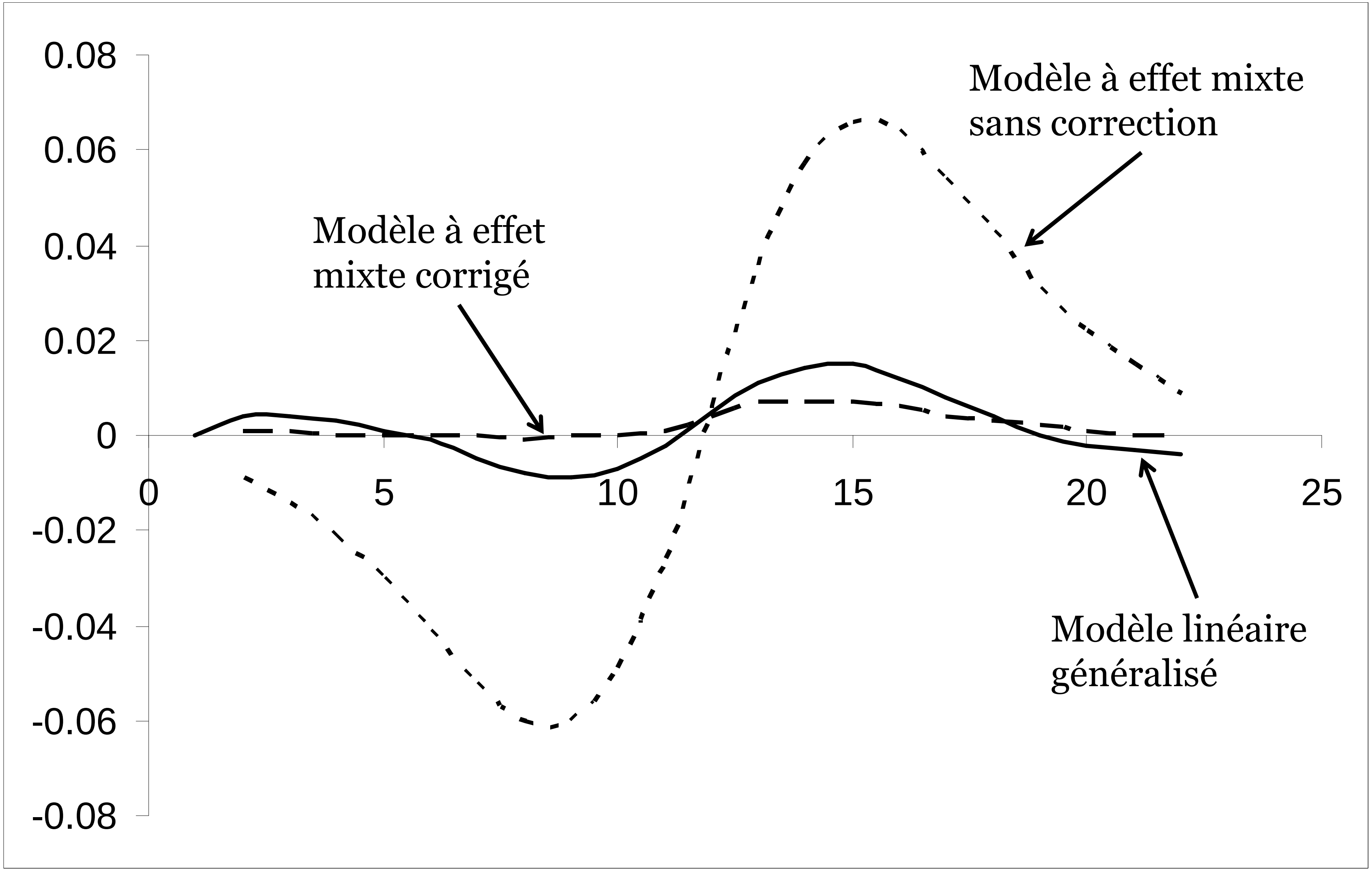
$$E[y_{ij}|\mathbf{x}_{ij}, u_i] = f(\mathbf{x}_{ij}\boldsymbol{\beta} + z_{ij}u_i) = \frac{e^{\beta_0+\beta_1 x_{ij}+u_i}}{1 + e^{\beta_0+\beta_1 x_{ij}+u_i}}$$

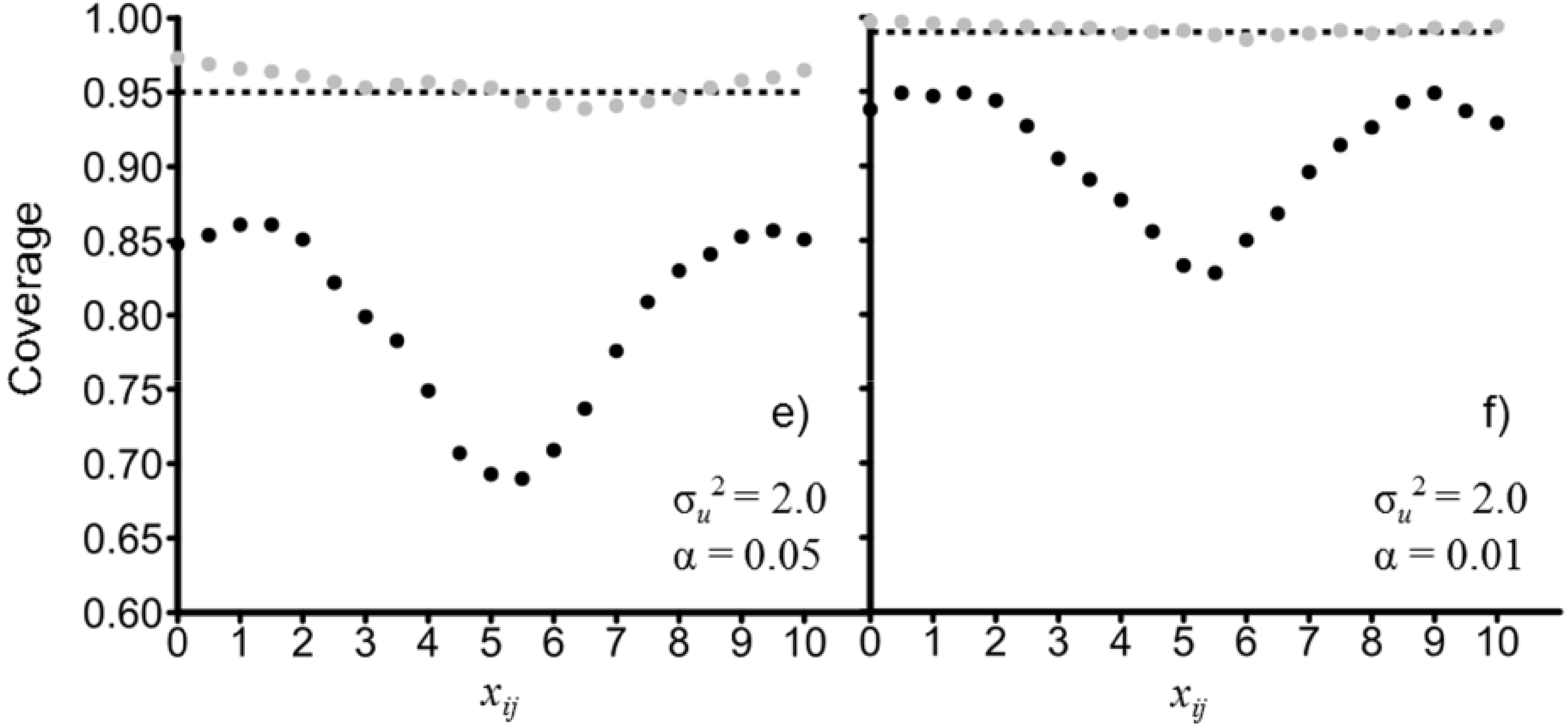
$$u_i \sim N(0, \sigma_u^2)$$

Etude par simulation

- On ajuste un modèle linéaire généralisé

- On ajuste un modèle linéaire généralisé à effet mixte
  1. On corrige avec le nouvel estimateur
  2. On ne corrige pas et on fixe l'effet aléatoire à zéro


0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
0
5
10
15
20
25
Modèle à effet mixte
sans correction
Modèle à effet
mixte corrigé
Modèle linéaire
généralisé


Coverage
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0 1 2 3 4 5 6 7 8 9 10
x_{ij}
e)
$\sigma_u^2 = 2.0$
$\alpha = 0.05$
f)
$\sigma_u^2 = 2.0$
$\alpha = 0.01$

Conclusions

- En ce qui concerne la moyenne
  - L'estimateur corrigé et la méthode traditionnelle sont sans biais
  - Un modèle mixte sans correction est la pire option
  - Plus la variance de l'effet aléatoire est élevée, plus la divergence est grande

Conclusions

- En ce qui concerne les intervalles de confiance
  - L'estimateur corrigé est sans biais
  - La méthode traditionnelle tend à sous-estimer la couverture des intervalles
  - Plus la variance de l'effet aléatoire est élevée, plus la sous-estimation est importante

Publication

- Fortin, M. 2013. *Population-averaged predictions with generalized linear mixed-effects models in forestry: an estimator based on Gauss-Hermite quadrature*. Canadian Journal of Forest Research 43: 129-138.

- Programmes disponibles pour adapter la fonction R glmer et la procédure SAS GLIMMIX