

Estimation de l'erreur quadratique moyenne du meilleur prédicteur linéaire empirique de la moyenne d'un petit domaine sous un modèle linéaire two-fold avec variances aléatoires

Marius STEFAN¹

RÉSUMÉ

Dans notre article nous proposons un modèle linéaire two-fold, contenant des effets petits domaines et unité primaires, avec variances aléatoires. Sous ce modèle nous obtenons la formule du meilleur prédicteur linéaire empirique de la moyenne d'un petit domaine, une approximation et un estimateur de son erreur quadratique moyenne. Nous prouvons que pour l'approximation, les termes négligés sont d'ordre $o(1/m)$, et que le biais de l'estimateur est du même ordre, m étant le nombre de petits domaines. Nous faisons ceci dans deux cas : *population infinie* et *population finie*.

1. INTRODUCTION

Un petit domaine est une sous-population pour laquelle, en appliquant les techniques classiques basées sur le plan de sondage, il est impossible d'obtenir des estimateurs ayant une bonne précision. Un tel estimateur est appelé *estimateur direct* et il utilise seulement les observations qui se trouvent dans le petit domaine. L'idée de la théorie des petits domaines est d'utiliser toutes les observations d'une enquête pour prédire μ_i , la moyenne du petit domaine i . Le prédicteur obtenu est dit *indirect* et s'obtient à partir d'un modèle statistique. Les modèles statistiques représentent un lien entre les petits domaines et les prédictions obtenues *borrow strength* des autres petits domaines.

Dans la section 2 nous proposons un nouveau modèle pour petits domaines et présentons des résultats théoriques concernant l'erreur quadratique moyenne du prédicteur obtenu. Dans la section 3 nous présentons quelques simulations réalisées qui confirment les résultats théoriques.

2. MODELE TWO FOLD AVEC VARIANCES ALEATOIRES

Stukel et Rao (1999) proposent un modèle two-fold avec variances constantes pour faire des prédictions pour petits domaines. Leur modèle se trouve ci-dessous :

$$y_{ijk} = \mathbf{x}_{ijk}'\boldsymbol{\beta} + v_i + u_{ij} + e_{ijk} \quad (1)$$
$$k=1, \dots, N_{ij}, j=1, \dots, M_i, i=1, \dots, m$$

Le modèle (1) est utile pour le cas où à l'intérieur de chaque petit domaine, les éléments sont regroupés dans des unités primaires ; v_i sont les effets petits domaines supposés indépendants et distribués selon $N(0, \sigma_v^2)$; u_{ij} sont les effets unité primaire j du petit domaine i , supposés indépendants est distribués selon $N(0, \sigma_u^2)$; e_{ijk} sont les erreurs du modèle, supposées indépendantes et distribuées selon $N(0, \sigma_e^2)$; m est le nombre de petits domaines sélectionnés, M_i est le nombre total d'unités primaires de i et N_{ij} est le nombre total d'éléments de l'unité primaire j du petit domaine i .

Pour sélectionner l'échantillon on prélève m petits domaines d'une population plus grande de petits domaines, à l'intérieur de chaque petit domaine on prélève m_i unités primaires, et à l'intérieur de

¹ Marius STEFAN, Université Libre de Bruxelles, CP124, Bruxelles, Belgique, mail : mastefan@ulb.ac.be

chaque unité primaire on prélève un échantillon de n_{ij} éléments. On considère que le plan de sondage est non informatif, à savoir les éléments sélectionnés vérifient le même modèle :

$$y_{ijk} = \mathbf{x}'_{ijk} \boldsymbol{\beta} + v_i + u_{ij} + e_{ijk} \quad (2)$$

$$k=1, \dots, n_{ij}, j=1, \dots, m_i, i=1, \dots, m$$

Les modèles (1) ou (2) supposent que les variances σ_u^2 et σ_e^2 sont les mêmes quelques soient i et j . En pratique cette hypothèse peut être restrictive si on a des petits domaines qui ne ressemblent pas du point de vue de leurs unités primaires. Nous proposons un modèle plus général où les variances dépendent de i :

$$y_{ijk} = \mu + v_i + u_{ij} + e_{ijk} \quad (3)$$

$$k=1, \dots, N, j=1, \dots, M', i=1, \dots, m$$

et sa forme pour les éléments sélectionnés :

$$y_{ijk} = \mu + v_i + u_{ij} + e_{ijk} \quad (4)$$

$$k=1, \dots, n, j=1, \dots, m', i=1, \dots, m$$

v_i sont indépendantes, distribuées selon $N(0, \sigma_v^2)$; u_{ij} sont conditionnellement indépendantes et $u_{ij} | \sigma_i^2 \sim N(0, \sigma_i^2)$; e_{ijk} sont conditionnellement indépendants et $e_{ijk} | \tau_i^2 \sim N(0, \tau_i^2)$; σ_i^2 sont indépendantes et $\sigma_i^2 \sim \text{loi}(\beta_1, \alpha_1)$; τ_i^2 sont indépendantes et $\tau_i^2 \sim \text{loi}(\beta_2, \alpha_2)$ ($\text{loi}(\beta, \alpha)$ signifie une loi quelconque de moyenne β et de variance α). σ_i^2 et τ_i^2 sont considérés aléatoires parce que les échantillons d'unités primaires et d'individus proviennent de populations plus grandes d'unités primaires, respectivement d'individus.

L'idée de passer d'un modèle avec variances constantes à un modèle avec variances aléatoires se trouve dans Kleffe et Rao (1992) qui l'ont fait pour un modèle one fold, à savoir un modèle où le niveau unité primaire n'existe pas.

On veut prédire μ_i , la moyenne du petit domaine i . On peut le faire sous deux approches différentes : *population infinie* et *population finie*. Dans le premier cas on suppose que M' est suffisamment grand pour que μ_i soit approximativement égale à une combinaison linéaire des effets du modèle :

$$\mu_i \approx \mu + v_i$$

Dans ce cas, prédire μ_i revient à prédire $\mu + v_i$ et on cherche le meilleur prédicteur linéaire non biaisé, qu'on notera $\tilde{\mu}_i$ (meilleur dans le sens de l'erreur quadratique moyenne). Dans le deuxième cas, μ_i est partagé en deux parties, les éléments observés et les éléments non observés :

$$\mu_i = \frac{1}{M'N} \left(\sum_{j=1}^{m'} \sum_{k=1}^n y_{ijk} + \sum_{j=1}^{m'} \sum_{k=n+1}^N y_{ijk} + \sum_{j=m'+1}^{M'} \sum_{k=1}^N y_{ijk} \right)$$

La deuxième et la troisième somme sont inconnues et seront prédites par un estimateur linéaire de manière à ce que le prédicteur de μ_i qui en résulte, noté $\tilde{\mu}_i^F$, soit de nouveau non biaisé et meilleur dans le sens de l'erreur quadratique moyenne (voir Bolfarine et Zacks (1992)).

On peut faire quelques remarques sur les modèles (3)-(4) :

- contrairement à (1)-(2) on suppose que l'échantillon et la population sont équilibrés (même nombre d'unités primaires pour tout domaine et même nombre d'éléments à l'intérieur de chaque unité primaire) ; l'hypothèse population équilibrée est nécessaire seulement pour l'approche *population finie* ;

- τ_i^2 sont supposées dépendre seulement de i , alors qu'un modèle plus général devrait travailler avec τ_{ij}^2 ;
- on n'utilise pas de variables auxiliaires.

Toutes ces suppositions ont des justifications théoriques. En minimisant les erreurs quadratiques moyennes on peut trouver les formules de $\tilde{\mu}_i$ et $\tilde{\mu}_i^F$ (F comme *population finie*) :

$$\tilde{\mu}_i = \bar{y}_{i..} - \frac{\beta}{m' \delta} (\bar{y}_{i..} - \bar{y})$$

$$\tilde{\mu}_i^F = \frac{1}{M'N} \{m'n\bar{y}_{i..} + \frac{1}{\delta} [(M'N - m'n)\sigma_v^2 + (N-n)\beta_1]\bar{y}_{i..} + \frac{1}{m'n\delta} [nN(M'-m')\beta_1 + (M'N - m'n)\beta_2]\bar{y}\}$$

$$\text{où } \beta = \beta_1 + \frac{\beta_2}{n}, \quad \delta = \sigma_v^2 + \frac{\beta_1}{m'} + \frac{\beta_2}{m'n}$$

On peut aussi calculer exactement les erreurs quadratique moyennes de $\tilde{\mu}_i$ et $\tilde{\mu}_i^F$:

$$EQM(\tilde{\mu}_i) = \frac{\beta\sigma_v^2}{m'\delta} + \frac{\beta^2}{mm'^2\delta}$$

$$EQM(\tilde{\mu}_i^F) = \frac{1}{(M'N)^2} \{ \sigma_v^2 (M'N - m'n)^2 + \beta_2 (M'N - m'n) + \beta_1 [m'(N-n)^2 + (M'-m')N^2] - \frac{1}{\delta} [\sigma_v^2 (M'N - m'n) + \beta_1 (N-n)]^2 + \frac{\delta}{m} [(M'N - m'n) - \frac{1}{\delta} (\sigma_v^2 (M'N - m'n) + \beta_1 (N-n))]^2 \}$$

$\tilde{\mu}_i$ et $\tilde{\mu}_i^F$ ne peuvent pas être utilisés en pratique pour obtenir des prédictions parce que leurs valeurs sont inconnues. On estime β , δ , β_2 , β_1 et σ_v^2 par les estimateurs non biaisés $\hat{\beta}$, $\hat{\delta}$, $\hat{\beta}_2$, $\hat{\beta}_1$ et respectivement $\hat{\sigma}_v^2$:

$$\hat{\beta} = \frac{1}{m(m'-1)} \sum_{i=1}^m \sum_{j=1}^{m'} (\bar{y}_{ij.} - \bar{y}_{i..})^2, \quad \hat{\delta} = \frac{1}{m-1} \sum_{i=1}^m (\bar{y}_{i..} - \bar{y})^2, \quad \hat{\beta}_2 = \frac{1}{mm'(n-1)} \sum_{i=1}^m \sum_{j=1}^{m'} \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2,$$

$$\hat{\beta}_1 = \hat{\beta} - \frac{\hat{\beta}_2}{n}, \quad \hat{\sigma}_v^2 = \hat{\delta} - \frac{\hat{\beta}}{m'}$$

Ensuite, on remplace ces estimateurs dans les formules de $\tilde{\mu}_i$ et $\tilde{\mu}_i^F$ pour obtenir les meilleurs prédicteurs linéaires empiriques de μ_i :

$$\hat{\mu}_i = \bar{y}_{i..} - \frac{\hat{\beta}}{m'\hat{\delta}} (\bar{y}_{i..} - \bar{y}),$$

$$\hat{\mu}_i^F = \frac{1}{M'N} \{m'n\bar{y}_{i..} + \frac{1}{\hat{\delta}} [(M'N - m'n)\hat{\sigma}_v^2 + (N-n)\hat{\beta}_1]\bar{y}_{i..} + \frac{1}{m'n\hat{\delta}} [nN(M'-m')\hat{\beta}_1 + (M'N - m'n)\hat{\beta}_2]\bar{y}\}$$

On veut maintenant calculer les erreurs quadratiques moyennes de $\hat{\mu}_i$ et $\hat{\mu}_i^F$. Contrairement à $\tilde{\mu}_i$ et $\tilde{\mu}_i^F$, ces erreurs ne peuvent pas être calculées exactement et ce serait naïf de considérer qu'ils ont la même précision (nous allons voir dans les simulations que ceci peut conduire à des sous-estimations importantes) :

$$EQM_N(\hat{\mu}_i) = EQM(\tilde{\mu}_i), \quad EQM_N(\hat{\mu}_i^F) = EQM(\tilde{\mu}_i^F)$$

Nous avons trouvé des approximations de $EQM(\hat{\mu}_i)$ et $EQM(\hat{\mu}_i^F)$ notées $EQM_A(\hat{\mu}_i)$ et $EQM_A(\hat{\mu}_i^F)$, et nous avons prouvé que les termes négligés sont d'ordre $o(1/m)$. Les résultats sont énoncés dans les deux théorèmes suivants :

Théorème 1

Soit le modèle (1) et les conditions de régularités suivantes :

- 1) $m' = O(1)$ et $n = O(1)$
- 2) σ_i^2 et τ_i^2 ont des moments d'ordre 12 finis

Alors :

$$EQM(\hat{\mu}_i) = \frac{3m'-1}{mm'^2(m'-1)} \frac{\beta^2}{\delta} + \frac{2}{mm'(m'-1)} \frac{\alpha}{\delta} + \frac{1}{m'} \frac{\sigma_v^2 \beta}{\delta} - \frac{3}{mm'^2} \frac{\sigma_v^4 \alpha}{\delta^3} + o\left(\frac{1}{m}\right)$$

$$\text{où } \alpha = \alpha_1 + \frac{\alpha_2}{n^2}.$$

Donc $EQM_A(\hat{\mu}_i)$ sera donnée par :

$$EQM_A(\hat{\mu}_i) = \frac{3m'-1}{mm'^2(m'-1)} \frac{\beta^2}{\delta} + \frac{2}{mm'(m'-1)} \frac{\alpha}{\delta} + \frac{1}{m'} \frac{\sigma_v^2 \beta}{\delta} - \frac{3}{mm'^2} \frac{\sigma_v^4 \alpha}{\delta^3}$$

Si on travaille avec un modèle one fold et si on a obtenu l'approximation $EQM_A(\hat{\mu}_i)$, alors on peut obtenir facilement l'approximation $EQM_A(\hat{\mu}_i^F)$. Pour un modèle two fold ceci n'est plus vrai. □

Théorème 2

Soit le modèle (1) et les conditions de régularités suivantes :

- 1) $M' = O(1)$, $m' = O(1)$, $N = O(1)$ et $n = O(1)$
- 2) σ_i^2 et τ_i^2 ont des moments d'ordre 12 finis

On définit :

$$T_1 = \frac{\beta_2}{m'n} - \frac{\beta_2^2}{m(m'n)^2 \delta} \left[m - 3 - \frac{2}{m'(n-1)} \right] + \frac{6\alpha_2 \beta_2}{m(m'n)^3 \delta^2} - \frac{3\alpha \beta_2^2}{mm'^4 n^2 \delta^3} - \frac{\alpha_2}{m(m'n)^2 \delta} \left[1 - \frac{2}{m'(n-1)} \right],$$

$$T_2 = \frac{\beta}{m'} - \frac{\beta^2}{m'^2 \delta} - \frac{3\alpha}{mm'^2 \delta} + \frac{\beta^2}{mm'^2 \delta} + \frac{2(\alpha + \beta^2)}{mm'(m'-1)\delta} + \frac{6\alpha\beta}{mm'^3 \delta^2} - \frac{3\beta^2 \alpha}{mm'^4 \delta^3},$$

$$T_3 = \frac{\beta_2}{m'n} - \frac{\beta_2 \beta}{m'^2 n \delta} \left(1 - \frac{3}{m} \right) - \frac{\alpha_2}{mm'^2 n^2 \delta} + \frac{3\beta_2 \alpha}{mm'^3 n \delta^2} + \frac{3\beta \alpha_2}{mm'^3 n^2 \delta^2} - \frac{3\beta_2 \beta \alpha}{mm'^4 n \delta}$$

Alors :

$$EQM(\hat{\mu}_i^F) = \frac{m'^2(N-n)^2}{(M'N)^2} T_1 + \frac{(M'-m')^2}{M'^2} T_2 + \frac{2m'(N-n)(M'-m')}{M'^2 N} T_3 + \frac{M'-m'}{M'^2} \beta + \frac{1}{M'^2} \left(\frac{M'N - m'n}{N^2} - \frac{M'-m'}{n} \right) \beta_2 + o\left(\frac{1}{m}\right)$$

Donc $EQM_A(\hat{\mu}_i^F)$ sera donnée par :

$$EQM_A(\hat{\mu}_i^F) = \frac{m'^2(N-n)^2}{(M'N)^2} T_1 + \frac{(M'-m')^2}{M'^2} T_2 + \frac{2m'(N-n)(M'-m')}{M'^2 N} T_3 + \frac{M'-m'}{M'^2} \beta + \frac{1}{M'^2} \left(\frac{M'N - m'n}{N^2} - \frac{M'-m'}{n} \right) \beta_2$$

Nous voulons maintenant trouver des estimateurs de $EQM(\hat{\mu}_i)$ et $EQM(\hat{\mu}_i^F)$. L'approche naïve est de remplacer dans les formules de $EQM(\hat{\mu}_i)$ et $EQM(\hat{\mu}_i^F)$ les paramètres inconnus par leurs estimateurs :

$$eqm_N(\hat{\mu}_i) = \frac{\hat{\beta} \hat{\sigma}_v^2}{m' \hat{\delta}} + \frac{\hat{\beta}^2}{mm'^2 \hat{\delta}},$$

$$eqm_N(\hat{\mu}_i^F) = \frac{1}{(M'N)^2} \{ \hat{\sigma}_v^2 (M'N - m'n)^2 + \hat{\beta}_2 (M'N - m'n) + \hat{\beta}_1 [m'(N-n)^2 + (M'-m')N^2] - \frac{1}{\hat{\delta}} [\hat{\sigma}_v^2 (M'N - m'n) + \hat{\beta}_1 (N-n)]^2 + \frac{\hat{\delta}}{m} [(M'N - m'n) - \frac{1}{\hat{\delta}} (\hat{\sigma}_v^2 (M'N - m'n) + \hat{\beta}_1 (N-n))]^2 \}$$

Par simulation on constate que $eqm_N(\hat{\mu}_i)$ et $eqm_N(\hat{\mu}_i^F)$ peuvent donner des sous estimations importantes. Nous avons trouvé des estimateurs dont le biais est d'ordre $o(1/m)$. Les estimateurs sont repris dans les deux théorèmes suivants :

Théorème 3

Soit le modèle (1) et les conditions de régularités suivantes :

- 1) $m' = O(1)$ et $n = O(1)$
- 2) σ_i^2 et τ_i^2 ont des moments d'ordre 12 finis

On définit :

$$\hat{\gamma}_i^2 = \frac{1}{m'-1} \sum_{j=1}^{m'} (\bar{y}_{ij} - \bar{y}_{i.})^2, \quad \hat{\alpha} = \frac{m'-1}{m(m'+1)} \sum_{i=1}^m (\hat{\gamma}_i^2) - \hat{\beta}^2$$

$$eqm(\hat{\mu}_i) = \frac{1}{m'} (\hat{\beta} - \frac{(m-1)\hat{\beta}^2}{mm'}) + \frac{4}{mm'} (\frac{1}{m'-1} \frac{\hat{\beta}^2}{\hat{\delta}} + \frac{1}{m'^2} \frac{\hat{\beta}\hat{\alpha}}{\hat{\delta}^2} + \frac{1}{m'(m'-1)} \frac{\hat{\alpha}}{\hat{\delta}})$$

Alors le biais de $eqm(\hat{\mu}_i)$ comme estimateur de $EQM(\hat{\mu}_i)$ est d'ordre $o(1/m)$. □

Théorème 4

Soit le modèle (1) et les conditions de régularités suivantes :

- 1) $M' = O(1)$, $m' = O(1)$, $N = O(1)$ et $n = O(1)$
- 2) σ_i^2 et τ_i^2 ont des moments d'ordre 12 finis

On définit :

$$\hat{\tau}_i^2 = \frac{1}{m'(n-1)} \sum_{j=1}^{m'} \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2, \quad \hat{\alpha}_2 = \frac{m'(n-1)}{m'(n-1)+2} \frac{1}{m} \sum_{i=1}^m (\hat{\tau}_i^2) - \hat{\beta}_2^2,$$

$$\hat{T}_1 = \frac{\hat{\beta}_2}{m'n} - \frac{m-1}{m(m'n)^2} \frac{\hat{\beta}_2^2}{\hat{\delta}} + \frac{4}{m(m'n)^2} \left[\frac{\hat{\beta}_2^2}{\hat{\delta}} + \frac{\hat{\beta}_2 \hat{\alpha}_2}{m'n \hat{\delta}^2} + \frac{\hat{\alpha}_2}{m'(n-1)\hat{\delta}} + \frac{\hat{\beta}_2^2}{m'(n-1)\hat{\delta}} \right],$$

$$\hat{T}_2 = \frac{\hat{\beta}}{m'} - \frac{m-1}{mm'^2} \frac{\hat{\beta}^2}{\hat{\delta}} + \frac{4}{mm'} \left[\frac{\hat{\beta}^2}{(m'-1)\hat{\delta}} + \frac{\hat{\beta}\hat{\alpha}}{m'^2 \hat{\delta}^2} + \frac{\hat{\alpha}}{m'(m'-1)\hat{\delta}} \right],$$

$$\hat{T}_3 = \frac{\hat{\beta}_2}{m'n} - \frac{m-1}{mm'^2 n} \frac{\hat{\beta}_2 \hat{\beta}}{\hat{\delta}} + \frac{2}{mm'^2 n} \left(\frac{2\hat{\beta}_2 \hat{\beta}}{\hat{\delta}} + \frac{\hat{\alpha}_2 \hat{\beta}}{m'n \hat{\delta}^2} + \frac{\hat{\alpha} \hat{\beta}}{m' \hat{\delta}^2} \right),$$

$$eqm(\hat{\mu}_i^F) = \frac{(N-n)^2 m'^2}{(M'N)^2} \hat{T}_1 + \frac{(M'-m')^2}{M'^2} \hat{T}_2 + \frac{2(N-n)(M'-m')m'}{M'^2 N} \hat{T}_3 + \frac{M'-m'}{M'^2} \hat{\beta} + \frac{1}{M'^2} \left(\frac{M'N - m'n}{N^2} - \frac{M'-m'}{n} \right) \hat{\beta}_2$$

Alors le biais de $eqm(\hat{\mu}_i^F)$ comme estimateur de $EQM(\hat{\mu}_i^F)$ est d'ordre $o(1/m)$. □

3. ETUDE MONTE CARLO

Dans cette section nous allons présenter des simulations qui montrent la qualité des approximations et des estimateurs ci-dessus pour les cas *population infinie* et *population finie*.

Le cas *population infinie*

On a pris $\mu = 0$, $m = 30$, $m' = 2$, $n = 2$ et $\beta_2 = 300$. On a donné à σ_v^2 et β_1 les valeurs 15, 30, 60, 150, 300 et 600. Les valeurs Monte Carlo ont été calculées à partir de $G=10000$ échantillons y_{ijk} , chaque échantillon étant calculé de la manière suivante :

- on a généré $\sigma_1^2, \dots, \sigma_{30}^2$ de $\chi^2(\beta_1)$ et $\tau_1^2, \dots, \tau_{30}^2$ de $\chi^2(\beta_2)$

- on a généré e_{ijk} de $N(0, \tau_i^2)$, u_{ij} de $N(0, \sigma_i^2)$, v_i de $N(0, \sigma_v^2)$

- on a calculé $y_{ijk} = v_i + u_{ij} + e_{ijk}$

Pour chaque échantillon on a calculé $\hat{\mu}_{1g}$, $\mu_{1g} = v_{1g}$, $EQM_N(\hat{\mu}_1)_g$, $EQM_A(\hat{\mu}_1)_g$, $eqm_N(\hat{\mu}_1)_g$, $eqm(\hat{\mu}_1)_g$.

L'erreur quadratique moyenne et les moyennes des estimateurs ont été simulées utilisant les formules :

$$EQM_{MC}(\hat{\mu}_1) = \frac{1}{10000} \sum_{g=1}^{10000} (\hat{\mu}_{1g} - \mu_{1g})^2, E_{MC}(eqm) = \frac{1}{10000} \sum_{g=1}^{10000} eqm(\hat{\mu}_1)_g, \text{etc...}$$

Ensuite, on a calculé les erreurs relatives et les biais relatifs des approximations respectivement des estimateurs :

$$ER = 100 \times \frac{EQM_A(\hat{\mu}_1) - EQM_{MC}(\hat{\mu}_1)}{EQM_{MC}(\hat{\mu}_1)}, BR = 100 \times \frac{E_{MC}(eqm) - EQM_{MC}(\hat{\mu}_1)}{EQM_{MC}(\hat{\mu}_1)}, \text{etc...}$$

Dans les quatre tableaux suivants se trouvent les résultats des simulations :

Tableau 1

L'erreur relative (%) de $EQM_N(\hat{\mu}_1)$, l'approximation naïve de l'erreur quadratique moyenne de $\hat{\mu}_1$
($m=30$, $m'=2$, $n=2$)

		β_1 / β_2					
		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-45.73	-39.71	-46.22	-51.89	-60.30	-68.74
	0.1	-25.24	-27.30	-30.47	-39.91	-44.36	-51.52
	0.2	-13.60	-14.74	-18.37	-22.01	-32.19	-41.21
	0.5	-8.00	-6.58	-8.55	-9.70	-16.80	-25.04
	1	-2.81	-4.69	-3.61	-5.63	-10.24	-16.46
	2	-1.37	-2.28	-2.33	-3.36	-2.78	-6.06

Tableau 2

L'erreur relative (%) de $EQM_A(\hat{\mu}_1)$, l'approximation de l'erreur quadratique moyenne de $\hat{\mu}_1$
($m=30$, $m'=2$, $n=2$)

		β_1 / β_2					
		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-6.82	0.43	-2.91	-1.73	-2.97	-4.00
	0.1	0.01	-0.51	-0.87	-3.32	0.16	2.65
	0.2	1.39	1.34	-0.22	1.62	-1.32	-0.34
	0.5	-1.24	0.72	-0.15	1.79	-0.75	-1.45
	1	0.71	-0.86	0.81	0.54	-1.31	-2.58
	2	0.42	-0.32	-0.05	-0.15	1.96	1.51

Tableau 3
Le biais relatif (%) de $eqm_N(\hat{\mu}_1)$, l'estimateur naïf de
l'erreur quadratique moyenne de $\hat{\mu}_1$
($m=30, m'=2, n=2$)
 β_1 / β_2

		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-80.50	-84.10	-89.38	-107.02	-117.19	-129.07
	0.1	-52.35	-55.98	-60.88	-77.37	-88.72	-112.96
	0.2	-29.63	-32.05	-37.68	-47.08	-64.35	-83.28
	0.5	-14.69	-14.21	-17.31	-22.76	-33.81	-49.03
	1	-6.38	-8.39	-8.60	-12.41	-19.70	-30.47
	2	-3.07	-4.62	-5.09	-6.64	-7.62	-14.39

Tableau 4
Le biais relatif (%) de $eqm(\hat{\mu}_1)$, l'estimateur de
l'erreur quadratique moyenne de $\hat{\mu}_1$
($m=30, m'=2, n=2$)
 β_1 / β_2

		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-3.42	2.01	1.05	-0.04	1.86	2.23
	0.1	1.85	1.48	1.92	-1.16	5.86	6.08
	0.2	3.31	3.39	1.65	4.23	0.91	3.77
	0.5	-0.06	1.83	0.94	2.86	0.57	0.70
	1	1.45	-0.11	1.16	1.24	-0.45	-1.49
	2	0.92	-0.34	-0.07	0.37	2.93	2.30

A partir des tableaux 1 et 3 on peut voir que les approches naïves peuvent donner des sous estimations importantes. On observe que pour une même valeur de σ_v^2 / β_2 , quand on augmente β_1 / β_2 , l'erreur relative de $EQM_N(\hat{\mu}_1)$ augmente en valeur absolue, alors que pour une même valeur de β_1 / β_2 , si on augmente σ_v^2 / β_2 , alors l'erreur relative diminue en valeur absolue. Pour des valeurs de plus en plus grandes de σ_v^2 / β_2 , $EQM_N(\hat{\mu}_1)$ et $EQM_A(\hat{\mu}_1)$ donnent des valeurs de plus en plus proches. On a les mêmes conclusions pour les biais relatifs.

Le cas *population finie*

On a pris $\mu = 0$, $m = 30$, $M' = 8$, $N = 8$, $m' = 2$, $n = 2$; $\beta_2 = 300$. On a donné à σ_v^2 et β_1 les valeurs 15, 30, 60, 150, 300 et 600. Les valeurs Monte Carlo ont été calculées à partir de $G=10000$ échantillons y_{ijk} . Contrairement au cas *population infinie*, ici on génère d'abord toute la population et ensuite on a prélevé un échantillon dans chaque population générée. Les 10000 populations ont été générées de la manière suivante :

- on génère $\sigma_1^2, \dots, \sigma_{30}^2$ de $\chi^2(\beta_1)$ et $\tau_1^2, \dots, \tau_{30}^2$ de $\chi^2(\beta_2)$
- on génère e_{ijk} de $N(0, \tau_i^2)$, u_{ij} de $N(0, \sigma_i^2)$, v_i de $N(0, \sigma_v^2)$

- on calcule $y_{ijk} = v_i + u_{ij} + e_{ijk}$, $i=1, \dots, m$ $j=1, \dots, M'$ $k=1, \dots, N$

Les échantillons ont été obtenus en prélevant dans chaque petit domaine, m' unités primaires, et dans chaque unité primaire n individus, dans les deux cas les échantillons étant aléatoires simples. Pour chaque échantillon on a calculé $\hat{\mu}_{1g}^F$, μ_{1g} comme la moyenne de toute la population, $EQM_N(\hat{\mu}_1^F)_g$, $EQM_A(\hat{\mu}_1^F)_g$, $eqm_N(\hat{\mu}_1^F)_g$, $eqm(\hat{\mu}_1^F)_g$. L'erreur quadratique moyenne et les moyennes des estimateurs ont été simulées utilisant les mêmes formules que ci-dessus. Dans les quatre tableaux suivants se trouvent les résultats des simulations :

Tableau 5
L'erreur relative (%) de $EQM_N(\hat{\mu}_1^F)$, l'approximation naïve de
l'erreur quadratique moyenne de $\hat{\mu}_1^F$
($m=30$, $M'=8$, $N=8$, $m'=2$, $n=2$)

		β_1 / β_2					
		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-29.31	-26.84	-25.77	-27.21	-27.23	-26.14
	0.1	-21.97	-18.90	-19.18	-20.77	-23.04	-25.06
	0.2	-13.19	-11.06	-13.09	-14.74	-19.54	-20.34
	0.5	-4.91	-4.47	-8.22	-7.93	-10.78	-11.37
	1	-4.36	-2.32	-3.20	-4.96	-6.49	-10.48
	2	-2.78	-1.10	-2.10	-2.45	-3.49	-7.23

Tableau 6
L'erreur relative (%) de $EQM_A(\hat{\mu}_1^F)$, l'approximation de
l'erreur quadratique moyenne de $\hat{\mu}_1^F$
($m=30$, $M'=8$, $N=8$, $m'=2$, $n=2$)

		β_1 / β_2					
		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-2.99	-0.60	0.47	-0.89	-0.83	0.33
	0.1	-3.85	-0.47	0.03	0.12	-0.56	-1.09
	0.2	-2.12	0.49	-0.57	0.05	-2.17	-0.22
	0.5	0.18	0.97	-2.07	-0.04	-0.51	2.25
	1	-1.65	0.57	0.09	-0.52	-0.37	-1.64
	2	-1.38	0.38	-0.38	-0.09	-0.11	-2.01

Tableau 7
 Le biais relatif (%) de $eqm_N(\hat{\mu}_1^F)$, l'estimateur naïf de
 l'erreur quadratique moyenne de $\hat{\mu}_1^F$
 ($m=30, M'=8, N=8, m'=2, n=2$)
 β_1 / β_2

		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-55.93	-53.37	-53.67	-54.91	-54.62	-55.13
	0.1	-39.42	-38.15	-39.91	-42.52	-45.69	-50.35
	0.2	-24.15	-23.12	-26.14	-30.13	-37.33	-42.17
	0.5	-10.00	-10.20	-14.47	-16.61	-21.25	-26.52
	1	-7.17	-5.55	-6.65	-9.33	-13.26	-19.34
	2	-4.42	-2.57	-4.24	-4.74	-6.75	-12.57

Tableau 8
 Le biais relatif (%) de $eqm(\hat{\mu}_1^F)$, l'estimateur de
 l'erreur quadratique moyenne de $\hat{\mu}_1^F$
 ($m=30, M'=8, N=8, m'=2, n=2$)
 β_1 / β_2

		0.05	0.1	0.2	0.5	1	2
σ_v^2 / β_2	0.05	-1.82	1.65	2.51	0.60	1.15	1.71
	0.1	-2.61	0.83	1.20	2.09	1.22	0.07
	0.2	-1.16	1.87	0.64	1.33	-0.90	1.08
	0.5	1.01	1.65	-1.43	0.57	0.69	3.50
	1	-1.38	0.82	0.49	0.23	-0.12	-0.54
	2	-1.43	0.71	-0.50	0.45	0.70	-1.58

En comparant les tableaux 5 et 6, on constate la supériorité de $EQM_A(\hat{\mu}_1^F)$ par rapport à $EQM_N(\hat{\mu}_1^F)$. Cependant, l'erreur relative de $EQM_N(\hat{\mu}_1^F)$ diminue quand on augmente σ_v^2 / β_2 (comme dans le cas *population infinie*), mais garde plus ou moins la même valeur quand on augmente β_1 / β_2 (contrairement au cas *population infinie*). On a les mêmes conclusions pour le biais.

RÉFÉRENCES

- [1] Bolfarine, H. et Zacks, S. (1992), *Prediction theory for finite population*, New York, Springer-Verlag.
- [2] Kleffe, J. et Rao J.N.K. (1992), « Estimation of mean squared error of empirical best linear unbiased predictors under a random error variance linear model », *Journal of multivariate analysis*, Vol 43, 1-15.
- [3] Rao, J.N.K. (2003), *Small area estimation*, New York, Wiley
- [4] Stukel, D.M. et Rao J.N.K. (1999), « Small area estimation under two fold nested error regression models », *Journal of statistical planning and inference*, No.78, 131-147.