Techniques d'enquête, décembre 2000
Vol. 26, N° 2, pp. 197-206
Statistique Canada, N° 12-001 au catalogue

Estimation bayésienne hiérarchique des moyennes pour petites régions à l'aide de modèles à plusieurs niveaux

Yong You et J.N.K. Rao [1]

Résumé

On examine les modèles standard à plusieurs niveaux comportant des paramètres de régression aléatoires pour les estimations régionales. Les auteurs élargissent également les modèles en admettant une variance inégale de l'erreur ou en supposant des modèles à effets aléatoires tant pour les paramètres de régression que pour la variance de l'erreur. Les auteurs présentent ces modèles en fonction d'un cadre hiérarchique de Bayes et ils estiment une moyenne régionale en fonction de sa moyenne a posteriori. La variance a posteriori de la moyenne régionale sert de mesure de la précision de l'estimation. Elle tient compte automatiquement de l'incertitude supplémentaire associée aux hyperparamètres du modèle à plusieurs niveaux. Les auteurs utilisent l'échantillonnage de Gibbs pour calculer les moyennes a posteriori et la variance a posteriori de moyennes régionales. Ils obtiennent des estimateurs Rao-Blackwellisés réduisant les erreurs de Monte Carlo. On étudie également la sélection d'un modèle bayésien et l'analyse de la sensibilité. La procédure est illustrée à l'aide de données sur le revenu des ménages de quelques comtés (régions) du Brésil.

Mots clés : Échantillonnage de Gibbs; hiérarchique de Bayes; modèle à plusieurs niveaux; variance de l'erreur d'échantillonnage; région.

1. Introduction

Depuis quelques années, on accorde beaucoup d'attention à l'estimation régionale à cause de la demande accrue d'estimateurs régionaux fiables. Les estimateurs directs traditionnels propres à une région ne fournissent pas une précision adéquate puisque la taille des échantillons régionaux est rarement assez grande. Il est donc nécessaire d'utiliser des estimateurs indirects qui s'appuient sur des régions connexes, notamment des estimateurs indirects modélisés. Battese, Harter et Fuller (1988) ont proposé et appliqué un modèle de régression à erreur emboîtée afin d'obtenir des estimations régionales modélisées. Le modèle se présente comme suit :

$$y_{ij} = x_{ij}^T \beta + v_{0i} + e_{ij},\ j = 1, \ldots, n_i; i = 1, \ldots, m, \qquad (1)$$

où y_{ij} représente les observations associées aux unités échantillonnées de la $i^{\text{ième}}$ région, $i = 1, \ldots, m$, x_{ij} représente le vecteur $p \times 1$ des variables explicatives au niveau de l'unité, β est un ensemble de p paramètres de régression fixes, v_{0i} sont des effets régionaux indépendants avec $E(v_{0i}) = 0$ et $V(v_{0i}) = \sigma_v^2$. Nous supposons que les e_{ij} sont des variables de l'erreur aléatoire indépendantes avec $E(e_{ij}) = 0$ et $V(e_{ij}) = \sigma_e^2$. v_{0i} et e_{ij} sont également supposés indépendants. Pour l'ensemble de la population, le modèle (1) s'applique avec n_i remplacé par N_i, la taille de la population régionale. Le modèle (1) se laisse exprimer en fonction d'une notation matricielle comme suit :

$$Y_i = X_i \beta + v_{0i} 1_{n_i} + e_i, i = 1, \ldots, m,$$

où $Y_i = (y_{i1}, \ldots, y_{i,n_i})^T$, $X_i = (x_{i1}, \ldots, x_{i,n_i})^T$ est une matrice $n_i \times p$, $1_{n_i} = (1, \ldots, 1)^T$ est le vecteur unité de longueur n_i et $e_i = (e_{i1}, \ldots, e_{i,n_i})^T$.

Holt et Moura (1993) ont élargi le cadre ci-dessus pour en faire un modèle à plusieurs niveaux en introduisant des coefficients de régression aléatoires et en les rattachant à des variables explicatives au niveau de la région afin d'expliquer une partie des variations entre les régions. Le modèle se laisse énoncer comme suit :

$$Y_i = X_i \beta_i + e_i, \beta_i = Z_i \gamma + v_i \qquad (2)$$

où Z_i est la matrice du plan $p \times q$ des variables au niveau de la région, γ est un vecteur $q \times 1$ de coefficients fixes et $v_i = (v_{i1}, \ldots, v_{ip})^T$ est un vecteur $p \times 1$ d'effets aléatoires pour la $i^{\text{ième}}$ région. Les v_i sont indépendants d'une région à l'autre et comportent une distribution composée au sein de chaque région avec $E(v_i) = 0$ et $V(v_i) = \Phi$, où la matrice de variances-covariances Φ est inconnue. À noter que le modèle (2) intègre effectivement le recours à des covariables au niveau de l'unité et au niveau de la région en un même modèle. Holt et Moura (1993) et Moura et Holt (1999) ont élargi le cadre adopté par Prasad et Rao (1990) pour en faire le modèle à plusieurs niveaux ci-dessus de façon à obtenir le meilleur prédicteur linéaire sans biais (MPLSB) de la moyenne régionale $\mu_i = \bar{X}_i^T \beta_i$ en supposant N_i grand, où $\bar{X}_i$ est le vecteur $p \times 1$ des moyennes connues de la population des variables auxiliaires pour la $i^{\text{ième}}$ région. Ils ont également obtenu le MPLSB empirique et une approximation d'ordre deux de l'erreur quadratique moyenne (EQM) du MPLSB empirique pour le modèle à plusieurs niveaux. À l'aide de données sur le revenu des ménages de

1. Yong You, Division des méthodes des enquêtes auprès des ménages, Statistique Canada, Ottawa (Ontario) Canada K1A 0T6; J.N.K. Rao, School of Mathematics and Statistics, Université Carleton, Ottawa (Ontario) Canada K1S 5B6.

comtés (régions) du Brésil, ils ont obtenu un accroissement de l'efficacité des estimateurs de type MPLSB empirique relativement aux estimateurs MPLSB empiriques obtenus à l'aide de modèles de régression à erreur emboîtée. Ghosh et Rao (1994) et Rao (1999) ont donné un aperçu détaillé des méthodes d'estimation régionale modélisées.

Dans le présent exposé, nous examinons le modèle (2) à plusieurs niveaux selon un cadre hiérarchique de Bayes et nous l'étendons à des modèles à plusieurs niveaux plus généraux qui admettent des variances de l'erreur inégale ou des variances de l'erreur aléatoire fixes. La moyenne regionale μ_i est estimée à l'aide de sa moyenne a posteriori, et sa précision est mesurée en fonction de sa variance a posteriori. La variance a posteriori tient compte automatiquement de l'incertitude supplémentaire associée aux hyperparamètres du modèle à plusieurs niveaux. Nous utilisons la méthode d'échantillonnage de Gibbs afin d'obtenir les estimations bayésiennes hiérarchiques et les variances a posteriori connexes. À la section 2, nous présentons les modèles bayésiens hiérarchiques à plusieurs niveaux avec diverses hypothèses pour ce qui est de la variance de l'erreur et de l'inférence de l'échantillonnage de Gibbs connexe. À la section 3, nous illustrons notre méthode et nous abordons la sélection du modèle et l'analyse de la sensibilité à l'aide de données sur le revenu des ménages de comtés (régions) du Brésil. Enfin, à la section 4, nous présentons des remarques et nos conclusions.

2. Modèle à plusieurs niveaux et inférence de l'échantillonnage de Gibbs

2.1 Variance de l'erreur égale

Nous considérons une représentation bayésienne hiérarchique du modèle (2) à plusieurs niveaux comme suit :

Modèle 1 :

(i) Sous réserve que β_i et σ_e^2, les y_{ij} sont indépendants avec

$$y_{ij}|\beta_i, \sigma_e^2 \sim N(x_{ij}^T\beta_i, \sigma_e^2), \quad (i=1,\ldots,m;\ j=1,\ldots,n_i); \tag{3}$$

(ii) Sous réserve que γ et Φ, les β_i sont indépendants avec

$$\beta_i|\gamma, \Phi \sim N_p(Z_i\gamma, \Phi), (i=1,\ldots,m). \tag{4}$$

Afin de compléter notre description de modèle bayésien, nous adaptons les distributions a priori pour des paramètres comme suit :

(iii) Distributions a priori marginales : $\gamma \sim N_q(0, D)$, $\tau_e \sim G(a, b)$ et $\Omega \sim W_p(\alpha, R)$, où $\tau_e = \sigma_e^{-2}$, $\Omega = \Phi^{-1}$ et D, a, b, α et R connus.

À l'étape (iii) du modèle 1, $G(a, b)$ désigne une distribution gamma avec une densité donnée par $f(x) = b^a/\Gamma(a)\,x^{a-1}e^{-xb}$, $a>0, b>0, x\geq 0$, et $W_p(\alpha, R)$ est une distribution de Wishart avec comme fonction de densité

$$f(X) = \frac{|R|^{\frac{\alpha}{2}}}{2^{\alpha p/2}\Gamma_{p\left(\frac{\alpha}{2}\right)}}|X|^{\frac{\alpha-p-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}(RX)\right\},$$

où $X>0, R>0$ et $\Gamma_p(\alpha)$ est une fonction gamma à plusieurs variables définie comme

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left(\alpha+\frac{1}{2}(1-j)\right).$$

Remarque 1.1 : Les distributions a priori à l'étape (iii) et les distributions d'échantillonnage et de population données par (3) et par (4) sont conjuguées en ce sens qu'elles mènent à des distributions conditionnelles complètes pour γ, τ_e et Ω qui sont encore une fois des distributions normale, gamma et de Wishart respectivement. La distribution de Wishart est la version à plusieurs variables d'une distribution gamma pour la matrice de variances-covariances inverse d'effets aléatoires. L'importance de la conjugaison peut être exploitée comme suit: (1) à l'étape de l'échantillonnage de Gibbs, en l'absence de conjugaison, la distribution conditionnelle complète pour un paramètre quelconque sera connue jusqu'aux constantes normalisatrices; dans ce cas, une génération aléatoire plus perfectionnée sera requise; (2) des distributions conditionnelles complètes de forme fermée peuvent servir à trouver les estimateurs Rao-Blackwellisés des moyennes a posteriori et des variances a posteriori, et donc à améliorer l'estimation a posteriori. En général, pour une inférence bayésienne, le choix de distributions a priori n'est pas une tâche simple puisque toute distribution a priori appropriée sur les paramètres de modèle est un candidat plausible. C'est là une limitation des méthodes bayésiennes.

Remarque 1.2 : Il importe de noter que nous avons utilisé des distributions a priori appropriées sur tous les paramètres inconnus afin de nous assurer que toutes les distributions a posteriori sont appropriées (Hobert et Casella 1996). Ainsi, nous n'avons pas à craindre la présence de distributions a posteriori inappropriées. La valeur des paramètres des distributions a priori (les hyperparamètres) est choisie de façon à refléter une connaissance assez vague des distributions a priori. Des détails seront fournis à la section 3 sur l'analyse des données.

Remarque 1.3 : Dans le modèle 1, nous supposons une variance de l'erreur égale σ_e^2 pour toutes les régions. Dans la pratique, toutefois, les variances de l'erreur d'échantillonnage pourraient être différentes pour diverses régions. Un modèle plus général devrait admettre des variances d'erreur possiblement différentes. Dans les sections 2.2 et 2.3, nous introduirons des modèles de variance d'erreur inégale et de variance d'erreur aléatoire.

Nous voulons trouver les distributions a posteriori des β_i pour les données $Y = (\{y_{ij}\}, i = 1, \ldots, m;\ j = 1, \ldots, n_i)$ et, en particulier, trouver les estimations a posteriori des moyennes régionales $\mu_i = \bar{X}_i^T \beta_i$, qui dépendent des estimations de β_i. L'évaluation directe de la distribution a posteriori composée comporte une intégration numérique très dimensionnelle, et le calcul n'est pas faisable. Par conséquent, nous utilisons la méthode d'échantillonnage de Gibbs (Gelfand et Smith 1990) afin de produire des échantillons à partir des distributions a posteriori composées. Pour appliquer l'échantillonnage de Gibbs dans le cadre du modèle 1, il nous faut les distributions conditionnelles complètes données par :

(i) Pour $i = 1, \ldots, m$,

$$[\beta_i | Y, \gamma, \Omega, \tau] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i + \Omega)^{-1} (\tau_i X_i^T Y_i + \Omega Z_i \gamma), (\tau_i X_i^T X_i + \Omega)^{-1})$$

(ii) $$[\gamma | Y, \beta, \Omega, \tau_e] \sim N_q\left(\left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1} \left(\sum_{i=1}^m Z_i^T \Omega \beta_i\right), \left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iii) $$[\Omega | Y, \beta, \gamma, \tau] \sim W_p\left(\alpha + m, R + \frac{1}{2}\sum_{i=1}^m (\beta_i - Z_i\gamma)(\beta_i - Z_i\gamma)^T\right)$$

(iv) $$[\tau_e | Y, \beta, \gamma, \Omega] \sim G\left(a + \frac{1}{2}\sum_{i=1}^m n_i,\ b + \frac{1}{2}\sum_{i=1}^m (Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i)\right).$$

Toutes les distributions conditionnelles complètes sont de forme fermée et permettent facilement de produire des échantillons. La méthode d'échantillonnage de Gibbs se présente comme suit : a) À l'aide de valeurs de départ $\gamma^{(0)}, \Omega^{(0)}$ et $\tau_e^{(0)}$, il s'agit de tirer $\beta_i^{(1)}, i = 1, \ldots, m$, à partir de $[\beta_i | Y, \gamma, \Omega, \tau_e]$; b) de tirer $\gamma^{(1)}$ à partir de $[\gamma | Y, \beta, \Omega, \tau_e]$ à l'aide de $\beta_i^{(1)}, i = 1, \ldots, m, \Omega^{(0)}$ et $\tau_e^{(0)}$; c) de tirer $\Omega^{(1)}$ à partir de $[\Omega | Y, \beta, \gamma, \tau_e]$ à l'aide de $\beta_i^{(1)}, i = 1, \ldots, m$, $\gamma^{(1)}$ et $\tau_e^{(0)}$; d) de tirer $\tau_e^{(1)}$ à partir de $[\tau_e | Y, \beta, \gamma, \Omega]$ à l'aide de $\beta_i^{(1)}, i = 1, \ldots, m, \gamma^{(1)}$ et $\Omega^{(1)}$. Les étapes (a)-(d) permettent de compléter un cycle d'échantillonnage. On exécute un nombre élevé de cycles, disons t, période dite de rodage, jusqu'à la convergence, puis on traite $\{\beta_i^{(t+k)}, i = 1, \ldots, m;\ \gamma^{(t+k)}; \Omega^{(t+k)}; \tau_e^{(t+k)}; k = 1, \ldots, G\}$ comme des échantillons G à partir de la distribution a posteriori composée de $\beta_i, i = 1, \ldots, m, \gamma, \Omega$ et τ_e.

Supposons qu'un échantillon de taille G, de la forme $\{\beta_i^{(k)}, i = 1, \ldots, m; \gamma^{(k)}; \Omega^{(k)}; \tau_e^{(k)}; k = 1, \ldots, G\}$ soit obtenu. Pour obtenir un estimateur de la moyenne a posteriori de β_i, on peut utiliser la moyenne d'échantillon du $\{\beta_i^{(k)}\}$. Puisque β_i comporte une distribution conditionnelle complète de forme fermée, nous pouvons utiliser la moyenne d'échantillon des espérances conditionnelles $\{E[\beta_i | Y, \gamma^{(k)}, \Omega^{(k)}, \tau_e^{(k)}]\}$ afin d'améliorer notre estimation, puisque $E(\beta_i | Y) = E(E(\beta_i | Y, \gamma, \Omega, \tau_e))$, et $\text{Var}(\beta_i | Y) \geq \text{Var}(E(\beta_i | Y, \gamma, \Omega, \tau_e))$. Cette modification se fonde sur le théorème bien connu de Rao-Blackwell, et l'estimateur correspondant est l'estimateur soi-disant Rao-Blackwellisé (Gelfand et Smith 1990, 1991). Nous avons ainsi les deux autres estimateurs ci-dessous pour β_i :

$$\hat{\beta}_i^{(E)} = \frac{1}{G}\sum_{k=1}^G \beta_i^{(k)} \tag{5}$$

et

$$\hat{\beta}_i^{(\text{RB})} = \frac{1}{G}\sum_{k=1}^G E(\beta_i | Y, \gamma^{(k)}, \Omega^{(k)}, \tau_e^{(k)}) = \frac{1}{G}\sum_{k=1}^G (\tau_e^{(k)} X_i^T X_i + \Omega^{(k)})^{-1} (\tau_e^{(k)} X_i^T Y_i + \Omega^{(k)} Z_i \gamma^{(k)}), \tag{6}$$

où $\hat{\beta}_i^{(E)}$ est l'estimateur empirique et $\hat{\beta}_i^{(\text{RB})}$ est l'estimateur Rao-Blackwellisé. $\hat{\beta}_i^{(E)}$ et $\hat{\beta}_i^{(\text{RB})}$ sont tous deux sans biais pour la moyenne a posteriori. Toutefois, $\hat{\beta}_i^{(\text{RB})}$ est meilleur que $\hat{\beta}_i^{(E)}$ pour ce qui est de l'erreur-type de simulation (Gelfand et Smith 1991).

Les estimateurs correspondants pour la moyenne régionale μ_i sont donnés comme suit

$$\hat{\mu}_i^{(E)} = \bar{X}_i^T \hat{\beta}_i^{(E)} \frac{1}{G}\sum_{k=1}^G \bar{X}_i^T \beta_i^{(k)} \tag{7}$$

et

$$\hat{\mu}_i^{(\text{RB})} = \bar{X}_i^T \hat{\beta}_i^{(\text{RB})} = \frac{1}{G}\sum_{k=1}^G \bar{X}_i^T (\tau_e^{(k)} X_i^T X_i + \Omega^{(k)})^{-1} (\tau_e^{(k)} X_i^T Y_i + \Omega^{(k)} Z_i \gamma^{(k)}). \tag{8}$$

Nous nous attendons à ce que $\hat{\mu}_i^{(E)}$ et $\hat{\mu}_i^{(\text{RB})}$ donnent tous deux presque les mêmes estimations ponctuelles. Toutefois, il est intéressant de calculer et de comparer les erreurs-types de simulation de ces deux estimateurs en vue de l'évaluation des effets de la Rao-Blackwellisation (voir la section 3).

Afin d'obtenir la variance a posteriori de μ_i, nous trouvons d'abord la variance a posteriori de β_i, puisque $V(\mu_i | Y) = \bar{X}_i^T V(\beta_i | Y) \bar{X}_i$. À noter que

$$\begin{aligned} V(\beta_i | Y) &= E(V(\beta_i | Y, \gamma, \Omega, \tau_e)) + V(E(\beta_i | Y, \gamma, \Omega, \tau_e)) \\ &= E(V(\beta_i | Y, \gamma, \Omega, \tau_e)) + E(E(\beta_i | Y, \gamma, \Omega, \tau_e)^2) \\ &\quad - [E(E(\beta_i | Y, \gamma, \Omega, \tau_e))]^2. \end{aligned} \tag{9}$$

À l'aide de (9), l'estimateur Rao-Blackwellisé de la variance a posteriori de β_i, noté $\hat{V}(\beta_i)$, s'obtient à l'aide des échantillons de Gibbs $\{\beta_i^{(k)}, i = 1, \ldots, m; \gamma^{(k)}; \Omega^{(k)}; \tau_e^{(k)}; k = 1, \ldots, G\}$; voir l'annexe A1. La variance a posteriori de la moyenne régionale μ_i est alors estimée à l'aide de

$$\hat{V}(\mu_i) = \bar{X}_i^T \hat{V}(\beta_i)\bar{X}_i. \quad (10)$$

On peut appliquer la même procédure d'estimation à la variance de l'erreur d'échantillonnage σ_e^2. Puisque, conditionnellement, σ_e^2 comporte une distribution gamma inverse, l'estimateur Rao-Blackwellisé de la moyenne a posteriori de σ_e^2 s'obtient sous la forme

$$\hat{\sigma}_e^{2(\mathrm{RB})} = \frac{1}{G}\sum_{k=1}^{G}\left[b + \frac{1}{2}\sum_{i=1}^{m}(Y_i - X_i\beta_i^{(k)})^T(Y_i - X_i\beta_i^{(k)})\right] \left(a + \frac{1}{2}\sum_{i=1}^{m} n_i - 1\right)^{-1}. \quad (11)$$

Puisque nous nous intéressons surtout à l'estimation des moyennes régionales, le calcul de la variance d'échantillonnage sert uniquement à la sélection du modèle. On trouvera des détails sur la sélection du modèle à la section 3.2.

2.2 Variance de l'erreur inégale

Dans la pratique, il est plus réaliste d'admettre une variance d'erreur inégale pour les erreurs d'échantillonnage. Soit σ_i^2, la vraie variance de l'erreur d'échantillonnage pour la $i^{\text{ième}}$ région. Une extension directe du modèle 1 mène au modèle bayésien hiérarchique à plusieurs niveaux de la variance d'erreur inégale que voici :

Modèle 2 :

(i) Sous réserve que β_i et σ_i^2, les y_{ij} sont indépendants avec

$$y_{ij}|\beta_i, \sigma_i^2 \sim N(x_{ij}^T\beta_i, \sigma_i^2), \quad (i = 1, \ldots, m;\ j = 1, \ldots, n_i); \quad (12)$$

(ii) Sous réserve que γ et Φ, les β_i sont indépendants avec

$$\beta_i|\gamma, \Phi \sim N_p(Z_i\gamma, \Phi), (i = 1, \ldots, m); \quad (13)$$

(ii) Distributions a priori marginales : $\gamma \sim N_q(0, D)$, $\tau_i \overset{\text{ind}}{\sim} G(a_i, b_i)$ et $\Omega \sim W_p(\alpha, R)$, où $\tau_i = \sigma_i^{-2}$, $\Omega = \Phi^{-1}$ et D, a_i, b_i, α et R connus.

Remarque 2.1 : Le modèle 2 se réduit au modèle 1 lorsque $\sigma_i^2 = \sigma_e^2$ pour tous les i. D'un point de vue hiérarchique de Bayes, l'extension du modèle de la variance d'erreur égale au modèle de la variance d'erreur inégale va de soi. Il n'y a pas non plus de difficulté pour la mise en œuvre de l'échantillonnage de Gibbs.

Remarque 2.2 : Les τ_i sont supposés indépendants et comportent des distributions a priori $G(a_i, b_i)$, où a_i et b_i sont des hyperparamètres connus normalement choisis très petits afin de refléter une vague connaissance des τ_i.

Les distributions conditionnelles complètes pour un échantillonnage de Gibbs dans le cadre du modèle 2 sont données par :

(i) Pour $i = 1, \ldots, m$,

$$[\beta_i|Y, \gamma, \Omega, \tau] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i + \Omega)^{-1} (\tau_i X_i^T Y_i + \Omega Z_i\gamma), (\tau_i X_i^T X_i + \Omega)^{-1})$$

(ii) $$[\gamma|Y, \beta, \Omega, \tau] \sim N_q\left(\left(\sum_{i=1}^{m} Z_i^T\Omega Z_i + D^{-1}\right)^{-1} \left(\sum_{i=1}^{m} Z_i^T\Omega\beta_i\right), \left(\sum_{i=1}^{m} Z_i^T\Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iii) $$[\Omega|Y, \beta, \gamma, \tau] \sim W_p\left(\alpha + m, R + \frac{1}{2}\sum_{i=1}^{m}(\beta_i - Z_i\gamma)(\beta_i - Z_i\gamma)^T\right)$$

(iv) Pour $i = 1, \ldots, m$,

$$[\tau_i|Y, \beta, \gamma, \Omega] \sim G(a_i + \frac{1}{2}n_i, b_i + \frac{1}{2} \times (Y_i - X_i\beta_i)^T(Y_i - X_i\beta_i)).$$

Pour le modèle 2, les estimateurs empiriques des moyennes a posteriori de β_i et μ_i ont la même forme que (5) et (7). Les estimateurs Rao-Blackwellisés $\hat{\beta}_i^{(\mathrm{RB})}$ et $\hat{\mu}_i^{(\mathrm{RB})}$ sont les estimateurs donnés par (6) et (8) avec $\tau_e^{(k)}$ remplacé par $\tau_i^{(k)}$. L'estimateur de la variance a posteriori est $\hat{V}(\mu_i)$ donné par (10) avec $\tau_e^{(k)}$ remplacé par $\tau_i^{(k)}$.

En vue de la sélection du modèle et de la comparaison des modèles, nous trouvons également l'estimateur Rao-Blackwellisé de la moyenne a posteriori de σ_i^2 dans le cadre du modèle 2 de la forme

$$\hat{\sigma}_i^{2(\mathrm{RB})} = \frac{1}{G}\sum_{k=1}^{G}\left[b_i + \frac{1}{2}(Y_i - X_i\beta_i^{(k)})^T(Y_i - X_i\beta_i^{(k)})\right] \times\left(a_i + \frac{1}{2}n_i - 1\right)^{-1}. \quad (14)$$

2.3 Variance de l'erreur aléatoire

Dans le modèle 2, nous avons supposé une variance d'erreur inégale pour les erreurs d'échantillonnage. Kleffe et Rao (1992) ont utilisé un modèle simple de la variance d'erreur aléatoire afin d'obtenir les meilleurs prédicteurs linéaires sans biais pour des moyennes régionales. Dans la présente section, nous étendons leur modèle au cas comportant plusieurs niveaux. Nous supposons des modèles d'effets aléatoires sur les coefficient de régression β_i aussi bien que sur les variances de l'erreur d'échantillonnage σ_i^2, ce qui mène au modèle 3 donné ci-dessous.

Modèle 3 :

(i) Comme pour le modèle 2;

(ii) Comme pour le modèle 2;

(iii) Sous réserve de η et de λ, les τ_i sont indépendants avec

$$\tau_i|\eta, \lambda \overset{\text{iid}}{\sim} G(\eta, \lambda), \quad (15)$$

où $\tau_i = \sigma_i^{-2}$;

(iv) Distributions a priori marginales : $\gamma \sim N_q(0, D)$, $\Omega \sim W_p(\alpha, R)$, $\eta \sim U^+$ et $\lambda \sim U^+$, où U^+ désigne une distribution uniforme sur un sous-ensemble de R^+ avec une longueur grande mais finie, D, α et R connus.

Remarque 3.1 : Dans le modèle 3, nous supposons que les τ_i sont des variables aléatoires gamma indépendantes et distribuées de façon identique avec des hyperparamètres η et λ inconnus. Nous avons ainsi des modèles de population pour les coefficients de régression β_i aussi bien que pour la variance d'échantillonnage σ_i^2. Dans les modèles 1 et 2, nous avons considéré la modélisation de β_i seulement et nous avons supposé de vagues distributions a priori appropriées sur σ_e^2 ou σ_i^2.

Remarque 3.2 : L'hypothèse (iii) n'est peut-être pas un bon modèle de population pour tous les τ_i. Comme autre possibilité, nous pouvons modéliser τ_i de façon plus réaliste, comme dans le cas de β_i, en précisant un modèle de régression pour le logarithme de τ_i. Cela peut exiger des renseignements auxiliaires liés à τ_i. Dans l'analyse des données de la section 3, toutefois, nous utilisons simplement $G(\eta, \lambda)$ comme modèle de population pour τ_i. Il n'est généralement pas facile de modéliser les variances d'échantillonnage lorsqu'elles sont inconnues.

Les distributions conditionnelles complètes pour un échantillonnage de Gibbs dans le cadre du modèle 3 sont données par :

(i) Pour $i = 1, \ldots, m$,

$$[\beta_i | Y, \tau, \gamma, \Omega, \eta, \lambda] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i + \Omega)^{-1} (\tau_i X_i^T Y_i + \Omega Z_i \gamma), (\tau_i X_i^T X_i + \Omega)^{-1})$$

(ii) Pour $i = 1, \ldots, m$,

$$[\tau_i | Y, \beta, \gamma, \Omega, \eta, \lambda] \overset{\text{ind}}{\sim} G\left(\eta + \frac{n_i}{2}, \frac{1}{2}(Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i) + \lambda\right)$$

(iii) $$[\gamma | Y, \beta, \tau, \eta, \lambda] \sim N_q\left(\left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1} \left(\sum_{i=1}^m Z_i^T \Omega \beta_i\right), \left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iv) $$[\Omega | Y, \beta, \sigma^2, \gamma, \eta, \lambda] \sim W_p\left(\alpha + m, R + \frac{1}{2}\sum_{i=1}^m (\beta_i - Z_i\gamma)(\beta_i - Z_i\gamma)^T\right)$$

(v) $$[\eta | Y, \beta, \tau, \gamma, \Omega, \lambda] \propto [\Gamma(\eta)]^{-m} \lambda^{m\eta} \left(\prod_{i=1}^m \tau_i\right)^{\eta}$$

(vi) $$[\lambda | Y, \beta, \tau, \gamma, \Omega, \eta] \sim G\left(m\eta + 1, \sum_{i=1}^m \tau_i\right).$$

Pour le modèle 3, les estimateurs a posteriori de β_i et μ_i ont les mêmes formes que celles qui ont été données pour le modèle 2. Dans le cadre du modèle 3, l'estimateur Rao-Blackwellisé de la moyenne a posteriori de σ_i^2 est donné par

$$\hat{\sigma}_i^{2(\mathrm{RB})} = \frac{1}{G}\sum_{k=1}^G [\lambda^{(k)} + \frac{1}{2}(Y_i - X_i\beta_i^{(k)})^T \times (Y_i - X_i\beta_i^{(k)})]\left(\eta^{(k)} + \frac{1}{2}n_i - 1\right)^{-1}. \quad (16)$$

Dans le cadre du modèle 3, $[\eta | Y, \beta, \tau, \gamma, \Omega, \lambda]$ est connu jusqu'à une constante multiplicatrice seulement. Toutefois, puisque $[\eta | Y, \beta, \tau, \gamma, \Omega, \lambda]$ est une fonction concave logarithmique de η (voir l'annexe A2), il est possible d'utiliser la méthode d'échantillonnage à rejet adapté de Gilks, Best et Tan (1995) dans l'échantillonneur de Gibbs afin de produire des échantillons à partir de la distribution conditionnelle $[\eta | Y, \beta, \tau, \gamma, \Omega, \lambda]$.

3. Analyse des données

3.1 Description des données et des modèles

Suivant Holt et Moura (1993) et Moura et Holt (1999), nous avons considéré l'estimation du revenu des ménages de comtés (régions) du Brésil. Les données originales de Holt et Moura comportent 140 régions, les unités d'échantillonnage étant tirées de chaque région par échantillonnage aléatoire simple. La méthode hiérarchique de Bayes n'exige pas que le nombre de régions soit grand, contrairement à la méthode du MPLSB empirique, pour l'obtention des erreurs-types. Par conséquent, nous avons utilisé uniquement une faible partie de l'ensemble original de données dans notre analyse, à simple titre d'illustration. Notre ensemble de données contient un sous-ensemble de 10 régions comportant 28 unités d'échantillonnage obtenues par échantillonnage aléatoire simple dans chaque région.

Soit y_{ij}, le revenu du $j^{\text{ième}}$ ménage dans la $i^{\text{ième}}$ région. Il y a deux variables auxiliaires au niveau de l'unité, c'est-à-dire x_1 et x_2, où x_1 désigne le nombre de pièces dans un ménage et x_2 désigne le niveau de scolarité atteint par le chef du ménage. Le modèle d'échantillonnage est donné par

$$y_{ij} = x_{ij}^T \beta_i + e_{ij} = \beta_{0i} + x_{1ij}\beta_{1i} + x_{2ij}\beta_{2i} + e_{ij}, \quad (17)$$

où x_{1ij} désigne le nombre de pièces dans le $j^{\text{ième}}$ ménage de la région i et x_{2ij} désigne le niveau de scolarité correspondant atteint par le chef du ménage. Les valeurs de x_{1ij} et de x_{2ij} sont centrées autour de leurs moyennes d'échantillon global respectives et e_{ij} est la variable de l'erreur d'échantillonnage, sa distribution étant précisée par les trois modèles de la variance d'erreur discutées à la section 2.

Dans le modèle d'échantillonnage (17), β_i est le coefficient de régression aléatoire correspondant à la $i^{\text{ième}}$ région et se laisse modéliser comme

$$\beta_{0i} = \gamma_0 + v_{0i}, \beta_{1i} = \gamma_{10} + \gamma_{11} z_i + v_{1i}, \beta_{2i} = \gamma_{20} + \gamma_{21} z_i + v_{2i},$$

où $\gamma = (\gamma_0, \gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21})^T$ est le vecteur inconnu de paramètres de régression fixes, $v_i = (v_{0i}, v_{1i}, v_{2i})^T$ est le vecteur d'effets aléatoires de la $i^{\text{ième}}$ région distribué comme

$v_i \sim N_3(0, \Phi)$, et z_i est une variable au niveau de la région définie comme le nombre moyen de voitures par ménage dans chaque région. La valeur de z_i est également centrée autour de sa moyenne d'échantillon globale.

Nous avons utilisé les trois modèles discutés à la section 2 pour notre analyse des données. De vagues distributions a priori appropriées sur des paramètres inconnus sont spécifiées comme suit : $\gamma \sim N_5(0, D)$, où $D = \text{diag}(10^4, 10^4, 10^4, 10^4, 10^4)$, donc $\gamma_0, \gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21}$ sont supposées être des variables normales indépendantes de moyenne 0 et d'écart-type 100, de sorte qu'un intervalle a priori de 95% correspond à ±200 environ, et la distribution a priori est uniforme localement pour la région appuyée par la vraisemblance. Comme autre possibilité, une distribution a priori uniforme sur un intervalle convenablement large pourrait être donnée, par exemple $U(-200, 200)$. Une distribution a priori de Wishart $W_3(\alpha, R)$ est spécifiée pour la matrice de covariances inverse $\Omega = \Phi^{-1}$. Pour représenter une vague connaissance a priori, nous avons choisi les degrés de liberté α pour que cette distribution soit aussi petite que possible, $\alpha = 3$, le rang de Ω (Spiegelhalter, Thomas, Best et Gilks 1996). La matrice scalaire *R* est décrite avec des éléments diagonaux égaux à 1 et des éléments non diagonaux égaux à 0,001, ce qui représente notre estimation préliminaire pour l'ordre de grandeur de la matrice de covariances. Pour les modèles 1 et 2, une distribution a priori gamma G(0,001, 0,001) est supposée pour τ_e et les τ_i. Pour le modèle 3, $\tau_i \sim G(\eta, \lambda)$, et on suppose que η et λ sont distribués indépendamment comme U(0, 10 000), c'est-à-dire la distribution uniforme sur un grand intervalle. Nous nous attendons à ce que les vagues distributions a priori appropriées sur les hyperparamètres se rapprochent raisonnablement bien des distributions a priori uniformes et qu'elles exercent ainsi un effet minimal sur l'estimation a posteriori.

Nous avons mis en œuvre l'échantillonneur de Gibbs pour les trois modèles à l'aide du programme BUGS (Spiegelhalter et coll. 1996) et de la fonction CODA Splus (Best, Cowles et Vines 1996) pour l'évaluation de la convergence. Le programme BUGS permet de construire les distributions conditionnelles complètes nécessaires et d'exécuter l'échantillonnage de Gibbs pourvu que nos modèles soient décrits à l'aide du langage BUGS. Les distributions a priori et les valeurs initiales des paramètres doivent être décrites dans le programme. Pour chaque modèle, l'échantillonneur de Gibbs a d'abord été exécuté au cours d'une période dite de rodage de 2 000 itérations, puis 5 000 autres itérations ont été exécutées et conservées en vue de l'analyse et de l'estimation du modèle.

Nous cherchons à estimer la moyenne régionale $\mu_i = \bar{X}_i^T \beta_i = \beta_{0i} + \bar{X}_{1i}\beta_{1i} + \bar{X}_{2i}\beta_{2i}$, où $\bar{X}_{1i}$ et $\bar{X}_{2i}$ sont les moyennes de population de la $i^{\text{ième}}$ région des variables auxiliaires x_1 et x_2, respectivement. Pour ce faire, nous allons d'abord choisir un modèle optimal pour l'ensemble de données, et nous allons présenter les estimations modélisées pour les moyennes régionales fondées sur le modèle choisi.

3.2 Sélection du modèle

À la section 2, nous avons proposé trois modèles en fonction d'hypothèses différentes quant aux variances d'échantillonnage. Afin de déterminer quel modèle ajuste les données, nous avons d'abord obtenu les estimations a posteriori des variances d'échantillonnage dans le cadre des trois modèles. Nous avons également calculé les estimations des moindres carrés ordinaires (MCO) des variances d'échantillonnage au sein de chaque région à l'aide des données propres à la région seulement. Le tableau 1 indique les estimations Rao-Blackwellisées des variances d'échantillonnage dans le cadre des trois modèles de même que les estimations MCO.

Tableau 1
Variances estimatives de l'erreur d'échantillonnage

Région	MCO	Modèle 1	Modèle 2	Modèle 3
1	38,17	76,86	40,18	63,60
2	31,75	76,86	34,24	62,13
3	81,26	76,86	94,77	79,58
4	48,73	76,86	52,01	67,27
5	115,98	76,86	121,65	87,70
6	90,74	76,86	94,35	79,78
7	101,67	76,86	101,67	82,14
8	135,65	76,86	159,94	97,96
9	59,10	76,86	63,37	70,57
10	62,86	76,86	65,72	71,22

Dans le tableau 1, les estimations MCO indiquent de fortes variations parmi les 10 régions. Le modèle 1 suppose une variance de l'erreur égale σ_e^2 pour toutes les régions, et on estime σ_e^2 par $\hat{\sigma}_e^{2(\text{RB})} = 76,86$, valeur beaucoup plus faible que les estimations MCO pour certaines régions. Le modèle 2 suppose une variance de l'erreur inégale σ_i^2 d'une région à l'autre. Dans le cadre du modèle 2, les variances de l'erreur estimée $\hat{\sigma}_i^{2(\text{RB})}$ montrent dans une certaine mesure l'aspect des régions; les $\hat{\sigma}_i^{2(\text{RB})}$ concordent avec la tendance des estimations MCO. Le résultat le plus remarquable est $\hat{\sigma}_5^{2(\text{RB})} = 121,65$ et $\hat{\sigma}_8^{2(\text{RB})} = 159,94$, indiquant de plus fortes variations au sein des régions 5 et 8. Le modèle 3 suppose que les σ_i^2 sont des variables aléatoires distribuées comme $G(\eta, \lambda)$. Dans le cadre du modèle 3, toutes les $\hat{\sigma}_i^{2(\text{RB})}$ tendent à être égales à $\hat{\sigma}_e^{2(\text{RB})} = 76,86$ et y sont déplacées. Les résultats du tableau 1 indiquent que le modèle 2, c'est-à-dire le modèle de la variance de l'erreur inégale, pourrait être le modèle optimal pour notre ensemble de données. En guise d'approfondissement, nous présentons maintenant une étude de validation croisée en vue de la sélection d'un modèle d'ajustement optimal.

Afin d'examiner comment les données appuient chaque modèle, nous avons calculé la densité prédictive de validation croisée pour chaque donnée simple y_{ij}. La densité de validation croisée pour y_{ij} est la densité conditionnelle $f(y_{ij}|Y_{(ij)})$, où $Y_{(ij)}$ désigne toutes les données sauf y_{ij}. Nous

avons examiné la valeur de $f(y_{ij}|Y_{(ij)})$ pour la donnée simple observée, la soi-disant ordonnée prédictive conditionnelle (CPO) pour chacun des trois modèles. Nous avons

$$\text{CPO}_{ij} = f(y_{ij,\text{obs}}|Y_{(ij),\text{obs}}),$$

où $y_{ij,\text{obs}}$ désigne la donnée simple observée. Puisque les CPO ne sont que les vraisemblances observées, les modèles comportant des CPO plus élevées fournissent un meilleur ajustement pour les données observées. À l'aide de la sortie de l'échantillonneur de Gibbs, nous pouvons calculer les CPO pour toutes les données simples. Ainsi, dans le cadre du modèle 1, nous avons

$$\begin{aligned} f(y_{ij}|Y_{(ij)}) &= \frac{f(Y)}{f(Y_{(ij)})} \\ &= \frac{1}{\int \frac{f(Y_{(ij)},\beta_i,\sigma_e^2)}{f(Y,\beta_i,\sigma_e^2)} \cdot f(\beta_i,\sigma_e^2|Y)d\beta_i d\sigma_e^2} \\ &= \frac{1}{\int \frac{1}{f(y_{ij}|Y_{(ij)},\beta_i,\sigma_e^2)} \cdot f(\beta_i,\sigma_e^2|Y)d\beta_i d\sigma_e^2}. \end{aligned}$$

Nous notons maintenant que les y_{ij} sont conditionnellement indépendantes, donc $f(y_{ij}|Y_{(ij)},\beta_i,\sigma_e^2) = f(y_{ij}|\beta_i,\sigma_e^2)$, et les valeurs CPO sont calculées comme suit :

$$\widehat{\text{CPO}}_{ij} = \frac{1}{\frac{1}{G}\sum_{k=1}^{G} \frac{1}{f(y_{ij,\text{obs}}|\beta_i^{(k)},\sigma_e^{2(k)})}} \qquad (18)$$

où $f(y_{ij}|\beta_i,\sigma_e^2)$ est la fonction de densité normale donnée par (3). Pour les modèles 2 et 3, on calcule les CPO avec $\sigma_e^{2(k)}$ remplacée par $\sigma_i^{2(k)}$ en (18). On trouvera une discussion plus détaillée dans Gelfand (1995).

Nous présentons un tracé CPO pour les trois modèles à la figure 1. Le modèle 2 est clairement le meilleur, puisqu'une majorité de valeurs CPO pour le modèle 2 est appréciablement plus grande que celles des modèles 1 et 3. Le modèle 3 est légèrement meilleur que le modèle 1 pour ce qui est des valeurs CPO. Il y a aussi de petites valeurs CPO pour les trois modèles, ce qui indique que les hypothèses de nos modèles ne sont peut-être pas bien respectées par notre ensemble de données.

Compte tenu des estimations de la variance d'échantillonnage que l'on trouve dans le tableau 1 et le tracé CPO, nous concluons que le modèle 2 est le meilleur modèle d'ajustement pour nos données. Par conséquent, nous avons utilisé le modèle 2 pour obtenir des estimations modélisées de moyennes régionales et des erreurs-types a posteriori connexes.

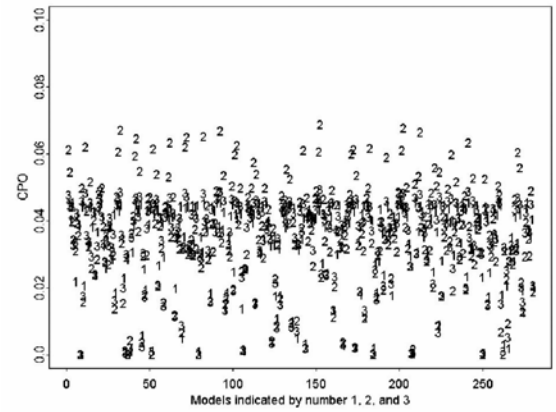

Figure 1. Sélection du modèle : trace de comparaison CPO.

3.3 Résultat de l'estimation

Nous présentons maintenant les estimations des moyennes régionales en fonction du modèle 2 seulement. Le tableau 2 montre les moyennes régionales a posteriori estimées et les erreurs-types a posteriori correspondantes. Notre étude a montré que l'estimateur empirique $\hat{\mu}_i^{(E)}$ et l'estimateur Rao-Blackwellisé $\hat{\mu}_i^{(\text{RB})}$ fournissent presque les mêmes estimations ponctuelles, et nous avons donc signalé uniquement les estimations obtenues à l'aide de $\hat{\mu}_i^{(\text{RB})}$. Pour la comparaison, nous avons calculé les estimations directes (moyennes d'échantillon) et les erreurs-types directes correspondantes pour les 10 régions. Le tableau 2 montre clairement que les estimations modélisées sont appréciablement plus efficaces que les estimations directes. Les erreurs-types a posteriori sont beaucoup plus petites que les erreurs-types directes.

Tableau 2
Estimations de moyennes régionales

Région	$\bar{y}_i$	é.-t.	$\hat{\mu}_i^{(\text{RB})}$	é.-t.
1	11,08	9,53	10,23	0,81
2	7,91	6,82	9,84	0,85
3	13,48	14,15	13,01	1,08
4	6,53	8,01	10,95	1,11
5	19,52	14,96	17,87	1,57
6	11,21	11,38	10,21	0,93
7	8,72	11,24	9,58	0,97
8	12,81	13,99	10,30	1,19
9	10,18	8,76	11,34	1,01
10	10,01	11,30	9,79	0,87

Afin d'étudier les effets de la Rao-Blackwellisation, nous avons calculé les erreurs-types de simulation de $\hat{\mu}_i^{(E)}$ et $\hat{\mu}_i^{(\text{RB})}$, qui sont respectivement les erreurs-types d'échantillon de $\{\bar{X}_i^T \hat{\beta}_i^{(k)}\}$ et $\{\bar{X}_i^T E[\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)}]\}$. Le tableau 3 montre les erreurs-types de simulation. Le tableau 3 indique clairement que l'estimateur Rao-Blackwellisé $\hat{\mu}_i^{(\text{RB})}$ comporte une erreur-type de simulation beaucoup plus

faible que l'estimateur empirique $\hat{\mu}_i^{(E)}$ pour toutes les régions. Dans tous les cas, l'erreur-type de $\hat{\mu}_i^{(\mathrm{RB})}$ représente de 50% à 75% environ de l'erreur-type de $\hat{\mu}_i^{(E)}$, ce qui confirme l'avantage de la Rao-Blackwellisation. Ainsi, $\hat{\mu}_i^{(\mathrm{RB})}$ est plus stable que $\hat{\mu}_i^{(E)}$ lorsqu'on l'utilise pour obtenir des estimations ponctuelles pour les moyennes a posteriori dans une analyse bayésienne computationnelle. Il convient de mentionner que l'erreur-type de simulation de $\hat{\mu}_i^{(E)}$ est également un estimateur de l'erreur-type a posteriori. Ainsi, l'erreur-type de simulation de $\hat{\mu}_i^{(E)}$ au tableau 3 est presque identique à l'erreur-type estimée $\hat{\mu}_i^{(\mathrm{RB})}$ au tableau 2

Tableau 3
Erreurs-types de simulation

Région	$\hat{\mu}_i^{(E)}$	$\hat{\mu}_i^{(\mathrm{RB})}$
1	0,817	0,506
2	0,862	0,498
3	1,090	0,548
4	1,101	0,604
5	1,583	0,878
6	0,930	0,481
7	0,978	0,480
8	1,208	0,842
9	0,997	0,524
10	0,869	0,513

3.4 Analyse de sensibilité

Dans le modèle 2, les variances d'erreur $\tau_i = \sigma_i^{-2}$ sont supposées indépendantes avec des distributions a priori $\mathrm{G}(a_i, b_i)$ ou σ_i^2 avec la distribution gamma inverse $\mathrm{IG}(a_i, b_i)$, où a_i et b_i sont des valeurs connues choisies de façon à refléter nos connaissances a priori de σ_i^2. Dans la pratique, il est toujours difficile d'obtenir des informations exactes au sujet des variances d'échantillonnage. De plus, à mesure que le nombre de régions *m* augmente, le nombre de composantes des variances σ_i^2 augmente. Nous nous intéressons aux effets possibles du choix des distributions a priori sur les σ_i^2 ; en particulier, nous aimerions évaluer la sensibilité des moyennes a posteriori au choix des distributions a priori sur les variances d'échantillonnage σ_i^2. Dans notre analyse des données, a_i et b_i ont été choisies égales à 0,001. Ainsi, nous avons utilisé des distributions a priori appropriées avec de très faibles valeurs paramétriques pour les composantes de variance de façon à refléter nos connaissances vagues de σ_i^2. Afin de vérifier la sensibilité des estimations a posteriori au choix de a_i et b_i dans le cadre du modèle 2, nous avons fixé six valeurs différentes de $a_i = b_i$ c'est-à-dire 0,0001, 0,001, 0,01, 0,1, 1 et 10. Puisque

$$\{\tau_i \mid Y, \beta, \gamma, \Omega] \sim G(a_i + \frac{1}{2}n_i, b_i + \frac{1}{2}(Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i)), \qquad (19)$$

les effets d'échantillon $n_i/2$ et $(Y_i - X_i\beta_i)^T(Y_i - X_i\beta_i)/2$ dominent l'information a priori a_i et b_i lorsque a_i et b_i sont petites. Ainsi IG(0,0001, 0,0001), IG(0,001, 0,001) et IG(0,01, 0,01) peuvent être considérées comme des distributions a priori non informatives tandis que IG(1, 1) et IG(10, 10) peuvent être considérées comme des distributions a priori informatives. Le tableau 4 présente des moyennes a posteriori dans le cadre du modèle 2 à l'aide des différentes distributions a priori sur σ_i^2, et le tableau 5 montre les variances a posteriori correspondantes.

Tableau 4
Comparaison de moyennes régionales estimées

Région	$\mathrm{IG}(a_i, b_i),\ a_i = b_i$					
	0,001	0,001	0,01	0,1	1	10
1	10,23	10,23	10,23	10,24	10,25	10,37
2	9,84	9,84	9,84	9,83	9,82	9,62
3	13,00	13,00	13,01	13,01	13,07	13,09
4	10,95	10,95	10,95	10,95	10,94	10,61
5	17,86	17,87	17,85	17,76	17,78	18,27
6	10,21	10,21	10,21	10,21	10,25	10,28
7	9,58	9,58	9,59	9,58	9,63	9,57
8	10,29	10,30	10,30	10,26	10,37	10,86
9	11,34	11,34	11,35	11,32	11,32	11,23
10	9,79	9,79	9,80	9,79	9,82	9,92

Le tableau 4 indique clairement que les estimations de moyennes régionales sont très stables; elles ne sont pas sensibles au choix de a_i et b_i. Toutefois, comme le montre le tableau 5, les variances a posteriori diminuent à mesure que les distributions a priori sur σ_i^2 deviennent plus informatives, et mènent à de plus petits coefficients de variation (cv), Cela indique que nous pouvons améliorer les résultats des estimations pour les régions en ce qui concerne le cv si nous avons plus de renseignements a priori sur les variances de l'erreur d'échantillonnage, Dans notre étude, nous avons considéré uniquement le cas $a_i = b_i$. Une étude plus approfondie comporterait différentes combinaisons de a_i et b_i.

Tableau 5
Comparaison de variances a posteriori de l'erreur d'échantillonnage estimée

Région	$\mathrm{IG}(a_i, b_i),\ a_i = b_i$					
	0,0001	0,001	0,01	0,1	1	10
1	0,658	0,658	0,658	0,656	0,653	0,499
2	0,724	0,724	0,724	0,711	0,684	0,462
3	1,167	1,167	1,167	1,161	1,152	0,917
4	1,220	1,220	1,218	1,217	1,202	0,919
5	2,455	2,455	2,454	2,462	2,139	1,335
6	0,871	0,870	0,870	0,830	0,826	0,699
7	0,933	0,933	0,931	0,930	0,914	0,779
8	1,418	1,417	1,418	1,375	1,351	1,337
9	1,015	1,014	1,014	1,011	0,975	0,790
10	0,760	0,760	0,760	0,750	0,745	0,613

Le tableau 6 présente les estimations a posteriori de σ_i^2 à l'aide des différentes distributions a priori sur σ_i^2. Comme l'indique le tableau 6, lorsque a_i et b_i sont petites (<= 0,01), il n'existe pratiquement pas de différence entre les estimations. À mesure que a_i et b_i augmentent, les estimations $\hat{\sigma}_i^{2(\mathrm{RB})}$ deviennent plus petites. Toutefois, en présence de solides informations a priori sur a_i et b_i, par exemple $a_i = b_i = 10$, les estimations a posteriori de σ_i^2 sont appréciablement différentes de celles qui sont obtenues pour des distributions a priori non informatives.

Tableau 6
Comparaison de variances de l'erreur d'échantillonnage estimée

Région	IG(a_i, b_i), $a_i = b_i$					
	0,0001	0,001	0,01	0,1	1	10
1	40,09	40,10	40,05	39,64	37,14	22,29
2	34,19	34,18	34,17	33,97	31,74	19,05
3	94,48	94,49	94,42	93,76	86,73	50,60
4	52,08	52,08	52,04	51,63	48,21	28,82
5	121,60	121,70	121,60	121,40	113,70	66,75
6	94,03	94,03	93,83	92,96	87,21	52,90
7	102,30	102,30	102,20	101,40	94,85	57,58
8	160,10	160,00	159,90	159,10	147,60	86,61
9	63,46	63,46	63,38	62,99	58,46	34,85
10	65,88	65,87	65,89	65,40	60,76	36,60

4. Conclusion

Dans le présent exposé, nous avons présenté des méthodes bayésiennes hiérarchiques pour l'estimation regionale, à l'aide de modèles à plusieurs niveaux. Il n'est clairement pas facile de fournir un modèle approprié pour toutes les régions de façon à obtenir des résultats satisfaisants, même si des méthodes bayésiennes de type MCCM (Monte Carlo à chaîne markovienne) comme l'échantillonnage de Gibbs nous permettent d'ajuster les données à l'aide de modèles bayésiens d'une complexité virtuellement illimitée. La taille et l'homogénéité des régions et la présence de solides informations auxiliaires auront un effet sur le résultat final. Des modèles appropriés dans certaines situations pourront ne pas se prêter à d'autres situations. La méthode bayésienne hiérarchique comporte également des limitations, par exemple le choix de distributions a priori sur les paramètres de modèle et certaines questions d'échantillonnage liées à la méthode d'échantillonnage de Gibbs. Néanmoins, la méthode bayésienne hiérarchique générale s'applique à tout un choix de situations pour l'estimation de paramètres régionaux. Le choix du modèle est un aspect important de l'analyse bayésienne hiérarchique. Il est également important de comparer la méthode bayésienne hiérarchique à d'autres méthodes largement utilisées pour l'estimation régionale, par exemple la méthode bayésienne empirique et la méthode du meilleur prédicteur linéaire sans biais empirique. Les travaux se poursuivent afin d'y intégrer les poids de plan d'enquête, en conformité avec You et Rao (1999).

Remerciements

Nous tenons à remercier deux examinateurs et le rédacteur de leurs remarques et suggestions utiles. Nous tenons également à remercier le professeur F. Moura de l'Université fédérale de Rio de Janeiro (Brésil) d'avoir fourni l'ensemble de données utilisé à la section 3. Nos travaux ont été appuyés en partie par une subvention du Conseil de recherches en sciences naturelles et en génie du Canada.

Annexe

A1 :

L'estimateur Rao-Blackwellisé de la variance a posteriori de β_i est donné par :

$$\hat{V}(\beta_i) = \frac{1}{G}\sum_{k=1}^{G} V(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})$$

$$+\frac{1}{G}\sum_{k=1}^{G}[E(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})]^2$$

$$-\left[\frac{1}{G}\sum_{k=1}^{G} E(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})\right]^2$$

$$=\frac{1}{G}\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}$$

$$+\frac{1}{G}\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})$$

$$\times(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})^T(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}$$

$$-\frac{1}{G^2}\left[\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})\right]$$

$$\times\left[\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})\right]^T.$$

A2 :

Lemme : $[\eta|Y,\beta,\tau,\gamma,\Omega,\lambda]$ est une fonction concave logarithmique de η.

Preuve : Soit $h(\eta) = \log[\eta|Y,\beta,\tau,\gamma,\Omega,\lambda]$. Il est suffisant de montrer que

$$\frac{\partial^2 h(\eta)}{\partial^2\eta} \le 0.$$

Manifestement,

$$\frac{\partial h(\eta)}{\partial\eta} = -m\frac{\Gamma'(\eta)}{\Gamma(\eta)} + m\log(\lambda) + \log\left(\prod_{i=1}^{m}\tau_i\right).$$

Soit $\psi(\eta) = \Gamma'(\eta)/\Gamma(\eta)$, nous avons alors

$$\frac{\partial^2 h(\eta)}{\partial^2 \eta} = -m\psi'(\eta) \leq 0$$

puisque $m > 0$ et $\psi'(\eta)$ est positif sur $(0, \infty)$ (Temme 1994, 54-55).

Bibliographie

Battese, G.E., Harter, R.M. et Fuller, W.A. (1988). An error components model for prediction of county crop area using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Best, N., Cowles, M.K. et Vines, K. (1996). CODA, *Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Version 0.30. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR.

Gelfand, A.E. (1995). Model determination using sampling-based methods. Dans *Markov Chain Monte Carlo in Practice* (Éds., W.R. Gilks, S. Richardson, et D.J. Spiegelhalter), 145-161. London : Chapman and Hall.

Gelfand, A.E., et Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelfand, A.E., et Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expections. *Communications In Statistics - Theory and Methods*, 20, 1747-1766.

Ghosh, M., et Rao, J.N.K. (1994). Small area estimation: An appraisal (avec discussion). *Statistical Science*, 9, 55-93.

Gilks, W.R., Best, N.G. et Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of Applied Statistics*, 44, 455-472.

Hobert, J.P., et Cassella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91, 1461-1473.

Holt, D., et Moura, F. (1993). Small area estimation using multi-level models. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 21-30.

Kleffe, J., et Rao, J.N.K. (1992). Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model. *Journal of Multivariate Analysis*, 43, 1-15.

Moura, F., et Holt, D. (1999). Production d'estimations régionales à partir de modèles multiniveau. *Techniques d'enquête*, 25, 81-89.

Prasad, N.G.N., et Rao, J.N.K. (1990). The estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Rao, J.N.K. (1999). Quelques progrès récents concernant l'estimation régionale fondée sur un modèle. *Techniques d'enquête*, 25, 199-212.

Spiegelhalter, D., Thomas, A., Best, N. et Gilks, W. (1996). BUGS 0.5, *Bayesian Inference Using Gibbs Sampling Manual*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR.

Temme, N.M. (1994). *Special Functions*: *An Introduction to the Classical Functions of Mathematical Physics*. New York : John Wiley & Sons, Inc.

You, Y., et Rao, J.N.K. (1999). Pseudo hierarchical Bayes small area estimation using sampling weights. *Proceedings of the Survey Methods Section*, *Statistical Society of Canada*, 117-122.