Nº 12-001-X au catalogue
ISSN 1712-5685

Techniques d'enquête

Estimation bayésienne robuste sur petits domaines

par Malay Ghosh, Jiyoun Myung et Fernando A.S. Moura

Date de diffusion : le 21 juin 2018

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence (p<0,05)

Techniques d'enquête, juin 2018
Vol. 44, nº 1, pp. 109-124
Statistique Canada, nº 12-001-X au catalogue

Estimation bayésienne robuste sur petits domaines

Malay Ghosh, Jiyoun Myung et Fernando A.S. Moura[1]

Résumé

Les modèles pour petits domaines conçus pour traiter les données au niveau du domaine reposent habituellement sur l'hypothèse de normalité des effets aléatoires. Cette hypothèse ne tient pas toujours. L'article présente un nouveau modèle pour petits domaines dont les effets aléatoires suivent une loi *t*. En outre, la modélisation conjointe des moyennes et des variances de petit domaine est examinée. Il est montré que cette approche donne de meilleurs résultats que les autres méthodes.

Mots-clés : Modèle à effets aléatoires; loi *t* de Student; lois a priori non subjectives; MCMC; échantillonnage de Gibbs; algorithme de Metropolis-Hastings.

1 Introduction

Depuis près de quatre décennies, l'article classique de Fay et Herriot (1979) est devenu une pierre angulaire de la recherche au sujet de l'estimation sur petits domaines. Le modèle de Fay-Herriot est un modèle à effets aléatoires qui s'appuie sur une hypothèse de normalité des effets aléatoires, ainsi que des erreurs. En outre, les variances d'erreur sont supposées connues. Cette dernière hypothèse est presque impérative en raison d'un problème d'identifiabilité. Si l'on dispose uniquement d'estimations sur petits domaines directes au niveau du domaine et que l'on ne possède pas de microdonnées, toute modélisation efficace des variances d'erreur est quasiment impossible.

De vaillantes tentatives en vue de remédier à ce problème ont été faites par W.R. Bell et ses collègues au *US Census Bureau* (Bell et Huang, 2006; Bell, 2008) pour le traitement de certaines données de recensement, mais des questions persistent quant à l'application universelle de leur approche. En outre, l'absence de microdonnées pour les utilisateurs secondaires tient surtout à des problèmes de confidentialité, principalement pour les organismes fédéraux. Quand des microdonnées deviennent disponibles, les modèles au niveau de l'unité conviennent mieux que ceux au niveau du domaine. L'article souvent cité de Battese, Harter et Fuller (1988) en est un exemple classique. Toutefois, les modèles au niveau du domaine sont d'usage très répandu en raison de la simplicité de leur mise en œuvre dans un contexte d'enquête complexe comparativement aux modèles au niveau de l'unité.

À mesure que s'est développée la branche de l'estimation sur petits domaines et que davantage de données ont été analysées, les chercheurs ont constaté que les hypothèses de normalité et de connaissance des variances d'erreur étaient inappropriées. Comme il est mentionné au paragraphe précédent, la dernière hypothèse est difficile à corriger sans aucune information supplémentaire. L'une des premières tentatives à cet égard est attribuable à Lahiri et Rao (1995) qui ont remplacé l'hypothèse de normalité des effets aléatoires par la finitude de leur huitième moment. Datta et Lahiri (1995) ont considéré pour les effets aléatoires un mélange général de lois normales incluant la loi t. Certains articles font entièrement abstraction de la normalité, mais maintiennent la linéarité du modèle, et font appel à des estimateurs

1. Malay Ghosh, Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, P.O. Box 118545, Gainesville, Floride, États-Unis, 32611. Courriel : ghoshm@stat.ufl.edu; Jiyoun Myung, Department of Statistics et Applied Probability, University of California, Santa Barbara, 5513 South Hall, Santa Barbara, Californie, États-Unis, 93106. Courriel : myung@pstat.ucsb.edu; Fernando A.S. Moura, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, CEP: 21941-909, RJ, Brésil. Courriel : fmoura@im.ufrj.br.

ANOVA des variances. On peut citer Butar et Lahiri (2002), ainsi que Jiang, Lahiri et Wan (2002) qui ont calculé les mesures d'incertitude correspondantes par les méthodes du jackknife ou du bootstrap. Bell et Huang (2006) ont utilisé des lois t pour les effets aléatoires ou les erreurs d'échantillonnage afin de diminuer les effets des valeurs aberrantes.

L'objectif du présent article est d'aborder ces deux questions importantes dans le contexte de l'estimation sur petits domaines. En premier lieu, nous considérons la modélisation des moyennes ainsi que des variances de population pour des petits domaines. Cette modélisation est possible grâce à l'existence de données additionnelles destinées à estimer les variances d'erreur. En deuxième lieu, afin d'induire une certaine robustesse dans notre procédure, nous considérons des lois a priori t pour les effets aléatoires.

L'ensemble de données utilisé dans le présent article provenait d'un essai de recensement démographique réalisé dans une municipalité brésilienne comptant 140 districts de recensement, que nous appellerons petits domaines. La variable étudiée était le revenu moyen des chefs de ménage pour chaque petit domaine, et l'objectif était de faire des prédictions pour les 140 moyennes de population des revenus des chefs de ménage. Les variables auxiliaires étaient les moyennes de population de petit domaine respectives du niveau d'études des chefs de ménage, et les moyennes de population respectives du nombre de pièces dans les logements des ménages pour chaque petit domaine. Seules des données au niveau du domaine nous ont été fournies.

Nous proposons une analyse bayésienne non subjective complète du problème général d'estimation sur petits domaines, où nous modélisons les moyennes ainsi que les variances de population. Au départ, l'idée était d'utiliser la loi a priori découlant de la règle générale de Jeffreys, en traitant tous les paramètres, y compris le nombre de degrés de liberté de la loi t de Student, comme étant inconnus. Cependant, la loi a priori résultante donnait une loi a posteriori impropre, qui a mené à une loi a priori de Jeffreys modifiée résultant en une loi a posteriori propre.

La présentation des autres sections de l'article est la suivante. À la section 2, nous décrivons le modèle, la matrice d'information de Fisher, ainsi que la loi a priori de Jeffreys et sa modification. L'impropriété de la loi a posteriori sous la loi a priori de Jeffreys et sa propriété sous sa version modifiée sont également traitées dans cette section. À la section 3, nous présentons une analyse de données réelles, ainsi qu'une étude en simulation. Enfin, à la section 4, nous formulons certaines observations finales.

Le caractère vraiment aléatoire des variances d'erreur est reconnu depuis longtemps. Les travaux d'Otto et Bell (1995), d'Arora et Lahiri (1997), de Wang et Fuller (2003), de Rivest et Vandal (2003) et d'autres ont tenté de tenir compte de ce fait de différentes façons. Slud et Maiti (2006), Dass, Maiti, Ren et Sinha (2012) et Maiti, Ren et Sinha (2014) ont adopté une approche bayésienne empirique à cette fin en estimant les hyperparamètres. Une analyse entièrement bayésienne en utilisant des méthodes bayésiennes hiérarchiques avec normalité des effets au niveau du domaine a été examinée dans You et Chapman (2006) et Sugasawa, Tamae et Kubokawa (2017). Au moyen d'une analyse des données et de simulations, nous montrerons que les lois a priori t pour les effets aléatoires donnent souvent de meilleurs résultats que les méthodes présentées dans ces deux derniers articles.

L'utilisation de lois a priori t pour les erreurs dans les modèles de régression normaux centrés réduits, mais non dans les modèles à effets mixtes, a été proposée dans Lange, Little et Taylor (1989), Fernandez et

Steel (1998), Vrontos, Dellaportas et Politis (2000), Jacquier, Polson et Rossi (2004), et Fonseca, Ferreira et Migon (2008), principalement en guise de protection contre les valeurs aberrantes. Toutefois, dans certaines situations, la normalité des erreurs est une hypothèse raisonnable, en raison du théorème central limite. En outre, il existe des techniques de diagnostic de modèle pour le vérifier. Malheureusement, l'hypothèse de normalité des effets aléatoires ne fonctionne pas toujours bien. Pour les données brésiliennes dont nous disposons, la modélisation conjointe des moyennes ainsi que des variances d'échantillon ainsi que des lois a priori t pour les effets aléatoires donne de meilleurs résultats que certains autres modèles au niveau du domaine. Comme l'a suggéré un examinateur, nous avons utilisé les données pour calculer les résidus en ajustant une régression avec des moyennes de domaine et des covariables réelles afin d'étudier la distribution des effets aléatoires pour cette application. Voir la section 3.

2 Le modèle

Un modèle au niveau du domaine type est donné par $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + u_i + e_i$, $(i = 1, \ldots, m)$, où m désigne le nombre de petits domaines, $\mathbf{x}_1, \ldots, \mathbf{x}_m$ est le vecteur de covariables de dimension $p\,(< m)$, et $\boldsymbol{\beta}\,(p \times 1)$ est le vecteur de coefficients de régression. Les distributions des effets aléatoires u_i et des erreurs d'échantillonnage e_i sont supposées indépendantes avec $u_i \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$ et $e_i \overset{\text{ind}}{\sim} N(0, v_i)$. Autrement dit, le modèle au niveau du domaine classique est

$$y_i \,|\, \theta_i \overset{\text{ind}}{\sim} N(\theta_i, v_i),$$

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma_u^2 \overset{\text{ind}}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_u^2), \quad i = 1, \ldots, m.$$

Les v_i sont supposées connues afin d'éviter la non-identifiabilité. L'hypothèse que les v_i sont connues devient presque obligatoire pour les utilisateurs secondaires des données qui n'ont accès à aucunes microdonnées pour les modéliser. Cependant, en réalité les variances sont aléatoires, basées sur des données échantillonnées. Dans les situations où l'on dispose de données additionnelles pour modéliser les v_i, ces dernières peuvent être estimées efficacement. En outre, dans de telles situations, il est possible d'obtenir des estimateurs à rétrécissement des moyennes de petit domaine θ_i, ainsi que des variances v_i.

Nous nous penchons sur les problèmes d'estimation sur petits domaines pour lesquels nous disposons de données additionnelles pour modéliser les v_i. En outre, pour la robustification, nous supposons que les effets aléatoires suivent une loi t plutôt qu'une loi normale. Nous énonçons notre modèle comme il suit,

$$y_i \,|\, \theta_i, v_i \overset{\text{ind}}{\sim} N(\theta_i, v_i),\ s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right)$$

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2, \nu \overset{\text{ind}}{\sim} t_\nu(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_\delta), \quad i = 1, \ldots, m, \tag{2.1}$$

où n_i est la taille de l'échantillon dans le i^{e} domaine, $t_\nu(\mu, \sigma)$ désigne la loi t de Student avec paramètres de position μ, d'échelle σ et de degrés de liberté ν, et $G(c, d)$ désigne la loi gamma avec la densité estimée par noyau $x^{c-1}\exp(-dx)$ pour $x > 0$.

Pour une analyse bayésienne complète, notre objectif est de trouver la loi a posteriori de $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$, sachant $\mathbf{y} = (y_1, \ldots, y_m)^T$ et $\mathbf{s}^2 = (s_1^2, \ldots, s_m^2)^T$. Pour cela, nous devons d'abord

trouver les lois a priori de tous les hyperparamètres, $\boldsymbol{\beta}$, $\mathbf{v} = (v_1, \ldots, v_m)^T$, σ_δ^2, et ν. Le premier essai habituel est la loi a priori de Jeffreys qui est proportionnelle à la racine carrée positive du déterminant de la matrice d'information de Fisher. Dans notre cas, cette matrice est

$$\mathbf{I}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) = \begin{bmatrix} \dfrac{(\nu+1)}{\sigma_\delta^2(\nu+3)}\mathbf{X}^T\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dfrac{m\nu}{2(\sigma_\delta^2)^2(\nu+3)} & \dfrac{-m}{\sigma_\delta^2(\nu+1)(\nu+3)} \\ \mathbf{0} & \mathbf{0} & \dfrac{-m}{\sigma_\delta^2(\nu+1)(\nu+3)} & mg(\nu) \end{bmatrix},$$

où $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$, $\mathbf{D} = \text{Diag}\left(\frac{n_1}{2v_1^2}, \ldots, \frac{n_m}{2v_m^2}\right)$ et $g(\nu) = \{\Psi'(\frac{\nu}{2}) - \Psi'(\frac{\nu+1}{2})\}/4 - (\nu+5)/\{2\nu(\nu+1)(\nu+3)\}$, avec $\Psi(z) = \Gamma'(z)/\Gamma(z)$ et $\Psi'(z) = d\Psi(z)/dz$ qui sont des fonctions digamma et trigamma. Donc, la loi a priori de Jeffreys est

$$\boldsymbol{\pi}_J(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) \propto (\sigma_\delta^2)^{-\frac{p}{2}-1} |\mathbf{X}^T\mathbf{X}|^{\frac{1}{2}} |\mathbf{D}|^{\frac{1}{2}} \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} \left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}. \qquad (2.2)$$

Cependant, la loi a priori de Jeffreys mène à une loi a posteriori impropre en raison du facteur $(\sigma_\delta^2)^{-\frac{p}{2}-1}$ dans (2.2).

Théorème 1. *La loi a priori de Jeffreys (2.2) mène à une loi a posteriori impropre.*

Preuve. Soit $\boldsymbol{\pi}_J \equiv \boldsymbol{\pi}_J(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2)$ la densité a posteriori avec la loi a priori de Jeffreys (2.2). En considérant les termes qui contiennent σ_δ^2 dans $\boldsymbol{\pi}_J$ et en prenant la transformation $w_i = (\theta_i - \mathbf{x}_i^T\boldsymbol{\beta})/\sigma_\delta$, c'est-à-dire $\theta_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sigma_\delta w_i$, nous avons

$$\int_0^\infty (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}(y_i - \mathbf{x}_i^T\boldsymbol{\beta} - \sigma_\delta w_i)^2\right] d\sigma_\delta^2$$

$$\geq \int_0^\infty (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left[-\sum_{i=1}^m \left\{\frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i} + \frac{\sigma_\delta^2 w_i^2}{v_i}\right\}\right] d\sigma_\delta^2$$

$$\geq \exp\left\{-\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i}\right\} \int_0^k (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left(-\sum_{i=1}^m \frac{\sigma_\delta^2 w_i^2}{v_i}\right) d\sigma_\delta^2 \text{ pour une constante } k > 0,$$

$$\geq \exp\left[-\sum_{i=1}^m \left\{\frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i} + \frac{kw_i^2}{v_i}\right\}\right] \int_0^k (\sigma_\delta^2)^{-\frac{p}{2}-1} d\sigma_\delta^2 = \infty.$$

Donc, la loi a priori de Jeffreys mène à une loi a posteriori impropre.

Cependant, une fois que la composante $(\sigma_\delta^2)^{-\frac{p}{2}-1}$ dans (2.2) est remplacée par $(\sigma_\delta^2)^{-\frac{p}{2}-1}\exp(-a/2\sigma_\delta^2)$ pour une valeur $a > 0$, cette version modifiée de la loi a priori de Jeffreys mène à une loi a posteriori propre sous la contrainte $\min(n_1, \ldots, n_m) > p$. Par conséquent, nous proposons pour notre modèle une loi a priori de Jeffreys modifiée comme il suit :

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) \propto (\sigma_\delta^2)^{-\frac{p}{2}-1}\exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^{m}\frac{1}{v_i}\right)\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}$$

$$\times\left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}} \text{ où } a > 0. \tag{2.3}$$

En combinant la vraisemblance de (2.1) et la loi a priori de Jeffreys modifiée (2.3), la loi a posteriori complète des paramètres, sachant les données, est

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2) \propto (\sigma_\delta^2)^{-\frac{p+m}{2}-1}\exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^{m} v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\frac{1}{v_i}(y_i-\theta_i)^2\right]$$

$$\times\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^{m}\left\{1+\frac{(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta})^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m$$

$$\times\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}. \tag{2.4}$$

Théorème 2. *Sous le modèle (2.1), la loi a posteriori* $\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2)$ *dans (2.4) est propre, à condition que* $\min(n_1, \ldots, n_m) > p$.

Preuve. Voir l'annexe A.

Le théorème 2 montre que la loi a priori de Jeffreys modifiée (2.3) mène à une loi a posteriori propre (2.4). L'idée essentielle est que nous avons besoin d'une loi a priori pour σ_δ^2 telle que $\int_0^\infty \boldsymbol{\pi}(\sigma_\delta^2)(\sigma_\delta^2)^{-\frac{p}{2}-1}\,d\sigma_\delta^2 < \infty$.

Remarque 1. $\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu)$ *peut être factorisé en quatre lois a priori indépendantes pour chaque paramètre.*

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) \propto \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\mathbf{v})\boldsymbol{\pi}(\sigma_\delta^2)\boldsymbol{\pi}(\nu)$$

où

$$\boldsymbol{\pi}(\boldsymbol{\beta}) \propto 1,\ \boldsymbol{\pi}(v_i) \propto \frac{1}{v_i} \quad \text{pour} \quad i = 1, \ldots, m,$$

$$\boldsymbol{\pi}(\sigma_\delta^2) \sim \mathrm{GI}\left(\frac{p}{2}, \frac{a}{2}\right)$$

et

$$\boldsymbol{\pi}(\nu) \propto \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} \left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}.$$

Ici, $\mathrm{GI}(c, d)$ désigne la loi gamma inverse avec la densité estimée par noyau $x^{-c-1}\exp(-d/x)$ pour $x > 0$.

Les lois conditionnelles complètes pour appliquer la méthode Monte Carlo par chaîne de Markov (MCMC) sont données à l'annexe B. Pour générer les échantillons, nous utilisons l'échantillonnage de Gibbs avec l'algorithme de Metropolis-Hastings, où la loi conditionnelle d'un paramètre est connue uniquement jusqu'à une constante multiplicative. Nous expliquons de façon détaillée comment appliquer un résultat de Chib et Greenberg (1995) pour l'algorithme de Metropolis-Hastings afin de générer les échantillons.

3 Application

3.1 Analyse de données réelles

L'ensemble de données est sélectionné par échantillonnage aléatoire de 10 % des ménages dans chaque domaine provenant d'un test de recensement démographique effectué dans une municipalité au Brésil. La municipalité compte 38 740 ménages dans 140 petits domaines au total, et le nombre de ménages par domaine dans la population varie de 57 à 588. Donc, les tailles d'échantillon de domaine dans l'ensemble de données varient de 6 à 59. Nous souhaitons estimer les 140 moyennes de population des revenus des chefs de ménage. La variable de réponse y_i est le revenu moyen des chefs de ménages dans le i^{e} domaine.

Cet ensemble de données comprend deux covariables auxiliaires centrées qui sont les moyennes de population de petit domaine respectives du niveau d'études des chefs de ménages (échelle ordinale de 0 à 5) et du nombre de pièces dans les logements des ménages (1 à 11+). Enfin, l'ensemble de données contient les variances d'échantillonnage respectives qui sont calculées de la manière habituelle. Puisque nous n'avons obtenu que des données au niveau du domaine et que les moyennes de domaine réelles sont connues, nous pouvons comparer les 140 prédictions sur petits domaines aux moyennes de domaine réelles, respectivement. L'analyse laisse entendre que notre modèle donne de meilleurs résultats que les autres modèles où les effets aléatoires sont fondés sur la loi normale. Aux fins de comparaison, nous utilisons trois autres modèles.

Le premier est le modèle de Fay-Herriot, noté FH, avec variances d'échantillonnage connues.

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \qquad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_\delta^2)$$

$$\pi(\boldsymbol{\beta}) \propto 1, \qquad \pi(\sigma_\delta^2) \sim \mathrm{GI}(a_0, b_0),$$

où a_0 et b_0 sont choisis égaux à 0,0001 (une petite constante) pour traduire la vague connaissance de σ_δ^2.

Le deuxième modèle, proposé par You et Chapman (2006), et noté YC, est un modèle bayésien hiérarchique donné par

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \quad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta^2)$$

$$s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right)$$

$$\pi(v_i) \sim \text{GI}(a_i, b_i), \quad \pi(\boldsymbol{\beta}) \propto 1, \quad \pi(\sigma_\delta^2) \sim \text{GI}(a_0, b_0),$$

où a_i, b_i, a_0 et b_0 sont également choisis égaux à 0,0001.

Le troisième modèle est un modèle bayésien à plusieurs degrés pour petits domaines proposé par Sugasawa et coll. (2017), désigné par STK. Le modèle STK produit une estimation à rétrécissement des moyennes ainsi que des variances.

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \quad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta^2)$$

$$s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right), \quad v_i \,|\, \gamma \sim \text{GI}(a_i, b_i\gamma),$$

$$\pi(\boldsymbol{\beta}, \sigma_\delta^2, \gamma) = 1,$$

où $a_i = 2$ et $b_i = 1/n_i$ comme l'ont proposé les auteurs pour un choix raisonnable.

Nous comparons les moyennes de petit domaine prédites par les modèles FH, YC, STK et par notre modèle, ci-après dénommé modèle RTS. Pour l'exécution de la méthode MCMC, nous générons une chaîne d'une longueur de rodage (*burn-in*) de 50 000 et la taille d'échantillonnage de $G = 50\,000$. Les estimations des θ_i sont données par

$$\hat{\theta}_i = \frac{1}{G}\sum_{g=1}^{G}\left(\gamma_i^{(g)} y_i + \left(1 - \gamma_i^{(g)}\right)\mathbf{x}_i^T \boldsymbol{\beta}^{(g)}\right)$$

où

$$\gamma_i^{(g)} = \frac{v_i^{-1(g)}}{v_i^{-1(g)} + \sigma_\delta^{-2(g)}\eta_i^{(g)}}.$$

Les critères de comparaison sont la moyenne des écarts quadratiques (MEQ), le biais absolu moyen (BAM), le biais relatif quadratique moyen (BRQM) et le biais relatif moyen (BRM). Ils sont définis comme il suit :

$$\text{MEQ} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\theta}_i - \theta_i\right)^2, \quad \text{BAM} = \frac{1}{m}\sum_{i=1}^{m}\left|\,\hat{\theta}_i - \theta_i\,\right|, \quad \text{BRQM} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right)^2,$$

et

$$\text{BRM} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{\theta}_i - \theta_i}{\theta_i}\right|,$$

où $\hat{\theta}_i$ et θ_i sont les valeurs estimées et réelles, respectivement, dans le i^{e} domaine. Le tableau 3.1 donne une comparaison des quatre modèles. Rappelons que la loi a priori de σ_δ^2 est $\boldsymbol{\pi}(\sigma_\delta^2) \sim \text{GI}\left(\frac{p}{2}, \frac{a}{2}\right)$. Avec le paramètre de forme, $\frac{p}{2} = 1$, nous considérons plusieurs valeurs de a. Si nous choisissons une valeur de a proche de p, le modèle RTS est mieux ajusté que les autres sous les quatre critères. Si nous choisissons $a = 1$, le modèle RTS est celui qui donne les meilleurs résultats. Le modèle YC donne de moins bons

résultats que les trois autres modèles. Si nous choisissons une très petite valeur de a, telle que 0,01 ou 0,001, le modèle RTS est celui qui donne les pires résultats.

Tableau 3.1
Comparaison entre le modèle RTS, le modèle FH, le modèle YC et le modèle STK

Modèle	MEQ	BAM	BRQM	BRM
Modèle RTS (a = 0,0001)	57,297	6,152	0,395	0,589
Modèle RTS (a = 0,01)	16,546	2,741	0,090	0,244
Modèle RTS (a = 0,5)	3,244	1,249	0,020	0,118
Modèle RTS (a = 0,2)	4,185	1,439	0,025	0,133
Modèle RTS (a = 1)	2,745	1,164	0,019	0,113
Modèle RTS (a = 2)	3,080	1,231	0,020	0,117
Modèle RTS (a = 3)	3,079	1,229	0,020	0,117
Modèle RTS (a = 5)	2,994	1,213	0,019	0,116
Modèle RTS (a = 10)	3,377	1,278	0,020	0,119
Modèle RTS (a = 50)	2,905	1,180	0,018	0,112
Modèle RTS (a = 100)	2,799	1,154	0,018	0,109
Modèle FH	4,484	1,448	0,026	0,133
Modèle YC	4,983	1,543	0,029	0,141
Modèle STK	3,199	1,257	0,021	0,121

De plus, comme l'a suggéré un examinateur, nous calculons les résidus en ajustant une régression à l'aide des moyennes de domaine et de covariables réelles pour visualiser la distribution des effets aléatoires pour ces données réelles. La figure 3.1 montre que la distribution s'écarte de la distribution normale.

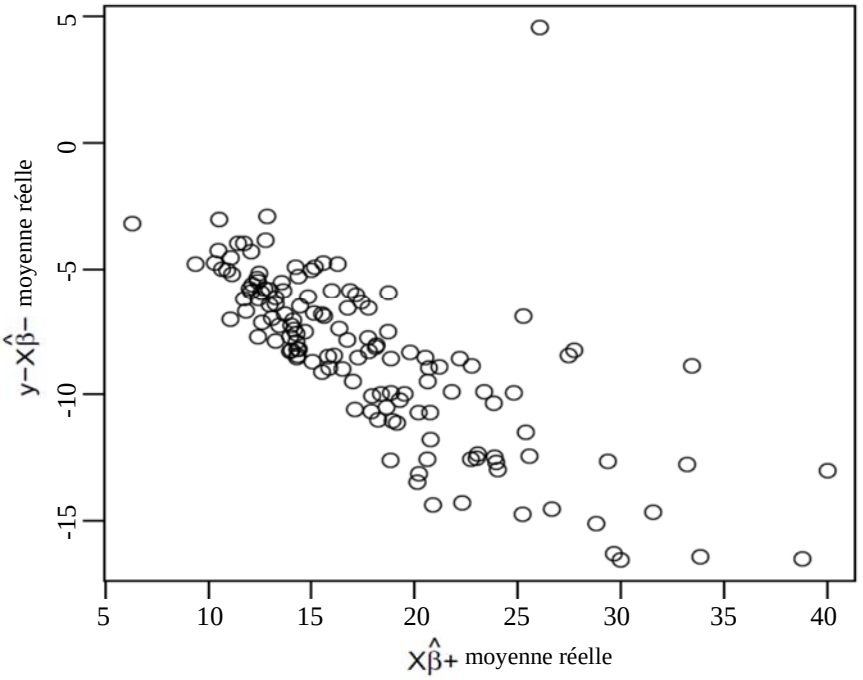

Figure 3.1 Résidus en ajustant une régression avec les moyennes et covariables réelles.

3.2 Étude en simulation

À la présente section, nous configurons une simulation proche de celle décrite dans Maiti et coll. (2014) (ou Sugasawa et coll. (2017)) pour comparer la précision de nos estimateurs à d'autres, dont ceux de You et Chapman (2006) et de Sugasawa et coll. (2017). Nous générons des observations pour chaque petit domaine à partir du modèle

$$y_{ij} = \beta_0 + \beta_1 x_i + u_i + e_{ij}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, m,$$

où $u_i \sim t_\nu(0, \sigma_\delta)$ et $e_{ij} \sim N(0, n_i v_i)$. Alors, le modèle des effets aléatoires pour la moyenne de petit domaine est

$$y_i = \beta_0 + \beta_1 x_i + u_i + e_i, \quad i = 1, \ldots, m,$$

où $y_i = \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ et $e_i = n_i^{-1} \sum_{j=1}^{n_i} e_{ij}$. Donc, $y_i \,|\, \theta_i \sim N(\theta_i, v_i)$, où $\theta_i = \beta_0 + \beta_1 x_i + u_i$; $\theta_i \sim t_\nu(\beta_0 + \beta_1 x_i, \sigma_\delta)$, et $e_i \sim N(0, v_i)$. Le paramètre d'intérêt est la moyenne θ_i, pour le i^{e} petit domaine. En outre, l'estimateur direct de v_i est

$$\frac{1}{n_i (n_i - 1)} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2.$$

Nous prenons $m = 30$ et $n_i = 7$ pour tous les domaines, et générons les covariables x_i à partir de la loi uniforme sur (2, 8). Les valeurs réelles des paramètres sont fixées à $\beta_0 = 0{,}5$; $\beta_1 = 0{,}8$; $\sigma_\delta = 1$; $\nu = 3$ et $v_i \sim \text{GI}(10, 5\exp(0{,}3x_i))$. En outre, nous choisissons $a = 3$ pour toutes les simulations.

Pour l'exécution de la méthode MCMC, nous avons généré 5 000 échantillons a posteriori après avoir écarté les 1 000 premiers pour $R = 2\,000$ simulations. Le tableau 3.2 donne une comparaison des quatre modèles. Les critères de comparaison sont la MEQ, le BAM et le BIAIS, ce dernier étant défini comme étant

$$\text{BIAIS} = \frac{1}{mR} \sum_{i=1}^{m} \left| \sum_{r=1}^{R} \left(\hat{\theta}_i^{(r)} - \theta_i^{(r)} \right) \right|.$$

Tableau 3.2
Résultat des simulations pour des effets aléatoires de loi t avec $\nu = 3$

Modèle	Moyenne			Variance		
	MEQ	BAM	BIAIS	MEQ	BAM	BIAIS
RTS	1,393	0,895	0,021	5,048	1,608	1,324
STK	1,821	0,933	0,025	5,042	1,514	1,367
FH	1,540	0,942	0,022			
YC	2,165	0,974	0,030	5,970	1,803	1,689

Le modèle RTS donne de meilleurs résultats que les autres sous les critères MEQ, BAM et BIAIS pour la moyenne. Tandis que le modèle RTS présente de petites améliorations par rapport aux autres modèles pour les critères BAM et BIAIS, il produit des améliorations d'environ 23,5 %, 10 % et 35,7 % par rapport

aux modèles STK, FH et YC, respectivement, pour le critère MEQ. Pour la variance, les modèles RTS et STK sont meilleurs que le modèle YC.

Les deux tableaux qui suivent donnent les résultats des simulations lorsque l'on fixe le nombre de degrés de liberté à $\nu = 2$ et $\nu = 4$.

Tableau 3.3
Résultat des simulations pour des effets aléatoires de loi t avec $\nu = 2$

Modèle	Moyenne			Variance		
	MEQ	BAM	BIAIS	MEQ	BAM	BIAIS
RTS	1,617	0,949	0,020	6,569	1,610	1,070
STK	7,566	1,107	0,038	11,144	1,441	0,996
FH	1,921	1,035	0,022			
YC	9,063	1,187	0,038	7,072	1,685	1,340

Tableau 3.4
Résultat des simulations pour des effets aléatoires de loi t avec $\nu = 4$

Modèle	Moyenne			Variance		
	MEQ	BAM	BIAIS	MEQ	BAM	BIAIS
RTS	1,265	0,862	0,019	4,876	1,619	1,428
STK	1,322	0,874	0,019	5,077	1,577	1,489
FH	1,350	0,894	0,020			
YC	1,509	0,905	0,022	6,201	1,869	1,802

Avec $\nu = 2$, le modèle RTS donne de meilleurs résultats que les autres sous les critères MEQ, BAM et BIAIS pour la moyenne. Dans cette simulation, les valeurs de MEQ pour les modèles STK et YC sont très grandes comparativement aux modèles RTS et FH. Le modèle RTS affiche des améliorations d'environ 78,6 % par rapport au modèle STK, 82,2 % par rapport au modèle YC, et 15,8 % par rapport au modèle FH. Pour les critères BAM et BIAIS, les valeurs du modèle RTS sont plus petites que celles des autres modèles. Si l'on examine la variance, le modèle RTS donne la valeur la plus petite pour MEQ.

Avec $\nu = 4$, le modèle RTS donne aussi de meilleurs résultats que les autres. En particulier, les valeurs de MEQ et de BIAIS indiquent que le modèle RTS améliore les résultats comparativement aux modèles STK et YC.

Les deux tableaux qui suivent représentent la situation lorsqu'on émet l'hypothèse de normalité des effets aléatoires. Ici, le modèle RTS donne des résultats un peu moins bons que les autres modèles.

Tableau 3.5
Résultat des simulations pour des effets aléatoires normaux avec $N(0, 5^2)$

Modèle	Moyenne			Variance		
	MEQ	BAM	BIAIS	MEQ	BAM	BIAIS
RTS	2,896	1,305	0,038	6,036	1,512	0,514
STK	2,560	1,229	0,051	1,851	0,961	0,114
FH	2,597	1,240	0,036			
YC	2,735	1,259	0,048	3,674	1,305	0,463

Tableau 3.6
Résultat des simulations pour des effets aléatoires normaux avec $N(0,10^2)$

Modèle	Moyenne			Variance		
	MEQ	BAM	BIAIS	MEQ	BAM	BIAIS
RTS	3,007	1,316	0,032	10,117	1,895	1,202
STK	2,784	1,272	0,031	2,221	1,038	0,155
FH	2,765	1,272	0,048			
YC	2,873	1,285	0,033	9,166	1,798	1,129

4 Observations finales

L'article décrit des modèles pour petits domaines en vue de traiter des données au niveau du domaine. L'élément nouveau de cet article est la modélisation à la fois des moyennes et des variances de petit domaine, parallèlement à l'utilisation d'une loi t pour les effets aléatoires. Il est montré par analyse de données et par simulation que la méthode proposée donne essentiellement de meilleurs résultats que les modèles de You et Chapman (2006) et de Sugasawa et coll. (2017) dans la plupart des situations.

Remerciements

Les travaux de recherche de Ghosh ont été financés en partie par la subvention SES-1327359 de la NSF.

Annexe A

Preuve

Théorème 2. Sous le modèle (2.1) avec la loi a priori de Jeffreys modifiée (2.3), la loi a posteriori (2.4) est propre, à condition que $\min(n_1, \ldots, n_m) > p$.

Preuve. Rappelons la loi a posteriori (2.4),

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \mid \mathbf{y}, \mathbf{s}^2) \propto (\sigma_\delta^2)^{-\frac{p+m}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}(y_i-\theta_i)^2\right]$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^m \left\{1+\frac{(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta})^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m$$

$$\times\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{\nu g(\nu)}{2(\nu+3)}-\frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}$$

où $g(\nu) = \left\{\Psi'\left(\frac{\nu}{2}\right) - \Psi'\left(\frac{\nu+1}{2}\right)\right\}\Big/4 - (\nu+5)\Big/\{2\nu(\nu+1)(\nu+3)\}$.

Avant tout, puisque $\log\left[\Gamma\left\{\frac{(\nu+1)}{2}\right\}\Big/\Gamma\left(\frac{\nu}{2}\right)\right] \doteq \{-\log(2e) + \nu\log(\nu+1) - (\nu-1)\log\nu\}/2$ en vertu de l'approximation de Stirling, nous avons

$$\frac{1}{2}\left[\Psi\left(\frac{\nu+1}{2}\right)-\Psi\left(\frac{\nu}{2}\right)\right] \doteq \frac{d}{d\nu}\log\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} = \frac{1}{2}\left[\log(\nu+1)-\log\nu\right]+\frac{1}{2\nu(\nu+1)}$$

et

$$\frac{1}{4}\left[\Psi'\left(\frac{\nu+1}{2}\right)-\Psi'\left(\frac{\nu}{2}\right)\right] \doteq \frac{1}{2}\left(\frac{1}{\nu+1}-\frac{1}{\nu}\right)-\frac{1}{2\nu^2}+\frac{1}{2(\nu+1)^2}.$$

Cela mène à

$$g(\nu) \doteq \frac{5\nu+3}{2\nu^2(\nu+1)^2(\nu+3)}$$

et

$$\left[\frac{\nu g(\nu)}{2(\nu+3)}-\frac{1}{(\nu+1)^2(\nu+3)^2}\right] \doteq \frac{1}{4\nu(\nu+1)^2(\nu+3)}.$$

Donc, cette approximation simplifie le dernier terme dans (2.4). La loi a posteriori correspondante est

$$\begin{aligned}\boldsymbol{\pi}_{\mathrm{MJ}}(\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{v},\sigma_\delta^2,\nu\,|\,\mathbf{y},\,\mathbf{s}^2) \propto{}& (\sigma_\delta^2)^{-\frac{p+m}{2}-1}\exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^m\frac{1}{v_i}(y_i-\theta_i)^2\right]\\ &\times\exp\left(-\frac{1}{2}\sum_{i=1}^m\frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^m\left\{1+\frac{(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta})^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m\\ &\times\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}. \qquad \text{(A.1)}\end{aligned}$$

D'abord, éliminons $\boldsymbol{\beta}$ par intégration. En prenant $w_i = (\theta_i - \mathbf{x}_i^T\boldsymbol{\beta})/\sigma_\delta$, c'est-à-dire $\theta_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sigma_\delta w_i$, nous avons

$$\begin{aligned}\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{w},\mathbf{v},\sigma_\delta^2,\nu\,|\,\mathbf{y},\,\mathbf{s}^2) \propto{}& (\sigma_\delta^2)^{-\frac{p}{2}-1}\exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m\frac{s_i^2}{v_i}\right)\\ &\times\left[\prod_{i=1}^m\left(1+\frac{w_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}|\,\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\,|^{-\frac{1}{2}}\\ &\times\exp\left[-\frac{1}{2}(\mathbf{Y}-\sigma_\delta\mathbf{w})^T\left\{\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right\}(\mathbf{Y}-\sigma_\delta\mathbf{w})\right].\end{aligned}$$

où $\mathbf{w} = (w_1, \ldots, w_m)^T$. Notons que $\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ est définie non négative. En utilisant $K(>0)$ comme une constante générique qui ne dépend que des données $(\mathbf{y}, \mathbf{s}^2)$, et en éliminant par intégration $\mathbf{w}$ et σ_δ^2 par ordre respectivement, nous obtenons

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2) \leq K(\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}} |\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}},$$

et puis

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v}, \nu \,|\, \mathbf{y}, \mathbf{s}^2) \leq K\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right) \times \left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}} |\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}}.$$

Notons que

$$\int_0^\infty \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}\,d\nu = \int_0^c \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}\,d\nu + \int_c^\infty \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}\,d\nu$$

$$\leq \int_0^c \nu^{-\frac{1}{2}}d\nu + \int_c^\infty \nu^{-2}d\nu = 2c^{\frac{1}{2}} + c^{-1} < \infty \text{ pour tout } c > 0.$$

Après élimination de ν par intégration, nous avons

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v} \,|\, \mathbf{y}, \mathbf{s}^2) \leq K\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}}.$$

Enfin, puisque $|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}} \leq (v_{\max})^{\frac{p}{2}}|\mathbf{X}^T\mathbf{X}|^{-\frac{1}{2}}$, où $v_{\max} = \max(v_1, \ldots, v_m)$, si $\min(n_1, \ldots, n_m) > p$, la loi a priori de Jeffreys modifiée mène à une loi a posteriori propre.

Annexe B

Lois conditionnelles complètes

La loi a posteriori complète des paramètres sachant les données est spécifiée en (A.1). Pour l'exécution de la méthode MCMC, il est commode d'utiliser les paramètres latents η_i $(i = 1, \ldots, m)$ tels que

$$\theta_i \,|\, \eta_i \overset{\text{ind}}{\sim} \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta}, \eta_i^{-1}\sigma_\delta^2)$$

$$\eta_i \overset{\text{iid}}{\sim} \mathrm{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

Soit $\boldsymbol{\Lambda} = \mathrm{Diag}(\eta_1, \ldots, \eta_m)$. Les lois conditionnelles complètes sont

I. $[\sigma_\delta^2 \,|\, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \nu, \mathbf{y}, \mathbf{s}^2] \sim \mathrm{GI}\left[\frac{p+m}{2}, \frac{1}{2}\left\{a + \sum_{i=1}^m \eta_i(\theta_i - \mathbf{x}_i^T\boldsymbol{\beta})^2\right\}\right]$;

II. $[v_i \,|\, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2] \overset{\text{ind}}{\sim} \mathrm{GI}\left[\frac{n_i}{2}, \frac{1}{2}\left\{(y_i - \theta_i)^2 + s_i^2\right\}\right]$;

III. $[\boldsymbol{\beta} \,|\, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2] \sim \mathrm{N}\left[(\mathbf{X}^T\boldsymbol{\Lambda}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Lambda}\boldsymbol{\theta}, \sigma_\delta^2(\mathbf{X}^T\boldsymbol{\Lambda}\mathbf{X})^{-1}\right]$;

IV. $[\eta_i \,|\, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2] \overset{\text{ind}}{\sim} \mathrm{G}\left[\frac{\nu+1}{2}, \frac{1}{2}\left\{\nu + \frac{(\theta_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{\sigma_\delta^2}\right\}\right]$;

$$\text{V.} \quad [\theta_i \,|\, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2] \overset{\text{ind}}{\sim} \mathrm{N}\left[\left(v_i^{-1} + \sigma_\delta^{-2}\eta_i\right)^{-1}\left(v_i^{-1} y_i + \sigma_\delta^{-2}\eta_i \mathbf{x}_i^T \boldsymbol{\beta}\right), \left(v_i^{-1} + \sigma_\delta^{-2}\eta_i\right)^{-1}\right];$$

et

$$\text{VI.} \quad [\nu \,|\, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \mathbf{y}, \mathbf{s}^2] \propto \nu^{\frac{m-1}{2}} \exp\left\{-\tfrac{\nu}{2}\sum_{i=1}^{m}(\eta_i - \log\eta_i - 1)\right\}$$

$$\times \left\{\frac{\nu^{\frac{m\nu-m}{2}}\exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}}\Gamma^m\left(\frac{\nu}{2}\right)}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}\right\}. \qquad \text{(B.1)}$$

Toutes les lois sauf (VI) requièrent la génération d'échantillons à partir de lois standards. Alors que nous utilisons la méthode d'échantillonnage de Gibbs pour les lois (I) à (V), nous utilisons l'algorithme de Metropolis-Hastings pour générer les échantillons à partir de (VI) comme il est décrit dans Chib et Greenberg (1995).

Application du résultat de Chib et Greenberg (1995) à (IV) dans (B.1)

Si la densité cible $\pi(t)$ peut s'écrire sous la forme $\pi(t) \propto \psi(t)\,h(t)$, où $h(t)$ est une densité qui peut être échantillonnée et $\psi(t)$ est uniformément bornée, alors on peut prendre $h(t)$ comme une densité candidate pour le tirage d'échantillons et utiliser $\psi(t)$ dans $\alpha(x, y) = \min\{\psi(y)/\psi(x), 1\}$ qui est la probabilité de mouvement.

Rappelons que la loi conditionnelle complète de ν est

$$\boldsymbol{\pi}(\nu \,|\, \cdot) \propto \nu^{\frac{m-1}{2}}\exp\left\{-\frac{\nu}{2}\sum_{i=1}^{m}(\eta_i - \log\eta_i - 1)\right\} \times \left\{\frac{\nu^{\frac{m\nu-m}{2}}\exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}}\Gamma^m\left(\frac{\nu}{2}\right)}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}\right\}.$$

Puisque $\Gamma\left(\frac{\nu}{2}\right) \geq \sqrt{2\pi}\,\exp\left(-\frac{\nu}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{\nu-1}{2}}$, le deuxième terme $\{\cdot\}$ ci-dessus est borné par $\left(\sqrt{2\pi}\right)^{-m}(\nu+1)^{\frac{p}{2}-1}(\nu+3)^{-\frac{p}{2}-\frac{1}{2}} \leq \left(\sqrt{2\pi}\right)^{-m}\Big/\sqrt{3}$. Donc, nous pouvons appliquer la troisième méthode de la section 5 de Chib et Greenberg (1995) avec

$$h(\nu) \sim \mathrm{G}\left(\frac{m+1}{2}, \frac{1}{2}\sum_{i=1}^{m}(\eta_i - \log\eta_i - 1)\right)$$

et

$$\psi(\nu) = \frac{\nu^{\frac{m\nu-m}{2}}\exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}}\Gamma^m\left(\frac{\nu}{2}\right)}\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}.$$

Ici, $h(\nu)$ est une densité génératrice de candidats, et $\psi(\nu)$ est uniformément bornée.

Bibliographie

Arora, V., et Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Stastica Sinica*, 7, 1053-1063.

Battese, G.E., Harter, R.M. et Fuller, W.A. (1988). An error component model for prediction of mean crop areas using survey and satellite data. *Journal of the American Statistical Association*, 95, 28-36.

Bell, W.R. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. Dans *Proceedings of the Survey Research Methods Section,* American Statistical Association, 327-333.

Bell, W.R., et Huang, E.T. (2006). Utilisation de la loi t pour traiter les valeurs aberrantes dans l'estimation pour de petits domaines. Recueil : *Symposium 2006, Enjeux méthodologiques reliés à la mesure de la santé des populations,* Statistique Canada.

Butar, F., et Lahiri, P. (2002). On the measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference,* 112, 63-76.

Chib, S., et Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician,* 49, 327-335.

Dass, S.C., Maiti, T., Ren, H. et Sinha, S. (2012). Estimation des intervalles de confiance des paramètres de petit domaine avec rétrécissement des moyennes et des variances. *Techniques d'enquête,* 38, 2, 187-203. Article accessible à l'adresse https://www150.statcan.gc.ca/n1/pub/12-001-x/2012002/article/11756-fra.pdf.

Datta, G.S., et Lahiri, P. (1995). Robust hierarchical Bayes estimation of small area characteristics in the presence of outliers. *Journal of Multivariate Analysis,* 54, 310-328.

Fay, R.E., et Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association,* 74, 269-277.

Fernandez, C., et Steel, M.F.J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association,* 93, 359-371.

Fonseca, T.C.O., Ferreira, M.A.R. et Migon, H.S. (2008). Objective Bayesian analysis for the student $-t$ regression model. *Biometrika,* 95, 325-333.

Jacquier, E., Polson, N.G. et Rossi, P.E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122, 185-212.

Jiang, J., Lahiri, P. et Wan, S. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Annals of Statistics,* 30, 1782-1810.

Lahiri, P., et Rao, J.N.K. (1995). Robust estimation of mean square error of small area estimators. *Journal of the American Statistical Association,* 90, 758-766.

Lange, K.L., Little, R.J.A. et Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association,* 84, 881-896.

Maiti, T., Ren, H. et Sinha, A. (2014). Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics,* 41, 775-790.

Otto, M.C., et Bell, W.R. (1995). Sampling error modelling of poverty and income statistics for states. Dans *Proceedings of the Section on Government Statistics,* American Statistical Association, 160-165.

Rivest, L.-P., et Vandal, N. (2003). Mean squared error estimation for small areas when the small area variances are estimated. Dans *Proceedings of the International Conference on Recent Advances in Survey Sampling.*

Slud, E.V., et Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of Royal Statistical Society, Series B*, 68, 239-257.

Sugasawa, S., Tamae, H. et Kubokawa, T. (2017). Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*, 44, 150-167.

Vrontos, I.D., Dellaportas, P. et Politis, D.N. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economic Statistics*, 18, 187-198.

Wang, J., et Fuller, W. (2003). The mean squared error of small area predictors constructed with estimated error variances. *Journal of the American Statistical Association*, 98, 716-723.

You, Y., et Chapman, B. (2006). Estimation pour petits domaines au moyen de modèles régionaux et d'estimations des variances d'échantillonnage. *Techniques d'enquête*, 32, 1, 107-114. Article accessible à l'adresse https://www150.statcan.gc.ca/n1/pub/12-001-x/2006001/article/9263-fra.pdf.