Composante du produit nº 12-001-X
Division des méthodes d'enquêtes auprès des entreprises

Article

Une approche intégrée de modélisation de l'estimation du taux de chômage pour les régions infraprovinciales au Canada

par Yong You

Juin 2008

Statistique Canada Statistics Canada

Canada

Techniques d'enquête, juin 2008
Vol. 34, N° 1, pp. 21-31
Statistique Canada, N° 12-001-X au catalogue

Une approche intégrée de modélisation de l'estimation du taux de chômage pour les régions infraprovinciales au Canada

Yong You [1]

Résumé

L'Enquête sur la population active (EPA) réalisée au Canada permet de produire des estimations mensuelles du taux de chômage aux niveaux national et provincial. Le programme de l'EPA diffuse aussi des estimations du chômage pour des régions infraprovinciales, comme les régions métropolitaines de recensement (RMR) et les centres urbains (CU). Cependant, pour certaines de ces régions infraprovinciales, les estimations directes ne sont pas fiables, parce que la taille de l'échantillon est assez petite. Dans le contexte de l'EPA, l'estimation pour de petites régions a trait à l'estimation des taux de chômage pour des régions infraprovinciales telles que les RMR/CU à l'aide de modèles pour petits domaines. Dans le présent article, nous discutons de divers modèles, dont celui de Fay-Herriot et des modèles transversaux ainsi que chronologiques. En particulier, nous proposons un modèle non linéaire intégré à effets mixtes sous un cadre hiérarchique bayésien (HB) pour l'estimation du taux de chômage d'après les données de l'EPA. Nous utilisons les données mensuelles sur les bénéficiaires de l'assurance-emploi (a.-e.) au niveau de la RMR ou du CU comme covariables auxiliaires dans le modèle. Nous appliquons une approche HB ainsi que la méthode d'échantillonnage de Gibbs pour obtenir les estimations des moyennes et des variances a posteriori des taux de chômage au niveau de la RMR ou du CU. Le modèle HB proposé produit des estimations fondées sur un modèle fiables si l'on s'en tient à la réduction du coefficient de variation. Nous présentons dans l'article une analyse d'ajustement du modèle et une comparaison des estimations fondées sur le modèle aux estimations directes.

Mots clés : Effet de plan; cadre hiérarchique bayésien; modèle loglinéaire à effets mixtes; vérification du modèle; variance d'échantillonnage; petit domaine.

1. Introduction

Le taux de chômage est généralement considéré comme un indicateur important des résultats économiques. Au Canada, les estimations du taux de chômage sont produites mensuellement par le programme de l'Enquête sur la population active (EPA) de Statistique Canada. Cette dernière est une enquête mensuelle réalisée auprès de 53 000 ménages sélectionnés conformément à un plan d'échantillonnage stratifié à plusieurs degrés. Chaque mois, un sixième de l'échantillon est remplacé. Donc, cinq sixièmes de l'échantillon sont communs à deux mois consécutifs. Ce chevauchement des échantillons crée des corrélations qui peuvent être exploitées pour produire de meilleures estimations au moyen d'autres méthodes que l'estimation directe, telles que les méthodes fondées sur un modèle, qui renforcent l'estimation à l'aide de données accumulées au fil du temps; cette approche sera examinée plus en détail à la section 2. Pour une description détaillée du plan de sondage de l'EPA, consulter Gambino, Singh, Dufour, Kennedy et Lindeyer (1998). Le programme de l'EPA diffuse des estimations mensuelles du taux de chômage pour de grandes régions, telles que l'ensemble du pays et les provinces, ainsi que pour des régions locales (petites régions ou petits domaines), telles que les régions métropolitaines de recensement (RMR, c'est-à-dire les villes comptant plus de 100 000 habitants) et les autres centres urbains du Canada. Bien que les estimations nationales et provinciales soient celles qui intéressent le plus les médias, les estimations infraprovinciales sont également très importantes. Le programme de l'assurance-emploi (a.-e.) les utilise pour établir les règles à appliquer pour administrer le programme. En outre, les taux de chômage dans les RMR et les CU sont examinés de près au niveau local. Cependant, pour nombre de régions locales, la taille de l'échantillon n'est pas suffisante pour calculer des estimations directes adéquates, car l'EPA est conçue pour produire des estimations adéquates ou fiables aux niveaux national et provincial. Pour les estimations nationales, le coefficient de variation (CV) est de l'ordre de 2 % et pour les estimations provinciales, de l'ordre de 4 % à 7 %. En revanche, pour les RMR et les CU, il varie d'environ 7 % à 50 %. En outre, pour certains CU, il est même supérieur à 50 %. Les estimations directes basées sur les données de l'EPA ne sont pas très fiables pour certaines régions municipales, le CV étant très grand à cause de la petite taille de l'échantillon pour ces régions. Par conséquent, d'autres estimateurs, en particulier des estimateurs fondés sur un modèle, sont envisagés pour améliorer les estimations directes d'après l'EPA pour les petites régions. L'objectif du présent article est d'obtenir un estimateur fondé sur un modèle fiable qui produit de meilleures estimations que l'estimateur direct d'après l'EPA, c'est-à-dire des estimations dont le CV est petit et stable.

1. Yong You, Division des méthodes d'enquête auprès des ménages, Statistique Canada, Ottawa, (Ontario), K1A 0T6. Courriel : yong.you@statcan.ca.

En général, on utilise les estimateurs directs, fondés uniquement sur des données d'échantillon sur un domaine particulier, pour estimer les paramètres pour de grands domaines, tels que le pays et les provinces. Par contre, les tailles d'échantillon pour les petits domaines, particulièrement les petites régions géographiques, sont rarement suffisantes pour fournir des estimations directes fiables. Quand on calcule des estimations pour des petits domaines, il est nécessaire d'emprunter de l'information supplémentaire à des domaines apparentés pour former des estimateurs indirects qui accroissent la taille effective d'échantillon et, donc, la précision. De tels estimateurs indirects sont fondés sur des modèles implicites ou explicites qui établissent un lien avec des petits domaines apparentés grâce à des données supplémentaires, tels que des dénombrements du recensement et des dossiers administratifs. Aujourd'hui il est généralement reconnu que, s'il faut recourir à des estimations indirectes, celles-ci devraient être fondées sur des modèles explicites qui relient les petits domaines d'intérêt grâce à des données supplémentaires; voir Rao (2003). Les estimateurs fondés sur un modèle sont des estimateurs indirects en ce sens qu'ils sont obtenus en utilisant des modèles pour petits domaines, des estimations directes et des variables auxiliaires connexes. Les estimateurs fondés sur un modèle sont établis en vue d'améliorer les estimateurs directs fondés sur le plan de sondage en ce qui concerne la précision et la fiabilité, c'est-à-dire la réduction des coefficients de variation. Habituellement, les estimateurs sur petits domaines sont renforcés par emprunt de données recueillies pour des petits domaines similaires à une période donnée ou pour le même domaine au fil du temps, mais non les deux. Ces dernières années, plusieurs méthodes ont été élaborées en vue d'emprunter simultanément des données de renfort transversales (dimension spatiale) et longitudinales (dimension temporelle). Les estimateurs fondés sur l'approche établie par Rao et Yu (1994), Ghosh, Nangia et Kim (1996), Datta, Lahiri, Maiti et Lu (1999), Datta, Lahiri et Maiti (2002), ainsi que You, Rao et Gambino (2000, 2003), permettent d'exploiter simultanément les deux dimensions pour produire des estimations améliorées ayant des propriétés souhaitables pour les petits domaines. En particulier, You et coll. (2000, 2003) ont étudié l'estimation fondée sur un modèle des taux de chômage pour des régions infra-provinciales locales, comme les RMR et les agglomérations de recensement (AR) au Canada. Ils ont obtenu des estimateurs fondés sur un modèle efficaces et un ajustement adéquat des modèles pour l'estimation des taux de chômage d'après l'EPA. Toutefois, le modèle proposé par You et coll. (2000, 2003) présente certaines limites. Dans le présent article, nous discutons de ces limites et proposons un nouveau modèle intégré pour l'estimation des taux de chômage d'après l'EPA sous un cadre hiérarchique bayésien (HB). L'idée est de modéliser ensemble les paramètres d'intérêt et les variances d'échantillonnage, comme l'ont suggéré You et coll. (2003), ainsi que You et Chapman (2006). Nous appliquons le modèle proposé aux données de l'EPA de 2005 et obtenons les estimations des taux de chômage fondés sur un modèle. Nous donnons aussi une comparaison des estimations HB aux estimations directes d'après l'EPA et l'analyse de l'ajustement du modèle.

Le plan de l'article est le suivant. À la section 2, nous présentons divers modèles pour petits domaines proposés dans la littérature pour l'estimation des taux de chômage et nous en discutons. À la section 3, nous examinons le problème du lissage et de la modélisation de la matrice de covariance d'échantillonnage. À la section 4, nous proposons un modèle non linéaire intégré à effets mixtes dans un cadre hiérarchique bayésien et décrivons l'utilisation de l'échantillonnage de Gibbs pour générer des échantillons à partir de la loi conjointe a posteriori. À la section 5, nous appliquons le modèle proposé aux données de l'EPA et obtenons les estimations HB pour les taux de chômage pour petits domaines. Nous donnons aussi l'analyse et l'évaluation du modèle. Enfin, à la section 6, nous présentons certaines conclusions et discutons de l'orientation de futurs travaux.

2. Modèles pour petits domaines

2.1 Modèle transversal

Les modèles transversaux, ou au niveau du domaine, sont utilisés pour produire des estimations fondées sur un modèle fiable par combinaison de l'information auxiliaire au niveau du domaine et des estimations directes au niveau du domaine. Un modèle transversal de base est celui, bien connu, de Fay-Herriot (Fay et Herriot 1979). Il possède deux composantes, à savoir 1) un modèle d'échantillonnage pour les estimations directes par sondage et 2) un modèle de lien qui relie les paramètres du petit domaine à des variables auxiliaires au niveau du domaine grâce à un modèle de régression linéaire. Pour l'estimation mensuelle du taux de chômage d'après l'EPA, dénotons par θ_{it} le taux de chômage réel pour la i^{e} RMR/CU à un point particulier dans le temps (mois) t, où $i = 1, ..., m$, où m est le nombre de RMR/CU, et soit y_{it} l'estimation directe d'après l'EPA de θ_{it}. Alors, le modèle d'échantillonnage pour y_{it} peut être exprimé sous la forme

$$y_{it} = \theta_{it} + e_{it}, \ \ i = 1, \ ..., \ m, \qquad (1)$$

où e_{it} est l'erreur d'échantillonnage associée à l'estimateur direct y_{it}. Nous supposons que cette erreur suit une loi normale telle que $e_{it} \sim N(0, \sigma_{it}^2)$ où σ_{it}^2 est la variance

d'échantillonnage. Le modèle de lien pour le taux de chômage réel θ_{it} peut s'écrire sous la forme

$$\theta_{it} = x'_{it}\beta + v_i, \quad i = 1, \ldots, m, \qquad (2)$$

où x_{it} est la variable auxiliaire et v_i est l'effet aléatoire particulier au domaine. Pour chaque point dans le temps (chaque mois), nous pouvons utiliser le modèle de Fay-Herriot pour produire les estimations directes mensuelles. Ce modèle combine l'information transversale, mais n'est pas renforcé par emprunt d'information recueillie au cours des périodes antérieures.

2.2 Modèle transversal et chronologique

Étant donné le plan d'échantillonnage du processus de renouvellement de l'échantillon de l'EPA, il existe, dans chaque domaine, un chevauchement important d'échantillons sur des périodes de six mois. Par conséquent, pour un domaine donné i, il est nécessaire de prendre en compte la corrélation entre les erreurs d'échantillonnage e_{it} et e_{is} $(t \neq s)$. You et coll. (2000, 2003) ont proposé un modèle transversal et chronologique pour estimer les taux de chômage d'après l'EPA. Ils n'ont utilisé que les données recueillies au cours des six mois précédents pour prédire le taux du mois courant, puisque le renouvellement de l'échantillon de l'EPA est fondé sur un cycle de six mois. Chaque mois, le sixième de l'échantillon de l'EPA est remplacé. Donc, après six mois, la corrélation entre les estimations est faible (voir la section 2.1 pour les coefficients de corrélation retardés). Soit $y_i = (y_{i1}, \ldots, y_{iT})'$, $\theta_i = (\theta_{i1}, \ldots, \theta_{iT})'$, et $e_i = (e_{i1}, \ldots, e_{iT})'$, où $T = 6$ ici. En supposant que e_i suit une loi normale multivariée de vecteur moyenne 0 et de matrice de covariance d'échantillonnage $\sum_i$, nous avons

$$y_i \sim N\left(\theta_i, \sum_i\right), \quad i = 1, \ldots, m.$$

Donc nous supposons que y_i est sans biais par rapport au plan pour θ_i. La matrice de covariance d'échantillonnage $\sum_i$ est inconnue dans le modèle. Des estimations directes des matrices de covariance d'échantillonnage sont disponibles. Dans le contexte de l'estimation sur petits domaines fondée sur un modèle au niveau du domaine, on suppose habituellement que la variance d'échantillonnage est connue (Rao 2003). Par exemple, le modèle classique de Fay-Herriot repose sur l'hypothèse que la variance d'échantillonnage est connue dans le modèle. Habituellement, on utilise un estimateur lissé de la variance d'échantillonnage. Toutefois, des progrès récents dans le domaine de la modélisation de la variance d'échantillonnage offrent une autre approche pour aborder le problème; par exemple, voir Wang et Fuller (2003), You et Dick (2004), ainsi que You et Chapman (2006). En ce qui concerne l'estimation des taux de chômage, nous fournissons des précisions sur le lissage et la modélisation des variances d'échantillonnage à la section 3.

Afin d'emprunter de l'information aux diverses régions et périodes, et en nous inspirant de You et coll. (2000, 2003), nous pouvons modéliser le taux réel de chômage θ_{it} par un modèle de régression linéaire avec effets aléatoires au moyen de variables auxiliaires x_{it}, c'est-à-dire

$$\theta_{it} = x'_{it}\beta + v_i + u_{it}, \quad i = 1, \ldots, m, \; t = 1, \ldots, T, \qquad (3)$$

où v_i est un effet aléatoire de domaine que nous supposons suivre une loi $N(0, \sigma_v^2)$ et u_{it} est une composante temporelle (période) et spatiale (domaine) aléatoire. Nous pouvons en outre supposer que u_{it} suit un processus de marche aléatoire sur les périodes $t = 1, \ldots, T$, autrement dit que

$$u_{it} = u_{i,\,t-1} + \varepsilon_{it}, \qquad (4)$$

où $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$. Alors, $\text{cov}(u_{it}, u_{is}) = \min(t, s)\,\sigma_\varepsilon^2$. Le vecteur de régression β et les composantes de la variance σ_v^2 et σ_ε^2 sont inconnus dans le modèle et doivent être estimés. En combinant les modèles (1), (3) et (4), nous obtenons un modèle mixte linéaire avec composantes temporelles de la forme

$$y_{it} = x'_{it}\beta + v_i + u_{it} + e_{it}, \quad i = 1, \ldots, m, \; t = 1, \ldots, T. \qquad (5)$$

You et coll. (2003) ont montré que le modèle transversal et chronologique (5) est supérieur au modèle de Fay-Herriot en ce qui concerne le lissage des estimations directes et la réduction des coefficients de variation des estimations directes pour l'estimation du taux de chômage d'après l'EPA.

Nous avons utilisé un modèle à marche aléatoire pour u_{it}. Rao et Yu (1994) ont utilisé un modèle autorégressif stationnaire pour u_{it}. You et coll. (2003) ont montré que le modèle à marche aléatoire sur u_{it} fournissait un modèle mieux ajusté à l'estimation du taux de chômage que le modèle autorégressif AR(1). Datta et coll. (1999) ont également utilisé un modèle à marche aléatoire pour estimer les taux de chômage au niveau de l'État aux États-Unis.

2.3 Modèle de lien loglinéaire

Cependant, l'une des limites du modèle (3) est due au fait que le modèle de lien pour le paramètre d'intérêt, c'est-à-dire le taux réel de chômage θ_{it}, est un modèle linéaire avec effets aléatoires normaux. Puisque θ_{it} est un nombre positif compris entre 0 et 1, et qu'il est proche de 0, le modèle de lien linéaire avec effets aléatoires normaux pourrait introduire des estimations négatives de θ_{it} pour certains petits domaines. Afin d'éviter ce problème, You, Chen et Gambino (2002) ont proposé un modèle de lien loglinéaire pour θ_{it} de la forme suivante :

$$\log(\theta_{it}) = x'_{it}\beta + v_i + u_{it}, \quad i = 1, \ldots, m, \; t = 1, \ldots, T. \qquad (6)$$

You et Rao (2002) ont également étudié le modèle de lien loglinéaire pour le modèle de Fay-Herriot en tant que modèles non appariés d'échantillonnage et de lien avec application à l'estimation du sous-dénombrement au recensement canadien. Les résultats de You et Rao (2002) et de You et coll. (2002) montrent que le modèle de lien loglinéaire donne de fort bons résultats lorsqu'il est appliqué à des problèmes d'estimation sur petits domaines. Ici, nous utiliserons par conséquent le modèle de lien loglinéaire (6) pour estimer le taux réel de chômage θ_{it}.

3. Variance d'échantillonnage

En général, nous pouvons obtenir des estimations directes de la variance d'échantillonnage d'après les données d'enquête. Toutefois, ces estimations directes sont instables si les tailles d'échantillon sont faibles. Dans les modèles au niveau du domaine pour l'estimation sur petits domaines, on suppose habituellement que les variances d'échantillonnage sont connues (par exemple, Fay et Herriot 1979; Datta et coll. 1999; You et Rao 2002). Sous cette hypothèse, des estimations fiables (lissées) des variances d'échantillonnage sont construites en utilisant d'autres données et modèles auxiliaires, habituellement en se servant de fonctions de variance généralisées (par exemple, Dick 1995; Datta et coll. 1999). Alternativement, dans le présent article, nous modélisons la matrice de variance-covariance d'échantillonnage en nous servant des estimations directes de manière particulière qui rend non nécessaire l'hypothèse que les variances et les covariances d'échantillonnage sont connues dans le modèle. Donc, nous simplifions le problème du lissage de la variance d'échantillonnage inconnue et intégrons la modélisation de la variance d'échantillonnage dans le modèle complet.

3.1 Lissage de la matrice de covariance d'échantillonnage

You et coll. (2000, 2003) ont suivi deux étapes pour lisser la matrice de covariance d'échantillonnage. La première consiste à obtenir un coefficient de variation lissé ou constant en calculant pour chaque RMR/CU le coefficient de variation moyen sur une certaine période, que nous dénotons $\overline{\mathrm{CV}}_i$, où $i = 1, 2, ..., m$. La deuxième étape consiste à obtenir le coefficient de corrélation avec décalage moyen sur le temps et toutes les RMR/CU, dénoté $\overline{\rho}_{|t-s|}$ pour le décalage temporel $|t-s|$. Cette étape requiert d'importantes ressources de calcul. Nous avons utilisé des données de l'EPA couvrant une période de trois années (1999 à 2001) pour calculer les coefficients de corrélation lissés. Dans le modèle, nous traitons les valeurs lissées temporellement et spatialement comme étant les valeurs réelles. Nous obtenons pour le coefficient de corrélation avec décalage d'un mois (décalage -1) la valeur $\overline{\rho}_1 = 0,48$, pour le coefficient de corrélation avec décalage -2, la valeur $\overline{\rho}_2 = 0,31$, pour le décalage -3, la valeur $\overline{\rho}_3 = 0,21$, pour le décalage -4, la valeur $\overline{\rho}_4 = 0,16$, pour le décalage -5, la valeur $\overline{\rho}_5 = 0,11$ et pour le décalage -6, $\overline{\rho}_6 = 0,1$. Après décalage -6, la valeur est inférieure à 0,1. Les coefficients de corrélation avec décalage diminuent à mesure qu'augmente le décalage, ce qui est en harmonie avec le processus de renouvellement de l'échantillon intégré dans le plan de sondage de l'EPA. La figure 1 montre les coefficients de corrélation avec décalage lissés pour les estimations du taux de chômage d'après l'EPA.

En utilisant ces coefficients de variation lissés et les coefficients de corrélation avec décalage, nous pouvons obtenir une matrice de covariance lissée $\hat{\Sigma}_i$ ayant pour éléments diagonaux $\hat{\sigma}_{it}^2 = (\overline{\mathrm{CV}}_i)^2 y_{it}^2$ et pour éléments non diagonaux $\hat{\sigma}_{its} = \overline{\rho}_{|t-s|}\hat{\sigma}_{it}\hat{\sigma}_{is}$. Nous traitons ensuite la matrice lissée $\hat{\Sigma}_i$ comme étant connue dans le modèle. L'étude de You et coll. (2000, 2003) donne à penser que l'utilisation de la matrice lissée $\hat{\Sigma}_i$ dans le modèle peut améliorer considérablement les estimations en ce qui concerne la réduction du CV comparativement aux estimations HB obtenues en utilisant dans le modèle les estimations directes par sondage de Σ_i. Pour plus de renseignements sur le résultat, voir You et coll. (2003).

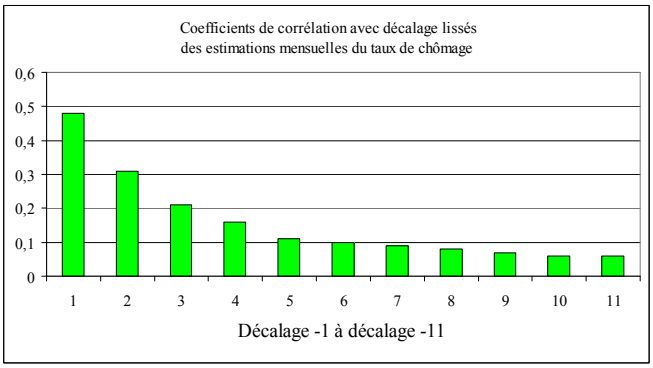

Figure 1 Coefficients de corrélation avec décalage lissés des taux de chômage

3.2 Approche de la modélisation avec CV égaux

Le principal problème de la méthode de You et coll. (2000, 2003) est que les matrices lissées de covariance d'échantillonnage dépendent des estimations directes par sondage y_{it}, alors que ces dernières ne sont pas fiables pour certaines petites régions. Il convient de souligner que la variance d'échantillonnage réelle peut s'écrire sous la forme $\sigma_{it}^2 = \theta_{it}^2(\mathrm{CV}_{it})^2$. Partant de l'hypothèse d'un CV constant au cours du temps pour un domaine donné, You et coll. (2003) ont proposé dans leurs conclusions d'utiliser des estimations de la forme $\tilde{\sigma}_{it}^2 = \theta_{it}^2(\overline{\mathrm{CV}}_i)^2$ et $\tilde{\sigma}_{its} = \overline{\rho}_{|t-s|}(\tilde{\sigma}_{it}\tilde{\sigma}_{is})$ pour les variances et les covariances lissées,

respectivement. Alors, la nouvelle matrice de covariance d'échantillonnage lissée $\tilde{\Sigma}_i$ a pour éléments diagonaux $\tilde{\sigma}_{it}^2$ et pour éléments non diagonaux $\tilde{\sigma}_{its}$. Cependant, sous cette méthode, la matrice de covariance d'échantillonnage $\tilde{\Sigma}_i$ devient inconnue dans le modèle, puisque $\tilde{\sigma}_{it}^2$ et $\tilde{\sigma}_{its}$ dépendent des paramètres inconnus θ_{it}, et que ceux-ci sont reliés à un modèle de lien. L'avantage de cette méthode tient au fait que la structure de la matrice de covariance d'échantillonnage est clairement spécifiée dans le modèle. Elle est meilleure que la méthode de lissage en ce sens que la covariance d'échantillonnage est spécifiée explicitement au lieu d'être considérée comme connue.

3.3 Approche de modélisation avec effets de plan égaux

Une autre approche de modélisation est fondée sur l'hypothèse d'effets de plan constants au cours du temps, comme l'ont proposé Singh, You et Mantel (2005), ainsi que Singh, Folsom et Vaish (2005) pour lisser la variance d'échantillonnage σ_{it}^2. L'effet de plan (deff) pour le i^{e} domaine au temps t peut s'écrire approximativement sous la forme

$$\text{deff}_{it} = \frac{\sigma_{it}^2}{\theta_{it}(1-\theta_{it})/n_{it}},$$

où n_{it} est la taille de l'échantillon correspondant. Alors, la variance d'échantillonnage σ_{it}^2 peut s'écrire sous la forme $\sigma_{it}^2 = \theta_{it}(1-\theta_{it}) \cdot \text{deff}_{it}/n_{it}$. Soit $\tau_{it} = \text{deff}_{it}/n_{it} = \sigma_{it}^2/(\theta_{it}(1-\theta_{it}))$. Alors, nous pouvons estimer τ_{it} en utilisant les estimations directes de θ_{it} et de σ_{it}^2 comme $\hat{\tau}_{it} = \hat{\sigma}_{it}^2/(y_{it}(1-y_{it}))$. Pour chaque domaine, sous l'hypothèse d'un effet de plan constant et d'une taille d'échantillon constante au cours du temps, nous pouvons obtenir un facteur moyen lissé $\bar{\tau}_i$ donné par $\bar{\tau}_i = \sum_{t=1}^{T} \hat{\tau}_{it}/T$. Alors, nous pouvons calculer une variance d'échantillonnage lissée de la forme $\tilde{\sigma}_{it}^2 = \theta_{it}(1-\theta_{it}) \cdot \bar{\tau}_i$, qui de nouveau dépend également de θ_{it}. La covariance d'échantillonnage reste de la forme $\tilde{\sigma}_{its} = \bar{\rho}_{|t-s|}(\tilde{\sigma}_{it}\tilde{\sigma}_{is})$, comme dans You et coll. (2003). Soulignons que $\bar{\tau}_i$ est une moyenne mobile de $\hat{\tau}_{it}$ sur la période T dans le modèle. Toutefois, en pratique, une autre option consiste à utiliser une série chronologique plus longue pour obtenir une estimation plus stable de $\bar{\tau}_i$ pour chaque domaine, au besoin. Ici, nous utiliserons le modèle à effet de plan constant pour l'estimation du taux de chômage basée sur le facteur à moyenne mobile lissé $\bar{\tau}_i$ car nous empruntons de l'information provenant de la période passée T.

4. Inférence bayésienne hiérarchique

À la présente section, nous proposons un modèle loglinéaire transversal et chronologique intégré pour estimer le taux de chômage. Nous appliquons l'approche hiérarchique bayésienne au modèle. Nous obtenons les estimations des moyennes et des variances a posteriori en utilisant la méthode d'échantillonnage de Gibbs.

4.1 Modèle hiérarchique bayésien intégré

Nous proposons maintenant le modèle loglinéaire transversal et chronologique intégré sous un cadre hiérarchique bayésien qui suit :

- Conditionnellement à $\theta_i = (\theta_{i1},...,\theta_{iT})'$, $[y_i|\theta_i] \sim \text{ind } N(\theta_i, \Sigma_i(\theta_i))$;
- Conditionnellement à β, u_{it} et σ_v^2, $[\log(\theta_{it})|\beta, u_{it}, \sigma_v^2] \sim \text{ind } N(x_{it}'\beta + u_{it}, \sigma_v^2)$;
- Conditionnellement à $u_{i,t-1}$ et σ_ε^2, $[u_{it}|u_{i,t-1}, \sigma_\varepsilon^2] \sim \text{ind } N(u_{i,t-1}, \sigma_\varepsilon^2)$;
- $\Sigma_i(\theta_i)$ dépend de θ_i avec les éléments diagonaux $\tilde{\sigma}_{it}^2 = \theta_{it}(1-\theta_{it}) \cdot \bar{\tau}_i$ et les éléments non diagonaux $\tilde{\sigma}_{its} = \bar{\rho}_{|t-s|}(\tilde{\sigma}_{it}\tilde{\sigma}_{is})$.
- Marginalement, β, σ_v^2 et σ_ε^2 sont mutuellement indépendants, leurs priors étant donnés par $\beta \propto 1$, $\sigma_v^2 \sim \text{IG}(a_1, b_1)$, et $\sigma_\varepsilon^2 \sim \text{IG}(a_2, b_2)$, où IG dénote une loi gamma inverse et a_1, b_1, a_2, b_2 sont des constantes positives connues auxquelles nous donnons habituellement une valeur très faible pour refléter nos connaissances vagues au sujet de σ_v^2 et σ_ε^2.

Remarques :

1. Nous utilisons dans le modèle HB proposé un modèle de lien loglinéaire pour le paramètre de petit domaine d'intérêt θ_{it}, comme l'ont suggéré You et coll. (2002) et You et Rao (2002).
2. La matrice de covariance d'échantillonnage Σ_i est inconnue dans le modèle et est spécifiée sous la forme d'une fonction du paramètre de petit domaine inconnu θ_i, comme l'ont suggéré You et Rao (2002) et You et coll. (2003).
3. Nous avons utilisé l'hypothèse d'effets de plan constants pour les petits domaines comme l'ont suggéré Singh, You et Mantel (2005).
4. Le modèle HB proposé surmonte les limites du modèle de You et coll. (2000, 2003) en ce qui concerne la modélisation loglinéaire et la spécification de la modélisation de la matrice de covariance d'échantillonnage inconnue. En particulier, nous modélisons cette matrice à l'aide de paramètres de petit domaine θ_i en utilisant les estimations lissées des effets de plan pour chaque domaine.

Nous souhaitons estimer le taux de chômage réel θ_{it} et, en particulier, le taux de chômage courant θ_{iT}. Dans l'analyse HB, θ_{iT} est estimé par sa moyenne a posteriori

$E(\theta_{iT} \mid y)$ et l'incertitude associée à l'estimateur est mesurée par la variance a posteriori $V(\theta_{iT} \mid y)$. Nous utilisons la méthode d'échantillonnage de Gibbs (Gelfand et Smith 1990; Gelman et Rubin 1992) pour obtenir la moyenne et la variance a posteriori de θ_{iT}.

4.2 Inférence sous échantillonnage de Gibbs

La méthode d'échantillonnage de Gibbs est une méthode itérative d'échantillonnage Monte Carlo par chaînes de Markov utilisée pour simuler le tirage d'échantillons à partir d'une loi conjointe de variables aléatoires en procédant à un échantillonnage à partir de densités sous un espace de dimension réduite pour faire des inférences au sujet des lois conjointe et marginale (Gelfand et Smith 1990). Cette méthode est appliquée principalement à l'inférence dans un cadre bayésien où l'on s'intéresse à la loi a posteriori des paramètres. Supposons que la densité conditionnelle de $y_i \mid \theta$ est $f(y_i \mid \theta)$ pour $i = 1, ..., n$ et que l'information a priori au sujet de $\theta = (\theta_1, ..., \theta_k)'$ est résumée par une densité a priori $\pi(\theta)$. Soit $\pi(\theta \mid y)$ la densité a posteriori de θ sachant les données $y = (y_1, ..., y_n)'$. En pratique, il pourrait être difficile de tirer directement des échantillons à partir de $\pi(\theta \mid y)$, à cause de l'intégration en grande dimension par rapport à θ. Cependant, nous pouvons utiliser l'échantillonneur de Gibbs pour construire une chaîne de Markov $\{\theta^{(g)} = (\theta_1^{(g)}, ..., \theta_k^{(g)})'\}$ avec $\pi(\theta \mid y)$ comme limite. À titre d'illustration, posons que $\theta = (\theta_1, \theta_2)'$. En partant d'un ensemble initial de valeurs $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})'$, nous générons $\theta^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)})'$ par échantillonnage de $\theta_1^{(g)}$ à partir de $\pi(\theta_1 \mid \theta_2^{(g-1)}, y)$ et $\theta_2^{(g)}$ à partir de $\pi(\theta_2 \mid \theta_1^{(g-1)}, y)$. Sous certaines conditions de régularité, $\theta^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)})'$ converge en loi vers $\pi(\theta \mid y)$ quand $g \to \infty$. Si la valeur de g est grande, l'inférence marginale au sujet de $\pi(\theta_i \mid y)$ peut être fondée sur les échantillons marginaux $\{\theta_i^{(g+k)}; k = 1, 2, ...\}$.

Pour les modèles HB intégrés proposés, afin d'obtenir l'estimation a posteriori du taux de chômage, nous mettons en œuvre la méthode d'échantillonnage de Gibbs en produisant des échantillons à partir des lois conditionnelles complètes des paramètres β, σ_v^2 et σ_ε^2, u_{it} et θ_i. Ces lois conditionnelles complètes sont données en annexe. Les lois de β, σ_v^2 et σ_ε^2, u_{it} sont des lois normales standard ou des lois gamma inverses qu'il est facile d'échantillonner. Cependant, la loi conditionnelle de θ_i n'a pas de forme explicite. Nous utilisons l'algorithme de Metropolis-Hastings dans l'échantillonneur de Gibbs (Chib et Greenberg 1995) pour mettre à jour θ_i. À l'instar de You et coll. (2002) et de You et Rao (2002), nous pouvons écrire la loi conditionnelle complète de θ_i dans l'échantillonneur de Gibbs sous la forme

$$\theta_i \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u \propto h(\theta_i) f(\theta_i),$$

où

$$h(\theta_i) = \left|\sum_i(\theta_i)\right|^{-1} \exp\{-\frac{1}{2}(y_i - \theta_i)'\sum_i^{-1}(y_i - \theta_i)\}$$

et

$$f(\theta_i) = \exp\left\{-\frac{1}{2\sigma_v^2}(\log(\theta_i) - \mathrm{x}_i'\beta - u_i)'(\log(\theta_i) - \mathrm{x}_i'\beta - u_i)\right\} \cdot \left(\prod_{t=1}^{T}\frac{1}{\theta_{it}}\right).$$

Pour mettre à jour θ_i, nous procédons comme il suit :

1. Pour $t = 1, ..., T$, tirer $\theta_{it}^{(k+1)} \sim \log N(x_{it}'\beta^{(k+1)} + u_{it}^{(k+1)}, \sigma_v^{2(k+1)})$, ce qui nous donne alors $\theta_i^{(k+1)} = (\theta_{i1}^{(k+1)}, ..., \theta_{iT}^{(k+1)})'$.
2. Calculer la probabilité de rejet

$$\alpha(\theta_i^{(k)}, \theta_i^{(k+1)}) = \min\{\frac{h(\theta_i^{(k+1)})}{h(\theta_i^{(k)})}, 1\}.$$

3. Générer $\lambda \sim \text{Uniforme}(0,1)$, si $\lambda < \alpha(\theta_i^{(k)}, \theta_i^{(k+1)})$, alors accepter $\theta_i^{(k+1)}$; sinon, rejeter $\theta_i^{(k+1)}$ et fixer $\theta_i^{(k+1)} = \theta_i^{(k)}$.

Pour appliquer l'échantillonnage de Gibbs, nous suivons les recommandations de Gelman et Rubin (1992) et nous exécutons indépendamment $L (L > 2)$ chaînes parallèles, chacune de longueur $2d$. Nous supprimons les d premières itérations de chaque chaîne. La surveillance de la convergence est fondée sur le facteur de réduction d'échelle possible proposé dans Gelman et Rubin (1992) et adopté par You et coll. (2003) pour estimer θ_{iT}. Une description détaillée figure dans You et coll. (2003). Les estimations de la moyenne *a posteriori* $E(\theta_{iT} \mid y)$ et de la variance a posteriori $V(\theta_{iT} \mid y)$ sont obtenues en se basant sur les échantillons créés à partir de l'échantillonneur de Gibbs.

5. Application aux données de l'EPA

5.1 Estimation

Nous utilisons les estimations du taux de chômage fondées sur les données de l'EPA recueillies de janvier à juin 2005, y_{it}, dans notre analyse de données. En plus des estimations directes y_{it} et des matrices de covariance d'échantillonnage utilisées dans les modèles pour petits domaines, nous avons besoin de variables auxiliaires administratives dans nos modèles. Pour l'estimation du taux de chômage, nous utilisons comme données auxiliaires les taux de bénéficiaires de l'assurance-emploi (a.-e.) au niveau local. Le taux de bénéficiaires est calculé sous forme du ratio du nombre de personnes faisant la demande de prestations d'assurance-emploi sur le nombre de personnes dans la population active. Le Canada compte 72 RMR/CU.

Un CU (Miramichi) ne possède pas de données sur l'assurance-emploi. Donc, nous considérons dans le modèle $m = 71$ RMR/CU. Dans chaque domaine, nous considérons six estimations mensuelles consécutives y_{it} allant de janvier 2005 à juin 2005, de sorte que $T = 6$. Pour les données recueillies de janvier à juin 2005, le taux de chômage moyen global (sur les 71 RMR/CU et les six mois) est de 0,076 et le taux moyen global de bénéficiaires de l'assurance-emploi est de 0,059. Pour le modèle pour petits domaines proposé, le paramètre d'intérêt θ_{iT} est le taux de chômage réel pour le domaine i en juin 2005, où $i = 1, \ldots, 71$. Pour appliquer l'échantillonneur de Gibbs, nous avons utilisés dix exécutions parallèles comptant chacune 2 000 itérations. Nous avons supprimé les 1 000 premières itérations à titre de période de « rodage ». Les hyperparamètres des composantes de variance incluses dans le modèle sont fixés à 0,0001 pour refléter nos connaissances vagues au sujet de σ_v^2 et de σ_e^2.

Nous présentons maintenant les estimations a posteriori des taux de chômage sous le modèle HB intégré proposé décrit à la section 4.1 en utilisant la méthode d'échantillonnage de Gibbs. La figure 2 donne les estimations directes d'après l'EPA et les estimations fondées sur le modèle HB des taux de chômage en juin 2005 pour les 71 RMR/CU au Canada. Ces 71 RMR/CU figurent par ordre de taille de population, en commençant à gauche par le plus petit CU (Dawson Creek, C.-B.) et en terminant à droite par la plus grande RMR (Toronto, Ont.). Dans le cas des estimations ponctuelles, les estimations HB correspondent à un lissage modéré des estimations directes d'après l'EPA. Pour les RMR ayant les populations les plus grandes et, par conséquent, les tailles d'échantillon les plus grandes, les estimations directes et les estimations HB sont très proches, comme il fallait s'y attendre, particulièrement pour Toronto, Montréal et Vancouver; pour les CU plus petits, l'écart entre les estimations directes et HB est important pour certaines régions.

La figure 3 donne les CV des estimations. Le CV de l'estimation HB correspond au ratio de la racine carrée de la variance a posteriori à la moyenne a posteriori. Il est manifeste, si l'on examine la figure 3, que les CV des estimations directes sont très grands, particulièrement pour les CU dont les CV sont très grands et instables. Comparativement aux estimations directes, les CV des estimations HB sont très petits et stables. Le gain d'efficacité obtenu grâce aux estimations HB est évident, particulièrement pour les CU dont la taille de population est faible. Plus précisément, nous avons calculé la réduction en pourcentage des CV pour les estimateurs HB en nous basant sur les données de juin 2005. La réduction en pourcentage du CV est calculée comme la différence entre le CV direct et le CV HB par rapport au CV direct. La réduction moyenne du CV est de 63 % pour les CU et de 35 % pour les RMR. Comme prévu, le modèle proposé a permis de réduire considérablement le CV comparativement aux estimations directes, particulièrement pour les petits CU dont la taille d'échantillon est faible.

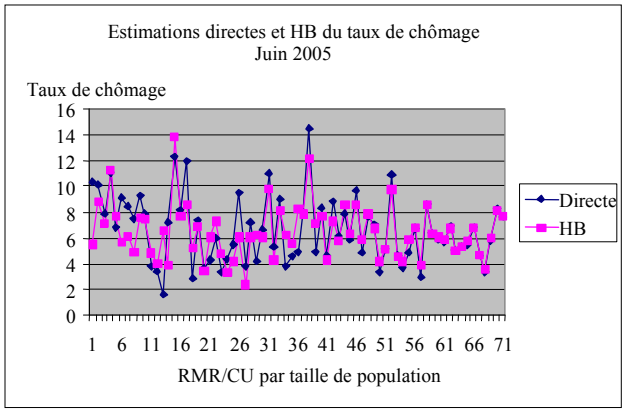

Figure 2 Comparaison des estimations directes et HB

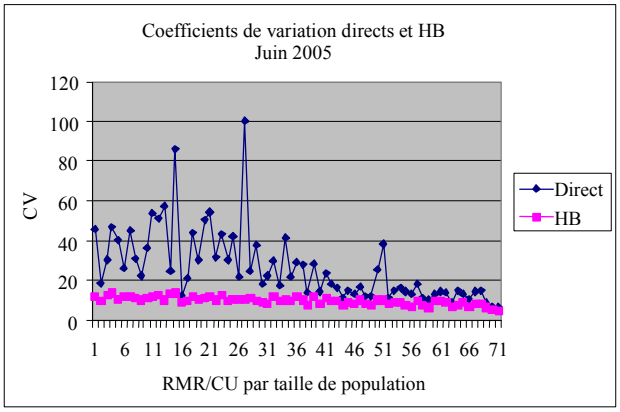

Figure 3 Comparaison des CV directs et HB

5.2 Ajustement du modèle au moyen de la loi prédictive a posteriori

Pour vérifier l'ajustement global du modèle proposé, nous utilisons la méthode de la loi prédictive a posteriori. Soit y_{rep} l'observation répétée sous le modèle. La loi prédictive a posteriori de y_{rep} sachant les données observées y_{obs} est définie comme étant

$$f(y_{\text{rep}} \mid y_{\text{obs}}) = \int f(y_{\text{rep}} \mid \theta) f(\theta \mid y_{\text{obs}}) d\theta.$$

Selon cette approche, nous pouvons définir une mesure de divergence $D(y, \theta)$ qui dépend des données y et du paramètre θ, et comparer la valeur observée $D(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ à la loi prédictive a posteriori de $D(y_{\text{rep}}, \theta \mid y_{\text{obs}})$, tout écart significatif indiquant une défaillance du modèle. Meng (1994), ainsi que Gelman, Carlin, Stern et Rubin (1995) ont proposé la valeur p prédictive a posteriori de la forme

$$p = P(D(y_{\text{rep}}, \theta) \geq D(y_{\text{obs}}, \theta) \mid y_{\text{obs}}).$$

Il s'agit d'une extension naturelle de la valeur p habituelle dans un contexte bayésien. Si un modèle est ajusté aux données observées, les deux valeurs de la mesure de divergence sont semblables. Autrement dit, si le modèle est ajusté correctement aux données observées, $D(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ devrait se situer près de la partie centrale de l'histogramme des valeurs $D(y_{\text{rep}}, \theta \mid y_{\text{obs}})$ si y_{rep} est généré de façon répétée à partir de la loi prédictive a posteriori. Conséquemment, la valeur p prédictive a posteriori devrait s'approcher de 0,5 si le modèle est bien ajusté aux données. Les valeurs p extrêmes (proches de 0 ou de 1) impliquent que l'ajustement n'est pas bon. Nous pouvons estimer la valeur p prédictive a posteriori de la façon suivante. Soit θ^* un tirage à partir de la loi a posteriori $f(\theta \mid y_{\text{obs}})$, et soit y^*_{rep} un tirage à partir de $f(y_{\text{rep}} \mid \theta^*)$. Alors, marginalement, y^*_{rep} est un échantillon provenant de la loi prédictive a posteriori $f(y_{\text{rep}} \mid y_{\text{obs}})$. Il est relativement facile de calculer la valeur p en utilisant les valeurs simulées de θ^* provenant de l'échantillonneur de Gibbs. Pour chaque valeur simulée θ^*, nous pouvons simuler y^*_{rep} à l'aide du modèle et calculer $D(y^*_{\text{rep}}, \theta^*)$ et $D(y_{\text{obs}}, \theta^*)$. Alors, la valeur p est estimée par la proportion de fois que $D(y^*_{\text{rep}}, \theta^*)$ excède $D(y_{\text{obs}}, \theta^*)$.

Pour le modèle HB proposé, la mesure de divergence utilisée pour l'ajustement global est donnée par $d(y, \theta) = \sum_{i=1}^{m} (y_i - \theta_i)' \sum_i^{-1} (y_i - \theta_i)$. Cette mesure a été utilisée par Datta et coll. (1999) et par You et coll. (2003). Nous avons calculé la valeur p en combinant les valeurs simulées de θ^* et y^* provenant des dix exécutions parallèles. Nous avons obtenu une valeur p moyenne estimée d'environ 0,38. Donc, rien n'indique que l'ajustement global du modèle est insuffisant.

Certains ont critiqué la vérification de l'ajustement du modèle à l'aide de la valeur p prédictive a posteriori, la jugeant modérée à cause du double usage des données observées. La double utilisation des données peut induire un comportement non naturel, comme l'ont démontré Bayarri et Berger (2000). Ces auteurs ont proposé deux mesures de la valeur p de rechange pour la vérification du modèle qu'ils ont nommées la valeur p prédictive a posteriori partielle et la valeur p prédictive conditionnelle. Cependant, leurs méthodes sont plus difficiles à appliquer et à interpréter (Rao 2002; Sinharay et Stern 2003). Comme le soulignent Sinharay et Stern (2003), la valeur p prédictive a posteriori est particulièrement utile si nous voyons dans le modèle courant un point final plausible auquel des modifications ne seront apportées que si un manque considérable d'ajustement est constaté.

Pour comparer le modèle proposé au modèle de You et coll. (2003), nous avons calculé la mesure de divergence de Laud et Ibrahim (1995) fondée sur la loi prédictive a posteriori. La mesure de divergence attendue de Laud et Ibrahim (1995) est donnée par $d(y^*, y_{\text{obs}}) = E(k^{-1} \| y^* - y_{\text{obs}} \|^2 \mid y_{\text{obs}})$, où k est la dimension de y_{obs} et y^* est un échantillon tiré de la loi prédictive a posteriori $f(y \mid y_{\text{obs}})$. De deux modèles, nous préférons celui qui donne la valeur la plus faible de cette mesure. Comme Datta, Day et Maiti (1998) et You et coll. (2003), nous avons utilisé, pour approximer la mesure de divergence $d(y^*, y_{\text{obs}})$, les échantillons simulés à partir de la loi prédictive a posteriori. En utilisant les sorties multiples de l'échantillonnage de Gibbs, nous avons obtenu une mesure de divergence de l'ordre de 8 à 9 pour le modèle proposé et d'environ 12 à 14 pour le modèle de You et coll. (2003). Donc, la mesure de divergence donne à penser que le modèle HB intégré proposé pour l'estimation du taux de chômage d'après l'EPA est mieux ajusté.

5.3 Diagnostic du biais par analyse de régression

Afin d'évaluer le biais que pourrait introduire le modèle, nous utilisons une méthode simple d'analyse de régression par les moindres carrés ordinaires pour les estimations directes d'après l'EPA ainsi que pour les estimations fondées sur le modèle HB. La méthode de régression est proposée par Brown, Chamber, Heady et Heasman (2001). Si les estimations fondées sur le modèle sont proches des taux réels de chômage, les estimateurs directs d'après l'EPA devraient se comporter comme des variables aléatoires dont les valeurs prévues correspondent aux valeurs des estimations fondées sur le modèle. Nous représentons graphiquement les estimations HB fondées sur le modèle en abscisse X et les estimations directes d'après l'EPA en ordonnée Y et nous voyons dans quelle mesure la droite de régression s'approche de $Y = X$. En termes de régression, fondamentalement, nous ajustons le modèle de régression $Y = \alpha X$ aux données et estimons le coefficient α. Des estimations fondées sur le modèle dont le biais est faible devraient donner une valeur de α proche de 1. Pour les données de juin 2005, soit Y les estimations directes du taux de chômage et X les estimations HB fondées sur le modèle. Nous obtenons une estimation de la valeur de α de 1,0207 avec une erreur-type de 0,0281. La figure 4 donne le diagramme de dispersion avec la droite de régression ajustée.

Les résultats de la régression ne révèlent aucun écart significatif par rapport à $Y = X$. Par conséquent, nous concluons que les estimations fondées sur le modèle calculées d'après le modèle proposé concordent avec les estimations directes d'après l'EPA sans inclusion d'un biais éventuel supplémentaire. Les résultats pourraient également indiquer qu'il n'existe aucune preuve d'un biais dû à une erreur de spécification possible du modèle.

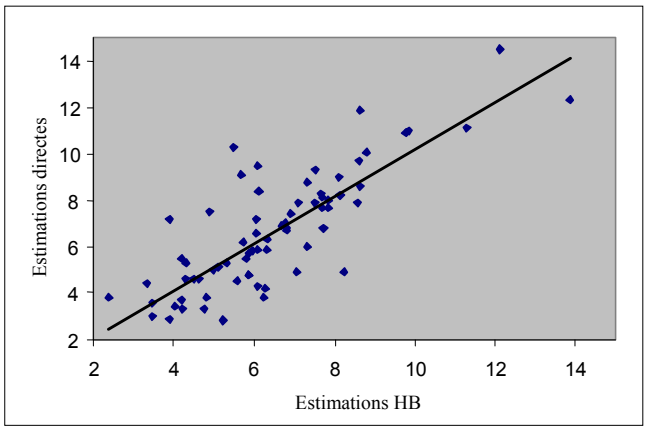

Figure 4 Diagramme de dispersion avec droite de régression

6. Conclusion et futurs travaux

Nous avons passé en revue certains modèles pour petits domaines, y compris le modèle de Fay-Herriot et le modèle transversal et chronologique de You et coll. (2003). Étant donné les travaux antérieurs, nous avons proposé un modèle non linéaire transversal et chronologique intégré en vue d'obtenir des estimations fondées sur un modèle des taux de chômage pour les RMR/CU au Canada en utilisant les données de l'EPA. Le modèle proposé surmonte les limites constatées lors des travaux antérieurs. En particulier, nous pouvons modéliser la variance d'échantillonnage sous forme d'une fonction de la moyenne de petits domaines en émettant l'hypothèse d'un coefficient de variation constant pour un domaine donné ou d'un effet de plan constant pour un domaine donné. Notre analyse des données révèle que le modèle proposé est relativement bien ajusté aux données. Les estimations hiérarchiques bayésiennes fondées sur le modèle améliorent significativement les estimations directes par sondage en ce qui concerne la réduction du coefficient de variation, surtout pour les CU dont la population est faible.

Nous prévoyons adopter l'approche de modélisation de rechange pour la variance d'échantillonnage. Récemment, You et Dick (2004), ainsi que You et Chapman (2006) ont utilisé l'approche HB pour modéliser directement la variance d'échantillonnage, sans spécifier la forme de cette dernière sous le cadre du modèle de Fay-Herriot. Le modèle tient compte automatiquement de la variabilité de l'estimation des variances d'échantillonnage. En particulier, You et Dick (2004) ont appliqué le modèle au problème de l'estimation du sous-dénombrement au recensement et ont obtenu des estimations HB efficaces de ce sous-dénombrement pour divers petits domaines au Canada. Il serait intéressant d'adapter la même idée au modèle transversal et chronologique et de comparer les résultats à ceux des présents travaux. L'objectif de la comparaison est d'établir un modèle fiable et facile à mettre en œuvre de l'estimation des taux de chômage fondés sur un modèle ajusté aux données de l'EPA pour les petits domaines.

Nous prévoyons produire les estimations fondées sur le modèle pour une période relativement longue, par exemple 24 mois allant de 2004 à 2005. Nous comparerons ces estimations aux estimations directes pour les 24 mois, particulièrement pour les grandes RMR, afin d'étudier les effets de lissage du modèle proposé. Les estimations fondées sur le modèle devraient suivre le profil des estimations directes d'après l'EPA pour les grandes RMR, ce qui indiquerait que le lissage des effets de la série temporelle sont raisonnables. Le but est de vérifier la robustesse des estimations fondées sur le modèle proposé au fil du temps.

Annexe

Nous présentons ci-après les lois conditionnelles complètes pour l'échantillonneur de Gibbs sous le modèle HB proposé. Soit $Y = (Y_1', ..., Y_m')'$, $X = (X_1', ..., X_m')'$, $\theta = (\theta_1', ..., \theta_m')'$ et $u = (u_1', ..., u_m')'$, avec $Y_i' = (y_{i1}, ..., y_{iT})$, $X_i' = (x_{i1}, ..., x_{iT})$, $\theta_i' = (\theta_{i1}, ..., \theta_{iT})$, et $u_i' = (u_{i1}, ..., u_{iT})$. Nous obtenons les lois conditionnelles complètes comme il suit :

- $\beta \mid Y, \sigma_v^2, \sigma_\varepsilon^2, u, \theta \sim$
  $N((X'X)^{-1}X'(\log(\theta) - u), \sigma_v^2(X'X)^{-1})$;

- $\sigma_v^2 \mid Y, \beta, \sigma_\varepsilon^2, u, \theta \sim$
  $\text{IG}\left(\left(a_1 + mT/2, b_1 + \sum_{i=1}^{m}\sum_{t=1}^{T}(\log(\theta_{it}) - x_{it}'\beta - u_{it})^2\right)/2\right)$;

- $\sigma_\varepsilon^2 \mid Y, \beta, \sigma_v^2, u, \theta \sim$
  $\text{IG}\left(\left(a_2 + m(T-1)/2, b_2 + \sum_{i=1}^{m}\sum_{t=2}^{T}(u_{it} - u_{i,t-1})^2\right)/2\right)$;

- Pour $i = 1, ..., m$,
  $u_{i1} \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i2}, \theta \sim$
  $N\left(\left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\log(\theta_{i1}) - x_{i1}'\beta}{\sigma_v^2} + \frac{u_{i2}}{\sigma_\varepsilon^2}\right), \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_\varepsilon^2}\right)^{-1}\right)$;

- Pour $i = 1, ..., m$, et $2 \le t \le T-1$,
  $u_{it} \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,t-1}, u_{i,t+1}, \theta \sim$
  $N\left(\left(\frac{1}{\sigma_v^2} + \frac{2}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\log(\theta_{it}) - x_{it}'\beta}{\sigma_v^2} + \frac{u_{i,t-1} + u_{i,t+1}}{\sigma_\varepsilon^2}\right),\right.$
  $\left.\left(\frac{1}{\sigma_v^2} + \frac{2}{\sigma_\varepsilon^2}\right)^{-1}\right)$;

- Pour $i = 1, ..., m$,

$$u_{iT} \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,T-1}, \theta \sim N\left(\left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\log(\theta_{iT})-x_{iT}'\beta}{\sigma_v^2}+\frac{u_{i,T-1}}{\sigma_\varepsilon^2}\right), \left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\right);$$

- Pour $i = 1, ..., m$,

$$\theta_i \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u \propto \left|\sum_i\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\theta_i)'\sum_i^{-1}(y_i-\theta_i)\right\} \times \exp\left\{-\frac{1}{2\sigma_v^2}\sum_{t=1}^{T}(\log(\theta_{it})-x_{it}'\beta-u_{it})^2\right\}\left(\prod_{t=1}^{T}\frac{1}{\theta_{it}}\right).$$

Remerciements

L'auteur remercie le rédacteur en chef, le rédacteur adjoint et un examinateur de leurs commentaires et suggestions. Les travaux ont été financés en partie par le fonds global de financement de la recherche de la Direction de la méthodologie de Statistique Canada.

Bibliographie

Bayarri, M.J., et Berger, J.O. (2000). *P* values for composite null models. *Journal of the American Statistical Association*, 95, 1127-1142.

Brown, G., Chambers, R., Heady, P. et Heasman, D. (2001). Evaluation of small area estimation methods - An application to unemployment estimates from the UK LFS. *Proceedings of Statistics Canada Symposium 2001 Achieving Data Quality in a Statistical Agency*: *A Methodological Perspective*, CD-ROM.

Chib, S., et Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49, 327-335.

Datta, G.S., Day, B. et Maiti, T. (1998). Multivariate Bayesian small area estimation: An application to survey and satellite data. *Sankhyā*, 60, 344-362.

Datta, G.S., Lahiri, P., Maiti, T. et Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Datta, G.S., Lahiri, P. et Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102, 83-97.

Dick, P. (1995). Modélisation du sous-dénombrement net dans le recensement du Canada de 1991. *Techniques d'enquête*, 21, 51-61.

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.

Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). *Methodology of the Canadian Labour Force Survey*, Statistique Canada, Catalogue No. 71-526.

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.

Gelman, A., Carlin, J.B., Stern, H.S. et Rubin, D.B. (1995). *Bayesian Data Analysis.* Londres : Chapman and Hall.

Gelman, A., et Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.

Ghosh, M., Nangia, N. et Kim, D.H. (1996). Estimation of median income of four-person families: A Bayesian time series approach. *Journal of the American Statistical Association*, 91, 1423-1431.

Laud, P., et Ibrahim, J. (1995). Predictive model selection. *Journal of Royal Statistical Society*, Séries B, 57, 247-262.

Meng, X.L. (1994). Posterior predictive *p* value. *The Annals of Statistics*, 22, 1142-1160.

Rao, J.N.K., et Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics*, 22, 511-528.

Rao, J.N.K. (2003). *Small Area Estimation*. New York : John Wiley & Sons, Inc.

Singh, A.C., Folsom, R.E., Jr. et Vaish, A.K. (2005). Small area modeling for survey data with smoothed error covariance structure via generalized design effects. *Federal Committee on Statistical methods Conference proceedings*, *Washington*, *D.C.*, *www.fcsm.gov*.

Singh, A., You, Y. et Mantel, H. (2005). Use of generalized design effects for variance function modeling in small area estimation from survey data. Présentation au 2005 Statistical Society of Canada Annual Meeting, Regina, SK.

Sinharay, S., et Stern, H.S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference*, 111, 209-221.

Wang, J., et Fuller, W.A. (2003). The mean square error of small area predictors constructed with estimated area 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, 107-114.

You. Y., Chen, E. et Gambino, G. (2002). Nonlinear mixed effects cross-sectional and time series models for unemployment rate estimation. *2002 Proceedings of the American Statistical Association, Section on Government Statistics*, Alexandria, VA : American Statistical Association. 3883-3888.

You, Y., et Dick, P. (2004). Hierarchical Bayes small area inference to the 2001 census undercoverage estimation. *2004 Proceedings of the American Statistical Association*, *Section on Government Statistics*, Alexandria, VA : American Statistical Association, 1836-1840.

You, Y., et Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30, 3-15.

You, Y., Rao, J.N.K. et Gambino, J. (2000). Hierarchical Bayes estimation of unemployment rates for sub-provincial regions using cross-sectional and time series data. *American Statistical Association 2000 Proceedings of the Section on Government Statistics and Section on Social Statistics*, 160-165.

You, Y., Rao, J.N.K. et Gambino, J. (2003). Estimation du taux de chômage fondée sur un modèle pour l'Enquête sur la population active du Canada : une approche bayésienne hiérarchique. *Techniques d'enquête*, 29, 27-36.