Estimation du maximum de vraisemblance avec contrainte (MVC) dans le lissage des taux de sous-dénombrement du recensement selon l'approche empirique de Bayes

NOEL CRESSIE¹

RÉSUMÉ

Une façon de calculer le sous-dénombrement au niveau infra-national (par ex.: pour un État) est de prendre des données-échantillon d'une enquête postcensitaire et de les lisser suivant un modèle linéaire de variables explicatives. Le rapport entre la variance de l'erreur d'échantillonnage et la variance de l'erreur de modèle correspondante détermine le degré de lissage. L'estimation par la méthode du maximum de vraisemblance peut mener à un lissage excessif et, par conséquent, rendre le calcul du sous-dénombrement trop tributaire du modèle linéaire. Les estimateurs du maximum de vraisemblance avec contrainte (MVC) ne présentent pas de tels inconvénients. Dans cet article, on traite la prévision empirique de Bayes du sous-dénombrement fondée sur l'estimation MVC et on la compare, par des exemples et des simulations, à celle fondée sur la méthode du maximum de vraisemblance et à celle fondée sur une méthode des moments. Les propriétés de distribution pour grand échantillon des estimateurs MVC permettent un calcul précis de l'erreur quadratique moyenne de prévision des filtres de lissage fondés sur l'estimation MVC.

MOTS CLÉS: Modèle linéaire; maximum de vraisemblance; maximum de vraisemblance avec contrainte; composantes de la variance.

1. INTRODUCTION

Malgré tous les efforts que l'on déploie dans un recensement pour dénombrer toute la population, les chiffres finals sont inexacts pour une multitude de raisons. Le personnel affecté au recensement – depuis le directeur de la division jusqu'aux milliers d'employés temporaires engagés comme recenseurs – est chargé d'une tâche gigantesque dont le succès dépend de la capacité de chacun d'accomplir son travail à la perfection.

En outre, il faut espérer que les manifestations de phénomènes naturels incontrôlables (par ex.: conditions atmosphériques, catastrophes naturelles) n'auront pas de conséquences inattendues sur les résultats. De toute évidence, dans un pays de la taille des É.-U. (en population comme en superficie), beaucoup de facteurs peuvent contribuer à fausser les chiffres du recensement. Cependant, la taille n'est pas le seul problème; l'hétérogénéité démographique et géographique peut fausser les chiffres de différentes façons.

Les inexactitudes s'expriment normalement par un sous-dénombrement, de sorte qu'une valeur négative signifie un surdénombrement. Supposons que les É.-U. sont divisés en i régions, $i=1,\ldots,n$, (par ex.: les 50 États plus le District de Columbia). Pour la région i, soit T_i l'effectif réel (inconnu) de la population et C_i l'effectif recensé. Alors, le sous-dénombrement, exprimé en pourcentage de l'effectif réel, est défini par l'expression:

$$U_i \equiv \{ (T_i - C_i)/T_i \} 100. \tag{1.1}$$

Les différences de taux de sous-dénombrement posent un problème sérieux lorsque les chiffres du recensement doivent servir à la répartition du pouvoir et des ressources monétaires entre les régions et les sous-régions. (Pour un examen détaillé de la question, se référer à Ericksen

¹ Noel Cressie, Département de statistique, Université Iowa State, Ames, (IA), U.S.A. 50011.

et Kadane 1985, Freedman et Navidi 1986 et Cressie 1988.) Des États comme la Californie, le Texas et New York profiteraient largement d'un redressement en fonction du sous-dénombrement, c.-à-d. de la substitution de F_iC_i à C_i où F_i est un facteur de redressement. Le facteur de redressement approprié est

$$F_i = T_i/C_i, (1.2)$$

et ce facteur est lié au sous-dénombrement par la relation suivante,

$$F_i = \{1 - U_i/100\}^{-1}.$$

L'équation (1.2) se prête mal à un redressement car on ne connaît pas l'effectif réel de la population, T_i . Afin d'obtenir de l'information supplémentaire qui permettra d'estimer F_i , le Census Bureau des É.-U. effectue une enquête post-censitaire (EP) qui permet de savoir si certaines personnes, en l'occurrence celles incluses dans l'échantillon de l'EP, ont été recensées ou non (voir, par exemple, Wolter 1986). L'enquête implique la participation de plusieurs centaines de milliers de ménages et produit des facteurs de redressement "bruts" $\{Y_i: i=1,\ldots,n\}$ qui doivent être lissés.

Supposons que, étant donné F_i ,

$$Y_i \sim \operatorname{Gau}(F_i, \delta_i^2),$$
 (1.3)

c'est-à-dire que Y_i , moyennant F_i , suit une distribution gaussienne de moyenne F_i et de variance δ_i^2 . Si on pose en plus l'hypothèse d'indépendance, on obtient

$$Y \sim \operatorname{Gau}(F,\Delta),$$
 (1.4)

où $Y = (Y_1, \ldots, Y_n)'$, $F = (F_1, \ldots, F_n)'$ et Δ est la matrice diagonale $n \times n$ diag $\{\delta_1^2, \ldots, \delta_n^2\}$.

Supposons maintenant que

$$F \sim \text{Gau}(X\beta,\Gamma(\tau^2)),$$
 (1.5)

où X est une matrice $n \times p$ de variables explicatives, β est un vecteur $p \times 1$ de coefficients (inconnus) du modèle linéaire, $\Gamma(\tau^2)$ est une matrice diagonale $n \times n$:

$$\Gamma(\tau^2) = \tau^2 D \tag{1.6}$$

et $D = \text{diag}\{1/C_1, \ldots, 1/C_n\}$. Le modèle hétéroscédastique défini par les expressions (1.5) et (1.6) est analysé en profondeur dans Cressie (1990). Intuitivement, il est raisonnable de penser que dans le cas des régions à forte population, le facteur de redressement a une variance moins élevée; Cressie (1989) en fait la démonstration tant d'un point de vue bayesien que "fréquentiste".

Le modèle défini par les expressions (1.4) et (1.5) peut aussi s'écrire:

$$\underline{Y} = X\beta + y + \epsilon, \tag{1.7}$$

où les vecteurs $n \times 1$ \underline{y} et $\underline{\varepsilon}$ sont statistiquement indépendants, $\underline{y} \sim \text{Gau}(0,\Gamma(\tau^2))$ et $\underline{\varepsilon} \sim \text{Gau}(0,\Delta)$. Si nous supposons maintenant que les valeurs $\delta_1^2, \ldots, \delta_n^2$ sont calculées à l'aide de formules de la variance d'échantillonnage adaptées à la base de sondage de l'EP, il ne reste plus que les paramètres $\underline{\beta}$ et τ^2 à estimer. Par conséquent, les deux composantes de

variance Δ et $\Gamma(\tau^2)$ ne renferment qu'un seul paramètre inconnu, soit τ^2 . Il est utile de souligner ici que les méthodes élaborées dans cet article peuvent être facilement généralisées au delà de ce simple problème de composantes de la variance. Le modèle linéaire général est considéré dans la section 3.

Dans la section 2, nous définissons le prédicteur de Bayes et le prédicteur empirique de Bayes de F. L'estimation de g est simple mais il y a plusieurs manières d'estimer r^2 . Dans la section 3, nous présentons trois méthodes d'estimation particulières: la méthode du maximum de vraisemblance (m.v.), la méthode des moments et la méthode du maximum de vraisemblance avec contrainte (MVC). Dans la section suivante, nous analysons l'effet de l'estimation de r^2 sur l'erreur quadratique moyenne de prévision. Enfin, nous comparons les diverses méthodes par une simulation et un exemple dans la section 5 et présentons nos conclusions et une analyse dans la section 6.

2. PRÉDICTION EMPIRIQUE DE BAYES

Dans cet article, l'effectif réel de la population d'une petite région est réputé inconnu. Une fois que l'effectif recensé est connu, on révise le degré d'incertitude qui avait été associé à l'origine au "chiffre réel de population". Les modèles statistiques du sous-dénombrement **dépendent** donc des chiffres du recensement. Le modèle que nous avons défini dans la section 1 au moyen des expressions (1.4), (1.5) et (1.6) servira dans les sections 2, 3 et 4.

En utilisant un équivalent matriciel de la fonction quadratique de perte, on a pour prédicteur optimal $E(F \mid Y)$ (Cressie 1990), qui est,

$$\underline{p}^{*}(\underline{Y}) = \Gamma(\tau^{2}) (\Delta + \Gamma(\tau^{2}))^{-1} \underline{Y} + \{I - \Gamma(\tau^{2}) (\Delta + \Gamma(\tau^{2}))^{-1}\} X \beta$$
 (2.1)

et la matrice des erreurs quadratiques moyennes de prévision est,

$$E\{(\bar{F} - p^*(\bar{Y}))(\bar{F} - p^*(\bar{Y}))'\} = \{I - \Gamma(\tau^2)(\Delta + \Gamma(\tau^2))^{-1}\}\Gamma(\tau^2). \tag{2.2}$$

Pour ce qui a trait à la matrice de perte, $L(F,p) \equiv (F-p)(F-p)'$, on voit facilement que (2.1) est un prédicteur de **Bayes** de F. De fait, $g \in T^2$ sont inconnus et l'expression (2.1) n'est donc pas une statistique (c'est-à-dire qu'elle n'est pas strictement une fonction des observations). Selon l'approche bayesienne, il serait indiqué ici de définir d'autres distributions et hyperdistributions a priori pour tous les paramètres inconnus. (Cette solution au problème des paramètres inconnus est parfois appelée méthode hiérarchique de Bayes et exige une connaissance préalable de la variabilité de processus que de nombreux scientifiques estiment ne pas avoir. Néanmoins, on peut souvent obtenir des estimateurs raisonnables avec les distributions et hyperdistributions a priori non informatives.) Il arrive souvent que les distributions a posteriori posent des difficultés insolubles sur le plan analytique. Si le modèle et la distribution a priori étaient définis en fonction de leurs distributions conditionnelles, on pourrait se servir du modèle de Gibbs pour obtenir numériquement toutes les distributions marginales et les distributions conjointes voulues (voir, par exemple, Gelfand et Smith 1990).

La solution proposée dans cet article consiste à tenir tous les paramètres, sauf F, pour fixes mais inconnus et à utiliser les observations Y pour estimer ces paramètres. C'est ce qu'on appelle la méthode **empirique de Bayes**. Bien que nous supposions dans cet article une distribution a priori paramétrique (conjuguée), nous pourrions aussi bien utiliser une distribution a priori non paramétrique (voir, par exemple, Laird et Louis 1987).

Supposons maintenant que β est inconnu mais que τ^2 dans (1.6) est (pour l'instant) connu. En utilisant encore une fois l'équivalent matriciel de la fonction quadratique de perte, on obtient le prédicteur linéaire non biaisé optimal en substituant l'estimateur par les moindres carrés généralisés:

$$\hat{\beta} = \{X'(\Delta + \Gamma(\tau^2))^{-1}X\}^{-1}X'(\Delta + \Gamma(\tau^2))^{-1}Y$$

dans (2.1), ce qui donne

$$\hat{p}(Y; \tau^{2}) = \Gamma(\tau^{2}) (\Delta + \Gamma(\tau^{2}))^{-1} Y + \{I - \Gamma(\tau^{2}) (\Delta + \Gamma(\tau^{2}))^{-1}\}$$

$$X\{X'(\Delta + \Gamma(\tau^{2}))^{-1}X\}^{-1} X'(\Delta + \Gamma(\tau^{2}))^{-1} Y \equiv \Lambda(\tau^{2}) Y$$
(2.3)

(Cressie 1990). La matrice des erreurs quadratiques moyennes de prévision est,

$$M_{1}(\tau^{2}) \equiv E\{ (F - \hat{p}(Y; \tau^{2})) (F - \hat{p}(Y; \tau^{2}))' \}$$

$$= \Lambda(\tau^{2}) \Delta \Lambda(\tau^{2})' + (\Lambda(\tau^{2}) - I) \Gamma(\tau^{2}) (\Lambda(\tau^{2}) - I)'. \tag{2.4}$$

De manière plus réaliste, τ^2 est aussi inconnu. On obtient un **prédicteur empirique de Bayes** en substituant un estimateur $\hat{\tau}^2$ dans $\Lambda(\tau^2)$, ce qui donne

$$\hat{p}(\underline{Y}; \hat{\tau}^2) = \Lambda(\hat{\tau}^2)\underline{Y}. \tag{2.5}$$

On voit facilement que lorsque $\hat{\tau}^2$ est l'estimateur du maximum de vraisemblance de τ^2 , l'équation (2.5) est l'estimateur du maximum de vraisemblance du prédicteur de Bayes.

Le prédicteur (2.5) a été proposé par Ericksen et Kadane (1985), puis critiqué par Freedman et Navidi (1986). À propos, les prédicteurs de Freedman et Navidi peuvent sembler différents des prédicteurs (2.1), (2.3) et (2.5) mais en réalité, ils leur sont identiques si l'on tient compte de l'équation $A(A + B)^{-1}B = (A^{-1} + B^{-1})^{-1}$, où A et B sont des matrices carrées telles que A, B et A + B ont des inverses.

En substituant $\hat{\tau}^2$ dans (2.4), on obtient un estimateur de la matrice des erreurs quadratiques moyennes de prévision:

$$M_1(\hat{\tau}^2) = \Lambda(\hat{\tau}^2) \Delta \Lambda(\hat{\tau}^2)' + (\Lambda(\hat{\tau}^2) - I) \Gamma(\hat{\tau}^2) (\Lambda(\hat{\tau}^2) - I)'. \tag{2.6}$$

Comme (2.6) ne tient pas compte de l'estimation de τ^2 dans $\hat{p}(\underline{Y}; \hat{\tau}^2)$, il risque d'être un estimateur biaisé de $E\{(\underline{F} - \hat{p}(\underline{Y}; \hat{\tau}^2))(\underline{F} - \hat{p}(\underline{Y}; \hat{\tau}^2))'\}$. Nous traitons plus en détail cette importante question dans la section 4.

Une fois calculés $\hat{\beta}$ et $\hat{\tau}^2$, on peut faire un diagnostic de modèle pour vérifier l'ajustement du modèle estimé. Par exemple, un graphique quantile-quantile mettant en relation les résidus normalisés $(\Delta + \Gamma(\hat{\tau}^2))^{-1/2}(Y - X\hat{\beta})$ et les statistiques d'ordre théoriques d'une distribution gaussienne d'unités n'a révélé aucun manque d'ajustement apparent pour le modèle utilisé dans la section 5. Le diagnostic de modèle est traité plus en détail dans la section 6.

3. ESTIMATION DES PARAMÈTRES DE LA MATRICE DE VARIANCES

Dans cette section, nous suppposerons le modèle linéaire général,

$$\underline{Y} \sim \operatorname{Gau}(X\beta, \Sigma(\gamma)),$$
 (3.1)

où γ est le vecteur $k \times 1$ des paramètres de la matrice de variances. En particulier, le modèle défini par (1.4), (1.5) et (1.6) donne,

$$\Sigma (\gamma) = \Delta + \Gamma(\tau^2), \tag{3.2}$$

où γ consiste en un seul paramètre, τ^2 .

Si γ est connu, on peut estimer facilement β :

$$\hat{\mathcal{G}}(\gamma) \equiv (X' \sum (\gamma)^{-1} X)^{-1} X' \sum (\gamma)^{-1} \underline{Y}. \tag{3.3}$$

Dans la réalité toutefois, γ est inconnu et il faut l'estimer; en substituant son estimateur dans (3.3), on obtient un estimateur par les moindres carrés généralisés estimés de β . Dans le reste de cette section, nous examinons trois méthodes d'estimation de γ .

3.1 Estimation par la méthode du maximum de vraisemblance

La fonction de vraisemblance logarithmique négative de β et γ est:

$$L(\underline{\beta}, \underline{\gamma}) = (n/2)\log(2\pi) + (\frac{1}{2})\log(|\sum (\underline{\gamma})|) + (\frac{1}{2})(\underline{Y} - X\underline{\beta})' \sum (\underline{\gamma})^{-1}(\underline{Y} - X\underline{\beta}).$$
(3.4)

En minimisant cette fonction, on obtient les estimations les plus vraisemblables $\hat{\beta}_{m\ell}$ et $\hat{\gamma}_{m\ell}$. La partie délicate de cette minimisation est de déterminer $\hat{\gamma}_{m\ell}$. L'algorithme de Gauss-Newton (ou de caractérisation) est décrit notamment dans Harville (1977) et dans Mardia et Marshall (1984), et nous le reproduisons ici dans tous ses détails.

Définissons,

$$\sum_{i} (\underline{\gamma}) \equiv \partial \sum_{i} (\underline{\gamma}) / \partial \gamma_{i}; i = 1, ..., k,$$

$$\sum_{i} (\underline{\gamma}) \equiv \partial \sum_{i} (\underline{\gamma}) / \partial \gamma_{i} = -\sum_{i} (\underline{\gamma})^{-1} \sum_{i} (\underline{\gamma}) \sum_{i} (\underline{\gamma})^{-1}; i = 1, ..., k,$$
(3.5)

le vecteur L_{γ} , de dimension $k \times 1$, dont l'élément i est défini:

$$(\underline{L}_{\gamma})_{i} \equiv (\frac{1}{2})\operatorname{tr}(\sum (\gamma)^{-1}\sum_{i}(\gamma)) + (\frac{1}{2})(\underline{Y} - X\underline{\beta})'\sum_{i}(\gamma)(\underline{Y} - X\underline{\beta}), \tag{3.6}$$

et la matrice J_{γ} , de dimension $k \times k$, dont l'élément (i,j) est défini:

$$(J_{\gamma})_{ij} \equiv (\frac{1}{2}) \operatorname{tr} \left(\sum_{i} (\underline{\gamma})^{-1} \sum_{i} (\underline{\gamma}) \sum_{i} (\underline{\gamma})^{-1} \sum_{j} (\underline{\gamma}) \right). \tag{3.7}$$

Alors,

$$\gamma^{(\ell+1)} = \gamma^{(\ell)} - (J_{\gamma}^{(\ell)})^{-1} \underline{L}_{\gamma}^{(\ell)}, \tag{3.8}$$

où $J_{\gamma}^{(\ell)}$ et $\underline{L}_{\gamma}^{(\ell)}$ désignent respectivement J_{γ} et \underline{L}_{γ} , évalués à $\underline{\gamma} = \underline{\gamma}^{(\ell)}$ et à $\underline{\beta} = \hat{\underline{\beta}}(\underline{\gamma}^{(\ell)})$.

Lorsque γ consiste uniquement en τ^2 dans (1.6), l'algorithme (3.8) est particulièrement simple. Dans la simulation et l'exemple présentés dans la section 5, nous avons utilisé la valeur de départ

$$(\tau^{2})^{(0)} = \{1/(n-p)\} (\underline{Y} - X(X'D^{-1}X)^{-1}X'D^{-1}\underline{Y})'D^{-1}$$

$$(Y - X(X'D^{-1}X)^{-1}X'D^{-1}Y).$$
(3.9)

Alors (3.8) est,

$$(\tau^{2})^{(\ell+1)} = (\tau^{2})^{(\ell)} - \{(\frac{1}{2}) \sum_{i=1}^{n} 1/(C_{i}\delta_{i}^{2} + (\tau^{2})^{(\ell)})^{2}\}^{-1}L_{\tau}^{(\ell)}; \ell = 0, 1, \dots$$
 (3.10)

où

$$L_{\tau}^{(\ell)} = (\frac{1}{2}) \sum_{i=1}^{n} \frac{1}{(C_{i}\delta_{i}^{2} + (\tau^{2})^{(\ell)})}$$

$$- (\frac{1}{2}) \{ Y - X \hat{\beta} ((\tau^2)^{(\ell)}) \} ' \operatorname{diag} \{ C_i / (C_i \delta_i^2 + (\tau^2)^{(\ell)})^2 \} \{ Y - X \hat{\beta} ((\tau^2)^{(\ell)}) \}.$$
 (3.11)

En itérant (3.8) jusqu'à convergence, on obtient l'estimateur du maximum de vraisemblance $\hat{\gamma}_{m\ell}$, qui, une fois substitué dans l'équation (3.3), donne l'estimateur du maximum de vraisemblance $\hat{\beta}(\hat{\gamma}_{m\ell})$. Suivant des conditions de régularité appropriées (voir, par ex. Mardia et Marshall 1984), $(\hat{\beta}(\hat{\gamma}_{m\ell})', \hat{\gamma}'_{m\ell})'$ suit approximativement une distribution gaussienne multidimensionnelle qui a pour moyenne $(\beta', \gamma')'$ et pour matrice de variances asymptotiques

$$\begin{bmatrix} (X' \sum (\gamma)^{-1} X)^{-1} & 0 \\ 0 & J_{\gamma}^{-1} \end{bmatrix};$$
 (3.12)

lorsque γ consiste uniquement en τ^2 dans (1.6), la matrice (3.12) devient,

$$\begin{bmatrix} (X' \sum (\tau^2)^{-1} X)^{-1} & 0 \\ 0 & \left\{ (\frac{1}{2}) \sum_{i=1}^{n} \frac{1}{(C_i \delta_i^2 + \tau^2)^2} \right\}^{-1} \end{bmatrix}.$$
 (3.13)

En pratique, on obtient les variances et les covariances estimées en évaluant (3.12) en fonction de l'estimation la plus vraisemblable $\hat{\gamma}_{m\ell}$.

3.2 Estimation par la méthode des moments

Il n'existe pas d'estimateur par la méthode des moments particulier pour $\tilde{\gamma}$, l'essentiel est d'apparier les moments d'ordre inférieur des observations avec les moments empiriques correspondants. Si on n'utilise que les moments du premier ordre et du deuxième ordre, il est clair que l'hypothèse gaussienne en (3.1) est superflue.

Soit U une matrice symétrique définie positive. Considérons l'estimateur par régression pondéré, $\hat{g}_U \equiv (X'U^{-1}X)^{-1}X'U^{-1}Y$, et les résidus pondérés,

$$\varrho_U \equiv U^{-\frac{1}{2}} (I - X(X'U^{-1}X)^{-1}X'U^{-1}) \tilde{Y}. \tag{3.14}$$

Alors, par des opérations simples de l'algèbre matricielle, nous avons

$$E(\varrho_U'\varrho_U) = \operatorname{tr}(\Sigma(\gamma)\Pi_U), \tag{3.15}$$

où $\Pi_U \equiv U^{-1} - U^{-1}X(X'U^{-1}X)^{-1}X'U^{-1}$. En supposant que $\Sigma(\gamma) = \Delta + \gamma_1\Gamma_1 + \ldots + \gamma_k\Gamma_k$, où les Γ_i sont connus, on obtient,

$$\sum_{i=1}^{k} \gamma_{i} \operatorname{tr}(\Gamma_{i} \Pi_{U}) = E(\varrho'_{U} \varrho_{U}) - \operatorname{tr}(\Delta \Pi_{U}).$$

En choisissant k U_j différents, j = 1, ..., k (par ex.: $U_1, U_1^2, ..., U_1^k$), on obtient k équations avec k inconnues:

$$\sum_{i=1}^{k} \gamma_i \operatorname{tr}(\Gamma_i \Pi_{U_j}) = \varrho'_{U_j} \varrho_{U_j} - \operatorname{tr}(\Delta \Pi_{U_j}); j = 1, \dots, k,$$
(3.16)

lesquelles peuvent être résolues en $\hat{\gamma}_1, \ldots, \hat{\gamma}_k$. Il est essentiel de vérifier si la solution $\hat{\gamma}$ appartient à l'espace des paramètres $\{\hat{\gamma}: \sum_{i=1}^k \gamma_i \Gamma_i \text{ est définie positive}\}$.

Lorsque γ consiste uniquement en τ^2 dans (1.6), une seule matrice U dans (3.16) est nécessaire. Avec les prédicteurs antérieurs du sous-dénombrement, l'estimation de τ^2 reposait sur U = I (Ericksen et Kadane 1985; Freedman et Navidi 1986; Ericksen, Kadane et Tukey 1989); cependant, une petite étude de sensibilité concernant le modèle hétéroscédastique (1.6) a permis de croire qu'il existait un meilleur estimateur.

Choisissons $U_{\alpha} = \Delta + \Gamma(\alpha)$ dans (3.15) comme une reproduction du modèle (1.7). Alors, quand $\alpha = \tau^2$ (la valeur vraie), Fay et Herriot (1979) montrent que

$$E(\varrho'_{U_{\alpha}}\varrho_{U_{\alpha}}) = n - p, \tag{3.17}$$

où n est le nombre de régions, p est le nombre de variables prédictives dans la matrice X (par ex.: p=3 pour le modèle utilisé dans la section 5) et g_U est le résidu normalisé défini en (3.14). Par conséquent, l'estimateur par la méthode des moments proposé pour τ^2 est la valeur de pour laquelle α

$$\varrho'_{U_{\alpha}}\varrho_{U_{\alpha}} = n - p, \tag{3.18}$$

cette équation peut être résolue à l'aide de la méthode itérative de Newton-Raphson ou d'une méthode de fractionnement simple; désignons l'estimateur en question par $\hat{\tau}_{mm}^2$.

Fay et Herriot (1979) notent que les estimateurs $\hat{\tau}_{mm}^2$ et $\hat{\tau}_{m\ell}^2$ se distinguent l'un de l'autre surtout par la façon dont les régions avec de faibles valeurs δ_i^2 sont pondérées dans le processus d'estimation; pour de telles régions, les carrés des résidus ont relativement plus de poids avec $\hat{\tau}_{m\ell}^2$ qu'avec $\hat{\tau}_{mm}^2$. Compte tenu de cette observation et des résultats d'une petite étude de simulation sur le biais, Cressie (1990) a donné la préférence à $\hat{\tau}_{mm}^2$. Néanmoins, du point de vue asymptotique, $\hat{\tau}_{m\ell}^2$ est parfaitement efficient et répond à une loi de distribution connue. En revanche, le fait qu'il n'existe aucune loi de distribution (asymptotique) particulière pour $\hat{\tau}_{m\ell}^2$ pose des problèmes propres à ce paramètre, par ex.: comment faire de l'inférence sur $\hat{\tau}_{mm}^2$ et comment corriger l'erreur quadratique moyenne de prévision (section 4). Nous proposons ci-dessous un estimateur plus acceptable, supérieur à l'estimateur du maximum de vraisemblance au point du vue du biais.

3.3 Estimation par la méthode du maximum de vraisemblance avec contrainte (MVC)

Il s'agit, ici, de trouver un estimateur convenable des paramètres de la matrice de variances γ , de (3.1). La méthode du maximum de vraisemblance avec contrainte (MVC), élaborée par Patterson et Thompson (1971, 1974), applique le principe du maximum de vraisemblance aux

contrastes d'erreur plutôt qu'aux données proprement dites. (Rao 1979, appelle cette méthode MVM – maximum de vraisemblance marginal – en ce qui regarde l'estimation des composantes de variance. Récemment, des auteurs l'ont appelée méthode du maximum de vraisemblance résiduel, bien qu'ils aient conservé le sigle REML en anglais (restricted maximum likelihood – residual maximum likelihood.) Une combinaison linéaire $\underline{a}'\underline{Y}$ est appelée contraste d'erreur si $\underline{E}(\underline{a}'\underline{Y}) = 0$ pour tous $\underline{\beta}$ et $\underline{\gamma}$; ainsi, $\underline{a}'\underline{Y}$ est un contraste d'erreur si et seulement si $\underline{a}'X = \underline{0}'$.

Soit W = A'Y un vecteur de n - p contrastes d'erreur linéairement indépendants, c'està-dire que les (n - p) colonnes de A sont linéairement indépendantes et A'X = 0. Suivant l'hypothèse gaussienne (3.1), $W \sim \text{Gau}(0, A' \sum (\gamma)A)$, qui ne dépend pas de β . La fonction de vraisemblance logarithmique négative est donc,

$$L_{W}(\underline{\gamma}) = ((n-p)/2)\log(2\pi) + (\frac{1}{2})\log(|A'\sum(\underline{\gamma})A|) + (\frac{1}{2})\underline{W}'(A'\sum(\underline{\gamma})A)^{-1}\underline{W}.$$

Si l'on définissait W, au moyen d'une autre série de (n-p) contrastes linéairement indépendants, la nouvelle fonction de vraisemblance logarithmique négative ne différerait de $L_W(\gamma)$ que par une constante additive (Harville 1974). En effet, pour la matrice A qui satisfait l'équation $AA' = I - X(X'X)^{-1}X'$ (et A'A = I),

$$L_{W}(\underline{\gamma}) = ((n-p)/2)\log(2\pi) - (\frac{1}{2})\log(|X'X|) + (\frac{1}{2})\log(|\Sigma(\underline{\gamma})|) + (\frac{1}{2})\log(|X'\Sigma(\underline{\gamma})|^{-1}X|) + (\frac{1}{2})\underline{Y'}\Pi(\underline{\gamma})\underline{Y},$$
(3.19)

où $\Pi(\gamma) \equiv \sum (\gamma)^{-1} - \sum (\gamma)^{-1} X(X'\sum (\gamma)^{-1}X)^{-1} X'\sum (\gamma)^{-1}$; voir Harville (1974). On obtient une estimation du maximum de vraisemblance avec contrainte de γ , désignée par $\hat{\gamma}_{r\ell}$, en minimisant (3.19) par rapport à γ . La distinction entre l'estimation par la méthode MVC et l'estimation par la méthode du maximum de vraisemblance devient importante lorsque p est grand par rapport à n.

La méthode MVC a été conçue dans le but d'estimer les paramètres des composantes de la variance; c'est dans cette perspective qu'ont été élaborés des algorithmes numériques (Harville 1977), des ajustements robustes (Fellner 1986) et des lois de distribution (Cressie et Lahiri 1991). Kitanidis (1983) et Zimmerman (1989) donnent les détails mathématiques de la minimisation de (3.19) par itération.

Harville (1974) donne une justification bayesienne de la méthode MVC en supposant une distribution a priori non informative pour β , qui est statistiquement indépendant de γ , et en montrant que la distribution a posteriori marginale de γ est proportionnelle au produit de (3.19) par la distribution a priori de γ . Lorsque cette dernière est non informative, les estimations MVC correspondent aux estimations a posteriori maximum marginales. Par conséquent, lorsque les distributions a priori non informatives de β et de γ sont indépendantes, l'estimation par la méthode MVC peut être vue comme un compromis entre l'estimation par la méthode du maximum de vraisemblance et l'estimation bayesienne avec fonction quadratique de perte. En ce qui a trait au modèle défini par (1.4), (1.5) et (1.6), la seconde méthode donnerait une estimation de Bayes, $\int_0^\infty \tau^2 \exp\{-L_W(\tau^2)\}d\tau^2$, que l'on peut aussi obtenir en moyennant sur τ^2 , pondéré par la fonction de vraisemblance complète, $\exp\{-L(\beta, \tau^2)\}$. Par ailleurs, la méthode du maximum de vraisemblance produit comme estimation de τ^2 la valeur $\hat{\tau}_{m\ell}^2$, que l'on obtient en maximisant la fonction de vraisemblance complète. Quant à la méthode MVC, elle fait la moyenne de la fonction de vraisemblance complète par rapport à β mais maximise la fonction de vraisemblance résultante (avec contrainte) par rapport à τ^2 .

L'estimation du maximum de vraisemblance de τ^2 tend à être biaisée vers zéro parce que la fonction de vraisemblance, en tant que fonction de τ^2 , est étalée vers la droite. Lorsque cette fonction est normalisée de manière à ce que son intégrale soit un, la moyenne de cette fonction est généralement plus grande que le mode (voir, par ex., Groeneveld et Meeden 1977).

L'estimation la plus vraisemblable dépend du profil de la surface de vraisemblance de β et de τ^2 , et à cause de cela, on a plus de chances d'obtenir des valeurs moins élevées pour τ^2 . (Par contraste, on calcule l'estimation MVC en intégrant la fonction de vraisemblance par rapport à β , puis en maximisant le résultat obtenu par rapport à τ^2 . Notons que les adeptes de l'approche bayesienne préconiseraient en plus une intégration par rapport à τ^2 .)

Bien que l'interprétation bayesienne de l'estimation MVC aide à en expliquer les propriétés, $\hat{\gamma}_{r\ell}$ est souvent considéré dans l'optique "fréquentiste" comme un estimateur fondé sur de l'information restreinte.

On peut minimiser (3.19) par rapport à γ au moyen de n'importe lequel des algorithmes de gradient. Rappelons-nous que,

$$W = A'Y \tag{3.20}$$

et supposons que A satisfait:

$$AA' = I - X(X'X)^{-1}X' \text{ et } A'A = I.$$

Concentrons-nous pour l'instant sur les (n-p) "données" W; leur distribution conjointe dépend uniquement de γ et la fonction de vraisemblance logarithmique négative correspondante (avec contrainte) est $L_W(\gamma)$, définie en (3.19).

Définissons le vecteur $(k \times 1)$ M_{γ} , qui a pour élément i:

$$(\underline{M}_{\gamma})_{i} \equiv \partial L_{W}(\underline{\gamma})/\partial \gamma_{i} = (\frac{1}{2})\operatorname{tr}\{\Pi(\underline{\gamma})\sum_{i}(\underline{\gamma})\} - (\frac{1}{2})\underline{Y}'\Pi(\underline{\gamma})\sum_{i}(\underline{\gamma})\Pi(\underline{\gamma})\underline{Y},$$
(3.21)

et la matrice $(k \times k)$ G_{γ} , qui a pour élément (i, j):

$$(G_{\gamma})_{ij} = E(\partial^{2}L_{W}(\gamma)/\partial\gamma_{i}\partial\gamma_{j}) = (\frac{1}{2})\operatorname{tr}\{\Pi(\gamma)\sum_{i}(\gamma)\Pi(\gamma)\sum_{j}(\gamma)\}, \tag{3.22}$$

où $\Pi(\gamma)$ est défini au-dessous de l'équation (3.19) et $\sum_i(\gamma)$ est défini en (3.5). (Les expressions (3.21) et (3.22) sont attribuables à Harville 1977.) Alors, l'algorithme de Gauss-Newton (caractérisation) servant à déterminer $\hat{\gamma}_{r\ell}$ est:

$$\gamma^{(\ell+1)} = \gamma^{(\ell)} - (G_{\gamma}^{(\ell)})^{-1} \underline{M}_{\gamma}^{(\ell)}, \tag{3.23}$$

où $G_{\gamma}^{(\ell)}$ et $M_{\gamma}^{(\ell)}$ désignent respectivement G_{γ} et M_{γ} , évalués à $\gamma = \gamma^{(\ell)}$.

Lorsque γ consiste uniquement en τ^2 dans (1.6), l'algorithme (3.23) est particulièrement simple. Dans la simulation et l'exemple présentés dans la section 5, nous avons utilisé la valeur de départ (3.9). Alors, (3.23) est,

$$(\tau^2)^{(\ell+1)} = (\tau^2)^{(\ell)} - (G_{\tau}^{(\ell)})^{-1} M_{\tau}^{(\ell)}, \tag{3.24}$$

où

$$M_{\tau} = (\frac{1}{2}) \operatorname{tr} \{ \Pi(\tau^2) D \} - (\frac{1}{2}) Y' \Pi(\tau^2) D \Pi(\tau^2) Y, \tag{3.25}$$

$$G_{\tau} = (\frac{1}{2}) \operatorname{tr} \{ \Pi(\tau^2) D \Pi(\tau^2) D \},$$
 (3.26)

$$\Pi(\tau^2) = \Sigma(\tau^2)^{-1} - \Sigma(\tau^2)^{-1} X(X'\Sigma(\tau^2)^{-1}X)^{-1} X'\Sigma(\tau^2)^{-1}, \tag{3.27}$$

sont évaluées à $\tau^2 = (\tau^2)^{(\ell)}$. Rappelons-nous, en outre, que $\Sigma(\tau^2) = \Delta + \tau^2 D$ et $D = \text{diag}\{1/C_1, \ldots, 1/C_n\}$.

En itérant (3.23) jusqu'à convergence, on obtient l'estimateur MVC $\hat{\gamma}_{r\ell}$. Cressie et Lahiri (1991) ont démontré que cet estimateur suit approximativement une distribution gaussienne multidimensionnelle ayant comme moyenne $\hat{\gamma}_{r\ell}$ et comme matrice de variances asymptotiques

$$G_{\gamma}^{-1}.\tag{3.28}$$

Lorsque γ consiste uniquement en τ^2 dans (1.6), la matrice (3.28) devient un scalaire,

$$[(\frac{1}{2}) \operatorname{tr} \{ \Pi(\tau^2) D \Pi(\tau^2) D \}]^{-1}. \tag{3.29}$$

Dans la pratique, on calcule les variances et les covariances estimées en évaluant (3.28) à $\gamma = \hat{\gamma}_{r\ell}$. De plus, l'estimateur par les moindres carrés généralisés normalisés (estimés) $\hat{\beta}(\hat{\gamma}_{r\ell})$ suit approximativement une distribution gaussienne ayant comme matrice de variances asymptotiques $(X'\Sigma(\gamma)X)^{-1}$.

4. ESTIMATION PLUS PRÉCISE DES ERREURS QUADRATIQUES MOYENNES DE PRÉVISION

Dans cette section, nous allons étudier l'effet que peut avoir sur la prévision l'estimation de γ dans $\Sigma(\gamma)$, qui figure dans l'expression (3.1). En généralisant (1.5) par

$$F \sim \operatorname{Gau}(X\beta, \Gamma(\gamma)),$$
 (4.1)

il est clair que

$$\Sigma(\underline{\gamma}) = \Delta + \Gamma(\underline{\gamma}). \tag{4.2}$$

En principe, Δ pourrait aussi dépendre de paramètres inconnus (par ex. dans un modèle pour variances d'échantillonnage) et les résultats de cette section seraient tout aussi pertinents. Le prédicteur linéaire non biaisé optimal est

$$\hat{p}(\underline{Y};\underline{\gamma}) = \Gamma(\underline{\gamma}) (\Delta + \Gamma(\underline{\gamma}))^{-1} \underline{Y} + \{I - \Gamma(\underline{\gamma}) (\Delta + \Gamma(\underline{\gamma}))^{-1}\}$$

$$X\{X' (\Delta + \Gamma(\underline{\gamma}))^{-1}X\}^{-1} X' (\Delta + \Gamma(\underline{\gamma}))^{-1} \underline{Y} = \Lambda(\underline{\gamma}) \underline{Y}. \tag{4.3}$$

Alors, la matrice des erreurs quadratiques moyennes de prévision de $\hat{p}(Y; \gamma)$, désignée par $M_1(\gamma)$, est définie

$$M_1(\gamma) = \Lambda(\gamma)\Delta\Lambda(\gamma)' + (\Lambda(\gamma) - I)\Gamma(\gamma)(\Lambda(\gamma) - I)'. \tag{4.4}$$

En réalité, γ est inconnu et doit être estimé par $\hat{\gamma}$, par exemple. Le prédicteur empirique de Bayes de F est alors $\hat{p}(Y; \hat{\gamma})$, qui correspond à (4.3) lorsque $\gamma = \hat{\gamma}$. Dans ce cas, $M_1(\gamma)$ est une mesure inadéquate de la précision du prédicteur; il faudrait utiliser à la place:

$$M_2(\hat{y}) = E\{ (\hat{F} - \hat{p}(\hat{Y}; \hat{y})) (\hat{F} - \hat{p}(\hat{Y}; \hat{y}))' \}. \tag{4.5}$$

C'est cette matrice de risque, ou une estimation de celle-ci, ainsi que le prédicteur $\hat{p}(\hat{Y}; \hat{\gamma})$ qu'il faudrait connaître. Or, on ne connaît habituellement que $M_1(\hat{\gamma})$; il convient donc de se demander quelles erreurs peuvent découler de l'utilisation de $M_1(\hat{\gamma})$ et s'il n'existerait pas un estimateur de $M_2(\gamma)$ plus approprié.

Par ailleurs, suivant les hypothèses (4.1) et (4.2) (la caractéristique gaussienne est importante ici), et pourvu que $\hat{\gamma}$ soit une fonction paire et invariante par translation des observations, on peut se servir des résultats de Harville (1985) pour établir que $M_2(\hat{\gamma}) - M_1(\hat{\gamma})$ est définie non négative. (Un estimateur est pair si $\hat{\gamma}(\hat{Y}) = \hat{\gamma}(-\hat{Y})$ et est invariant par translation si $\hat{\gamma}(\hat{Y} + X\hat{\lambda}) = \hat{\gamma}(\hat{Y})$, pour n'importe quel vecteur $p \times 1\hat{\lambda}$.) Lorsque $\hat{\gamma}$ consiste uniquement en $\hat{\tau}^2$ dans (1.6), les estimateurs $\hat{\tau}^2_{m\ell}$, $\hat{\tau}^2_{mm}$ et $\hat{\tau}^2_{r\ell}$ sont tous pairs et invariants par translation. Intuitivement, l'estimation des paramètres inconnus $\hat{\gamma}$ engendre des erreurs quadratiques moyennes de prévision plus élevées; les résultats ci-dessus confirment cette intuition.

Cependant, il existe une autre source potentielle de biais du fait que $M_1(\hat{\gamma})$, et non $M_1(\gamma)$, sert à estimer la matrice de risque. Supposons que l'on choisisse $\hat{\gamma}$ en vue d'obtenir un estimateur sans biais de la matrice de variances de (Y', F')', ce qui, on en conviendra, est un objectif louable. On peut alors se servir des résultats de Eaton (1985) et de Zimmerman et Cressie (1991) pour établir que $M_1(\hat{\gamma}) - E(M_1(\hat{\gamma}))$ est définie non négative. (La preuve repose sur une version multidimensionnelle de l'inégalité de Jensen et sur le fait que $\hat{p}(Y; \gamma)$, qui peut s'écrire aussi $\Lambda(\gamma)Y$, minimise la matrice de risque pour tous les prédicteurs linéaires sans biais.)

Étant donné l'expression,

$$M_{2}(\hat{y}) - M_{1}(\hat{y}) = \{M_{2}(\hat{y}) - M_{1}(\hat{y})\} + \{M_{1}(\hat{y}) - E(M_{1}(\hat{y}))\} + \{E(M_{1}(\hat{y})) - M_{1}(\hat{y})\},$$

$$(4.6)$$

les résultats précédents permettent d'établir que la sous-estimation de $M_2(\gamma)$ a deux causes. Même si on connaissait une expression pour $M_2(\gamma)$, $M_2(\hat{\gamma})$ serait probablement biaisée pour $M_2(\gamma)$; cela illustre davantage les difficultés inhérentes à l'estimation des erreurs quadratiques moyennes de prévision.

Prasad et Rao (1990) ont proposé une solution fondée sur un développement asymptotique de $M_2(\gamma)$. Envisageons la prédiction du sous-dénombrement dans la région i et désignons par $[M_2(\gamma)]_{ii}$ et $[M_1(\hat{\gamma})]_{ii}$ les éléments (i,i) des matrices de risque $M_2(\gamma)$ et $M_1(\hat{\gamma})$, respectivement. En appliquant rigoureusement la proposition de Prasad et Rao, on obtient l'estimateur de $[M_2(\gamma)]_{ii}$,

$$[M_2(\hat{y})]_{ii}^* = [M_1(\hat{y})]_{ii} + 2\operatorname{tr}\{A_{ii}(\hat{y})B(\hat{y})\}; i = 1, ..., n.$$
(4.7)

Dans l'équation ci-dessus, $A_{ii}(\hat{y})$ est une matrice $k \times k$ définie par l'expression

$$A_{ii}(\gamma) = \operatorname{var} \{ \partial \hat{p}_i(Y; \gamma) / \partial \gamma \}$$
 (4.8)

et $B(\tilde{y})$ est une matrice qui équivaut précisément ou approximativement à la matrice $k \times k$,

$$E\{(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)'\}. \tag{4.9}$$

Pour ce qui a trait à l'estimation par la méthode du maximum de vraisemblance,

$$B(\underline{\gamma}) = J_{\gamma}^{-1}, \tag{4.10}$$

où J_{γ} est défini en (3.7), et en ce qui concerne l'estimation MVC,

$$B(\underline{\gamma}) = G_{\gamma}^{-1}, \tag{4.11}$$

où G_{γ} est défini en (3.22).

Kass et Steffey (1989) donnent des approximations (de la variance conditionnelle) qui se rapprochent fondamentalement de (4.7), pour des distributions de probabilités qui ne sont pas nécessairement gaussiennes. Cependant, leur approche exige des répétitions indépendantes, ce qui n'est pas une caractéristique des distributions définies en (3.1).

Si l'on devait regrouper de petites régions, il est essentiel d'avoir un estimateur approximativement sans biais de tous les éléments de $M_2(\gamma)$. On peut facilement généraliser (4.7) par l'expression

$$[M_2(\gamma)]_{ij}^* = [M_1(\hat{\gamma})]_{ij} + 2\operatorname{tr}\{A_{ij}(\hat{\gamma})B(\hat{\gamma})\}; i, j = 1, ..., n,$$

où $A_{ij}(\underline{\gamma}) \equiv \text{cov}\{\partial p_i(\underline{Y}; \underline{\gamma})/\partial \underline{\gamma}, \partial p_j(\underline{Y}; \underline{\gamma}); \partial \underline{\gamma}\}$. Prasad et Rao (1990) montrent que, dans le même ordre de grandeur, on peut remplacer $A_{ij}(\underline{\gamma})$ par $\text{cov}\{\partial p_i^*(\underline{Y})/\partial \underline{\gamma}, \partial p_j^*(\underline{Y})/\partial \underline{\gamma}\}$, où $p^*(\underline{Y})$ est défini en (2.1); ces dérivées sont probablement plus faciles à calculer.

Lorsque γ consiste uniquement en τ^2 dans (1.6), le calcul de $B(\gamma)$ est simple; voir (3.13) et (3.26). Considérons maintenant

$$\operatorname{var}(\partial \hat{p}(\underline{Y};\tau^2)/\partial \tau^2) = (\partial \Lambda(\tau^2)/\partial \tau^2) \Sigma(\tau^2) (\partial \Lambda(\tau^2)/\partial \tau^2)', \tag{4.12}$$

où $\Lambda(\tau^2)$ est défini en (2.3). Lorsque ce dernier terme est exprimé en fonction de $\Pi(\tau^2)$, défini en (3.27), et de Δ , défini en (1.4), nous avons

$$\Lambda(\tau^2) = I - \Delta \Pi(\tau^2). \tag{4.13}$$

Par conséquent, on peut calculer (4.12) à l'aide de (4.13) en se servant des équations (3.4) et (3.5). Alors, $A_{ii}(\tau^2)$, défini en (4.8), est l'élément (i,i) de

$$\Delta (\partial \Pi(\tau^2)/\partial \tau^2) \Sigma(\tau^2) (\partial \Pi(\tau^2)/\partial \tau^2)' \Delta', \tag{4.14}$$

οù

$$\partial \Pi(\tau^{2})/\partial \tau^{2} = -\Sigma(\tau^{2})D\Sigma(\tau^{2})\{I - X(X'\Sigma(\tau^{2})^{-1}X)^{-1}X'\Sigma(\tau^{2})^{-1}\} - \Sigma(\tau^{2})^{-1}X(X'\Sigma(\tau^{2})^{-1}X)^{-1}\{X'\Sigma(\tau^{2})^{-1}D\Sigma(\tau^{2})^{-1}X\}$$

$$(X'\Sigma(\tau^{2})^{-1}X)^{-1}X'\Sigma(\tau^{2})^{-1} + \Sigma(\tau^{2})^{-1}X(X'\Sigma(\tau^{2})^{-1}X)^{-1}$$

$$X'\Sigma(\tau^{2})^{-1}D\Sigma(\tau^{2})^{-1}; \tag{4.15}$$

il convient de rappeler que $\Sigma(\tau^2) = \Delta + \tau^2 D$ et $D = \text{diag}\{1/C_1, \ldots, 1/C_n\}$.

On suppose que l'estimateur de l'erreur quadratique moyenne de prévision, $[M_2(\tau^2)]_{ii}^*$, est approximativement non biaisé (Prasad et Rao 1990, avaient étudié un modèle plus particulier que le nôtre). On obtient cet estimateur en combinant les expressions (4.7), (4.14) et (4.10), dans le cas de l'estimation du maximum de vraisemblance, ou (4.7), (4.14) et (4.11), dans le cas de l'estimation MVC. Dans la section qui suit, nous comparerons cet estimateur à l'estimateur courant $[M_1(\hat{\tau}^2)]_{ii}$ à l'aide de données du recensement et de l'enquête postcensitaire de 1980 aux États-Unis.

5. COMPARAISON D'ESTIMATEURS AU MOYEN D'UN EXEMPLE ET D'UNE SIMULATION

5.1 Exemple

Les données de la série PEP 3-8 de l'Enquête postcensitaire de 1980 pour les n=51 "États" des É.-U. (y compris Washington, DC) serviront à illustrer la méthode empirique de Bayes. Ces données figurent dans Cressie (1989, tableau 1, colonnes de totaux) et les variances $\delta_1^2, \ldots, \delta_{51}^2$ dans (1.3) sont tirées de la colonne de totaux qui a pour titre MSE (et dont les éléments mis au carré sont désignés par MSE₁, ..., MSE₅₁). En utilisant la relation $F_i = \{1 - U_i/100\}^{-1}$ et la méthode δ , nous avons $\delta_i^2 \simeq (Y_i)^4 (\text{MSE}_i)/10^4$. Huit variables explicatives, définies par Ericksen, Kadane et Tukey (1989), ont été adjointes aux 51 États (qui étaient divisés, globalement, en 66 sous-régions comprenant des villes, des portions d'État et des États entiers). Les variables explicatives étaient les suivantes:

- 1. Pourcentage de minorités.
- 2. Taux de criminalité.
- 3. Pourcentage de la population vivant sous le seuil de la pauvreté.
- 4. Proportion de personnes qui maîtrisent difficilement la langue anglaise.
- 5. Niveau d'instruction.
- 6. Logement.
- 7. Proportion de la population vivant dans l'une ou l'autre de 16 villes centrales définies au préalable.
- 8. Proportion de la population recensée selon la méthode classique.

Afin de trouver parmi ces variables celles qui définiraient un bon modèle du sousdénombrement, nous nous sommes servis de la méthode de sélection d'Ericksen, Kadane et Tukey (1989) mais nous avons pondéré les données proportionnellement à la racine carrée de l'effectif recensé des petites régions. Les variables choisies étaient le pourcentage de minorités et le niveau d'instruction, sans oublier la constante. Ces trois variables seront donc les seules, ici, à entrer dans le modèle linéaire, c'est-à-dire que seuls les coefficients de régression β_0 , β_1 et β_5 seront ajustés.

Suivant le modèle défini par les expressions (1.4), (1.5) et (1.6), les paramètres inconnus sont β et τ^2 . D'après l'algorithme de caractérisation (3.8), l'estimation du maximum de vraisemblance de τ^2 est:

$$\hat{\tau}_{m\ell}^2 = 47.32,$$

tandis que d'après l'algorithme de caractérisation (3.23), l'estimation MVC de τ^2 est:

$$\hat{\tau}_{r\ell}^2 = 58.53.$$

Ces résultats illustrent le phénomène que nous observons plus bas dans la simulation, à savoir que $\hat{\tau}_{m\ell}^2 < \hat{\tau}_{r\ell}^2$; dans la section 3.3, nous avions donné une explication intuitive de ce phénomène. (Incidemment, Cressie (1990) a calculé la valeur $\hat{\tau}_{mm}^2 = 94.96$, mais on ne peut dire vraiment s'il existe une relation d'inégalité générale entre les trois estimations.)

Les formules de la section 3 nous ont permis de calculer les estimations suivantes (les erreurs types estimées étant entre parenthèses):

m.v.	MVC
$\hat{\beta}_0 = 1.03227 \ (0.00708)$	$\hat{\beta}_0 = 1.03246 \ (0.00724)$
$\hat{\beta}_1 = 0.0006878 \ (0.0001402)$	$\hat{\beta}_1 = 0.0006941 \ (0.0001436)$
$\hat{\beta}_5 = -0.001070 \ (0.000231)$	$\hat{\beta}_5 = -0.001078 \ (0.000236)$
$\hat{\tau}^2 = 47.32 \ (32.87)$	$\hat{\tau}^2 = 58.53 \ (38.1).$

Notons qu'il y a très peu de différence entre les deux séries d'estimations, sauf pour ce qui a trait à τ^2 . Au moyen des estimations du m.v. et des estimations MVC dans $\hat{p}_i(\underline{Y}; \hat{\tau}^2)$ défini en (2.5), $[M_1(\hat{\tau}^2)]_{ii}$, défini en (2.6), et $[M_2(\tau^2)]_{ii}^*$, défini en (4.7), $i=1,\ldots,n$, nous pouvons calculer les prédicteurs pour petites régions ainsi que la racine carrée des erreurs quadratiques moyennes de prévision estimées. Le tableau 1 donne les résultats pertinents pour les n=51 États; on trouve aussi dans ce tableau le taux de sous-dénombrement brut, Y_i , le résultat du modèle linéaire ajusté, $(X\hat{\beta})_i$, et le poids,

$$w_i \equiv \hat{\tau}^2 / (C_i \delta_i^2 + \hat{\tau}^2), \tag{5.1}$$

de sorte que

$$\hat{p}_i(\hat{Y}; \hat{\tau}^2) = w_i Y_i + (1 - w_i) (X \hat{\beta})_i; i = 1, \dots, 51.$$
 (5.2)

Notons que w_i est systématiquement plus élevé pour l'estimation MVC que pour l'estimation du m.v., ce qui est logique intuitivement compte tenu du biais négatif notoirement élevé de $\hat{\tau}_{m\ell}^2$. Ainsi, l'estimation MVC de τ^2 attribue moins de poids au terme de modèle $(X\hat{\beta})_i$, mais d'une manière telle que l'on peut tenir compte de l'effet de l'estimation de τ^2 .

Nous notons avec intérêt qu'il y a un inconvénient à utiliser l'estimation MVC; la racine carrée de l'erreur quadratique moyenne de prévision est toujours plus élevée dans ce cas que dans le cas de l'estimation du m.v. Cela n'est pas étonnant si l'on tient compte du fait que l'estimateur MVC est (asymptotiquement) moins efficace que l'estimateur du m.v. Notons en outre que la racine carrée de l'erreur quadratique moyenne de prévision corrigée, $\sqrt{[M_2(\tau^2)]_{ii}^*}$, est de 1 à 9% plus élevée que $\sqrt{[M_1(\hat{\tau}^2)]_{ii}}$.

En ce qui a trait à la prévision, on peut évaluer l'importance globale des deux méthodes d'estimation de τ^2 comparées l'une à l'autre en calculant la somme des carrés pondérée,

$$\sum_{i=1}^{51} \{\hat{p}_i(Y;\hat{\tau}_{m\ell}^2) - \hat{p}_i(Y;\hat{\tau}_{r\ell}^2)\}^2 C_i = 15.$$

Lorsqu'on compare cette somme à,

$$\sum_{i=1}^{51} (Y_i - 1)^2 C_i = 70,421$$

et à

$$\sum_{i=1}^{51} \{Y_i - \hat{p}_i(Y; \hat{\tau}_{m\ell}^2)\}^2 = 26,033,$$

on vient bien que, d'un point de vue national, la prévision n'est pas très sensible aux méthodes d'estimation de τ^2 . (Cressie 1990 arrive à la même conclusion après avoir comparé de la même manière les estimateurs $\hat{\tau}_{m\ell}^2$ et $\hat{\tau}_{mm}^2$.) Cependant, on voit bien aussi d'après le tableau 1 que la racine carrée de l'erreur quadratique moyenne de prévision estimée est beaucoup plus sensible.

Tableau 1: De gauche à droite: les 51 États, identifiés par un code à trois lettres; le taux de sousdénombrement brut $\{Y_i\}$; le résultat du modèle ajusté $\{(X\hat{\mathbb{Q}})_i\}$; le poids $\{w_i\}$, défini en (5.1); le prédicteur (5.2) (désigné par F12); la racine carrée de l'erreur quadratique moyenne de prévision $\{\sqrt{[M_1(\hat{\tau}^2)]_{ii}}\}$ (désignée par RMPE1); et la racine carrée de l'erreur quadratique moyenne de prévision corrigée $\{\sqrt{[M_2(\tau^2)]_{ii}^*}\}$ (désignée par RMPE2).

Tableau 1

	Y	MVC				
ÉTAT		AJUST	POIDS	F12	RMPE1	RMPE2
ala	0.9965	1.0037	0.1431	1.0026	0.00439	0.00453
aka	1.0288	1.0175	0.4767	1.0229	0.00896	0.00976
arz	1.0204	1.0158	0.0742	1.0162	0.00487	0.00500
ark	0.9895	0.9962	0.1398	0.9953	0.00541	0.00562
cal	1.0307	1.0225	0.0682	1.0231	0.00322	0.00327
col	1.0033	1.0199	0.1926	1.0167	0.00473	0.00495
con	0.9886	1.0079	0.1029	1.0059	0.00435	0.00451
del	0.9938	1.0107	0.4571	1.0030	0.00739	0.00811
fla	1.0144	1.0120	0.0785	1.0122	0.00289	0.00295
gga	0.9955	1.0046	0.1639	1.0031	0.00391	0.00403
hai	1.0111	1.0105	0.2785	1.0107	0.00678	0.00730
idh	1.0125	1.0070	0.5627	1.0101	0.00531	0.00579
ill	1.0211	1.0103	0.1170	1.0116	0.00257	0.00265
ind	0.9936	1.0026	0.1413	1.0013	0.00334	0.00349
iow	0.9932	1.0033	0.1478	1.0018	0.00452	0.00475
kan	1.0056	1.0092	0.2215	1.0084	0.00466	0.00496
kty	0.9845	0.9872	0.1519	0.9868	0.00507	0.00524
lou	1.0234	1.0086	0.0263	1.0090	0.00476	0.00480
mne	1.0201	0.9992	0.3703	1.0069	0.00593	0.00645
mld	1.0242	1.0140	0.0712	1.0147	0.00406	0.00415
mas	0.9882	1.0068	0.1945	1.0032	0.00323	0.00341
mch	1.0079	1.0081	0.1601	1.0081	0.00259	0.00271
min	1.0111	1.0049	0.2793	1.0066	0.00359	0.00383
mis	1.0097	1.0086	0.1279	1.0087	0.00557	0.00575
mou	1.0080	1.0010	0.1681	1.0022	0.00350	0.00367 0.00761
mon	1.0144	1.0059	0.3785	1.0091	0.00699	0.00781
neb	1.0008	1.0071	0.5117	1.0039	0.00441 0.00744	0.00480
nev	1.0265	1.0151	0.2852	1.0183	0.00684	0.00740
nwh	0.9842	1.0033	0.3080	0.9974	0.00305	0.00740
nwj	1.0130	1.0105	0.0895	1.0107	0.00303	0.00514
nwm	1.0236	1.0256	0.3276	1.0249	0.00243	0.00047
nwy	1.0166	1.0119	0.0807	1.0123 1.0007	0.00243	0.0024
noc	1.0118	0.9998	0.0748	1.0007	0.00313	0.00430
nod	1.0005	0.9969	0.8931	1.0051	0.00253	0.0032
oho	1.0108	1.0044	0.1273 0.1625	1.0032	0.00233	0.0020
okl	0.9977 1.0027	1.0018 1.0089	0.1623	1.0071	0.00429	0.00464
ore		1.0089	0.2833	1.0007	0.00253	0.00263
pen	0.9972	0.9939	0.4167	1.0007	0.00625	0.00678
rhi	1.0089	1.0040	0.0216	1.0053	0.00555	0.00559
soc	1.0632 1.0008	0.9985	0.7538	1.0003	0.00353	0.00496
sod	0.9717	0.9966	0.0755	0.9947	0.00439	0.00449
ten	1.0037	1.0149	0.0482	1.0144	0.00341	0.0034
tex	1.0040	1.0149	0.4010	1.0101	0.00524	0.0056
uth	0.9889	1.0018	0.8232	0.9912	0.00324	0.00479
vmt	1.0009	1.0018	0.1753	1.0049	0.00338	0.0035
vir	1.0142	1.0038	0.1305	1.0123	0.00418	0.0043
was	0.9942	0.9877	0.1452	0.9887	0.00603	0.0062
wev	1.0173	1.0032	0.2877	1.0073	0.00325	0.0034
wis	1.0173	1.0032	0.3992	1.0221	0.00882	0.00963
wyo dol						0.0112
dcl	1.0375	1.0474	0.2191	1.0452	0.01081	

Tableau 1 (fin)

ÉTAT	Y	m.v.				
ETAT		AJUST	POIDS	F12	RMPE1	RMPE
ala	0.9965	1.0037	0.1190	1.0028	0.00415	0.00427
aka	1.0288	1.0175	0.4241	1.0223	0.00850	0.00933
arz	1.0204	1.0157	0.0608	1.0160	0.00448	0.00459
ark	0.9895	0.9963	0.1161	0.9955	0.00506	0.00525
cal	1.0307	1.0224	0.0559	1.0228	0.00314	0.00319
col	1.0033	1.0198	0.1617	1.0171	0.00446	0.00466
con	0.9886	1.0079	0.0849	1.0063	0.00398	0.00412
del	0.9938	1.0107	0.4050	1.0039	0.00697	0.00771
fla	1.0144	1.0120	0.0644	1.0121	0.00271	0.00276
gga	0.9955	1.0046	0.1368	1.0034	0.00375	0.00385
hai	1.0111	1.0105	0.2378	1.0106	0.00629	0.00679
idh	1.0125	1.0070	0.5099	1.0098	0.00507	0.00559
ill	1.0211	1.0103	0.0967	1.0113	0.00242	0.00248
ind	0.9936	1.0026	0.1174	1.0015	0.00309	0.00323
iow	0.9932	1.0034	0.1230	1.0021	0.00418	0.00438
kan	1.0056	1.0091	0.1870	1.0085	0.00432	0.00460
kty	0.9845	0.9874	0.1264	0.9870	0.00486	0.00502
lou	1.0234	1.0086	0.0214	1.0089	0.00446	0.00449
mne	1.0201	0.9993	0.3222	1.0060	0.00557	0.00608
mld	1.0242	1.0139	0.0583	1.0145	0.00376	0.00384
mas	0.9882	1.0068	0.1634	1.0037	0.00302	0.00319
mch	1.0079	1.0081	0.1335	1.0081	0.00242	0.00252
min	1.0111	1.0049	0.2386	1.0064	0.00339	0.00362
mis	1.0097	1.0085	0.1060	1.0087	0.00526	0.00541
mou	1.0080	1.0011	0.1404	1.0021	0.00326	0.00341
mon	1.0144	1.0059	0.3299	1.0087	0.00656	0.00717
neb	1.0008	1.0071	0.4587	1.0042	0.00420	0.00461
nev	1.0265	1.0150	0.2439	1.0178	0.00692	0.00746
nwh	0.9842	1.0033	0.2646	0.9983	0.00637	0.00691
nwj	1.0130	1.0105	0.0736	1.0106	0.00283	0.00290
nwm	1.0236	1.0254	0.2826	1.0249	0.00582	0.00617
nwy	1.0166	1.0119	0.0663	1.0122	0.00231	0.00235
noc	1.0118	0.9998	0.0614	1.0005	0.00401	0.00408
nod	1.0005	0.9970	0.8710	1.0000	0.00310	0.00324
oho	1.0108	1.0045	0.1055	1.0051	0.00236	0.00245
okl	0.9977	1.0018	0.1356	1.0013	0.00396	0.00416
ore	1.0027	1.0088	0.2421	1.0074	0.00408	0.00436
pen	0.9972	1.0014	0.1227	1.0008	0.00239	0.00248
rhi	1.0089	0.9940	0.3660	0.9995	0.00591	0.00645
soc	1.0632	1.0041	0.0176	1.0051	0.00519	0.00523
sod	1.0008	0.9985	0.7122	1.0002	0.00452	0.00490
ten	0.9717	0.9967	0.0619	0.9951	0.00413	0.00422
tex	1.0037	1.0148	0.0393	1.0144	0.00329	0.00332
uth	1.0040	1.0141	0.3512	1.0105	0.00498	0.00536
vmt	0.9889	1.0019	0.7901	0.9916	0.00445	0.00477
vir	1.0009	1.0058	0.1467	1.0051	0.00317	0.00330
was	1.0142	1.0120	0.1082	1.0123	0.00391	0.00406
wev	0.9942	0.9879	0.1207	0.9886	0.00567	0.00590
wis	1.0173	1.0033	0.2461	1.0067	0.00306	0.00328
wyo	1.0361	1.0127	0.3494	1.0209	0.00829	0.00909
dcl	1.0375	1.0470	0.1849	1.0452	0.01036	0.01078

Cressie (1990) définit des expressions pour déterminer le risque lié à un redressement par $\hat{p}(Y;\tau^2)$ et le risque lié à l'absence de redressement. Lorsqu'on substitue $\hat{\tau}_{r\ell}^2$ et $\hat{g}(\hat{\tau}_{r\ell}^2)$ dans ces expressions, le risque lié à un redressement est de 3,253 tandis que le risque lié à l'absence de redressement est de 34,134. Autrement dit, le fait de ne pas redresser l'effectif recensé accroît le risque de 949% (à la condition que le modèle défini en (1.4), (1.5) et (1.6) soit valable).

5.2 Simulation

Afin de vérifier les propriétés de distribution asymptotique de l'estimateur MVC (et de l'estimateur du m.v.) de τ^2 , nous avons soumis le modèle linéaire décrit dans la section 5.1 à une simulation comprenant les valeurs de paramètres:

$$\beta_0 = 1.0330, \quad \beta_1 = 0.000712, \quad \beta_5 = -0.000110, \quad \tau^2 = 95.00.$$
 (5.3)

La simulation

$$Y \sim \operatorname{Gau}(X\beta, \Delta + \tau^2 D),$$
 (5.4)

où Δ est défini en (1.4) (nous avons utilisé les mêmes valeurs $\delta_1^2, \ldots, \delta_{51}^2$ que dans la section 5.1 et Cressie (1990)) et D est défini en (1.6), a été exécutée 500 fois et les estimations $\hat{\tau}_{m\ell}^2, \hat{\tau}_{mm}^2$ et $\hat{\tau}_{r\ell}^2$ calculées à chaque fois. (Lorqu'une valeur négative était calculée, elle était ramenée à zéro.) Les diagrammes arborescents des trois séries d'estimations sont présentés dans les figures 1a, 1b et 1c respectivement. Il convient de remarquer que le nombre de zéros est relativement plus élevé dans le cas des estimations du m.v. (figure 1a).

Figure 1: Diagrammes arborescents de la composante de variance estimée τ^2 , selon 500 simulations du modèle (5.4): a) méthode du maximum de vraisemblance (section 3.1); b) méthode des moments (section 3.2) et c) méthode du maximum de vraisemblance avec contrainte (section 3.3).

```
00000000000001155556667
0
    0001223566667889
 1
 2
    000112356677899
 3
    001112234455555779999
    00111111122223334444555556666777788888899999
    0000122233333334455566666778888899999999
 5
    0000011111111122222222233333444455666777777888889999
    00011111111222223344444455556666777778888888999
 7
    00111222222333333333344455566777777888889999
8
    00001111222222333334555677777788
9
    00001111111233334444456677777888899
10
    0001112222234444456667899
11
    000111122223333336677788899
12
13
    1223345556677999
    0001222334445666799
14
    000012223344558999
15
16
    157899
    001122233589
17
    2568
18
19
    145
20
    7
21
    88
22
```

```
000000377778
 0
    011113344446679999
 2
    11144455557778888
    0000222222233333355555666666888899999
 4
    12222224466667777777999999
 5
    0002222333333455577777778888
    0000000111333334444444446666677779999999999
    111122222224444455555555777778888888
 8
    00000222333333355555556666666688999
 9
    111222222244444466666777779999
10
    000000222233333355577777888888
11
    0000000111113333444444666677789999
12
    11111222222244444444445777778
13
    00002233666888888999
14
    11122244667777999
15
    002233558888
16
    000001133444777799
17
    122222245555788
18
    00003335566899
19
    26799
20
    02258
21
    37
22
    11558
23
24
    79
25
26
27
    5
    5
28
29
    2
30
    78
31
    3
32
33
    2
```

Figure 1b

Les moyennes (\bar{X}) et les écarts-types des distributions présentées dans la figure 1 sont:

$\hat{ au}_{m\ell}^2$	$\hat{ au}_{mm}^2$	$\hat{ au}_{r\ell}^2$
$\bar{X} = 83.56$	$\bar{X} = 96.85$	$\bar{X} = 94.27.$
S = 45.65	S = 57.46	S = 49.17.

Comparons ces moyennes à la valeur vraie de τ^2 , soit 95.00. Le biais de $\hat{\tau}^2_{m\ell}$ est évident; $\hat{\tau}^2_{r\ell}$ a un très faible biais et présente un léger avantage par rapport à $\hat{\tau}^2_{mm}$. Pour ce qui a trait aux écarts-types, $\hat{\tau}^2_{r\ell}$ est beaucoup plus avantageux que $\hat{\tau}^2_{mm}$ mais désavantageux par rapport à $\hat{\tau}^2_{m\ell}$. Pour des raisons que nous avons exposées dans la section 3.3, et qui ne sont pas toutes de nature statistique, le biais est plus important que la variance; par conséquent, la méthode du maximum de vraisemblance avec contrainte pourrait remplacer avantageusement la méthode du maximum de vraisemblance pour l'estimation de τ^2 .

On peut vérifier les propriétés de distribution asymptotique de l'estimateur du m.v. et de l'estimateur MVC à l'aide des simulations. (La méthode des moments ne présente aucun intérêt ici car aucune loi de distribution asymptotique n'a encore été définie pour cette méthode.) En substituant $\tau^2 = 95.00$ dans (3.13), on obtient

```
00000001234777799
    0012334567789
2
    012225556888899
    112234444556889
    0013344445555666777888888899
    000011222333333444445566777788888999
    000111222222334444444445666677778888999999
    0000001111111222222233334444555677778888899999
    00000001111222333455555566666777778889
8
    0001112222222333333444455555566689999999
9
    0000000011122333444555566777888899
10
    00011222333444556666777888888899
11
    00011111122233344455567789
12
    000133334555555556788
13
    0001112344445667789
14
    00111222344566788
15
16
    0011122223355557999
17
    011235556
    00112566777779
18
19
    117
20
    013478
21
    123
22
23
    6
24
25
    02
```

Figure 1c

$$\{\operatorname{var}(\hat{\tau}_{m\ell}^2)\}^{1/2} \simeq 48.73,$$

qui doit être comparé à S=45.65. Enfin, en substituant $\tau^2=95.00$ dans (3.29), on obtient

$$\{\operatorname{var}(\hat{\tau}_{r\ell}^2)\}^{1/2} \simeq 50.14,$$

qui doit être comparé à S = 49.17.

La simulation nous permet aussi d'étudier les erreurs de prévision "réelles" et d'évaluer le rendement de $M_1(\hat{\tau}^2)$ et de $M_2(\tau^2)^*$. Si les valeurs des paramètres (5.3) étaient estimées à l'aide des observations initiales, on parlerait alors d'une "bootstrap" paramétrique.

6. CONCLUSIONS ET ANALYSE

La prédiction du sous-dénombrement fondée sur un modèle repose sur une vérification rigoureuse de l'ajustement. Des graphiques diagnostiques basés sur les résidus normalisés ont été proposés à la fin de la section 2 de cet article. Les résidus normalisés BLUP, $\{Y_i - \hat{p}_i(Y_i; \hat{\tau}^2)\}/\{[M(\hat{\tau}^2)]_{ii}\}^{\frac{1}{2}}$; $i=1,\ldots,n$, peuvent eux aussi être utiles. Ils peuvent être utilisés dans un graphique quantile-quantile (voir, par ex. Cressie 1991, p. 255) ou, comme le proposent Calvin et Sedransk (1991), être représentés graphiquement en fonction de $\hat{p}_i(Y_i; \hat{\tau}^2)$; $i=1,\ldots,n$.

On pourrait également élargir le modèle (1.4) en y incluant un paramètre de composante de variance inconnu, σ^2 :

$$Y \sim \text{Gau}(F, \sigma^2 \Delta),$$
 (6.1)

où $\Delta = \text{diag}\{\delta_1^2, \ldots, \delta_n^2\}$. En ajustant le modèle plus général défini par (6.1), (1.5) et (1.6), on pourrait alors vérifier si l'estimation MVC $\sigma_{r\ell}^2$ est significativement différente de $\sigma^2 = 1$, ce qui serait un moyen de détecter les erreurs de spécification. (Dans ce cas, il est préférable de recourir à l'estimation MVC plutôt qu'à l'estimation du m.v. puisque tout biais aura une influence considérable sur l'induction pour σ^2 .)

En estimant les paramètres de la matrice de variances par la méthode du maximum de vraisemblance avec contrainte (MVC), on a moins de chances d'obtenir des prédicteurs empiriques de Bayes qui mettent trop de poids sur le modèle de régression (1.5). En contrepartie, on aura une erreur quadratique moyenne de prévision légèrement plus grande. En se servant des propriétés de distribution asymptotique des estimateurs MVC (que l'on vérifie par simulation), on peut aussi obtenir des estimateurs plus précis de l'erreur quadratique moyenne de prévision. En vertu du modèle défini par les expressions (1.4), (1.5) et (1.6), on peut conclure qu'il existe des moyens fiables de faire de l'induction sur les facteurs de redressement $\{F_i: i=1,\ldots,n\}$; les prédicteurs $\{\hat{p}_i(Y;\hat{\tau}_{r\ell}^2): i=1,\ldots,n\}$ donnent les prédicteurs de l'effectif réel et du sous-dénombrement

$$T_i^{\text{prd}} = \hat{p}_i(\underline{Y}; \hat{\tau}_{r\ell}^2) C_i \text{ et } U_i^{\text{prd}} = 100\{1 - (\hat{p}_i(\underline{Y}; \hat{\tau}_{r\ell}^2))^{-1}\}; i = 1, ..., n,$$

respectivement. On peut calculer le biais et l'erreur quadratique moyenne de prévision de ces prédicteurs à l'aide de la méthode (voir Cressie 1991, section 3.2.2).

REMERCIEMENTS

L'auteur tient à exprimer sa reconnaissance à Robert Parker pour l'aide que ce dernier lui a apportée dans sa recherche. Il remercie aussi le rédacteur associé et les deux arbitres qui lui ont fait des commentaires utiles. Cette étude a été rendue possible grâce à la convention sur la statistique no JSA 90-41 intervenue entre le Bureau of the Census des É.-U. et l'université Iowa State. Les conclusions et les opinions formulées dans cet article sont celles de l'auteur et ne reflètent pas nécessairement la position du Bureau of the Census.

BIBLIOGRAPHIE

- CALVIN, J.A., et SEDRANSK, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *Journal of the American Statistical Association*, 86, 36-48.
- CRESSIE, N. (1988). Estimating census undercount at national and subnational levels. Dans *Proceedings* of Bureau of the Census Fourth Annual Research Conference, Washington, DC: U. S. Bureau of the Census, 127-150.
- CRESSIE, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *Journal of the American Statistical Association*, 84, 1033-1044.
- CRESSIE, N. (1990). Weighted smoothing of estimated undercount. In *Proceedings of Bureau of the Census 1990 Annual Research Conference*, Washington, DC: U. S. Bureau of the Census, 301-325.
- CRESSIE, N. (1991). Statistics for Spatial Data. New York: Wiley.
- CRESSIE, N., et LAHIRI, S.N. (1991). The asymptotic distribution of REML estimators. *Statistical Laboratory Preprint 91-20*, Iowa State University, Ames, Iowa.
- EATON, M.L. (1985). The Gauss-Markov Theorem in multivariate analysis. Dans *Multivariate Analysis VI*, (Éd. P.R. Krishnaiah). Amsterdam: Elsevier, 177-201.
- ERICKSEN, E.P., et KADANE, J.B. (1985). Estimating the population in a census year: 1980 and beyond. *Journal of the American Statistical Association*, 80, 98-109.

- ERICKSEN, E.P., KADANE, J.B., et TUKEY, J.W. (1989). Adjusting the 1980 Census of population and housing. *Journal of the American Statistical Association*, 84, 927-944.
- FAY, R.E., III, 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.
- FELLNER, W.H. (1986). Robust estimation of variance components. Technometrics, 28, 51-60.
- FREEDMAN, D.A., et NAVIDI, W.C. (1986). Regression models for adjusting the 1980 Census. *Statistical Science*, 1, 3-17.
- 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.
- GROENEVELD, R.A., et MEEDEN, G.D. (1977). The mode, median and mean inequality. *American Statistician*, 31, 120-121.
- HARVILLE, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61, 383-385.
- HARVILLE, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.
- HARVILLE, D.A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association*, 80, 132-138.
- KASS, R.E., et STEFFEY, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, 84, 717-726.
- KITANIDIS, P.K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*, 19, 909-921.
- LAIRD, N.M., et LOUIS, T.A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association*, 82, 739-750.
- MARDIA, K.V., et MARSHALL, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135-146.
- MILLER, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics*, 5, 746-762.
- PATTERSON, H.D., et THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, 58, 545-554.
- PATTERSON, H.D., et THOMPSON, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*. Washington, DC: Biometric Society, 197-207.
- PRASAD, N.G.N., et RAO, J.N.K. (1990). On the estimation of mean square error of small area predictors. *Journal of the American Statistical Association*, 85, 163-171.
- RAO, C.R. (1979). MINQE theory and its relation to ML and MML estimation of variance components. *Sankhyā B*, 41, 138-153.
- WOLTER, K.M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81, 338-346.
- ZIMMERMAN, D.L. (1989). Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Mathematical Geology*, 21, 655-672.
- ZIMMERMAN, D.L., et CRESSIE, N. (1991). Mean-squared prediction error in the spatial linear model. *Annals of the Institute of Statistical Mathematics*, 43, forthcoming.