Techniques d'enquête, juin 1987
Vol. 13, n° 1, pp. 83–91
Statistique Canada

Sur l'estimation efficace des taux de chômage à l'aide de données de l'enquête sur la population active

S. KUMAR et A.C. SINGH[1]

RÉSUMÉ

Tout comme avec la méthode d'estimation de X^2 min de Neyman (1949) pour les échantillons aléatoires simples, la méthode de $Q^{(T)}$ minimum proposée par Singh (1985) pour les plans de sondage complexes produit des estimateurs de paramètres de modèle asymptotiquement efficaces. La variable $Q^{(T)}$ peut être considérée comme une variable X^2 pour des données d'enquête qualitatives. Les estimateurs $Q^{(T)}$ min peuvent remplacer avantageusement les estimateurs obtenus par les moindres carrés pondérés qui ont souvent un comportement instable dans le cas des échantillons complexes. Les auteurs exposent tout d'abord la méthode de $Q^{(T)}$ min puis l'appliquent à l'estimation des paramètres d'un modèle logit pour des estimations de taux de chômage calculées à partir des données de l'EPA d'octobre 1980 classées selon deux critères: l'âge et le niveau d'instruction. Il appert que l'efficacité de trace des estimations lissées calculées par Kumar et Rao (1986) à l'aide de la méthode des pseudo-EMV peut être légèrement améliorée à l'aide de la méthode de $Q^{(T)}$ min. Détail digne de mention, les pseudo-EMV pour les cellules prises individuellement ont un comportement très comparable à celui des estimateurs efficaces $Q^{(T)}$ min utilisés pour l'EPA.

MOTS CLÉS: Pseudo-EMV; estimateur par les moindres carrés pondérés (MCP); estimateur $Q^{(T)}$ min; efficacité asymptotique; fonction de vraisemblance approximative; cote généralisée.

1. INTRODUCTION

En se basant sur les données de l'enquête sur la population active (EPA) d'octobre 1980, Kumar et Rao (1984, 1986) ont proposé et analysé un modèle de régression logistique (modèle logit) pour les taux de chômage. Ils ont utilisé la théorie élaborée par Roberts (1985) et Roberts, Rao et Kumar (1987). Ces derniers ont généralisé la méthode proposée par Rao et Scott (1981, 1984) pour corriger le X^2 en fonction des effets du plan de sondage et ont vérifié la validité de l'ajustement du modèle logit. Kumar et Rao ont considéré des taux de chômage pour diverses cellules (ou domaines) obtenues en classant simultanément la population visée en fonction de deux critères: l'âge et le niveau d'instruction. Le modèle logit renfermait des effets linéaires et quadratiques pour l'âge tandis qu'il ne renfermait que des effets linéaires pour le niveau d'instruction. Singh et Kumar (1986) ont analysé les mêmes données d'enquête au moyen d'une autre méthode, le test $Q^{(T)}$, qui a été proposée par Singh (1985). Le test $Q^{(T)}$ est un test du chi-carré fondé sur une cote généralisée de composantes principales. Ce test a produit des résultats comparables à ceux obtenus par la méthode du X^2 corrigé.

Dès qu'on a construit un modèle convenable, il faut trouver de bons estimateurs de ces paramètres. Ces estimateurs permettent ensuite d'estimer assez fidèlement les taux réels des domaines. Les estimations ainsi obtenues, souvent appelées "estimations lissées", sont particulièrement utiles dans les cas où les estimations d'enquête ne sont pas assez précises à cause d'un nombre insuffisant d'observations. Comme les estimations lissées sont calculées après qu'on a vérifié la validité de l'ajustement du modèle, il faut s'attendre que le biais des estimations soit négligeable. Kumar et Rao (1986) ont utilisé la méthode des pseudo-EMV (pseudo-estimateurs du maximum de vraisemblance) dans une situation fictive d'échantillons

[1] S. Kumar, méthodologiste principal, Division des méthodes d'enquêtes sociales, Statistique Canada, 4-D2, Immeuble Jean Talon, Parc Tunney, Ottawa (Ontario), K1A 0T6. A.C. Singh, Département de mathématique et de statistique, Memorial University of Newfoundland, St. John's (Terre-Neuve), Canada, A1C 5S7.

binomiaux indépendants pour estimer les paramètres d'un modèle logit après avoir vérifié la validité de l'ajustement de ce modèle pour les données de l'EPA d'octobre 1980. Ils ont constaté que les estimations lissées de taux de chômage étaient beaucoup plus efficaces que les estimations d'enquête dans l'exemple de l'EPA.

Les pseudo-EMV sont particulièrement utiles lorsqu'il est impossible ou difficile de calculer la fonction de vraisemblance à cause de la complexité du plan de sondage. Dans des conditions de régularité acceptables, la méthode des pseudo-EMV produit des estimations convergentes et asymptotiquement normales (Imrey, Koch et Stokes 1982). Dans cet article, nous tentons de produire des estimations asymptotiquement efficaces (au sens défini à la section 3) des paramètres de modèle et, partant, des estimations de domaines. Nous décrivons aussi l'estimateur $Q^{(T)}$ min, qui a été défini dans Singh (1985), à l'aide de la méthode des cotes généralisées; on peut considérer cet estimateur comme analogue à l'estimateur X^2 min de Neyman pour les échantillons aléatoires simples. Il convient de souligner que la méthode des MCP (moindres carrés pondérés) produit également des estimateurs asymptotiquement efficaces dans le cas des plans de sondage complexes (Koch, Freeman et Freeman 1975). Toutefois, ces estimateurs sont habituellement instables avec des échantillons de taille moyenne à cause de la quasi-singularité de la matrice des covariances estimées des estimations de cellules (voir Imrey, Koch et Stokes 1982; Fay 1985). Par contre, les estimateurs $Q^{(T)}$ min sont à l'abri de cette instabilité. Nous verrons que la méthode de $Q^{(T)}$ min peut permettre de contourner le problème de l'instabilité en utilisant une version modifiée de la matrice des covariances estimées, où l'on élimine les valeurs propres relativement minimes de la décomposition spectrale de la matrice.

La section 2 présente la notation des termes utilisés dans l'étude ainsi qu'une brève analyse du test $Q^{(T)}$. Dans la section 3, nous décrivons l'estimateur $Q^{(T)}$ min et son comportement asymptotique. La section 4 contient l'exemple d'application aux données de l'EPA. Cet exemple nous permet de constater avec intérêt que les pseudo-EMV sont presque aussi efficaces que les estimations efficaces $Q^{(T)}$ min au niveau de la cellule. Si nous prenons une mesure globale, comme l'efficacité de trace, nous constatons que les pseudo-EMV ne sont que légèrement inférieures aux estimations $Q^{(T)}$ min. Enfin, la section 5 sert de conclusion.

2. LE TEST $Q^{(T)}$: BRÈVE ANALYSE

Nous allons décrire brièvement le test $Q^{(T)}$ afin de motiver l'utilisation de la méthode d'estimation de $Q^{(T)}$ min (pour plus de détails, voir Singh 1985, Singh et Kumar 1986). Soit I le nombre de domaines disjoints et v_i le paramètre d'intérêt pour le i-ième domaine. Considérons un modèle pour $v = (v_1, v_2, \ldots, v_I)'$ tel que

$$H_0 \colon h(v) = X\theta \tag{2.1}$$

où X est une matrice connue $I \times r$ de plein rang r, θ est un r-vecteur de paramètres inconnus et h est une fonction univalente continûment différentiable, par exemple log ou logit.

Posons $\hat{v}$ comme I-vecteur des estimations d'enquête et supposons qu'en vertu du théorème de la limite centrale approprié,

$$\hat{v} \mathrel{\dot{\sim}} MVN(v, \Gamma/n) \tag{2.2}$$

où " $\dot{\sim}$ " signifie "distribué asymptotiquement comme", n est la taille de l'échantillon et Γ est la matrice des covariances asymptotiques de $\sqrt{n}(\hat{v} - v)$.

Choisissons maintenant un faible facteur de réduction de dimension $\epsilon(>0)$ (.01 ou .005 peuvent être des valeurs pratiques d'ϵ). Trouvons un nombre T de telle sorte qu'avec les valeurs propres $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_I$ de la matrice des covariances estimées $\hat{\Gamma}$, nous ayons

$$T = \max \left\{ t: t > r \text{ et } \sum_{i=t}^{I} \hat{\lambda}_i / \sum_{i=1}^{I} \hat{\lambda}_i \geq \epsilon \right\}. \tag{2.3}$$

Bien qu'aléatoire, la variable T peut être considérée comme indépendante pour nos fonctions asymptotiques. Il convient de souligner qu'en l'absence de valeurs propres relativement minimes (c.-à-d. que $\hat{\Gamma}$ n'est pas mal conditionnée), une faible valeur de ϵ n'aura normalement pas d'effet de réduction de dimension et T coïncidera alors avec I.

Examinons le problème qui consiste à tester l'hypothèse nulle H_0 par rapport à l'hypothèse alternative K_0: $h(v) \neq X\theta$ dans la catégorie de tests fondés sur les T premières composantes principales W de $\hat{v}$. Soit P_i le vecteur propre normé correspondant à $\hat{\lambda}_i$ (ce vecteur n'est pas nécessairement unique), et soit M_T la matrice $(I \times T)$ des vecteurs propres P_i correspondant aux T premières valeurs propres les plus élevées. Alors,

$$W = M_T'\hat{v} \mathrel{\dot{\sim}} MVN(\mu, D_T/n), \tag{2.4}$$

où

$$\mu = M_T'v, \; D_T = diag(\lambda_1, \ldots, \lambda_T).$$

Lorsqu'il est fondé sur les composantes principales W, le test d'hypothèse qui portait à l'origine sur un vecteur v de dimension I se ramène à un test d'hypothèse sur un paramètre μ de dimension T, où l'hypothèse est définie comme suit:

$$H_0': \mu = M_T'h^{-1}(X\theta) \text{ vs } K_0': \mu \neq M_T'h^{-1}(X\theta). \tag{2.5}$$

On peut calculer la statistique $Q^{(T)}$ comme une cote des composantes principales utilisant la fonction de vraisemblance approximative de θ, qui est définie par la distribution limite de W (2.4) pour le calcul des cotes efficaces (voir Cox et Hinkley 1974, p. 321-324). Nous allons désigner $Q^{(T)}$ comme un test de cote généralisée en vertu duquel H_0 serait rejetée pour les valeurs fortes de la formule quadratique

$$Q^{(T)}(\theta^o) = Y(\theta^o)'\Delta_T Y(\theta^o) - Z_T(\theta^o)'\wedge_T Z_T(\theta^o) \tag{2.6}$$
$$\mathrel{\dot{\sim}} \chi^2_{T-r}$$

où

$$Y(\theta^o) = \hat{v} - v(\theta^o), \; \Delta_T = n \sum_{i=1}^{T} (P_i P_i' / \hat{\lambda}_i),$$

$$Z_T(\theta^o) = B'\Delta_T Y(\theta^o), \; B = (\partial v / \partial \theta), \; \wedge_T = (B'\Delta_T B)^{-1},$$

et θ^o est un point fixe dans l'espace des paramètres nuls. Dans le calcul de $Q^{(T)}$, n'importe quelle estimation racine n-convergente de θ, sous H_0, peut être substituée à θ^o, comme peut l'être la pseudo-EMV de θ. Il est à noter que $Q^{(T)}$ dans (2.6) est en réalité définie par une formule quadratique dans W mais que, pour des raisons de commodité, elle est exprimée en fonction de $\hat{v}$.

Pour tester H_0 par rapport à K_0 dans la catégorie de tests fondés sur W, l'optimalité asymptotique du test $Q^{(T)}$ provient de celle de la cote. Pour les faibles valeurs de $\epsilon > 0$, $\hat{v}$ et W seront comparables en ce sens que les composantes principales, dans ce cas, assureront la meilleure forme de réduction de dimension qui soit, avec une perte minimum d'information. Le test $Q^{(T)}$ (pour de faibles valeurs de ϵ) devrait donc être robuste par rapport au test Q correspondant où il n'y a aucune réduction de dimension. En revanche, Q peut être instable (au sens d'une probabilité d'erreur de première espèce gonflée) dans le cas des échantillons finis à cause de la quasi-singularité possible de $\hat{\Gamma}$. Le test $Q^{(T)}$ est censé atténuer ce problème d'instabilité mais cela ne peut se faire sans qu'il y ait une perte d'information; Q peut donc renfermer des composantes peu fiables dans la direction des vecteurs propres correspondant aux valeurs propres relativement minimes. Cette perte d'information entraîne une diminution de la puissance du test $Q^{(T)}$ pour des alternatives en direction de la (quasi-) singularité. Cependant, cette perte de puissance est compensée par un gain au niveau du contrôle de l'erreur de première espèce. Comme H_0 est un sous-ensemble de H_0', $Q^{(T)}$ sera un test conservatif pour H_0; cela sert à assurer le contrôle de l'instabilité.

Une version spéciale de $Q^{(T)}\ (\theta^o)$, asymptotiquement équivalente et décrite par une formule plus simple et semblable à celle du critère X^2 standard de Pearson-Fisher, peut être obtenue en remplaçant θ^o par un estimateur $\tilde{\theta}$ qui minimise l'expression $(\hat{v} - v(\theta))'\Delta_T(\hat{v} - v(\theta))$. Nous avons donc,

$$\begin{aligned} Q^{(T)}\ (\tilde{\theta}) &= Y(\tilde{\theta})'\Delta_T Y(\tilde{\theta}) \\ &= \sum_{i=1}^{T} [P_i'(\hat{v} - v(\tilde{\theta}))]^2/\hat{\lambda}_i \\ &\dot{\sim}\ \chi^2_{T-r} \end{aligned} \tag{2.7}$$

Supposons maintenant que le test $Q^{(T)}$ ou un autre test, comme le test X^2 corrigé, nous a amené à croire qu'un modèle H_0 convient à un vecteur de données $\hat{v}$. Dans la section suivante, nous définissons une méthode asymptotiquement efficace pour estimer les paramètres θ sous H_0, utilisant la statistique $Q^{(T)}$. Les estimations ainsi obtenues permettent d'établir une série d'estimations lissées de v qui correspondent aux estimations d'enquête $\hat{v}$.

3. L'ESTIMATEUR $Q^{(T)}$ MIN

Considérons la fonction de vraisemblance approximative pour la moyenne μ des T premières composantes principales W de $\hat{v}$, qui sont définies par l'équation (2.4). Supposons que le modèle H_0: $h(v) = X\theta$ est accepté. Alors, la fonction du noyau $K(\theta)$ de la fonction de vraisemblance approximative pour $\mu(\theta)$ est donnée par

$$\begin{aligned} K(\theta) &= (W - \mu(\theta))'D_T^{-1}(W - \mu(\theta)) \\ &= (\hat{v} - v(\theta))'\Delta_T(\hat{v} - v(\theta)) \end{aligned} \tag{3.1}$$

La valeur $\tilde{\theta}$ qui minimise $K(\theta)$ correspond à l'EMV de θ pour la fonction de vraisemblance approximative de μ sous H_0. Cet estimateur $\tilde{\theta}$ sera asymptotiquement efficace (ou le meilleur estimateur asymptotiquement normal (MEAN) au sens de Neyman, 1949) dans une catégorie restreinte, notamment dans la catégorie de tests fondés sur W. D'après l'estimateur de X^2 min de Neyman (1949), l'estimateur $\tilde{\theta}$ est donc devenu dans Singh (1985) l'estimateur de $Q^{(T)}$ min. Il convient de souligner que l'estimateur $\tilde{\theta}$ dépend du facteur de réduction de dimension ϵ par Δ_T. Par conséquent, $\tilde{\theta}$ varie avec ϵ.

Étant donné les composantes principales W et sous H_0, nous pouvons calculer les estimations lissées de v de la façon suivante. Nous devons trouver tout d'abord la valeur $\tilde{\theta}$ qui minimise $K(\theta)$; en d'autres termes, la valeur recherchée est la solution de r équations

$$B'\Delta_T(\hat{v} - v(\theta)) = 0 \tag{3.2}$$

où $B(=\partial v/\partial\theta)$ et v ont tous deux rapport avec θ. Une méthode itérative comme celle de Newton-Raphson peut servir à résoudre l'équation (3.2). La valeur θ peut être estimée par la méthode des moindres carrés pondérés (MCP) ou la méthode des pseudo-EMV. Il est alors possible de calculer l'estimateur $Q^{(T)}$ min de v à l'aide de la formule suivante

$$\tilde{v} = h^{-1}(X\tilde{\theta}). \tag{3.3}$$

Les comportements asymptotiques de $\tilde{\theta}$ et de $\tilde{v}$ sont définis par la proposition suivante.

Proposition 3.1 Désignons l'expression $(B'\Delta_T B)^{-1}$ par le symbole $\wedge_T$. Nous avons

$$\begin{aligned}&\text{(a)}\quad \tilde{\theta} - \theta \approx \wedge_T B'\Delta_T(\hat{v} - v(\theta)) \stackrel{\cdot}{\sim} MVN(0,\wedge_T)\\ &\text{(b)}\quad \tilde{v} - v \approx B\wedge_T B'\Delta_T(\hat{v} - v(\theta)) \stackrel{\cdot}{\sim} MVN(0,B\wedge_T B')\end{aligned} \tag{3.4}$$

où " $\approx$ " signifie qu'il y a très peu de possibilité que la différence entre les deux membres de l'expression soient importante.

Cette proposition se démontre par l'application de la méthode-δ aux fonctions $B'\Delta_T(\hat{v} - v(\theta))$ et $\tilde{v} - v(\theta)$, ce qui donne

$$B'\Delta_T(\hat{v} - v(\theta)) - (B'\Delta_T B)(\tilde{\theta} - \theta) = o_p(1),$$

$$\tilde{v} - v(\theta) - B(\tilde{\theta} - \theta) = o_p(1).$$

La proposition ci-dessus nous amène à conclure que la matrice des covariances asymptotiques de l'estimateur $Q^{(T)}$ min $\tilde{\theta}$ est l'inverse de la matrice d'information $B'\Delta_T B$ pour θ, qui a été déterminée à l'aide de la fonction de vraisemblance approximative de θ définie en (2.4). Par conséquent, lorsqu'il n'y a pas de réduction de dimension, l'estimateur $\tilde{\theta}$ sera asymptotiquement équivalent à l'estimateur par les MCP de Koch, Freeman et Freeman (1975). Comme nous l'avons dit dans l'introduction, l'estimateur par les MCP a habituellement un comportement instable dans le cas des échantillons finis à cause d'une estimation inefficace de Γ. Par contre, pour un $\tilde{\theta}$ donné ($\epsilon > 0$), l'estimateur $\tilde{\theta}$ a normalement un comportement stable avec les échantillons finis, en ce sens que son comportement asymptotique peut fournir une bonne approximation de son comportement en général. Cette condition ne peut être réalisée sans compromettre l'optimalité asymptotique de $\tilde{\theta}$ car, pour que cette condition se réalise, cet estimateur doit être limité à une catégorie moindre, c'est-à-dire à la catégorie des estimateurs fondés sur les T premières composantes principales W. En revanche, l'estimateur par les MCP conserve son optimalité asymptotique dans une catégorie plus large, notamment celle des estimateurs fondés sur le vecteur complet de données $\hat{v}$. Si, pour une valeur faible de ϵ, la statistique $Q^{(T)}$ donne un résultat non significatif pour H_0, l'estimateur $Q^{(T)}$ min correspondant ($\tilde{v}$) est susceptible de se comparer avantageusement à l'estimateur par les MCP au point de vue de la robustesse.

4. ESTIMATIONS DES TAUX DE CHÔMAGE PAR LA MÉTHODE DE $Q^{(T)}$ MIN

Considérons les données de l'enquête sur la population active (EPA) d'octobre 1980 que Kumar et Rao (1984, 1986) et Roberts, Rao et Kumar (1987) ont analysées en appliquant l'extension de la méthode du X^2 corrigé de Rao-Scott à la régression logistique. Ces auteurs ont montré que le modèle logit défini ci-dessous s'ajustait bien aux estimations d'enquête de taux d'emploi ($v_{j\ell}$) pour le tableau de 60 cellules formées simultanément en fonction de l'âge (10 catégories) et du niveau d'instruction (6 catégories). Le modèle est

$$log \frac{v_{j\ell}}{1 - v_{j\ell}} = \beta_0 + \beta_1 A_j + \beta_2 A_j^2 + \beta_3 E_\ell \tag{4.1}$$

où A_j représente le point médian $12 + 5j$ pour le j-ième groupe d'âge ($j = 1, \ldots, 10$), et $E_\ell (\ell = 1, \ldots, 6)$ représente le nombre médian d'années de scolarité pour chaque catégorie, soit 7, 10, 12, 13, 14 et 16.

On peut exprimer le modèle (4.1) au moyen de la notation définie à la section 2 en numérotant les soixante cellules par ordre lexicographique. Ainsi, (4.1) peut être récrite $h(v) = X\theta$, où v est le vecteur des taux d'emploi, h est la fonction logit, X est une matrice 60×4 dont la i-ième ligne est $(1, A_i, A_i^2, E_i)$, et θ est $(\beta_0, \beta_1, \beta_2, \beta_3)'$. Nous avons aussi

$$H = (\partial h / \partial v) = D_v^{-1} D_{1-v}^{-1}, B = H^{-1} X, \tag{4.2}$$

où D_v et D_{1-v} sont des matrices diagonales dont les éléments de la diagonale sont définis par les indices inférieurs.

À l'aide de la pseudo-fonction de vraisemblance appliquée à la distribution binomiale mixte, Kumar et Rao (1984) ont calculé les pseudo-EMV de θ pour le modèle (4.1) et ont obtenu les valeurs suivantes:

$$\bar{\theta} = (-3.10, 0.211, -0.00218, 0.1509)'. \tag{4.3}$$

Ils ont également calculé la valeur du X^2 corrigé de premier ordre de Rao-Scott (qu'ils désignent par G_c^2); la valeur obtenue, 55.3, entraîne l'acceptation du modèle (2.1) quand on le compare à la distribution X_{56}^2.

Le modèle (4.1) a aussi été testé à l'aide de la méthode $Q^{(T)}$ (voir Singh 1985, Singh et Kumar 1986); là encore, le test s'est soldé par l'acceptation du modèle. Pour $\epsilon = .01$, nous avons $T = 51$ selon la matrice des covariances estimées $\hat{\Gamma}$, qui a été calculée par Kumar et Rao (1984). Si nous prenons maintenant les pseudo-EMV $\bar{\theta}$, nous avons

$$Q^{(51)}(\bar{\theta}) = 58.665 - 4.454 = 54.211 \tag{4.4}$$

Lorsque $\epsilon = .005$, T est égal à 54 et

$$Q^{(54)}(\bar{\theta}) = 67.774 - 2.343 = 65.431 \tag{4.5}$$

Lorsque $\epsilon = 0$, $T = 58$ parce que deux cellules renferment des taux de chômage nuls. Dans ce cas,

$$Q^{(58)}(\bar{\theta}) = 87.302 - 0.812 = 86.49 \tag{4.6}$$

En comparant les valeurs de $Q^{(51)}$, $Q^{(54)}$ et $Q^{(58)}$ aux distributions χ_{47}^2, χ_{50}^2 et χ_{54}^2 respectivement, nous constatons que les deux premiers tests entraînent l'acceptation de (4.1) tandis que le dernier entraîne son rejet. Il est possible de vérifier le degré d'instabilité des estimateurs en examinant l'écart qui existe entre $Q^{(58)}$ et $Q^{(T)}$ (pour $T = 51, 54$); cet écart peut être jugé très significatif en se référant à la distribution χ_{58-T}^2. L'analyse révèle une certaine instabilité pour le test Q où il n'y a aucune réduction de dimension. Il est clair que le test des MCP renfermerait aussi des problèmes d'instabilité à cause de la difficulté que l'on aurait à inverser la matrice $\hat{\Gamma}$ qui est singulière. La méthode de $Q^{(T)}$ min serait donc préférable à la méthode de Q min ou à celle des MCP. Dans le but de minimiser la perte d'information, il est recommandé d'utiliser la méthode qui a la valeur de T la plus élevée à la condition, bien sûr, que la valeur de $Q^{(T)}$ correspondante entraîne l'acceptation du modèle.

Nous passons maintenant au calcul d'estimations asymptotiquement efficaces. Les estimations de Q min et les estimations par les MCP n'ont pas été calculées parce que $\hat{\Gamma}$ est singulière. Les estimations $\tilde{\theta}$ de $Q^{(T)}$ min ont été calculées pour $\epsilon = .005$ et $\epsilon = .01$ en appliquant la méthode itérative de Newton-Raphson et en utilisant $\bar{\theta}$ comme l'estimateur initial de θ dans la résolution de l'équation (3.2). Les valeurs de $\tilde{\theta}_T$ et de $Q^{(T)}(\tilde{\theta})$ (dans ce dernier cas, le terme négatif de (2.6) disparaît) pour $\epsilon = .005$ et $T = 54$ sont les suivantes

$$\tilde{\theta}_{54} = (-2.7112,\ 0.1944,\ -0.00196,\ 0.1432)',\ \text{et}$$

$$Q^{(54)}(\tilde{\theta}_{54}) = 63.4737. \tag{4.7}$$

Pour $\epsilon = .01$, $T = 51$, nous avons

$$\tilde{\theta}_{51} = (-2.6739,\ 0.19702,\ -0.00202,\ 0.1364)',\ \text{et}$$

$$Q^{(51)}(\tilde{\theta}_{51}) = 55.2518. \tag{4.8}$$

Tableau 1

Efficacité des estimations lissées des taux de chômage par rapport aux estimations d'enquête[a]

N° de cellule	$Q^{(51)}$ Min	$Q^{(54)}$ Min	Pseudo-EMV	N° de celule	$Q^{(51)}$ Min	$Q^{(54)}$ Min	Pseudo-EMV
1	5.87	5.74	5.44	31	9.01	9.32	8.65
2	3.62	3.62	3.28	32	8.76	9.46	10.68
3	3.45	3.55	3.12	33	36.93	42.93	51.59
4	52.45	51.65	43.46	34	51.55	60.23	81.12
5	104.77	114.30	96.21	35	69.76	79.93	98.37
7	5.33	5.14	4.38	36	9.17	11.01	15.07
8	9.36	9.53	8.09	37	3.48	3.01	3.45
9	6.85	7.16	6.70	38	13.74	15.91	18.00
10	25.65	28.40	26.31	39	66.87	80.98	97.30
11	13.34	14.13	17.73	40	154.81	187.73	221.50
12	27.74	30.85	30.85	41	49.14	67.56	80.61
13	8.64	8.84	7.15	42	17.32	21.73	24.98
14	13.84	13.84	12.37	43	8.57	9.28	8.49
15	8.20	8.49	9.47	44	27.42	31.65	30.74
16	23.14	24.09	27.75	45	58.55	70.67	75.72
17	18.20	18.20	21.49	46	94.11	114.13	121.49
18	9.87	11.14	12.51	47	82.12	112.65	108.52
19	15.87	16.03	13.66	48	26.54	39.41	41.22
20	11.44	11.98	12.56	49	4.95	5.37	4.41
21	12.39	12.39	15.53	50	12.11	14.10	11.17
22	24.83	24.83	32.02	51	6.75	8.61	7.50
23	16.43	18.16	21.55	52	8.83	11.45	9.90
24	6.98	7.83	10.06	53	52.64	71.49	61.14
25	7.49	7.74	6.99	55	3.59	3.93	3.03
26	10.33	11.33	12.32	56	7.33	8.96	8.23
27	6.47	7.18	8.69	57	23.50	29.83	22.11
28	125.81	140.57	172.91	58	221.23	294.59	208.77
29	33.88	38.13	52.00	59	6.45	8.82	6.62
30	14.89	15.24	20.43	60	38.90	52.84	41.96

[a] Les cellules 6 et 54 ne sont pas inclues parce que les taux de chômage observés correspondants étaient zéro.

Les valeurs de $Q^{(T)}(\tilde{\theta})$ pour $T = 54$ et 51 et celles de $Q^{(T)}(\bar{\theta})$ pour les mêmes valeurs de T permettent de tirer des conclusions similaires.

Le tableau 1 donne les efficacités relatives aux estimations du taux de chômage $1 - v$ pour les 60 cellules correspondant aux trois estimations lissées. Les trois sortes d'estimations lissées sont: les pseudo-EMV, les estimations de $Q^{(51)}$ min et celles de $Q^{(54)}$ min. Les variances des pseudo-EMV sont tirés de Kumar et Rao (1986) tandis que celles des estimations de $Q^{(T)}$ min ont été établis à l'aide des éléments diagonaux de $B \wedge_T B'$ dans (3.4). Comme le soulignent Kumar et Rao (1986) à propos des pseudo-EMV, les estimations lissées obtenues par la méthode de $Q^{(T)}$ min produisent des gains d'efficacité considérables par rapport aux estimations d'enquête. Le rapport entre l'efficacité de trace des estimations lissées et l'efficacité de trace des estimations d'enquête est de 17.9, 18.95 et 19.88 respectivement pour la méthode des pseudo-EMV, celle de $Q^{(51)}$ min et celle de $Q^{(54)}$ min. Ainsi, la méthode de $Q^{(T)}$ min produit des estimations lissées légèrement plus efficaces que la méthode des pseudo-EMV. En ce qui a trait à l'efficacité au niveau de la cellule, le tableau 1 indique que les pseudo-EMV se comparent très avantageusement aux estimations efficaces de $Q^{(T)}$ min.

5. CONCLUSION

En ce qui a trait au calcul des pseudo-EMV, la forme usuelle de la fonction de vraisemblance se rapporte à des échantillons aléatoires simples (c.-à-d., échantillonnage multinomial ou échantillonnage multinomial mixte). Les pseudo-EMV produisent effectivement des estimations de paramètres de modèle convergentes sans que l'on ait à estimer la matrice des covariances Γ. Toutefois, ce genre d'estimations ne sont pas asymptotiquement efficaces dans le cas de données d'échantillons complexes. En revanche, les estimations de $Q^{(T)}$ min sont asymptotiquement efficaces en ce qui concerne la catégorie d'estimateurs fondés sur W (les T premières composantes principales du vecteur $\hat{v}$ des estimations d'enquête). Pour examiner la performance relative obtenue à l'aide du pseudo-EMV et du $Q^{(T)}$ min, il serait souhaitable d'effectuer une étude de simulation pour comparer les efficacités. On adapte les estimations aux plans de sondage complexes en utilisant une matrice $\hat{\Gamma}$. Si $\hat{\Gamma}$ n'est pas mal conditionnée, c'est-à-dire qu'elle n'a aucune valeur propre relativement minime, les estimateurs par les MCP sont parfaitement stables et, par conséquent, asymptotiquement efficaces. Dans ce genre de situation, on constatera souvent qu'il n'y a pas de réduction de dimension pour de faibles valeurs de ϵ et que T coïncide avec I. De plus, les estimations de $Q^{(T)}$ min ne perdront pas de leur efficacité par rapport aux estimations par les MCP. Toutefois, en raison des problèmes d'instabilité qui sont si souvent liés aux données qualitatives recoupées, les estimations de $Q^{(T)}$ min devraient bien valoir en robustesse les estimations par les MCP.

REMERCIEMENTS

Les recherches du premier auteur ont été appuyées par Statistique Canada et le Conseil de recherches en sciences naturelles et en génie du Canada.

BIBLIOGRAPHIE

COX, D.R., et HINKLEY, D.W. (1974). *Theoretical Statistics*. London: Chapman and Hall.

FAY, R.E. (1985). A jackknifed chi-squared test for complex samples. *Journal of the American Statistical Association*, 80, 148-157.

IMREY, P.B., KOCH, G.G., et STOKES, M.E. (1982). Categorical data analysis: Some reflections on the log-linear model and logistic regression. Part II: Data analysis. *International Statistical Review*, 50, 35-63.

KOCH, G.G., FREEMAN, D.H. Jr., et FREEMAN, J.L. (1975). Strategies in the multivariate analysis of data from complex surveys. *International Statistical Review*, 43, 59-78.

KUMAR, S., et RAO, J.N.K. (1984). Régression logistique et analyse de données de l'enquête sur la population active. *Techniques d'enquête*, 10, 62-81.

KUMAR, S., et RAO, J.N.K. (1986). On smoothed estimates of unemployment rates from labour force survey data. Dans *Small Area Statistics: An International Symposium '85* (éds. R. Platek et M.P. Singh), Ottawa: Carleton University.

NEYMAN, J. (1949). Contribution to the Theory of the X^2 test. Dans *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability* (éd. J. Neyman), Berkeley: University of California Press, 230-273.

RAO, J.N.K., et SCOTT, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two way tables. *Journal of the American Statistical Association*, 76, 221-230.

RAO, J.N.K., et SCOTT, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, 12, 46-60.

ROBERTS, G.R. (1985). *Contributions to chi-squared tests with survey data*. Thèse de doctorat, Carleton University, Ottawa.

ROBERTS, G., RAO, J.N.K., et KUMAR, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74, 1-12.

SINGH, A.C. (1985). On optimal asymptotic tests for analysis of categorical data from sample surveys. Document de travail, Division des méthodes d'enquêtes sociales, Statistique Canada.

SINGH, A.C., et KUMAR, S. (1986). Categorical data analysis for complex surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (à paraître).