Techniques d'enquête, décembre 1986
Vol. 12, n° 2, pp. 127–144
Statistique Canada

Analyse de régression pour des données d'enquête avec plan de sondage endogène

ARIE TEN CATE[1]

RÉSUMÉ

Cette étude traite de l'influence du plan de sondage sur l'estimation d'un modèle de régression linéaire. L'auteur étudie plus particulièrement les plans de sondage qui dépendent des valeurs de la variable endogène dans la population, c'est-à-dire les plans endogènes (ou informatifs). L'auteur définit aussi un estimateur convergent des coefficients de régression. La variance de cet estimateur correspond à la somme d'un élément lié au plan de sondage et d'un autre lié à la perturbation. L'auteur examine aussi brièvement la régression sans modèle. L'estimateur utilisé dans une régression de ce genre équivaut à l'estimateur utilisé dans une régression avec modèle lorsqu'il y a un plan de sondage endogène.

MOTS CLÉS: Régression; enquête par sondage; plan de sondage endogène.

1. INTRODUCTION

Tout modèle statistique repose fondamentalement sur l'hypothèse selon laquelle la valeur d'une ou de plusieurs variables est tirée d'une distribution de probabilité (par exemple, un modèle de régression avec des perturbations distribuées suivant une loi normale). Dans ce document, nous étudions un ensemble fini d'éléments qui répondent aux règles d'un modèle de ce genre. Cet ensemble d'éléments est appelé la population. Un échantillon est prélevé dans cette population, sans remise. Le présent document a pour objet d'étudier l'influence du plan de sondage sur l'estimation des paramètres du modèle. Cette influence variera essentiellement selon que le plan de sondage sera exogène ou endogène par rapport au modèle. Si le plan de sondage est endogène (ou informatif), les probabilités d'échantillonnage dépendront de la valeur des variables endogènes (dépendantes). Dans ce cas, le plan de sondage devra être pris en considération dans l'estimation des paramètres du modèle. La nature du problème est illustrée à la figure 1 où l'on peut voir un plan d'échantillonnage stratifié. Trois strates sont définies pour la variable endogène du modèle de régression. La strate médiane a une fraction de sondage plus élevée que les deux autres strates. Nous constatons que la pente de la droite de régression estimée, fondée uniquement sur des données échantillonnées, est biaisée par défaut si l'on fait abstraction du plan de sondage. Ce biais persiste avec de grands échantillons. Cela peut être vérifié intuitivement en imaginant que chacun des points (noirs ou blancs) de la figure 1 représente une multitude de points identiques. Même si cette multitude tend vers l'infini, la pente de la droite de régression estimée demeurera biaisée par défaut parce que la forme du nuage ne changera pas.

L'application des méthodes de régression à l'échantillonnage de populations finies fait l'objet d'études de plus en plus nombreuses qui portent sur des problèmes variés. Par exemple, on cherche à savoir comment appliquer des méthodes de régression à l'estimation d'un agrégat de population finie. On s'intéresse aussi à l'estimation de paramètres de population tels que $\Sigma xy / \Sigma x^2$, où la sommation porte sur tous les éléments de la population finie. Nathan (1981) et Smith (1981) passent en revue la littérature sur ce sujet. Un troisième problème auquel on s'intéresse est l'estimation des paramètres d'un modèle de régression au moyen d'un échantillon prélevé dans une population finie. Ce problème peut être résolu assez facilement lorsque le plan de sondage est exogène. Voir Porter (1973, section 1.2), DuMouchel et Duncan (1983) et des ouvrages comme celui de Cramer (1971, p.143). Kmenta (1971, section 8.3)

[1] Bureau central de planification, Van Stolkweg 14, 2585 JR La Haye, Pays-Bas.

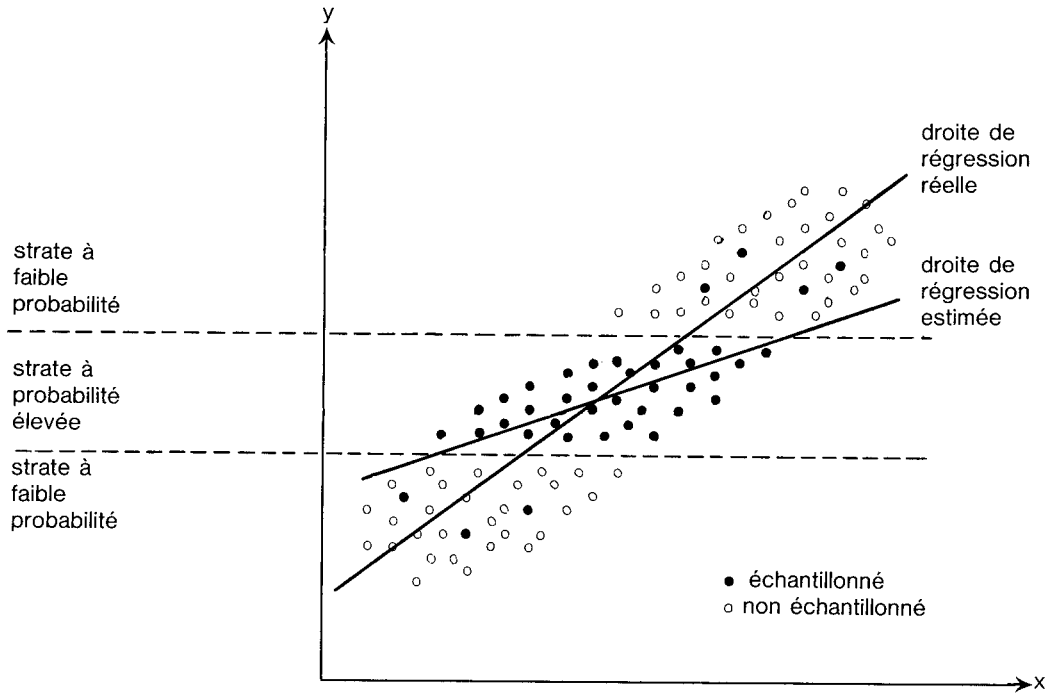

Figure 1. Effet de la stratification endogène sur la droite de régression estimée.

et Johnston (1972, section 9.2) examinent la question des variables explicatives aléatoires, qui est étroitement liée à la question précédente. Voir aussi White (1980a) pour la régression non linéaire. Le sujet qui nous occupe, soit l'analyse de la régression avec plan de sondage endogène, est plus complexe. Hausman et Wise (1981) étudient des plans de sondage endogènes stratifiés dans des conditions très simples: deux strates et un modèle de régression ne comportant qu'une constante. Jewell (1985) définit quelques estimateurs itératifs pour la stratification endogène.

L'analyse de régression avec plan de sondage endogène se rattache au problème de la non-réponse endogène dans l'analyse de régression normale. Voir Heckman (1979). Toutefois, l'analyse de régression avec plan endogène pose un peu moins de difficultés puisque les probabilités d'échantillonnage sont censées être connues: elles font partie intégrante du plan de sondage choisi. Par ailleurs, comme nous le verrons dans la section 6.1, il est plutôt difficile, en règle générale, d'estimer la variance lorsque nous avons un plan endogène.

L'analyse de régression avec plan de sondage endogène peut être comparée à l'analyse par la méthode des logits avec plan de sondage endogène, aussi appelée analyse par la méthode des logits avec échantillonnage fondé sur le choix ou échantillonnage de cas avec groupe témoin. Voir Manski et McFadden (1981, chapitres 1 et 2) et Breslow et Day (1980, section 6.3).

La présente étude se divise comme suit. Dans les sections 2 et 3, nous énonçons les principaux théorèmes, lesquels définissent des estimateurs convergents des paramètres d'un modèle de régression linéaire; ces estimateurs sont établis à partir d'un échantillon prélevé au moyen d'un plan endogène. La convergence est définie à peu près de la même façon que dans l'analyse du biais faite plus haut, sauf que la définition est légèrement plus raffinée. On effectue plusieurs itérations des valeurs de x et on observe les valeurs de y obtenues selon le modèle de régression. Dans les sections 4 et 5, nous étudions la variance des estimateurs des coefficients de régression. Dans la section suivante, nous analysons l'estimation de ces variances. La section 7 porte sur la régression sans modèle tandis que la section 8 traite des divers motifs qui justifient l'utilisation de la régression pondérée. Enfin, la section 9 présente la conclusion.

2. LE MODÈLE, L'ÉCHANTILLON ET UN ESTIMATEUR PAR RÉGRESSION

Dans cette section, nous étudions les propriétés asymptotiques d'un estimateur d'un modèle de régression appliqué à l'échantillonnage sans remise d'une population finie. Il peut sembler contradictoire de parler de théorie asymptotique des échantillons prélevés sans remise dans une population finie puisque de tels échantillons doivent être bornés. La contradiction disparaît si l'on accroît dans la même proportion la taille de la population et celle de l'échantillon, celui-ci n'étant aucunement borné. L'interdépendance des probabilités d'inclusion des unités d'une population dans l'échantillon constitue un autre problème, surtout lorsqu'il s'agit de plans de sondage complexes. Pour résoudre ce problème, nous nous inspirons de Brewer (1979). Le modèle de Brewer nous permet d'utiliser des théorèmes de limites pour des suites de variables indépendantes puis d'appliquer néanmoins les résultats à des plans de sondage complexes. Essentiellement, ce modèle repose sur la notion d'itération évoquée précédemment. Cette notion sera très souvent appliquée dans cette étude. Pour une approche différente, voir Robinson (1982).

Tout d'abord, la structure de la population et le modèle sont connus. Considérons un ensemble fini de N_0 éléments. Chaque élément possède r caractéristiques non aléatoire exogènes de valeur réelle, formant pour l'ensemble fini d'éléments une matrice X_0 de dimension $(N_0 \times r)$. Nous posons l'une des hypothèses fondamentales de cette étude: la population est constituée de K itérations de l'ensemble fini de N_0 éléments, c'est-à-dire qu'elle possède $N \equiv KN_0$ éléments. La matrice des variables exogènes est X, où

$$X = \iota_K \otimes X_0. \tag{1}$$

Dans cette équation, ι_K est le K-vecteur dont tous les éléments égalent l'unité, et $\otimes$ désigne le produit de Kronecker. On obtient des résultats asymptotiques en faisant tendre K vers l'infini.

Les hypothèses du modèle décrivent un modèle linéaire standard. À chacun des N éléments de la population correspond une valeur de la variable aléatoire endogène. On obtient ainsi un N-vecteur y. Nous supposons que

$$E_\xi(y) = X\beta \tag{2}$$

pour un r-vecteur fixe et inconnu β. E_ξ désigne l'espérance mathématique pour toutes les valeurs $y \in R^N$. Définissons ensuite

$$\varepsilon = y - X\beta \tag{3}$$

Nous supposons que les N éléments de ε sont indépendants et identiquement distribués. D'après l'équation (2), nous pouvons affirmer que tous les éléments de ε ont une espérance mathématique nulle. Leur variance est σ^2, c'est-à-dire

$$E_\xi(\varepsilon\varepsilon') = \sigma^2 I. \tag{4}$$

Dans le cas présent, l'échantillonnage est sans remise, conformément à la pratique courante. L'échantillon est défini par une matrice diagonale T de dimension $(N \times N)$ de telle manière que

$$t_{ii} = \begin{cases} 1 \text{ si l'élément } i \text{ de la population est inclus dans l'échantillon} \\ 0 \text{ dans le cas contraire;} \end{cases}$$

pour tout $i = 1, \ldots, N$. Naturellement, T est idempotente. L'espace-échantillon S est l'ensemble de toutes les matrices T ainsi définies. Cet ensemble est défini. Le plan d'échantillonnage est une distribution de probabilité sur les éléments de l'espace-échantillon S. En l'occurrence, ce plan est endogène, c'est-à-dire qu'il dépend de y. Par conséquent, le plan de sondage proprement dit est stochastique. (Un plan qui ne dépend pas de y est appelé exogène ou non informatif.) Partitionnons T en une matrice carrée $K \times K$ formée de blocs $(N_0 \times N_0)$. Supposons que T_k désigne le k-ième bloc diagonal correspondant à la k-ième itération. De même, supposons que y est décomposé en K vecteurs de dimension N_0, de telle manière que $y' = (y_1', y_2', \ldots, y_k', \ldots, y_K')$. En outre, nous supposons que le plan de sondage tient à y dans le sens suivant: les K paires $(T_1, y_1), \ldots, (T_K, y_K)$ sont indépendantes et identiquement distribuées.

L'espérance de S, qui dépend de y (ou de ε), occupe une place importante dans cette étude. Cette espérance est désignée par E_p. Nous définissons donc

$$\Pi \equiv E_p(T). \tag{5}$$

Nous supposons que Π est connue. Ses éléments diagonaux sont appelés probabilités d'inclusion, c'est-à-dire la probabilité que les éléments d'une population soient inclus dans l'échantillon. La matrice Π est partitionnée en une matrice carrée $K \times K$ formée de blocs $(N_0 \times N_0)$. Définissons Π_k le k-ième bloc diagonal, correspondant à la k-ième itération. Soulignons que chaque Π_k est stochastique puisqu'il dépend de y_k. Comme les K paires, $(T_1, y_1), \ldots, (T_K, y_K)$ sont indépendantes et identiquement distribuées, les blocs diagonaux $\Pi_1, \ldots, \Pi_k$ le sont aussi. La relation de dépendance qui existe entre Π_k et y est désignée par une fonction F, de sorte que

$$\Pi_k = F(y_k) \tag{6}$$

pour tous les $k = 1, \ldots, K$. Nous supposons que $F(y_k)$ est non singulière pour chaque y_k. Autrement dit, les probabilités d'inclusion sont toujours positives.

Ce modèle diffère quelque peu de celui de Brewer (1979). En effet, le modèle de Brewer ne comporte pas de variable endogène et, par conséquent, tous les blocs diagonaux Π_k sont égaux et non-aléatoires. On peut aussi comparer ce modèle avec la notion de "constante dans des échantillons itérés", que l'on trouve dans les ouvrages d'économétrie. Voir, par exemple, Theil (1971, p. 364).

Nous sommes maintenant prêts à estimer β. Nous allons considérer les propriétés stochastiques des estimateurs pour toutes les paires $(y, T) \in (R^N \times S)$. L'espérance mathématique correspondante sera désignée par $E_\xi E_p$. Nous considérons donc un estimateur des moindres carrés généralisé de β, disons $\hat{\beta}$, dont les poids sont égaux à la racine carrée des probabilités d'inclusion. Ainsi, nous avons

$$\begin{aligned} \hat{\beta} &\equiv [(\Pi^{-1/2}X)'T(\Pi^{-1/2}X)]^{-1}(\Pi^{-1/2}X)'T(\Pi^{-1/2}y) \\ &= (X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}Ty. \end{aligned} \tag{7}$$

Il faut se rappeler que la matrice Π est connue. Soulignons aussi que X et y ont trait à la population tandis que T opère une sommation sur les éléments de l'échantillon. Au lieu de considérer $\hat{\beta}$ comme un estimateur des moindres carrés généralisé, nous pouvons supposer que tous les éléments de π^{-1} sont des nombres entiers. Alors, si chaque observation i de l'échantillon est reproduite π_{ii}^{-1} fois, $\hat{\beta}$ devient l'estimateur des moindres carrés ordinaire appliqué à cet échantillon grossi. Dans ce contexte, il n'est plus question de racine carrée des probabilités. Voir aussi Hausman et Wise (1981, p. 373). Le théorème principal de cette étude est le suivant.

Théorème 1. Suivant les hypothèses énoncées ci-dessus ((1), (2) et la distribution de ε et de T), l'estimateur des moindres carrés généralisé $\hat{\beta}$, défini dans l'équation (7), est convergent lorsque $K \to \infty$.

Le reste de la section est consacré à la démonstration de ce théorème. Pour les besoins de cette démonstration, nous allons nous servir du lemme ci-dessous, qui sera aussi utilisé dans les démonstrations des théorèmes énoncés plus loin.

Lemme 1. Nous considérons un N-vecteur z, de telle sorte que $z = \iota_k \otimes z_0$, où z_0 est un N_0-vecteur fixe. Nous considérons également un N-vecteur η décomposé de telle sorte que $\eta' = (\eta_1', \eta_2', \ldots, \eta_K')$. Chaque η_k possède N_0 éléments. Supposons que chaque η_k est une fonction de X_0, β et de ε_k, toutes ces fonctions étant identiques. Alors

$$\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}z'\Pi^{-1}T\eta\right) = z_0'E_\xi(\eta_0), \tag{8}$$

où $E_\xi(\eta_0)$ est l'espérance mathématique de n'importe quelle valeur de η_k, la même pour tout k.

Démonstration du lemme 1: Considérons l'espérance de $\Pi_k^{-1}T_k\eta_k$:

$$E_\xi E_p(\Pi_k^{-1}T_k\eta_k) = E_\xi[\Pi_k^{-1}E_p(T_k)\eta_k] = E_\xi(\eta_k), \tag{9}$$

pour tous les k. Comme la distribution de η_k est la même pour chaque k, nous pouvons écrire

$$E_\xi E_p(\Pi_k^{-1}T_k\eta_k) = E_\xi(\eta_0) \tag{10}$$

pour tous les k. De plus, les K vecteurs $z_0'\Pi^{-1}T_k\eta_k$ sont indépendants et identiquement distribués. Par conséquent, le théorème de Khintchine s'applique comme suit,

$$\begin{aligned}\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}z'\Pi^{-1}T\eta\right) &= \operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\sum_k z_0'\Pi_k^{-1}T_k\eta_k\right) = E_\xi E_p(z_0'\Pi_1^{-1}T_1\eta_1)\\ &= z_0'E_\xi E_p(\Pi_1^{-1}T_1\eta_1).\end{aligned} \tag{11}$$

En appliquant l'équation (10) au membre de droite de l'équation (11), on obtient l'équation du lemme. Il est désormais facile de faire la démonstration du théorème 1.

Démonstration du théorème 1: L'estimateur des moindres carrés généralisé du théorème peut s'écrire de la façon suivante:

$$\hat\beta = (X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}Ty = \beta + (X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}T\varepsilon. \tag{12}$$

Donc

$$\begin{aligned}\operatorname*{plim}_{K\to\infty}\hat\beta &= \beta + \left[\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}TX\right)\right]^{-1}\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}T\varepsilon\right)\\ &= \beta + (X_0'X_0)^{-1}X_0'0 = \beta.\end{aligned} \tag{13}$$

L'expression $(X_0'X_0)$ est obtenue par l'application répétée du lemme 1, où les valeurs de z et de η sont remplacées par les colonnes de X. Il convient de souligner que $E_\xi(X_0) = X_0$ étant donné que X_0 est une constante. L'expression $X_0'0$ est obtenue par l'application répétée du lemme 1, où les valeurs de z sont remplacées par les colonnes de X et celles de η *par* ε.

3. ESTIMATION DE LA VARIANCE DE LA PERTURBATION

Le modèle de régression défini dans la section précédente contient deux paramètres: β et σ^2. Le théorème 1 concernait l'estimation de β; cette fois, nous nous penchons sur l'estimation de σ^2 . Dans la présente section, on cherche à démontrer le théorème suivant:

Théorème 2. La variance de l'échantillon pondéré des résidus de y est un estimateur convergent de la variance σ^2 de la perturbation si les poids de l'échantillon égalent l'inverse de la racine carrée des probabilités d'inclusion.

Démonstration: L'estimateur de la variance défini dans le théorème est

$$\hat{\sigma}^2 = (\iota_N' \Pi^{-1} T \iota_N)^{-1} \tilde{e}' \tilde{e} \tag{14}$$

où

$$\tilde{e} \equiv \Pi^{-\frac{1}{2}} T(y - X\hat{\beta}). \tag{15}$$

Posons

$$\tilde{y} \equiv \Pi^{-\frac{1}{2}} T y, \tag{16}$$

$$\tilde{X} \equiv \Pi^{-\frac{1}{2}} T X, \tag{17}$$

et

$$\tilde{\varepsilon} \equiv \Pi^{-\frac{1}{2}} T \varepsilon. \tag{18}$$

Alors

$$\tilde{e} = \tilde{y} - \tilde{X}\hat{\beta} = \tilde{y} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \tag{19}$$

et

$$\tilde{e}'\tilde{e} = \tilde{y}'[I_N - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}']\tilde{y} = (\tilde{X}\beta + \tilde{\varepsilon})'[I_N - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'](\tilde{X}\beta + \tilde{\varepsilon})$$

$$= \tilde{\varepsilon}'\tilde{\varepsilon} - \tilde{\varepsilon}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\varepsilon}. \tag{20}$$

La convergence en probabilité du premier terme du membre de droite de l'équation (20) s'exprime de la façon suivante,

$$\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\tilde{\varepsilon}'\tilde{\varepsilon}\right) = \operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\varepsilon'\Pi^{-1}T\varepsilon\right) = \operatorname*{plim}_{K\to\infty}\left[\frac{1}{K}\iota_N'\Pi^{-1}T\,\mathrm{diag}(\varepsilon)\varepsilon\right]$$

$$= \iota_{N_0}'(\sigma^2\iota_{N_0}) = N_0\sigma^2. \tag{21}$$

Dans cette équation, diag(ε) désigne la matrice diagonale et les valeurs de ε constituent la diagonale principale. Pour en arriver à l'expression ci-dessus, nous nous sommes servis de l'équation (4) et avons appliqué la formule du lemme 1 en substituant ι_N à z et diag(ε)ε à η. Considérons maintenant le second terme du membre de droite de l'équation (20).

$$\operatorname*{plim}_{K\to\infty}\left[\frac{1}{K}\tilde{\varepsilon}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\varepsilon}\right]$$

$$= \left[\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\tilde{X}'\tilde{\varepsilon}\right)\right]'\left[\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\tilde{X}'\tilde{X}\right)\right]^{-1}\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\tilde{X}'\tilde{\varepsilon}\right)$$

$$= \left[\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}T\varepsilon\right)\right]'\left[\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}TX\right)\right]^{-1}\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}T\varepsilon\right)$$

$$= 0'(X_0'X_0)^{-1}0 = 0. \tag{22}$$

Pour en arriver à l'équation (22), nous avons appliqué le lemme 1 de la même manière que dans la détermination de l'équation (13). En combinant les équations (20), (21) et (22), nous obtenons

$$\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\tilde{e}'\tilde{e}\right) = N_0\sigma^2. \tag{23}$$

Enfin, nous appliquons le lemme 1 au premier facteur de l'équation (14) en remplaçant z et η, dans la formule du lemme, par ι_N. Nous obtenons donc

$$\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}\iota_N'\Pi^{-1}T\iota_N\right) = N_0. \tag{24}$$

Étant donné les équations (23) et (24), nous avons

$$\operatorname*{plim}_{K\to\infty}(\hat{\sigma}^2) = \sigma^2, \tag{25}$$

ce qui démontre le théorème. Enfin, il peut être utile de signaler, comme corollaire de l'équation (23), que

$$\left(\frac{1}{N}\right)\tilde{e}'\tilde{e} \tag{26}$$

est un estimateur convergent de σ^2.

4. VARIANCE DE $\hat{\beta}$

Dans cette section, nous définissons la variance asymptotique de l'estimateur $\hat{\beta}$.

Théorème 3. La variance asymptotique de $\hat{\beta}$ est définie

$$\operatorname{Var}(\hat{\beta}) = (X'X)^{-1}X'VX(X'X)^{-1}, \tag{27}$$

où

$$V \equiv E_\xi\left[\operatorname{diag}(\varepsilon)\Pi^{-1}P\Pi^{-1}\operatorname{diag}(\varepsilon)\right], \tag{28}$$

et

$$P \equiv E_p(T\iota\iota' T). \tag{29}$$

Les éléments de P sont désignés les probabilités d'inclusion du second ordre, c'est-à-dire la probabilité qu'une paire quelconque d'éléments de la population soit incluse dans l'échantillon. P a la même diagonale que Π. Le reste de la section est consacré à la démonstration du théorème.

Démonstration: Considérons la distribution asymptotique (lorsque $K \to \infty$) de

$$\begin{aligned} K^{1/2}(\hat{\beta} - \beta) &= K^{1/2}[(X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}Ty - \beta] \\ &= K^{1/2}(X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}T\varepsilon. \end{aligned} \tag{30}$$

Puisque

$$\operatorname*{plim}_{K\to\infty}\left(\frac{1}{K}X'\Pi^{-1}TX\right) = X_0'X_0, \tag{31}$$

$K^{1/2}(\hat{\beta} - \beta)$ a la même distribution asymptotique que δ, ce dernier étant défini

$$\delta \equiv K^{-1/2}(X_0'X_0)^{-1}X'\Pi^{-1}T\varepsilon = K^{-1/2}(X_0'X_0)^{-1}\sum_k X_0'\Pi_k^{-1}T_k\varepsilon_k = K^{-1/2}\sum_k \delta_k, \tag{32}$$

et

$$\delta_k \equiv (X_0'X_0)^{-1}X_0'\Pi_k^{-1}T_k\varepsilon_k, \tag{33}$$

pour tous les $k = 1, \ldots, K$. (Voir, par exemple, Rao (1973), p. 122.) Comme les vecteurs δ_k $(k = 1, \ldots, K)$ sont indépendants et identiquement distribués et qu'en plus

$$\begin{aligned} E_\xi E_p(\delta_k) &= (X_0'X_0)^{-1}X_0'E_\xi E_p(\Pi_k^{-1}T_k\varepsilon_k) \\ &= (X_0'X_0)^{-1}X_0'E_\xi[\Pi_k^{-1}E_p(T_k)\varepsilon_k] \\ &= (X_0'X_0)^{-1}X_0'E_\xi(\varepsilon_k) = 0, \end{aligned} \tag{34}$$

la variance de δ, désignée Var(δ), est la même pour tous les K et est égale à la variance de la distribution asymptotique de δ lorsque $K \to \infty$. Elle peut être exprimée comme suit

$$\operatorname{Var}(\delta) = E_\xi E_p(\delta_k\delta_k') \tag{35}$$

pour tout $k \in \{1, \ldots, K\}$. Puisque les vecteurs δ_k sont indépendants et identiquement distribués, l'équation ci-dessus peut être reformulée comme suit

$$\text{Var}\ (\delta) = \frac{1}{K} \sum_k E_\xi E_p(\delta_k \delta_k')$$

$$= \frac{1}{K}(X_0'X_0)^{-1}\left[E_\xi E_p\left(\sum_k X_0'\Pi_k^{-1}T_k\varepsilon_k\varepsilon_k'T_k\Pi_k^{-1}X_0\right)\right](X_0'X_0)^{-1}$$

$$= K(X'X)^{-1}\left[E_\xi E_p(X'\Pi^{-1}T\varepsilon\varepsilon'T\Pi^{-1}X)\right](X'X)^{-1}$$

$$= K(X'X)^{-1}X'\left\{E_\xi E_p\left[\text{diag}\,(\varepsilon)\Pi^{-1}T\iota\iota'T\Pi^{-1}\text{diag}\,(\varepsilon)\right]\right\}X(X'X)^{-1}$$

$$= K(X'X)^{-1}X'\left\{E_\xi\left[\text{diag}\,(\varepsilon)\Pi^{-1}E_p(T\iota\iota'T)\Pi^{-1}\text{diag}(\varepsilon)\right]\right\}X(X'X)^{-1}. \qquad (36)$$

En divisant Var(δ) par K, on obtient Var($\hat{\beta}$) ce qui conclut la démonstration.

5. DÉCOMPOSITION DE VAR ($\hat{\beta}$)

L'équation (27) peut être formulée autrement:

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} + (X'X)^{-1}X'V^*X(X'X)^{-1} \qquad (37)$$

où, par l'équation (4),

$$V^* \equiv E_\xi\left[\text{diag}(\varepsilon)\,(\Pi^{-1}P\Pi^{-1} - \iota\iota')\,\text{diag}(\varepsilon)\right], \qquad (38)$$

Le premier terme du membre de droite de l'équation (37) pourrait être désigné à bon droit comme la ξ-composante de la variance de $\hat{\beta}$. Cette composante représenterait, à elle seule, la variance de $\hat{\beta}$ si toute la population était échantillonnée. Elle dépend entièrement des variations de la perturbation ε et est l'expression habituellement utilisée pour ce genre de situation. Le deuxième terme du membre de droite de l'équation (37) peut être désigné la p-composante de la variance de $\hat{\beta}$. Cette composante renferme les matrices Π et P, qui décrivent le plan de sondage. Elle ressemble à la formule de la variance de l'estimateur d'un agrégat ou d'une moyenne d'une population finie. Pour mieux comprendre la signification de la p-composante de Var($\hat{\beta}$), nous analysons brièvement dans cette section la théorie sur laquelle repose ce genre d'estimateur.

Nous considérons une population finie de N éléments. (Dans le cas présent, on ne suppose pas l'existence d'un mode d'itération.) À chaque élément de la population correspond une valeur d'une variable non aléatoire réelle, qui est tirée d'un N-vecteur x. Un échantillon est prélevé, sans remise, dans cette population. Comme précédemment, l'échantillon est défini par la matrice diagonale T. De même,

$$\Pi \equiv E_p(T) \qquad (39)$$

et

$$P \equiv E_p(T\iota\iota'T), \qquad (40)$$

ces deux équations désignant respectivement les probabilités d'inclusion du premier ordre et les probabilités d'inclusion du second ordre. Comme, dans ce cas-ci, il n'y a pas de modèle de régression, Π et P sont des matrices dont les éléments sont connus et fixes. Horvitz et Thompson (1952) ont proposé d'estimer l'agrégat de population $X'\iota$ par

$$\hat{X} = x'\Pi^{-1}T\iota \qquad (41)$$

Étant donné l'équation (39), il s'agit évidemment d'un estimateur sans biais. La variance de $\hat{X}$ est

$$\text{Var}(\hat{X}) = E_p(\hat{X}^2) - [E_p(\hat{X})]^2 = E_p(x'\Pi^{-1}T\iota\iota'T\Pi^{-1}x) - x'\iota\iota'x$$

$$= x'(\Pi^{-1}P\Pi^{-1} - \iota\iota')x. \tag{42}$$

Le dernier membre de l'équation (42) est la formule de la variance de l'estimateur d'Horvitz-Thompson, que l'on retrouve dans des ouvrages sur l'échantillonnage, comme celui de Cochran (1977). La forme matricielle est toutefois peu usitée dans ces ouvrages. L'expression entre parenthèses dans le dernier membre de l'équation ci-dessus est la même que celle qui figure dans l'équation (38), qui définit V^*. Comme V^* est contenu dans la formule de la p-composante de Var($\hat{\beta}$), les éléments diagonaux de la p-composante de la matrice de variances Var($\hat{\beta}$) peuvent être considérés comme la ξ-espérance de la p-variance de l'estimateur d'Horvitz-Thompson des totaux de ligne de $(X'X)^{-1}X'$ diag (ε). Ces totaux sont les éléments du vecteur $(X'X)^{-1}X'\varepsilon$.

6. ESTIMATION DE VAR($\hat{\beta}$)

6.1 Cas général

Cette section porte sur l'estimateur de la variance asymptotique Var($\hat{\beta}$). Il est plutôt difficile de trouver un estimateur convergent de Var($\hat{\beta}$) puisque, pour cela, il faut connaître la relation F qui existe entre y et le plan de sondage, telle qu'elle est définie dans la matrice V. En pratique, seul le plan d'échantillonnage pour les valeurs réelles de y peut être connu. En règle générale, le plan ne permet pas, à lui seul, de prévoir ce qu'il deviendrait si y prenait d'autres valeurs. Dans une certaine mesure, on ne parle pas uniquement d'un modèle de régression mais aussi d'un modèle du concepteur même du plan!

Pour le moment, nous supposons que la fonction F est connue et, par conséquent, V est une fonction connue de X et des paramètres du modèle. (Voir la section 6.2 pour un cas particulier.) La fonction V s'exprime comme suit:

$$V = V(\beta, \sigma^2; X), \tag{43}$$

Nous supposons que $V(\beta, \sigma^2, X)$ est une fonction continue. Par souci de concision, nous définissons $\hat{V}$

$$\hat{V} \equiv V(\hat{\beta}, \hat{\sigma}^2; X), \tag{44}$$

où $\hat{\beta}$ et $\hat{\sigma}^2$ sont des estimateurs convergents de β et de σ^2 respectivement. Le reste de cette section est consacré à la démonstration d'un théorème concernant un estimateur convergent de la variance. Dans le cas présent, un estimateur Vâr($\hat{\beta}$) est dit convergent si:

$$\operatorname*{plim}_{K\to\infty} K\text{vâr}(\hat{\beta}) = \lim_{K\to\infty} K\,\text{Var}(\hat{\beta}). \tag{45}$$

Théorème 4. Suivant les hypothèses énoncées ci-dessus, un estimateur convergent de la variance asymptotique Var($\hat{\beta}$) est donné par

$$\text{vâr}(\hat{\beta}) = (X'\Pi^{-1}TX)^{-1}X'T\left(\frac{\hat{V}}{P}\right)TX(X'\Pi^{-1}TX)^{-1}, \tag{46}$$

où $(\hat{V}/P)$ désigne la matrice formée des quotients des éléments de $\hat{V}$ par les éléments correspondants de P.

Démonstration: Considérons tout d'abord la structure de V. Partitionnons la matrice V en une matrice carrée $K \times K$ formée de blocs de dimension $(N_0 \times N_0)$. Le (k, r)-ième bloc non diagonal de V est égal à

$$\begin{aligned} & E_\xi \left[\operatorname{diag}(\varepsilon_k) \Pi_k^{-1} E_p(T_k \iota \iota' T_r) \Pi_r^{-1} \operatorname{diag}(\varepsilon_r) \right] \\ & = E_\xi \left[\operatorname{diag}(\varepsilon_k) \Pi_k^{-1} E_p(T_k) \iota \iota' E_p(T_r) \Pi_r^{-1} \operatorname{diag}(\varepsilon_r) \right] \\ & = E_\xi (\varepsilon_k \varepsilon_r') = 0, \end{aligned} \tag{47}$$

si l'on applique le mode d'itération de la population et le plan d'échantillonnage proposés antérieurement. Les blocs diagonaux de V sont identiques et dépendent de X_0. Par conséquent, $V(\beta, \sigma^2; X)$ peut s'exprimer comme suit:

$$V(\beta, \sigma^2; X) = I_K \otimes V_0(\beta, \sigma^2; X_0), \tag{48}$$

où $V_0(\beta, \sigma^2; X_0)$ est une fonction matricielle $N_0 \times N_0$. En nous servant des équations (1) et (48), nous pouvons reformuler $K\operatorname{Var}(\hat{\beta})$ de la façon suivante,

$$K\operatorname{Var}(\hat{\beta}) = (X_0' X_0)^{-1} X_0' V_0 X_0 (X_0' X_0)^{-1}, \tag{49}$$

où V_0 désigne $V_0(\beta, \sigma^2; X_0)$. Le membre de droite de l'équation (49) est indépendant de K et par conséquent, égal à sa limite lorsque K tend vers l'infini. Considérons maintenant le membre de gauche de l'équation (45).

$$K\hat{\operatorname{var}}(\hat{\beta}) = \left(\frac{1}{K} X' \Pi^{-1} T X \right)^{-1} \left[\frac{1}{K} X' T \left(\frac{\hat{V}}{P} \right) T X \right] \left(\frac{1}{K} X' \Pi^{-1} T X \right)^{-1}. \tag{50}$$

Pour déterminer les équations (13) et (22), nous avons utilisé l'équation

$$\operatorname*{plim}_{K \to \infty} \left(\frac{1}{K} X' \Pi^{-1} T X \right) = X_0' X_0. \tag{51}$$

Donc, compte tenu de l'hypothèse voulant que $V(\beta, \sigma^2; X)$ soit une fonction continue, nous avons

$$\operatorname*{plim}_{K \to \infty} \hat{V}_0 = V_0, \tag{52}$$

où $\hat{V}_0$ désigne $V_0(\hat{\beta}, \hat{\sigma}^2; X_0)$. Par les équations (1), (48) et (52), nous avons

$$\begin{aligned} \operatorname*{plim}_{K \to \infty} \frac{1}{K} X' T \left(\frac{\hat{V}}{P} \right) T X &= \operatorname*{plim}_{K \to \infty} \frac{1}{K} \sum_k \left[X_0' T_k \left(\frac{\hat{V}_0}{P_0} \right) T_k X_0 \right] \\ &= \operatorname*{plim}_{K \to \infty} \frac{1}{K} \sum_k \left[X_0' T_k \left(\frac{V_0}{P_0} \right) T_k X_0 \right] = X_0' V_0 X_0. \end{aligned} \tag{53}$$

Dans l'équation ci-dessus, P_0 désigne $E_p\,(T_k \iota\iota' T_k)$, qui est la même pour tous les $k = 1, \ldots, K$. Le dernier signe d'égalité résulte de l'application du théorème de Khintchine, puisque les termes de la seconde sommation sur k sont indépendants et différemment distribués avec une p-espérance égale à $X_0' V_0 X_0$. Enfin, en combinant les équations (50), (51) et (53), nous obtenons

$$\operatorname*{plim}_{K\to\infty} K\hat{\text{var}}(\hat{\beta}) = (X_0'X_0)^{-1}X_0'V_0X_0(X_0'X_0)^{-1}, \tag{54}$$

ce qui correspond au membre de droite de l'équation (49).

6.2 Échantillonnage stratifié

Ici, le calcul de la matrice $T(\hat{V}/P)T$ est fait pour un cas particulier: 1) les perturbations sont distribuées suivant une loi normale et 2) le plan d'échantillonnage est fondé sur une stratification endogène de telle manière que la probabilité d'inclusion π_{ii} de l'élément i de la population est uniquement fonction du i-ème élément de y, par exemple de $y_{(i)}$. Ainsi,

$$\pi_{ii} = f(y_{(i)}), \tag{55}$$

pour $i = 1, \ldots, N$. À titre d'exemple, prenons l'échantillon stratifié qui est illustré à la figure 1. Le plan d'échantillonnage, en l'occurrence, comprend trois strates. Les éléments de la strate médiane sont ceux pour lesquels la probabilité d'inclusion est la plus élevée. La figure 2 illustre la fonction f correspondante.

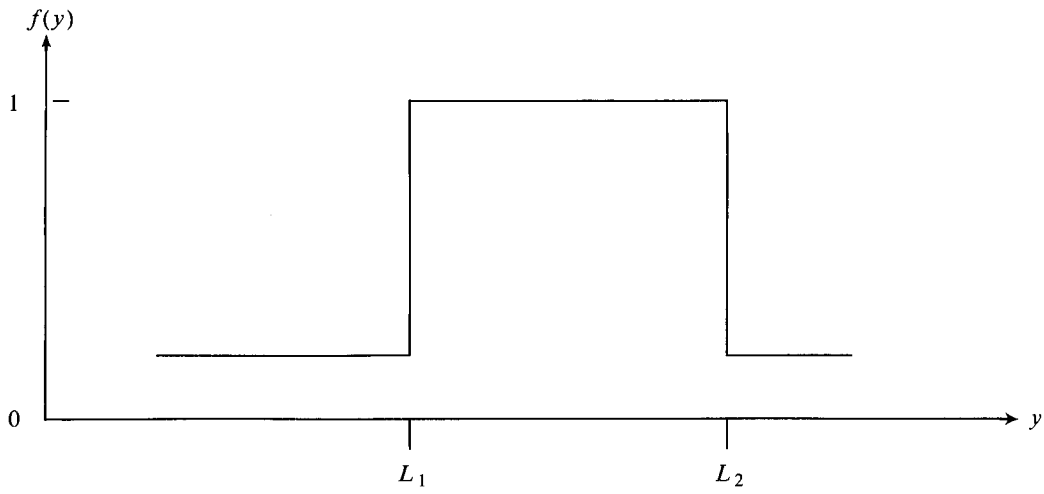

Figure 2. Fonction de probabilité f correspondant à la figure 1.

D'une manière générale, supposons H strates, désignées respectivement $h = 1, \ldots, H$. Définissons $L_0, L_1, \ldots, L_H$ comme les bornes de ces strates. Normalement, $L_0 = -\infty$ et $L_H = +\infty$. Posons $\pi_{(h)}$ comme la probabilité d'inclusion des éléments de la population dans la strate h. De façon plus formelle, la fonction $f(\cdot)$ est telle que $f(y)$ égale $\pi_{(h)}$ si $L_{h-1} \leq y < L_h$. Dans la pratique, on connaît habituellement les valeurs de $\pi_{(h)}$ et de L_h puisque le plan de sondage utilisé en dépend.

Dans l'échantillonnage stratifié, la probabilité d'inclusion de n'importe quelle paire d'éléments d'une population qui n'appartiennent pas à la même strate (probabilité d'inclusion du second ordre) est égale au produit des probabilités d'inclusion respectives de ces éléments (probabilité d'inclusion du premier ordre): l'inclusion du premier élément et l'inclusion du deuxième élément sont des évènements indépendants. Cette règle s'applique à peu près identiquement à n'importe quelle paire d'éléments d'une population qui appartiennent à la même strate. Ainsi, les éléments non diagonaux de P sont approximativement égaux aux éléments non diagonaux de $\Pi\iota\iota'\Pi$. Comme précédemment, P et Π ont la même diagonale. Donc, d'une manière approximative,

$$P = \Pi\iota\iota'\Pi - \Pi^2 + \Pi. \tag{56}$$

Alors

$$V = E_\xi[\mathrm{diag}(\varepsilon)(\iota\iota' - I + \Pi^{-1})\mathrm{diag}(\varepsilon)]$$

$$= E_\xi[\varepsilon\varepsilon' - \mathrm{diag}^2(\varepsilon) + \mathrm{diag}^2(\varepsilon)\Pi^{-1}] = E_\xi[\mathrm{diag}^2(\varepsilon)\Pi^{-1}], \tag{57}$$

compte tenu de l'hypothèse (4). Dans le cas présent, V est donc une matrice diagonale. Alors

$$T\left(\frac{V}{P}\right)T = T\Pi^{-1}E_\xi[\mathrm{diag}^2(\varepsilon)\Pi^{-1}], \tag{58}$$

qui est aussi une matrice diagonale. Considérons maintenant un élément i de la population, qui appartient à l'échantillon. Alors, en appliquant l'équation (58) et en supposant une distribution normale des perturbations,

$$\left[T\left(\frac{\hat{V}}{P}\right)T\right]_{ii} = \frac{1}{\pi_{ii}}\sum_{h=1}^{H}\frac{1}{\pi_{(h)}}\int_{L_{h-1}-x_i'\hat{\beta}}^{L_h-x_i'\hat{\beta}}\varphi(\varepsilon_i;\hat{\sigma}^2)\varepsilon_i^2 d\varepsilon_i$$

$$= \frac{\hat{\sigma}^2}{\pi_{ii}}\left\{\frac{1}{\pi_{(H)}} + \sum_{h=1}^{H-1}\left(\frac{1}{\pi_{(h)}} - \frac{1}{\pi_{(h+1)}}\right)\Psi[(L_h - x_i'\hat{\beta})/\hat{\sigma}]\right\}. \tag{59}$$

Dans l'équation précédente, $\phi(\cdot;\hat{\sigma}^2)$ désigne la fonction de distribution d'une Loi normale, de moyenne nulle et de variance $\hat{\sigma}^2$. La fonction $\Psi(\cdot)$ est définie

$$\Psi(x) \equiv \int_{-\infty}^{x}\varphi(\varepsilon;1)\varepsilon^2 d\varepsilon = \Phi(x) - x\varphi(x;1), \tag{60}$$

où $\Phi(\cdot)$ représente la fonction de distribution d'une Loi normale centrée réduite. Pour en arriver à l'équation (59), nous avons utilisé $\Psi(L_0) = 0$ et $\Psi(L_H) = 1$.

7. RÉGRESSION SANS MODÈLE

7.1 Estimateur convergent

Dans cette section, nous nous éloignons du sujet principal de notre analyse pour traiter de la régression sans modèle. Tout d'abord, disons que la régression sans modèle peut être utile lorsqu'on doute de la validité d'un modèle linéaire. Se référer à Fuller (1975), qui étudie la régression sans modèle pour des plans de sondage particuliers. Van Praag (1981, 1982) examine la régression sans modèle dans le cas d'un échantillonnage répété appliqué à une

distribution de probabilité quelconque. Voir aussi DuMouchel et Duncan (1983). White (1980b, section 3) analyse des sujets connexes. En deuxième lieu, disons que l'estimateur par régression d'un agrégat de la population repose sur la régression sans modèle. Consulter des ouvrages comme celui de Cochran (1977), les comptes rendus de Nathan (1981) et de Smith (1981) et l'ouvrage de Bethlehem et Keller (1983).

La régression sans modèle sert à estimer le vecteur des paramètres d'une population

$$b \equiv (X'X)^{-1}X'y, \tag{61}$$

en l'absence d'hypothèses concernant la distribution de probabilité de y. En fait, X et y sont considérées l'une et l'autre comme des variables non aléatoires. En outre, on utilise le même mode d'itération que dans la section 2, c'est-à-dire,

$$X = \iota_K \otimes X_0, \tag{62}$$

et

$$y = \iota_K \otimes y_0, \tag{63}$$

où y_0 désigne un vecteur fixe de dimension N_0. Comme précédemment, les K matrices diagonales T_k ($k = 1, \ldots, K$) sont indépendantes et identiquement distribuées. Comme dans la section 2, elles décrivent l'échantillon. Les matrices T_k constituent la matrice T. Aucune autre hypothèse n'est faite sur la distribution de T.

Nous pouvons démontrer assez facilement, un peu comme nous l'avons fait dans la section 2, que l'estimateur pondéré $\hat{\beta}$ défini en (7) est un estimateur convergent de b défini par l'équation (61). Voir aussi Jönrup et Rennermalm (1976), qui identifient $\hat{\beta}$ comme un estimateur "approximativement sans biais" de b, et Van Praag (1982, Section 4d), qui analyse, dans le contexte de la régression sans modèle, le "biais dû à la sélection" lorsque les probabilités d'inclusion sont connues.

Comme dans la section 4, il s'ensuit que la variance asymptotique de $\hat{\beta}$, désignée par $\mathrm{Var}_{\mathrm{SM}}(\hat{\beta})$, est définie de la façon suivante pour la régression sans modèle.

$$\mathrm{Var}_{\mathrm{SM}}(\hat{\beta}) = (X'X)^{-1}X'VX(X'X)^{-1}, \tag{64}$$

Notons que

$$e \equiv y - Xb, \tag{65}$$

$$V = \mathrm{diag}(e)\Pi^{-1}P\Pi^{-1}\mathrm{diag}(e), \tag{66}$$

et que P est défini comme dans l'équation (29). De plus, l'équation (66) diffère de l'équation (28) par l'élimination de la ξ-espérance et la substitution de e à ε.

Il serait intéressant de reformuler $\mathrm{Var}_{\mathrm{SM}}(\hat{\beta})$ comme nous l'avons fait pour $\mathrm{Var}(\hat{\beta})$ dans la section 5. Pour cela, nous utiliserons

$$X'e = 0, \tag{67}$$

qui découle directement des équations (61) et (65). Ainsi, $\mathrm{Var}_{\mathrm{SM}}(\hat{\beta})$ peut être reformulée comme suit

$$\begin{aligned}\mathrm{Var}_{\mathrm{SM}}(\hat{\beta}) &= (X'X)^{-1}X'\mathrm{diag}(e)\,(\Pi^{-1}P\Pi^{-1} - \iota\iota')\,\mathrm{diag}(e)X(X'X)^{-1} \\ &\quad + (X'X)^{-1}X'ee'X(X'X)^{-1} \\ &= (X'X)^{-1}X'\mathrm{diag}(e)\,(\Pi^{-1}P\Pi^{-1} - \iota\iota')\,\mathrm{diag}(e)X(X'X)^{-1}. \end{aligned} \tag{68}$$

Le dernier membre de l'équation (68) correspond à la p-composante de la variance définie par l'équation (37). Nous pouvons en conclure que, dans la régression sans modèle, la variance de l'estimateur des coefficients de régression est constituée uniquement de la p-composante, la ξ-composante étant éliminée.

Enfin, sur la base des commentaires exprimés à la fin de la section 5, signalons que le dernier membre de l'équation (68) peut être exprimé comme suit,

$$(X'X)^{-1}\Sigma(X'X)^{-1}, \tag{69}$$

où la matrice Σ est la p-variance-covariance des totaux de ligne de $X'\text{diag}(e)$. Binder (1983, section 4) en arrive à la même conclusion mais selon un raisonnement différent.

8. ANALYSE

Cette section présente des considérations d'ordre pratique sur l'utilisation de poids dans l'analyse de régression. Nous allons examiner brièvement plusieurs raisons qui justifient l'utilisation de poids dans ce genre d'analyse, en égard à ce qui a été traité dans les sections précédentes..

En premier lieu, signalons qu'il peut y avoir une grande différence entre une régression pondérée et une régression non pondérée. Cela est particulièrement vrai lorsque les unités étudiées sont des entreprises commerciales, soit des exploitations agricoles, des entreprises industrielles ou tout autre genre d'entreprises commerciales très différentes au point de vue du nombre d'employés. Au Bureau central de la statistique des Pays-Bas, par exemple, le nombre d'employés est un critère de stratification standard dans les plans d'échantillonnage appliqués aux entreprises commerciales, ce qui crée une très grande diversité de probabilités d'inclusion, les grosses unités ayant relativement plus de chances d'être choisies. Dans les études où la variable endogène est l'emploi, le plan de sondage est nécessairement endogène et, de ce fait, exige une régression pondérée, par laquelle les grosses unités ont des poids peu élevés.

En deuxième lieu, lorsque l'analyse de régression porte sur des unités qui diffèrent beaucoup par leur taille, l'hétéroscédasticité des résidus pose un problème majeur. Il faut alors recourir à la régression pondérée, de même que nous l'avons fait dans la section 2 à cause de l'existence d'un plan endogène. On attribue alors des poids peu élevés aux grosses unités.

Enfin, la troisième raison qui justifie la pondération des données d'échantillon concerne la régression sans modèle, qui a été analysée dans la section 7. Dans ce cas, les poids sont du même genre que ceux définis dans la section 2.

En définitive, il ne semble y avoir aucune raison qui empêche d'incorporer le plan de sondage à l'analyse de régression.

9. CONCLUSION

Cette étude nous a permis d'analyser l'estimation d'un modèle de régression au moyen de données d'une enquête par sondage. Nous avons étudié, plus particulièrement, les échantillons prélevés à l'aide d'un plan de sondage endogène, par exemple des échantillons stratifiés en fonction de la variable endogène. Nous avons vu dans ce cas que la pondération des observations de l'enquête par l'inverse de la racine carrée des fractions de sondage produisait un estimateur convergent. La notion de convergence appliquée ici est une version modifiée de la notion proposée par Brewer (1979). Nous avons défini la variance asymptotique de l'estimateur de même qu'un estimateur convergent de cette variance, celle-ci étant la somme d'un élément lié à l'échantillonnage et d'un élément lié au modèle.

Nous avons aussi considéré la régression sans modèle, laquelle exige la même pondération que la stratification endogène. Dans ce type de régression, la variance de l'estimateur des coefficients de régression n'est formée que de l'élément lié à l'échantillonnage, celui lié au modèle étant éliminé.

Enfin, nous avons présenté des considérations d'ordre pratique sur la pondération des données.

REMERCIEMENTS

L'auteur tient à remercier Abby Israëls, Albert Verbeek et plusieurs arbitres anonymes pour les commentaires qu'ils ont faits sur des versions antérieures de ce document.

BIBLIOGRAPHIE

BETHLEHEM, J.G., et KELLER, W.J. (1983). Weighting sample survey data using linear models. Rapport interne, Département des méthodes statistiques, Bureau central de la statistique des Pays-Bas, Voorburg.

BINDER, D.A. (1983). On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review,* 51, 279-292.

BRESLOW, N., et DAY, N.E. (1980). *Statistical Methods in Cancer Research, Volume 1: the Analysis of Case-Control Studies.* Centre international de recherche sur le cancer, Lyon.

BREWER, K.R.W. (1979). A class of robust sampling designs for large scale surveys. *Journal of the American Statistical Association,* 74, 911-915.

COCHRAN, W.G. (1977). *Sampling Techniques.* New York: Wiley.

CRAMER, J.S. (1971). *Empirical Econometrics.* Amsterdam: North-Holland.

DuMOUCHEL, W.H., et DUNCAN, G.J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association,* 78, 535-543.

FULLER, W.A. (1975). Regression analysis for sample survey. *Sankhyā,* C37, 117-132.

HAUSMAN, J.A., et WISE, D.A. (1981). Stratification on endogenous variables and estimation: the Gary income maintenance experiment. Dans *Structural Analysis of Discrete Data with Econometric Applications,* (éd. C.F. Manski et D. McFadden), Cambridge: MIT Press.

HECKMAN, J.J. (1979). Sample selection bias as a specification error. *Econometrica,* 47, 153-161.

HORVITZ, D.G., et THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* 47, 663-685.

JEWELL, N.P. (1985). Least squares regression with data arising from stratified samples of the dependent variable. *Biometrika,* 72, 11-21.

JOHNSTON, J. (1972). *Econometric Methods.* Tokyo: McGraw-Hill Kogakusha.

JONRUP, H., et RENNERMALM, B. (1976). Regression analysis in samples from finite populations. *Scandinavian Journal of Statistics,* 33-36.

KMENTA, J. (1978). *Elements of Econometrics.* New York: McMillan.

MANSKI, C.F., et McFADDEN, D. (ed.) (1981). *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: MIT Press.

NATHAN, G. (1981). L'inférence statistique basée sur des plans d'échantillonnage complexes. *Techniques d'enquête,* 7, 109-130.

PORTER, R.D. (1973). On the use of survey sample weights in the linear model. *Annals of Economic and Social Measurement,* 2, 141-158.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley.

ROBINSON, P.M. (1982). On the convergence of the Horvitz-Thompson estimator. *Australian Journal of Statistics,* 24, 234-238.

SMITH, T.M.F. (1981). Regression analysis for complex surveys. Dans *Current Topics in Survey Sampling,* (éd. D. Krewski, R. Platek, et J.N.K. Rao), New York: Academic Press, 267-292.

THEIL, H. (1971). *Principles of Econometrics.* New York: Wiley.

VAN PRAAG, B.M.S. (1981). Model-free regression. *Economics Letters,* 7, 139-144.

VAN PRAAG, B.M.S. (1982). The population-sample decomposition with an application to minimum distance estimators. Rapport n° 8218, Centre de recherche en économie publique, Université Leyden.

WHITE, H.,(1980a). Nonlinear regression on cross section data. *Econometrica,* 48, 721-746.

WHITE, H., (1980b). Using least squares to approximate unknown regression functions. *International Economic Review,* 12, 149-170.