Estimateur d'un modèle à variables qualitatives à l'aide d'estimateurs paramétriques et semi-paramétriques : différences théoriques et empiriques

A. Laouar[1]

La présente contribution tente de répondre à la question suivante : quelles sont les différences, tant théoriques qu'empiriques, entre l'estimateur d'un modèle à variables qualitatives à l'aide d'estimateurs paramétriques et semi-paramétriques.

Pour ce faire, nous commençons par présenter les deux estimateurs utilisés, soit le probit et le maximum score.

La présentation qui suit nous permet d'établir que le *maximum score* est théoriquement recommandé dans les situations où le chercheur ne peut, en toute logique, supposer que les aléas sont distribués comme une loi normale ou logistique, dans des situations où celui-ci croit avoir à faire face à des problèmes d'hétéroscédasticité dans les résidus de la régression, ainsi que dans les situations où l'on privilégie le pouvoir prédictif d'un modèle, au détriment de la précision des estimateurs.

Nous constatons également que la littérature n'offre que quelques exemples de comparaisons empiriques entre les deux classes de modèles traitées ici. Nous comparons ensuite *le probit* et le maximum score sur la base d'un jeu de données.

Les méthodes d'estimation de variables dépendantes au caractère discret constituent certainement la classe de modèles non-linéaires la plus utilisée. Dans ces modèles, la variable dépendante est une transformation d'une variable latente qui ne peut être mesurée par elle-même. Ainsi, on n'observe pas le phénomène en tant que tel, mais sa manifestation. Par exemple, la décision d'abandonner. C'est une manifestation de l'utilité que l'individu retire de demeurer aux études, utilité qui n'est pas directement mesurable. Nous estimons un tel modèle en supposant que la

---

[1] Chargé de Cours INPS.

variable latente dépend d'une fonction F de variables exogènes observables (sexe, scolarité des parents, etc.) et d'un terme d'erreur u inobservable.

Par le passé, cette estimation se faisait nécessairement en supposant une structure paramétrique à la fonction F et une de distribution conditionnelle aux variables exogènes pour les aléas u. La justification de ces hypothèses très restrictives réside principalement dans la simplicité. Supposer une loi normale, comme le probit, ou une loi logistique, comme le logit, rend les calculs beaucoup plus simples et permet de faire de l'inférence et des tests basés sur ces lois dont les propriétés sont très connues et documentées.

Cependant, comme en économie, la simplicité a un prix. Ces modèles sujets à de graves problèmes, si la loi supposée de u n'est pas la bonne, c'est-à-dire, celle que suivent réellement les aléas. L'exemple le plus flagrant est le biais inévitable, si l'on ne maximise pas la bonne fonction de vraisemblance, lors de l'estimation d'un modèle comme le probit ou le logit. Et comme nécessairement la loi supposée aux aléas entre dans cette fonction, les estimateurs sont biaisés, si les erreurs de la régression ne suivent pas réellement la loi qui leur est imputée.

Pour éviter ces inconvénients, certains auteurs ont imaginé des méthodes d'estimation qui ne nécessitent pas la supposition d'une loi précise pour u. Ces estimateurs sont nommés semi-paramétriques et ont donné naissance à une riche littérature. Cette littérature a cependant le défaut de souvent demeurer dans le domaine théorique. Ceci est compréhensible, quand on pense que calculer des estimateurs semi-paramétriques ou non paramétrique demande l'élaboration d'algorithmes de calcul numérique, car les équations sont non linéaires, non différentiables. Mais dans les faits, avant de choisir s'il utilise un estimateur paramétrique ou semi-paramétrique, un chercheur doit pouvoir savoir quelles sont les grandes différences empiriques qui existent entre ces méthodes. C'est pourquoi l'idée d'une comparaison entre un estimateur paramétrique et un estimateur semi-paramétrique nous semble intéressante et justifiée.

Ainsi, dans cet article, nous nous proposons d'effectuer une telle comparaison ,en utilisant un jeu de données. La question à laquelle nous désirons répondre est fort simple : dans les faits, quelles sont les différences majeures dans les résultats de l'estimation d'un modèle paramétrique de données qualitatives et d'un modèle semi- paramétrique, comme le maximum score. Nous procéderons en deux étapes : premièrement nous présenterons la théorie avant le probit et le maximum score . Dans la deuxième partie, nous présenterons les données utilisées et rapporterons les résultats de l'estimation et comparerons les résultats de l'estimation avec le probit et avec le maximum score.

1. Discussion théorique

- le probit

La structure standard d'un problème de choix binaire consiste à supposer une variable latente Y_i^* définie par la relation :

$$Y_i^* = \beta' x_i + \mu_i$$

où la variable latente n'est pas observable. Ce que le chercheur observe est une variable binaire de la forme :

$$Y = 1 \quad si \quad y_i^* > 0$$
$$Y = 0 \quad sinon$$

En termes de probabilités associées à cette relation, nous voyons que :

$$Prob(y_i = 1) = Prob(\mu_i > -\beta' x_i)$$
$$= 1 - F(-\beta' x_i)$$
$$Prob(y_i = 0) = 1 - (1 - F(-\beta' x_i))$$
$$= F(-\beta' x_i)$$

où F(.) représente la fonction de distribution cumulative pour μ_i.

Dans le probit, nous supposons que $\mu_i \rightarrow \mathrm{N}(0,\partial^2)$ et nous standardisons les variables telles que :

$$y_i^*/\partial = \beta' x_i/\partial + \mu_i/\partial$$

$$y_i^* = (\beta'/\partial)x_i + e_i \quad ; \quad var(e_i) = var(\mu/\partial) = 1/\partial^2\, var(\mu_i)$$

Nous pouvons en conséquence, exprimer la relation en probabilités comme :

$$F(-\beta' x_i/\partial) = \int_{-\infty}^{-\beta' x_i} 1/\sqrt{(2\pi\partial^2)}\exp\{-1/2e_i^2\}\,de_i$$

Puisque le modèle à estimer est non linéaire, la méthode d'estimation qui s'impose est celle du *maximum de vraisemblance*. Nous cherchons les valeurs de β qui maximisent la vraisemblance tel que, à partir d'un échantillon aléatoire, nous puissions expliquer y_i. La fonction à maximiser s'écrit

$$L = \prod_{i=1}^{n}\left[1 - F(-\beta' x_i/\partial)\right]^{y_i}\left[F(-\beta' x_i/\partial)\right]^{1-y_i}$$

En considérant le logarithme, nous avons :

$$L = \log L = \sum_{i=1}^{n} y_i \log\left[1 - F(-\beta' x_i/\partial)\right] + (1-y_i)\left[F(\beta' x_i/\partial)\right]$$

Rappelons aussi qu'une propriété importante de l'estimateur du maximum de vraisemblance est que pour un vecteur de paramètre à estimer :

$$\hat{\theta} \rightarrow \mathrm{N}\left(\theta, I^{-1}(\theta)\right)$$

Et donc que nous pouvons obtenir une matrice de variance – covariance pour les paramètre à estimer en évaluant $I^{-1}(\theta)$ comme

$$I(\theta) = -E\left[1/\partial\theta\ \partial\theta'\right]$$

C'est-à-dire, en inversant la matrice dite Hessienne

$$\partial^2 L/\partial\theta\ \partial\theta' = \begin{bmatrix} \partial^2 L/\partial^2\theta_1 & \partial^2 L/\partial\theta_1\theta_2 & \cdots & \partial^2 L/\partial\theta_1\theta_k \\ \partial^2 L/\partial\theta_2\theta_1 & \partial^2 L/\partial^2\theta_2 & \cdots & \partial^2 L/\partial\theta_2\partial\theta_k \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2 L/\partial\theta_k\theta_1 & \cdots & \cdots & \partial^2 L/\partial^2\theta_k \end{bmatrix}$$

- Le maximum score

Introduit par Manski (1975), l'estimateur du maximim score est l'estimateur semi-paramétrique le plus connu et utilisé par les économètres dans les situations où la variable dépendante est de type *binaire*. Tout ce dont on a besoin pour dériver l'estimateur du maximum score est de supposer que la distribution F(.) possède une médiane unique de valeur zéro. Cette caractéristique rend l'estimateur intéressant dans des situations particulières ; mais elle complique sérieusement la démonstration mathématique. Pour dériver le probit nous pouvions nous fier aux propriétés de la loi normale qui sont bien documentées et rendent le probit plus facilement compréhensible, en supposant que l'utilisateur est familier avec la loi normale. C'est pourquoi avant d'entrer dans une démonstration formelle, nous nous attarderons à l'intuition derrière l'estimateur.

Exprimons le maximum score, sans le prouver pour l'instant, comme :

$$Max_b S_{N\alpha}(b) = 1/n\sum_{i=1}^{n}\left[Z_i - (1-2\alpha)sgn(x_i b)\right]$$

Ainsi, l'estimateur du maximum score (avec $\alpha' = 0{,}5$), choisira β qui maximise le nombre de fois où la prédiction est du même signe que Z.

Et parce que le signe de $\beta' x$ est le même pour chaque multiple positifs, l'estimateur doit être calculé par rapport à la contrainte :

$$\beta' \beta = 1.$$

Nous procédons à la démonstration du maximum score. Nous supposerons que pour un quantile α compris entre 0 et 1, le α^e quantile de la stribution F(.) est unique et égal à zéro. Mathématiquement, nous dirons que :

$$Prob(y \geq x\beta / x) = 1 - \alpha$$

Ce qui implique que :

$$x\beta > 0 \Leftrightarrow Prob(y = 1/x) > 1 - \alpha$$
$$x\beta = 0 \Leftrightarrow Prob(y = 1/x) = 1 - \alpha$$
$$x\beta < 0 \Leftrightarrow Prob(y = 1/x) < 1 - \alpha$$

et qui lie le paramètre β aux probabilités estimables par le modèle binaire. Posons maintenant que pour $b \in R^k$.

$$X_b = [x \in X : sgn(xb) \neq sgn(x\beta)]$$

Et qu'ainsi, nous pouvons identifier β par rapport à b si et seulement si :

$$Prob(x \in X_b) > 0$$

En sachant que $\| \ \|$ désigne la distance Euclidienne dans l'espace R, posons :

$$B^* = B / \|B\|$$

$$B_0 \equiv \left[b \in R^k : b / \|b\| \neq B^* \right]$$

Manski (1985) démontre qu'il es possible d'identifier le paramètre normalisé B^*, par rapport aux éléments de l'ensemble B, si la ditribution de x n'est pas comprise dans un sous ensemble propre de R^k, qu'il existe au moins un $k \in \{1,2,3,.....K\}$ tel que : $B_k \neq 0$ et qu'au moins une composante du vecteur x ait une densité de Lebesgue positive en tout point. Ces conditions étant satisfaites, un estimateur peut être dérivé en posant β comme un sous-ensemble de $b \in R^k : \|b\| = 1$ avec $\beta^* \in B$, le score dans la population est donné par :

$$\max_b S_\alpha(b) = E[\{y - (1 - 2\alpha)\} sgn(xb)]$$

Et son analogue au niveau de l'échantillon

$$\max_b S_{n\alpha}(b) = 1/n \sum_{i=1}^{n} [y_i - (1 - 2\alpha)] \, sgn(x_i b)$$

Notons que l'estimateur du maximum score peut aussi être exprimé comme :

$$\max_b S_{n\alpha}(b) = 1/n \sum_{i=1}^{n} [z_i - (1 - 2\alpha)] \, sgn(x_i b)$$

$$ou \quad z_i = 2y_i - 1 \ ; \quad donc \quad z_i = -1 \quad si \quad y_i = 0$$

La convergence faible de l'estimateur est prouvée par Manski (1975), dans le même papier où il a introduit la méthode de dérivation du maximum score et la convergence forte est prouvée encore par Manski, en 1985. Il fallut cependant attendre 1990, pour obtenir la preuve que

l'estimateur convergence à la vitesse $N^{1/3}$ ce qui est plus lent que les estimateurs dits standard qui convergent habituellement à la vitesse $N^{1/2}$.

Cela étant, nous voyons que non seulement l'expression du maximum score n'est pas linéaire, mais elle n'est pas non plus dérivable. Or, dans l'estimation du probit, une matrice variance–covariance des éléments du vecteur b peut être obtenue à partir des dérivées secondes de la fonction de vraisemblance. Dans l'estimation d'un estimateur semi-paramétrique et non paramétrique, une telle méthode ne peut de toute évidence être utilisée.

Pour avoir une idée de la précision de nos estimés, nous utiliserons une technique statistique appelée *bootstrapping* développée par Effron (1979). Elle permet de mimer la distribution d'un estimateur ou d'une statistique de test en ré-échantillonnant des données. Il s'agit d'une méthode simple pour trouver des approximations de quantités qui peuvent être difficilement calculables. L'idée est de remplacer la fonction de distribution habituellement connue, du terme d'erreur laquelle est souvent la loi normale par la fonction de distribution des résidus.

L'échantillon utilisé est considéré comme une population dans laquelle on fait des tirages. En ré-échantillonnant, de nouveaux échantillons sont crées et en répétant cette procédure un certain nombre de fois, il est possible d'utiliser la *moyenne* de ces quantités calculées afin d'obtenir une estimation de la valeur de cette quantité « *boostrapée* ». Avec cette technique, le chercheur se sert des données elles-mêmes pour tenter de décrire les propriétés d'un estimateur. Il n'est pas inutile d'en faire un bref rappel.

Soit $\hat{\theta}_n$ l'estimé d'un vecteur θ tirés d'un échantillon $X = \{x_1, x_2, ...x_n\}$. Les estimateurs du boutstrap, $\hat{\theta}(b)m$ avec $b = 1, ..., B$ sont obtenus en tirant un échantillon avec remise de m observations de l'ensemble $X = \{x_1, x_2, ...x_n\}$ et en calculant $\hat{\theta}$ à

chaque tirage. En tirant B échantillon, les caractéristiques de l'estimateur sont calculées à partir de l'ensemble variance :

$$\hat{\theta} = \left[\hat{\theta}(1)_m, ..., \hat{\theta}(\beta)_m\right]$$

Et, en supposant n assez grand, une approximation asymptotique de la matrice de variance-covariance de l'estimateur du maximum score est obtenue par la dérivation moyenne carrée

$$MSD(b) = 1/B \sum_{b=1}^{B} \left[(b_m(b) - b_n)(b_m(b) - b_n)'\right]$$

Avec une estimation de la matrice de variance-covariance, nous avons une idée sur la précision de nos estimateurs ; mais cette précision est loin d'être absolue.

La distribution des estimés obtenus par le maximum score n'est pas bien connue et ses propriétées quand n tend vers l'infini sont encore peu explorées. Ainsi, puisque le bootstap est une technique asymptotique, les variances de nos estimateurs ne sont pas aussi robustes que celle obtenues avec des techniques utilisant les fonctions de vraisemblance comme le probit. Cependant Manski et Thompson (1986) testèrent cette technique à l'aide des simulations de Monté Carlo et conclurent que le bootstrap fonctionne relativement bien et que les éstimés obtenus semblent non biaisés. Ils notent qu'une approche plus conservatrice constitue à utiliser les estimés du boostrap comme borne supérieure de la vraie variance.

1.2. Résultats empiriques

Maintenant que les paramètres théoriques sont posés, il convient de présenter les données qui seront utilisées dans la procédure d'estimation. Les données sont relatives à un établissement scolaire de la région d'Alger. L'objectif visé par cette étude porte sur la modélisation de la performance à l'examen du Baccalauréat et l'évaluation de la pertinence de certains variables explicatives.

Les variables retenues sont :

- Moyenne obtenue à l'examen du BEF (MGN 1) ;
- Moyenne obtenue en fin de cycle secondaire (MGN 2) ;
- Sexe ;
- Education des parents (Edu.p) ;
- Difficultés en maths (Diff.m) ;
- Age.

Probit

Variables	Coefficient	Ecart-type	b/écart
Cte	- 0.953	2.053	0.464
MGN 1	0.150	0.159	0.948
MGN 2	0.016	0.029	0.566
Sexe	- 0.102	0.154	- 0.662
Edu.p	- 0.400	0.240	-1.669
Diff.m	0.095	0.143	0.663
Age	- 0.137	0.090	-1.547

Maximum score

Variables	Coefficient	Ecart-type	b/écart
Cte	0.124	0.329	0.378
MGN 1	0.016	0.307	0.053
MGN 2	0.005	0.054	0.010
Sexe	- 0.355	0.468	- 0.758
Edu.p	0.283	0.437	0.644
Diff.m	0.083	0.302	0.277
Age	-0.028	0.117	-0.242

Pouvoir prédictif

Probit = 262 bonnes prédictions (70.6%) ; Max score : 279 bonnes prédictions (73%)

Conclusion

Comparer un modèle paramétrique et un modèle semi-paramétrique n'est pas chose aisée. La première chose qui nous frappe dans les résultats, c'est la différence entre les coefficients. Autre observation quant aux écarts-types, nous constatons que ceux-ci sont plutôt grands, lorsque nous utilisons le maximum Score. Nos estimations du Maximum Score sont imprécis. Au niveau des écarts-types, nous comprenons mieux ce qui se produit. Comme la théorie asymptotique nous le suggère, les écarts-types du probit sont plus petits que ceux du maximum score, ce qui correspond avec le fait que le Maximum score, de par l'estimation par bootstrap des écarts types et sa distribution complexe, converge moins que le probit.

Compte tenu de ces différences fondamentales entre les résultats avec nos deux types d'estimateurs, la question qui nous a guidé depuis le début de notre analyse est de savoir quelle méthode choisir entre le probit ou le Maximum score demeure ouverte.

Si l'on se base uniquement sur le pouvoir prédictif des méthodes d'estimation, nous pourrions dire que le maximum score fait *un peu mieux* que le probit.

Si le chercheur a de bonnes raisons de croire que l'hypothèse de normalité des erreurs est contredite par ses données et /ou que les données présentent des éléments hétéroscédastiques dans leur lien avec les erreurs de la régression, le maximum est théoriquement supérieur. Mais choisir le Maximum score est une décision coûteuse en termes de précision.

Bibliographie

- **Dagenais, Marcel ; Montmarquette, Nathalie** ; Le retour à l'école ; un modèle économétrique , Rapport de Recherche CIRANO 2000.

- **Green,William H;** Econometric Analysis,third Edition, Upper Saddle River,1997.

- **Horowitz, Joel**; A smoothed Maximum Score Estimator for binary Response Model, Econometrica, Vol 60,1992;

- **Maddala. G**; Limited dependant and qualitative variable in econometrics, Cambridge University Press, NY, 1985.

- **Manski, Charles F**; Maximum score estimation of the stochastic utility function, Journal of econometric, Vol3,1975.

$$I(\theta) = -E[1/\partial\theta\ \partial\theta']$$

C'est-à-dire, en inversant la matrice dite Hessienne

$$\partial^2 L/\partial\theta\ \partial\theta' = \begin{bmatrix} \partial^2 L/\partial^2\theta_1 & \partial^2 L/\partial\theta_1\theta_2 & \cdots & \partial^2 L/\partial\theta_1\theta_k \\ \partial^2 L/\partial\theta_2\theta_1 & \partial^2 L/\partial^2\theta_2 & \cdots & \partial^2 L/\partial\theta_2\partial\theta_k \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2 L/\partial\theta_k\theta_1 & \cdots & \cdots & \partial^2 L/\partial^2\theta_k \end{bmatrix}$$

- *Le maximum score*

Introduit par Manski (1975), l'estimateur du maximim score est l'estimateur semi-paramétrique le plus connu et utilisé par les économètres dans les situations où la variable dépendante est de type *binaire*. Tout ce dont on a besoin pour dériver l'estimateur du maximum score est de supposer que la distribution F(.) possède une médiane unique de valeur zéro. Cette caractéristique rend l'estimateur intéressant dans des situations particulières ; mais elle complique sérieusement la démonstration mathématique. Pour dériver le probit nous pouvions nous fier aux propriétés de la loi normale qui sont bien documentées et rendent le probit plus facilement compréhensible, en supposant que l'utilisateur est familier avec la loi normale. C'est pourquoi avant d'entrer dans une démonstration formelle, nous nous attarderons à l'intuition derrière l'estimateur.

Exprimons le maximum score, sans le prouver pour l'instant, comme :

$$Max_b S_{N\alpha}(b) = 1/n \sum_{i=1}^{n} [Z_i - (1-2\alpha)\, sgn(x_i b)]$$

Ainsi, l'estimateur du maximum score (avec $\alpha' = 0{,}5$), choisira β qui maximise le nombre de fois où la prédiction est du même signe que Z.

Et parce que le signe de $\beta' x$ est le même pour chaque multiple positifs, l'estimateur doit être calculé par rapport à la contrainte :

$\beta' \beta = 1.$

Nous procédons à la démonstration du maximum score. Nous supposerons que pour un quantile α compris entre 0 et 1, le α^e quantile de la stribution F(.) est unique et égal à zéro. Mathématiquement, nous dirons que :

$$Prob(y \geq x\beta / x) = 1 - \alpha$$

Ce qui implique que :

$$x\beta > 0 \Leftrightarrow Prob(y = 1/x) > 1 - \alpha$$
$$x\beta = 0 \Leftrightarrow Prob(y = 1/x) = 1 - \alpha$$
$$x\beta < 0 \Leftrightarrow Prob(y = 1/x) < 1 - \alpha$$

et qui lie le paramètre β aux probabilités estimables par le modèle binaire. Posons maintenant que pour $b \in R^k$.

$$X_b = [x \in X : sgn(xb) \neq sgn(x\beta)]$$

Et qu'ainsi, nous pouvons identifier β par rapport à b si et seulement si :

$$Prob(x \in X_b) > 0$$

En sachant que $\| \ \|$ désigne la distance Euclidienne dans l'espace R, posons :

$$B^* = B / \|B\|$$

$$B_0 \equiv \left[b \in R^k : b / \|b\| \neq B^* \right]$$

Manski (1985) démontre qu'il es possible d'identifier le paramètre normalisé B^*, par rapport aux éléments de l'ensemble B, si la ditribution de x n'est pas comprise dans un sous ensemble propre de R^k, qu'il existe au moins un $k \in \{1,2,3,.....K\}$ tel que : $B_k \neq 0$ et qu'au moins une composante du vecteur x ait une densité de Lebesgue positive en tout point. Ces conditions étant satisfaites, un estimateur peut être dérivé en posant β comme un sous-ensemble de $b \in R^k : \|b\| = 1$ avec $\beta^* \in B$, le score dans la population est donné par :

$$\max_b S_\alpha(b) = E[\{y - (1 - 2\alpha)\} sgn(xb)]$$

Et son analogue au niveau de l'échantillon

$$\max_b S_{n\alpha}(b) = 1/n \sum_{i=1}^{n} [y_i - (1 - 2\alpha)] \, sgn(x_i b)$$

Notons que l'estimateur du maximum score peut aussi être exprimé comme :

$$\max_b S_{n\alpha}(b) = 1/n \sum_{i=1}^{n} [z_i - (1 - 2\alpha)] \, sgn(x_i b)$$

$$ou \quad z_i = 2y_i - 1 \ ; \quad donc \quad z_i = -1 \quad si \quad y_i = 0$$

La convergence faible de l'estimateur est prouvée par Manski (1975), dans le même papier où il a introduit la méthode de dérivation du maximum score et la convergence forte est prouvée encore par Manski, en 1985. Il fallut cependant attendre 1990, pour obtenir la preuve que

l'estimateur convergence à la vitesse $N^{1/3}$ ce qui est plus lent que les estimateurs dits standard qui convergent habituellement à la vitesse $N^{1/2}$.

Cela étant, nous voyons que non seulement l'expression du maximum score n'est pas linéaire, mais elle n'est pas non plus dérivable. Or, dans l'estimation du probit, une matrice variance–covariance des éléments du vecteur b peut être obtenue à partir des dérivées secondes de la fonction de vraisemblance. Dans l'estimation d'un estimateur semi-paramétrique et non paramétrique, une telle méthode ne peut de toute évidence être utilisée.

Pour avoir une idée de la précision de nos estimés, nous utiliserons une technique statistique appelée *bootstrapping* développée par Effron (1979). Elle permet de mimer la distribution d'un estimateur ou d'une statistique de test en ré-échantillonnant des données. Il s'agit d'une méthode simple pour trouver des approximations de quantités qui peuvent être difficilement calculables. L'idée est de remplacer la fonction de distribution habituellement connue, du terme d'erreur laquelle est souvent la loi normale par la fonction de distribution des résidus.

L'échantillon utilisé est considéré comme une population dans laquelle on fait des tirages. En ré-échantillonnant, de nouveaux échantillons sont crées et en répétant cette procédure un certain nombre de fois, il est possible d'utiliser la *moyenne* de ces quantités calculées afin d'obtenir une estimation de la valeur de cette quantité « *boostrapée* ». Avec cette technique, le chercheur se sert des données elles-mêmes pour tenter de décrire les propriétés d'un estimateur. Il n'est pas inutile d'en faire un bref rappel.

Soit $\hat{\theta}_n$ l'estimé d'un vecteur θ tirés d'un échantillon $X = \{x_1, x_2, ...x_n\}$. Les estimateurs du boutstrap, $\hat{\theta}(b)m$ avec $b = 1, ..., B$ sont obtenus en tirant un échantillon avec remise de m observations de l'ensemble $X = \{x_1, x_2, ...x_n\}$ et en calculant $\hat{\theta}$ à

chaque tirage. En tirant B échantillon, les caractéristiques de l'estimateur sont calculées à partir de l'ensemble variance :

$$\hat{\theta} = \left[\hat{\theta}(1)_m, ..., \hat{\theta}(\beta)_m\right]$$

Et, en supposant n assez grand, une approximation asymptotique de la matrice de variance-covariance de l'estimateur du maximum score est obtenue par la dérivation moyenne carrée

$$MSD(b) = 1/B \sum_{b=1}^{B} \left[(b_m(b) - b_n)(b_m(b) - b_n)'\right]$$

Avec une estimation de la matrice de variance-covariance, nous avons une idée sur la précision de nos estimateurs ; mais cette précision est loin d'être absolue.

La distribution des estimés obtenus par le maximum score n'est pas bien connue et ses propriétées quand n tend vers l'infini sont encore peu explorées. Ainsi, puisque le bootstap est une technique asymptotique, les variances de nos estimateurs ne sont pas aussi robustes que celle obtenues avec des techniques utilisant les fonctions de vraisemblance comme le probit. Cependant Manski et Thompson (1986) testèrent cette technique à l'aide des simulations de Monté Carlo et conclurent que le bootstrap fonctionne relativement bien et que les éstimés obtenus semblent non biaisés. Ils notent qu'une approche plus conservatrice constitue à utiliser les estimés du boostrap comme borne supérieure de la vraie variance.

1.2. Résultats empiriques

Maintenant que les paramètres théoriques sont posés, il convient de présenter les données qui seront utilisées dans la procédure d'estimation. Les données sont relatives à un établissement scolaire de la région d'Alger. L'objectif visé par cette étude porte sur la modélisation de la performance à l'examen du Baccalauréat et l'évaluation de la pertinence de certains variables explicatives.

Les variables retenues sont :

- Moyenne obtenue à l'examen du BEF (MGN 1) ;
- Moyenne obtenue en fin de cycle secondaire (MGN 2) ;
- Sexe ;
- Education des parents (Edu.p) ;
- Difficultés en maths (Diff.m) ;
- Age.

Probit

Variables	Coefficient	Ecart-type	b/écart
Cte	- 0.953	2.053	0.464
MGN 1	0.150	0.159	0.948
MGN 2	0.016	0.029	0.566
Sexe	- 0.102	0.154	- 0.662
Edu.p	- 0.400	0.240	-1.669
Diff.m	0.095	0.143	0.663
Age	- 0.137	0.090	-1.547

Maximum score

Variables	Coefficient	Ecart-type	b/écart
Cte	0.124	0.329	0.378
MGN 1	0.016	0.307	0.053
MGN 2	0.005	0.054	0.010
Sexe	- 0.355	0.468	- 0.758
Edu.p	0.283	0.437	0.644
Diff.m	0.083	0.302	0.277
Age	-0.028	0.117	-0.242

Pouvoir prédictif

Probit = 262 bonnes prédictions (70.6%) ; Max score : 279 bonnes prédictions (73%)

Conclusion

Comparer un modèle paramétrique et un modèle semi-paramétrique n'est pas chose aisée. La première chose qui nous frappe dans les résultats, c'est la différence entre les coefficients. Autre observation quant aux écarts-types, nous constatons que ceux-ci sont plutôt grands, lorsque nous utilisons le maximum Score. Nos estimations du Maximum Score sont imprécis. Au niveau des écarts-types, nous comprenons mieux ce qui se produit. Comme la théorie asymptotique nous le suggère, les écarts-types du probit sont plus petits que ceux du maximum score, ce qui correspond avec le fait que le Maximum score, de par l'estimation par bootstrap des écarts types et sa distribution complexe, converge moins que le probit.

Compte tenu de ces différences fondamentales entre les résultats avec nos deux types d'estimateurs, la question qui nous a guidé depuis le début de notre analyse est de savoir quelle méthode choisir entre le probit ou le Maximum score demeure ouverte.

Si l'on se base uniquement sur le pouvoir prédictif des méthodes d'estimation, nous pourrions dire que le maximum score fait *un peu mieux* que le probit.

Si le chercheur a de bonnes raisons de croire que l'hypothèse de normalité des erreurs est contredite par ses données et /ou que les données présentent des éléments hétéroscédastiques dans leur lien avec les erreurs de la régression, le maximum est théoriquement supérieur. Mais choisir le Maximum score est une décision coûteuse en termes de précision.

Bibliographie

- **Dagenais, Marcel ; Montmarquette, Nathalie** ; Le retour à l'école ; un modèle économétrique , Rapport de Recherche CIRANO 2000.

- **Green,William H;** Econometric Analysis,third Edition, Upper Saddle River,1997.

- **Horowitz, Joel**; A smoothed Maximum Score Estimator for binary Response Model, Econometrica, Vol 60,1992;

- **Maddala. G**; Limited dependant and qualitative variable in econometrics, Cambridge University Press, NY, 1985.

- **Manski, Charles F**; Maximum score estimation of the stochastic utility function, Journal of econometric, Vol3,1975.