TECHNIQUES ALTERNATIVES D'ESTIMATION ET TESTS EN PRÉSENCE D'ERREURS DE MESURE SUR LES VARIABLES EXPLICATIVES

Rapport de recherche

présenté par

Francois-E. Racicot

Pour l'obtention du grade de

MAITRISE ES SCIENCES (ÉCONOMIQUE)

(Spécialisation Économétrie)

Dirigé par

Prof. Marcel G. Dagenais

Département de Sciences Économiques

Faculté des Arts et des Sciences

UNIVERSITÉ DE MONTRÉAL

(Mars 1993)

TABLES DES MATIERES

INTRODUCTION

La plupart des données utilisées dans les études économiques empiriques contiennent des erreurs de mesure. Ces erreurs sont relativement plus importantes dans les données macro-économiques [Morgenstern (1972), Langanskens et Van Rickeghem (1974)], mais elles sont aussi présentes dans la plupart des études micro-économiques. Comme l'a souligné Morgenstern (1972), ce phénomène n'a jamais été pris sérieusement en considération dans le développement des techniques économétriques. Actuellement, la plupart des livres d'économétrie contiennent une section où il est démontré que dans les modèles de régression, la présence d'erreurs de mesure dans les variables explicatives mènera à la non-convergence de l'estimateur des moindres carrés ordinaires (M.C.O). Bien que dans les applications, les auteurs avertissent généralement les lecteurs du fait que la présence possible d'erreurs de mesure peut biaiser les résultats, aucun effort particulier n'est habituellement fait pour atténuer ces biais et aucun test sur la présence de ces erreurs n'est effectué en utilisant, par exemple, le test de Hausman(1978). L'attitude de la plupart des chercheurs qui font de l'application est probablement dûe, dans un bon nombre de cas, au fait qu'il n'est pas toujours facile de vérifier si les variables instrumentales disponibles utilisées pour obtenir des estimateurs sans biais, satisfont les conditions requises pour justifier leur utilisation. Dans d'autres cas, les instruments ne sont simplement pas accessibles aux chercheurs, ou encore, les coûts à défrayer pour obtenir ces données semblent trop grands par rapport aux bénéfices provenant de l'utilisation de ces instruments.

Le but premier de ce rapport de recherche est donc de présenter:

1) un nouvel estimateur β_H (et sa version alternative β_E), qui est une combinaison linéaire optimale de deux estimateurs déjà connus (β_D et β_P). Alors que l'estimateur des M.C.O. est calculé à partir de moments échantillonnaux d'ordre deux, les estimateurs β_H, β_E, β_D et β_D utilisent des moments échantillonnaux d'ordre 3 et 4. Par ailleurs, alors que les estimateurs β_D et β_P semblent en pratique très erratiques, celà n'est pas le cas pour β_H et β_E. l'estimateur β_H a la qualité d'être sans biais sous l'hypothèse H0 (absence d'erreurs de mesure sur les variables explicatives) et d'être convergent sous l'hypothèse H1 (présence d'erreurs de mesure sur les variables explicatives). De plus, en présence d'erreurs de mesure sur les variables explicatives, la probabilité d'erreur de type I associée aux tests t (t de Student) effectués sur les paramètres, est toujours voisine de la valeur désirée lorsqu'on utilise β_H, même s'il y a des erreurs de mesure. Ce n'est pas le cas pour les M.C.O.

2) un test pour déceler la présence d'erreurs de mesure sur les variables explicatives.

3) les d'expériences de Monte Carlo sur les estimateurs β_H et β_E. Le but de ces expériences est de mesurer la performance relative de ces estimateurs par rapport à l'estimateur des M.C.O. La puissance du nouveau test a également été investiguée à partir des expériences de Monte Carlo.

Une introduction aux erreurs de mesure est aussi présentée au début du rapport, dans le but d'expliciter la provenance et les causes de ces erreurs.

CHAPITRE 1

Ce chapitre se divise en deux parties: I Provenance et causes des erreurs sur les variables, II Conséquences des erreurs sur les variables explicatives pour l'inférence statistique.

Partie I:

I.a) INFORMATIONS SOUSTRAITES DE FAÇON DÉLIBÉRÉE ET MENSONGES:

Souvent les statistiques économiques et sociales sont basées sur des réponses évasives et des mensonges délibérés de différents types. Les mensonges que l'on retrouve le plus fréquemment s'expliquent dans la plupart des cas, par une peur des pouvoirs fiscaux, par une incompréhension ou un par désir de tromper ses concurrents.

Il est souvent plus payant pour le monde des affaires, de cacher des informations. Il existe dans ces milieux, des secrets bien gardés et la loi ne peut pas toujours forcer la divulgation de ces informations.

L'incitation à les dissimuler est aussi fortement influencée par la situation de compétition: " plus grande est la prédominance des monopoles, quasi monopoles ou oligopoles, moins dignes de confiance sont de nombreuses statistiques provenant de ces industries " (Morgenstern 1972, p.17). Celà vaut en particulier pour les statistiques de prix à cause des rabais accordés à certains clients.

Dans les gouvernements, il arrive aussi qu'il y ait des falsifications de statistiques. On peut citer le cas de la Grande-Bretagne où, durant la deuxième guerre mondiale, " l'exchange equalisation account " du gouvernement aurait supprimé pendant un certain temps, toutes les statistiques sur ses avoirs en or.

Enfin, notons qu'il est difficile de falsifier certaines informations quand un système (ou une organisation) est bien compris et décrit; mais souvent, les organisations sociales, sont loin d'être bien comprises et décrites; alors il devient possible de falsifier certaines données statistiques. Certaines expériences sur des firmes individuelles ont montré, par exemple, que de nombreuses falsifications de tableaux de production ne peuvent être découvertes, même au moyen des contrôles comptables les plus précis. Quand il s'agit de l'enregistrement de prix, des mouvements de biens, les possibilités d'erreurs et falsification sont accrues.

En résumé, on peut voir qu'il y a trois sources possibles de fausses statistiques:

1. " L'observateur en faisant un choix de quoi et combien observer, introduit un biais qui est impossible à éviter car un phénomène complexe ne peut jamais être décrit de façon exhaustive "(Morgenstern 1972, P.17). Ce biais est commun à toutes les sciences et ne nous intéresse pas ici.

2. L'observateur peut dissimuler de l'information pour satisfaire à ses hypothèses ou à des fins politiques.

3. Les répondants peuvent mentir à l'enquêteur ou ne pas savoir la réponse exacte. Ce sera le cas, par exemple, de la richesse personnelle, où l'individu ne sait pas exactement combien valent ses biens matériels au prix du marché; mais n'a seulement qu'une approximation de cette valeur.

I.b) QUALIFICATIONS DES OBSERVATEURS:

Souvent les observateurs choisis pour la collecte de statistiques économiques n'ont pratiquement aucune formation dans le domaine ou en possèdent très peu et fréquemment, ils sont engagés seulement à cette seule occasion. " Ceci est une source de la forme la plus sérieuse d'erreur en masse " (Morgenstern 1972, P.24).

Même des personnes entraînées spécialement pour faire des enquêtes ne sont pas des observateurs au sens scientifique. " Un observateur scientifique est l'astronome à son télescope, un chimiste observant les changements de concentration d'une solution dans le temps, le biologiste déterminant le comportement héréditaire de certaines cellules, etc " (Morgenstern 1972, P.25). Toutes ces personnes sont des hommes de science, c'est-à-dire qu'ils n'agissent pas par l'intermédiaire d'agents qui chagent fréquemment. Généralement, dans les sciences sociales, on ne retrouvera pas cette situation, sauf dans des cas précis. Par exemple, on ne verra pas un économiste ou un statisticien aller cueillir lui-même les statistiques. La masse de chiffres nécessaires rendrait cela physiquement impossible. On doit donc faire confiance aux observateurs (non-scientifiques). " Mais il est bien connu d'après l'expérience de l'échantillonnage (où, si le travail est fait correctement, on traite des statistiques strictement déterminées même si elles ne sont toujours parfaitement appropriées), que la réponse est très différente suivant le type de l'observateur, même si ce dernier est qualifié, et devrait se montrer -miraculeusement- minimalement exempt de biais. " (Morgenstern 1972, P.25).

I.c) ERREURS CAUSÉES PAR LES QUESTIONNAIRES:

Pour obtenir des statistiques économiques et sociales, on a souvent recours aux questionnaires, oraux ou écrits. Dans ces derniers, on peut retrouver plusieurs centaines de questions posées à la même firme ou à un même individu. Alors il n'est pas impossible que des erreurs puissent apparaître (et effectivement apparaissent) dans la mise au point de ces questionnaires et dans les réponses.

Il arrive souvent que les questionnaires ne soient pas assez clairs et donnent lieu à des ambiguïtés de sorte que plusieurs réponses existent pour une seule question. Il peut arriver aussi que les questions posées demandent aux répondeurs une connaissance plus élevée qu'ils ne possèdent. Certaines questions posées invitent au mensonge ou aux échappatoires, surtout si elles sont très nombreuses. Alors elles incitent le ou les répondants à donner des réponses sommaires, ceci pour économiser effort, temps et argent. On arrive aussi à des résultats très différents selon que l'on pose les questions soit par la poste, le téléphone, le porte-à-porte, etc. Cela est bien connu; chaque forme d'interrogation introduit un biais de nature différente.

" Le processus consistant à poser des questions et à obtenir des réponses est délicat sur le plan psychologique. En dehors du mensonge et du refus de donner de l'information, il y a l'oubli, les suggestions de l'interrogateur, avec leur propre biais comme conséquence, le manque de compréhension de la question, etc "(Morgenstern 1972, P.26).

En statisque, le domaine des questionnaires a évolué ainsi que théorie qui les concernes; celle-ci n'est cependant pas complète. " En fait on peut même douter que la description qualitative et l'énumération de ses caractéristiques soient complètes "(Morgenstern 1972, P.27).

L'idée ici est de voir que les formes d'enquêtes mentionnées plus haut, peuvent créer des données biaisées et dans certains cas aberrantes. Il existe des techniques pour éliminer les données aberrantes, mais un problème se pose: si on prend une mauvaise méthode, alors il y a risque d'introduire d'autres erreurs dans les données qui ne sont pas particulièrement faciles à detecter.

" On a introduit ce sujet pour montrer que l'on peut quelquefois découvrir un fait évident ou frappant, et l'identifier comme une erreur ou une distorsion; mais derrière celle-ci, il y a généralement de nombreuses autres erreurs du même type qui restent cachées et insaisissables "(Morgenstern 1972, P.29)

I.d) ERREURS CAUSÉES PAR LES INSTRUMENTS:

En sciences économiques et sociales, les erreurs d'instruments jouent un rôle dans la formation des statistiques économiques; mais ce rôle n'est peut-être pas aussi important que dans les sciences naturelles. Encore aujourd'hui, les instruments principalement utilisés dans la formation des statistiques économiques sont des multitudes d'êtres humains et ceci pour: compiler, interpréter, classer, ou dénombrer les données obtenues. Évidement les erreurs causées par les humains sont plus grandes que celles faites par les machines comme les ordinateurs, les trieuses, les codeuses, etc.

" Les instruments nécessitent souvent des transcriptions ordinaires mais fréquentes, pendant lesquelles des erreurs se produisent. De façon générale, ces erreurs ajoutent ce que l'on appelle un <<bruit>> à la variance propre des données "(Morgenstern 1972, P.36).

Ces erreurs sont surtout dues aux interactions hommes-machines. C'est-à-dire, la machine n'introduit pas d'erreur (sauf erreur d'arrondi); mais c'est l'homme qui en crée par mégarde car l'homme n'est pas une machine. " Mais le poids relatif des erreurs humaines et des erreurs non-humaines, se modifie au cour du temps "(Morgenstern 1972, P.38).

I.e) AUTOCORRELATIONS:

Il y a un grand nombre de raisons justifiant également l'apparition de l'autocorrélation des erreurs de mesure dans les séries chronologiques. Une cause fréquente serait le biais des statistiques réapparaissant au cours du temps. Un grand nombre de causes pourront faire que les statisques sur-estiment ou sous-estiment les vraies valeurs des variables, dans une proportion plus ou moins constante. Le biais apparaît pour de multiples raisons. Les genres d'erreurs pouvant mener à un biais seraient par exemple: des omissions et des doubles comptages, des classes d'objets négligées alors qu'elles devraient être prises en considération.

" Même en l'absence d'erreurs dans les séries temporelles, des observations successives tendent fréquemment à être autocorrélées; par exemple le volume de production de mai dépend de celui d'avril, etc. En d'autre termes, les séries temporelles économiques ne sont normalement pas des suites de nombres aléatoires, chacune indépendante des autres " (Morgenstern 1972, P.50).

Pour avoir de l'autocorrélation des erreurs, il faut que le biais soit persistant et de direction constante (positive ou négative). " La gravité du biais peut être grande ou non. Elle dépend des utilisations que l'on veut faire des données. Si le chiffre absolu est important, le biais peut enlever tout sens aux conclusions "(Morgenstern 1972, P.51)

Il existe plusieurs autres sources d'autocorrélation des erreurs qui pourraient se rencontrer dans les données économiques.

L'orsqu'une méthode de compilation de données dans une période est essentiellement liée à une période précédente, il est fort probable que l'autocorrélation apparaisse. " Des pratiques considérant par exemple à utiliser le même facteur (comme les pondérations obtenues à partir d'un recensement), pour, période après périodes, <<souffler>> les résultats d'un échantillons en chiffres sur la population tout entière, risquent fréquemment de conduire à l'autocorrélation. Lorsque des parties d'un chiffre sont basées sur des méthodes d'estimation plutôt grossières, comme cela arrive pour les statistique du commerce extérieure et les chiffres du revenu national "(Morgenstern 1972, P.52). Alors les erreurs provoquées par les estimations dans les périodes suivantes ne peuvent pas êtres indépendantes. Si on utilise des échantillons qui ont une dépendance entre période, il est fort probable qu'il y ait apparition d'autocorrélation.

Enfin, comme toute source d'erreur, l'autocorrélation peut généralement mener à vicier l'utilité des données économiques, a moins que ces données ne soient utilisées que d'une façon approximative. " Il est peu probable que les erreurs soient liées de façon si pratique et si stable qu'elles finissent par ne plus avoir aucune importance "(Morgenstern 1972, P.55). Et si on ignore tout du genre et du type d'erreurs liant les statistiques, il ne faut pas supposer que ces erreurs sont si bien liées que leurs influences s'annulent.

I.f) MAUVAISE CLASSIFICATION:

Il arrive souvent qu'on soit confronté à un problème de classification ou de définition d'un phénomène qui doit être mesuré ou enregistré. Ce genre de problème est bien connu et plusieurs efforts ont été investis en ce sens, en vue de l'établissement d'une classification uniforme. Mais il existe encore de vastes domaines où très peu de recherche ont été faites et où il y a quasi impossibilité d'avancement. Par exemple, on peut introduire la possibilité d'un double comptage ou bien les difficultés de classification dues au manque ou à l'impossibilité de précision des définitions. On voit l'apparition de ce genre d'erreurs lorsqu'on considère de grandes sociétés. La compagnie G.M.C.(General Motor Company) est un ensemble de sociétés faisant des activités dans de nombreuses industries. Cette compagnie produit des automobiles, des moteurs d'avion, des locomotives diesel, des appareils électriques, de l'équipement de chauffage, etc. Ces différentes activités placent la compagnie dans des industries distinctes. Les questions qu'on peut se poser maintenant sont: comment la compagnie G.M.C. doit-elle être classée? Comme producteur automobile seulement? Est-ce que certaines parties doivent être enregistrées dans des industries différentes? Maintenant, comment ceci peut-il être réalisé? D'après les capitaux investis dans les différentes filiales, d'après le volume des ventes ou d'après le nombre d'ouvriers? Comment peut-on déterminer leurs parts respectives dans les profits totaux qui ne sont donnés que sous forme globale? On peut raisonner de la sorte pour la plupart des grandes compagnies du monde. Le problème augmente lorsque la diversification (dans les entreprises) grandit. Le problème augmente aussi lorsque le progrès technologique rapide provoque des liens inattendus entre des industries qui sont technologiquement séparées. On peut retrouver aujourd'hui un fabricant de pneus qui peut être un des principaux producteurs de carburant de fusée, ou bien un fabricant d'avion qui produit aussi des réacteurs nucléaires, etc. On voit par ces exemples qu'il devient difficile de classer correctement ces firmes, à moins que l'on considère ces entreprises comme des sociétés de "holding", ce qui ne serait pas exact.

Les erreurs de classifications apparaissent aussi, dans les statistiques de commerce extérieur. Les progrès technologiques réalisés régulièrement, ont une grande importance dans le problème de classification des données servant à former les statistiques du commerce international. Les pays qui éprouvent un retard technologique et qui importent des produits technologiquement avancés, ont fréquemment de grandes difficultés à les identifier et ne savent pas toujours comment les insérer adéquatement dans leurs statisques.

Enfin, les problèmes de classification sont une des nombreuses causes de l'inéxactitude des statistiques. Le biais introduit par ce genre d'erreur n'est toutefois pas constant. Cela peut être dû à l'influence de la politique. Considèrons: " les <<guerres de porcs>> entre l'Autriche-Hongrie et la Serbie où les porcs importés par l'Autriche-Hongrie étaient libellés différemment pour prevenir ou gêner l'importation, etc., falsifiant également par là toutes les statistiques intéressantes de la période "(Morgenstern 1972, P.35).

I. g) VARIABLES NON OBSERVABLES

En économique, il est fréquent que les variables que l'on désire utiliser pour expliquer un phénomène, ne soient pas observables ou bien qu'elles soient observées de façon erronée. Considérons les modèles économiques de recherche scientifique incluant le cas des variables théoriques ou abstraites pour lesquelles les instruments de mesure sont reconnus comme étant imparfaits et pour lesquels l'échelle de mesure est inexistante. L'utilité, l'habileté, l'ambition, l'attitude politique, le concept de capital humain, sont des variables difficilement mesurables. Pratiquement personne n'a jamais mesuré l'habilité; mais elle est utilisée pour expliquer le revenu de quelqu'un ou bien pour établir son statut social. Dans le modèle de Cagans(1956) sur les anticipations adaptatives, les quantités offertes peuvent dépendre de variables non observables comme le prix anticipé, les anticipations étant révisées en fonction de l'erreur associée au niveau précédent des anticipations. Dans certains cas, les variables que l'on mesure ne sont pas celles que l'on voudrait vraiment mesurer. Par exemple, les examens finaux que l'on fait passer aux étudiants sont utilisés pour mesurer approximativement les progrès réalisés au cours d'une année (semestre) d'éducation et conséquemment la variable qui en resulte est approximative et sujette à des erreurs de mesure.

Enfin, même si les variables sont observables, les données sont sujettes à une variété très grande d'erreurs. Comme nous l'avons vu précédemment, ces erreurs arrivent parce que l'analyste et l'observateur sont séparés, parce qu'il y a un manque de précision sur ce que l'on voudrait observer et que les phénomènes que l'on essaie de mesurer sont complexes. (Judge, G.G (1985))

PARTIE II : CONSÉQUENCES DES ERREURS SUR LES VARIABLES

Mentionnons deux points importants avant d'entamer l'exposé des conséquences des erreurs de mesures sur les variables explicatives.

1) Les exemples d'erreurs de mesure donné à la Partie I du CHAPITRE 1, introduisent un terme d'erreur qui, dans plusieurs cas, peut être représenté par un terme constant plus un terme aléatoire. L'espérance mathématique de ce terme d'erreur n'est donc pas nulle. Si on utilise un modèle de regression, le terme constant de l'erreur dans la variable explicative causera un biais systématique qui n'affectera que l'estimation de la constante de la régression. Comme cette estimation est souvent de moindre intérêt, on ne s'intéressera ici qu'à la partie aléatoire de l'erreur, qu'on supposera de moyenne nulle. L'exemple suivant illustre le problème, soit:

$$Y = \alpha + \tilde{X}\beta + e \qquad \text{où}$$

$X = \tilde{X} + w$ et $w = E(w) + w - E(w)$, $E(w) = c$, $v = w - E(w)$ et

$w = c + v$, où c est la partie constante de l'erreur (w) et v la partie aléatoire. En faisant les subtitutions appropriées, on obtient:

$Y = \alpha + X\beta + e - w\beta$

$= \alpha + X\beta + e - c\beta - v\beta$

$= \alpha - c\beta + X\beta + e - v\beta$. L'estimation de la constante sera biaisée et $c\beta$ est la partie qui biaise cette estimation (ce que l'on obtient est une estimation de $\alpha - c\beta$). Même si l'estimation de la constante est biaisée, nous n'en tiendrons pas compte. Nous nous intéressons seulement à l'estimation du paramètre β. On supposera donc que $X = \tilde{X} + v$ et $E(v) = 0$.

2) Les problèmes de corrélations entre les erreurs d'observation ne seront pas considérés dans l'exposé qui suit. Dans cette analyse, on se contentera de traiter du problème plus simple des erreurs de mesure qui ne sont pas autocorrélées.

PREMIÈRE CONSÉQUENCE:

Si il y a présence d'erreurs de mesure aléatoires sur les variables dépendantes, cette erreur se combinera à l'erreur de la régression sans créer de biais d'estimation. Mais dans le cas où les variables explicatives sont affectées d'erreurs de mesure, alors les conséquences sont beaucoup plus graves.

Pour illustrer le problème, considérons un modèle de régression linéaire univarié sans terme constant (NOTE: pour le cas multivarié, voir Appendice A):

(1) $$Y = \tilde{X}\beta + e$$

où Y est un vecteur de variables dépendantes de dimension $N\times 1$ et $\tilde{X}$ est un vecteur de variables explicatives de dimension $N\times 1$ avec $E(\tilde{X}_i^2) = \sigma_{\tilde{X}}^2$,

e est un vecteur d'erreurs de dimension $N\times 1$ avec $E(e) = 0$ et $E(ee') = \sigma_e^2 I, E(e_i^2) = \sigma_e^2$ enfin β est un paramètre a estimer. On suppose que les vraies variables $\tilde{X}$ ne sont pas observables. On observe:

(2) $$X = \tilde{X} + v$$

où v est un vecteur d'erreurs de mesure avec $E(v) = 0$ et $E(v_i^2) = \sigma_v^2$. On suppose aussi que: $E(ve')=E(\tilde{X}v')=E(\tilde{X}e')=0$. Alors en remplacant $\tilde{X}$ par $X-v$ dans l'equation (1), on obtient:

$$Y = \beta X - \beta v + e$$

$$= \beta X + e^*$$

où $e^* = e - \beta v$. Si on applique les M.C.O. sur la dernière équation, on a:

$$\hat{\beta} = \beta + (X'X)^{-1}X'e^*$$

et l'estimateur résultant sera non-convergent; puique:

$$plim(\hat{\beta}) = \beta + plim(X'X/N)^{-1}plim(X'e^*/N)$$

En utilisant les théorèmes sur le calcul du $plim$ présenté en Appendice B, on a, sous les hypothèses appropriées:

$plim(X'X/N) = E_{N\to\infty}(X'X/N)$ et $plim(X'e^*/N)=E_{N\to\infty}(X'e^*/N)$. Maintenant, si on développe ces deux expressions:

$$E_{N\to\infty}(X'X/N) = E_{N\to\infty}[\Sigma_{i=1}^{N}(\tilde{X}_i + v_i)(\tilde{X}_i + v_i)]/N = E_{N\to\infty}[\Sigma_{i=1}^{N}(\tilde{X}_i^2 + 2\tilde{X}_i v_i + v_i^2)]/N$$

$$= \sigma_{\tilde{X}}^2 + \sigma_v^2$$

$$E_{N\to\infty}(X'e^*/N) = E_{N\to\infty}[\Sigma_i^N(\tilde{X}_i + v_i)(e_i - \beta v_i)/N] = -\beta\sigma_v^2$$

donc le $plim(\beta) = \beta - \beta\sigma_v^2/(\sigma_{\tilde{X}}^2 + \sigma_v^2) = \beta/(1 + \lambda)$, où $\lambda = \sigma_v^2/\sigma_{\tilde{X}}^2$. On peut conclure que les M.C.O. sont asymptotiquements biaisés et non-convergents et que le biais dépend de λ (le biais dépend de l'importance relative de la variance des erreurs de mesure sur les variables explicatives). On définit λ comme étant le ratio de la variance des erreurs de mesure par rapport à la variance des variables explicatives non observées ($\tilde{X}$). λ peut être vu aussi comme l'importance relative des erreurs de mesure dans les variables explicatives.

DEUXIÈME CONSÉQUENCE:

Quels sont les effets des erreurs de mesure dans les variables explicatives sur les estimateurs du maximum de vraisemblance. Pour répondre à cela, reconsidérons le modèle de régression linéaire univariée présenté plus haut et supposons que e et v sont normalement et indépendamment distribués soit: $E(e)=E(v)=0, E(ev')=0, E(ee')=\sigma_e^2 I, E(e_i^2)=\sigma_v^2, E(vv')=\sigma_v^2 I, E(v_i^2)=\sigma_v^2$.

Le paramètre β ne peut pas être estimé de façon convergente par le maximun de vraisemblance, dû au fait qu'il n'y a pas assez d'information pour identifier le paramètre. Le problème de manque d'information peut être détecté lorsqu'on applique la méthode du maximum de vraisemblance. La fonction de densité conjointe de e et v est:

$$f(e,v)=(2\pi\sigma_e^2)^{-N/2}\exp[-(e'e)/2\sigma_e^2]\times(2\pi\sigma_v^2)^{-N/2}\exp[-(v'v)/2\sigma_v^2]$$

et par le Jacobien on trouve la fonction de vraisemblance de Y et X soit:

$$\mathcal{L}(Y,X;\beta,\sigma_e^2,\sigma_v^2,\tilde{X})=$$

$$(2\pi\sigma_e^2)^{-N/2}(2\pi\sigma_v^2)^{-N/2}\exp[-((Y-X\beta)'(Y-X\beta))/2\sigma_e^2]\exp[-((X-\tilde{X})'(X-\tilde{X}))/2\sigma_v^2]$$

Le logarithme de cette fonction est égal à:

$$\ln\mathcal{L}(Y,X;\beta,\sigma_e^2,\sigma_v^2,\tilde{X})= -N(\ln 2\pi)\ -N/2(\ln\sigma_e^2+\ln\sigma_v^2)-(X-\tilde{X})'(X-\tilde{X})/2\sigma_v^2-\ \ldots$$

$$\ldots\ -(Y-\tilde{X}\beta)'(Y-\tilde{X}\beta)/2\sigma_{\epsilon}^{2}$$

Si on dérive $\mathfrak{L}$ par rapport aux paramètres que l'on cherche, on obtient:

$$\partial\mathfrak{L}/\partial\beta=0 \rightarrow \tilde{X}^{*\prime}(Y-\tilde{X}^{*}\beta^{*})=0$$

$$\partial\mathfrak{L}/\partial\sigma_{v}^{2}=0 \rightarrow (X-\tilde{X}^{*})'(X-\tilde{X}^{*})=N\sigma_{v}^{2}$$

$$\partial\mathfrak{L}/\partial\sigma_{\epsilon}^{2}=0 \rightarrow (Y-\tilde{X}^{*}\beta^{*})'(Y-\tilde{X}^{*})=N\sigma_{\epsilon}^{2*}$$

$$\partial\mathfrak{L}/\partial\tilde{x}=0 \rightarrow (X-\tilde{X}^{*})\sigma_{\epsilon}^{2*}+\beta^{*}(Y-\tilde{X}^{*}\beta^{*})\sigma_{v}^{2*}=0$$

et si on résoud le système d'équations on obtient que:

$$\sigma_{v}^{2*} = \beta^{2*}\sigma_{\epsilon}^{2*},$$

ce qui n'est pas un résultat satisfaisant (voir APPENDICE A). L'incapacité de la méthode du maximum de vraisemblance de donner une seule estimation est le résultat d'un manque d'information. Le système est sous-identifié. (ref:Judge)

CHAPITRE 2

PRÉSENTATION DES ESTIMATEURS:

Jusqu'ici nous n'avons pas discuté de méthode permettant d'obtenir des estimateurs convergents lorsqu'il y a présence d'erreurs de mesure sur les variables explicatives. Une méthode préconisée dans la plupart des livres d'économétrie est celle des variables instrumentales. Cette méthode consiste à trouver une matrice d'instrument pour éliminer le problème de corrélation entre la matrice des variables explicatives et les erreurs de régression. Le problème est, qu'il est très difficile de trouver une telle matrice.

Prenons un exemple pour illustrer le problème.

Si on reprend le modèle de régression linéaire présenté au chapitre précédent avec les mêmes hypothèses, soit:

$$Y = X\beta + e$$

et qu'on suppose en premier lieu qu'il n'y a pas d'erreur sur les variables explicatives, alors en appliquant les M.C.O sur cette équation on trouve:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

et si on substitue Y par sa valeur

pour montrer que $\hat{\beta}$ est convergent on calcule le $plim\hat{\beta}$:

$$\hat{\beta} = \beta + (X'X)^{-1}X'e$$

$$plim\hat{\beta} = \beta + (plim(X'X/N))^{-1}\ plim(X'e/N)$$

$$= \beta + Q^{-1} \cdot 0$$

où $plim(X'X/N) = \lim_{N\to\infty} (X'X/N) = Q$, est une matrice non-singulière (une hypothèse bien connues du modèle classique linéaire général) et $plim(X'e/N) = \lim_{N\to\infty} E(X'e/N)$ où $E(X'e) = 0$ car X est non-stochastique et que $E(e) = 0$ (une autre hypothèse du modèle classique). Le fait que $E(X'e) = 0$ nous indique que les éléments de X et ceux de e sont non-corrélés. Mais dans le cas où $plim(X'e/N) \neq 0$ (corrélation entre X et e) on voit automatiquement que les M.C.O. deviennent non-convergents. La méthode des variables instrumentales consiste donc à trouver une matrice, par exemple W, tel que $plim(W'e/N) = 0$. L'estimateur des M.C.O devient, si on remplace la matrice X par W:

$$\hat{\beta} = (W'X)^{-1}W'Y$$

où $plim\hat{\beta} = \beta$. La matrice W est formée de variables fortement correlées avec celles contenues dans X mais pas avec e. En pratique, il est très difficile de trouver une matrice, comme la matrice W, qui possède toutes les caractéristiques voulues.

Une alternative à l'approche des variables instrumentales mais qui n'a reçu que très peu d'attention dans la littérature récente est discutée en particulier par PAL (1980). PAL présente pour un modèle de regression qui ne contient qu'une seule variable explicative six estimateurs basés sur les moments du troisième ordre des observations. Ces estimateurs ont, sous des hypothèses très raisonnables, la propriété d'êtres convergents même s'il y a presence d'erreurs de mesure sur les variables explicatives. Il mentionne qu'un de ces estimateurs a été suggèré par DURBIN (1954), un autre par DRION (1951) et un troisième par GEARY (1942). Un des estimateurs que nous allons considèrer sera celui de DURBIN et dans le reste de l'exposé, on l'appelera β_d. Cet estimateur a la propriété d'être aussi sans biais lorsqu'il n'y a pas d'erreurs sur les variables explicatives, ce qui n'est pas le cas pour les autres estimateurs mentionné par PAL. PAL fait allusion au fait que l'estimateur de DURBIN peut être facilement généralisé au cas où le modèle contiendrait plus d'un régresseur.

Il existe aussi d'autres estimateurs convergents (même s'il y a presence d'erreurs de mesure sur les variables explicatives) basés sur des moments échantillonaux d'ordre plus élevés. Pour le cas où un modèle de régression aurait seulement une variable explicative, des estimateurs basés sur les quatrièmes moments ont été proposés par GEARY (1942) et PAL (1980). Un de ces estimateurs sera aussi considéré, car il demeure sans biais dans le cas où il n'y a pas d'erreurs dans les variables indépendantes. Cet estimateur, de même que β_d, sera généralisé au cas où il y a plus d'une variable explicative et s'appelera β_p pour le reste du texte.

β_d et β_p dans leur forme univariée(forme initiale), prenaient la forme suivante:

$$\beta_d = \frac{\Sigma x_i^2 y_i}{\Sigma x_i^3}$$

$$\beta_p = \frac{\Sigma x_i^3 y_i - 3(\Sigma x_i^2/N)\Sigma x_i y_i}{\Sigma x_i^4 - 3(\Sigma x_i^2/N)\Sigma x_i^2}$$

On peut facilement démontrer que, si les erreurs de mesure sont indépendantes et suivent une loi normale, $plim \beta_d = \beta$, $plim \beta_p = \beta$. Le modèle de régression univarié sous-jacent s'écrit:

$$Y = \alpha + \tilde{X}\beta + u$$

$X=\tilde{X}+v$, $v \sim N(0,\sigma_v^2 I)$ $E(u)=0$, $V(u)=\sigma_u^2 I$, $V(u_i)=\sigma_u^2$ et les variables x et y, telles qu'exprimées plus haut sont en déviation de la moyenne. La particularité de β_p est le terme de correction $3\,(\Sigma x_i^2/N)\,\Sigma x_i y_i$ du numérateur et celui du dénominateur $3\,(\Sigma x_i^2/N)\,\Sigma x_i^2$, introduits pour tenir compte du fait que les erreurs v sont normalement distribuées.

Sous l'hypothèse H0 (absence d'erreur de mesure sur les variables explicatives), β_d et β_p sont sans biais. Mais le problème que pose ces estimateurs, et ceci est reconnu depuis longtemps, est qu'ils sont beaucoup plus erratiques que les estimateurs des M.C.O correspondants [voir: Kendall et Stuart(1963), Malinvauld(1978) et la partie II du chapitre II, tableau 1]. Ceci explique probablement le fait que ces estimateurs ne soient pas utilisés dans les applications actuelles.

Le but de ce chapitre est de présenter les nouveaux estimateurs (Dagenais, 1991) basés sur les moments échantillonnaux d'ordre supérieur à deux, qu'on appelera β_H et β_E (β_E est une extension de β_H). β_H est une combinaison linéaire matricielle pondérée de la version généralisée de β_p et β_d (Dagenais 1991). Cet estimateur, en plus d'être sans biais sous H0 et convergent sous l'alternative H1 (cela sera montré dans l'appendice B), a la qualité d'être beaucoup moins erratique que β_d où β_p pris séparément (cela sera montré au chapitre III).

Comme il a été mentionné par PAL(1980), tous ces estimateurs peuvent être considérés comme des types spéciaux d'estimateurs à variables instrumentales, où les instruments contiennent les variables explicatives élevées à des puissances particulières.

Le test de l'hypothèse H0 peut être fait avec l'estimateur β_H. Mais il est à noter que le même test peut être fait de deux autres façons. Une de ces façons est analogue au test proposé par HAUSMAN(1978). L'autre façons est basée sur la propriété d'indépendance entre les erreurs de la régression et les régresseurs, sous H0. Le test en question est un test exact dans les petits échantillons car sous H0 l'estimateur β_H est sans biais (la distribution exacte de la statistique est dérivable). Il apparaît d'après les expériences effectuées, que le test proposé peut être utile pour détecter la présence d'erreurs dans les variables explicatives dans les grands échantillons, même si ces erreurs sont petites et le coefficient corrélation multiple est faible ($\sqrt{R^2}$ faible): une situation typique en microéconomique (voir CHAPITRE 3). Le test peut aussi avoir de la puissance dans les petits échantillons quand le coefficient corrélation multiple est fort ($\sqrt{R^2}$ fort) et les erreurs de mesure sont importantes: une situation typique en macroéconomique (voir CHAPITRE 3).

Dans le cas où l'on suspecte fortement la présence d'erreurs de mesure sur les variables explicatives, il est possible d'utiliser la version étendue de β_H soit β_E, qui est aussi convergente et asymptotiquement plus efficace, quand H1 est acceptée.

Après cette brève introduction des nouveaux estimateurs, on peut maintenant passer à leur présentation.

L'estimateur β_H:

On suppose qu'on a le modèle de régression suivant:

$$Y = \alpha i + \tilde{X}\beta + u$$

où $\tilde{X}$ est une matrice de dimension $N \times K$ de variables non-aléatoires. Cette matrice est la matrice des variables explicatives mesurées sans erreur et on suppose que $\lim_{N\to\infty} \frac{\tilde{X}'\tilde{X}}{N} = Q$ où Q est finie et non-singulière. Le vecteur u est un vecteur d'erreurs résiduelles normalement distribuées, de dimension $N \times 1$ avec $E(u) = 0$, $E(uu') = \sigma_u^2 I_N$. Y est un vecteur d'observations sur la variable dépendante de dimension $N \times 1$; enfin β est un vecteur de paramètres à estimer de dimension $K \times 1$, α est la constante à estimer et i est un vecteur unité de dimension $N \times 1$. Explicitement:

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ . \\ Y_N \end{pmatrix}, \quad \beta=\begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ . \\ \beta_K \end{pmatrix}, \quad i=\begin{pmatrix} 1 \\ 1 \\ . \\ . \\ . \\ . \\ 1_N \end{pmatrix}$$

et

$$\tilde{X}=\begin{pmatrix} \tilde{X}_{11} & . & . & . & . & \tilde{X}_{1K} \\ \tilde{X}_{21} & \tilde{X}_{22} & . & . & . & \tilde{X}_{2K} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ \tilde{X}_{N1} & . & . & . & . & \tilde{X}_{NK} \end{pmatrix}$$

$$u=\begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ . \\ u_N \end{pmatrix}, \qquad var(u)=E(uu')=\sigma_u^2\begin{pmatrix} 1 & 0 & 0 & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . & 0 \\ . \\ . \\ . \\ . \\ 0 & 0 & 0 & . & . & . & . & 1_N \end{pmatrix}.$$

$var(u)=\sigma_u^2 I_N$ implique que les erreurs de la régression ne sont pas corrélées entre elles et que la variance est homoscédastique.

On suppose que $\tilde{X}$ est non-observable et que la matrice X est observée à la place, soit:

$$X = \tilde{X} + v$$

où v est une matrice d'erreurs de mesure (normalement distribuées) sur les variables explicatives, de dimension $N\times K$. La forme matricielle de v est la suivante:

$$v=\begin{pmatrix} v_{11} & . & . & . & . & v_{1K} \\ v_{21} & v_{22} & . & . & . & v_{2K} \\ . \\ . \\ . \\ . \\ v_{N1} & . & . & . & . & v_{NK} \end{pmatrix}.$$

On suppose aussi que $var[vect(v)]=\Sigma\otimes I_N$ où $var[.]$ veut dire: matrice de covariances et Σ est une matrice symétrique, positive semi-definie. Soit:

$$var[vect(v)]=\begin{pmatrix} \sigma_{11}I_N & . & . & . & . & \sigma_{1K}I_N \\ \sigma_{21}I_N & \sigma_{22}I_N & . & . & . & \sigma_{2K}I_N \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ \sigma_{KK}I_N & . & . & . & . & \sigma_{KK}I_N \end{pmatrix} \quad \text{et} \quad \Sigma=\begin{pmatrix} \sigma_{11} & . & . & . & . & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2K} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ \sigma_{1K} & . & . & . & . & \sigma_{KK} \end{pmatrix}$$

Comme on peut le constater ci-dessus, les hypothèses faites sur les erreurs de mesure des variables explicatives impliquent que les erreurs sont indépendantes entre observations, soit: $cov(v_{ij}, v_{i+k,j'})=0,\ i=1,...,N,\ j,\ j'=1,...,K,\ k\neq 0$, mais qu'elles sont dépendantes entre variables pour une même observation, soit: $cov(v_i, v_j)\neq 0,\ i\neq j$. Par exemple: la matrice

de covariance des erreurs de mesure de la variable X_1 est: $\begin{pmatrix} \sigma_{11} & . & . & . & . & . & 0 \\ 0 & \sigma_{11} & 0 & . & . & . & 0 \\ . & & & & & & \\ . & & & & & & \\ . & & & & & & \\ 0 & 0 & 0 & . & . & . & \sigma_{11} \end{pmatrix}$ et cette matrice est de dimension $N \times N$. Cette hypothèse implique aussi que pour une même variable, les variances des erreurs de mesure sont homoscédastiques.

β_p n'a pas été généralisé au cas où il y aurait plus d'une variable explicative. (La généralisation de β_d est donnée dans PAL). Alors voici la présentation multivariée de ces estimateurs (Dagenais M.G. 1992) qui servent à la construction de β_H et β_E, sous forme d'estimateurs à variables instrumentales:

$\beta_D = (z_1'x)^{-1}z_1y$ où $z_1 = \begin{pmatrix} x_{11}^2 & . & . & . & . & x_{1K}^2 \\ x_{21}^2 & x_{22}^2 & . & . & . & x_{2K}^2 \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ x_{N1}^2 & x_{N2}^2 & x_{N3}^2 & . & . & x_{NK}^2 \end{pmatrix}$. Les x_{ij} sont les éléments de la matrice x

et $x = AX$ où $A = I_N - ii'/N$. La matrice x correspond à la matrice X calculée en déviation de la moyenne. Même chose pour y où $y = AY$.

$\beta_P = (z_2'x)^{-1}z_2'y$ où $z_2' = z_3' - 3D(x'x/N)x'$ et $z_3 = \begin{pmatrix} x_{11}^3 & . & . & . & . & x_{1K}^3 \\ x_{21}^3 & x_{22}^3 & . & . & . & x_{2K}^3 \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ x_{N1}^3 & x_{N2}^3 & x_{N3}^3 & . & . & x_{NK}^3 \end{pmatrix}$. Encore une fois les éléments de x et y sont pris en déviation de la moyenne, $D(x'x/N)$ est une matrice diagonale de dimension $K \times K$ avec sur la diagonale les éléments correspondants de la matrice $x'x/N$. Pour calculer $D(x'x/N)$, on prend chaque élément de $x'x/N$ et on le multiplie par l'élément correspondant de la matrice identité. Soit: $x = (x_1 x_2 \ldots x_k)$ où les éléments de x sont des vecteurs colonnes et

$$x'x/N = 1/N \begin{pmatrix} x_1' \\ x_2' \\ \cdot \\ \cdot \\ \cdot \\ x_K' \end{pmatrix} \cdot (x_1 \;\; x_2 \;\; \ldots \;\; x_K) = 1/N \begin{pmatrix} x_1'x_1 & \cdot & \cdot & \cdot & \cdot & x_1'x_K \\ x_2'x_1 & x_2'x_2 & & \cdot & \cdot & \cdot & x_2'x_k \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ x_K'x_1 & x_2'x_K & x_3'x_K & \cdot & \cdot & x_K'x_K \end{pmatrix}$$

et donc

$$D(x'x/N) = \begin{pmatrix} x_1'x_1/N & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & x_2'x_2/N & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & x_K'x_K/N \end{pmatrix}$$

Connaissant maintenant β_D et β_P on peut passer à β_H:

$\beta_H = W \begin{pmatrix} \beta_D \\ \beta_P \end{pmatrix}$ où $W = (C'S^{-1}C)^{-1}C'S^{-1}$, $C = \begin{pmatrix} I_K \\ I_K \end{pmatrix}$ et S est la matrice de covariances de $\begin{pmatrix} \beta_D \\ \beta_P \end{pmatrix}$

sous H0. Montrons plus précisément d'où vient β_H: β_H est obtenu en applicant les M.C.G. sur une combinaison de β_D et β_P; β_D est sans biais sous H0 et a comme matrice de covariance S_D, soit:

$$\beta_D = (z_1'x)^{-1}z_1y$$

et, $y = x\beta + Au$ où $y=AY$ et $x=AX$. Si on substitue dans β_D, y par sa valeur on obtient :

$$\beta_D = (z_1'x)^{-1}z_1'x\beta + (z_1'x)^{-1}z_1'Au$$

alors pour montrer que β_D est sans biais sous H0 on calcule:

$E(\beta_D) = \beta + (z_1'x)^{-1}z_1'AE(u)$ et $E(u)=0$ donc $E(\beta_D) = \beta$, la matrice de covariances de β_D est:

$S_D = E[(\beta_D-\beta)(\beta_D-\beta)']$

$=E[((\beta+(z_1'x)^{-1}z_1'Au)-\beta)((\beta+(z_1'x)^{-1}z_1'Au)-\beta)']$

$=E[((z_1'x)^{-1}z_1'Au)((z_1'x)^{-1}z_1'Au)']$

$=E[(z_1x)^{-1}z_1'Auu'A'z_1(x'z_1)^{-1}]$

$=(z_1'x)^{-1}z_1'Az_1(x'z_1)^{-1}\sigma_u^2$ car $AA'=A$, $E(uu')=\sigma_u^2 I_N$. On procède de la même façon pour β_P:

$$\beta_P = (z_2'x)^{-1}z_2'y \quad , y=AY \ , x=AX$$

en substituant y dans β_P on trouve: $\beta_P = \beta + (z_2'x)^{-1} z_2'Au$ et donc β_P est sans biais sous H0 car $E(\beta_P) = \beta$. La matrice de covariance de β_P est:

$S_P = E[(\beta_P - \beta)(\beta_P - \beta)']$

$= (z_2'x)^{-1} z_2'Az_2(x'z_2)^{-1}\sigma_u^2$. Maintenant pour obtenir β_H dérivé de β_P et β_D, il faut appliquer les M.C.G. sur la forme suivante:

$\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix} = C\beta + \begin{pmatrix}u_D\\ u_P\end{pmatrix}$ où $C = \begin{pmatrix}I_K\\ I_K\end{pmatrix}$, $u_D = (z_1'x)^{-1} z_1'Au$ et $u_P = (z_2'x)^{-1} z_2'Au$ ce qui donne:

$$\beta_H = (C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$$

ou plus précisément, si on applique les M.C.O sur la forme suivante:

$P\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix} = PC\beta + P\begin{pmatrix}u_D\\ u_P\end{pmatrix}$ où P est une matrice non-singulière de dimension $2K \times 2K$ telle que

$P'P = S^{-1}$, alors

$(PC)'P\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix} = (PC)'PC\ \beta_H$

$\beta_H = (C'P'PC)^{-1}(C'P'P)\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$ et donc $\beta_H = (C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$ où S est la matrice d-covariance de $\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$ et est donnée par: $\begin{pmatrix}S_D & S_{DP}\\ S'_{DP} & S_P\end{pmatrix}$ où

$S_{DP}=(z_1'x)^{-1}z_1'Az_2(x'z_2)^{-1}\sigma_u^2$. La matrice de covariance de β_H est donnée par: $V(\beta_H) = (C'S^{-1}C')^{-1}$ (σ_u^2 est inclus dans S^{-1}). β_H sera sans biais car β_D et β_P sont sans biais sous H0 soit:

$$E(\beta_H) = (C'S^{-1}C)^{-1}C'S^{-1}\ E\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$$

$$= (C'S^{-1}C)^{-1}C'S^{-1}C\beta \ \text{ car } E\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix} = c\beta$$

$$= \beta$$

L'application des M.C.G. sur la forme décrite plus haut donne un estimateur qui est une combinaison linéaire matricielle optimale de β_D et β_P. Il est possible de montrer que l'estimateur résultant d'un arrangement de la sorte aura une variance plus petite ou égale à la plus petite des variances des deux estimateurs β_D ou β_P (Theil, H. et A.S. Goldbeger (1961)). On peut montrer que β_H demeure convergent (convergence en probabilité) en présence d'erreurs de mesure sur les variables explicatives (voir APPENDICE B).

Il existe une autre façon (plus simple) pour déduire β_H, ceci en utilisant

la méthode des régressions artificielles (MacKinnon 1992). Il s'agit, dans un premier temps, de former K régressions artificielles en prenant X comme variable dépendante et (z_1, z_2) comme variables indépendantes soit:

$$X = (i, z_1, z_2)\Gamma + h$$

où $\Gamma = \begin{pmatrix} \gamma_A' \\ \gamma_B' \\ \gamma_C' \end{pmatrix}$, explicitement: $\Gamma = \begin{pmatrix} \gamma_{A_{11}} & \cdot & \cdot & \cdot & \gamma_{A_{1K}} \\ \gamma_{B_{11}} & \gamma_{B_{12}} & \cdot & \cdot & \gamma_{B_{1K}} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \gamma_{B_{K1}} & \gamma_{B_{K2}} & \cdot & \cdot & \gamma_{B_{KK}} \\ \gamma_{C_{11}} & \gamma_{C_{12}} & \cdot & \cdot & \gamma_{C_{1K}} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \gamma_{C_{K1}} & \gamma_{C_{K2}} & \cdot & \cdot & \gamma_{C_{KK}} \end{pmatrix}$ et

$$h=\begin{pmatrix} h_{11} & . & . & . & . & h_{1K} \\ h_{21} & h_{22} & . & . & . & h_{2K} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ h_{N1} & h_{N2} & . & . & . & h_{NK} \end{pmatrix}$$ est une matrice d'erreurs de dimension $N \times K$ avec $E(h)=0$. On

applique les M.C.O. sur $X=(i, z_1, z_2)\Gamma + h$ et on trouve ainsi $\hat{\Gamma}$ et on calcule $\hat{X}=(i, z_1, z_2)\hat{\Gamma}$. Pour terminer, on transforme $\hat{X}$ en $\hat{x}$ (en déviation de la moyenne) et on utilise $\hat{x}$ comme matrice d'instrument soit:

$$\beta_H = (\hat{x}'x)^{-1}\hat{x}'y \ .$$

On peut remplacer x par $\hat{x}$, soit:

$$\beta_H = (\hat{x}'\hat{x})^{-1}\hat{x}'y \ ;$$

On peut faire cela parce que $x=(z_1, z_2)\hat{\Gamma}+ \hat{h}$ et $\hat{x}'\hat{h}=0$. Enfin, la matrice de covariance de β_H sous H0 est: $V(\beta_H) = (\hat{x}'\hat{x})^{-1}\sigma_u^2$ où σ_u^2 est estimé par $\hat{\sigma}_u^2=(u_H'u_H/N-K-1)$ et $u_H=y-x\beta_H$.

Connaissant maintenant β_H on peut passer à la présentation de β_E comme il a été mentionné plus haut dans le chapitre.

L'estimateur β_E:

L'estimateur β_H dérivé d'une combinaison linéaire matricielle optimale de β_D et β_P sous H0, demeure convergent lorsqu'il y a des erreurs sur les variables explicatives. Mais il est possible de déduire un estimateur alternatif (β_E) qui serait une combinaison linéaire matricielle asymptotiquement optimale de β_D et β_P sous l'hypothèse H1 (présence d'erreurs de mesure sur les variables explicatives). On peut trouver cet estimateur en substituant dans β_H la matrice $\hat{S}^*$ à la matrice S. Soit:

$$\beta_E = W^* \begin{pmatrix} \beta_D \\ \beta_P \end{pmatrix}$$

où $W^* = (C'\hat{S}^{*-1}C)^{-1}C'\hat{S}^{*-1}$ et $\hat{S}^* = \begin{pmatrix} \hat{S}_D^* & \hat{S}_{DP}^* \\ \hat{S}_{DP}^{*\prime} & \hat{S}_P^* \end{pmatrix}$. On peut montrer maintenant comment on calcule la matrice $\hat{S}^*$.

Sous H1, $\Psi = \sqrt{N}(\beta_D - \beta) = (z_1'x/N)^{-1} z_1' A\eta/\sqrt{N}$. Ψ a la même distribution asymptotique que $\Lambda = [plim(z_1'x/N)]^{-1} z_1' A\eta/\sqrt{N}$, puisque $plim(\Lambda - \Psi) = 0$. Donc la matrice asymptotique de covariance de $\sqrt{N}(\beta_D - \beta)$ est:

$$\lim_{N\to\infty} V(\Lambda) = \lim_{N\to\infty} E(\Lambda\Lambda') = \lim_{N\to\infty} E[([plim\ z_1'x/N]^{-1} z_1'A\eta/\sqrt{N})([plim\ z_1'x/N]^{-1} z_1'A\eta/\sqrt{N})']$$

$$= ([plim\ z_1'x/N]^{-1} \lim_{N\to\infty} E\frac{z_1'A\eta\eta'Az_1}{N} [plim\ z_1'x/N]^{-1}).$$ Cette

expression peut être estimée de façon convergente (White 1982) par:

$$N\hat{S}_D^* = [z_1'x/N]^{-1} \frac{\sum_{i=1}^{N} \tilde{z}_{i1}' \eta_{Hi}^2 \tilde{z}_{i1}}{N} [x'z_1/N]^{-1}$$

où $\tilde{z}_{i1} = z_{i1} - \bar{z}_1$, z_{i1} est la ième ligne de z_1 et $\bar{z}_1$ est un vecteur ($1\times$K) de moyennes des K variables contenues dans z_1, $\eta_{Hi} = y_i - x_i\beta_H$ est le ième élément de η_H, y_i et x_i sont respectivement les éléments de la ième ligne de y et x. De la même façon on trouve que

$$N\hat{S}_P^* = [z_2'x/N]^{-1} \frac{\sum_i^N \tilde{z}_{i2}' \eta_{Hi}^2 \tilde{z}_{i2}}{N} [x'z_2/N]^{-1} \text{ et } N\hat{S}_{DP}^* = [z_1'x/N]^{-1} \frac{\sum_i^N \tilde{z}_{i1}' \eta_{Hi}^2 \tilde{z}_{i2}}{N} [x'z_2/N]^{-1}.$$

Présentation du test:

Pour terminer le chapitre II, nous allons maintenant passer à la présentation du test pour la présence d'erreurs de mesure sur les variables explicatives.

L'approche des régressions artificielles utilisée pour dériver β_H et β_E peut être utilisée pour tester l'hypothèse H0. On définit:

$$\hat{h} = X - \hat{X}$$

où $\hat{X}=(i,z_1,z_2)\hat{\Gamma}$ et $X=(i,z_1,z_2)\hat{\Gamma}+\hat{h}$. Si on suppose qu'il n'y a pas d'erreur sur les variables explicatives ($X=\tilde{X}$), on peut écrire:

$$Y= \alpha i+X\beta+u = \alpha i+(\hat{h}+\hat{X})\beta+u$$

$$=\alpha i+\hat{h}\beta+\hat{X}\beta+u$$

si on pose $\beta=\gamma \rightarrow Y = \alpha i+\hat{X}\beta+\hat{h}\gamma+u$,

on peut écrire aussi que

$$Y = \alpha i+\hat{X}\beta+\hat{h}\beta+\hat{h}\gamma-\hat{h}\beta+u$$

$$= \alpha i+\hat{X}\beta+\hat{h}\beta+\hat{h}(\gamma-\beta)+u$$

$$= \alpha i+\hat{X}\beta+\hat{h}\beta+\hat{h}(\theta)+u \quad \text{où } \theta=\gamma-\beta=0$$

$$= \alpha i+(\hat{X}+\hat{h})\beta+\hat{h}(\theta)+u$$

$$= \alpha i+X\beta+\hat{h}\theta+u.$$

Alors sous H0, on régresse Y sur X et $\hat{h}$. Le résultat attendu est que les éléments du vecteur de coefficient associés à la matrice $\hat{h}$ soient tous égaux à zéro. Le test "F" habituel, qui a (K, N-2K-1) degré de liberté (K restriction et 2K+1 paramètres dans la régression), est utilisé pour tester l'hypothèse nulle (H0 : $\theta = 0$).

Deux façons alternatives existent pour tester exactement la même chose:

1) Un test qui repose sur le fait que, sous H0 β_H et β_L sont sans biais donc $E(\beta_H-\beta_L)=0$, alors que sous H1, $plim(\beta_H-\beta_L)\neq 0$.

Le test est donné par:

$$m = (\beta_H-\beta_L)' G^{-1}(\beta_H-\beta_L)/\hat{\sigma}^2$$

où $G\hat{\sigma}^2$ est l'estimateur la matrice de covariance de $(\beta_H-\beta_L)$ sous H0. On calcule cette matrice de covariance par $V[\beta_H-\beta_L]=E[(\beta_H-\beta_L)(\beta_H-\beta_L)']=G\sigma^2$ et $G=JJ'$

où $J=(C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}(z_1'x)^{-1}z_1'A\\(z_2'x)^{-1}z_2'A\end{pmatrix}-(x'x)^{-1}x'$. On estime σ^2 par $\hat{\sigma}^2=\dfrac{e_L'e_L}{N-K-1}$ où

$e_L=y-x\beta_L$. On peut estimer aussi $\hat{\sigma}^2$ par $\dfrac{e_H'e_H}{N-K-1}$ où $e_H=y-x\beta_H$.

Enfin, on calcule exactement la "p-value" associée à m. Cette valeur est définie comme étant la probabilité que sous H0 la variable aléatoire correspondant à m ait une valeur plus grande que celle observée pour l'échantillon considéré. Pour un test effectué à un niveau de 5% de signification, H0 sera rejeté si la "p-value" de m est plus petite que 0.05 (voir APPENDICE B pour le calcule de la "p-value" associée à m).

2) La deuxième façon, serait sous H0, de faire la régression de e_H dérivé de β_H sur x soit : $e_H=x\theta+\xi$ où ξ est un vecteur de dimension $N\times 1$ d'erreurs de régression. θ est un vecteur de dimension $K\times 1$ de paramètres inconnus. Sous H0, la corrélation entre e_H et x est nulle donc tous les éléments de θ devraient être égaux à zéro. S'il y avait des erreurs dans les variables explicatives, les éléments de θ seraient significativement différents de zéro. On estime ici θ par $\hat{\theta}$ en utilisant les M.C.G., car sous H0:

$$\xi=e_H$$

$$e_H=y-x\beta_H$$

$$=x\beta+Au-x\beta+x(C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}(z_1'x)^{-1}z_1'\\(z_2'x)^{-1}z_2'\end{pmatrix}Au$$

$$=Mu \quad \text{où } M=(I-x(C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}(z_1'x)^{-1}z_1'\\(z_2'x)^{-1}z_2'\end{pmatrix})A$$

et $V(e_H)=E(e_He_H')=\sigma_u^2MM'$. L'estimation de θ par les M.C.G. est $\hat{\theta}=(x'(MM')^{-1}x)^{-1}x'(MM')^{-1}y$. Le problème maintenant est de trouver l'inverse de MM' car cette matrice est singulière. La méthode de l'inverse généralisé de Moore-Penrose résoud ce problème. L'inverse généralisé de MM' est $(MM)^+$ et on le calcule comme suit:

$(MM')^+=HD^{-1}H'(MM')'$ où D est une matrice diagonale de dimension $r\times r$ $(r=N-K-1)$ et les éléments de la diagonale sont les valeurs propres (différentes de zéro) de $(MM')'MM'$. H est de dimension $N\times r$ et est une matrice de vecteurs propres correspondants au r valeurs propres de $(MM')'MM'$ soit: $(MM')'MM'$ est de dimension $N\times N$ et de rang $N-K-1$

$$D=\begin{pmatrix} \lambda_1 & . & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ 0 & 0 & . & . & . & \lambda_{N-K-1} \end{pmatrix} \text{ et } H=\begin{pmatrix} m_{11} & . & . & . & . & m_{1r} \\ m_{21} & m_{22} & . & . & . & m_{2r} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ m_{N1} & m_{N2} & . & . & . & m_{Nr} \end{pmatrix}$$ (Theil, H. (1971)).

Enfin le test "F" habituel peut être utilisé pour tester $\theta=0$. Ce test a (K,N-2K-1) degré de liberté (K restriction et 2K+1 paramètres à estimer). Il est à noter que si $N\rightarrow\infty$ alors $Mu\rightarrow u$ donc pour les grands échantillons les M.C.O peuvent être utilisés pour estimer θ.

CHAPITRE 3

Ce chapitre se divise en deux parties:

a) Trouver des éléments d'explication (l'explication est valable pour β_E qui est une extension de β_H) de la meilleur performance de β_H sur β_D et β_P. Dans le cas où il y a présence d'erreurs de mesure dans les variables explicatives, essayer de comprendre pourquoi β_H (et β_E extension de β_H) semble, encore une fois, avoir une meilleure performance que β_L (M.C.O).

b) Expériences de Monte Carlo permettant de comparer β_H et β_E relativement à β_L.

Partie a):

ÉLÉMENTS EXPLICATION DE LA MEILLEURE PERFORMANCE DE β_H PAR RAPPORT A β_D ET β_P

a) Dans le but d'expliquer cette meilleure performance, il peut être utile de calculer la grandeur relative de l'écart type des estimateurs β_D, β_P et β_H quand il n'y a pas d'erreur sur les variables explicatives. Cela se fait en divisant les écarts types de ces estimateurs par ceux des β_L correspondants (voir Tableau 1).

Tableau 1

ÉCART TYPE DES ESTIMATEURS INDIQUÉS / ÉCART TYPE DES M.C.O.

Paramètres	β_D	β_P	β_H
β_1	10.4	4.6	3.6
β_2	9.9	18.5	2.0
β_3	4.0	5.6	2.1
β_4	25.3	6.2	2.4

On constate au Tableau 1, le gain remarquable d'utiliser β_H au lieu de β_D et β_P. Par exemple, si on prend l'estimation du coefficient β_4, les ratios de β_D et β_P sont respectivement de 25 et 6 pour 1 tandis que pour β_H le ratio est seulement de 2.4 pour 1. Cependant, il est important de noter que les résultats présentés dans ce tableau dépendent des données utilisés pour ces calculs, ceci n'est qu'une illustration. La description de ces données est présentée à la partie b) de ce chapitre.

Sachant que les trois estimateurs peuvent être interprétés comme des estimateurs à variables instrumentales obtenus en remplaçant les variables explicatives X par $\hat{X}$ (ceci revient à utiliser une approche de variables instrumentales) où $\hat{X}$ est obtenu en régressant X sur z_1 ou sur z_2 ou bien sur les deux et sachant que les variances de ces estimateurs dépendent de la corrélation de ces variables avec les variables explicatives (voir APPENDICE C), alors il peut être utile d'examiner la racine du coefficient de détermination ($\sqrt{R^2}$: coefficient de corrélation multiple de la régression) de la régression des quatres variables $\tilde{X}_i$ sur z_1 ou z_2 ou bien sur les deux (voir

Tableau 2).

Tableau 2

COEFFICIENT DE CORRÉLATION MULTIPLE DE LA RÉGRESSION DE $\tilde{X}_i$ SUR z_i

Variables dépendantes	Variables indépendantes		
	z_1	z_2	z_1 et z_2
	estimateurs correspondants		
	(β_D)	(β_P)	(β_H)
$\tilde{X}_1$	0.42	0.26	0.44
$\tilde{X}_2$	0.51	0.16	0.69
$\tilde{X}_3$	0.63	0.25	0.78
$\tilde{X}_4$	0.27	0.44	0.51

Pour mieux comprendre comment β_H peut avoir une meilleure performance sur β_D et β_P, il peut être utile d'étudier, au Tableau 2, la contribution marginale sur le $\sqrt{R^2}$ de l'ensemble de variables contenues dans z_2 une fois que l'ensemble de variables contenues dans z_1 a déjà été introduit comme instrument pour $\tilde{X}$ et aussi la contribution marginale au $\sqrt{R^2}$ de z_1 une fois que z_2 a déjà été introduit comme instrument pour $\tilde{X}$. En d'autres termes, on essaie de voir si le fait d'ajouter z_1 à z_2 ou z_2 à z_1 contribue à augmenter le $\sqrt{R^2}$ par rapport au fait d'avoir z_2 ou z_1 seul. On constate que dans le cas de $\tilde{X}_1$, $\tilde{X}_2$ et $\tilde{X}_3$ l'introduction de z_2 n'augmente pas fortement la corrélation multiple en comparaison de celle obtenue seulement avec z_1. Cependant, le cas est différent pour z_2.

Partie b):

EXPÉRIENCES DE MONTE CARLO

Différentes mesures comme: la Racine Carrée de l'Écart Quadratique Moyen, L'Erreur de Type I, le Biais, seront utilisées et donc dans un premier temps nous allons expliquer ce que sont et a quoi servent ces différentes mesures.

Enfin, nous allons présenter les différentes expériences de Monte Carlo servant à mesurer la performance relative de β_H et β_E par rapport à β_L.

<u>Racine Carrée l'Écart Quadratique Moyen:</u> mesure la dispersion de l'estimateur autour de la vraie valeur. Un bon estimateur par rapport à un moins bon, a un REQM plus faible. On le difinit comme suit:

$$REQM(\beta_L) = \sqrt{E[\beta_L-\beta]^2} = \sqrt{BIAIS(\beta_L)^2+VARIANCE(\beta_L)}$$

<u>Biais:</u> Mesure la différence entre les vraies valeurs d'un ensemble de paramètres et la moyenne des estimations de ces paramètres. Si cette différence est non nulle alors on dit que l'estimateur est biaisé. Définition: $BIAIS(\beta_L) = E(\beta_L) - \beta$

Erreur De Type I: La probabilité de rejeter l'hypothèse nulle (H0), qui suppose qu'un paramètre k est égal à une valeur donnée, alors que l'hypothèse est vraie. On la calcule de la façon suivante:

[1 - (nombre fois qu'H0 est accepté / nombre d'échantillons)] $\times$ 100. Pour avoir le nombre de fois qu'Ho est accepté, un test-t est effectué au niveau 5% pour un paramètre k sur chaque échantillon.

Les différentes expériences de Monté Carlo discutées plus bas, concernent β_H et β_B relativement à β_L. La performance relative de ces estimateurs, comme il a été expliqué, est mesurée en terme de REQM de Biais et d'Erreurs De Type I. La Puissance du Test est aussi discutée (définie comme suit: le nombre de fois qu'H1 est acceptée alors qu'H1 est vraie / nombre d'échantillons avec erreurs de mesure). Seulement les cas où les erreurs de mesure affectent une ou deux variables ont été considérés et dans ce dernier cas, on a supposé que les erreurs de mesure des deux variables n'étaient pas correlées. Il est à noter que les expériences effectuées plus bas ont été effectuées en fixant la probabilité que les échantillons comportent des erreurs de mesure, à 50%. Il est probable que si les expériences avaient été effectuées avec 100% des échantillons affectés d'erreurs, la performance des estimateurs proposés aurait augmenté relativement celle de β_L (M.C.O). Mais probablement pas la performance relative de β_B par rapport à β_H. D'autres éléments comme la matrice de corrélation des variables indépendantes ou la matrice de corrélation des erreurs de mesure peuvent affecter la performance relative de β_H et β_B.

Expériences avec les données d'une enquête sur les consommateurs

Les données utilisées pour faire les expériences sont tirées de l'enquête faite par Statistiques Canada sur les Finances des Consommateurs en 1986.

On utilise un modèle simple reliant la consommation totale annuelle aux variables suivantes:

$\tilde{X}_1$: revenu annuel du chef du ménage

$\tilde{X}_2$: âge du chef du ménage

$\tilde{X}_3$: revenu annuel des autres membres du ménage

$\tilde{X}_4$: nombre de personnes-semaines formant le ménage durant l'année

Pour garder une certaine homogénéité de l'ensemble des observations, on a gardé les observations pour lesquelles le revenu total du ménage se situait entre \$25000 et \$55000. Les paramètres utilisés pour générer la variable dépendante sont ceux montrés dans l'équation suivante:

$$Y = 1 + 1\tilde{X}_1 + 1\tilde{X}_2 + 1\tilde{X}_3 + 1\tilde{X}_4 + u$$

où u est un terme d'erreur de régression. Les variables $\tilde{X}$ observées dans l'échantillon sont supposées être les vraies valeurs des variables (Note: les paramètres de l'équation ci-dessus ont été fixés à 1 dans le but de faciliter l'interprétation des résultats). Les Y ont donc été générés en utilisant l'équation ci-dessus avec un terme d'erreur normalement distribué ayant pour matrice de covariance $\sigma_u^2 I$. Dans les différentes expériences, σ_u a été fixé à différentes valeurs dans le but d'obtenir différentes valeurs du coefficient de détermination théorique qui est défini comme suit:

$$\tilde{R}^2 = (\beta'\tilde{X}'\tilde{X}\beta) / (\beta'\tilde{X}'\tilde{X}\beta + N\sigma_u^2)$$

(les $\tilde{X}_i$ sont en déviations de la moyenne). Aussi, des erreurs de mesure normalement distribuées on été ajoutées aux variables $\tilde{X}$. L'importance relative de ces erreurs de mesure a été évaluée pour chaque variable par le ratio de la variance des erreurs de mesure divisé par la variance de la variable originale concernée ($\lambda_i = \frac{var(v_i)}{var(\tilde{X}_i)}$ où i=1,...4). Des échantillons aléatoires de différentes tailles on été obtenus de l'ensemble total disponible de 4400 observations. Dans toutes les expériences discutées plus bas, les erreurs de mesures introduites, affectent la variable $\tilde{X}_1$ (revenu annuel du chef du ménage) ou bien les variables $\tilde{X}_1$ et $\tilde{X}_3$ (revenu annuel des autres membres du ménage). Pour mieux comprendre l'importance relative des erreurs de mesure affectant les variables explicatives dans les différentes expériences, il est utile de regarder le Tableau 4. Par exemple, dans les expériences où seulement $\tilde{X}_1$ est affecté d'erreurs de mesure avec une variance de 5% de la variance de $\tilde{X}_1$, il est implicitement supposé que le revenu du chef du ménage est rapporté avec une erreur moyenne absolue de \$1526, ce qui correspond à 6.1% de la valeur moyenne de $\tilde{X}_1$. Il est aussi supposé que 95% des répondants vont rapporter la valeur de $\tilde{X}_1$ avec une erreur absolue plus petite que \$3825 (approximativement). Cela suggère qu'un cas d'erreur de 5% est probablement une représentation plausible des erreurs de mesure présentes dans les enquêtes de revenu. Un pourcentage plus haut correspondrait à des erreurs de mesure sur les variables qui sont plus difficiles à appréhender, comme la richesse personnelle nette ou bien sur d'autres types de variables économiques, comme les variables agrégées macro-économiques.

Tableau 4

Statistiques Sommaires pour L'échantillon Complet de 4400 Données

Variables	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_4$
	($)	(Années)	($)	(Pers.-Semaines)
Valeur min.	-17802	20	-15000	52
Valeur max.	67832	80	62713	468
Ecart type	8553	13.06	8183	67.32
Moyenne	25116	42.36	8116	167.49

Taille var. des erreurs		Val.	% de la moyenne	Val.	% de la moyenne	Val.	% de la moyenne	Val.	% de la moyenne
5%	$2\sigma_v$	3825	15.2	5.841	13.8	3659	45.1	30.106	17.97
	$E(\lvert v \rvert)$	1526	6.1	2.330	8.9	1460	17.9	12.01	7.2
10%	$2\sigma_v$	5409	21.5	8.259	19.5	5175	63.8	42.58	25.4
	$E(\lvert v \rvert)$	2158	8.5	3.285	7.8	2064	25.4	16.98	10.1
15%	$2\sigma_v$	6625	26.4	10.116	23.9	6338	78.1	52.15	31.1
	$E(\lvert v \rvert)$	2643	10.5	4.036	9.5	2528	31.2	20.8	12.4
25%	$2\sigma_v$	8553	34.1	13.060	30.8	8183	100.8	66.319	39.6
	$E(\lvert v \rvert)$	3412	13.6	5.210	12.3	3265	40.2	26.457	15.8
50%	$2\sigma_v$	12095	48.2	14.306	33.8	8964	110.4	73.745	44.0
	$E(\lvert v \rvert)$	4825	19.2	5.707	13.5	3576	44.1	29.419	17.6

Pour évaluer les performances relatives de β_H et β_E, trois critères on été retenus: le Biais, la Racine Carrée De L'Écart Quadratique Moyen (REQM), la Probabilité D'Erreur Du Type I lorsqu'on teste l'hypothèse (H0) que le paramètre estimé est égal à sa vraie valeur. Trois expériences on été effectuées soient:

1) expérience avec erreurs de mesure sur $\tilde{X}_1$ seulement, en fonction de λ pour un $\tilde{R}^2$ donné et avec un échantillon de 500 observations;

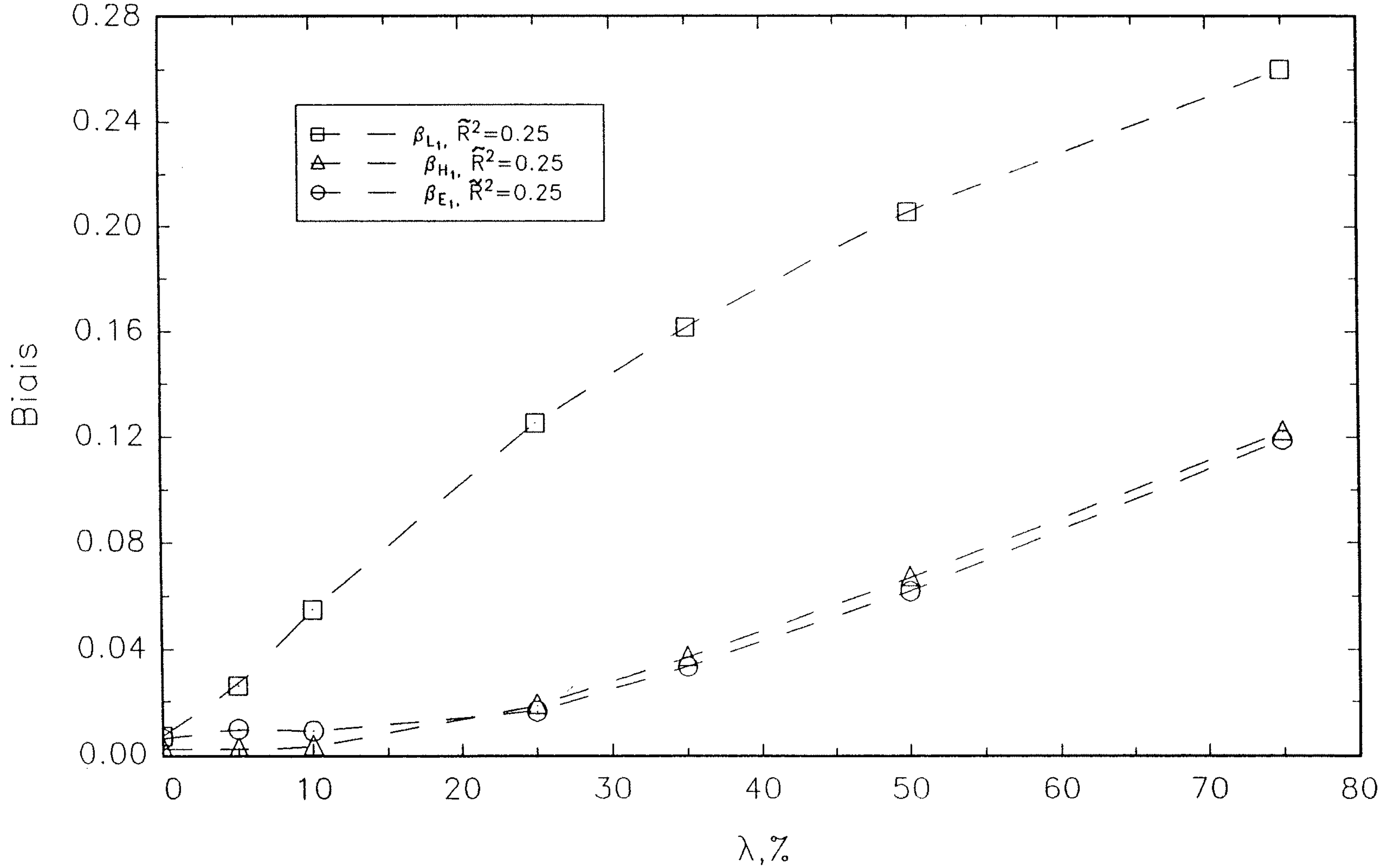
2) expérience avec erreurs de mesure sur $\tilde{X}_1$ et $\tilde{X}_3$ ($\lambda=5\%$) en fonction de la taille de l'échantillon pour un $\tilde{R}^2$ donné;

3) expérience avec erreurs sur $\tilde{X}_1$ seulement ($\lambda=25\%$) en fonction de la taille de l'échantillon (Note: chaque expérience a été faite avec 250 échantillons).

Comme il a été montré au Chapitre 1, le Biais asymptotique pour les M.C.O (β_L) provient du fait qu'il y a présence d'erreurs de mesure sur les variables explicatives. Soit, dans le cas où il n'y a qu'une seule variable explicative: $plim\beta_L=\beta/(1+\lambda)$. En faisant varier λ on peut voir l'évolution du Biais. Les figures $5a_1$ et $5a_2$ présentent le Biais des estimateurs $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$ en fonction de λ pour un $\tilde{R}^2$ donné, pour un échantillon de 500 observations. Par exemple, pour $\tilde{R}^2$=0.25 (Figure $5a_1$) si λ=10%, les Biais respectifs de $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$ sont environ de 0.06, 0.002, 0.01 ce qui veut dire: β_{H_1} et β_{E_1} ont une meilleure performance (pour ce critère) que β_{L_1}. On constate aussi que β_{E_1} ne performe pas mieux β_{H_1} pour tout λ. Si on reprend le même exemple mais avec $\tilde{R}^2$=0.85 (Figure $5a_2$), la performance de β_{H_1}, β_{E_1} demeure supérieure à celle de β_{L_1} et cette fois β_{E_1} surpasse β_{H_1} pour tous les λ, mais de peu.

Figure 5a$_1$

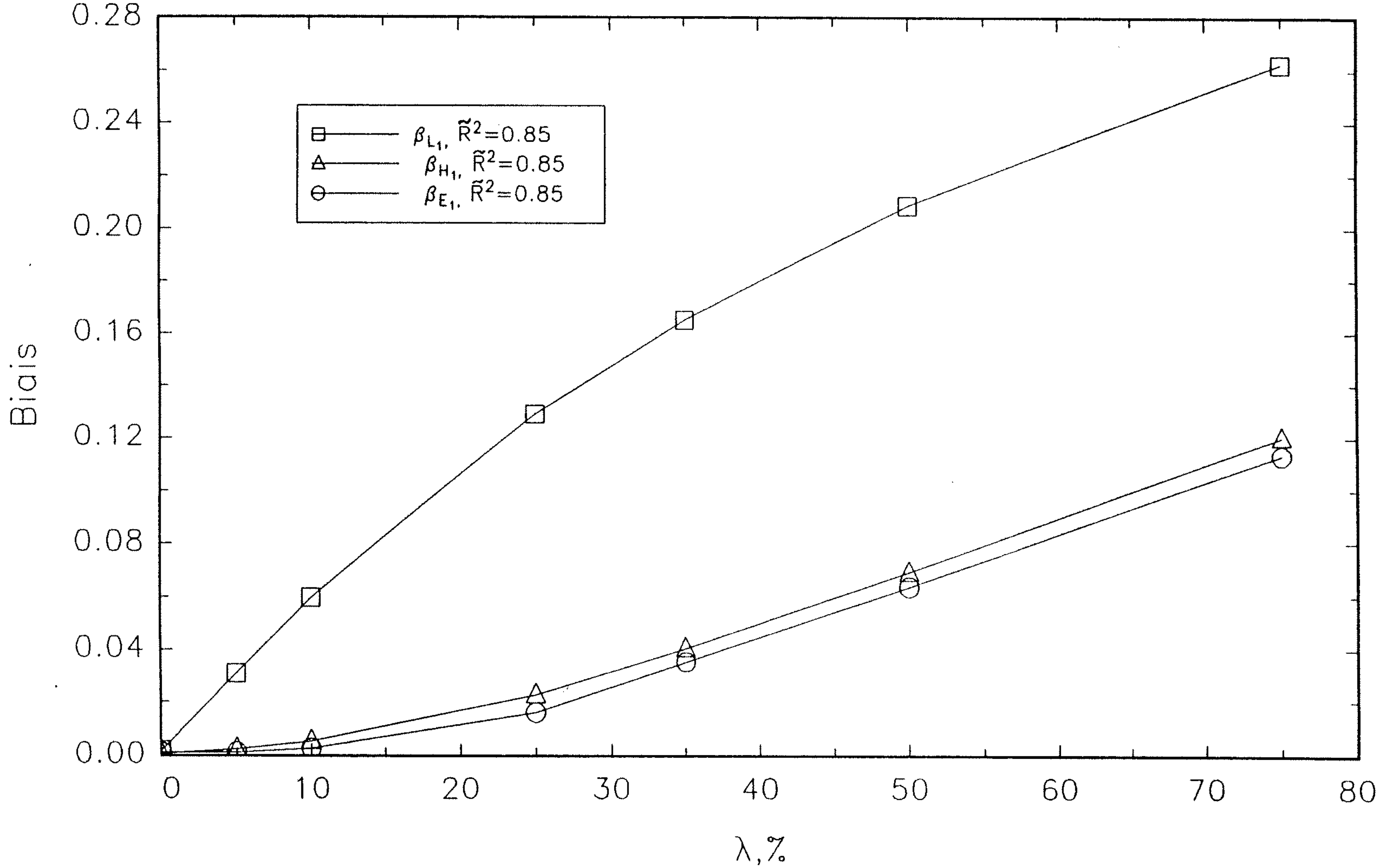
Biais en valeur absolue pour $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$, en fct. de λ



GAUSS Mon Mar 15 19:21:55 1993

Figure $5a_2$

Biais en valeur absolue pour $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$, en fct. de λ

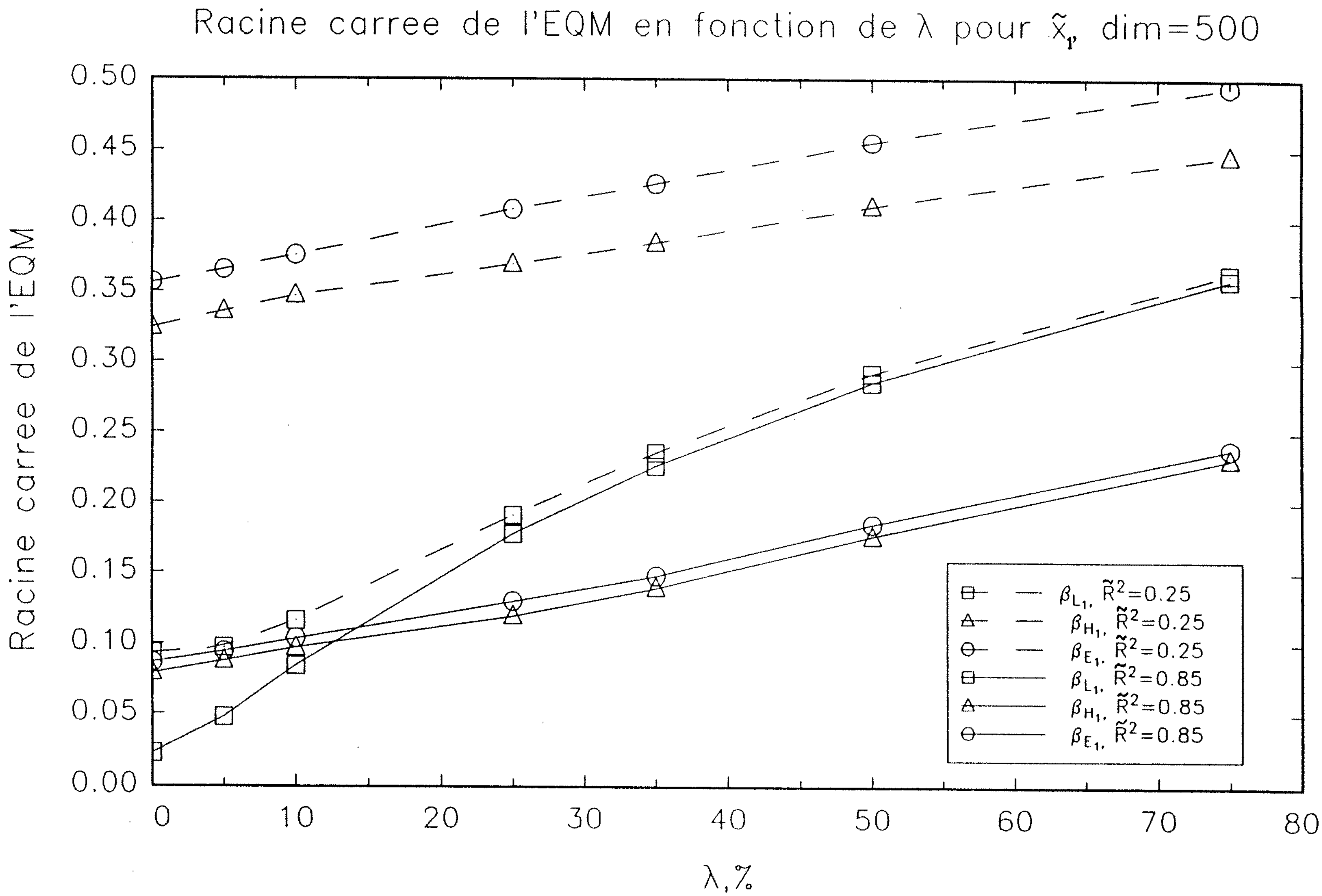


La figure 5b présente aussi les résultats de la première expérience mais cette fois-ci pour le deuxième critère retenu soit: la Racine Carrée De l'Écart Quadratique Moyen (REQM=$\sqrt{Biais^2 + Variance}$). On constate que les M.C.O (β_{L_1}) pour un $\bar{R}^2$=0.25, ont une meilleure performance que β_{H_1} et β_{E_1} même si la taille des erreurs de mesure est très grande. Par exemple, pour λ=75% le REQM de β_{L_1} est d'environ 0.36 alors que celui de β_{H_1} et de β_{E_1} est d'environ 0.50 et 0.45. Maintenant si le $\bar{R}^2$=0.85, on observe que β_{H_1} et β_{E_1} ont une meilleure performance que β_{L_1} à partir du moment où λ est légèrement supérieur à 10%, mais on remarque aussi que β_{E_1} n'est toutefois pas plus performant que β_{H_1}. La figure 5c présente le troisième critère retenu pour l'expérience 1) soit: la Probabilité d'Erreur Du Type I, lorsqu'un test-t est effectué sur un paramètre. On remarque que la probabilité de commettre une telle erreur, lorsqu'un test est effectué sur un paramètre estimé par les M.C.O (β_L), est très élevé relativement à celle de β_{H_1} et de β_{E_1}. Par exemple, si $\bar{R}^2$=0.85 et λ=10% et que β_{L_1} est utilisé pour estimer le paramètre au lieu de β_{H_1} ou de β_{E_1}, la probabilité de commettre une erreur du type I est d'environ 56% alors que cette probabilité est de seulement 2% pour β_{H_1} et 1.2% pour β_{E_1}.

La deuxième expérience a été effectuée en introduisant des erreurs de mesure sur $\tilde{X}_1$ et $\tilde{X}_3$ d'une taille λ=5% et les trois critères de performance des estimateurs sont évalués en fonction de la taille de l'échantillon. Les figures $6a_1$ et $6a_2$ présentent donc le premier critère d'évaluation soit: le biais moyen de $\beta_{L_{13}}$, $\beta_{H_{13}}$ et $\beta_{E_{13}}$ en fonction de la taille de l'échantillon pour un $\bar{R}^2$ donné (les indices 1 et 3 de $\beta_{L_{13}}$, $\beta_{H_{13}}$, $\beta_{E_{13}}$, signifient que c'est la moyenne des

GAUSS Mon Mar 15 19:16:53 1993

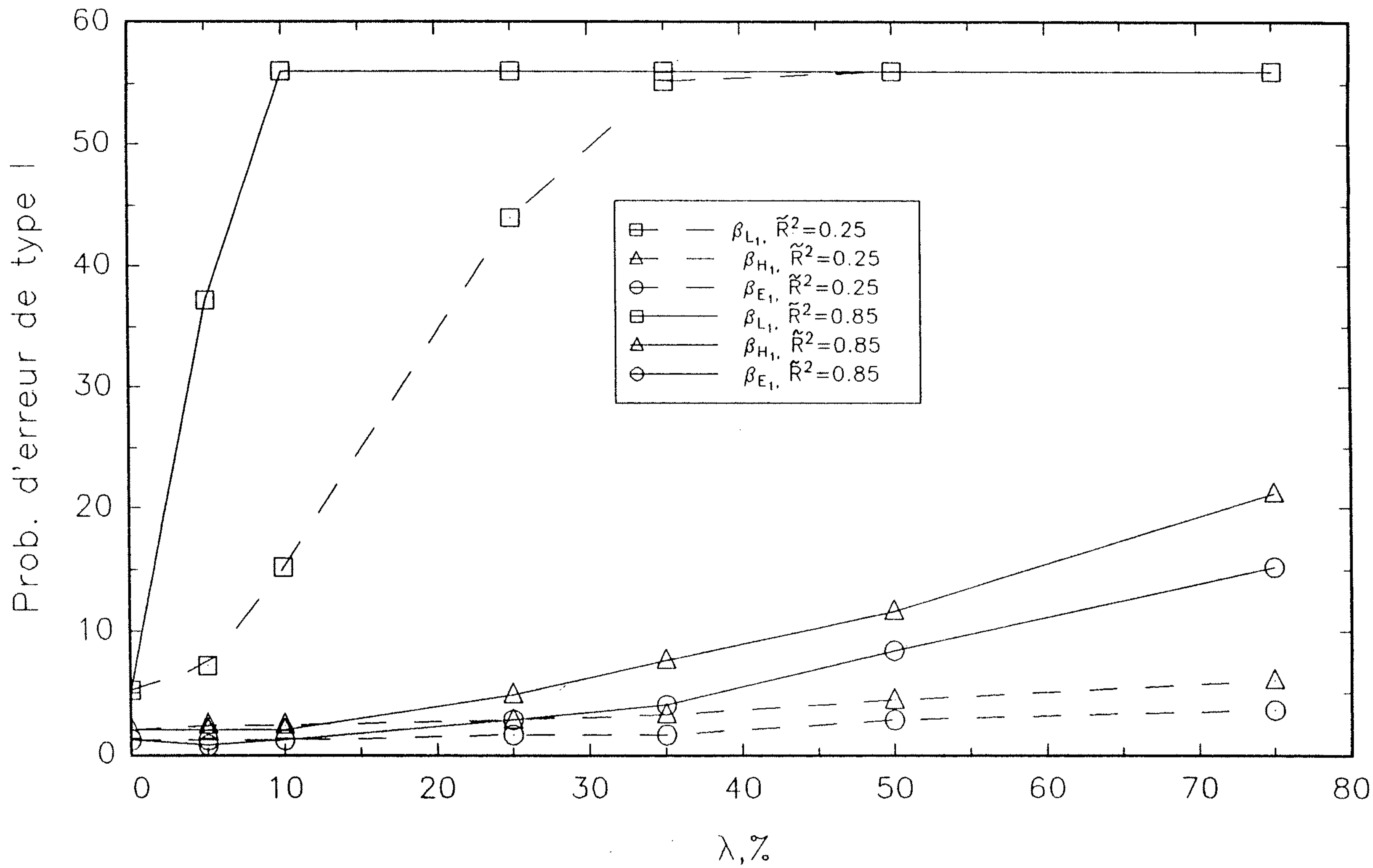
Figure 5b



GAUSS Mon Mar 15 19:15:28 1993

Figure 5c

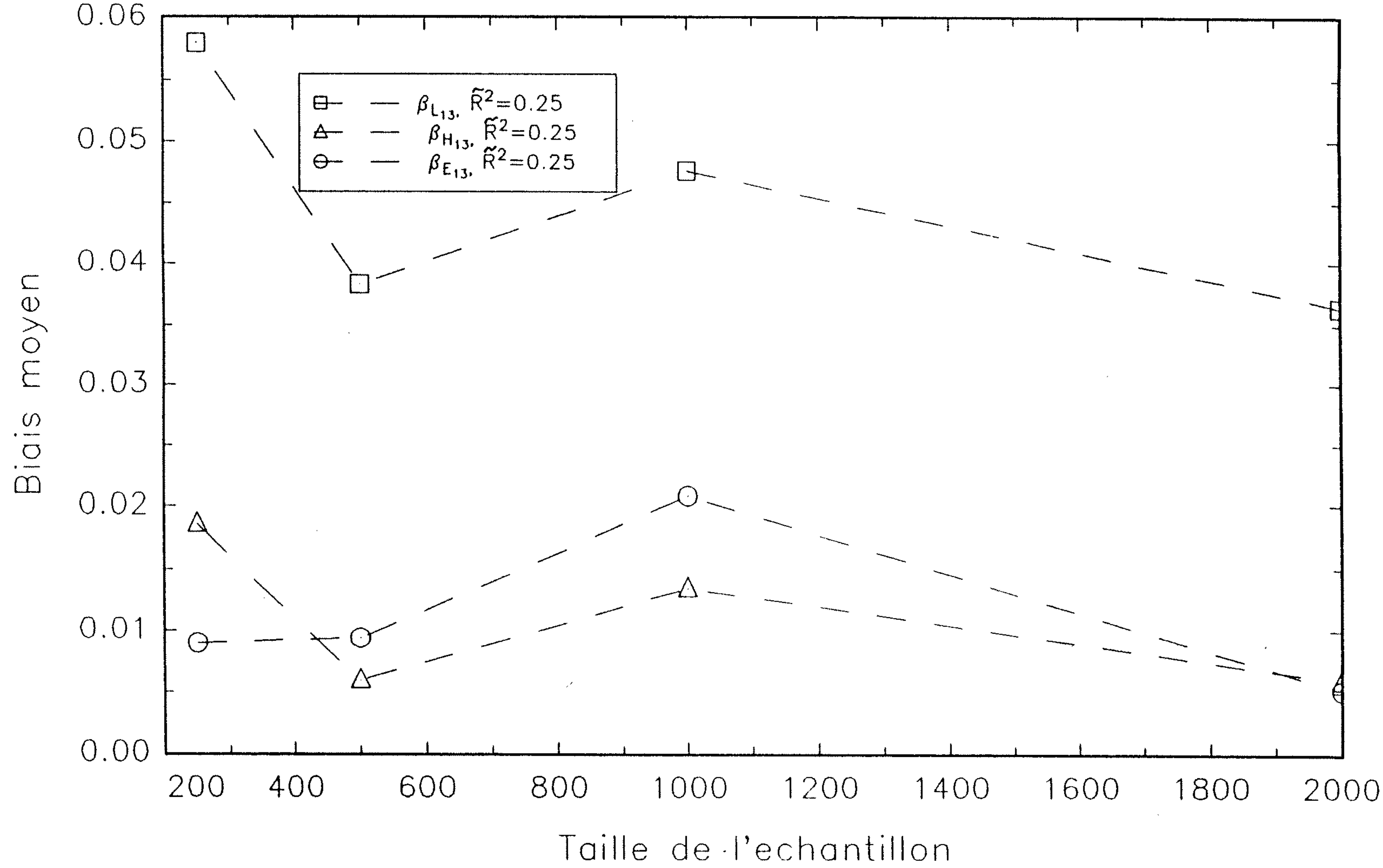
Prob. d'erreur de type I $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$ erreurs sur $\tilde{x}_1$, dim=500



GAUSS Mon Mar 15 19:22:53 1993

Figure $6a_1$

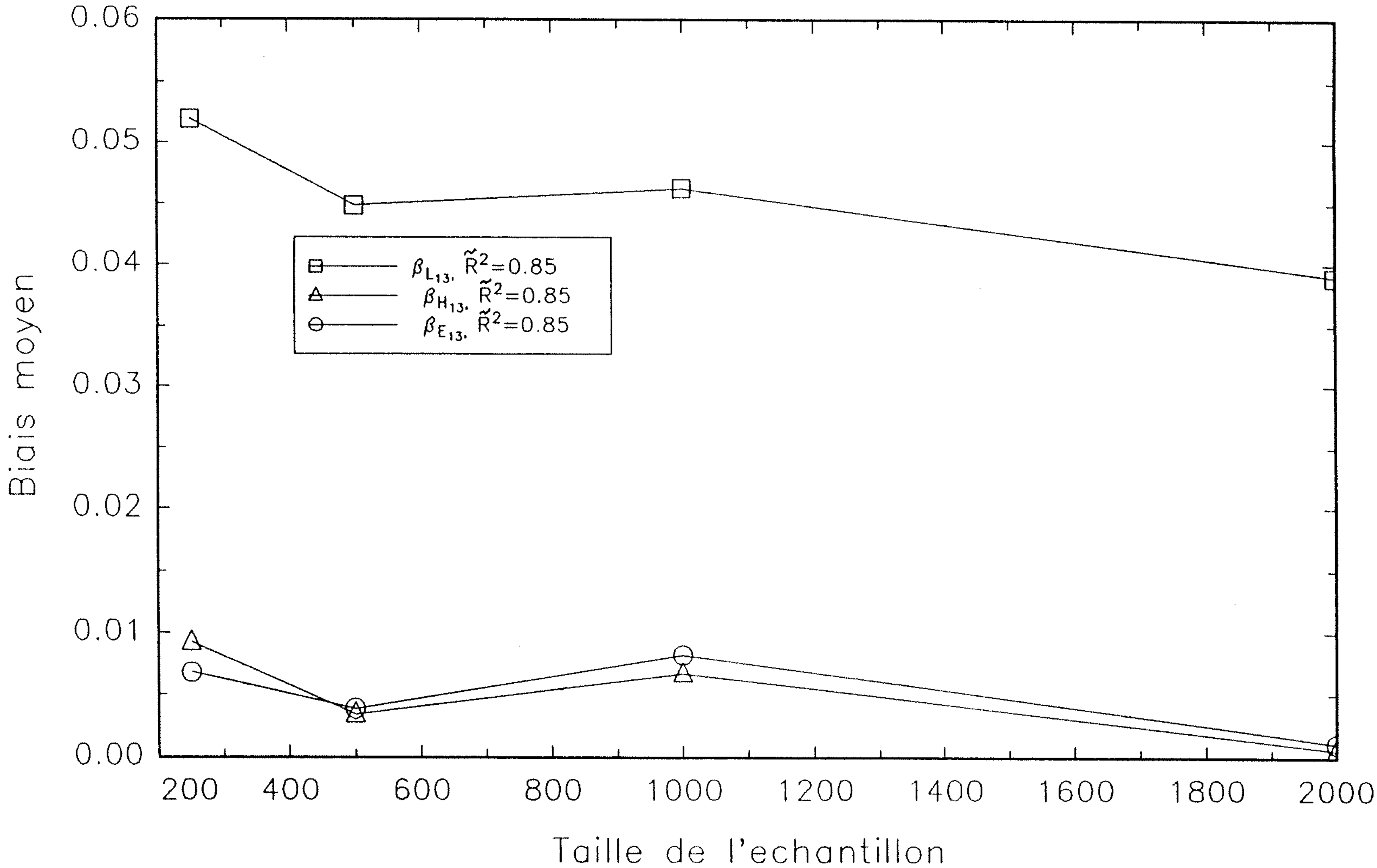
Biais moyen pour $\beta_{L_1}, \beta_{H_1}, \beta_{E_1}$ avec $\lambda=5\%$ pour $\tilde{x}_1$ et $\tilde{x}_3$



GAUSS Mon Mar 15 19:25:35 1993

Figure $6a_2$

Biais moyen de $\beta_{L_{13}}, \beta_{H_{13}}, \beta_{E_{13}}$ avec $\lambda=5\%$ pour $\tilde{x}_1$ et $\tilde{x}_3$



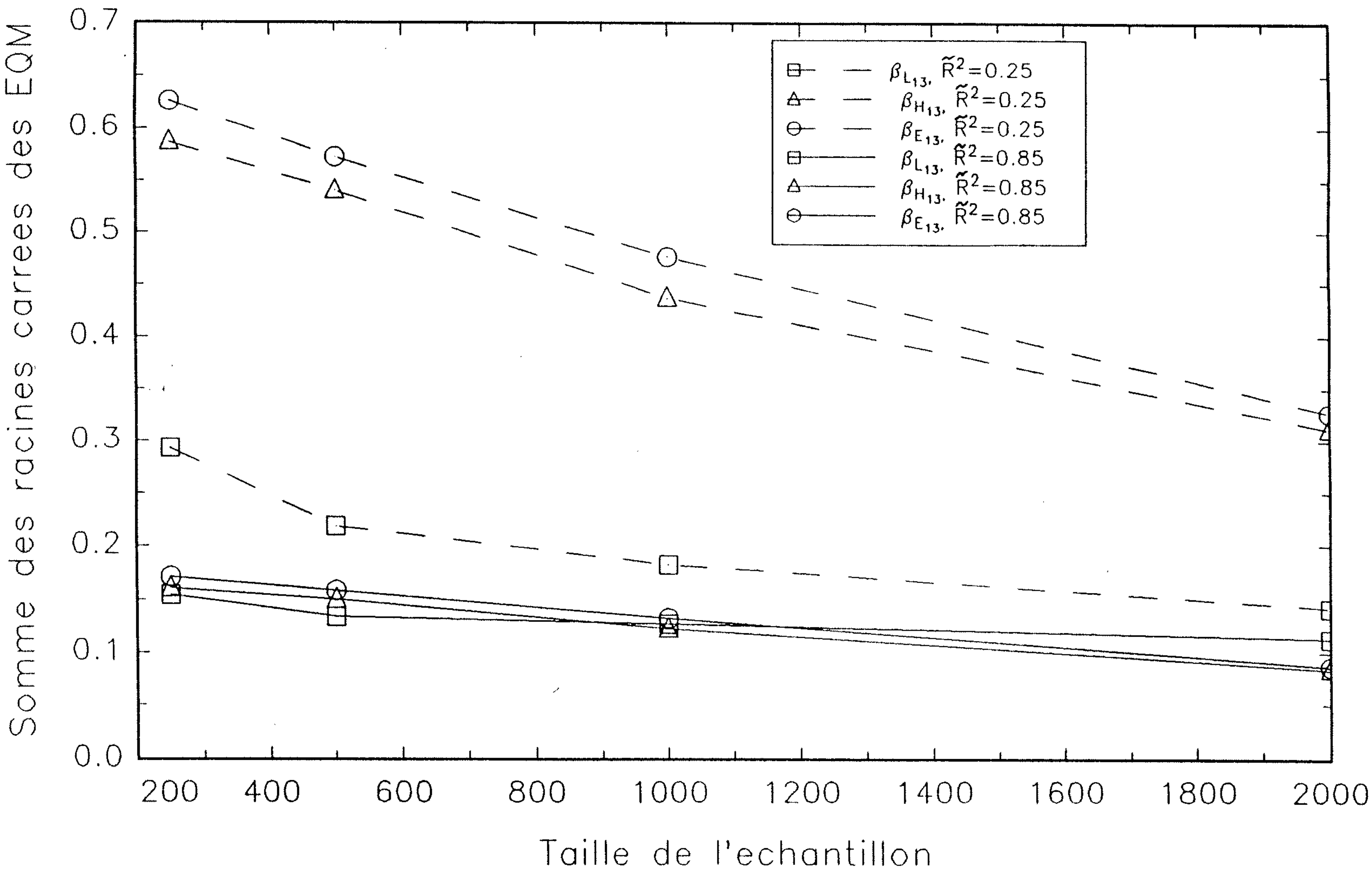
estimations des paramètres associés aux variables $\tilde{X}_1$ et $\tilde{X}_3$ qui est utilisée pour les différents critères d'évaluation). À la figure $6a_1$ ($\tilde{R}^2$=0.25), on remarque que le Biais de $\beta_{L_{13}}$ est plus grand que celui de $\beta_{H_{13}}$ et de $\beta_{E_{13}}$ pour toutes tailles d'échantillons et pour un échantillon de 2000 observations, le biais de $\beta_{E_{13}}$ est plus petit que celui de $\beta_{H_{13}}$. Maintenant pour un $\tilde{R}^2$=0.85, les estimateurs sont encore plus performants, car le Biais, par exemple pour un échantillon de 2000 observations, s'approche de zéro alors que celui de $\beta_{L_{13}}$ a augmenté par rapport à l'expérience faite avec le $\tilde{R}^2$=0.25. On voit aussi que $\beta_{E_{13}}$ semble être moins performant que $\beta_{H_{13}}$ pour l'ensemble des échantillons. La figure 6b montre la performance relative de $\beta_{L_{13}}$, $\beta_{H_{13}}$ et $\beta_{E_{13}}$ en terme de Racine Carrée De l'Écart Quadratique Moyen (REQM) en fonction de la taille de l'échantillon pour un $\tilde{R}^2$ donné. Lorsque le $\tilde{R}^2$=0.25 on constate que la performance de $\beta_{L_{13}}$ est nettement supérieure à celle de $\beta_{H_{13}}$ et à celle de $\beta_{E_{13}}$ pour l'ensemble des échantillons testés. Dans le cas où $\tilde{R}^2$=0.85, $\beta_{H_{13}}$ devient meilleur que $\beta_{L_{13}}$ lorsque la taille de l'échantillon est supérieure à 1000. Alors que $\beta_{E_{13}}$ surpasse $\beta_{L_{13}}$ seulement lorsque taille de l'échantillon est supérieure à 1200 observations.

La troisième expérience a été faite seulement pour la Racine Carrée de l'Écart Quadratique Moyen des estimateurs β_{L_1}, β_{H_1} et β_{E_1} avec λ=25% pour $\tilde{X}_1$ en fonction de la taille de l'échantillon pour un $\tilde{R}^2$ donné. A la figure 7, on constate que β_{L_1} est supérieur à β_{H_1} et β_{E_1} avec $\tilde{R}^2$=0.25% pour les différents échantillons expérimentés. Mais on remarque aussi que l'écart entre les nouveaux estimateurs et β_{L_1} diminue à mesure que l'échantillon augmente. Pour

GAUSS Mon Mar 15 19:30:40 1993

Figure 6b

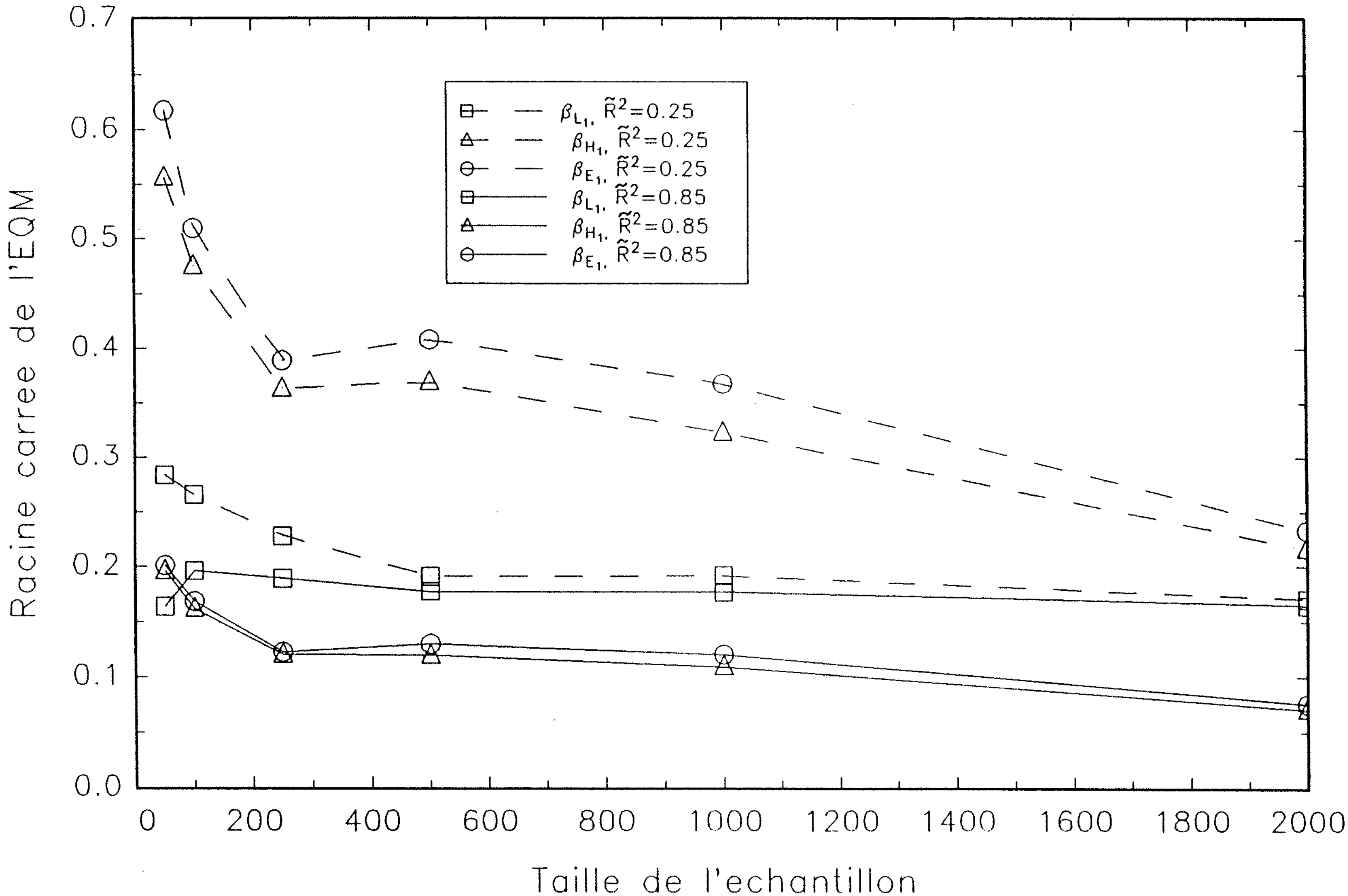
Somme des REQM avec $\lambda=5\%$ pour $\tilde{x}_1$ et $\tilde{x}_3$



GAUSS Mon Mar 15 19:33:48 1993

Figure 7

Racine carree de l'EQM avec $\lambda=25\%$ pour $\tilde{x}_1$



$\tilde{R}^2$=0.85, β_{H_1} et β_{E_1} on une performance supérieure à β_{L_1} lorsque l'échantillon a une taille supérieure à 75 observations. On remarque aussi que la performance de β_{H_1} semble supérieure à celle de β_{E_1} pour l'ensemble des échantillons.

Deux expériences ont été faites pour évaluer la puissance du test d'erreurs de mesure sur les variables explicatives.

1) La première expérience a été effectuée avec un échantillon d'une taille égale à 500. La Puissance du test fut évaluée en fonction de λ pour un $\tilde{R}^2$ donné où seulement la variable $\tilde{X}_1$ était affectée d'erreurs de mesure. On peut voir les résultats de cette expérience à la figure 8. On remarque que le test est puissant seulement quand λ est très grand pour un $\tilde{R}^2$=0.85. Dans le cas où le $\tilde{R}^2$=0.25, le test n'est pas puissant même pour des erreurs très grandes. Par exemple, si λ=75 % la Puissance du test n'est que d'environ 17%.

2) La deuxième expérience qui a été faite avec des erreurs de mesure λ=5% sur $\tilde{X}_1$ et $\tilde{X}_3$ fût évaluée en fonction de la taille de l'échantillon pour un $\tilde{R}^2$ donné. A la figure 9, on constate que le test n'est pas très puissant lorsque le $\tilde{R}^2$=0.25 même pour un échantillon d'une taille égale à 2000. Dans le cas où le $\tilde{R}^2$=0.85, le test a une puissance de 100% quand l'échantillon est de 2000 observations. Pour que la puissance du test dépasse 50%, il faut que l'échantillon dépasse 700 observations.

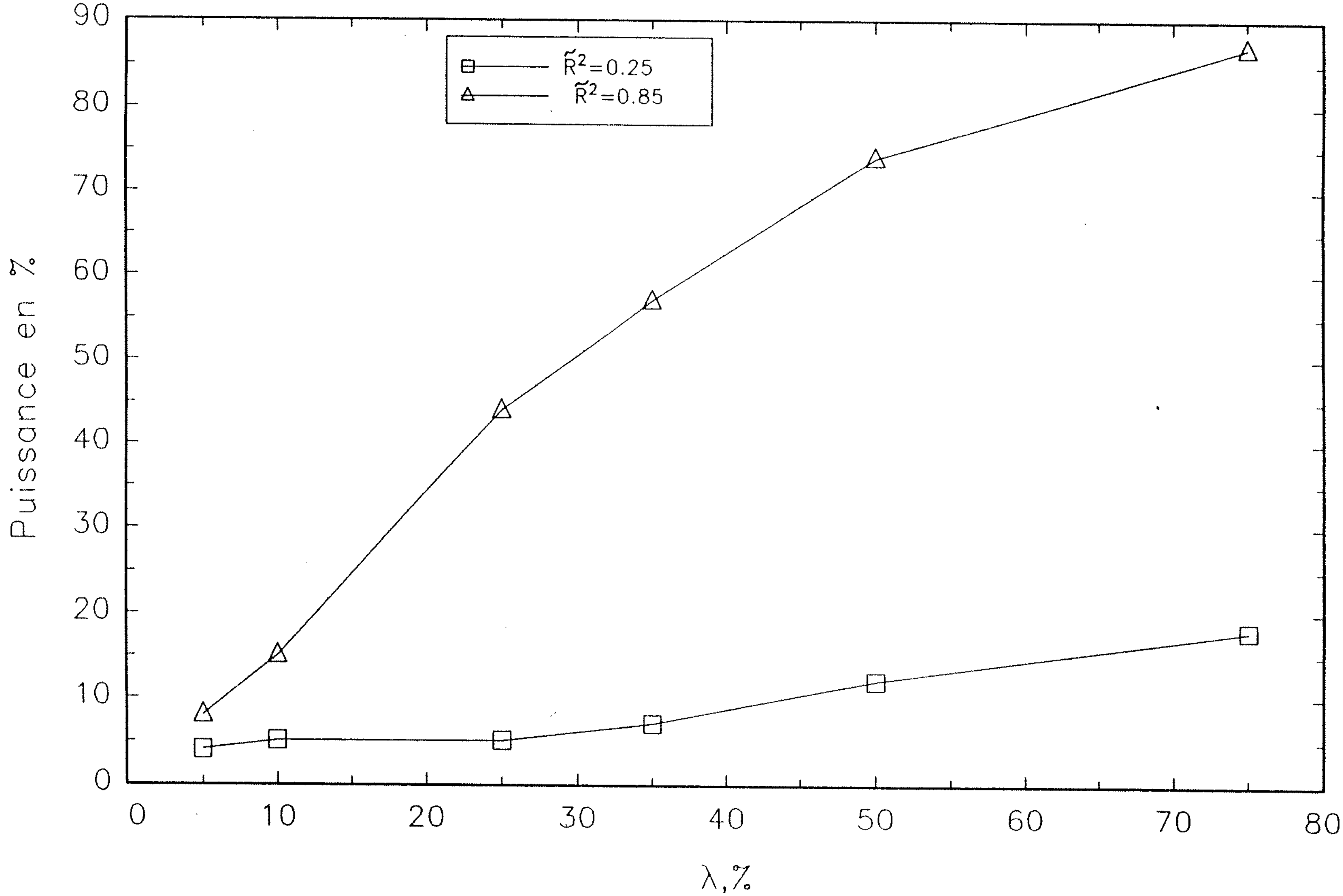
Pour résumer les résultats des expériences avec les données de l'enquête sur les consommateurs on peut dire que:

i) β_E ne semble pas tellement meilleur que β_H pour les expériences effectuées jusqu'ici. (Note

GAUSS Mon Mar 15 19:35:24 1993

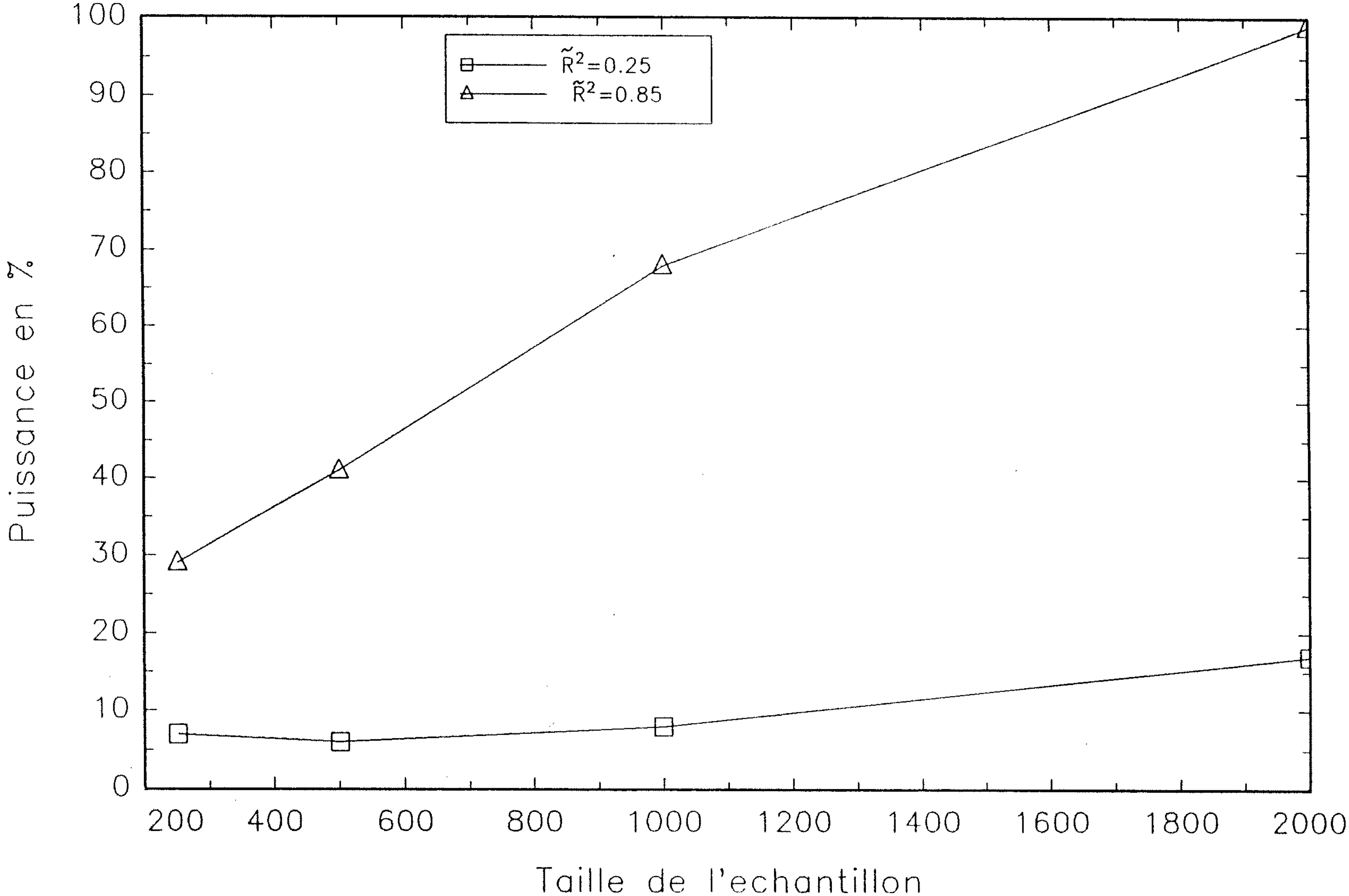
Figure 8

Puissance du Test, erreurs sur $\tilde{x}_1$ (dimen=500)



GAUSS Mon Mar 15 19:36:52 1993

Figure 9
Puissance du Test, $\lambda=5\%$ pour $\tilde{x}_1$ et $\tilde{x}_3$



: les expériences effectuées supposaient qu'environ la moitié des échantillons comportaient des erreurs de mesure. La performance de β_E devrait s'améliorer si 100% des échantillons comportaient des erreurs de mesure).

ii) β_H est sans biais lorsqu'il n'y a pas d'erreur de mesure. Donc quand il y a un doute sur la présence d'erreurs de mesure, il n'y a pas de risque sérieux d'utiliser β_H au lieu de β_L.

iii) Quand il y a présence d'erreurs de mesure, β_H et β_E sont moins biaisés que β_L. Ceci demeure vrai même si la taille des erreurs de mesure (λ) est aussi petite que 5% .

iv) Lorsqu'il y a présence d'erreurs de mesure, la probabilité d'erreur du type I avec β_H et β_E semble toujours près de la valeur choisie (dans notre cas, la valeur choisie est de 5%). Alors que si on utilise β_L, la Probabilité d'Erreur De Type I demeure grande même si la taille des erreurs de mesure est aussi petite que λ=5%. β_E ne semble pas performer tellement mieux que β_H pour ce critère. (Note 2: La note 1 peut expliquer ce fait)

v) Le REQM de β_H et β_E est plus petit que celui de β_L pour des erreurs relativement petites si le $\bar{R}^2$ est fort et l'échantillon est suffisamment grand, lorsqu'une seule des variables est affectée d'erreurs de mesure. Pour le cas où plus d'une variable est affectée d'erreurs de mesure, la performance relative des estimateurs β_H et β_E est améliorée.

Enfin, le test montre suffisamment de puissance si plus d'une variable explicative est affectée d'erreurs et si la taille de l'échantillon est relativement grande pour un $\bar{R}^2$ fort. Dans le cas où un seule des variables explicatives est affectée d'erreurs, ces erreurs doivent êtres grandes et le $\bar{R}^2$ fort pour que le test puisse les détecter. Donc le test n'est pas très puissant dans ce cas.

CONCLUSION

La présence d'erreurs de mesure dans les variables explicatives est un problème majeur en analyse économétrique. Les expériences de Monte Carlo effectuées dans ce rapport ont montré que le fait d'ignorer cette réalité peut mener à des erreurs d'inférence importante, si les techniques d'estimations utilisées ne sont pas appropriées. Nous avons suggéré un nouvel estimateur β_H et son extension β_E. Cet estimateur a été plus performant que l'estimateur des moindres carrés ordinaires (M.C.O.) lorsqu'il y avait présence d'erreurs de mesure dans les variables, si ces erreurs étaient relativement importantes. Nous avons aussi suggéré un test pour la présence d'erreurs dans les variables. Ce test s'est montré suffisamment puissant si plus d'une variable explicative était affectée d'erreurs et si ces erreurs étaient relativement grandes. De plus, la puissance du test augmente quand le $\tilde{R}^2$ augmente. Enfin, d'autres expériences devraient être fait sur la puissance du test pour arriver à des conclusions définitives. Il est à noter aussi que des expériences, non décrites dans ce rapport, suggèrent que la corrélation entre les vraies variables indépendantes peut avoir un effet important sur la performance du nouvel estimateur (β_H). De plus, l'effet de la corrélation entre les erreurs de mesure des différentes variables n'a pas été présenté dans ce rapport.

REMERCIEMENTS

Je tiens à remercier le Professeur Marcel G. Dagenais d'avoir accepté de diriger ce travail et aussi pour les nombreux conseils qu'il m'a procurés tout au long de cette rédaction.

Je remercie également Norma Cosaya, Denise d'Aragon, Lucille Robillard, Stéphane Gouin et Boussad Amis pour les nombreux conseils concernant la forme de ce rapport de recherche. Cependant, il va sans dire que je prends sur moi l'entière responsabilisé de toutes les erreurs qui peuvent se glisser dans ce texte.

RÉFÉRENCES

Dagenais M.G. and Denyse Laberge Dagenais (1991), "Estimation and Testing in Regression Models with Errors in the Variables", manuscrit, C.R.D.E., Université de Montréal.

Dagenais M.G. and Denyse Laberge Dagenais (1992), "Discriminating Between Consumption and Market Betas, Using a New estimator for Regressions with Errors in the Variables", manuscrit, C.R.D.E., Université de Montréal.

Dagenais M.G. and Denyse Laberge Dagenais (1992), "Estimation and Testing in Regression Models with Errors in the Variables", manuscript, C.R.D.E., Université de Montréal.

Drion,E.F. (1951), "Estimation of Parameters of a Straitht Line and of the Variance of the Variables, If They Are Both Subject to Error", INDAGATIONES MATHEMATICAE 13, 265-260.

Durbin, J. (1954), "Errors in Variables", INTERNATIONAL STATISTICAL REVIEW 22, 23-32.

Fomby T.B., Carter Hill R. and Johnson S.R. (1984), ADVANCED ECOCOMETRIC METHODES, Springer-Verlag, New York.

Johnston, J. (1985), METHODES ECONOMETRIQUES, 3e édition, Économica, Paris.

Judge G.G, Griffiths W.E., Carter Hill R., Lütkepohl H., Lee T.-C. (1985), THE THEORY AND PRACTICE OF ECONOMETRICS, Second Edition, John Wiley & Sons, New York.

Geary, R.C (1942), "Inherent Relations Between Random Variables", PROC. ROYAL IRISH ACADEMY 47, 63-76.

Hausman, J.D. (1978), "Specification Tests in Econometrics", ECONOMETRICA 46(6), November, 1251-1271.

Imhof, P.J. (1961), "Computing the Distridbution of Quadratic Forms in Normal Variables", BIOMETRIKA 48, 419-426.

Langaskens, Y. and M. Van Rickeghem (1974), "A New Method to Estimate Measurement Errors in National Income Account Statistics : The Belgian Case", INT. STAT. REV. 42(3), 283-290.

MacKinnon, E. (1992), "Model Specification Tests and Artificial Regressions", JOURNAL OF ECONOMIC LITERATURE, 102-146.

Morgenstern, O. (1972), L'ILLUSION STATISTIQUE : PRÉCISION ET INCERTITUDE DES DONNÉES ÉCONOMIQUES, Dunod, Paris.

Pal, M. (1980), "Consistent Moment Estimators of Regression Coeficients in the Presence of Errors in Variables", Journal of Econometrics 14(3), December, 349-364.

Statistics Canada (1988), "Family Expenditure in Canada, 1986", Survey of Family Expenditure, Puplic Use Microdata file.

Theil, H. (1971), PRINCIPLES OF ECONOMETRICS, John Wiley & Sons, New York.

Theil, H. and A.S. Goldbeger (1961), "On Pure and Mixed Statistical Estimation in Economics", INTERNATIONAL ECONOMIC REVIEW, 65-78.

White, H. (1984), ASYMPTOTIC THEORY FOR ECONOMETRICIANS, Academic Press inc., Orlando.

APPENDICE A

1) Comme il a été mentionné à la partie II du chapitre 1, la preuve multivariée de la non-convergence des M.C.O. n'a pas été présentée. Nous la présentons ici. Considérons :

$$Y = \tilde{X}\beta + u$$

le modèle de régression linéaire multivariée (sans terme constant) où $\tilde{X}$ est une matrice de variables explicatives dimension $N \times K$ avec $plim(\tilde{X}'\tilde{X}/N) = \Sigma_{\tilde{X}\tilde{X}}$, Y est un vecteur de dimension $N \times 1$ (variable dépendante) et u est un vecteur d'erreurs de régression de dimension $N \times 1$ avec $E(u) = 0$, $E(uu') = \sigma_u^2 I$ et supposons aussi que $\tilde{X}$ n'est pas observé mais plutôt X où $X = \tilde{X} + v$ et v est une matrice d'erreurs de mesures sur les variables explicatives de dimension $N \times K$ avec $E(v) = 0$, $E(v_i, u) = 0$, $plim(v'v/N) = \Sigma_{vv}$. Si on substitue $\tilde{X}$ par sa valeur dans la première équation on obtient:

$Y = (X - v)\beta + u$

$= \beta X - v\beta + u = \beta X + e$ où $e = u - v\beta$ si on applique les M.C.O. sur $Y = X\beta + e$ et qu'on remplace Y par sa valeur on obtient :

$\beta = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'e = \beta + (X'X)^{-1}X'e$. Pour montrer que β n'est pas convergent lorsqu'il y a des erreurs de mesure sur les variables explicatives il faut calculer le $plim\beta$:

$plim\beta = \beta + (plim(X'X/N))^{-1}(plim(X'e/N))$

où $plim(X'e/N) = plim(v'(u - v\beta)/N)$

$=plim(v'u-v'v\beta/N)=-plim(v'v/N)\beta=-\Sigma_{vv}\beta$ et

$plim(X'X/N)=plim[(\tilde{X}'+v')(\tilde{X}+v)/N]=plim[(\tilde{X}'\tilde{X}/N)+(\tilde{X}'v/N)+(v'\tilde{X}/N)+(v'v/N)]$

$=\Sigma_{\tilde{X}\tilde{X}}+\Sigma_{vv}$ donc $plim\hat{\beta}=\beta-(\Sigma_{\tilde{X}\tilde{X}}+\Sigma_{vv})^{-1}\Sigma_{vv}\beta=\beta(I-\lambda)$ où

$\lambda=(\Sigma_{vv}+\Sigma_{\tilde{X}\tilde{X}})^{-1}\Sigma_{vv}$

2)

SOLUTION AU SYSTÈME D'ÉQUATIONS DU MAXIMUM DE VRAISEMBLANCE

On a les quatres équations suivantes:

1) $X^*(Y-X^*\beta)=0$

2) $(X-X^*)'(X-X^*)=N\sigma_v^2$

3) $(Y-X^*\beta)'(Y-X^*\beta)=N\sigma_e^2$

4) $(X-X^*)\sigma_e^2+\beta(Y-X^*\beta)\sigma_v^2=0$

Si on substitue σ_e^2 et σ_v^2 par leur valeurs dans 4) on obtient :

$$(X-X^*)\frac{(Y-X^*\beta)'(Y-X^*\beta)}{N}+\beta(Y-X\beta)\frac{(X-X^*)'(X-X^*)}{N}=0$$

si on simplifie on obtient: $((Y-X^*\beta)')'=(-\beta(X-X^*)')'=Y-X^*\beta=-\beta(X-X^*)=H$ et si on multiplie 4) par β on obtient:

$\beta(X-X^*)\sigma_e^2+\beta^2(Y-X^*\beta)\sigma_v^2=0$ qu'on appelle 4').

Enfin, on a qu'à substituer dans 4') les coefficients de σ_v^2, σ_e^2 par H soit: $-H\sigma_v^2+\beta^2 H\sigma_e^2=0$, en simplifiant on trouve $\sigma_v^2=\beta^2\sigma_e^2$.

APPENDICE B

a) Pour montrer que β_H et β_E sont convergents, nous utilisons le concept de convergence en probabilité et de ce fait, nous utiliserons l'opérateur $plim$. Pour ce faire, il existe des théorèmes généraux qui sous des hypothèses appropriées permettent de calculer le $plim$ simplement en calculant l'espérance asymptotique, sans calculer la variance asymptotique, ce qui signifie qu'on a qu'à vérifier la nullité du biais asymptotique.

Il existe pour cela, entre autres, deux théorèmes:

Th.1: Observations indépendantes et distribuées de façon identiques. Si $z_1, z_2, \ldots, z_t$ est un ensemble de variables aléatoires et que $E(z_t) < \infty$ soit $E(z_t) = u$

et $cov(z_t, z_{t+n}) = 0$ alors $plim(\frac{\sum_{t=1}^{N} z_t}{N}) = \lim_{N \to \infty} E(\frac{\sum_{t=1}^{N} z_t}{N})$

Th.2: Observations indépendantes et distribuées de façon hétérogènes . Si $z_1, z_2, \ldots, z_t$ est un ensemble de variables aléatoires et que

$\exists\ \sigma > 0,\ 0 < \Delta < \infty\ /\ E(|z_t|^{1+\sigma}) < \Delta^{N \to \infty} < \infty, \forall t$ soit $E(z_t) = u_t$ alors

$$plim \frac{\sum_{i=1}^{N} z_t}{N} = \lim_{N \to \infty} E \frac{\sum_{t=1}^{N} z_t}{N}$$

(Note: Th.1, Th.2 sont expliqués dans le livre de WHITE: ASYMPTOTIC THEORY FOR

ECONOMETRICIANS).

b) Comme il a été mentionné dans le CHAPITRE 2, β_H et β_E sont convergents mais celà n'a pas été montré. Alors montrons maintenant que ces deux estimateurs sont effectivement convergents sous l'hypothèse H1. Pour cela, il faut montrer que β_D et β_P sont convergents ensuite l'on pourra montrer que β_H et β_E le sont aussi. Sous l'hypothèse H1 on a:

$X=\tilde{X}+v$ avec $V(vect(v))=\Sigma\otimes I_N$

$$Y=\alpha i+\tilde{X}\beta+u$$

si on substitue $\tilde{X}$ dans Y, alors on a :

$$Y=\alpha i+X\beta-v\beta+u$$

en déviation de la moyenne, Y devient :

$$y=x\beta+Au-Av\beta$$

donc pour β_D :

$$\beta_D=(z_1'x)^{-1}z_1'y$$

(Note: il est supposé que $(z_1'x/N)^{-1}$ existe).

Si on substitut y dans β_D :

$\beta_D=\beta+(z_1'x/N)^{-1}z_1'Au/N-(z_1'x/N)^{-1}(z_1'Av/N)\beta$

Calcul du $plim\beta_D$:

$$plim\beta_D=\beta+[plim(z_1'x/N)]^{-1}plim(z_1'Au/N)-[plim(z_1'x/N)]^{-1}plim(z_1'Av/N)\beta$$

où

$plim(z_1'Au/N)=E_{N\to\infty}(z_1'Au/N)=0$ parce que les éléments de z_1 et de Au sont indépendants et que $E(u)=0$. Il nous reste a montrer que $plim(z_1'Av/N)=0$ soit :

$plim(z_{1j}'Av_K/N)$ pour $(j,K=1,\dots,K)$ et $\overline{v}_K=\frac{\sum_{i=1}^N v_{ik}}{N}$ alors,

$$plim\frac{\sum_{i=1}^N x_{ij}^2(v_{ik}-\overline{v}_k)}{N}=plim(\Sigma x_{ij}^2 v_{ik}/N)-plim\overline{v}_K\ plim\Sigma x_{ij}^2/N$$ sachant que

$plim\overline{v}_K=E_{N\to\infty}(\Sigma_{i=1}^N v_{ik}/N)=0$ parce que $E(v_{ik})=0$ donc

$plim\ \Sigma_{i=1}^N x_{ij}^2(v_{iK}-\overline{v}_k)/N=plim\ \Sigma(\tilde{x}_{ij}+v_{ij})^2 v_{iK}/N$

$=plim\Sigma\tilde{x}_{ij}^2 v_{iK}/N+2plim\Sigma\tilde{x}_{ij}v_{ij}v_{ik}/N+plim\Sigma v_{ij}^2 v_{ik}/N$ où

1) $plim\ \Sigma(\tilde{x}_{ij}^2 v_{iK}/N)=E_{N\to\infty}(\tilde{x}_{ij}^2 v_{iK}/N)=0$ car $\tilde{x}_{ij}^2$ et v_{ik} sont indépendants.

2) $plim\ \Sigma\tilde{x}_{ij}v_{ij}v_{ij}/N=E_{N\to\infty}\ (\Sigma\tilde{x}_{ij}v_{ij}v_{iK}/N)={}_{N\to\infty}\Sigma E(v_{ij}v_{iK})E(\tilde{x}_{ij})/N$

$=\mu_{jK}(1,1)\Sigma\ E(\tilde{x}_{ij})/N=0$ car $E(\tilde{x}_{ij})=E(\tilde{X}_{ij})-\overline{X}_j=0$ et $\mu_{jK}(1,1)=E(v_{ij}v_{iK})$ où

$\mu_{iK}(l,m)=Ev_{ij}^l v_{iK}^m$

3) $plim\ \Sigma\ v_{ij}^2 v_{iK}/N=E_{N\to\infty}\ \Sigma v_{ij}^2 v_{iK}/N={}_{N\to\infty}\Sigma\ \mu_{jK}(2,1)/N=0$ où

$\mu_{jK}(2,1)=E(v_{ij}^2 v_{iK})=0$ p.c.q les moments d'ordres impairs de v.a. normalement distribués sont nuls (THEIL p.62-63).

Donc $plim\ \beta_D = \beta$

Maintenant il faut montrer que β_P est convergent sous Hi.

On a :

$$\beta_P = (z_2'x)^{-1} z_2' y$$

(Note: il est que supposé que $(z_2'x)^{-1}$ existe)

où $z_2' = z_3' - 3D(x'x/N)x'$ comme pour β_D, les éléments de β_P sont en déviations de la moyenne et donc $y = x\beta - Av\beta + Au$ est le modèle de régression sous Hi mis en déviation de la moyenne. β_P est égal à $\beta_P = \beta + (z_2'x)^{-1} z_2'Au - (z_2'x)^{-1} z_2'Av\beta$. Pour voir si β_P est convergent, il faut calculer le $plim\ \beta_P$ soit: $plim\ \beta_P = \beta + [plim(z_2'x/N)]^{-1} plim[(z_2'Au/N) - (z_2'Av\beta/N)]$ et β_P sera convergent si (1) $plim\ [(z_2Au/N) - (z_2'Av\beta/N)] = 0$. En substituant z_2' par sa valeur dans (1), on obtient (1'):

$$plim(z_3'Au/N) - plim(z_3'Av\beta/N) - 3D[plim(x'x/N)]\,plim(x'Au/N)$$

$$+\ 3D[plim(x'x/N)]\,plim(x'Av\beta/N).$$

Sachant que u est indépendant de z_3 et de x alors on a que :

$plim(z_3'Au/N) = E_{N\to\infty}(z_3'Au/N) = 0$ et que $plim(x'Au/N) = E_{N\to\infty}(x'Au/N) = 0$.

Et:

$$plim(z_3'Av\beta/N) = plim(z_3'[I_N - ii'/N]v\beta/N) = plim(z_3'v\beta) - plim(\frac{z_3'[ii'/N]v\beta}{N})$$

où $plim(z_3'[ii'/N]v\beta) = plim(z_3'i/N)plim(i'v\beta/N) = 0$ et donc

$plim(z_3'Av\beta/N) = plim(z_3'v\beta/N)$ de la même façon :

$plim(x'Av\beta/N) = plim(x'v\beta/N)$.

Enfin pour montrer que $plim\ \beta_P = \beta$ il faut montrer que:

$$plim\ z_3'v/N + 3D(plim x'x/N)\ plim(x'v/N) = 0$$

Pour un élément représentatif (j,K) on a par exemple :

$$plim(z_{3j}'v_K/N) = plim(\Sigma_{i=1}^N(\tilde{x}_{ij}+v_{ij}-\overline{v}_j)^3 v_{iK}/N$$

$$= plim(\Sigma_{i=1}^N \tilde{x}_{ij}^3 v_{iK}/N) + plim(\Sigma_{i=1}^N v_{ij}^3 v_{iK}/N) - plim(\Sigma_{i=1}^N \overline{v}_j^3 v_{ik}/N)$$

$$+3plim(\Sigma_{i=1}^N \tilde{x}_{ij}^2 v_{ij} v_{iK}/N) - 3plim(\Sigma_{i=1}^N \tilde{x}_{ij}^2 \overline{v}_j v_{iK}/N) + 3plim(\Sigma_{i=1}^N \tilde{x}_{iK} v_{ij}^2 v_{iK}/N)$$

$$-3plim(\Sigma_{i=1}^N v_{ij}^2 \overline{v}_j v_{iK}/N) + 3plim(\tilde{x}_{ij}\overline{v}_j^2 v_{iK}/N) + 3plim(\Sigma_{i=1}^N v_{ij}\overline{v}_j^2 v_{iK}/N)$$

$$-6plim(\Sigma_{i=1}^N \tilde{x}_{ij} v_{ij}\overline{v}_j v_{iK}/N).$$

4) le $plim(\Sigma_{i=1}^N \tilde{x}_{ij}^3 v_{ik}/N) = E_{N\to\infty}(\Sigma \tilde{x}_{ij}^3 v_{ik}/N) = 0$ parce que $\tilde{x}_{ij}^3$ et v_{iK} sont indépendants et que $E(v_{iK}) = 0$.

5) $plim(\Sigma_{i=1}^N v_{ij}^3 v_{iK}/N) = E_{N\to\infty}(\Sigma v_{ij}^3 v_{iK}/N) = \mu_{jK}(3,1) \neq 0$ car les moments pairs d'une distribution normale sont différents de zéro.

6) $plim(\Sigma_{i=1}^N \overline{v}_j^3 v_{ik}/N) = plim(\overline{v}_j^3 \Sigma_{i=1}^N v_{iK}/N) = plim \overline{v}_j \cdot plim(\Sigma v_{ik}/N)$ et donc

$$plim(\Sigma_{i=1}^{N} v_{ik}/N) = E_{N\to\infty}\Sigma(v_{ik}/N) = {}_{N\to\infty}\Sigma(E(v_{ik})/N) = 0$$

$$plim(\Sigma_{i=1}^{N} \overline{v}_j^3 v_{iK}/N) = 0$$

7) $3plim(\Sigma_{i=1}^{N} \tilde{x}_{ij}^2 v_{ij} v_{ik}/N) = 3E_{N\to\infty}(\Sigma_{i=1}^{N} \tilde{x}^2 v_{ij} v_{ik}/N) = 3_{N\to\infty}(\Sigma\sigma_{\tilde{x}_j}^2 \mu_{iK}(1,1)/N)$

$$= 3\sigma_{\tilde{x}_j}^2 \mu_{jK}(1,1)$$

8) $3plim(\Sigma_{i=1}^{N} \tilde{x}_{iK} v_{ij}^2 v_{iK}/N) = -3plim(\Sigma_{i=1}^{N} (v_{ij}^2 \overline{v}_j v_{iK}/N) = 0$ parce que $E(v_{ij}^2 v_{iK}) = \mu_{jK}(2,1) = 0$ (moments impairs d'une loi normale sont égaux à zéro).

9)

$$3plim(\Sigma_{i=1}^{N} \tilde{x}_{ij} \overline{v}_j^2 v_{iK}/N) = 3plim(\overline{v}_j^2 \cdot \Sigma_{i=1}^{N} \tilde{x}_{ij} v_{iK}/N) = 3plim\overline{v}_j^2 \cdot plim(\Sigma_{i=1}^{N} \tilde{x}_{ij} v_{iK}$$

et $plim(\Sigma_{i=1}^{N} \tilde{x}_{ij} v_{iK}/N) = E_{N\to\infty}(\Sigma_{i=1}^{N} \tilde{x}_{ij} v_{iK}/N) = 0$ car $E(v_{iK}) = 0 = E(\tilde{x}_{ij})$

10)

$$3plim(\Sigma_{i=1}^{N} v_{ij} \overline{v}_j^2 v_{iK}/N) = 3plim(\overline{v}_j^2) \cdot plim(\Sigma_{i=1}^{N} v_{ij} v_{iK}/N)$$

$$= 3plim(\overline{v}_j) \cdot plim(\overline{v}_j) \cdot plim(\Sigma_{i=1}^{N} v_{ij} v_{iK}/N) = 3 \cdot 0 \cdot 0 \; plim(\Sigma v_{ij} v_{iK}/N) = 0$$

car $plim(\overline{v}_j) = plim(\Sigma_{i=1}^{N} v_{ij}/N) = E_{N\to\infty}(\Sigma_{i=1}^{N} v_{ij}/N) = 0$ et $E(v_{ij}) = 0$.

11)

$$-6plim(\Sigma_{i=1}^{N}\tilde{x}_{ij}v_{ij}\overline{v}_{j}v_{iK}/N= -6plim(\overline{v}_{j})\cdot plim(\Sigma_{i=1}^{N}v_{ij}v_{iK}/N)=0$$

car $plim(\overline{v}_j)=0$.

On a aussi pour un élément représentatif (j,K) que:

$$plim(x_j v_k/N)=plim(\Sigma_{i=1}^{N}(\tilde{x}_{ij}+v_{ij})\,v_{iK}/N)=plim(\Sigma_{i=1}^{N}(\tilde{x}_{ij}v_{iK}+v_{ij}v_{iK}/N)$$

$$=E_{N\to\infty}(\Sigma_{i=1}^{N}(\tilde{x}_{ij}v_{iK}+v_{ij}v_{iK})/N)=\mu_{iK}(1,1)$$

Et que pour un terme représentatif (j,j) de la matrice diagonale $D(x'x/N)$ on a que:

$$plim(x_j'x_j/N)=plim(\Sigma_{i=1}^{N}(\tilde{x}_{ij}+v_{ij})^2/N)=plim(\Sigma_{i=1}^{N}(\tilde{x}_{ij}^2+2\tilde{x}_{ij}v_{ij}+v_{ij}^2)/n)$$

$$=plim(\Sigma_{i=1}^{N}x_{ij}^2/N)+plim(\Sigma_{i=1}^{N}2v_{ij}\tilde{x}_{ij}^2/N)+plim(\Sigma_{i=1}^{N}v_{ij}^2/N)$$

$$=E_{N\to\infty}(\Sigma_{i=1}^{N}\tilde{x}_{ij}^2/N)+E_{N\to\infty}(2\tilde{x}_{ij}v_{ij}/N)+E_{N\to\infty}(\Sigma_{i=1}^{N}(v_{ij}^2/N))$$

$$=\sigma_{\tilde{x}_j}^2+0+\sigma_{v_j}^2$$

donc

$$-plim(z_3'v/N)+3Dplim(x'x/N)\;\cdot plim(x'u/N)=$$

$$=-\mu_{jK}(3,1)\;-3\sigma_{\tilde{x}_j}\mu_{jK}(1,1)\;+3(\sigma_{v_j}^2+\sigma_{\tilde{x}_j}^2)\,\mu_{jk}(1,1)$$

$$=-u_{jK}(3,1)-3\sigma_{\tilde{x}_j}^2\mu_{jK}(1,1)+3\sigma_{v_j}^2\mu_{jK}(1,1)+3\sigma_{\tilde{x}_j}^2\mu_{jK}(1,1)$$

$$=\mu_{jK}(3,1)\;+3\sigma_{v_j}^2\mu_{jK}(1,1)$$

Sachant que les v_{iK} suivent une loi normale, alors (ici on utilise les propriétés de la loi normale)

$$\mu_{jK}(3,1)=E(v_{ij}^3 v_{iK}) \quad =3E(v_{ij}^2)\cdot E(v_{ij}v_{iK})$$

$$=3\sigma_{v_j}^2\cdot\mu_{jK}(1,1) \quad \text{donc } 3\sigma_{v_j}^2\mu_{jK}(1,1)-3\sigma_{v_j}^2\mu_{jK}(1,1)=0, \quad (j=1,K=1,\ldots,N).$$

Enfin, pour montrer que β_H et β_E sont convergents

$$\beta_H=(C'S^{-1}C)^{-1}C'S^{-1}\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}=W\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$$

$$\beta_E=(C'\hat{S}^{*-1}C)^{-1}C'\hat{S}^{*-1}\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$$

il faut calculer le $plim\beta_H$ et le $plim\beta_E$ soit:

$$plim\beta_H=[C'(plim(S/N))^{-1}C]^{-1}C'[plim(S/N)]^{-1}plim\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix}$$

où $plim\begin{pmatrix}\beta_D\\ \beta_P\end{pmatrix} =\begin{pmatrix}plim\beta_D\\ plim\beta_P\end{pmatrix}= \begin{pmatrix}\beta\\ \beta\end{pmatrix}=C\beta$ et où $C=\begin{pmatrix}I_K\\ I_K\end{pmatrix}$ donc

$$plim\beta_H=[C'(plim(S/N))^{-1}C]^{-1}C'(plim(S/N))^{-1}C\beta=\beta \text{ et}$$

$$plim\beta_E=[C'(plim(\hat{S}^*/N))^{-1}C]^{-1}C'(plim(\hat{S}^*/N))^{-1}C\beta=\beta\ .$$

c) On a supposé, au CHAPITRE 2, que les erreurs de régression et les erreurs de mesure sur les variables explicatives sont normalement distribuées. Cela peut sembler à priori une hypothèse peu raisonnable mais si on étudie plus en profondeur la composition des erreurs de mesure des variables explicatives et les erreurs de régression, alors il devient plus raisonnable de faire de telles suppositions si on considères ceci:

$Y=\alpha i+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+u$ où u représente, par exemple, les variables explicatives omises alors si on a omis :

$u=\beta_4 X_4+...+\beta_N X_N$, donc u est une somme de facteurs omis où chaqu'un des facteurs n'est pas nécessairement normalement distribué mais la somme des facteurs, (on a de bonnes raisons de croire que le nombres facteurs tend vers l'infini) par le théorème central limite, convergera en distribution vers une loi normale soit:

$N^{-1/2}\sum_{i=1}^{N}(X_i\beta_i-E(X_i\beta_i))\xrightarrow{d} N(0,\sigma_u^2 I)$. Maintenant pour les erreurs de mesure sur les variables explicatives, l'explication est la même en ce qui concerne l'intuition sous-jacente. Les v_i sont des sommes de facteurs omis et par le théorème central limite chaque vecteur v_i est normalement distribué ce qui implique que l'ensemble des vecteurs v_i est normalement distribué (Johnston (1985), Judge (1985)).

d) On calcule la "Valeur-P" associée à m comme suit:

$Pr(\tilde{m} > m)=Pr[\eta'[JG^{-1}J-m[I-x(X'X)^{-1}X']/(N-K-1)]\eta]>0$ où $\tilde{m}=(u'J'G^{-1}Ju/\sigma^2)/(\hat{\sigma}^2/\sigma^2)$. Alternativement $\tilde{m}$ peut être vu comme: $\tilde{m}=((\eta'J'G^{-1}J\eta)/[\eta'(I-X(X'X)^{-1}X')\eta/(N-K-1)])$ où e_L est exprimé en terme de u et η. η est une variable normale réduite: $\eta=u/\sigma$.

Enfin, cette probabilité $(Pr(\tilde{m}>m))$ peut se calculer par la procédure d'Imhof. (Imhof, P.J. (1961))