L'impasse de la non-normalité : estimateurs d'ordres supérieurs pour modèles de régression linéaire avec erreurs de mesure

J. Sébastien Blais
Département de sciences économiques
Université de Montréal

Résumé

La construction d'estimateurs convergents pour les modèles de régression linéaire avec erreurs sur les variables fait l'objet de ce rapport de recherche. Nous présentons les résultats de simulations exposant les performances de l'estimateur d'ordres supérieurs (EOS) proposé récemment par Dagenais et Dagenais (1997) lorsque l'hypothèse de normalité de erreurs de mesure n'est pas vérifiée. Nous montrons par ailleurs que l'estimateur ne peut en général s'étendre à d'autres distributions. Quoiqu'il soit non convergent, l'estimateur original se comporte très bien en échantillon fini sous des erreurs suivant une loi t de Student multivariée.

Ce rapport de recherche fut réalisé sous la direction de Marcel G. Dagenais avec le support financier du C.R.D.E.

Table des Matières

Liste des tableaux

I. Introduction

Bien qu'il soit souvent difficile d'en apprécier quantitativement l'importance, il est généralement admis qu'aucune donnée n'est totalement exempte d'erreurs de mesure. Malgré l'intérêt porté aux problèmes des erreurs sur les variables dès les années 1920 en statistique, les économètres semblent les avoir négligés par la suite. D'abord sans doute parce que, en économétrie plus qu'en physique où des modèles exacts étaient souvent déjà connus, les biais engendrés par une mauvaise spécification des équations sont d'un ordre nettement supérieur (Malinvaud, 1981). Aujourd'hui, alors que la théorie économique s'est raffinée, ces problèmes prennent une importance relative moindre. De plus, les méthodes d'abord proposées pour traiter ces maux n'étaient certainement pas d'application aisée. Les biais peuvent de fait généralement être corrigés par l'utilisation de variables instrumentales, quoique peu de chercheurs le fassent en pratique. On se contente souvent d'aviser le lecteur de l'existence possible de telles erreurs, sans toutefois la tester formellement, et d'en rappeler brièvement les conséquences. Ce manquement provient, au moins partiellement, de l'absence d'outils facilement accessibles. Il est en effet habituellement coûteux et peu pratique de recueillir des variables supplémentaires, satisfaisant les conditions distributionnelles requises, pour parfaire nos résultats. Dans d'autres cas, l'exercice n'est tout simplement pas possible. Nous aurions besoin d'un outil qui ne nécessite aucune autre collecte de données.

Dagenais et Dagenais (1997) (DD) proposent un tel outil, inspiré des travaux de Geary (1942), Drion (1951) et Durbin (1954), sous l'hypothèse d'erreurs normalement distribuées. Si l'erreur sur la variable est une somme de plusieurs composantes vérifiants les conditions usuelles nécessaires à l'application du théorème limite central, l'hypothèse de normalité est alors bien fondée. Par ailleurs, si l'erreur est le fruit de l'utilisation d'une variable approximative (*proxies*), l'économètre peut difficilement justifier cette hypothèse. Nous étudions ici l'importance de cette hypothèse de normalité.

La section II fait un bref survol de la littérature et introduit la problématique de l'étude. La section III présente le modèle et détaille comment l'estimateur proposé par DD repose directement sur une propriété particulière de la loi normale qui prévient toute généralisation de l'estimateur à d'autres distributions en général, et à la distribution t de Student multivariée en particulier. La section IV est consacrée aux résultats de simulations exposant les performances de leur estimateur original lorsque les erreurs de mesure sont distribuées selon une loi t de Student multivariée avec différents degrés de liberté. La section V conclut.

II. Revue de la littérature

Les premières études traitant des erreurs sur les variables remontent aux années 1920 et portaient sur les conditions d'identification des paramètres d'une relation linéaire entre deux variables lorsque celles-ci sont mesurées avec erreur[1]. Gini (1921) montre en effet que les paramètres d'une relation linaire entre deux variables mesurées avec erreur ne peuvent être identifiés et ne peuvent donc être estimés sans une connaissance *a priori* de la variance d'au moins une des erreurs ou du ratio de ces variances. Lindley (1947) généralise ce résultat au cas multivarié et présente des estimateurs lorsque cette information est disponible. Lorsqu'une telle information n'est pas disponible, la technique bien connue des variables instrumentales, proposée d'abord par Reiersol (1941), permet d'obtenir des estimateurs convergents. Évidemment, le chercheur doit avoir accès à ces variables, hautement corrélées aux régresseurs et non corrélées aux erreurs de mesure. Geary (1942) propose d'utiliser l'asymétrie de la densité des régresseurs[2] pour construire des variables instrumentales qui conduisent à des estimateurs convergents, quoique peu efficaces, des paramètres d'une régression. Il propose aussi l'utilisation du quatrième moment empirique dans le cas univarié. Pal (1980) montre dans une couverture exhaustive de cette littérature que les estimateurs de Drion (1951) et Durbin (1954) sont aussi dérivés de cette condition d'asymétrie. Intuitivement, on peut présenter les fondements de cette approche de la manière suivante :

$$f(X) \text{ non symétrique } \leftrightarrow E\left(X^3\right) \neq 0 \leftrightarrow \operatorname{cov}\left(X^2, X\right) \neq 0 .$$

X^2 constitue donc une variable instrumentale possible pour X puisqu'elle est de surcroît, par nos hypothèses, non corrélée à l'erreur de mesure.

Ces problèmes d'identification impliquent que l'usage de la méthode des moindres carrés ordinaires n'est pas approprié. En effet, dès qu'un seul régresseur est mesuré avec erreur et que celui-ci est corrélé aux autres régresseurs, l'estimateur MCO de chacun des paramètres sera biaisé. Contrairement au modèle de régression simple où le biais est toujours de signe opposé à celui de la

---

[1] Madansky (1959) présente une revue complète.

[2] Une condition suffisante à la non-normalité. Reiersol (1950) montre que, si les erreurs sont distribuées normalement, les paramètres ne peuvent être identifiés si et seulement si les variables X et Y sont aussi normales ou fixes.

vraie valeur du coefficient et de valeur absolue moindre, aucune information aisément interprétable sur ce biais n'est disponible, sinon qu'il est généralement décroissant en la corrélation des régresseurs et croissant en la variance de l'erreur de mesure[3].

Notons ici qu'un plus grand échantillon ne sera d'aucun apport. En effet, le biais restera inchangé alors que la variance de l'estimateur sera réduite entraînant des conséquences désastreuses en ce qui concerne la taille effective des tests sur les paramètres.

Dagenais et Dagenais proposent une généralisation de la méthode introduite par Geary et illustrent par simulations de Monte Carlo l'ampleur des problèmes associés à l'utilisation de l'estimateur MCO. Ils montrent que, dans le cas multivarié, l'utilisation conjointe des troisièmes et quatrièmes moments des variables explicatives et de la variable dépendante comme variables instrumentales est nettement plus efficace que ne le suggérait la littérature. Sous une hypothèse de normalité des erreurs de mesure, ils montrent l'absence de corrélation entre ces dernières et leurs instruments. La section suivante expose comment la convergence de l'estimateur proposé par DD repose de façon directe sur cette hypothèse de normalité et montre qu'il ne sera pas possible de le généraliser à d'autres distributions.

III. Analyse théorique

Nous présentons ici, dans un premier temps, le cadre théorique de l'étude ainsi que l'instrument original proposé, puis nous montrons comment l'hypothèse de normalité est essentielle.

1. Le modèle

Soit le modèle :

$$Y = \alpha\iota_N + \widetilde{X}\beta + u \qquad (1.1)$$

où $\widetilde{X}$ est une matrice N×K de variables exogènes stochastiques qu'on ne peut observer directement telle que

$$E\left(\frac{\widetilde{X}'\widetilde{X}}{N}\right) = \operatorname*{plim}_{N\to\infty}\left(\frac{\widetilde{X}'\widetilde{X}}{N}\right) = Q \qquad (1.2)$$

où Q est non singulière et finie. u est un vecteur N×1 d'erreurs indépendantes de $\widetilde{X}$, de matrice de covariance $\sigma^2 I_N$, et Y est le vecteur N×1 des observations de la variable dépendante.

[3] Ces affirmations sont illustrées en annexe.

Le vecteur K×1 β et les scalaires α et σ^2 sont les paramètres à estimer. Ce que l'économètre observe en pratique est $X = \widetilde{X} + V$, où V est une matrice N×K d'erreurs aléatoires de densité symétrique. De plus, nous supposons que V est non corrélée avec u et que $\mathrm{var}[vec(V)] = \Sigma \otimes I_N$ où la matrice Σ de dimension K×K est symétrique définie positive. Cette dernière hypothèse permet une corrélation entre les variables mais préserve l'indépendance entre les observations. Le modèle peut alors s'écrire :

$$Y = \alpha\iota_N + (X - V)\beta + u = \alpha\iota_N + X\beta + (u - V\beta) = \alpha\iota_N + X\beta + \eta . \qquad (1.3)$$

La variable instrumentale proposée est $Z = (\iota_N, z_1, z_2, z_3, z_4, z_5, z_6, z_7)$ où

$$z_1 = x * x \qquad (1.4)$$

$$z_2 = x * y \qquad (1.5)$$

$$z_3 = y * y \qquad (1.6)$$

$$z_4 = x * x * x - 3x[E(x'x/N) * I_K] \qquad (1.7)$$

$$z_5 = x * x * y\iota_N' - 2x[E(x'y\iota_N'/N) * I_K] - y\{\iota_K'[E(x'x/N) * I_K] \qquad (1.8)$$

$$z_6 = x * y\iota_K' * y\iota_K' - x[E(y'y/N)] - 2y[E(y'x/N)] \qquad (1.9)$$

$$z_7 = y * y * y - 3y[E(y'y/N)] \qquad (1.10)$$

avec x et y représentant respectivement X et Y en déviation de leur moyenne et $*$ le produit élément-par-élément d'Hardamard. Les sept éléments qui composent la variable instrumentale seront ci-après appelés les *instruments*.

2. La généralisation

DD indiquent que l'estimateur pourrait, sous certaines hypothèses, être généralisé pour d'autres distributions d'erreurs symétriques, telle une loi t de Student multivariée, en modifiant la constante 3 apparaissant dans l'équation *(1.7)* par une autre constante propre à la nouvelle distribution. Nous montrons ici qu'en général cette constante est en fait une matrice diagonale de dimension K et qu'alors cette généralisation n'est pas possible.

Remplaçons d'abord l'équation *(1.7)* comme suggéré.

$$z_4 = x * x * x - x[E(x'x/N) * C] \qquad (2.1)$$

où C est la matrice diagonale de dimension K propre à notre nouvelle distribution symétrique.

Nous devons montrer que $\underset{N \to \infty}{E}\left(\frac{z_4'\eta}{\sqrt{N}}\right) = 0$. (2.2)

Un élément typique de $E\left(\frac{z_4'\eta}{\sqrt{N}}\right)$ est

$$E\left(\frac{z_{4j}'\eta}{\sqrt{N}}\right) = E\left[\sum_{i=1}^{N}\left(\frac{x_{ij}^3\eta_i}{\sqrt{N}}\right) - C_j\sum_{i=1}^{N}\left(\frac{c_{jj}x_{ij}\eta_i}{\sqrt{N}}\right)\right] \tag{2.3}$$

$$\text{où } c_{jj} = E\left[\sum_{i=1}^{N}\left(\frac{x_{ij}^2}{N}\right)\right] \quad = E\left[\sum_{i=1}^{N}\tilde{x}_{ij}^2\right] + \sigma_{v_j}^2 - \frac{\sigma_{v_j}^2}{N}$$

$$= E\left[\sum_{i=1}^{N}\tilde{x}_{ij}^2\right] + \sigma_{v_j}^2 - O\left(N^{-1}\right) \tag{2.4}$$

et C_j l'élément de la $j^{\text{ième}}$ colonne de la matrice C.

Nous avons donc

$$E\left[\sum_{i=1}^{N}\left(\frac{x_{ij}^3\eta_i}{\sqrt{N}}\right) - C_j\sum_{i=1}^{N}\left(\frac{c_{jj}x_{ij}\eta_i}{\sqrt{N}}\right)\right]$$

$$= \left(1/\sqrt{N}\right)E\left[\sum_{i=1}^{N}\tilde{x}_{ij}^3\eta_i + 3\sum_{i=1}^{N}\tilde{x}_{ij}^2 v_{ij}\eta_i + 3\sum_{i=1}^{N}\tilde{x}_{ij}v_{ij}^2\eta_i + \sum_{i=1}^{N}v_{ij}^3\eta_i - C_j c_{jj}\left(\sum_{i=1}^{N}\tilde{x}_{ij}\eta_i + \sum_{i=1}^{N}v_{ij}\eta_i\right)\right] \tag{2.5}$$

$$= \left(1/\sqrt{N}\right)E\left[3\sum_{i=1}^{N}\tilde{x}_{ij}^2 v_{ij}\eta_i + \sum_{i=1}^{N}v_{ij}^3\eta_i - C_j c_{jj}\sum_{i=1}^{N}v_{ij}\eta_i\right] \tag{2.6}$$

puisque les autres termes sont égaux à zéro par nos hypothèses d'indépendance. Ensuite,

$$\left(1/\sqrt{N}\right)E\left[3\sum_{i=1}^{N}\tilde{x}_{ij}^2 v_{ij}\eta_i + \sum_{i=1}^{N}v_{ij}^3\eta_i - C_j c_{jj}\sum_{i=1}^{N}v_{ij}\eta_i\right]$$

$$= 3\left(1/\sqrt{N}\right)E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right) - \sqrt{N}\sum_{l=1}^{K}E\left(v_{ij}^3 V_{il}\beta_l\right)$$

$$+ C_j\sqrt{N}E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2/N\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right) + C_j\sqrt{N}\sigma_{v_j}^2\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right) + O\left(N^{-1/2}\right) \tag{2.7}$$

Ce terme est égal à zéro pour

$$C_j = \frac{3\left(1/\sqrt{N}\right)E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)+\sqrt{N}\sum_{l=1}^{K}E\left(v_{ij}^3V_{il}\beta_l\right)}{\sqrt{N}E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2/N\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)+\sqrt{N}\sigma_{v_j}^2\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)} \tag{2.8}$$

Dans le cas où V est multinormale, on a l'égalité (Kendall and Stuart, 1963, p. 91)

$$\mu_{31} = 3\mu_{20}\mu_{11} \tag{2.9}$$

et l'équation *(2.8)* peut s'écrire

$$C_j = \frac{3\left(1/\sqrt{N}\right)E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)+3\sqrt{N}\sigma_{v_j}^2\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)+O\left(N^{-1/2}\right)}{\sqrt{N}E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2/N\right)\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)+\sqrt{N}\sigma_{v_j}^2\sum_{l=1}^{K}E\left(v_{ij}V_{il}\beta_l\right)}. \tag{2.10}$$

$$= \frac{3\left(1/\sqrt{N}\right)E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2\right)+3\sqrt{N}\sigma_{v_j}^2+O\left(N^{-1/2}\right)}{\sqrt{N}E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2/N\right)+\sqrt{N}\sigma_{v_j}^2} \xrightarrow{N\to\infty} 3.$$

Pour une distribution autre, la constante reliant μ_{31} à $\mu_{20}\mu_{11}$ sera généralement différente de 3. En particulier, cette constante est fonction des degrés de liberté pour une loi t de Student[4] :

$$\mu_{31} = \left(\frac{6}{\upsilon-4}+3\right)\mu_{20}\mu_{11} \tag{2.11}$$

Nous obtenons donc

$$C_j = \frac{3\left(1/\sqrt{N}\right)E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2\right)+\left(\frac{6}{\upsilon-4}+3\right)\sqrt{N}\sigma_{v_j}^2+O\left(N^{-1/2}\right)}{\sqrt{N}E\left(\sum_{i=1}^{N}\tilde{x}_{ij}^2/N\right)+\sqrt{N}\sigma_{v_j}^2}. \tag{2.12}$$

Cette constante dépend bien de la colonne j, par l'intermédiaire de la variance de v_j. Nous devrions donc estimer les K éléments distincts de la matrice diagonale C. Toutefois, puisque l'instrument z_4 n'ajoute que K équations, l'estimation de ces K paramètres additionnels le rend inutile.

[4] Hengartner offre l'expression de cette constante.

Dans la mesure où il ne semble pas possible de généraliser l'estimateur pour une distribution autre que la normale, il devient essentiel de quantifier les problèmes associés à son utilisation lorsque que l'hypothèse n'est pas vérifiée. Nous présentons dans la section qui suit les résultats de simulations qui indiquent que l'estimateur original performe relativement bien même lorsque les erreurs sont distribuées selon une loi t de Student multivariée. Nous présentons aussi les performances d'un estimateur excluant z_4, le seul instrument requérant l'hypothèse de normalité.

IV. Analyse empirique

Cadre de la simulation

Puisque les résultats présentés ne sont valides qu'asymptotiquement, nous comparons entre eux par simulations de Monte Carlo quatre estimateurs : l'estimateur des moindres carrées ordinaires (β_{MCO}), l'estimateur original de DD (β_Z), un estimateur constitué d'une constante et des premier et quatrième instruments ($\beta_{Z1,Z4}$), puis l'estimateur original auquel on a retiré le quatrième instrument ($\beta_{Z/Z4}$). Ce dernier estimateur est considéré puisque seul l'instrument z_4 prévient la convergence de l'estimateur original et qu'il est donc souhaitable d'examiner son apport de manière individuelle.

Des erreurs distribuées selon une loi t de Student pour différents degrés de liberté sont générées pour différentes tailles d'échantillons. Le choix de la distribution multivariée t de Student s'impose comme généralisation symétrique de la loi normale puisqu'elle l'inclut comme cas particulier avec une infinité de degrés de liberté. Ainsi, s'il appert que l'estimateur original performe mal avec une distribution si près de la normale, il nous sera permis de croire que ses déficiences ne seront que plus importantes pour une distribution quelconque. L'objectif poursuivi ici est d'étudier le comportement de l'estimateur lorsque la distribution des erreurs de mesure dévie de la normale d'une manière particulière : préservant la symétrie tout en augmentant le kurtosis. Nous ne présentons aucune justification théorique ou empirique de ce choix et la distribution t de Student n'est utilisée qu'à titre d'exemple d'une telle loi.

Les estimateurs seront comparés selon trois critères : le biais, l'erreur quadratique moyenne (EQM) et la taille effective de l'erreur de type I des tests. L'EQM est bien sûr une mesure standard de performance d'un estimateur et n'a nul besoin d'autre justification. Le biais étant la problématique même de l'utilisation de l'estimateur MCO en présence d'erreurs sur les variables, nous voudrons nous y attarder de façon distincte. L'importance de la taille effective des tests de type I sur les paramètres, illustrée par DD, sensibilise à l'effet le plus pervers de l'utilisation aveugle de l'estimateur MCO.

Puisque la performance de l'estimateur est sensible à la distribution des régresseurs, nous utilisons des données réelles auxquelles nous ajoutons une erreur. Les données proviennent du recensement 1986 de Statistique Canada sur les finances des ménages. Nous considérons un modèle linéaire simple reliant les dépenses totales annuelles d'un ménage à son revenu total annuel, à l'âge du chef de famille et au nombre de personnes-semaines le constituant. Seuls les ménages de revenu total annuel se situant entre 25 000$ et 55 000$ furent conservés de sorte qu'une certaine homogénéité de l'échantillon soit assurée. Afin de rendre les résultats aisément interprétables, nous avons transformé les régresseurs pour obtenir $\beta = (1,1,1,1)'$ comme vecteur de coefficients. Pour ce faire, nous avons utilisé les coefficients d'une première régression sur le logarithme des régresseurs bruts. Cette régression nous permet aussi d'obtenir une valeur du coefficient de corrélation multiple, près de 0.40, valeur que nous utiliserons pour fixer la taille de l'erreur de régression.

Expérience 1.

Pour cette première simulation, première base de comparaison pour les expériences 2 et 3, nous fixons la taille de l'échantillon à 2000 observations et la variance de l'erreur sur la première variable de sorte que le ratio de cette variance à la variance du régresseur affecté $\left(\lambda_1 = \sigma^2_{V_1}/\sigma^2_{\tilde{X}_1}\right)$ soit égal à 0,3.

Malgré sa divergence, l'estimateur original se comporte de manière satisfaisante. Bien que de façon moins spectaculaire, il surclasse toujours l'estimateur MCO lorsque les erreurs suivent une loi t avec 5 degrés de liberté selon tous nos critères. Tel qu'attendu, ces performances s'améliorent lorsqu'on augmente les degrés de liberté, la distribution t se rapprochant ainsi de la distribution normale.

Nos simulations mettent aussi en évidence les faiblesses de l'estimateur $\beta_{Z1,Z4}$. Bien que DD suggéraient qu'il puisse être recommandable dans certaines situations, ce ne semble plus être le cas lorsque l'hypothèse de normalité n'est pas vérifiée. Ainsi, pour une distribution t à 5 degrés de liberté, ses performances seront inférieures à celle de l'estimateur MCO en termes d'EQM. Nous pourrions ici blâmer aveuglément l'instrument z_4 pour cette piètre performance mais les résultats de l'estimateur $\beta_{Z/Z4}$ nous rappellent à l'ordre. Bien qu'il permette de réduire les biais substantiellement, son EQM est toujours élevée et d'un ordre comparable à celui de l'estimateur MCO. On note cependant, en comparant les résultats de l'estimateur original, que l'instrument z_4 assure une réduction importante de l'EQM. En fait, il s'agit ici d'une illustration supplémentaire que l'utilisation conjointes des moments de troisième et quatrième ordres conduit à des estimateurs moins erratiques. Lorsque tous les autres instruments sont utilisés, l'ajout de l'instrument z_4 réduit la variance de l'estimateur et donne des erreurs de type 1 effectives près des valeurs théoriques. Remarquons aussi que l'erreur de type 1 effective de l'estimateur $\beta_{Z1,Z4}$ est largement supérieure à celle de l'estimateur β_Z lorsque les erreurs de mesure suivent une loi t à 5 degrés de liberté. En somme, l'estimateur original β_Z semble être le plus performant à tous les niveaux.

Tableau 1 : Expérience 1. N=2000, λ_1=0,3, R^2=0,40, 1000 échantillons.

A. Biais

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,944	0,942	0,947	0,930	0,938	0,942	-0,760	-0,260	-0,145	-0,091	-0,074	0,003
β_1	-0,237	-0,235	-0,236	-0,234	-0,235	-0,236	0,207	0,074	0,041	0,024	0,019	0,000
β_2	-0,038	-0,030	-0,025	-0,038	-0,035	-0,034	0,135	0,067	0,039	0,012	-0,001	0,009
β_3	0,089	0,086	0,087	0,091	0,092	0,088	-0,161	-0,065	-0,032	-0,014	-0,011	-0,004
Biais moyen	0,327	0,324	0,324	0,323	0,325	0,325	0,316	0,116	0,064	0,035	0,026	0,004

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,046	0,004	0,011	0,005	0,004	0,006	-0,068	-0,095	-0,046	-0,031	-0,023	0,020
β_1	-0,010	0,002	-0,001	0,003	0,000	0,002	0,018	0,027	0,013	0,008	0,005	-0,004
β_2	0,032	0,034	0,026	0,062	0,033	0,044	0,014	0,029	0,017	0,010	-0,004	0,013
β_3	0,015	-0,001	0,005	0,005	0,011	0,003	-0,010	-0,023	-0,005	0,003	0,004	0,002
Biais moyen	0,026	0,010	0,011	0,019	0,012	0,014	0,028	0,044	0,020	0,013	0,009	0,010

B. Erreur quadratique moyenne

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,952	0,951	0,955	0,938	0,946	0,951	1,242	0,503	0,410	0,377	0,370	0,336
β_1	0,239	0,237	0,238	0,235	0,237	0,238	0,338	0,137	0,110	0,100	0,097	0,087
β_2	0,114	0,114	0,111	0,116	0,116	0,118	0,318	0,247	0,231	0,220	0,229	0,220
β_3	0,118	0,117	0,117	0,121	0,123	0,120	0,291	0,172	0,149	0,145	0,151	0,145
EQM moyenne	0,356	0,355	0,355	0,352	0,356	0,357	0,547	0,265	0,225	0,210	0,212	0,197

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,524	0,464	0,463	0,445	0,453	0,464	0,414	0,345	0,320	0,309	0,315	0,302
β_1	0,127	0,114	0,108	0,104	0,105	0,107	0,110	0,092	0,083	0,080	0,081	0,077
β_2	0,573	0,532	0,586	0,563	0,557	0,588	0,226	0,219	0,218	0,210	0,219	0,215
β_3	0,192	0,191	0,185	0,181	0,183	0,186	0,153	0,148	0,139	0,138	0,143	0,141
EQM moyenne	0,354	0,325	0,336	0,323	0,325	0,336	0,225	0,201	0,190	0,184	0,189	0,184

C. Taille de l'erreur de type 1, en %

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	37,8%	10,6%	7,7%	6,7%	5,0%	5,9%
β_1	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	39,1%	12,4%	7,8%	6,6%	6,4%	5,9%
β_2	6,3%	5,9%	4,6%	6,5%	5,7%	7,1%	9,5%	5,9%	5,4%	4,3%	4,5%	5,0%
β_3	19,6%	18,5%	19,1%	20,5%	22,0%	19,9%	18,7%	9,0%	5,1%	4,8%	6,5%	5,8%

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	12,6%	7,9%	8,6%	8,1%	8,8%	9,0%	10,3%	6,5%	6,5%	6,4%	6,1%	6,0%
β_1	12,9%	9,7%	9,2%	8,0%	8,1%	8,8%	10,1%	8,1%	5,7%	6,3%	6,0%	5,9%
β_2	9,8%	8,4%	10,2%	9,7%	8,9%	9,6%	7,0%	6,6%	6,8%	6,1%	6,4%	5,8%
β_3	8,0%	7,1%	6,8%	6,2%	5,8%	6,7%	7,2%	7,0%	4,5%	5,4%	5,4%	6,4%

Ce tableau et les suivants pésentent les résultats de simulations où les erreurs de mesure suivent des lois t de Student à 5, 10, 15, 20 et 25 degrés de liberté et une loi normale. La partie C donne la taille effective de l'erreur de type 1 basée sur une valeur attendue de 5%.

Tableau 2 : Expérience 2. N=2000, λ_1=0,1, R^2=0,40, 1000 échantillons.

A. Biais

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,375	0,377	0,375	0,372	0,373	0,370	-0,151	-0,038	-0,025	-0,007	-0,009	-0,009
β_1	-0,093	-0,094	-0,094	-0,093	-0,093	-0,093	0,042	0,010	0,006	0,001	0,003	0,002
β_2	-0,008	-0,009	-0,012	-0,007	-0,012	-0,015	0,033	0,005	0,001	-0,002	0,010	-0,006
β_3	0,031	0,035	0,040	0,038	0,032	0,035	-0,040	-0,010	0,000	0,002	-0,006	-0,003
Biais moyen	0,127	0,129	0,130	0,127	0,127	0,128	0,067	0,016	0,008	0,003	0,007	0,005

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	-0,037	-0,026	-0,031	-0,016	-0,029	-0,017	-0,098	-0,032	-0,025	-0,013	-0,015	-0,014
β_1	0,011	0,009	0,010	0,009	0,011	0,009	0,027	0,009	0,007	0,003	0,005	0,003
β_2	0,022	0,025	0,037	0,060	0,039	0,050	0,021	0,013	0,011	0,007	0,014	0,001
β_3	-0,006	-0,003	0,002	-0,002	-0,009	-0,006	-0,025	-0,006	0,002	0,004	-0,005	-0,001
Biais moyen	0,019	0,016	0,020	0,022	0,022	0,021	0,043	0,015	0,011	0,007	0,010	0,005

B. Erreur quadratique moyenne

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,398	0,402	0,403	0,396	0,396	0,393	0,347	0,248	0,241	0,230	0,225	0,222
β_1	0,098	0,099	0,100	0,098	0,098	0,098	0,092	0,062	0,059	0,056	0,056	0,055
β_2	0,111	0,109	0,106	0,108	0,110	0,110	0,214	0,215	0,215	0,221	0,216	0,216
β_3	0,083	0,087	0,085	0,087	0,084	0,083	0,140	0,131	0,126	0,127	0,130	0,124
EQM moyenne	0,172	0,174	0,174	0,172	0,172	0,171	0,198	0,164	0,160	0,159	0,157	0,154

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,377	0,377	0,372	0,377	0,369	0,382	0,257	0,235	0,237	0,225	0,222	0,217
β_1	0,086	0,085	0,083	0,084	0,083	0,084	0,066	0,059	0,058	0,055	0,055	0,054
β_2	0,555	0,547	0,552	0,540	0,547	0,562	0,198	0,207	0,202	0,211	0,213	0,206
β_3	0,168	0,175	0,167	0,167	0,172	0,167	0,130	0,130	0,126	0,126	0,128	0,125
EQM moyenne	0,297	0,296	0,294	0,292	0,293	0,299	0,162	0,158	0,156	0,155	0,155	0,150

C. Taille de l'erreur de type 1, en %

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	78,5%	78,0%	75,7%	78,0%	77,1%	78,6%	10,5%	6,5%	5,8%	4,9%	4,0%	4,1%
β_1	83,9%	82,8%	81,6%	82,0%	83,0%	83,8%	11,6%	6,6%	5,7%	5,1%	4,1%	3,9%
β_2	5,1%	5,8%	5,0%	5,3%	6,1%	5,4%	4,8%	5,0%	5,5%	5,9%	5,1%	5,0%
β_3	7,4%	9,1%	8,1%	7,6%	7,6%	6,2%	5,4%	5,6%	5,3%	4,5%	5,4%	4,3%

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	8,4%	9,0%	8,8%	10,3%	7,4%	9,9%	6,5%	6,1%	6,7%	4,9%	4,7%	4,9%
β_1	10,8%	9,4%	8,6%	10,1%	9,1%	10,8%	7,4%	6,0%	6,7%	5,3%	4,6%	5,1%
β_2	11,2%	9,9%	10,8%	9,7%	10,5%	11,0%	5,5%	6,7%	6,0%	6,7%	8,7%	6,7%
β_3	6,5%	7,8%	6,0%	6,4%	7,5%	5,4%	5,4%	6,2%	5,4%	5,6%	6,2%	4,4%

Tableau 3 : Expérience 3. N=700, $\lambda 1=0,3$, $R^2=0,40$, 1000 échantillons.

A. Biais

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	1,025	1,028	1,024	1,030	1,015	1,025	-0,400	-0,196	-0,088	-0,046	-0,025	0,043
β_1	-0,234	-0,235	-0,234	-0,235	-0,233	-0,234	0,105	0,053	0,024	0,012	0,004	-0,011
β_2	-0,017	-0,021	-0,023	-0,014	-0,016	-0,021	0,096	0,056	0,020	0,018	-0,002	-0,015
β_3	0,115	0,113	0,109	0,118	0,130	0,112	-0,129	-0,075	-0,039	-0,007	0,024	0,007
Biais moyen	0,348	0,349	0,348	0,349	0,348	0,348	0,183	0,095	0,043	0,021	0,014	0,019

	$\beta_{Z,Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,082	0,057	0,065	0,050	0,012	0,043	-0,047	-0,035	0,019	0,019	0,005	0,066
β_1	-0,009	-0,002	-0,006	0,001	0,006	0,000	0,015	0,012	-0,003	-0,003	-0,002	-0,015
β_2	0,119	0,126	0,106	0,160	0,137	0,114	0,036	0,033	0,014	0,030	0,017	0,005
β_3	0,015	0,008	0,012	0,026	0,033	0,005	-0,018	-0,016	-0,002	0,014	0,029	0,016
Biais moyen	0,056	0,048	0,047	0,059	0,047	0,040	0,029	0,024	0,010	0,016	0,013	0,026

B. Erreur quadratique moyenne

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	1,050	1,055	1,050	1,055	1,041	1,051	1,123	0,738	0,642	0,619	0,594	0,579
β_1	0,239	0,240	0,239	0,240	0,239	0,240	0,292	0,186	0,161	0,153	0,146	0,142
β_2	0,165	0,168	0,161	0,165	0,173	0,170	0,411	0,328	0,327	0,325	0,320	0,328
β_3	0,206	0,211	0,202	0,203	0,220	0,203	0,473	0,380	0,362	0,333	0,350	0,357
EQM moyenne	0,415	0,418	0,413	0,416	0,418	0,416	0,575	0,408	0,373	0,357	0,352	0,351

	$\beta_{Z,Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,754	0,750	0,726	0,725	0,697	0,723	0,601	0,581	0,539	0,548	0,532	0,531
β_1	0,182	0,176	0,167	0,170	0,167	0,172	0,151	0,141	0,131	0,135	0,128	0,132
β_2	0,778	0,768	0,770	0,786	0,794	0,778	0,319	0,297	0,306	0,309	0,301	0,319
β_3	0,436	0,423	0,410	0,400	0,419	0,440	0,354	0,346	0,338	0,327	0,329	0,353
EQM moyenne	0,538	0,529	0,518	0,520	0,519	0,528	0,356	0,341	0,328	0,330	0,323	0,334

C. Taille de l'erreur de type 1, en %

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	99,3%	99,2%	99,1%	99,1%	99,2%	99,2%	9,4%	6,6%	5,7%	5,7%	5,3%	6,7%
β_1	99,9%	99,5%	99,8%	99,6%	99,6%	99,6%	10,3%	6,3%	6,1%	6,4%	5,6%	6,8%
β_2	4,5%	5,1%	4,8%	5,1%	5,8%	4,9%	6,8%	3,8%	4,9%	4,7%	4,6%	4,8%
β_3	10,1%	9,2%	8,0%	8,8%	11,1%	9,7%	6,8%	4,5%	6,0%	3,6%	4,9%	5,2%

	$\beta_{Z,Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	8,4%	9,1%	8,8%	10,1%	7,4%	8,6%	6,8%	8,1%	6,9%	8,4%	7,8%	7,5%
β_1	9,5%	10,1%	8,5%	9,3%	8,5%	10,3%	8,8%	8,4%	7,2%	8,4%	8,0%	9,6%
β_2	9,6%	8,7%	8,5%	9,7%	9,5%	8,2%	7,3%	4,5%	5,3%	6,8%	5,3%	7,0%
β_3	6,5%	7,8%	6,1%	4,9%	6,2%	8,0%	6,4%	6,6%	5,8%	5,1%	5,8%	8,1%

Tableau 4 : Expérience 4. N=2000, λ_1=0,3, λ_2=0,3, R^2=0,40, 1000 échantillons.

A. Biais

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,873	0,875	0,874	0,874	0,881	0,874	-0,643	-0,250	-0,114	-0,087	-0,060	0,034
β_1	-0,237	-0,238	-0,238	-0,237	-0,239	-0,237	0,192	0,080	0,039	0,029	0,021	-0,010
β_2	-0,254	-0,259	-0,255	-0,259	-0,254	-0,248	0,301	0,185	0,103	0,070	0,050	-0,011
β_3	0,094	0,096	0,096	0,091	0,096	0,098	-0,155	-0,063	-0,037	-0,031	-0,028	0,009
Biais moyen	0,365	0,367	0,366	0,365	0,368	0,364	0,322	0,145	0,073	0,054	0,040	0,016

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,024	-0,023	-0,034	-0,015	-0,025	-0,013	-0,032	-0,090	-0,049	-0,028	-0,017	0,033
β_1	-0,013	0,003	0,006	0,002	0,004	0,004	0,012	0,027	0,016	0,009	0,006	-0,010
β_2	-0,058	-0,013	-0,015	-0,010	-0,016	0,018	0,048	0,059	0,040	0,021	0,010	-0,008
β_3	0,026	0,021	0,009	0,011	0,003	0,008	-0,011	-0,015	-0,014	-0,009	-0,013	0,012
Biais moyen	0,030	0,015	0,016	0,009	0,012	0,011	0,026	0,048	0,030	0,017	0,012	0,016

B. Erreur quadratique moyenne

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,882	0,884	0,882	0,884	0,889	0,884	1,405	0,561	0,442	0,417	0,391	0,367
β_1	0,239	0,240	0,239	0,239	0,241	0,239	0,389	0,152	0,115	0,103	0,097	0,089
β_2	0,271	0,276	0,273	0,277	0,272	0,268	1,408	0,547	0,483	0,451	0,433	0,394
β_3	0,122	0,125	0,123	0,120	0,123	0,123	0,336	0,183	0,166	0,149	0,151	0,142
EQM moyenne	0,378	0,381	0,379	0,380	0,381	0,379	0,885	0,361	0,301	0,280	0,268	0,248

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,530	0,505	0,481	0,475	0,461	0,481	0,445	0,375	0,346	0,341	0,336	0,335
β_1	0,124	0,115	0,112	0,110	0,108	0,111	0,105	0,094	0,085	0,082	0,081	0,081
β_2	0,650	0,617	0,626	0,627	0,609	0,616	0,552	0,394	0,391	0,386	0,376	0,340
β_3	0,195	0,197	0,189	0,189	0,195	0,184	0,156	0,154	0,148	0,141	0,144	0,137
EQM moyenne	0,375	0,359	0,352	0,350	0,343	0,348	0,315	0,254	0,243	0,237	0,234	0,223

C. Taille de l'erreur de type 1, en %

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	23,9%	10,4%	7,1%	6,8%	5,7%	6,8%
β_1	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	31,6%	13,4%	8,2%	8,1%	6,3%	6,4%
β_2	74,0%	75,9%	75,3%	75,1%	74,7%	72,5%	12,9%	6,0%	6,8%	6,0%	4,5%	4,4%
β_3	21,1%	22,3%	22,0%	20,2%	22,9%	20,6%	17,2%	8,8%	6,8%	4,9%	5,0%	4,8%

	$\beta_{Z/Z4}$						β_Z					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	11,2%	11,5%	9,1%	9,3%	7,3%	9,2%	9,8%	7,9%	6,1%	5,4%	6,0%	7,4%
β_1	12,3%	10,9%	9,4%	9,7%	8,4%	8,6%	7,8%	8,0%	5,8%	4,8%	5,7%	7,3%
β_2	11,4%	9,5%	9,3%	9,2%	9,6%	7,5%	13,2%	7,4%	8,9%	10,0%	8,6%	7,4%
β_3	6,6%	7,4%	6,4%	5,5%	6,7%	5,4%	5,9%	7,0%	6,1%	5,1%	5,2%	4,9%

Tableau 5 : Expérience 5. N=2000, λ_1=0,3, λ_2=0,3, ρ_{12}=0,2, R^2=0,40, 1000 échantillons.

A. Biais

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,880	0,877	0,885	0,876	0,883	0,881	-0,620	-0,234	-0,147	-0,087	-0,059	0,001
β_1	-0,248	-0,248	-0,250	-0,248	-0,249	-0,249	0,191	0,083	0,047	0,028	0,018	-0,002
β_2	-0,371	-0,369	-0,369	-0,370	-0,369	-0,368	0,357	0,261	0,106	0,069	0,027	-0,016
β_3	0,101	0,104	0,107	0,105	0,103	0,106	-0,154	-0,075	-0,038	-0,019	-0,018	0,007
Biais moyen	0,400	0,400	0,403	0,400	0,401	0,401	0,330	0,163	0,085	0,050	0,031	0,006

	$\beta_{Z/Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,004	0,016	-0,036	-0,022	-0,007	-0,018	-0,014	-0,061	-0,058	-0,025	-0,012	0,018
β_1	-0,016	-0,009	0,003	0,000	0,000	0,000	0,002	0,022	0,015	0,007	0,004	-0,008
β_2	-0,135	-0,028	-0,053	-0,033	-0,002	-0,032	-0,010	0,074	0,011	0,017	0,006	-0,026
β_3	0,044	0,027	0,018	0,027	0,017	0,022	0,005	-0,016	-0,006	0,005	-0,002	0,016
Biais moyen	0,050	0,020	0,027	0,021	0,006	0,018	0,008	0,043	0,023	0,013	0,006	0,017

B. Erreur quadratique moyenne

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,888	0,887	0,894	0,886	0,893	0,891	1,389	0,543	0,452	0,416	0,405	0,371
β_1	0,250	0,250	0,252	0,250	0,251	0,251	0,394	0,150	0,119	0,106	0,100	0,093
β_2	0,383	0,381	0,382	0,383	0,382	0,380	1,600	0,642	0,465	0,447	0,431	0,402
β_3	0,130	0,131	0,131	0,131	0,130	0,132	0,356	0,187	0,157	0,150	0,146	0,151
EQM moyenne	0,413	0,412	0,415	0,413	0,414	0,413	0,935	0,380	0,298	0,280	0,270	0,254

	$\beta_{Z/Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	0,546	0,479	0,483	0,474	0,472	0,469	0,439	0,360	0,356	0,344	0,347	0,329
β_1	0,128	0,112	0,111	0,110	0,109	0,107	0,106	0,089	0,088	0,083	0,085	0,081
β_2	0,684	0,625	0,630	0,629	0,629	0,627	0,583	0,416	0,365	0,349	0,366	0,344
β_3	0,199	0,199	0,188	0,186	0,197	0,189	0,157	0,154	0,141	0,143	0,140	0,145
EQM moyenne	0,389	0,354	0,353	0,350	0,352	0,348	0,321	0,255	0,237	0,230	0,234	0,225

C. Taille de l'erreur de type 1, en %

	β_{MCO}						$\beta_{Z1,Z4}$					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	24,4%	7,7%	7,3%	6,8%	6,8%	6,0%
β_1	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	31,5%	9,5%	9,1%	6,2%	6,4%	7,9%
β_2	97,1%	96,7%	96,4%	96,4%	96,8%	97,1%	17,6%	9,4%	4,1%	5,3%	5,3%	5,4%
β_3	24,2%	26,4%	26,6%	27,5%	24,4%	26,6%	15,7%	6,7%	5,0%	4,6%	3,6%	5,6%

	$\beta_{Z/Z4}$						β_{Z}					
	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale	t_5	t_{10}	t_{15}	t_{20}	t_{25}	normale
α	12,6%	8,3%	8,2%	7,6%	7,6%	8,1%	8,7%	4,6%	6,7%	5,8%	5,6%	5,9%
β_1	11,5%	8,8%	9,6%	8,4%	8,5%	8,0%	9,1%	4,5%	6,9%	6,4%	6,3%	6,3%
β_2	13,6%	9,2%	9,3%	8,5%	10,0%	9,2%	15,2%	8,6%	6,8%	6,7%	8,2%	7,2%
β_3	7,6%	6,8%	6,2%	5,0%	6,9%	5,4%	7,1%	6,2%	4,9%	4,5%	3,7%	5,3%

Expérience 2.

Dans cette expérience l'importance de l'erreur est réduite pour obtenir un ratio de 0,1 et les caractéristiques observées précédemment sont confirmées. L'estimateur sans la variable z_4 permet un biais moyen inférieur et l'instrument utilisant l'ensemble des variables procure les meilleures performances. En fait, seul cet instrument performe mieux que l'estimateur des MCO selon tous nos critères sous toutes les distributions considérées. Il est à noter toutefois que l'estimateur $\beta_{Z1,Z4}$ donne les erreurs de type 1 effectives les plus près des valeurs attendues de 5% lorsque la variance de l'erreur est ainsi réduite.

Expérience 3.

Le ratio est ramené à 0,3 et l'échantillon est ici réduit à 700 observations. Cet échantillon fut généré de manière indépendante au premier et il est donc nécessaire d'exercer une certaine prudence lorsque l'on désire comparer entre eux les résultats de ces expériences.

Curieusement, une réduction de la taille de l'échantillon révèle que l'estimateur $\beta_{Z/Z4}$ perd son avantage pour la correction des biais. L'estimateur original permet les meilleures performances.

Expérience 4.

Pour une seconde base de comparaison, nous ajoutons ici une erreur sur le second régresseur, de même variance mais indépendante de la première. Cette simulation renforce les conclusions de la première : l'estimateur $\beta_{Z/Z4}$ donne les biais minimum, seul l'estimateur β_Z surclasse l'estimateur β_{MCO} dans tous les cas et donne les plus petites erreurs de type 1 lorsque la distribution de erreurs de mesure est très leptokurtotique.

Expérience 5.

Nous introduisons maintenant une corrélation des erreurs de mesure de 20%. Ceci donne des résultats surprenants : l'estimateur original β_Z surclasse tous les autres selon tous les critères et toutes les distributions considérées.

En somme, comme DD l'avaient suggéré, un certain sous-ensemble des instruments permet un meilleur ajustement lorsque l'erreur de mesure est distribuée normalement. D'autre part, nos résultats montrent aussi que la totalité des instruments devrait être utilisée si les erreurs suivent une loi t de Student multivariée.

Il est important de rappeler que nous avons présenté les résultats de simulations où l'on a testé le comportement d'un estimateur construit sur une hypothèse de normalité lorsque cette dernière n'est pas vérifiée d'une façon très particulière. La distribution t de Student étant sous plusieurs angles similaire à la normale, une distribution plus éloignée pourrait engendrer des résultats moins satisfaisants.

V. Conclusion

Un regard attentif sur la construction de l'EOS nous montre qu'il n'est généralement pas possible de l'étendre à des distributions autres que la normale. En effet, la convergence de l'estimateur repose sur une propriété très particulière de la distribution normale, propriété qui n'est généralement pas vérifiée par d'autre distributions[5], dont la distribution t de Student. Si l'EOS est utilisé lorsque les erreurs de mesures suivent une autre loi, il sera biaisé.

D'autre part, des simulations de Monte Carlo suggèrent que ses performance en échantillon fini sont tout à fait respectables. Dans tous les cas considérés, l'EOS produit une erreur quadratique moyenne inférieure à celle des moindres carrés ordinaires même pour une distribution aussi leptokurtique qu'une distribution t de Student avec 5 degrés de liberté. En fait, retirer le seul instrument non convergent de l'EOS affecte négativement ses performances. L'importance de ce quatrième instrument, celui qui utilise le quatrième moment des régresseurs, suggère que l'ajout d'autres moments à l'EOS pourrait le raffiner davantage.

D'autres simulations seraient utiles. D'une part, le caractère quelque peu erratique de nos résultats, lorsqu'on les compare entre eux dans une même expérience pour différents degrés de liberté, suggèrent qu'un plus grand nombre d'échantillons serait nécessaire. D'autre part, d'autres distributions d'erreurs devraient être simulées. Par exemple, une erreur d'arrondissement de distribution uniforme, loi de kurtosis inférieur à la normale, est certainement un cas d'intérêt empirique important. Dans un contexte où le régresseur et son erreur sont des prix, cette erreur pourrait aussi suivre une loi asymétrique. L'instrument z_1 requiert la symétrie de la distribution de l'erreur de mesure.

[5] Il n'a pu être démontré que seule la loi normale possédait cette propriété, quoiqu'aucun contre-exemple n'ait pu être identifié.

Il est utile de remarquer que, en regard de nos résultats et de ceux obtenus par DD, les performances de l'EOS, et conséquemment des instruments qu'on y inclut, dépendent de la distribution conjointe des régresseurs, des erreurs et du régressand.

Annexe A : Biais de l'estimateur MCO

On peut aisément montrer (Johnston, 1984) que la limite en probabilité de l'écart entre l'estimateur MCO et la vraie valeur du paramètre s'exprime :

$$\underset{N\to\infty}{\text{plim}}\left(\beta_{MCO}-\beta\right)=\left(Q+\Omega\right)^{-1}\Omega\beta$$

où $\Omega=\underset{N\to\infty}{\text{plim}}\left(\frac{V'V}{N}\right)$ et $Q=\underset{N\to\infty}{\text{plim}}\left(\frac{\tilde{X}'\tilde{X}}{N}\right)$. Cette expression étant peu révélatrice d'un point de vue intuitif, nous présentons ici les biais sur des coefficients normalisés à l'unité pour différentes valeurs de Q et Ω.

Tableau A.1. Illustration du biais de l'estimateur des moindres carrés ordinaire.

Ω : \ Q :	1 0 0	1 0.1 0.1	1 0.2 0.2	1 0.2 0.1	1 0.1 0.2
	0 1 0	0.1 1 0.1	0.2 1 0.2	0.2 1 0.2	0.1 1 0.1
	0 0 1	0.1 0.1 1	0.2 0.2 1	0.1 0.2 1	0.2 0.1 1
0.1 0 0	0.090909091	0.092436975	0.096774194	0.094674556	0.094918504
0 0 0	0	-0.0084033613	-0.016129032	-0.017751479	-0.0076701822
0 0 0	0	-0.0084033613	-0.016129032	-0.0059171598	-0.018216683
0.1 0.01 0.01	0.10714286	0.090909091	0.078947368	0.085934986	0.083850932
0.01 0.1 0.01	0.10714286	0.090909091	0.078947368	0.076279369	0.092320723
0.01 0.01 0.1	0.10714286	0.090909091	0.078947368	0.085934986	0.083850932
0.1 0.02 0.02	0.12280702	0.104477610	0.090909091	0.098939929	0.096402473
0.02 0.1 0.02	0.12280702	0.104477610	0.090909091	0.087696756	0.106239460
0.02 0.02 0.1	0.12280702	0.104477610	0.090909091	0.098939929	0.096402473
0.1 0.02 0.01	0.11489920	0.096912917	0.083469722	0.090909091	0.089364113
0.02 0.1 0.02	0.12309457	0.106128090	0.093884839	0.090909091	0.107775100
0.01 0.02 0.1	0.11489920	0.096912917	0.083469722	0.090909091	0.089364113
0.1 0.01 0.02	0.11511609	0.098497496	0.086376302	0.093954379	0.090909091
0.01 0.1 0.01	0.10699789	0.089391410	0.076110867	0.073217419	0.090909091
0.02 0.01 0.1	0.11511609	0.098497496	0.086376302	0.093954379	0.090909091

Bibliographie

Dagenais, M.G. and D.L. Dagenais, "Higher moment estimators for linear regression models with errors in the variables", *Journal of Econometrics* 76, 1997, 190-221.

Drion, E.F., "Estimation of the parameters of a straight line and of the variance of the variables, if they are both subject to error", *Indagationes Mathematicae* 13, 1951, 256-260.

Durbin, J., "Errors in the variables", *International Statistical Review* 22, 1954, 23-32.

Geary, R.C., "Inherent relations between random variables", *Proceedings of the Royal Irish Academy* 47, 1942, 63-76.

Gini, C., "Sull'interpolazione di una retta quando I valori della variabile indipendente sono affetti da errori accidentali", *Metron* 1, Nos. 3, 1921, 63-82.

Hengartner, O., "Ausgewählte zwer und dreidimensionale Prüfverteilungen", Part 2, *Metrika* 9, 1965, 105-148.

Madansky, A., "The fitting of straight lines when both are subject to error", *American Statistical Association Journal*, March 1959, 173-205.

Lindley, D.V., "Regression lines and the linear functional relationship", *Journal of the Royal Statistical Society* 9 (Supplement), Nos.1-2, 1947, 218-244.

Malinvaud, E., *Méthodes statistiques de l'économétrie*, Bordua, Paris, 1981.

Pal, M., "Consistent moment estimators of regression coefficients in the presence of errors in the variables", *Journal of Econometrics* 14, 1980, 349-364.

Reiersol, O., "Confluence analysis by means of lag moments and other methods of confluence analysis", *Econometrica* 9, 1941, 1-24.

Reiersol, O., "Identifiability of a linear relation between variables which are subject to error", *Econometrica* 18, 1950, 375-389.