HAL
open science

Cours d'économétrie

Francois Bonnieux

▶ **To cite this version:**

Francois Bonnieux. Cours d'économétrie. Economie et sociologie rurales (Econométrie), 1971, 105 p. hal-02858594

HAL Id: hal-02858594

https://hal.inrae.fr/hal-02858594v1

Submitted on 8 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

386775

ECOLE NATIONALE SUPERIEURE AGRONOMIQUE

DE RENNES

Section spécialisée en Economie et Sociologie Rurales.

I. N. R. A. - RENNES
ÉCONOMIE RURALE
BIBLIOTHÈQUE

Cours d'Econométrie

par

F. BONNIEUX
Assistant à l'I.N.R.A.
(Station d'Economie Rurale de Rennes)

Année 1970-71.

- Avertissement -

Ce document présente les notes qui ont servi de base à un cours d'econométrie à l'E.N.S.A.R. Pour leconstruire, nous avons essentiellement utilisé le livre de JOHNSTON (*). Les travaux pratiques et les lectures d'articles qui l'ont accompagné donneront ultérieurement lieu à une publication.

L'exposé traite d'abord des modèles à une équation, puis des modèles à équations multiples. Sa lecture suppose une bonne connaissance des éléments de Probabilités, Statistique Mathématique et Algèbre linéaire. Il est nécéssaire d'avoir à l'esprit la théorie de l'estimation et celle de la régression généralisée.

En ce qui concerne des compléments, nous renvoyons à :

MALINVAUD E., 1969 - Méthodes statistiques de l'économétrie. Dunod, Paris.

où tous les points que nous abordons sont développés.

La bibliographie de base peut être complètée par deux ouvrages :

CHRIST C. F., 1966 - Econometric models and methods. Wiley, New York.

GOLDBERGER A. S., 1964 - Econometric theory. Wiley, New-York.

(*) JOHNSTON J., 1963, Econometric methods - Mc Graw Hill, London.

Chapitre 1

MODELES LINEAIRES A ERREURS SUR LES VARIABLES

I. N. R. A. - RENNES
ÉCONOMIE RURALE
BIBLIOTHÈQUE

Nous avons admis que les variables intervenant dans le modèle de la régression étaient observées sans erreur. Cette hypothèse est souvent justifiée en fait. Car, si les données statistiques sont généralement sujettes à de nombreuses imperfections, leur précision est néanmoins suffisante pour l'estimation de relations qui ne sont pas exactes. Autrement dit, les erreurs qui affectent les équations sont le plus souvent prépondérantes vis-à-vis de celles qui peuvent affecter la mesure des variables.

Cependant, les variables d'un modèle ne sont pas toujours observées directement. On est souvent obligé de leur substituer d'autres variables en relation étroite, mais non rigoureuse, avec elles. Par exemple en vue d'estimer la relation entre le revenu du ménage et telle ou telle de leur consommation, on dispose des résultats d'une enquête indiquant, pour un échantillon restreint, et pour une certaine période, les dépenses totales du ménage et les dépenses pour la catégorie de consommation considérée. On retient alors souvent les dépenses totales comme une évaluation des revenus bien qu'il existe entre les unes et les autres une différence inconnue, parfois importante.

Les erreurs sur les variables sont donc tout d'abord des erreurs de mesure, elles sont dues aussi au fait que l'on travaille sur un échantillon restreint, mais la source d'erreur la plus importante provient de ce que l'on n'observe pas directement les variables de la théorie.

Quels sont les effets de ces erreurs sur les estimations des paramètres du modèle linéaire ? Ce modèle reste-t-il valable ? Sinon quel modèle doit-on adopter ? C'est là les questions auxquelles nous allons essayer d'apporter une réponse.

1 - Le cas du modèle linéaire à deux variables

Soit pour illustrer le cas du modèle linéaire à deux variables. Supposons que :

$$(1) \quad \boxed{X = \chi + u}$$

$$(2) \quad \boxed{Y = \Psi + v}$$

où X et Y sont les valeurs mesurées, χ et Ψ les vraies valeurs des variables considérées et où u et v désignent les erreurs d'observation. On suppose en outre que l'on a :

$$(3) \quad \boxed{\Psi = \alpha + \beta \chi}$$

relation fonctionnelle exacte entre les vraies valeurs. Analysons le modèle défini par les relations (1), (2) et (3), nous avons :

$$(4) \quad \boxed{Y = \alpha + \beta X + w}$$

avec $w = v - \beta u$.

Nous pouvons supposer que les erreurs u_t et v_t sont mutuellement indépendantes et suivent des distributions indépendantes de t, centrées de variances constantes, en outre indépendantes des vraies valeurs χ et Ψ . Les hypothèses nécessaires à l'utilisation d'une régression simple à partir de (4) ne sont cependant pas vérifiées, car w et X sont dépendantes. En effet calculons la covariance de X et w :

$$E\left\{w\left[X - E(X)\right]\right\} = E\left\{(v - \beta u)\, u\right\}$$

$$= - \beta \operatorname{Var}(u) \neq 0$$

car $E(v) = E(u) =$ donc $E(X) = \chi$

Il en résulte, que l'application des formules du modèle de la régression simple conduit à une estimation biaisée de α et β . De plus, les estimateurs sont asymptotiquement biaisés, ils ne sont donc pas corrects.* Montrons ce résultat pour l'estimateur de β obtenu par la méthode des moindres carrés, le désignant par b_n, nous avons :

$$b_n = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

$$= \frac{\sum(\chi - \overline{\chi})(\Psi - \overline{\Psi}) + \sum(\chi - \overline{\chi})(v - \overline{v}) + \sum(\Psi - \overline{\Psi})(u - \overline{u}) + \sum(u - \overline{u})(v - \overline{v})}{\sum(\chi - \overline{\chi})^2 + 2\sum(\chi - \overline{\chi})(u - \overline{u}) + \sum(u - \overline{u})^2}$$

Introduisons les hypothèses suivantes : les erreurs sont indépendantes les unes des autres et des vraies valeurs.

$$p \lim b_n = \frac{\beta \sum(\chi - \overline{\chi})^2}{\sum(\chi - \overline{\chi})^2 + \sum(u - \overline{u})^2}$$

comme $\hat{\beta} = \dfrac{\sum(\chi - \overline{\chi})^2 (\Psi - \overline{\Psi})^2}{\sum(\chi - \overline{\chi})^2}$ on obtient :

$$(5) \qquad p \lim b_n = \frac{\beta}{1 + \sigma_u^2 / \sigma_\chi^2}$$

Donc $\boxed{p \lim b_n \leqslant \beta}$, on a donc une sous-estimation de β . Si la variance de l'erreur est 10 % de celle de la vraie valeur on aboutit par une application directe des moindres carrés à une sous-estimation de 10 % de β .

---

* Un estimateur t_n de θ est correct si :

$$\lim_{n \to \infty} P\,rob \left\{ | t_n - \theta | \geq \delta \right\} = 0$$

On dit encore que l'estimateur t_n converge en probabilité (ou stochastiquement) vers θ :

$$p \lim_{n \to \infty} t_n = \theta = st \lim_{n \to \infty} t_n = \theta$$

On peut étudier ce problème de trois façons :

(i) Approche classique qui suppose vérifiée des hypothèses fortes sur les distributions de probabilité des erreurs.

(ii) Développements basés sur un regroupement des observations et des hypothèses plus faibles sur les erreurs.

(i, i,i) Introduction de variables instrumentales.

2 - Approche classique

On peut considérer deux cas selon les hypothèses faites sur les vraies valeurs, les formules d'estimation sont cependant identiques.

Partons des relations (1), (2) et (3) :

$$\boxed{\begin{array}{ll} (1) & X = \chi + u \\ (2) & Y = \Psi + v \\ (3) & \Psi = \alpha + \beta X \end{array}}$$

et supposons que u et v sont des variables aléatoires normales, indépendantes, centrées et de variances σ_u^2 et σ_v^2 . La combinaison des trois relations de base conduit à :

$$X = \chi + u$$
$$Y = \alpha + \beta X + v$$

La fonction de vraisemblance d'un échantillon de taille n s'écrit :

$$L = \frac{1}{(2\pi\ \sigma_u^2)^{n/2}} e^{-\sum (X_i - \chi_i)^2 / 2\ \sigma_u^2} \quad \frac{1}{(2\pi\ \sigma_v^2)^{n/2}} e^{-\sum (Y_i - \alpha - \beta\ \chi_i)^2 / 2\sigma_v^2}$$

$$(6)\ \text{Log}\ L = L^* = \text{const} - \frac{n}{2} \text{Log}\ \sigma_u^2 - \frac{n}{2} \text{Log}\sigma_v^2 - \frac{1}{2\ \sigma_u^2} \sum (X_i - \chi_i)^2 - \frac{1}{2\sigma_v^2} \sum (Y_i - \alpha - \beta\ \chi_i)^2$$

Nous cherchons des estimateurs de α, β, σ_u^2 et σ_v^2. Mais les x_i étant inconnus, nous devons aussi les estimer. Calculons donc les dérivées partielles de L* :

$$\frac{\partial L^*}{\partial \sigma_u^2} = -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum (X_i - x_i)^2$$

$$\frac{\partial L^*}{\partial \sigma_v^2} = -\frac{n}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum (Y_i - \alpha - \beta x_i)^2$$

$$\frac{\partial L^*}{\partial x_i} = \frac{1}{\sigma_u^2}(X_i - x_i) + \frac{\beta}{\sigma_v^2}(Y_i - \alpha - \beta x_i) \quad (i = 1, 2, \ldots, n)$$

$$\frac{\partial L^*}{\partial \alpha} = \frac{1}{\sigma_v^2}\sum (Y_i - \alpha - \beta x_i)$$

$$\frac{\partial L^*}{\partial \beta} = \frac{1}{\sigma_v^2}\sum x_i (Y_i - \alpha - \beta x_i)$$

En annulant les deux premières dérivées partielles, nous obtenons :

(7) $$\tilde{\sigma}_u^2 = \frac{1}{n}\sum (X_i - \tilde{x}_i)^2$$

(8) $$\tilde{\sigma}_v^2 = \frac{1}{n}\sum (Y_i - \tilde{\alpha} - \tilde{\beta}\tilde{x}_i)^2$$

Considérons pour tout i, l'équation :

$$\frac{\partial L^*}{\partial x_i} = 0 \quad (i = 1, 2, \ldots, n)$$

(9) $$Y_i - \tilde{\alpha} - \tilde{\beta}\tilde{x}_i = -\frac{\tilde{\sigma}_v^2}{\tilde{\beta}\tilde{\sigma}_u^2}(X_i - \tilde{x}_i)$$

En égalant les carrés des deux memebres, puis en sommant pour i = 1, 2, ...n membres à membres, et enfin en divisant par n, on obtient ;

$$\tilde{\sigma}_v^2 = \frac{\tilde{\sigma}_v^4}{\tilde{\beta}^2\tilde{\sigma}_u^2}$$

c'est-à-dire :

$$\tilde{\beta}^2 = \frac{\tilde{\sigma}_v^{\,2}}{\tilde{\sigma}_u^{\,2}}$$

Relation vraie quelle que soit la taille de l'échantillon : nous savons que les estimateurs du maximum de vraisemblance sont asymptotiquement corrects, ce qui conduit à :

$$\beta^2 = \frac{\sigma_v^{\,2}}{\sigma_u^{\,2}}$$

relation qui n'a aucune raison d'être vraie, <u>donc la méthode du maximum de vraisemblance ne convient pas à cette application.</u>

Supposons qu'il existe un λ connu tel que :

$$(10) \qquad \boxed{\lambda = \frac{\sigma_u^{\,2}}{\sigma_v^{\,2}}}$$

Ce qui paraît possible, dans la mesure où l'on connait les sources d'erreurs.

La relation (9) peut alors s'écrire sous la forme :

$$(X_i - \tilde{X}_i) + \lambda\,\tilde{\beta}\,(Y_i - \tilde{\alpha} - \tilde{\beta}\,\tilde{X}_i) = 0$$

$$(11) \qquad \tilde{X}_i = \frac{X_i + \lambda\,\tilde{\beta}\,Y_i - \lambda\,\tilde{\alpha}\,\tilde{\beta}}{1 + \lambda\,\tilde{\beta}^2}$$

Annulons $\partial L^*/\partial\alpha$, on a :

$$(12) \qquad \tilde{\alpha} = \overline{Y} - \tilde{\beta}\,\overline{\tilde{X}}$$

où $\overline{\tilde{X}}$ désigne la moyenne des n estimateurs $\tilde{X}_i$. Puis en annulant $\partial L^*/\partial\beta$, on obtient :

$$\sum \tilde{X}_i\,(Y_i - \tilde{\alpha} - \tilde{\beta}\,\tilde{X}_i) = 0$$

qui s'écrit encore :

$$(13) \qquad \sum(\tilde{X}_i - \overline{\tilde{X}})\,(Y_i - \tilde{\alpha} - \tilde{\beta}\,\tilde{X}_i) = 0$$

En effet :

$$\sum \bar{\tilde{X}} \left(Y_i - \tilde{\alpha} - \tilde{\beta} \tilde{X}_i \right) = \bar{\tilde{X}} \sum (Y_i - \tilde{\alpha} - \tilde{\beta} \tilde{X}_i) = 0 \quad \text{car} \quad \partial L^* / \partial \alpha = 0$$

Substituons (12) dans (13), nous obtenons :

$$\tilde{\beta} = \frac{\sum (\tilde{X}_i - \bar{\tilde{X}}) (Y_i - \bar{Y})}{\sum (\tilde{X}_i - \bar{\tilde{X}})^2}$$

formule habituelle lorsqu'on calcule une régression simple de Y par rapport à $\tilde{X}$

La relation (11) peut s'écrire sous la forme :

$$\tilde{X}_i - \bar{\tilde{X}} = \frac{1}{1 + \lambda \tilde{\beta}^2} \left[(X_i - \bar{X}) + \lambda \tilde{\beta} (Y_i - \bar{Y}) \right]$$

Dé là :

$$\sum (\tilde{X}_i - \bar{\tilde{X}}) (Y_i - \bar{Y}) = \frac{1}{1 + \lambda \tilde{\beta}^2} \left[\sum (X_i - \bar{X}) (Y_i - \bar{Y}) + \lambda \tilde{\beta} \sum (Y_i - \bar{Y})^2 \right]$$

Et :

$$\sum (\tilde{X} - \bar{\tilde{X}})^2 = \frac{1}{(1 + \lambda \tilde{\beta}^2)^2} \left[\sum (X_i - \bar{X})^2 + 2 \lambda \tilde{\beta} \sum (X_i - \bar{X}) (Y_i - \bar{Y}) + \lambda^2 \tilde{\beta}^2 \sum (Y_i - \bar{Y})^2 \right]$$

Introduisons les notations :

$$m_{XY} = \frac{1}{n} \sum (X_i - \bar{X}) (Y_i - \bar{Y})$$

$$m_{XX} = \frac{1}{n} \sum (X_i - \bar{X})^2$$

$$m_{YY} = \frac{1}{n} \sum (Y_i - \bar{Y})^2$$

nous avons :

$$\tilde{\beta} = \frac{(1 + \lambda \tilde{\beta}^2) (m_{XY} + \lambda \tilde{\beta} m_{YY})}{m_{XX} + 2 \lambda \tilde{\beta} m_{XY} + \lambda^2 \tilde{\beta}^2 m_{YY}}$$

Forme quadratique en $\tilde{\beta}$ qui s'écrit :

(14) $$\boxed{\lambda\tilde{\beta}\, m_{XY} - \tilde{\beta}(\lambda\, m_{YY} - m_{XX}) - m_{XY} = 0}$$

et qui a toujours deux racines distinctes : (*)

(15) $$\boxed{\tilde{\beta}_1 = \theta + \sqrt{\theta^2 + \frac{1}{\lambda}} \qquad \tilde{\beta}_2 = \theta - \sqrt{\theta^2 + \frac{1}{\lambda}}}$$

où

(16) $$\boxed{\theta = \frac{m_{YY} - (1/\lambda)\, m_{XX}}{2\, m_{XY}}}$$

θ est calculée à partir des valeurs observées et de λ. Il apparaît que quelque soit le signe de θ , $\tilde{\beta}_1$ est positif tandis que $\tilde{\beta}_2$ est négatif. Comme $\tilde{\beta}$ mesure la pente de la droite de regression de Ψ sur χ , on choisira comme estimé $\tilde{\beta}_1$ lorsque la covariance entre valeurs observées c'est-à-dire m_{XY} sera positive et $\tilde{\beta}_2$ dans le cas contraire. Ainsi (15) fournit un estimateur de $\tilde{\beta}$ et (12) un estimateur de α :

(12) $$\tilde{\alpha} = \overline{Y} - \tilde{\beta}\,\overline{\tilde{\chi}}$$

Rappelons la relation (11) :

(11) $$\tilde{\chi}_i = \frac{X_i + \lambda\tilde{\beta}\, Y_i - \lambda\,\tilde{\alpha}\tilde{\beta}}{1 + \lambda\,\tilde{\beta}^2}$$

(*) $A\tilde{\beta}^2 - B\tilde{\beta} - C = 0$

$\theta = -\frac{B}{2A} \qquad \frac{1}{\lambda} = -\frac{C}{A}$

$A = B^2 - 4\,AC = 4\,A^2\,\theta^2 + 4\,\frac{A^2}{\lambda}$

$\tilde{\beta}_1 = \frac{+B + \sqrt{\Delta}}{2\,A} = \theta + \sqrt{\theta^2 + \frac{1}{\lambda}}$

et on a : $\sqrt{\theta^2 + \frac{1}{\lambda}} > |\theta|$

qui entraîne :

$$\tilde{\bar{\chi}} = \frac{\bar{X} + \lambda \tilde{\beta} \bar{Y} - \lambda \tilde{\alpha} \tilde{\beta}}{1 + \lambda \tilde{\beta}^2} = \frac{\bar{X} + \lambda \tilde{\beta} \bar{Y} - \lambda \tilde{\beta} (\bar{Y} - \tilde{\beta} \tilde{\bar{\chi}})}{1 + \lambda \tilde{\beta}^2}$$

$$\boxed{\tilde{\bar{\chi}} = \bar{X}}$$

L'estimateur du maximum de vraisemblance de α est donc :

$$(17) \qquad \boxed{\tilde{\alpha} = \bar{Y} - \tilde{\beta} \bar{X}}$$

Considérons deux cas particuliers de ce modèle. Supposons tout d'abord que $\lambda = 0$, c'est-à-dire que $\sigma_u^2 = 0$. Dans ce cas χ est observé sans erreur :

$$X = \chi$$

et l'équation (14) se réduit à :

$$\tilde{\beta} = \frac{m_{XY}}{m_{XX}}$$

qui est la pente de la droite de régression de Y sur X. Soit maintenant le cas où $\lambda \to \infty$, ou encore $\sigma^2_v = 0$. Alors :

$$\tilde{\beta} = \frac{m_{YY}}{m_{XY}}$$

qui est l'inverse de la pente de la droite de régression de X sur Y.

Voyons maintenant s'il est possible de généraliser ce modèle en incluant un terme aléatoire dans la relation entre les vraies valeurs. Le modèle s'écrit alors :

$$\boxed{\begin{array}{ll} (1) & X_i = \chi_i + u_i \\ (2) & Y_i = \Psi_i + v_i \\ (3\text{-a}) & \Psi_i = \alpha + \beta \chi_i + \varepsilon_i \end{array}}$$

il peut encore s'écrire sous la forme :

$$X_i = \chi_i + u_i$$

$$Y_i = \alpha + \beta \chi_i + \varepsilon_i + v_i$$

Si l'on suppose que u, v, et ε sont independantes, centrées, suivent une distribution normale et ont pour variance σ^2_u, σ^2_v et σ^2_ε respectivement, alors il est facile de voir que les résultats précédents se généralisent en remplaçant v par $\varepsilon + v$. En effet $v + \varepsilon$ est centré, suit une distribution normale et sa variance est égale à $\sigma^2_v + \sigma^2_\varepsilon$. Si l'on peut alors obtenir en estimé à priori de :

$$\boxed{\lambda = \frac{\sigma^2_u}{\sigma^2_\varepsilon + \sigma^2_v}}$$

les estimations basées sur (15) s'appliquent. Il faut remarquer qu'il est beaucoup plus difficile d'estimer à priori λ ici que précédemment.

On peut approcher le problème par la méthode du maximum de vraisemblance en supposant que σ^2_u est connu (*). Le logarithme de la fonction de vraisemblance s'écrit :

$$(18) \qquad L^* = \text{const} - \frac{n}{2} \operatorname{Log} (\sigma^2_\varepsilon + \sigma^2_v) - \frac{1}{2\,\sigma^2_u} \sum (X_i - x_i)^2$$

$$- \frac{1}{2(\sigma^2_\varepsilon + \sigma^2_v)} \sum (Y_i - \alpha - \beta\, x_i)^2$$

On obtient en annulant les dérivées partielles :

$$(19) \qquad \tilde{\sigma}_\varepsilon{}^2 + \tilde{\sigma}^2_v = \frac{1}{n} \sum (Y_i - \tilde{\alpha} - \tilde{\beta}\, \tilde{x}_i)^2$$

$$\frac{1}{\sigma^2_u} (X_i - \tilde{x}_i) + \frac{\tilde{\beta}}{\tilde{\sigma}^2 + \tilde{\sigma}^2_v} (Y_i - \tilde{\alpha} - \tilde{\beta}\, \tilde{x}_i) = 0 \qquad (i = 1, 2, \ldots, n)$$

$$\sum (Y_i - \tilde{\alpha} - \tilde{\beta}\, \tilde{x}_i) = 0$$

$$\sum \tilde{x}_i (Y_i - \tilde{\alpha} - \tilde{\beta}\, \tilde{x}_i) = 0$$

Pour simplifier, posons $\sigma^2 = \sigma^2_\varepsilon + \sigma^2_v$ et estimons globalement cette somme. La deuxième équation du système (19) donne :

$$\tilde{\sigma}^2 (X_i - \tilde{x}_i) + \tilde{\beta}\, \sigma^2_u (Y_i - \tilde{\alpha} - \tilde{\beta}\, \tilde{x}_i) = 0$$

(*) On pose quelquefois que σ^2_v est aussi connu, ce qui permet d'estimer σ^2_ε Ici on estime simplement la somme $\sigma^2_v + \sigma^2_\varepsilon$

Qui donne en sommant pour tout i et en faisant la moyenne

$$\sigma^2(\overline{X}_i - \overline{\tilde{\chi}}_i) + \tilde{\beta}\,\sigma_u^2\,(\overline{Y} - \tilde{\alpha} - \tilde{\beta}\,\overline{\tilde{\chi}}) = 0$$

et par soustraction nous obtenons :

$$(20) \qquad \tilde{\chi}_i - \overline{\tilde{\chi}} = \frac{\tilde{\sigma}^2 x_i + \tilde{\beta}\,\sigma_u^2\, y_i}{\tilde{\sigma}^2 + \tilde{\beta}^2\,\sigma_u^2} \quad \text{où } x_i = X_i - \overline{X} \text{ et } y_i = Y_i - \overline{Y}$$

Les deux dernières équations du système (19) donne :

$$\tilde{\beta} = \frac{\sum y_i\,(\tilde{\chi}_i - \overline{\tilde{\chi}})}{\sum (\tilde{\chi}_i - \overline{\tilde{\chi}})^2}$$

et en utilisant (20) on obtient :

$$\tilde{\beta} = \frac{(\tilde{\sigma}^2 \sum x_i\, y_i + \tilde{\beta}\,\sigma_u^2 \sum y_i^2)\,(\tilde{\sigma}^2 + \tilde{\beta}^2\sigma_u^2)}{\tilde{\sigma}^4 \sum x_i^2 + \tilde{\beta}^2\,\sigma_u^4 \sum y_i^2 + 2\,\tilde{\beta}\tilde{\sigma}^2\sigma_u^2 \sum x_i\, y_i}$$

ce qui donne :

$$(21) \qquad \tilde{\beta}^2\,\sigma_u^2\, m_{YY} - \tilde{\beta}(\sigma_u^2\, m_{YY} - \tilde{\sigma}^2 m_{XX}) - \tilde{\sigma}^2\, m_{XY} = 0$$

L'estimation du maximum de vraisemblance de σ^2 s'écrit d'après la première équation du système (19) :

$$\tilde{\sigma}^2 = \frac{1}{n} \sum (Y_i - \tilde{\alpha} - \tilde{\beta}\,\tilde{\chi}_i)^2$$

$$\tilde{\sigma}^2 = \frac{1}{n}\left[\sum y_i^2 - \tilde{\beta} \sum y_i\,(\tilde{\chi}_i - \overline{\tilde{\chi}})\right]$$

$$= m_{YY} - \frac{\tilde{\beta}(\tilde{\sigma}^2 m_{XY} + \tilde{\beta}\,\sigma_u^2\, m_{YY})}{\tilde{\sigma}^2 + \tilde{\beta}^2\sigma_u^2}$$

$$(22) \qquad \tilde{\sigma}^2 = m_{YY} - \tilde{\beta} m_{XY} - \tilde{\beta}^2\sigma_u^2$$

En utilisant (22) dans (21), et en simplifiant on obtient :

$$(23) \qquad \boxed{a_o\tilde{\beta}^3 + a_1\,\tilde{\beta}^2 + a_2\,\tilde{\beta} + a_3 = 0}$$

où (24) $a_o = -\,\sigma_u^2\, m_{XX}$

$$a_1 = m_{XY}\,(2\,\sigma_u^2 - m_{XX})$$

$$a_2 = m_{XY}^{\,2} + m_{YY}\,(m_{XX} - \sigma_u^2)$$

$$a_3 = -\,m_{XY}\, m_{YY}$$

On remarque que si l'on fait $\sigma_u^2 = 0$, la relation (21) se réduit à :

$$\tilde{\beta} = \frac{m_{XY}}{m_{XX}}$$

pente de la droite de la régression de Y par rapport à X, et (22) donne :

$$\tilde{\sigma} = m_{YY} - \tilde{\beta}\, m_{XY}$$

formule habituelle pour estimer la variance résiduelle.

Considérons l'équation du troisième degré en $\tilde{\beta}$, ces coefficients se calculent à partir des observations. On choisit pour estimateur de β, la racine réelle qui maximise la fonction L^* définie par la relation (18).

On peut obtenir plus simplement des estimateurs corrects, à partir des relations (1), (2) et (3-a). Nous avons :

$$\frac{1}{n}\sum (X_i - \overline{X})^2 = \frac{1}{n}\sum(x_i - \overline{x})^2 + \frac{1}{n}\sum(u_i - \overline{u})^2 + \frac{2}{n}\sum(x_i - \overline{x})(u_i - \overline{u})$$

$$\frac{1}{n}\sum (Y_i - \overline{Y})^2 = \beta^2\frac{1}{n}\sum(x_i - \overline{x})^2 + \frac{1}{n}\sum(\varepsilon_i - \overline{\varepsilon})^2 + \frac{1}{n}\sum(v_i - \overline{v})^2 + \text{termes rectangles}$$

$$\frac{1}{n}\sum (Y_i - \overline{Y})(X_i - \overline{X}) = \beta\frac{1}{n}\sum(x_i - \overline{x})^2 + \text{termes rectangles}$$

Prenons les espérances des membres de ces relations :

$$(25) \qquad E\left\{\frac{1}{n}\sum (X_i - \overline{X})^2\right\} = \frac{1}{n}\sum (x_i - \overline{x})^2 + \frac{n-1}{n}\sigma_u^2$$

$$(26) \qquad E\left\{\frac{1}{n}\sum (Y_i - \overline{Y})^2\right\} = \beta^2\frac{1}{n}\sum (x_i - \overline{x})^2 + \frac{n-1}{n}(\sigma_\varepsilon^2 + \sigma_v^2)$$

$$(27) \qquad E\left\{\frac{1}{n}\sum (Y_i - \overline{Y})(X_i - \overline{X})\right\} = \beta\frac{1}{n}\sum(x_i - \overline{x})^2$$

en effet si l'on suppose les erreurs mutuellement indépendantes et indépendantes des vraies valeurs alors les espérances des termes rectangles sont nulles. On peut estimer les premiers membres de (25), (26) et (27) par m_{XX}, m_{YY} et m_{XY} respectivement. On obtient alors en supposant σ_u^2 connu :

$$(28) \qquad \boxed{\tilde{\tilde{\beta}} = \frac{m_{XY}}{m_{XX} - \frac{n-1}{n}\sigma_u^2}}$$

puis un estimateur pour $\sigma^2_\varepsilon + \sigma^2_v$

$$(29) \quad \boxed{\tilde{\tilde{\sigma}}^2 = \frac{n}{n-1}\left(m_{YY} - \frac{m^2_{XY}}{m_{XX} - \frac{n-1}{n}\sigma^2_u}\right)}$$

Ces estimateurs sont corrects, car m_{XX} , m_{YY} et m_{XY} convergent en probabilité vers les limites de leurs espérances. Finalement, on estime α par la relation habituelle :

$$\boxed{\tilde{\tilde{\alpha}} = \overline{Y} - \tilde{\tilde{\beta}}\,\overline{X}}$$

Supposons maintenant que χ est une variable aléatoire. L'approche précédente reste valable si et seulement si, l'ensemble des valeurs de $\times$ peut être répété dans plusieurs tirages successifs, ce qui est irréaliste. Aussi posons que χ , u et v sont indépendantes, centrées, normales de variances σ^2_χ , σ^2_u et σ^2_v . D'après :

$$\boxed{\begin{array}{ll} (1) & X = \chi + u \\ (2) & Y = \Psi + v \\ (3) & \Psi = \alpha + \beta\,\chi \end{array}}$$

Comme X et Y sont des formes linéaires de variables normales, leur distribution conjointe est normale à deux dimensions.

et

$$\boxed{\begin{array}{l} \sigma^2_X = \sigma^2_\chi + \sigma^2_u \\ \sigma^2_Y = \beta^2\,\sigma^2_\chi + \sigma^2_v \\ \sigma_{XY} = \beta\,\sigma^2_\chi \end{array}}$$

Comme X et Y sont normales les estimateurs du maximum de vraisemblance de σ^2_X , σ^2_Y et σ^2_{XY} sont m_{XX}, m_{YY} et m_{XY}. En reprenant la relation (10) comme hypothèse :

$$\lambda = \frac{\sigma^2_u}{\sigma^2_v}$$

nous obtenons trois équations :

$$(31) \quad \begin{cases} m_{XX} = \tilde{\sigma}^2_\chi + \lambda\,\tilde{\sigma}^2_v \\ m_{YY} = \tilde{\beta}^2\,\tilde{\sigma}^2_\chi + \tilde{\sigma}^2_v \\ m_{XY} = \tilde{\beta}\,\tilde{\sigma}^2_\chi \end{cases}$$

qui se résoud par rapport à $\tilde{\sigma}^2_\chi$, $\tilde{\sigma}^2_v$ et $\tilde{\beta}$, fournissant ainsi les estimateurs du maximum de vraisemblance de σ^2_χ, σ^2_v et β. Le système (31) donne l'équation du second degré en $\tilde{\beta}$ suivante :

$$\boxed{\lambda\tilde{\beta}^2 m_{XY} - \tilde{\beta}(\lambda\ m_{YY} - m_{XX}) - m_{XY} = 0}$$

qui est la même que celle obtenue en (14), pour le cas où les valeurs de χ sont fixées. Les relations (15), (16) et (17) s'appliquent donc :

Considérons pour finir le modèle

$$\boxed{\begin{aligned} X &= \chi + u \\ Y &= \psi + v \\ \psi &= \alpha + \beta\ \chi + \varepsilon \end{aligned}}$$

en supposant que χ, u, v et ε sont normales et indépendantes. Dans ce cas X et Y sont normales et nous avons :

$$(32) \quad \boxed{\begin{cases} \sigma^2_X = \sigma^2_\chi + \sigma^2_u \\ \sigma^2_Y = \beta^2\ \sigma^2_\chi + \sigma^2_\varepsilon + \sigma^2_v \\ \sigma_{XY} = \beta\ \sigma^2_\chi \end{cases}}$$

en remplaçant σ^2_X, σ^2_Y et σ_{XY} par leurs estimateurs du maximum de vraisemblance en supposant σ^2_u connu et en posant $\sigma^2 = \sigma^2_\varepsilon + \sigma^2_v$ nous obtenons :

$$(33) \quad \begin{cases} m_{XX} = \tilde{\sigma}^2_\chi + \sigma^2_u \\ m_{YY} = \tilde{\beta}^2\ \sigma^2_x + \tilde{\sigma}^2_v \\ m_{XY} = \tilde{\beta}\ \tilde{\sigma}^2 \end{cases}$$

qui donne pour estimateur du maximum de vraisemblance de β

$$(34) \quad \boxed{\tilde{\beta} = \frac{m_{XY}}{m_{XX} - \sigma^2_u}}$$

formule analogue à celle de la relation (28).

3 - Problèmes de prévision

X étant donné, posons nous le problème de la prévision de Y (ou Ψ). Pour fixer les idées, raisonnons sur l'exemple suivant ou Y mesure la consommation et X le revenu avec :

$$\Psi_t = \alpha + \beta X_{t-1} + \varepsilon_t$$

Connaissant X_n, on peut utiliser comme estimateur de Y_{n+1} la quantité définie par :

$$(35) \qquad \tilde{Y}_{n+1} = \tilde{\alpha} + \tilde{\beta} \; X_n$$

où $\tilde{\alpha}$ et $\tilde{\beta}$ sont obtenues par l'une ou l'autre des méthodes précédentes.

Calculons l'espérance de Y_{n+1} sachant X_n :

$$E\,(Y_{n+1} \,/\, X_n) = E\,(\alpha + \beta \; X_n + \varepsilon_{n+1} + v_{n+1} \,/\, X_n)$$
$$= \alpha + \beta \, E\,(X_n \,/\, X_n)$$

ce qui nécéssite de calculer $E\,(X_n \,/\, X_n)$. Si X et u sont normales indépendantes on a : (*)

$$E\,(X_n \,/\, X_n) = \frac{\mu_X \, \sigma_u^2 + X_n \, \sigma_X^2}{\sigma_X^2 + \sigma_u^2}$$

d'où

$$(36) \qquad E\,(Y_{n+1} \,/\, X_n) = \alpha + \beta \, \frac{\mu_X \, \sigma_u^2 + X_n \sigma_X^2}{\sigma_X^2 + \sigma_u^2}$$

on constate en comparant (35) et (36) que mis à part les fluctuations de $\tilde{\alpha}$ et $\tilde{\beta}$, $\tilde{Y}_{n+1}$ est biaisé sauf si $X_n = \mu_X$. Si $X_n > \mu_X$ le biais est positif, il est négatif si $X_n < \mu_X$.

(*) Si X et Y sont normales et indépendantes d'espérances μ_X et μ_Y (avec $\mu_Y = 0$) d'écarts types σ_X Et σ_Y, si on définit $Z = X + Y$ on montre que :

$$E\,(X/Z) = \frac{\mu_X \; \sigma_Y^2 + Z \; \sigma_X^2}{\sigma_X^2 + \sigma_Y^2}$$

Pour le montrer, on définit d'abord la distribution de Z, puis celle de X sachant Z en remarquant que :

$$f\,(X \,/\, Z) = \frac{f\,(X, Z)}{f\,(Z)}$$

Nous avons vu au premier paragraphe que l'application directe des moindres carrés aux valeurs de X et de Y conduisait à un estimateur b_n de β convergeant en probabilité ;

$$\underset{n \to \infty}{\text{st lim}}\, b_n = \frac{\beta}{1 + \sigma_u^2 / \sigma_X^2}$$

prenons alors pour prédicteur :

$$\hat{Y}_{n+1} = Y_n + b_n (X_n - \overline{X})$$

et étudions sa limite en probabilité ;

$$\underset{n \to \infty}{\text{st lim}}\, \hat{Y}_{n+1} = \alpha + \beta \mu_X + \frac{\beta}{1 + \sigma_u^2 / \sigma_X^2} (X_n - \mu_X)$$

$$= \alpha + \beta \frac{\mu_X \sigma_u^2 + X_n \sigma_X^2}{\sigma_X^2 + \sigma_u^2}$$

$$= E (Y_{n+1} / X_n)$$

Ainsi bien que l'application directe des moindres carrés ne convienne pas pour estimer le modèle, elle permet d'obtenir de bonnes prévisions.

4 - Regroupement des observations

Dans ce domaine Wald* et Bartlett* ont apporté la contribution la plus importante. Wald* pose que :

$$X = \chi + u$$
$$Y = \Psi + v$$
$$\Psi = \alpha + \beta\chi$$

il suppose que u et v sont sériellement et mutuellement indépendantes, que le nombre d'observations est pair: $n = 2m$. Supposons que les X_i soient rangés en ordre croissant :

$$X_1, X_2, \dots X_m, X_{m+1}, \dots X_n$$

et soit la série de Y_i correspondante ;

$$Y_1, Y_2, \dots Y_m, Y_{m+1}; \dots Y_n$$

---

* A. WALD, "The fitting of straight lines if both variables are subject to error" AMS, Vol 11, pp. 284-300, 1940

M. S. BARTLETT "The fitting of straight lines if both variables are subject to error"- Biometrics, vol 5, pp. 207-242, 1949.

On définit les moyennes :

$$\overline{X}_1 = \frac{1}{m} \sum_{1}^{m} X_i \qquad \overline{X}_2 = \frac{1}{m} \sum_{m+1}^{n} X_i$$

$$\overline{Y}_1 = \frac{1}{m} \sum_{1}^{m} Y_i \qquad \overline{Y}_2 = \frac{1}{m} \sum_{m+1}^{n} Y_i$$

Wald estime β et α par

(37) $$b = \frac{\overline{Y}_1 - \overline{Y}_2}{\overline{X}_1 - \overline{X}_2}$$

(38) $$a = \overline{Y} - b\,\overline{X}$$

Si l'hypothèse suivante est vérifiée : (*)

(39) $$\liminf_{n \to \infty} \left| \frac{1}{m} \sum_{i=1}^{m} X_i - \frac{1}{m} \sum_{i=m+1}^{n} X_i \right| > 0$$

Ces estimateurs sont corrects. Wald a déterminé des intervalles de confiance pour α et β. L'hypothèse (39) n'est pas vraie pour des variables normales

Bartlett a proposé quelques modifications :

$$\overline{X}_1 = \frac{1}{k} \sum_{1}^{k} X_i \qquad \overline{X}_2 = \frac{1}{k} \sum_{i=n-k+1}^{n} X_i$$

$$\overline{Y}_1 = \frac{1}{k} \sum_{1}^{k} Y_i \qquad \overline{Y}_2 = \frac{1}{k} \sum_{bn-k+1}^{n} Y_i$$

et les estimateurs

(40) $$b' = \frac{\overline{Y}_3 - \overline{Y}_1}{\overline{X}_3 - \overline{X}_1}$$

(41) $$a' = \overline{Y} - b'\,\overline{X}$$

Dans le cas particulier où les valeurs de X sont régulièrement espacées, la variance de b' est minimum pour $k = n/3$.

Ces méthodes ont le mérite de la simplicité et peuvent être généralisées(**).

---

* (Soit une suite (u_n) on définit $\liminf_{n\to\infty} u_n = \lim_{p\to\infty} (\inf_{n \geq p} u_n)$

(**) J. W. HOOPER H. THEIL, "The extension of Wald's method of fitting straight lines to multiple regression".

Rev. Intern. Statist. Inst. 1958

5 - Introduction de variables instrumentales

Reprenons les relations (1), (2) et (3) et posons :

(4) $Y = \alpha + \beta X + w$

où $w = v - \beta u$

nous avons vu que l'application de la méthode des moindres carrés à la relation (4) n'est pas satisfaisante.

Il est possible de trouver une autre variable Z qui est indépendante de u et v. Soit l'estimateur :

(42) $$\hat{\beta} = \frac{\sum y_i \; z_i}{\sum x_i \; z_i}$$

$$\hat{\beta} = \frac{\beta \sum x_i \; z_i + \sum z_i \; (w_i - \overline{w}_i)}{\sum x_i \; z_i}$$

Comme z est indépendante de u et de v, $\sum z_i \; (w_i - \overline{w})$ converge en probabilité vers zéro. Donc l'estimateur fournit par la relation (42) est correct.

Z est appelé variable instrumentale ; elle permet d'obtenir des estimateurs corrects. On peut généraliser au cas d'un nombre quelconque de variables. On choisira Z de telle sorte que sa corrélation avec X soit forte.

L'introduction de variables instrumentales pose trois problèmes :

- part importante d'arbitraire dans leur choix d'où il résulte des variations des estimateurs correspondants,
- difficultés pour vérifier l'indépendance entre erreurs et variables instrumentales,
- on insiste beaucoup sur le "consistency".

exemple (*)

Soit le problème de l'estimation des élasticités de consommation à partir des résultats d'enquêtes sur les budjets familiaux. Pour les ménages d'un échantillon de taille n, on connait les dépenses totales X et les dépenses Y effectuées sur le type d'articles considéré. Le modèle spécifie une relation linéaire entre Y et le revenu χ du ménage (entièrement dépensé) :

$$Y = \alpha + \beta \chi + w$$

(*) N. LIVIATAN "Errors in variables and Engel cuve analysis" Econometrica, juillet 1961.

Par ailleurs l'écart entre la dépense totale et le revenu est considéré comme bien décrit par une variable aléatoire centrée :

$$X = \chi + \varepsilon$$

on se trouve ainsi en face d'un modèle à erreurs sur les variables.

Il arrive que l'on dispose également du revenu déclaré Z ; c'est une évaluation médiocre du revenu réel qu'il sous estime. Une régression entre Y et Z conduirait à une mauvaise estimation du modèle, aussi introduit-on la variable instrumentale Z. En effet la corrélation revenu réel, revenu déclaré doit être élevée, (tout au moins pour les tranches moyennes de revenu) alors qu'elle est sans doute faible entre le revenu déclaré et ε. Car ce dernier est dû aux aléas dans le rythme des dépenses lesquelles jouent de la même manière à divers niveaux de revenu. D'où :

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})(Z_i - \overline{Z})}{\sum_{i=1}^{n} (X_i - \overline{X})(Z_i - \overline{Z})}$$

Chapitre 2 - AUTOCORRELATION

1 - Le cas du modèle linéaire à deux variables

Une des hypothèses introduite dans l'étude de la régression pose que les termes résiduels ne sont pas autocorrélés :

$$E\,(u_t\, u_{t+s} = 0 \quad \forall\, t,\ \forall s \neq 0.$$

c'est-à-dire en notations matricielles :

$$E\,(u\, u') = \sigma^2$$

Dans de nombreux cas il semble très hasardeux d'admettre cette hypothèse. Supposons, par exemple, que la relation entre la variable exogène et la variable endogène soit mal spécifiée. Nous avons posé que Y dépend linéairement de X, alors qu'elles sont liées par une forme du second degré. Même si les erreurs ne sont pas autocorrelées dans la relation exacte, le résidu de la relation linéaire contiendra un terme en X^2 . S'il existe une autocorrélation des X, les résidus sont aussi autocorrelés. Cet exemple est un cas particulier d'omission d'une variable explicative. En général on ne tient compte que de l'influence de certaines variables et les termes résiduels représentent aussi l'influence des variables omises. Toutefois une autocorrélation des variables omises n'implique pas nécéssairement une autocorrélation des erreurs, il peut se produire une compensation. Cependant si l'autocorrélation a lieu sur toutes les variables omises, et si celles ci se modifient en phase, on a toutes les chances d'avoir des erreurs autocorrelées. Ajoutons que l'autocorrélation des erreurs peut provenir des erreurs de mesure.

Considérons le cas du modèle

(1) $$\boxed{Y_t = \alpha + \beta X_t + u_t}$$

et supposons que les erreurs sont liées par la relation autorégressive :

(2) $$\boxed{u_t = \rho\, u_{t-1} + \varepsilon_t}$$

où $|\rho| < 1$ où ε_t satisfait aux hypothèses suivantes :

(3) $$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t\, \varepsilon_{t+s}) = \begin{cases} \sigma^2_\varepsilon & \forall\, t,\ s \quad t = s \\ 0 & \forall\, t,\ s \quad t \neq s \end{cases}$$

Nous avons alors ;

$$u_t = \rho\, u_{t-1} + \varepsilon_t$$

$$= \rho(\rho\, u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\ \varepsilon_{t-2} + \dots$$

(4) $$u_t = \sum_{r=0}^{\infty} \rho\varepsilon_{t-r}$$

et $E(U_t) = 0$ car $E(\varepsilon_t) = 0$

$$E(u_t^2) = E(\varepsilon_t{}^2) + \rho^2 E(\varepsilon^2_{t-1}) + \rho^4 E(\varepsilon^2_{t-2}) + \dots$$

car les ε ne sont pas liés et finalement :

$$E(u_t{}^2) = (1 + \rho^2 + \rho^4 + \dots)\sigma^2_\varepsilon$$

(5) $$\boxed{\sigma^2_u = \frac{\sigma^2_\varepsilon}{1 - \rho^2}}$$

Calculons :

$$E(u_t\, u_{t-1}) = E\left[(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots)(\varepsilon_{t-1} + \rho\, \varepsilon_{t-2} + \rho^2\varepsilon_{t-3} \dots)\right]$$

$$= E\left[(\varepsilon_t + \rho\,(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots)\right]\left[(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots)\right]$$

$$= \rho E\left[(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \dots)^2\right]$$

$$= \rho\sigma^2_u$$

et de façon générale :

$$(6) \qquad \boxed{E\,(u_t\ u_{t-s}) = \rho^s\ \sigma_u^2}$$

qui peut s'écrire sous la forme de :

$$\frac{E\,(u_t\ u_{t-s})}{E\,(u_t^2)} = \rho^s$$

Le membre de gauche de cette relation définit le coefficient d'autocorrélation de rang s de la série u_t.

Le schéma autorégressif définit par la relation (2) est le plus simple possible.

2 - Conséquences de la liaison entre erreurs

Supposons qu'il y ait une liaison entre les erreurs et que l'on applique les formules classiques de la régression, il en résulte un certain nombre de conséquences :

1° - On obtient des estimateurs biaisés de α et β dont les variances sont très grandes comparées à celles que l'on obtiendrait par des méthodes légèrement différentes.

2° - les variances sont sous-estimées, aussi les tests ne sont pas valables

3° - les prévisions sont innefficaces, car leurs variances sont trop élevées.

3 - Moindres carrés généralisés

Utilisons la notation matricielle :

$$(7) \qquad \boxed{Y = X\beta + u}$$

où

$$(8) \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \qquad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Nous supposons maintenant que l'on a :

$$\boxed{\begin{array}{ll} (9) & E(u) = 0 \\ (10) & E(u\,u') = V \end{array}}$$

V est donc la matrice des variances - covariances des termes résiduels, elle est carrée symétrique d'ordre n, et est régulière. Si par exemple l'erreur respecte le schéma autorégressif ;

$$(2) \qquad u_t = \rho u_{t-1} + \varepsilon_t$$

$$\boxed{V = \sigma^2_u \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}}$$

Définissons :

$$(11) \quad \boxed{\beta^* = A\,Y}$$

comme estimateur linéaire de β, A est donc une matrice rectangulaire k x n. Pour que β^* soit un estimateur sans biais il faut que :

$$E(AY) = \beta$$

$$\text{Mais } E(AY) = E\left[A(X\beta + u)\right]$$

$$= A\,X\,\beta \quad \text{car } E(u) = 0$$

et $E(AY) = \beta$ si et seulement si

(12) $\boxed{AX = I}$

La matrice des variances covariances des estimateurs β_i^* est donnée par $E\left[(\beta^* - \beta)(\beta^* - \beta)'\right]$, où les termes de la diagonale principale représentent les variances. Faisons la somme des variances, nous obtenons ;

$$E\left[(\beta^* - \beta)'(\beta^* - \beta)\right] = E\left[(Au)'(Au)\right]$$

car $\beta^* = AY = AX\beta + Au = \beta + Au$ d'après (12)

Ainsi $E\left[(\beta^* - \beta)'(\beta^* - \beta)\right] = E(u' A' Au)$

d'où (voir annexe) :

(13) $E\left[(\beta^* - \beta)'(\beta^* - \beta)\right] = tr\,(A'AV)$

La recherche d'un estimateur linéaire, sans biais, tel que la somme des variances soit minimum se ramène à :

Trouver une matrice A, telle que tr (A'AV) soit minimum et telle que la condition $AX = I$ soit vérifiée. Nous introduisons k^2 multiplicateurs de Lagrange μ_{ij} $(i, j = 1, \ldots k)$ et le problème se ramène à la minimisation de :

$$\phi = tr\,(A'AV) - \mu_{11}\left(\sum_{i=1}^{n} a_{1i} X_{1i} - 1\right) - \mu_{12}\left(\sum_{i=1}^{n} a_{1i} X_{2i}\right) \ldots$$

$$- \mu_{1k}\left(\sum_{i=1}^{n} a_{1i} X_{ki}\right) - \mu_{21}\left(\sum_{i=1}^{n} a_{2i} X_{1i}\right) - \mu_{22}\left(\sum_{i=1}^{n} a_{2i} X_{2i} - 1\right) \ldots$$

$$- \mu_{2k}\left(\sum_{i=1}^{n} a_{2i} X_{ki}\right)$$

$$- \mu_{k1}\left(\sum_{i=1}^{n} a_{ki} X_{1i}\right) - \mu_{k2}\left(\sum_{i=1}^{n} a_{ki} X_{2i}\right) \ldots - \mu_{kk}\left(\sum_{i=1}^{n} a_{hi} X_{ki} - 1\right)$$

ou encore :

$$\phi = tr\,(A'AV) - tr\left[M'(AX - i)\right]$$

$$\text{où } M = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1k} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2k} \\ & & & \\ \mu_{k1} & \mu_{k2} & \cdots & \mu_{kk} \end{bmatrix}$$

il ne reste plus qu'à dériver ϕ par rapport aux éléments de A.

$$\text{tr } (A'AV) = v_{11} \sum_{i=1}^{k} a_{i1}^2 + v_{22} \sum_{i=1}^{k} a_{i2}^2 + \ldots + v_{nn} \sum_{i=1}^{k} a_{in}^2$$

$$+ 2 v_{12} \sum_{i=1}^{k} a_{i1} a_{i2} + \ldots + 2 v_{1n} \sum_{i=1}^{k} a_{i1} a_{in} + \ldots + 2 v_{n-1} v_{n} \sum_{i=1}^{k} a_{i,n-1} a_{i,n}$$

d'où l'on obtient :

$$\frac{\partial \phi}{\partial a_{11}} = 2 (v_{11} a_{11} + v_{12} a_{12} + \ldots + v_{1n} a_{1n}) - (\mu_{11} X_{11} + \mu_{12} X_{21} + \ldots + \mu_{1k} X_{k1})$$

$$\frac{\partial \phi}{\partial a_{21}} = 2 (v_{11} a_{21} + v_{12} a_{22} + \ldots + v_{1n} a_{2n}) - (\mu_{21} X_{11} + \mu_{22} X_{21} + \ldots + \mu_{2k} X_{k1})$$

et en général :

$$\frac{\partial \phi}{\partial a_{ij}} = 2 (v_{j1} a_{i1} + v_{j2} a_{i2} + \ldots + v_{jn} a_{in}) - (\mu_{i1} X_{1j} + \mu_{i2} X_{2j} + \ldots + \mu_{ik} X_{kj}$$

$$i = 1, \ldots, k$$

$$j = 1, \ldots, n$$

ou encore :

$$\frac{\partial \phi}{\partial a_{ij}} = 2 (i^{e} \text{ ligne de } A \; X_j^{\,e} \text{ colonne de } v) - (i^{e} \text{ ligne de } M \; X_j^{\,ie} \text{ ligne de } X)$$

ce qui donne en annulant les dérivées l'équation matricielle :

(15) $\boxed{2 AV = M X'}$

Multiplions à droite par $V^{-1} X$:

$$2 AV V^{-1} X = M X' V^{-1} X$$

or $AX = \dot{I}$ donc

$$2 \dot{I} = M (X'V^{-1} X)$$

$$M = 2 (X'V^{-1} X)^{-1}$$

$$2 AV = 2 (X'V^{-1} X)^{-1} X' \text{ car } 2AV = M X' \quad (15)$$

et enfin :

$$(16) \qquad \boxed{A = (X'V^{-1}X)^{-1}\, X'V^{-1}}$$

Nous obtenons donc une valeur extrême de la somme des variances pour la valeur de A définie par la relation (16), il ne reste plus qu'à montrer qu'il s'agit bien d'un minimum.

Définissons un estimateur linéaire sans biais :

$$b = (A+B)Y$$

B matrice k x n, non nulle. b est sans biais donc :

$$E(b) = E\left[(A+B)\,(X\beta + u)\right]$$
$$= \beta + (B \times \beta)$$

Ce qui implique :

$$B\,X = 0$$

La somme des variances pour ce nouvel estimateur s'écrit ;

$$E\left[(b-\beta)'\,(b-\beta)\right] = E\left\{u'\left[(X'V^{-1}X)^{-1}\,X'V^{-1} + B\right]'\left[X'V^{-1}X)^{-1}\,X'V^{-1}+B\right]u\right\}$$
$$= E\left\{u'\left[V^{-1}X\,(X'V^{-1}X)^{-1} + B'\right]\left[(X'V^{-1}X)^{-1}\,X'V^{-1}+B\right]u\right\}$$
$$= E\,(u'A'\,Au) + E\left\{u'\left[V^{-1}X\,(X'V^{-1}X)^{-1}\right]Bu\right\}$$
$$+ E\left[u'B'\,(X'V^{-1}X)^{-1}\,X'V^{-1}\,u\right] + E\,(u'B'Bu)$$

on montre alors que ;

$$E\left\{u'\left[V^{-1}X\,(X'V^{-1}X)^{-1}\right]Bu\right\} = tr\left[V^{-1}X\,(X'V^{-1}X)^{-1}\,BV\right]$$
$$E\left[u'\,B'\,(X'V^{-1}X)^{-1}\,X'V^{-1}u\right] = tr\left[B'\,(X'V^{-1}X)^{-1}\,X'\right]$$

d'autre part :

$$tr\,(ABCD) = tr\,(BCDA) = \ldots \text{ etc.}$$
$$tr\left[V^{-1}\,X\,(X'V^{-1}X)^{-1}\,BV\right] = tr\left[(X'V^{-1}X)^{-1}\,B\,X\right]$$
$$= 0 \text{ car } B\,X = 0$$

Donc :

$$(17) \qquad E\left[(b-\beta)'\,(b-\beta)\right] = E\,(u'\,A'\,Au) + E\,(u'\,B'\,Bu)$$

Considérons le deuxième membre de la relation (17)

- le premier terme est la somme des variances pour l'estimateur A, définie par la relation (16)

- le deuxième terme est l'espérance de la forme quadratique $u'B'Bu$, qui pour toute matrice réelle B est définie positive ou semi-définie positive. Elle sera positive si le rang de B est égal à k. Ainsi, la somme des variances pour l'estimateur A, est au moins aussi petite que celle de tout autre estimateur linéaire.

On estime donc β par AY où A est donnée par la relation (16).

Notons que lorsque les erreurs ne sont pas liées : $V^{-1} = 1/\sigma_u^2 \;) \; I$ et $A = (X'X)^{-1}X'$, relation que nous avons rencontrée dans la régression à un nombre quelconque de variables.

La formule (16) peut aussi être obtenue en recherchant un estimateur qui minimise

$$(Y - Xb)' \, V^{-1} \, (Y - Xb) \quad (*)$$

$$(Y - Xb)' \, V^{-1} \, (Y - Xb) = (Y' - b' \, X') \, V^{-1} \, (Y - X \, b)$$

$$= Y' \, V^{-1} Y - 2 \, b' \, X' \, V^{-1} \, Y + b' \, X' \, V^{-1} \, Xb$$

en différenciant par rapport aux composantes de b et en annulant on obtient :

$$2 \, X' \, V^{-1} Y = 2 \, X'V^{-1} \, Xb$$

$$b = (X'V^{-1}X)^{-1}X'V^{-1} \, Y$$

Etudions l'estimateur $\beta^* = AY$

$$\beta^* = AY = (X'V^{-1} \, X)^{-1} \, X'V^{-1} \, (X\beta + u)$$

$$= \beta + (X' \, V^{-1} \, X)^{-1} \, X' \, V^{-1} u$$

La matrice des variances - covariances des éléments de β s'écrit :

$$E \left[(\beta^* - \beta) \, (\beta^* - \beta)' \right] = E \, (Au \, u' \, A')$$

$$(18) \qquad E \left[(\beta^* - \beta) \, (\beta^* - \beta)' \right] = (X'V^{-1}X)^{-1}$$

(*) si on suppose que les erreurs sont distribuées normalement, cette somme apparaît comme exposant de l'exponentielle de la fonction de vraisemblance.

Une fois encore, nous définissons un estimateur linéaire sans biais :

(19) $b = (A+B)\,Y$

comme on l'a montré $BX = 0$. La matrice des variances - covariances de b s'écrit :

$$E\left[(b-\beta)\,(b-\beta)'\right] = E\left[(A+B)\,uu'\,(A'+B')\right]$$

$$= AVA' + BVA' + AVB' + BVB'$$

(20) $$E\left[(b-\beta)\,(b-\beta)'\right] = (X'\,V^{-1}X)^{-1} + BVB'$$

car $BVA' = BVV^{-1}\;X\,(X'V^{-1}X)^{-1} = 0$

et $AVB' = 0$

Les éléments de la diagonale principale de (18) et (20) définissent les variances des éléments de β^* et de b. Les éléments de la diagonale principale de BVB' sont positifs donc :

$$\mathrm{Var}\;\beta_i^* < \mathrm{Var}\;b_i \qquad (i = 1, 2, \ldots k)$$

<u>β^* est donc un estimateur sans biais, la somme des variances de ses éléments est minimum, chacun de ses éléments a une variance minimum. D'autre part β^* est l'estimateur du maximum de vraisemblance, si les erreurs suivent une distribution normale.</u>

exemple Soit le modèle autorégressif défini par les relations (1), (2) et (3).

$$Y_t = \alpha + \beta X_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad |\rho| < 1$$

D'après (5) et (6), nous avons :

$$E\,(u\,u') = V = \frac{\sigma_\varepsilon^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

d'où $$V^{-1} = \frac{1}{\sigma_\varepsilon^2}\begin{bmatrix} 1 & -\rho & & & 0 \\ -\rho & 1+\rho^2 & -\rho & & \\ & -\rho & 1+\rho^2 & -\rho & \\ & 0 & -\rho & \ddots & \ddots \\ & & & -\rho & 1 \end{bmatrix}$$

On peut montrer que la méthode généralisée des moindres carrés équivaut à un processus en deux étapes :

(1) transformer les variables initiales selon le modèle autorégressif,

(2) appliquer la méthode simple des moindres carrés aux variables transformées.

Transformons par la matrice T la relation :

$$Y = X\beta + u$$

$$(23) \qquad TY = TX\beta + Tu$$

L'estimateur des moindres carrés de β est :

$$(24) \qquad \beta^* = \left[(TX)'\,(TX)\right]^{-1}\,(TX)'(TY)$$
$$= (X'T'T\,X)^{-1}\,X'\,T'\,TY$$

en comparant (24) et (16) on voit que les deux estimateurs sont identiques si (25) $\quad T'T = V^{-1}$

Prenons pour simplifier $\sigma^2_\varepsilon = 1$ dans (22). La relation (25) est satisfaite pour la matrice T, d'ordre $(n-1) \times n$:

$$T = \begin{bmatrix} -\rho & 1 & & & \\ & 1 & & 0 & \\ & -\rho & \ddots & & \\ & & & \ddots & 1 \\ & & & -\rho & \\ & 0 & & & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ | \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ | \\ Y_n - \rho Y_{n-1} \end{bmatrix}$$

et T'T est la matrice n x n

$$(26) \qquad T'T = \begin{bmatrix} \rho^2 & -\rho & & & \\ -\rho & 1+\rho^2 & & -\rho & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & -\rho & & -\rho \\ & & & & 1 \end{bmatrix}$$

en comparant avec V^{-1}, on sait que la seule différence provient du premier terme de la première ligne.

T transforme les variables en :

$$\begin{bmatrix} Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ - - - - \\ Y_n - \rho Y_{n-1} \end{bmatrix} \text{ et } \begin{bmatrix} X_2 - \rho X_1 \\ X_3 - \rho X_2 \\ - - - - \\ X_n - \rho X_{n-1} \end{bmatrix}$$

Cette transformation donne des variables auxquelles on peut appliquer la méthode simple des moindres carrés. La matrice des variances-covariances des erreurs transformées s'écrit :

$$E\left[(Tu)\,(Tu)'\right] = T\,V\,T'$$

$$= \sigma_\varepsilon^2 \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

et l'estimateur des moindres carrés minimise

$$(Y - Xb)'\,T'T\,(Y - X\,b)$$

quantité sensiblement égale à :

$$(Y - X\,b)'\,V^{-1}\,(Y - X\,b)$$

Lorsqu'on a lieu de penser que les erreurs satisfont à un modèle autorégressif simple, de paramètre ρ connu, on adoptera cette transformation des variables initiales. Certains économetriciens ont pris ρ égal à l'unité.

<u>Si on applique la méthode simple des moindres carrés pour estimer les paramètres d'une relation à erreurs autocorrelées on obtient des estimateurs sans biais, dont on sous-estime les variances qui ne sont en aucune façon minimales.</u>

Considérons le modèle définit par les relations (7) à (10).

$$Y = X\,\beta + u$$

et soit $\hat{\beta}$ l'estimateur de β obtenu par la méthode simple des moindres carrés :

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

$$E\,(\hat{\beta}) = E\left[(X'X)^{-1}\,X'\,(X\beta + u)\right]$$

$$= \beta$$

car $E\,(u) = 0$

Déterminons la matrice des variances-covariances de $\hat{\beta}$:

$$(28) \qquad E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = (X'X)^{-1} X'VX (X'X)^{-1}$$

si les erreurs ne sont pas liées $V = \sigma_u^2 I$ et (28) se réduit à $\sigma_u^2 (X'X)^{-1}$ comme pour la relation (11)

L'emploi de la formule (11) pour calculer les variances peut être trompeur lorsque les erreurs sont autocorrelées et ceci pour deux raisons :

1° - on ignore de nombreux termes, prenons le cas du modèle autorégressif d'ordre un, la relation (28) donne :

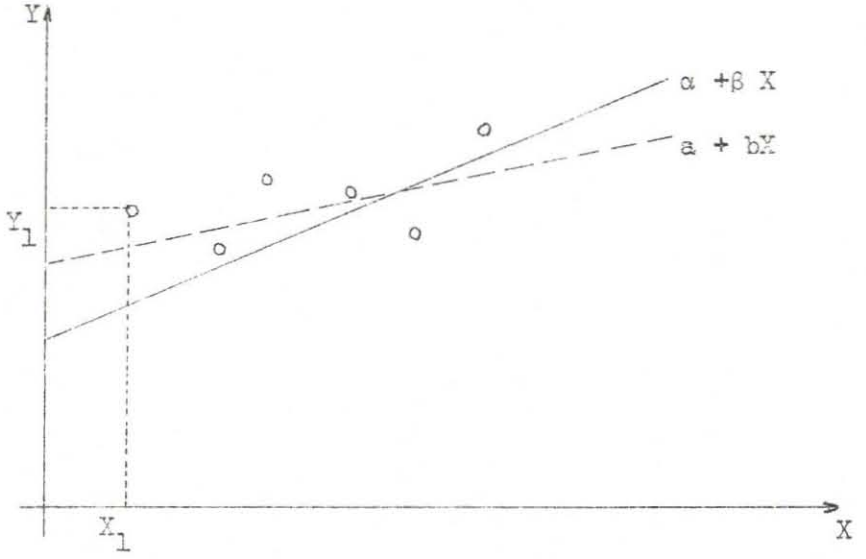
$$(29) \qquad Var(\hat{\beta}) = \frac{\sigma_u^2}{\sum_1^n x_i^2} \left(1 + 2\rho \frac{\sum_1^{n-1} x_i x_{i+1}}{\sum_1^n x_i^2} + 2\rho^2 \frac{\sum_1^{n-2} x_i x_{i+2}}{\sum_1^n x_i^2} + \ldots + 2\rho^{n-1} x_1 x_n\right.$$

La formule des moindres carrés simple s'écrit :

$$(30) \qquad Var(\hat{\beta}) = \frac{\sigma_u^2}{\sum_1^n x_i^2}$$

si ρ est positif et si la variable explicative X est autocorrelée positivement, la somme des termes dans la parenthèse de la relation (29) est supérieure à un et la variance calculée plus grande que celle obtenue par la formule (30)

2° - l'estimation de σ_u^2 à partir d'une série autocorellée conduit à un biais négatif.



Supposons que u et x soient autocorrelés. Si au point (X_1, Y_1) correspond un u positif, il est possible que les points observés se répartissent dans leur ensemble au-dessus de la droite représentant la relation exacte. Donc une mesure de la dispersion autour de la droite ajustée conduira à une sous estimation de σ^2_u .

Ces points sont confirmés par l'examen des résultats expérimentaux. La recherche de tels résultats peut d'ailleurs être conduite de manière systématique. Il suffit de fabriquer à partir d'un modèle autorégressif connu un certain nombre d'échantillons artificiels et de déterminer les estimations que l'on aurait calculées sur chacun des échantillons si l'on avait effectivement observé cet échantillon et lui seul. On dispose aussi pour chaque estimateur d'autant de valeurs qu'il y a d'échantillons. La distribution empirique de ces valeurs fournit une approximation de la loi théorique que l'on souhaite connaître. Cette méthode dite "méthode de Monte-Carlo" a été adoptée notamment par Cochrane et Orcutt (*) et par Malinvaud (**). Cochrane et Orcutt ont considéré la relation ;

(31) $X_1 = 2 X_2 + X_3 + u$

et ils ont engendré X_2, X_3 et u selon les cinq modèles suivants ;

(A) $Z_{t+1} = Z_t + 0,3 (Z_t - Z_{t-1}) + \varepsilon_{t+1}$ (processus autorégressif du second ordre),

(B) $Z_{t+1} = Z_t + \varepsilon_{t+1}$ (limite du processus autorégressif du premier ordre lorsque ρ tend vers un),

(C) $Z_{t+1} = 0,3 Z_t + \varepsilon_{t+1}$ (processus autorégressif du premier ordre)

(D) $Z_{t+1} = \varepsilon_{t+1}$ (processus aléatoire)

(E) $Z_{t+1} = \varepsilon_{t+1} - \varepsilon_t$ (différence des deux processus aléatoires).

Ils appliquèrent la méthode des moindres carrés à 20 ensembles de 20 observations de X_2, X_3 et u pour chaque type de processus générateur (A, B, C, D ou E) et ils comparèrent les variances des résidus par rapport aux variances des erreurs vraies.

(*) D. Cochrane, G.H. Orcutt - "Application of least squares regressions to relationship containing auto-correlated error terms", J. Am. Statist. Assoc. , vol 44, pp. 32-61, 1949.

(**) E. MALINVAUD - "Estimation et prévision dans les modèles économiques autorégressifs" Rev. Inst. Intrn. de Stat. , vol 29, n°2, 1961.

Tableau : Comparaison des variances

n°	processus générateur des variables explicatives	nombre de variables explicatives	variance moyenne des résidus	variance moyenne des erreurs vraies
1	B	1	5 142	1 725
2	D	1	1 375	1 466
3	B	2 +temps	784	4 386
4	B	2	1 690	4 386
5	D	2	634	749

on constate qu'il y a toujours sous estimation de la variance.

Utilisons les formules (18) et (28) pour comparer les variances relatives au cas des moindres carrés simples et à celui des moindres carrés généralisés. Considérons pour illustrer le cas d'une seule variable explicative et d'un erreur suivant un processus autorégressif d'ordre un. Nous avons :

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

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1} X'VX (X'X)^{-1} \qquad \mathrm{var}(\beta^*) = (X'V^{-1} X)^{-1}$$

Ce qui donne dans le cas particulier étudié :

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum_1^n x_i^2} \left(1 + 2\rho \frac{\sum_1^{n-1} x_i x_{i+1}}{\sum_1^n x_i^2} + 2\rho^2 \frac{\sum_1^{n-2} x_i x_{i+2}}{\sum_1^n x_i^2} + \dots + 2\rho^{n-1} x_1 x_n\right)$$

$$\mathrm{Var}(\beta^*) = \frac{\sigma_u^2 (1 - \rho^2)}{(1+\rho^2) \sum_1^n x_i^2 - \rho^2(x_1^2 + x_n^2) - 2\rho \sum_1^{n-1} x_i x_i + 1}$$

d'après (5) $\sigma_u^2 (1-\rho^2) = \sigma_\varepsilon^2$ d'où, si les X ne sont pas autocorrelées approximativement :

$$\frac{\mathrm{var}(\beta^*)}{\mathrm{var}(\hat{\beta})} = \frac{1-\rho^2}{1+\rho^2}$$

en négligeant les covariances sérielles. Si $\rho = 0,5$ l'efficacité des moindres carrés simples est égale à 60 % de celle des moindres carrés généralisés, tandis que si $\rho = 0,9$ le rapport n'est que de 10 %.

4 - La statistique d de Durbin - Watson (*)

Nous avons vu quelles pouvaient être les conséquences d'une application brutale de la méthode des moindres carrés simples à une relation comptant des liaisons entre erreurs. Pour tester l'existence de telles autocorrélations l'emploi de la statistique d de Durbin-Watson est particulièrement simple.

Soit z_t ($t = 1, 2, \ldots n$) les résidus d'un nuage d'observations par rapport à une régression ajustée. Nous définissons :

$$d = \frac{\sum_{t=r}^{n} (z_t - z_{t-1})^2}{\sum_{t=1}^{n} z_t^2}$$

Durbin et Watson ont calculé des bornes inférieure d_i et supérieure d_s en fonction du nombre d'observation n et du nombre de variables explicatives k.

on rejette l'hypothèse d'erreurs aléatoires

si $d < d_i$ on admet l'hypothèse d'une autocorrélation positive.

si $d > d_s$ on ne rejette pas l'hypothèse d'erreurs aléatoires

si $d_i < d < d_s$ on ne conclut pas.

5 - Méthodes d'estimation

Lorsqu'il y a autocorrélation des erreurs, trois méthodes principales d'estimation peuvent être employées :

(i) on connaît V

cette situation est idéale mais innaccessible, cependant dans ce cas, on a :

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

$$\text{var}(\beta) = (X' V^{-1} X)^{-1}$$

(*) J. Durbin, G.S. Watson "Testing for serial correlation in least squares regressions" Biometrika, 1950, partie 1; 1951, partie 2.

Parfois on dispose d'un estimateur V_1 de V, les estimateurs obtenus en remplaçant V par V_1 seront sans biais mais la matrice des variances-covariances?

$$(X' V_1^{-1} X)^{-1}$$

connue avec inexactitude, il faudrait en fait prendre :

$$(X'V_1^{-1} X)^{-1} X'V_i^{-1} V V_1^{-1} X (X' V_i^{-1} X)^{-1}$$

qui fait intervenir V qui est inconnu. Toutefois, il est préférable d'utiliser un estimateur imparfait de V plutôt que d'utiliser les moindres carrés simples.

(ii) procédé itératif

On utilise les moindres carrés simples en calculant,

$$b = (X' X)^{-1} X'Y$$

puis on calcule les résidus :

$$\hat{u} = Y - X b$$

que l'on utilise pour estimer la matrice des variances-covariances des erreurs :

$$V_2 = \hat{u} \hat{u}'$$

puis

$$b = (X' V_2^{-1} X)^{-1} X' V_2^{-1} Y$$

Mais malheureusement V_2 est singulière ;

$$V_2 = \begin{vmatrix} \hat{u}_1^2 & \hat{u}_1 \hat{u}_2 & \cdots & \hat{u}_1 \hat{u}_n \\ \hat{u}_2\hat{u}_1 & \hat{u}_2^2 & \cdots & \hat{u}_2 \hat{u}_n \\ \cdots & \cdots & \cdots & \cdots \\ \hat{u}_n\hat{v}_1 & \hat{u}_n \hat{v}_2 & \cdots & \hat{u}_n^2 \end{vmatrix} = \hat{u}_1 \hat{u}_2 \ldots \hat{u}_n \begin{vmatrix} \hat{u}_1 & \hat{u}_1 \ldots \hat{u}_1 \\ \hat{u}_2 & \hat{u}_2 \ldots \hat{u}_2 \\ \cdots & \cdots \\ \hat{u}_n & \hat{u}_n \ldots \hat{u}_n \end{vmatrix} = 0$$

Il est nécéssaire d'admettre que les erreurs obéissent à un modèle autorégressif fixé par exemple, le modèle simple :

$$u_t = \rho u_{t-1} + \varepsilon_t$$

puis d'estimer ρ en calculant la régression :

$$\hat{u}_t = r \hat{u}_{t-1} + e_t$$

le coefficient r est alors utilisé pour transformer les variables en :

$$(Y_t - rY_{t-2}), (X_{2t} - rX_{2t-1}), \ldots (X_{kt} - rX_{kt-1})$$

et on applique les moindres carrés simples aux variables transformées. Si les résidus qui en résultent ne sont pas autocorrélées, le calcul est arrêté, sinon on effectue une deuxième estimation de ρ qui permet de calculer un deuxième ensemble de variables transformées, et un nouvel ajustement.. jusqu'à ce que les résidus ne soient plus autocorrélés.

(iii) Estimation de tous les paramètres

Soit le modèle:

$$Y = X\beta + u$$

où X satisfait un modèle autorégressif, on peut trouver une matrice telle que :

$$T u = \varepsilon$$

où les ε sont indépendants, centrés, de variance constante, si de plus ils suivent une distribution normale, on peut rechercher des estimateurs du maximum de vraisemblance. On minimise alors ;

$$\varepsilon'\varepsilon = (Y' - \beta'X') T'T (Y - X\beta)$$

$$(32) \qquad = Y' T'TY - 2\beta' X' T'TY + \beta' X' T'TX\beta$$

en dérivant par rapport aux éléments de β et de T.

Durbin (*) a proposé une méthode d'estimation en deux étapes, qui est plus simple. Elle fournit des estimateurs qui ont asymptotiquement la même espérance, et la même matrice des variances-civariances que ceux obtenus directement à partir de (32). Durbin raisonne sur le modèle :

$$(33) \qquad Y_t = \beta_1 X_{1t} + \ldots \beta_q X_{qt} + u_t \qquad (t = 1, 2 \ldots, n)$$

où $\{u_t\}$ est un processus autorégressif stationnaire engendrée par :

$$(34) \qquad u_t + \alpha_1 u_{t-1} + \ldots + \alpha_p u_{t-p} = \varepsilon_t \qquad t = \ldots, -1, 0, 1, \ldots$$

et où chaque série $X_{i1}, X_{i2}, \ldots$ $(i=1, \ldots q)$ est une série de constantes données. On suppose que les ε_t sont indépendants, équidistribués, centrés, de variance constante.

(*) J. Durbin "Estimation of parameters in time-series regression models" J. Royal Statist. Soc., série B, vol 22, n° 1, 139 -153, 1960.

De (33) et (34) on tire :

(35) $Y_t + \alpha_1 Y_{t-1} + \ldots + \alpha_p Y_{t-p} = \beta_1 X_{1t} + \ldots + \beta_q X_{qt} + \alpha_1 \beta_1 X_{1,t-1} + \ldots + \alpha_p \beta_q X_{q,t-p} + \varepsilon_t$

soient $-a_1, \ldots, -a_p$ les coefficients de $Y_{t-1}, \ldots, Y_{t-p}$ dans la régression de Y_t par rapport à Y_{t-1}, Y_{t-p}, $X_{1t}, \ldots, X_{qt}, X_{1t-1}, \ldots X_{q\,t-1}, \ldots X_{1\,t-p}, \ldots X_{q\,t-p}$.

Soit

$$v_t = Y_t + a_1 Y_{t-1} + \ldots a_p X_{i\,t-p}$$

et $w_{it} = X_{it} + a_1 X_{i\,t-1} + \ldots + a_p X_{i\,t-p}$ $(i=1, \ldots)$

alors $\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_q$ sont les coefficients de $w_{1t}, \ldots w_{qt}$ dans la régression de v_t par rapport à $w_{1t}, \ldots w_{qt}$.

<u>exemple</u> Soit :

$$Y_t = \beta X_t + u_t$$

$$u_t + \alpha_1 u_{t-1} = \varepsilon_t$$

à la première étape nous calculerions la régression de Y_t par rapport à Y_{t-1}, X_t et X_{t-1}, en désignant par $-a$ le coefficient de Y_t. Puis nous calculerions les séries :

$$v_t = Y_t + a\, Y_{t-1}$$

et la régression de v_t par rapport à w_t qui fournirait les estimés de β.

6 - Problèmes de prévision

Les liaisons entre erreurs se rencontrent fréquemment dans l'estimation de relations à partir de données chronologiques. Nous avons vu, que les estimations étaient améliorées si l'on connaissait la structure des liaisons il en est de même pour les prévisions.

Soient les observations X_t, Y_t (t=1, 2, ...n) et le modèle :

(1) $Y_t = \alpha + \beta X_t + u_t$

(2) $u_t = \rho u_{t-1} + \varepsilon_t$

et où satisfait aux hypothèses (3). Prévoir Y_{n+1} sachant X_{n+1}. On a :

(37) $E(Y_{n+1} / u_1, \ldots u_n) = \alpha + \beta X_{n+1} + E(u_{n+1}/u_1, \ldots u_n)$

$$= \alpha + \beta X_{n+1} + \rho u_n$$

d'où en employant la relation (1) :

(38) $E(Y_{n+1}/u_1, \ldots u_n) = \alpha(1-\rho) + \beta(X_{n+1} - \rho X_n) + \rho Y_n$

(39) $E\left[(Y_{n+1} - \rho Y_n)/(u_1, \ldots u_n)\right] = \alpha(1-\rho) + \beta(X_{n+1} - \rho X_n)$

Si l'on connaît ρ, on écrit (1) sous la forme :

(40) $Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \varepsilon_t$

qui satisfait les hypothèses de la régression simple. En appliquant la méthode des moindres carrés simples, on obtient le meilleur estimateur linéaire sans biais de $Y_{n+1} - \rho Y_n$ c'est-à-dire :

$$\alpha^*(1-\rho) + \beta^*(X_{n+1} - \rho X_n)$$

d'où :

(41) $Y^*_{n+1} = \alpha^*(1-\rho) + \beta^*(X_{n+1} - \rho X_n) + \rho Y_n$

(42) $= \alpha^* + \beta^* X_{n+1} + \rho\left[Y_n - (\alpha^* + \beta^* X_n)\right]$

Ainsi le meilleur estimateur de Y_{n+1} a deux propriétés intéressantes :

- il utilise la connaissance du processus autorégressif,
- on corrige $\alpha^* + \beta^* X_n$ par le produit de ρ par un estimateur de la perturbation pour la période précédente.

L'utilisation brutale des moindres carrés à partir de la relation (1) aurait conduit à :

(43) $\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta} X_{n+1}$

qui est d'une part biaisé, d'autre part moins efficace que (42).

Ces résultats se généralisent à un nombre quelconque de variables explicatives.

Exercice

Soient les données sur la consommation Y et le revenu X aux Etats-Unis, par la méthode des moindres carrés on a calculé :

$$\hat{Y} = 7,0 + 0,0025\ X$$

année (1)	consommation Y (2)	revenu X (3)	$\hat{Y}$ (4)	$\hat{u}$ (5)	$\Delta\hat{u}$ (6)
1948	199	212	198,3	0,7	
					3,2
1949	204	214	200,1	3,9	
					-3,4
1950	216	231	215,5	0,5	
					-3,4
1951	218	237	220,9	-2,9	
					-0,3
1952	224	244	227,2	-3,2	
					1,1
1953	235	255	237,1	-2,1	
					1,2
1954	238	257	238,9	-0,9	
					3,5
1955	256	273	253,4	2,6	
					-1,9
1956	264	284	263,3	0,7	
					0,6
1957	270	290	268,7	1,3	

(milliards de dollars 1954)

$$\hat{u} = Y - \hat{Y}$$

nous calculons la statistique de Durbin-Watson :

$$d = \frac{\sum(\Delta\ \hat{u})^2}{\sum\hat{u}^2} = 1,07$$

valeur qui indique une autocorrélation positive des erreurs. Considérons alors le modèle :

$$\hat{u}_t = r\ \hat{u}_{t-1} + e_t$$

$$r = \frac{\sum_2^n \hat{u}_t\ \hat{u}_{t-1}}{\sum_2^n \hat{u}_{t-1}^2} = 0,457$$

On définit alors les variables ;

$$Y'_t = Y_t - 0,457\ Y_{t-1}$$

$$X'_t = X_t - 0,457\ X_{t-1}$$

année	Y't	X't	$\hat{Y}$'t	û't	Δ u't
1949	113,1	117,1	109,3	3,8	-5,0
1950	122,0	133,2	124,0	-1,2	-1,9
1951	119,3	131,4	122,4	-3,1	1,2
1952	124,4	135,7	126,3	-1,9	1,1
1953	132,6	143,5	133,4	-0,8	0,7
1954	130,6	140,5	130,7	-0,1	2,9
1955	147,2	155,6	144,4	2,8	-3,5
1956	147,0	159,2	147,7	-0,7	1,5
1957	149,4	160,2	148,6	0,8	

L'application des moindres carrés donne :

$$\hat{Y}'t = 2,6 + 0,9114\ X't$$

on calcule alors les résidus û't, puis la nouvelle valeur d' de la statistique de Durbin-Watson :

$$d' = 1,41$$

qui signifie que l'autocorrélation a presque disparu.

— ANNEXE —

$$u'A'Au = \left[u_1, u_2, \ldots, u_n\right] \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_n \end{bmatrix}$$

$$= \left[u_1 w_{11} + w_2 w_{21} + \ldots + u_n w_{n1}; \ldots u_1 w_{1n} + u_2 w_{2n} + \ldots u_n w_{nn}\right] \begin{bmatrix} u_1 \\ u_2 \\ u_n \end{bmatrix}$$

$$= w_{11} u_1^2 + w_{22} u_2^2 + \ldots + w_{nn} u_n^2 + 2 w_{12} u_1 u_2 + \ldots + 2 w_{1n} u_1 u_n$$

$$+ \ldots + 2 w_{n-1,n} u_{n-1} u_n$$

$$A'Auu' = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ & & & \\ & & & \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} u_1^2 & u_1 u_2 & \ldots & u_1 u_n \\ u_2 u_1 & u_2 & \ldots & u_2 u_n \\ & & & \\ u_n u_1 & \ldots & & u_n^2 \end{bmatrix}$$

$$tr\,(A'A\,u\,u') = w_{11} u_1^2 + w_{22} u_2^2 + \ldots + w_{nn} u_n^2$$

$$+ 2 w_{12} u_1 u_2 + \ldots + 2 w_{1n} u_1 u_n + \ldots + 2 w_{n-1,n} u_{n-1} u_n$$

$$= u' A' A u$$

D'où :

$$E\,(u'A'Au) = E\left[tr\,(A'Auu')\right]$$
$$= tr\left[A'A \quad E(u\,u')\right]$$
$$= tr\,(A'A'V)$$

Chapitre 3 QUELQUES POINTS PARTICULIERS DE L'ETUDE DES MODELES A UNE EQUATION

Nous aborderons successivement :

- la collinéarité,
- l'hétéroscédasticité,
- les modèles à retards échelonnés,
- l'introduction de variables fictives dans les modèles.

1 - Collinéarité

Lorsque les variables explicatives, ou une partie d'entre elles, sont hautement corrélées, il devient presque impossible, ou pour le moins très difficile, de séparer leurs effets. A cette situation correspond le problème de la collinéarité des variables explicatives.

Nous considérerons trois cas :

(i) Supposons que la théorie nous conduise à supposer que Y soit liée à X_2 et X_3 pour une relation linéaire :

(1) $$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t.$$

A partir de n observations, nous désirons estimer les paramètres de la relation (1). Nous savons que si nous écrivons ;

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} x_{21} & x_{31} \\ x_{22} & x_{32} \\ \hdashline x_{2n} & x_{3n} \end{bmatrix} \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

(en introduisant les variables centrées)

les estimateurs des moindres carrés sont donnés par :

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

avec

$$\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

cette méthode est prise en défaut si $|X'X| = 0$. Supposons que X_2 et X_3 soient liés par une relation linéaire stricte :

$$X_{2t} = k_2 + k_3 X_{3t} \quad \text{pour tout } t,$$

d'où

$$x_{2t} = k_3 x_{3t}$$

$$\text{alors } X'X = \begin{bmatrix} \sum x_2^2 & \sum x_2 x_3 \\ \sum x_2 x_3 & \sum x_3^2 \end{bmatrix} = \sum x_3^2 \begin{bmatrix} k_3^2 & k_3 \\ k_3 & 1 \end{bmatrix}$$

d'où $\quad |X'X| = 0$

exemple d'un cas parfait de collinéarité, où il est absolument impossible de séparer les effets de X_2 et X_3. L'interprétation géométrique est très révélatrice. Les points sont alignés dans le plan (X_2, X_3), tenter d'ajuster un plan, revient à rechercher un plan vertical.

Revenons au cas general où X est d'ordre n x k, une relation stricte entre certaines variables explicatives réduit à une valeur r<k le rang de la matrice X'X. Après arrangements des lignes de cette matrice on peut considérer que les k-r dernières lignes sont linéairement dependantes des r premières, et au lieu de raisonner sur l'ensemble complet d'équations normales :

(2) $\quad (X'X)\,\hat{\beta} = X'Y$

on raisonne sur :

(3) $\quad S\hat{\beta} = X'r\,Y$

où S est une matrice d'ordre r x k formée par les r premières lignes de X'X, et X'r une matrice d'ordre rxn comprenant les observations sur les r premières variables explicatives. S a au moins un mineur d'ordre r, non nul. Supposons par exemple que les r premières colonnes de S soient indépendantes. Nous pouvons écrire (3) sous la forme de :

(4) $\quad S_1\,\hat{\beta}_r + S_2\,\hat{\beta}_{k-r} = X'_r\,Y$

en partitionnant S en une matrice régulière S_1 d'ordre r et une matrice d'ordre r x (k-r), ainsi que $\hat{\beta}$. Supposons que $\hat{\beta}_{k-r}$ soit connu, on peut alors résoudre (4) :

$$(5) \qquad \hat{\beta}_r = S_1^{-1}\left[X'_r Y - S_2\,\hat{\beta}_{k-r}\right]$$

Une solution analogue peut être obtenu pour tout ensemble de r colonnes indépendantes de S. Il est possible, cependant, que même si k-r coefficients sont connus à priori, les r coefficients restant puissent correspondre à un en- ensemble de colonnes dépendantes, auquel cas il n'y a pas de solution.

Pour illustrer cette technique, considérons le cas de trois variables, la relation (2) peut s'écrire :

$$\begin{bmatrix} \sum x_2^2 & \sum x_2\, x_3 \\ \sum x_2\, x_3 & \sum x_3^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum x_2 y \\ \sum x_3\, y \end{bmatrix}$$

si $x_3 = kx_2$, le système d'équations est redondant et (3) s'écrit :

$$\begin{bmatrix} \sum x_2^2 & \sum x_2\, x_3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \sum x_2\, y$$

$$\hat{\beta}_2 \sum x_2^2 + \hat{\beta}_3 \sum x_2\, x_3 = \sum x_2\, y$$

si $\hat{\beta}_3$ est connu, on trouve :

$$(6) \qquad \hat{\beta}_2 = \frac{\sum x_2\, y - \hat{\beta}_3 \sum x_2 x_3}{\sum x_2^2} = \frac{\sum x_2\,(y - \hat{\beta}_3\, x_3)}{\sum x_2^2}$$

qui est l'estimateur de la pente de la droite de régression de $(y - \hat{\beta}_3 x_3)$ par rapport à x_2. La technique d'estimation déduite de la relation (5), équivaut à corriger la variable dépendante de l'influence des variables explicatives correspondant aux coefficients connus, puis de faire une régression de ces résidus par rapport aux variables explicatives restantes.

Cette technique ne convient cependant pas toujours, prenons le cas de trois variables et supposons que l'on ait $X_3 = k\,X_2$. L'équation (3) devient alors :

$$\begin{bmatrix} n & \sum X_2 & k \sum X_2 \\ \sum X_2 & \sum X_2^2 & k \sum X_2^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_2\, Y \end{bmatrix}$$

si $\hat{\beta}_2$ ou $\hat{\beta}_3$ sont connus on peut en déduire les deux autres, mais si $\hat{\beta}_1$ est connu le mineur :

$$\begin{bmatrix} \sum X_2 & n\sum X_2 \\ \sum X_2^2 & n\sum X_2^2 \end{bmatrix} = 0$$

(ii) Un cas moins extrême et plus vraisemblable est celui où il existe une corrélation forte (sans liaison stricte) entre les variables explicatives. Dans le cas de trois variables ;

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2.3}^2} \qquad \text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3.2}^2}$$

où $x_{2.3} = x_2 - b_{2\,3}\, x_3$, $x_{3.2} = x_3 - b_{3\,2}\, x_2$ (b_{23} et b_{32} coefficient de régression simple)

Lorsqu'il y a une corrélation forte entre X_2 et X_3, $\sum x_{2.3}^2$ et $\sum x_{3.2}^2$ sont petites et décroissent lorsque la corrélation croit. Aussi y-a-t-il une très forte imprécision quant à l'estimation des effets des variables.

σ^2 est en général inconnu, et on l'estime sur la base de $\sum x_{1.23}^2$. Il n'y a pas de raison apparente pour que cet estimateur soit biaisé en baisse lorsque la corrélation entre variables explicatives croit. Haavelmo pense que "l'estimateur de σ^2 n'est pas diminué du fait que les variables indépendantes sont fortement correlées". Cependant Stone, se basant sur le fait que :

$$\text{var}(\hat{\beta}_2) \simeq \frac{1 - R_{1.23}^2}{1 - r_{23}^2}$$

$R_{1.23}^2$ = coefficient de corrélation multiple

r_{23}^2 = coefficient de corrélation simple.

et que $R_{1.23}$ et r_{23} sont voisins de l'unité pense que var $(\hat{\beta}_2)$ peut être faible par rapport à $\hat{\beta}_2$ ce qui peut entraîner la conclusion erronée selon laquelle $\hat{\beta}_2$ est relativement bien déterminé. Stone pose en fait implicitement qu'une corrélation croissante entre variables explicatives entrainent nécéssairement une diminution de $\sum x_{1.23}^2$ (donc un accroissement de $R_{1.23}^2$). Ce qui n'est pas à priori exacte ni même confirmé par l'exemple numérique suivant.

Exemple numérique de collinéarité

X_2	10	20	30	40	50
v	-6	3	-3	12	-12
$X_3 = X_2+v$	4	23	27	52	38
ε	1	0	4	1	2
$X_1 = 2X_2+X_3+\varepsilon$	25	63	91	133	140

On a construit les valeurs de X_3 de telle sorte qu'elle soient corrélées avec celles de X_2.
Les moindres carrés appliqués à ces données donne :

$$r^2_{23} = 0.74 \qquad R^2_{1.23} = 0.9991 \qquad \sum x^2_{1.23} = 8.1 \qquad \sum x^2_{2.3} = 261.9$$

$$\hat{\beta}_2 = 2.006 \qquad \text{var}\,(\hat{\beta}_2) = 0.0155$$

Examinons ce qu'il résulte d'un accroissement de la corrélation entre X_2 et X_3. Prenons pour v les valeurs :

-2, 1, -1, 4, -4

c'est-à-dire diminuons les valeurs initiales en les divisant par trois.
Les moindres carrés donnent sur les nouvelles données :

$$r^2_{23} = 0.96 \qquad R^2_{1.23} = 0.9985 \qquad \sum x^2_{1.23} = 13.5 \qquad \sum x^2_{2.3} = 36.5$$

$$\hat{\beta}_2 = 2.20 \qquad \text{var}\,(\hat{\beta}_2) = 0.1849$$

La corrélation entre variables explicatives est passée de 0.74 à 0.96, mais elle ne s'est pas accompagnée d'une réduction de $\sum x^2_{1.23}$ qui en réalité augmente de 8;1 à 13.5. Quant à la variance de $\hat{\beta}_2$ elle a été multipliée par 12.

(iii) Il s'agit d'un problème étudié par Frisch. Supposons que les variables se décomposent en deux parties :

- une composante systématique X'_{it}
- une composante d'erreur X''_{it}

$$X_{it} = X'_{it} + X''_{it}$$

et que la théorie spécifie une relation linéaire entre composantes systématiques :

$$X'_1 = \beta_1 + \beta_2 X'_2 + \beta_3 X'_3$$

S'il existe une relation linéaire stricte entre X'_2 et X'_3, comme dans (i), il est impossible de déterminer $\hat{\beta}_2$ et $\hat{\beta}_3$. Mais, ce fait, peut être masqué en pratique par la présence d'erreurs d'observation, et l'application de la méthode des moindres carrés peut conduire à des estimations en apparence bonnes. Alors que les vraies valeurs sont alignées dans un espace à trois dimensions les valeurs observées sont dispersées du fait des erreurs, et ce sont celles-ci qui servent de base à l'estimation.

S'il existe une corrélation élevée entre X_2 et X_3 l'estimation risque d'être encore dominée par les erreurs. Frisch a mis au point une méthode d'analyse dans ce cas. Il propose de déterminer différents plans de régression, en minimisant la somme des carrés des résidus dans différentes directions. Si les résidus sont aléatoires, les plans obtenus seront très différents alors qu'ils seront proches dans le cas contraire.

(iv) Le remède à la collinearité est dans l'acquisition de nouvelles données. Les premières études de la demande, basées sur des données chronologiques, rencontrèrent des difficultés dues à la corrélation entre prix et revenu. L'utilisation de données ponctuelles a permis de déterminer avec une bonne précision l'influence du revenu. Cette utilisation conjointe de séries chronologiques et de données ponctuelles est maintenant classique dans ce type d'études. Certaines difficultés de spécification et d'interprétation demeurent.

Lorsque l'on se propose de faire des prévisions, la corrélation entre variables explicatives est moins grave, dans la mesure où elle a toutes les chances de se perpétuer.

2 - <u>Hétéroscédasticité</u>

Une des hypothèses du modèle linéaire de la régression est que :

$$E(uu') = \sigma^2 I$$

ce qui signifie que les erreurs sont indépendantes et de variance constante σ^2. <u>La condition : variance constante, est appelée homoscédasticité, sa négation hétéroscédasticité.</u>

En notation matricielle on a :

(7) $\boxed{Y = X\beta + u}$

posons que :

(8) $E(u\,u') = \sigma^2 \begin{bmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_n \end{bmatrix}$

et définissons la matrice diagonale :

(9) $\Lambda = \begin{bmatrix} \sqrt{\lambda 1} & & & 0 \\ & \sqrt{\lambda 2} & & \\ & & \ddots & \\ 0 & & & \sqrt{\lambda n} \end{bmatrix}$

La relation (7) s'écrit sous la forme :

(10) $\boxed{\Lambda Y = \Lambda X\,\beta + \Lambda u}$

et la matrice des variances-covariances des erreurs transformées u est un multiple de l'unité, en effet :

$$E(\Lambda u\,u'\Lambda') = \sigma^2 I$$

Ainsi peut on appliquer les moindres carrés simples à la relation (10).

(11) $\boxed{\hat{\beta} = (X'\Lambda^2 X)^{-1}\,X'\Lambda^2 Y}$

et

$$\mathrm{var}(\hat{\beta}) = \sigma^2 (X'\Lambda^2 X)^{-1}$$

Dans le cas d'une seule variable explicative :

$$X'\Lambda^2 X = \begin{bmatrix} \sum\lambda_i & \sum \lambda_i X_i \\ \sum\lambda_i X_i & \sum\lambda_i X_i^2 \end{bmatrix}$$

et

(13) $\begin{cases} \hat{\beta}_1 \sum \lambda_i + \hat{\beta}_2 \sum \lambda_i X_i = \sum \lambda_i Y_i \\ \hat{\beta}_1 \sum\lambda_i X_i + \hat{\beta}_2 \sum\lambda_i X_i^2 = \sum\lambda_i X_i Y_i \end{cases}$

(14) $\mathrm{var}(\hat{\beta}_2) = \dfrac{\sigma^2 \sum\lambda_i}{(\sum \lambda_i)(\sum \lambda_i X_i^2) - (\sum\lambda_i X_i)^2}$

Si nous avions appliqué directement les moindres carrés à la relation (7) nous aurions obtenu :

(15) $\beta^* = (X'X)^{-1}\,X'Y$

qui est sans biais et dont la matrice des variances-covariances est donné par :

$$(16) \qquad E\left[(\beta^* - \beta)(\beta^* - \beta)'\right] = (X'X)^{-1} X' E(uu') X(X'X)^{-1}$$

Lorsqu'il y a une seule variable explicative (15) et (16) donnent :

$$(17) \qquad \begin{cases} n\,\beta_1^* + \beta_2^* \sum X = \sum Y \\ \beta_1^* \sum X + \beta_2^* \sum X^2 = \sum XY \end{cases}$$

$$(18) \quad \text{var}\,(\beta_2^*) = \frac{\sigma^2\left[(\sum \frac{1}{\lambda_i})(\sum X_i)^2 - 2\,n(\sum \frac{1}{\lambda_i} X_i)(\sum X_i) + n^2 (\sum \frac{1}{\lambda_i} X_i^2)\right]}{\left[n \sum X_i^2 - (\sum X_i)^2\right]^2}$$

Lorsque la matrice Λ est connue le choix entre une application directe des moindres carrés et une application après transformation se fait en comparant (12) et (16). Illustrons cette comparaison dans le cas d'une variable explicative en faisant des hypothèses sur la forme de l'hétéroscédasticité. Supposons que :

$$(19) \qquad E\,(u_i^2) = \sigma^2 X_i^2 \qquad i=1,\ 2 \ldots\ n$$

situation que l'on rencontre dans les enquêtes sur budjets.

$$(20) \qquad \sqrt{\lambda_i} = \frac{1}{X_i}$$

Les relations (14) et (18) donnant :

$$(21) \qquad \text{var}\,(\hat{\beta}_2) = \frac{\sigma^2 \sum (1/X_i^2)}{n \sum (1/X_i^2) - \left[\sum 1/X_i\right]^2}$$

$$(22) \qquad \text{var}\,(\beta_2^*) = \sigma^2 \frac{(\sum X_i^2)(\sum X_i)^2 - 2\,n(\sum X_i^3)(\sum X_i) + n^2(\sum X_i^4)}{\left[n \sum X_i^2 - (\sum X_i)^2\right]^2}$$

prenons pour valeurs de X : 1, 2, 3, 4 et 5, on trouve alors :

$$\frac{\text{var}\,\hat{\beta}_2}{\text{var}\,\beta_2^*} = \frac{0{,}69}{1{,}24} = 0{,}56$$

on choisit donc $\hat{\beta}_2$.

Si nous appliquons la relation (20) à la relation (10), on obtient :

$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_1} + \beta_2 + v_i \qquad i = 1,\ 2, \ldots\ n \quad v_i = u_i/X_i$$

Traitons maintenant le cas où

$$E(u_i^2) = \sigma^2 X_i$$

et $$\lambda_i = \frac{1}{X_i}$$

Les relations (14) et (18) donnent :

$$\text{var}(\hat{\beta}_2) = \sigma^2 \frac{\sum(1/X_i)}{\sum(1/X_i)\sum X_i - n^2}$$

$$\text{var}(\beta_2^*) = \sigma^2 \frac{(\sum X_i)^3 - 2n(\sum X_i^2)(\sum X_i) + n^2(\sum X_i^3)}{\left[n\sum X_i^2 - (\sum X_i)^2\right]^2}$$

et l'application numérique :

$$\frac{\text{var}(\hat{\beta}_2)}{\text{var}(\beta_2^*)} = 0,83$$

, montre que les moindres carrés appliqués directement donne de meilleurs résultats que précédemment.

3 - Modèles à retards échelonnés

Supposons que les valeurs de la variable dépendante soient liées, le cas le plus simple est :

(23) $$\boxed{Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t}$$

On sait qu'une des conditions nécéssaires pour appliquer les moindres carrés est que, dans le cas d'une seule variable explicative on ait :

(24) $$E\left\{\varepsilon_t \left[X_{t+r} - E(X_{t+r})\right]\right\} = 0 \qquad \forall t, \forall r = \ldots -1, 0, 1 \ldots$$

Cette condition signifie que ε_t est indipendante des valeurs futures , présentes et passées des variables explicatives. D'après le modèle (23) $X_t = Y_{t-1}$, donc ε_t influence $Y_t = X_{t+1}$, aussi ε_t n'est-elle pas indépendante des valeurs futures de la variable explicative. L'application des moindres carrés dans ce cas conduit à des estimateurs biaisés ; ils sont cependant asymptotiquement efficaces et corrects.

Raisonnons sur le modèle :

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$$

et supposons que les ε_t soient normales, indépendantes, centrées de variance σ^2 . Supposons que Y_1 soit fixée, et donc que la suite des valeurs dépend des n-1 valeurs $\varepsilon_2, \ldots \varepsilon_n$. La fonction de vraisemblance s'écrit :

$$f(\varepsilon_2, \varepsilon_3 \ldots \varepsilon_n) = f(\varepsilon_2)\, f(\varepsilon_3) \ldots f(\varepsilon_n)$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^{n-1}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n} \varepsilon_i^2\right)$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^{n-1}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=2}^{n} (Y_i - \alpha - \beta Y_{i-1})^2\right]$$

maximiser la fonction de vraisemblance équivaut à minimiser :

$$\sum_{i=2}^{n} (Y_i - \alpha - \beta Y_{i-1})^2$$

Ce qui conduit aux estimateurs des moindres carrés $\hat{\alpha}$ et $\hat{\beta}$ en résolvant :

$$\begin{cases} \sum_{i=2}^{n} Y_i = (n-1)\,\hat{\alpha} + \hat{\beta}\sum_{i=2}^{n} Y_{i-1} \\ \sum_{i=2}^{n} Y_i\, Y_{i-1} = \hat{\alpha}\sum_{i=2}^{n} Y_{i-1} + \hat{\beta}\sum_{i=2}^{n} Y_{i-1}^2 \end{cases}$$

ainsi lorsque Y_1 est fixé, nous obtenons des estimateurs asymptotiquement efficaces et corrects.

Supposons que Y_1 soit aléatoire, nous obtenons à partir de (23) :

(26) $$Y_t = \frac{\alpha}{1-\beta} + \sum_{r=0}^{\infty} \beta^r \varepsilon_{t-r} \quad |\beta| < 1$$

(27) $$E(Y_t) = \frac{\alpha}{1-\beta}$$

(28) $$\mathrm{var}(Y_t) = \frac{\sigma^2}{1-\beta^2}$$

comme ε est normal, Y_t l'est aussi et sa distribution est entièrement définie par les relations (27) et (28). La densité de probabilité de Y_1 s'écrit :

$$L_1 = \frac{\sqrt{1-\beta^2}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1-\beta^2}{2\sigma^2}\left(Y_1 - \frac{\alpha}{1-\beta}\right)^2\right]$$

et la fonction de vraisemblance d'un échantillon :

$$L_1 \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=2}^{n} (Y_t - \alpha - \beta Y_{t-1})^2\right]$$

Les estimateurs du maximum de vraisemblance sont différents de ceux des moindres carrés, mais ils convergent asymptotiquement les uns vers les autres.(*)

Les économétriciens utilisent très souvent de petits échantillons aussi les propriétés asymptotiques ne les intéressent-ils pas beaucoup. Hurwicz (**) a étudié le biais des estimateurs, sans en étudier la valeur. Déterminons en le signe. Soit :

$$(23) \qquad Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$$

l'estimatedr des moindres carrés de β:

$$(29) \qquad \hat{\beta} = \frac{\sum_{t=2}^{n} (Y_t - \overline{Y})\,(Y_{t-1} - \overline{\overline{Y}})}{\sum_{t=2}^{n} (Y_{t-1} - \overline{\overline{Y}})^2}$$

avec $\overline{Y} = \frac{1}{n-1}\,(Y_2 + \ldots + Y_n)$

$$\overline{\overline{Y}} = \frac{1}{n-1}\,(Y_1 + \ldots + Y_{n-1})$$

d'après (23) $\overline{Y} = \alpha + \beta\,\overline{\overline{Y}} + \overline{\varepsilon}$

$$Y_t - \overline{Y} = \beta\,(Y_{t-1} - \overline{\overline{Y}}) + (\varepsilon_t - \overline{\varepsilon})$$

$$(30) \text{ et } \quad \hat{\beta} = \beta + \frac{\sum_{t=2}^{n} (Y_{t-1} - \overline{\overline{Y}})\,\varepsilon_t}{\sum_{t=2}^{n} (Y_{t-1} - \overline{\overline{Y}})^2}$$

E ($\hat{\beta}$) comprend donc l'espérance du quotient qui est difficile à évaluer. Considérons son signe, c'est celui du numérateur. (23) donne :

(*) Tjalling C. Koopmans (editeur...) "Statistical inference in dynamic economic models ; Cowles Commission Monograph 10, Wiley, New-York, 1950, chapitre 7 par L. Hurvicz "Prédiction and least squares" page 298 note n°3.

(**) in Koopmans (op. cit.) chapitre 15 par L. Hurvicz "Least squares bias in time series" pages 365 - 383.

$$Y_1 = Y_1$$

$$Y_2 = \alpha + \beta Y_1 + \varepsilon_2$$

$$Y_3 = \alpha + \alpha\beta + \beta^2 Y_1 + \beta\varepsilon_2 + \varepsilon_3$$

$$Y_4 = \alpha + \alpha\beta + \alpha\beta^2 + \beta^3 Y_1 + \beta^2 \varepsilon_2 + \beta \varepsilon_3 + \varepsilon_4$$

- - - - - - - - - - - - -

$$Y_{n-1} = \alpha + \alpha\beta + \ldots + \alpha\beta^{n-3} + \beta^{n-2} Y_1 + \beta^{n-3} \varepsilon_2 + \ldots + \varepsilon_{n-1}$$

Ainsi $\bar{\bar{Y}}$ = termes en α , β et Y_1 plus $B_2 \varepsilon_2 + B_3 \varepsilon_3 + \ldots + B_{n-1} \varepsilon_{n-1}$ où les B sont des moyennes des puissances de β. Si $\beta > 0$ tous les β sont positifs. Si l'on suppose que les ε sont indépendants :

$$E\left[\sum_{t=2}^{n} (Y_{t-1} - \bar{\bar{Y}})\, \varepsilon_t\right] = -\sigma^2 \sum_{i=2}^{n-1} B_i$$

donc le biais est négatif.

Combinant des variables décalées et des résidus autocorrélés, Orcutt et Cochrane (*) ont examiné le modèle :

$$x_t = 0{,}4\, x_{t-1} + u_t$$

où les u_t sont corrélés positivement.

biais dus à des variables décalées et des erreurs autocorrélées

	vraie valeur	estimation	écart type
α	0	-19,36	4,71
β	0,4	0,90	0,02

Remarquons que la corrélation des résidus ne conduit pas à un biais, que les variables décalées en entrainent un qui est négatif et que simultanément ces deux influences entrainent un biais positif.

Les variables décalées sont de plus en plus utilisées dans les travaux d'économétrie (**). Soit le modèle :

(31) $Y_t = \alpha_c X_t + \alpha_i X_{t-1} + \ldots + u_t$

(*) G.H. Orcutt et D. Cochrane : "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis" J. Am. Statist. Assoc., vol 44, pp. 356-372 , 1949.

(**) L. M. KOYCK : "Distributed lags and investment analysis" North-Holland Publishing Company, Amsterdam, 1954.

Marc Nerlove : "Distributed and demand analysis for agricultural and other commodities "Agriculture Hanbook 145, US Djtof agriculture, 1958.

et soit l'hypothèse

(32) $\alpha_j = \alpha \lambda^j \qquad j = 1, 2 \ldots$

$0 \leq \lambda < 1$

(31) peut s'écrire :

$$Y_{t+1} = \alpha X_{t+1} + \alpha \lambda X_t + \alpha \lambda^2 X_{t-1} + \ldots + u_{t+1}$$

$$\lambda Y_t = \alpha \lambda X_t + \alpha \lambda^2 X_{t-1} + \ldots + \lambda u_t$$

et delà :

(33) $Y_{t+1} = \lambda Y_t + \alpha X_{t+1} + (u_{t+1} - \lambda u_t)$

(34) $\Delta Y_t = \alpha X_{t+1} - \gamma Y_t + (u_{t+1} - \lambda u_t)$

où $\Delta Y_t = Y_{t+1} - Y_t$

$\gamma = 1 - \lambda$

Même l'étude du modèle (33) est difficile. Si u suit le modèle autorégressif;

(35) $u_t - \lambda u_{t-1} = \varepsilon_t$

où les ε sont indépendantes, centrées de même variance. Les résidus du modèle (33) sont indépendants et les estimateurs des moindres carrés de α et λ sont corrects, mais biaisés pour des échantillons finis. Il reste cependant peu vraisemblable que (35) soit vérifiée.

On a recherché des estimateurs corrects des paramètres du modèle (33). Soit :

(36) $u_t = \xi u_{t-1} + \varepsilon_t \qquad 0 \leq \xi < 1$

et supposons que l'on ait calculé les estimateurs $\hat{\alpha}$ et $\hat{\lambda}$ des moindres carrés. Calculons la somme des carrés des résidus :

$$\sum_{t=1}^{n} z_t^2 = \sum_{t=1}^{n} (Y_t - \hat{\alpha} X_t - \hat{\lambda} Y_{t-1})^2$$

Pour obtenir les estimateurs corrects, Koyck (*) propose le système :

$$(37) \quad \begin{cases} \bar{\alpha} \sum X_t^2 + \bar{\lambda} \sum X_t Y_{t-1} = \sum X_t Y_t \\ \bar{\alpha} \sum X_t Y_{t-1} + \bar{\lambda} \sum Y_{t-1}^2 = \sum Y_t Y_{t-1} + \dfrac{(\bar{\lambda} - \xi) \sum z_t^2}{1 - \xi \bar{\lambda} + \hat{\lambda} (\bar{\lambda} - \xi)} \end{cases}$$

Il ne propose pas de méthode pour estimer ξ, mais suggère de résoudre (37) pour différentes valeurs de ξ, et d'en déduire les variations de $\bar{\alpha}$ et $\bar{\lambda}$. Les études empiriques de Koyck montrent que ces variations sont faibles.

(*) op. cit.

Lorsque $\xi = 0$, Klein (*) a montré que les estimateurs de Koyck pouvaient être obtenus plus simplement. Ecrivons (33) sous la forme :

$$(38) \quad Y_t - u_t = \alpha X_t + \lambda (Y_{t-1} - u_{t-1})$$

Comme on l' a vu (**) les estimateurs des paramètres d'une telle relation dépendent du rapport des variances. On peut prendre ce rapport égal à un. Si u_t est normal on obtient un estimateur du maximum de vraisemblance pour λ en résolvant ;

$$(39) \quad \lambda^2 \left[\frac{(\sum X_t Y_{t-1})(\sum X_t Y_t)}{\sum X_t^2} - \sum Y_t Y_{t-1} \right] + \lambda \left[\sum Y_t^2 - \sum Y_{t-1}^2 + \frac{(\sum X_t Y_{t-1})^2 (\sum X_t Y_t)^2}{\sum X_t^2} \right]$$

$$+ \sum Y_t Y_{t-1} - \left[\frac{(\sum X_t Y_{t-1})(\sum X_t Y_t)}{\sum X_t^2} \right] = 0$$

l'autre estimateur est donné par :

$$(40) \quad \text{est } \alpha = \frac{-\text{est } \lambda \sum X_t Y_{t-1} + \sum X_t Y_t}{\sum X_t^2}$$

On obtient ainsi des estimateurs en une étape au lieu de deux, ils sont identiques à ceux qui sont données par (37) pour $\xi = 0$. Klein (***) propose d'autre part une méthode itérative pour calculer des estimateurs lorsque $\xi \neq 0$. Toutes ces méthodes donnent des estimateurs biaisés pour les tailles d'échantillon que l'on rencontre habituellement en économétrie.

Koyck avait obtenu la relation (31) à partir de considérations techniques, institutionnelles, et psychologiques quant aux réactions des consommateurs. Supposons que la variable χ ait été au niveau X_0 pendant suffisamment longtemps pour que Y soit ajustée au niveau $Y_0 = \alpha X_0$. Supposons qu'à la période un X atteigne le niveau $X_0 + \Delta X$ et qu'elle conserve indéfiniment cette valeur. Le système de réactions ;

$$Y_1 - Y_0 = \alpha_0 \Delta X$$

(*) L. R. KLEIN "The estimation of distributed lags" Econometrica, Vol 26, n° 4, pp. 553-565, octobre 1958

(**) chapitre 1.

(***) op. cit. pp. 557-559 voir aussi : E. MALINVAUD "The estimation of distributed lags : a comment". Econometrica vol 29, n° 3, pp. 430-433, juillet 1961. Malinvaud montre qu'en général les estimateurs sont corrects.

$$Y_2 - Y_1 = \alpha_1 \Delta X$$
$$Y_3 - Y_2 = \alpha_2 \Delta X$$

...

donne

$$Y_t = (\alpha - \sum_{i=0}^{t-1} \alpha_i) X_0 + (\sum_{i=0}^{t-1} \alpha_i) (X_0 + \Delta X)$$

Si nous supposons que X prend à partir de la période 1 les valeurs X_1, X_2, X_3 ... en changeant à chaque période, alors le système des réactions s'écrit :

$$Y_1 - Y_0 = \alpha_0 (X_1 - X_0)$$

$$Y_2 - Y_1 = \alpha_0 (X_2 - X_1) + \alpha_1 (X_1 - X_0)$$

$$Y_3 - Y_2 = \alpha_0 (X_3 - X_2) + \alpha_1 (X_2 - X_1) + \alpha_2 (X_1 - X_0)$$

d'où

$$Y_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + (\alpha - \sum_{i=0}^{t-1} \alpha_i) X_0$$

Ce système suppose donc que $Y_t - Y_{t-1}$ est la somme d'une composante fonction de $X_t - X_{t-1}$ et d'une composante fonction de toutes les variations antérieures de la variable explicative.

Supposons que sur la base de la connaissance historique de X, on s'attende à un niveau futur vraisemblable égal à X^* et que l'on désire ajuster Y de telle sorte que :

(41) $Y^* = b X^*$

soit vraie. La formation de X^* peut-être décrite par :

(42) $X_t^* - X_{t-1}^* = \beta(X_t - X_{y-1}^*)$

relation qui admet un ajustement basé sur la valeur courante et la valeur prévue précédente, elle équivaut à :

(43)

$$X_t^* = \sum_{\lambda=0}^{\infty} \beta(1-\beta)^\lambda X_{t-\lambda}$$

Il peut y avoir aussi des décalages dans l'ajustement de Y_t à Y_t^* :

(44) $Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1})$

ce qui équivaut à :

(45) $Y_t = \sum_{\lambda=0}^{\infty} \delta(1-\delta)^\lambda Y_{t-\lambda}^*$

La valeur attendue de X est donc une moyenne pondérée des valeurs réelles précédentes, alors que la valeur réelle de Y_t est une moyenne pondérée des valeurs attendues précédentes. Le modèle défini par (41), (42) et (45) peut s'écrire :

$$(46) \quad Y_t = b\,\beta\,\delta\Big\{X_t + \left[(1-\beta)+(1-\delta)\right] X_{t-1} + \left[(1-\beta)^2 + (1-\beta)(1-\delta) + (1-\delta)^2\right] X_{t-2} + \left[(1-\beta)^3 + (1-\beta)^2(1-\delta) + (1-\beta)(1-\delta)^2 + (1-\delta)^3\right] X_{t-3} + \dots\Big\}$$

équation difficile à estimer. On peut obtenir une autre réduction ;

$$Y_t = \delta Y_t^* + (1-\delta)\, Y_{t-1} \quad \text{(d'après (44))}$$

$$= b\,\delta X_t^* + (1-\delta)\, Y_{t-1} \quad \text{(d'après (41))}$$

$$= b\delta\left[\beta X_t + (1-\beta) X_{t-1}^*\right] + (1-\delta) Y_{t-1} \quad (42)$$

$$= b\delta\left[\beta X_t + (1-\beta)\left(\frac{1}{b\,\delta} Y_{t-1} - \frac{1-\delta}{b\,\delta} Y_{t-2}\right)\right] + (1-\delta)\, Y_{t-1}$$

$$(47) \quad Y_t = b\,\beta\,\delta X_t + \left[(1-\beta) + (1-\delta)\right] Y_{t-1} - (1-\beta)(1-\delta)\, Y_{t-2}$$

L'estimation de (47) n'est pas trop difficile, cependant on ne peut pas estimer séparément β et δ qui entrent symétriquement dans la relation. On peut estimer b, $\beta\,\delta$ et $\beta + \delta$. Nerlove (*) discute en détails ces modèles et les problèmes qu'ils posent.

4 - Variables fictives

De nombreuses recherches économétriques introduisent des variables fictives dans les analyses de régression. Elles servent à représenter des effets transitoires comme ceux des guerres, les différences entre saisons... Elles servent aussi à représenter des variables qualitatives comme le sexe, l'état matrimonial, la catégorie socio-professionnelle, ou quantitatives lorsqu'il y a une classification par exemple en classes d'âge pour l'âge.

Comme illustration, soit une fonction de consommation linéaire ;

(48 a) $C = \alpha_1 + \beta Y$ temps de guerre

(48 b) $C = \alpha_2 + \beta Y$ temps de paix

avec $\alpha_2 > \alpha_1$

(*) Nerlove op. cit. "The dynamics of supply : estimation of farmer's response to price" John Hop-kins, Baltimore, 1958

On peut bien sur estimer séparément ces deux relations en utilisant les données correspondantes. Cependant si nous faisons l'hypothèse que la propension marginale à consommer est constante, il est préférable d'estimer β en utilisant toutes les données, aussi combinons (48 a) et (48 b) :

$$(49) \quad C = \alpha_1 X_1 + \alpha_2 X_2 + \beta Y$$

où les X sont des variables fictives définies par :

$$X_1 \begin{cases} 1 & \text{temps de guerre} \\ 0 & \text{temps de paix} \end{cases} \qquad X_2 \begin{cases} 0 & \text{temps de guerre} \\ 1 & \text{temps de paix} \end{cases}$$

Mais il n'est pas possible d'estimer directement (49) car la matrice n'est pas inversible en effet, soit m le nombre de périodes de guerre, n le nombre de périodes de paix :

$$\begin{bmatrix} m+n & m & n & \sum_{i=1}^{m+n} Y_i \\ m & m & 0 & \sum_{i=1}^{m} Y_i \\ n & 0 & n & \sum_{i=m=1}^{m+n} Y_i \\ \sum_{i=1}^{m+n} Y_i & \sum_{i=1}^{m} Y_i & \sum_{i=m+1}^{m+n} Y_i & \sum_{i=1}^{m+n} Y_i^2 \end{bmatrix}$$

est singulière. On modifie alors (49), en éliminant une variable fictive :

$$(50) \quad C = \gamma_1 + \gamma_2 X_2 + \beta Y$$

$$X_2 = \begin{cases} 0 & \text{temps de guerre} \\ 1 & \text{temps de paix} \end{cases}$$

d'où

$$(51) \quad \gamma_1 = \alpha_1 \quad \text{et} \quad \gamma_2 = \alpha_2 - \alpha_1$$

Remarquons que l'on peut aussi supposer que les propensions marginales à consommer sont différentes :

$$(52) \quad C = \gamma_1 + \gamma_2 X_2 + \beta_1 Y + \beta_2 Z$$

$$Z = X_2 Y$$

$$X_2 = \begin{cases} 0 & \text{temps de guerre} \\ 1 & \text{temps de paix} \end{cases}$$

Ce qui donne en temps de guerre :

$$C = \gamma_1 + \beta_1 Y$$

et en temps de paix :

$$C = (\gamma_1 + \gamma_2) + (\beta_1 + \beta_2)\, Y$$

ces relations contiennent dans leur second membre des variables fictives et des variables explicatives habituelles. On peut aussi n'avoir que des variables fictives. Supposons par exemple que la variable expliquée Y soit le nombre d'heures passées par an à lire des romans qui ne sont pas de fiction et qu'elle dépende de deux variables qualitatives : le sexe et le niveau d'instruction. Le sexe est représenté par une variable fictive à deux niveaux, le degré d'instruction par une variable fictive à trois niveaux, il y a ainsi six cas possibles. Raisonnons sur le modèle suivant :

$$(53) \quad Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_6 X_6 + u$$

où les X_i sont définies par :

niveau d'instruction	sexe	X_1	X_2	X_3	X_4	X_5	X_6
I	I	1	0	0	0	0	0
I	II	1	0	0	1	0	0
II	I	1	1	0	0	0	0
II	II	1	1	0	1	1	0
III	I	1	0	1	0	0	0
III	II	1	0	1	1	0	1

Si E (Y/II,I) désigne l'espérance de Y sachant que le niveau d'instruction est le niveau II et que le sexe est I, nous avons :

$$E\,(Y/I,I) = \beta_1$$

$$E\,(Y/I,II) = \beta_1 + \beta_4$$

$$E\,(Y/II,I) = \beta_1 + \beta_2$$

$$E\,(Y/II,II) = \beta_1 + \beta_2 + \beta_4 + \beta_5$$

$$E(\,Y/III,I) = \beta_1 + \beta_3$$

$$E\,(Y/III,II) = \beta_1 + \beta_3 + \beta_4 + \beta_6$$

Ce modèle permet d'introduire les interactions. Ainsi la différence entre sexe pour une instruction de niveau I est égale à β_4, pour une instruction de niveau II à $\beta_4 + \beta_5$ et à $\beta_4 + \beta_6$ pour le niveau III. De la même façon pour

le sexe I, la différence entre le niveau d'instruction II et le niveau I est β_2 entre III et I β_3, entre III et II $\beta_3 - \beta_2$. Pour le sexe II on a les différences $\beta_2 + \beta_5$, $\beta_3 + \beta_6$, $\beta_3 + \beta_6 - \beta_2 - \beta_5$

On peut considérer des cas où la variable expliquée revêt ces formes :

- posséder ou non une automobile
- avoir une maison hypothèque ou non..

Dans tous les cas, la présence est notée par l'unité, l'absence par un zéro. Si nous faisons une régression multiple de Y par rapport à des variables explicatives X, nous pouvons interpréter la valeur calculée de Y, pour des X donnés comme un estimé de la probabilité conditionnelle de Y pour les X donnés. En économétrie, cette approche est très utilisée par l'équipe de Orcutt du Social Systems Research Institute de l'Université de Wisconsin (*). Elle essaie d'intégrer des variables sociologiques à côté de variables économiques dans l'étude dynamique des systèmes socio-économiques, ce qui conduit à utiliser des variables fictives.

Pour étudier les dettes hypothécaires des unités de consommation aux Etats-Unis, Orcutt et Rivlin ont procédé en deux étapes :

(i) estimation d'une équation pour estimer la probabilité qu'une unité de consommation ait une dette hypothécaire,

(ii) estimation d'une équation pour estimer le montant de la dette hypothécaire pour les unites de consommation qui ont une dette de cette nature.

La première équation, obtenue à partir des données de Survey of Consumer Finance, s'écrit :

$$\hat{Y}_1 = -0{,}08 + F_{1,1} + F_{1.2} + F_{1.3} + F_{1.4}$$

ou $\hat{Y}_1$ est la valeur estimée de la probabilité pour une unité de consommation d'avoir une dette hypothècaire et où les fonctions sont données par ;

(*) Guy Orcutt, Martin Greenberger, John Korbel et Alice M Rivlin, "Microanalysis of socio économic systems : a simulation study" Harper and Row, New-York, 1961.

Etat matrimonial	$F_{1.1}$
célibataire	0
durée du mariage : 0 à 2 ans	0,07
" 2 à 3 ans	0,15
" 3 à 4 ans	0,19
" 4 à 5 ans	0,27
" 5 à 9 ans	0,31
" 10 à 20 ans	0,25
" plus de 20 ans	0,14

Age (en années)	$F_{1.2}$
18 - 20	0
21 - 24	0,02
25 - 29	0,09
30 - 34	0,16
35 - 39	0,25
40 - 44	0,22
45 - 49	0,24
50 - 54	0,16
55 - 59	0,15
60 - 64	0,12
plus de 65	0,07

niveau d'instruction	$F_{1.3}$
sans	0
primaire	0,02
secondaire	0,05
diplôme secondaire	0,08
supérieur	0,12
diplôme supérieur	0,13

race	$F_{1.4}$
blanc	0
noir	-0,10
autre	-0,14

Ainsi la probabilité qu'un non de 32 ans ayant fréquenté l'école secondaire, marié depuis 3 ans ait une dette hypothécaire est estimée par :

$$\hat{Y}_1 = -0,08 + 0,19 + 0,16 + 0,05 - 0,10 = 0,22$$

Pour les unités de consommation ayant une dette hypothécaire son montant est estimé par :

$$\hat{Y}_2 = 3,040 + F_{2.1} + F_{2.2} + F_{2.3} + F_{2.4}$$

où les $F_{2.i}$ sont donnés par :

état matrimonial	$F_{2.1}$
célibataire	0
durée du mariage : 0 à 2ans	4 550
2 à 3 ans	1 380
3 à 4 ans	1 530
4 à 5 ans	167
5 à 9 ans	49
10 à 20 ans	744
plus de 20 ans	215

Age (en années)	$F_{2.2}$
18 - 20	0
21 - 24	526
25 - 29	1 290
30 - 34	1 560
35 - 39	1 200
40 - 44	964
45 - 49	134
50 - 54	166
55 - 59	-1 270
60 - 64	-1 1 40
plus de 65	-1 000

niveau d'instruction	$F_{2.3}$
sans	0
primaire	335
secondaire	1 190
diplôme secondaire	1 980
supérieur	3 610
diplôme supérieur	4 930

Race	$F_{2.4}$
blanc	0
noir	-1 040
autre	-1 010

Le montant de la dette hypothécaire pour l'unité de consommation précédente est estimé par :

$$\hat{Y}_2 = 3,040 + 1530 + 1\,560 + 1\,190 - 1\,040 = 6280 \text{ dollars.}$$

Chacune des équations précédentes est estimée en utilisant la méthode des moindres carrés et en introduisant des variables fictives, sept par exemple pour l'état matrimonial. Le modèle utilisé suppose que les effets sont additifs, rien ne s'oppose à ce que soient introduits des interactions.

Goldberger a montré que si Y est une variable dichotomique les erreurs ne sont pas homoscédastiques. Soit X' le vecteur des observations à l'instant t (c'est la tième ligne de X) nous avons :

(54) $u_t = Y_t - X'_t\beta$

comme Y_t est égal soit à 0, soit à 1, u_t prend soit la valeur $- X'_t\beta$ soit la valeur $1 - X'_t\beta$. La distribution de u_t sachant X'_t est alors :

u_t	Prob (u_t)
$- X'_t\beta$	$1 - X'_t\beta$
$1 - X'_t\beta$	$X'_t\beta$

si $E(u_t) = 0$. D'où :

(55) $\mathrm{Var}(u_t) = E(u_t^2) = (X'_t\beta)(1-X'_t\beta)$

on en déduit la matrice des variances - covariances de u :

$$(56)\quad E(uu') = \begin{bmatrix} X'_1\beta(1-X'_1\beta) & & & 0 \\ & X'_2\beta(1-X'_2\beta) & & \\ & & \ddots & \\ 0 & & & X'_n\beta(1-X'_n\beta) \end{bmatrix}$$

aussi pour obtenir le meilleur estimateur linéaire sans biais doit- on utiliser cette matrice dans (11). β étant inconnu Golberger propose une technique en deux étapes : il substitue dans (11) l'expression de (56) obtenue en remplaçant β par son estimé des moindres carrés simples.

Chapitre 4 L'ESTIMATION DANS LES MODELES A EQUATIONS MULTIPLES

1 - Généralités

Jusqu'à maintenant nous nous sommes intéressés à des modèles ne comportant qu'une seule équation. Nous allons étendre notre analyse dans deux directions :

(i) estimation d'une équation dans un modèle à équations multiples,

(ii) estimation de toutes les équations dans un modèle à équations multiples.

Soit le modèle :

(1) $C_t = \alpha + \beta Y_t + u_t$

(2) $Y_t = C_t + Z_t$

où : C = consommation,
Y = revenu,
Z = épargne,
u = terme aléatoire,
t = période de temps.

L'équation (1) est une fonction de consommation et la relation (2) exprime une identité. On considère que l'épargne est une variable exogène et que la consommation et le revenu sont endogènes : leurs valeurs sont déterminées par les relations du modèle à partir de celles de la variable exogène.

Introduisons une troisième équation :

$$Z_t = \gamma (Y_{t-1} - Y_{t-2}) + \delta r_t + v_t$$

où r_t désigne le taux d'intérêt. On définit alors un modèle à trois équations où le taux d'intérêt est exogène et où l'épargne est endogène. Il y a autant d'équations dans le modèle que de variables endogènes. La classification en

variables endogènes et variables exogènes est très relative, elle dépend du système étudié et des objectifs pour lesquels le modèle est construit.

Raisonnons sur le modèle à deux équations, dans lequel Z est exogène. On introduit les hypothèses suivantes sur le résidu :

(3-a) $E(u_t) = 0 \quad \forall t$

(3-b) $E(u_t\, u_{t+s}) = \begin{cases} 0 & \forall t, \forall s \neq 0 \\ \sigma^2 & \forall t,\ s = \theta \end{cases}$

(3-c) Z et u sont indépendantes

L'hypothèse (3-c) est en particulier vérifiée si les valeurs de Z sont fixées.

Pour estimer la fonction de consommation (1), on peut penser appliquer les moindres carrés à la relation entre C et Y. Les erreurs sont homoscédastiques (3-a) et ne sont pas autocorrélées (3-b). Il faut étudier la dépendance entre u et Y. En substituant (1) dans (2), nous obtenons :

$$Y_t = \alpha + \beta Y_t + Z_t + u_t$$

$$Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + \frac{u_t}{1-\beta}$$

donc Y_t dépend de u_t.

$$E(Y_t) = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t$$

$$E\left\{u_t \left[Y_t - E(Y_t)\right]\right\} = \frac{1}{1-\beta} E(u_t^2) \neq 0$$

Dans la fonction de consommation, revenu et résidu sont dépendants, nous sommes donc dans le cas d'un modèle à erreurs sur les variables. On ne peut appliquer directement les moindre carrés simples. Introduisons les moments empiriques du second ordre :

$$m_{CY} = \frac{1}{n} \sum_{t=1}^{n} (C_t - \overline{C})(Y_t - \overline{Y}) \qquad m_{YY},\ m_{Y2},\ m_{ZZ},\ m_{Zu} \text{ et } m_{uu}.$$

Les estimateurs des moindres carrés de α et β sont :

$$\hat{\beta} = \frac{m_{CY}}{m_{YY}} \quad \text{et} \quad \hat{\alpha} = \overline{C} - \hat{\beta}\ \overline{Y}$$

Résolvons les équations (1) et (2) par rapport aux variables endogènes :

$$\boxed{\begin{aligned} &(4) \qquad C_t = \frac{\beta}{1-\beta} Z_t + \frac{\alpha}{1-\beta} + \frac{u_t}{1-\beta} \\ &(5) \qquad Y_t = \frac{1}{1-\beta} Z_t + \frac{\alpha}{1-\beta} + \frac{u_t}{1-\beta} \end{aligned}}$$

ce qui s'écrit encore :

$$C_t - \overline{C} = \frac{\beta}{1-\beta}(Z_t - \overline{Z}) + \frac{1}{1-\beta}(u_t - \overline{u})$$

$$Y_t - \overline{Y} = \frac{1}{1-\beta}(Z_t - \overline{Z}) + \frac{1}{1-\beta}(u_t - \overline{u})$$

d'où :

$$m_{CY} = \frac{\beta}{(1-\beta)^2} m_{ZZ} + \frac{1+\beta}{(1-\beta)^2} m_{Zu} + \frac{1}{(1-\beta)^2} m_{uu}$$

$$m_{YY} = \frac{1}{(1-\beta)^2} m_{ZZ} + \frac{2}{(1-\beta)^2} m_{Zu} + \frac{1}{(1-\beta)^2} m_{uu}$$

ce qui donne :

$$\hat{\beta} = \frac{\beta\, m_{ZZ} + (1+\beta)\, m_{Zu} + m_{uu}}{m_{ZZ} + 2\, m_{Zu} + m_{uu}}$$

En supposant que lorsque n tend vers l'infini :

$$m_{uu} \to \sigma^2$$

$$m_{Zu} \to 0$$

$$m_{ZZ} \to \overline{m}_{ZZ}$$

on obtient :*

$$\underset{n \to \infty}{\text{plim}}\ \hat{\beta} = \frac{\beta\, \overline{m}_{ZZ} + \sigma^2}{\overline{m}_{ZZ} + \sigma^2} = \beta + \frac{(1-\beta)\, \sigma^2 / \overline{m}_{ZZ}}{1 + \sigma^2 / \overline{m}_{ZZ}}$$

---

* En effet $\text{p lim}\ \frac{P}{Q} = \frac{\text{plim}\ P}{\text{plim}\ Q}$

Donc si $0 < \beta < 1$ alors $\underset{n \to \infty}{p\lim} \hat{\beta} > \beta$. On a un biais positif qui n'est pas éliminé en augmentant la taille de l'échantillon.

Considérons une autre méthode d'estimation basée sur les équations (4) et (5) qui sont appelées <u>forme réduite</u> du modèle. Lorsqu'un modèle est mis sous forme réduite, il n'apparaît dans chaque équation qu'une seule variable endogène. Etant les hypothèses (3) on peut appliquer la méthode simple des moindres carrés aux équations réduites.

(6) $\frac{\beta}{1-\beta}$ est estimé par $\frac{m_{CZ}}{m_{ZZ}}$

(7) $\frac{1}{1-\beta}$ est estimé par $\frac{m_{YZ}}{m_{ZZ}}$

(8) $\frac{\alpha}{1-\beta}$ est estimé par $\overline{C} - \frac{m_{CZ}}{m_{ZZ}} \overline{Z}$

(9) $\frac{\alpha}{1-\beta}$ est estimé par $\overline{Y} - \frac{m_{YZ}}{m_{ZZ}} \overline{Z}$

et ce sont les meilleurs estimateurs linéaires sans biais. (6) ou (7) donne un estimateur de β, (8) ou (9) de α. De (6) on tire :

$$\beta^* = \frac{m_{CZ}}{m_{ZZ} + m_{CZ}}$$

et de (7) :

$$\beta^* = \frac{m_{YZ} - m_{ZZ}}{m_{YZ}}$$

montrons que ces deux estimateurs sont identiques :

$$Y = C + Z$$

donc $m_{YZ} = m_{CZ} + m_{ZZ}$

et $$\frac{m_{YZ} - m_{ZZ}}{m_{YZ}} = \frac{m_{CZ}}{m_{CZ} + m_{ZZ}} \quad \text{d'où :} \quad \boxed{\beta^* = \frac{m_{CZ}}{m_{YZ}}} \qquad (10)$$

ce qui prouve l'identité. Les relations (8) et (9) fournissent aussi des estimateurs identiques de α.

(11) $$\boxed{\alpha^* = \frac{m_{ZZ}\,\overline{C} - m_{CZ}\,\overline{Z}}{m_{YZ}}}$$

Les estimateurs (10) et (11) qui ont été obtenus à partir d'estimateurs sans biais des paramètres des équations réduites, sont des estimateurs biaisés des paramètres structuraux α et β du modèle défini par (1) et (2). Ce sont par contre des estimateurs corrects. Montrons ces résultats pour β^* .

(12) $$\beta^* = \frac{m_{CZ}}{m_{YZ}} = \frac{\beta\, m_{ZZ} + m_{Zu}}{m_{ZZ} + m_{Zu}}$$ d'après (4) et (5).

Lorsque $n \to \infty$, $m_{Zu} \to o$ donc :

(13) $$\operatorname*{plim}_{n \to \infty} \beta^* = \beta$$

Calculons son biais :

$$E(\beta^*) = E\left(\frac{\beta\, m_{ZZ} + m_{Zu}}{m_{ZZ} + m_{Zu}}\right) = E\left(\frac{\beta + m_{Zu} / m_{ZZ}}{1 + m_{Zu} / m_{ZZ}}\right)$$

prenons les Z constants, ce qui entraîne $E(m_{Zu}) = 0$ mais non $E(\beta^*) = 0$.
Pour s'en convaincre considérons l'exemple suivant :

m_{Zu} / m_{ZZ}	probabilité	β^*	$\beta = 0,5$
- 0,2	0,25	3/8 = 0,3750	
- 0,1	0,25	4/9 = 0,4440	
0,1	0,25	6/11 = 0,5454	
0,2	0,25	7/12 = 0,5833	
		$E(\beta^*) = 0,4870$	

La méthode d'estimation que nous venons d'utiliser s'appelle méthode indirecte des moindres carrés ; elle procède en deux étapes :

(i) estimation des paramètres des équations réduites par la méthode simple des moindres carrés,

(ii) estimation des paramètres structuraux,

comme nous le verrons par la suite, elle ne peut être appliquée que dans certains cas spéciaux.

Moindres carrés en deux étapes (ou doubles moindres carrés)

Dans le modèle précédent, la difficulté fondamentale est due à la corrélation entre le terme résiduel et la variable explicative dans la fonction de consommation (1). La méthode indirecte des moindres carrés fournit une méthode possible d'estimation, la méthode des moindres carrés en deux étapes est d'application plus générale. Nous en verrons une description assez complète, auparavant présentons en l'idée fondamentale.

Nous désirons éliminer de la variable explicative Y l'influence du terme résiduel u. Ce résultat peut être obtenu de la manière suivante :

(i) Déterminer la régression de Y par rapport à la seule variable exogène Z, et remplacer Y par son estimateur fonction de Z.

(ii) Appliquer la méthode simple des moindres carrés à la relation qui résulte de (i).

La première étape donne donc :

(14) $$Y = \hat{\Pi}_1 + \hat{\Pi}_2 Z + e$$

(15) où $$\begin{cases} \hat{\Pi}_2 = \dfrac{m_{YZ}}{m_{ZZ}} \\ \hat{\Pi}_1 = \bar{Y} - \hat{\Pi}_2 \bar{Z} \end{cases}$$

et (16) $$\hat{Y} = \hat{\Pi}_1 + \hat{\Pi}_2 Z$$

pour estimateur de Y. En ce qui concerne la seconde étape nous obtenons :

(17) $$C = \alpha + \beta (\hat{Y}) + (u + \beta e)$$

sous cette forme la relation (1) est telle que la variable explicative est indépendante du résidu. Pour finir, nous avons :

$$\hat{\beta} = \frac{m_{C\hat{Y}}}{m_{\hat{Y}\hat{Y}}}$$

D'après (14) : $\hat{y} = \hat{\Pi}_2 z$,

$$m_{C\hat{Y}} = \hat{\Pi}_2^2 m_{CZ}$$

$$m_{\hat{Y}\hat{Y}} = \hat{\Pi}_2 m_{ZZ}$$

$$\hat{\beta} = \frac{m_{CZ}}{\hat{\Pi}_2 \, m_{ZZ}}$$

$$= \frac{m_{CZ}}{m_{YZ}} \text{ d'après (15)}$$

on obtient la même expression que par la méthode indirecte des moindres carrés (10). On en déduit que l'estimateur de α obtenu par la méthode indirecte est identique à celui obtenu par la méthode en deux étapes.

Le rapport de la plus petite variance

Tel qu'il est spécifié par la relation (1), le modèle implique que la consommation est déterminée par le revenu et que l'épargne Z n'intervient pas dans la fonction de consommation. Comparons :

(18 -a) $C = \alpha + \beta Y + u$

(18 -b) $C = \alpha + \beta Y + \gamma Z + v$

on a $\gamma = 0$, cependant à partir d'un échantillon de taille finie l'inclusion de Z comme variable explicative conduira à un coefficient de l'épargne différent de zéro et à une variance résiduelle plus faible.

Le principe du rapport de la plus petite variance pose que les estimateurs des paramètres α et β doivent être choisis de telle sorte que le rapport de la variance résiduelle de (18-a) à celle de (18-b) soit minimum. Définissons ;

(19) $C^* = C - (\alpha^* + \beta^* Y)$

où α^* et β^* représentent les estimateurs obtenues par la méthode du rapport de la plus petite variance.

En introduisant les variables réduites, nous obtenons ;

(20) $c^* = c - \beta^* y$

La somme des carrés des résidus de (18 - a) est alors $\sum c^{*2}$. En utilisant la définition de C^*, on peut considérer que (18-b) est une relation entre C^* et Z. Z étant exogène, on peut lui appliquer les moindres carrés et la somme des carrés des résidus de (18-b) est alors :

$$\sum (c^* - \hat{\gamma} z)^2$$

où $\hat{\gamma}$ est l'estimateur des moindres carrés :

$$\hat{\gamma} = \frac{\sum c^* z}{\sum z^2}$$

enfin $\sum (c^* - \hat{\gamma} z)^2 = \sum c^{*2} - \frac{(\sum c^* z)^2}{\sum z^2}$

et le rapport des variances qu'il faut minimiser s'écrit ;

$$(21) \qquad 1 = \frac{\sum c^{*2}}{\sum c^{*2} - (\sum c^* z)^2 / \sum z^2}$$

le minimum de ce rapport est égal à un et il est atteint si et seulement si : $\sum c^* z = 0$, ce qui donne d'après (20) :

$$\sum (c - \beta^* y) z = 0$$

c'est-à-dire :

$$\beta^* = \frac{m_{CZ}}{m_{YZ}}$$

ainsi dans cet exemple simple les trois méthodes d'estimations :

(i) méthode indirecte des moindres carrés,

(ii) méthode des moindres carrés en deux étapes,

(iii) méthode du rapport de la plus petite variance,

conduisent à des estimateurs identiques. Comme nous le verrons ceci est dû au modèle considéré.

Les relations (1) et (2) sont les relations structurelles du modèle, la première est une relation de comportement qui traduit une hypothèse quant à la fonction de consommation tandis que la seconde est une simple identité. Ce modèle comprend deux relations structurelles et deux variables endogènes, les hypothèses sur les résidus complètent sa spécification. Un ensemble de valeurs numériques données aux paramètres inconnus α, β et σ_u^2 détermine une structure particulière du modèle. Les relations (4) et (5) en constituent la forme réduite, elles explicitent les variables endogènes comme des fonctions des autres variables. Dans la forme réduite les résidus sont des formes linéaires des résidus originaux. Les trois méthodes d'estimation introduites ont conduit à des résultats identiques, mais ce n'est pas en général vrai. Mise à part la brève discussion sur la transformation des estimateurs des paramètres de la forme réduite des estimateurs des paramètres structurels, nous n'avons pas encore abordé le délicat problème de l'identification. Nous verrons que lorsqu'il y a sous-identification, il est impossible d'estimer certains paramètres, par contre lorsqu'il y a identification exacte les trois méthodes d'estimation introduites ci-dessus conduisent aux mêmes résultats,

enfin un cas de sur-identification on ne peut pas utiliser la méthode indirecte des moindres carrés, si les deux autres méthodes sont applicables elles donnent des résultats différents.

Formulation Générale

Posons le modèle à équations simultanées dans toute sa généralité. Soit un modèle linéaire à G relations structurelles, la ième relation s'écrit, pour l'instant t :

$$(22) \qquad \beta_{i1}\, y_{1t} + \beta_{i2}\, y_{2t} + \dots + \beta_{iG}\, y_{Gt} + \gamma_{i1}\, x_{1t} + \dots + \gamma_{ik} x_{kt} = u_{it} \quad (*)$$

$$(t = 1, 2 \dots n)$$

où les y_{it} sont les variables endogènes, les x_{it} des variables exogènes ou des valeurs décalées des variables endogènes. Ces deux derniers groupes de variables forment l'ensemble des variables prédéterminées. Le modèle représente une théorie qui explique la détermination conjointe des variables dépendantes y_{it} ($i = 1, 2 \dots G$; $t = 1, 2, \dots n$) à partir des variables prédéterminées x_{it} ($i = 1, \dots K$; $t = 1, \dots n$) et des résidus u_{it} ($i = 1, \dots G$; $t = 1, \dots n$). En général on pose que certains des β et γ sont nuls, si on ne le faisait pas l'estimation du modèle serait impossible car on ne pourrait pas distinguer ses relations les unes des autres. Sous forme matricielle, le modèle s'écrit :

$$(23) \qquad B\, y_t + \Gamma\, x_t = u_t$$

B = matrice G x G des coefficients des variables endogènes,

Γ = matrice G x K des coefficients des variables prédéterminées;

y_t , x_t et u_t sont des vecteurs colonnes à G1,K et G éléments respectivement.

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \\ y_{Gt} \end{bmatrix} \qquad x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \\ x_{kt} \end{bmatrix} \qquad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \\ u_{Gt} \end{bmatrix}$$

Si nous supposons que B est régulière, la forme réduite du modèle peut être écrite :

$$(24) \qquad y_t = \Pi\, x_t + v_t$$

(*) nous notons par des miniscules les valeurs actuelles des variables.

(25) $\Pi = - B^{-1} \Gamma \quad v_t = B^{-1} u_t$

Π = matrice G x K des coefficients de la forme réduite,

v_t = vecteur colonne à G éléments, des résidus de la forme réduite.

2 - Identification

L'estimation des paramètres du modèle (23) ne peut se faire qu'en introduisant des hypothèses sur la distribution de probabilité des résidus. Supposons que l'on ait :

(26) $E(u_t) = 0$

(27) $$E(u_t \, u'_t) = \Phi = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1G} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2G} \\ & & & \\ \sigma_{G1} & \sigma_{G2} & . & \sigma_{GG} \end{bmatrix}$$

Φ est la matrice G x G des variances-covariances des u_{it}. On a supposé que Φ est indépendante du temps, mais aucune hypothèse n'est faite sur les éléments n'appartenant pas à la diagonale principale. Les résidus de différentes relations peuvent donc être liés. A toute identité, exprimée par le modèle, il correspond une ligne et une colonne de Φ nulles, ce qui entraine que Φ est singulière. Si le modèle a été résolu de telle sorte que les identités aient été éliminées, on peut supposer que les résidus sont linéairement indépendants donc que Φ est régulière. Supposons en outre que les résidus ne sont pas autocorrelées donc que :

(28) $p(u_t \, u_{t+s}) = p(u_t) \, p(u_{t+s})$

Considérons maintenant la relation (23) :

(23) $B \, y_t + \Gamma \, x_t = u_t$

u_t est indépendante de x_t, x_{t-1}, x_{t-2}... Ceci est évident pour les variables exogènes, pour les variables prédéterminées comme y_{t-1}, elles dépendent de u_{t-1},... qui sont indépendantes de u_t d'après (28).

On peut calculer la probabilité conditionnelle de y_t sachant x_t :

$$p\,(y_t \,/\, x_t) = p(u_t \,/\, x_t)\left|\frac{\partial\, u_t}{\partial\, y_t}\right|$$

$$(29) \qquad p\,(y_t \,/\, x_t) = p(u_t)\left|\frac{\partial\, u_t}{\partial\, y_t}\right|$$

où $\left|\frac{\partial u_t}{\partial y_t}\right|$ désigne la valeur absolue du déterminant de la matrice :

$$(30) \qquad \begin{bmatrix} \frac{\partial u_1 t}{\partial y_1 t} & \frac{\partial u_1 t}{\partial y_2 t} & \cdots & \frac{\partial u_1 t}{\partial y_G t} \\ \frac{\partial u_2 t}{\partial y_1 t} & \frac{\partial u_2 t}{\partial y_2 t} & \cdots & \frac{\partial u_2 t}{\partial y_G t} \\ \frac{\partial u_G t}{\partial y_1 t} & \frac{\partial u_G t}{\partial y_2 t} & \cdots & \frac{\partial u_G t}{\partial y_G t} \end{bmatrix}$$

cette matrice est simplement la matrice B donc :

$$(31) \qquad p\,(y_t \,/\, x_t) = |\det B| \;\; p(u_t)$$

En utilisant l'hypothèse exprimée par (28) ; on calcule la fonction de vraisemblance d'un échantillon de variables endogènes sachant les x_t :

$$(32) \qquad p(y_1, y_2, \ldots y_n \,/\, x_1, x_2, \ldots x_n) = |\det B|^{\,n} p(u_1)\, p(u_2) \ldots p(u_n)$$

Transformons la relation (23) en multipliant ses deux membres par une matrice régulière A. Ce qui revient à remplacer chaque équation de la structure initiale par une combinaison linéaire des équations de cette structure. La nouvelle structure s'exprime par :

$$(33) \qquad (AB)\, y_t + (A\,\Gamma)\, x_t = w_t$$

$$(34) \qquad w_t = A\, u_t$$

La matrice des variances-covariances de w_t est :

$$E\,(w_t\, w'_t) = E\,(A\, u_t\, u'_t\, A')$$

$$(35) \qquad E\,(w_t\, w'_t) = A\, \Phi\, A'$$

et on a :

$$p(w_t) = p(u_t)\left|\frac{\partial u_t}{\partial w_t}\right|$$

$$(36) \qquad p(w_t) = |\det A^{-1}|p(u_t)$$

La fonction de vraisemblance conditionnelle d'un échantillon $y_1, y_2, \dots y_n$ calculée à partir de la structure (33) s'écrit :

$$p(y_1, \dots y_n / x_1, \dots x_n) = |\det (AB)|^n \, p(w_1)\, p(w_2) \dots p(w_n)$$
$$= |\det A|^n \, |\det B|^n \, |\det A^{-1}|^n \, p(u_1)\, p(u_2) \dots p(u_n)$$

car $\det AB = \det A \times \det B$
et comme $\det A^{-1} = (\det A)^{-1}$, on obtient :

$$p(y_1, \dots y_n/x_1, \dots x_n) = |\det B|^n \, p(u_1) \; p(u_2) \dots p(u_n)$$

Les structures exprimées par les relations (23) et (33) sont donc équivalentes, du point de vue de l'observation : on dit que <u>ces structures ne sont pas identifiables</u>. Quand les structures ne sont pas identifiables, il est impossible de déterminer à partir des observations, tous les paramètres introduits dans la représentation du phénomène. Les caractéristiques internes du processus générateur des données ne seront pas connues complètement, on ne peut préciser exactement le modèle. Si nous calculons la forme réduite du modèle (33) nous constatons qu'elle est égale à celle du modèle (23), en effet :

$$(AB)^{-1} (AB)\, y_t + (AB)^{-1} (A\Gamma) x_t = (AB)^{-1} (Au_t)$$

équivaut à :

$$y_t + B^{-1} \Gamma x_t = B^{-1} u_t$$

La fonction de vraisemblance peut être écrite sous la forme :

$$p(y_1, \dots y_n/x_1, \dots x_n) = p(v_1)\, p(v_2) \dots p(v_n)$$

Comme $v_t = y_t - \Pi x_t$, les paramètres de la fonction de vraisemblance sont les éléments de la matrice Π des coefficients de la forme réduite. Les éléments de la matrice des variances-covariances des résidus de la forme réduite sont $B^{-1} \; \Phi(B^{-1})'$. <u>Le problème de l'identification revient à déduire les valeurs des paramètres des relations structurelles à partir d'une connaissance des paramètres de la forme réduite.</u>

Poser ainsi, l'identification est un problème insoluble. Si aucune hypothèse n'est faite sur les matrices B et Γ, les paramètres de la matrice Π de la forme réduite représentent un vaste ensemble de structures obtenues à partir de la structure initiale par multiplication par une matrice arbitraire A. L'identification n'est possible qu'en posant des hypothèses a priori sur B, Γ et Φ . Les plus simples expriment que certains de leurs éléments sont nuls. En d'autres termes, la spécification du modèle nécessite que certaines variables n'apparaissent pas dans certaines relations. Un modèle où toutes les variables apparaissent dans toutes les relations est une proposition statistique sans espoir. Cependant l'exclusion a priori de certaines variables, bien que nécéssaire , n'est pas une condition suffisante permettant l'identification comme le montrent les exemples suivants.

Exemple 1

Considérons le modèle ;

(37) $$y_{1t} + \beta_{12}\, y_{2t} + \gamma_{11}\, x_{1t} = u_{1t}$$

(38) $$\beta_{21}\, y_{1t} + y_{2t} + \gamma_{22}\, x_{2t} = u_{2t}$$

$$B = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \qquad \Phi = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

La forme réduite s'écrit :

(39) $$y_{1t} = \Pi_{11}\, x_{1t} + \Pi_{12}\, x_{2t} + v_{1t}$$

(40) $$y_{2t} = \Pi_{21}\, x_{1t} + \Pi_{22}\, x_{2t} + v_{2t}$$

En combinant (39) et (40) nous obtenons :

(41) $$y_{1t} + \beta_{12}\, y_{2t} = (\Pi_{11} + \beta_{12}\, \Pi_{21})\, x_{1t} + (\Pi_{12} + \beta_{12}\, \Pi_{22})\, x_{2t} + v_{1t} + \beta_{12}\, v_{2t}$$

comparons alors (41) et (37) nous avons nécéssairement ;

(42) $$-\gamma_{11} = \Pi_{11} + \beta_{12}\, \Pi_{21}$$

(43) $$0 = \Pi_{12} + \beta_{12}\, \Pi_{22}$$

et $$u_{1t} = v_{1t} + \beta_{12}\, v_{2t}$$

On en déduit :

(44) $$\beta_{12} = - \frac{\Pi_{12}}{\Pi_{22}}$$

(45) $$\gamma_{11} = \frac{\Pi_{12} \ \Pi_{21} - \Pi_{11} \ \Pi_{22}}{\Pi_{22}}$$

Les paramètres de (37) s'expriment de façon unique en fonction des éléments de Π. Il en est de même pour les paramètres de (38). Faisons le même calcul pour la matrice des variances-covariances.

$$v_t = B^{-1} u_t$$

(46) $$E(v_t \ v'_t) = \Psi = B^{-1} \ \Phi \ (B^{-1})'$$

(47) $$\Phi = B \ \Psi \ B'$$

on est donc dans un cas d'identification exacte.

Déterminons une autre structure compatible avec le modèle en prémultipliant (37) et (38) par une matrice régulière A d'ordre 2. Par compatible, nous entendons qu'aucune des hypothèses du modèle n'est violée : x_2 n'apparaît pas dans la première relation, x_1 dans la seconde ; le coefficient de y_1 est 1 dans la première relation celui de y_2 aussi mais dans la deuxième. On définit :

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

d'où :

$$AB = \begin{bmatrix} a_{11} + a_{12} \ \beta_{21} & a_{11} \ \beta_{12} + a_{12} \\ a_{21} + a_{22} \ \beta_{21} & a_{21} \ \beta_{12} + a_{22} \end{bmatrix} \qquad A\Gamma = \begin{bmatrix} a_{11} \ \gamma_{11} & a_{12} \ \gamma_{22} \\ a_{21} \ \gamma_{11} & a_{22} \ \gamma_{22} \end{bmatrix}$$

matrices des coefficients de la structure transformée. Les hypothèses quant aux coefficients de x_1 et x_2 impliquent que :

$$a_{12} = a_{21} = 0$$

celles concernant les coefficients de y_1 et y_2 :

$$a_{11} = 1 = a_{22}$$

La seule matrice A qui donne une structure compatible avec le modèle est donc la matrice unité. La forme réduite conduit à un ensemble unique de paramètres structurels pour le modèle. C'est un cas d'identification exacte.

Exemple 2 :

Soit le modèle :

(48) $y_{1t} + \beta_{12} y_{2t} = u_{1t}$

(49) $\beta_{21} y_{1t} + y_{2t} + \gamma_{21} x_{1t} = u_{2t}$

$$B = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 0 \\ \gamma_{21} \end{bmatrix} \qquad \Phi = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

En multipliant à gauche par une matrice A d'ordre 2, nous obtenons :

$$A\Gamma = \begin{bmatrix} a_{12} \gamma_{21} \\ a_{22} \gamma_{21} \end{bmatrix}$$

Les hypothèses du modèle conduisent à :

$$a_{12} \gamma_{21} = 0 \Longrightarrow a_{12} = 0$$

$$a_{11} + a_{12} \beta_{21} = 1 \Longrightarrow a_{11} = 1$$

$$a_{21} \beta_{12} + a_{22} = 1$$

D'où :

$$A = \begin{bmatrix} 1 & 0 \\ a_{21} & 1-a_{21} \beta_{12} \end{bmatrix}$$

avec a_{21} arbitraire.

Nous définissons ainsi un ensemble d'une infinité de matrices compatibles avec le modèle et nous avons :

$$AB = \begin{bmatrix} 1 & \beta_{12} \\ a_{21} (1-\beta_{12} \beta_{21}) + \beta_{21} & 1 \end{bmatrix} \qquad A\Gamma = \begin{bmatrix} 0 \\ (1-a_{21} \beta_{12}) \gamma_{21} \end{bmatrix}$$

Nous voyons que pour toutes les structures, les coefficients de la première relation restent inchangés, tandis que ceux de la seconde varient. Les paramètres de la première relation sont exactement identifiés alors que ceux de la seconde sont sous identifiés (ou indentifiés).

On peut mettre encore ce résultat en évidence en exprimant les paramètres structurels en fonction de ceux de la forme réduite. Pour la première relation il y a unicité, ce qui est faux en ce qui concerne la seconde. D'après (25) :

$$- B^{-1} \Gamma = \Pi$$

ce qui entraine

$$\frac{-1}{1 - \beta_{12} \beta_{21}} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \gamma_{21} \end{bmatrix} = \begin{bmatrix} \Pi_{11} \\ \Pi_{21} \end{bmatrix}$$

c'est-à-dire :

$$(50) \quad \begin{bmatrix} \dfrac{\beta_{12} \gamma_{21}}{\Delta} \\ \\ \dfrac{-\gamma_{21}}{\Delta} \end{bmatrix} = \begin{bmatrix} \Pi_{11} \\ \\ \Pi_{21} \end{bmatrix}$$

avec $\Delta = 1 - \beta_{12} \beta_{21}$.

$$\beta_{12} = \frac{-\Pi_{11}}{\Pi_{21}}$$

mais on ne peut pas déterminer β_{21} et γ_{21}.

On ne peut vérifier qu'il y a identification exacte en prenant :

$$\Phi = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$$

pour faire le calcul, on identifie Φ à la matrice $B\Psi B'$ où Ψ est la matrice des variances-covariances des résidus de la forme réduite.

Exemple 3

$$(51) \quad y_{1t} + \beta_{12} y_{2t} + \gamma_{11} x_{1t} = u_{1t}$$

$$(52) \quad \beta_{21} y_{1t} + y_{2t} + \gamma_{21} x_{1t} = u_{2t}$$

La matrice AB est la même que précédemment

$$\Gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \qquad A\Gamma = \begin{bmatrix} a_{11}\ \gamma_{11} + a_{12}\ \gamma_{21} \\ a_{21}\ \gamma_{11} + a_{22}\ \gamma_{22} \end{bmatrix}$$

les seules contraintes concernent les coefficients de y_1 et y_2 :

$$a_{11} + a_{12}\ \beta_{21} = 1$$
$$a_{21}\ \beta_{12} + a_{22} = 1$$

ces conditions laissent un élément arbitraire dans chaque ligne de A, aucune des relations n'est identifiée. Il est impossible d'exprimer les paramètres structurels en fonction de ceux de la forme réduite, d'après (25) :

$$\frac{\beta_{12}\ \gamma_{21} - \gamma_{11}}{\Delta} = \Pi_{11}$$

$$\frac{\beta_{21}\ \gamma_{11} - \gamma_{21}}{\Delta} = \Pi_{21}$$

Ce qui ne permet pas d'obtenir un seul des quatre paramètres.

Condition nécessaire et suffisante d'identification

Traitons de façon générale le problème de l'identification et intéressons nous à la première relation du modèle (23) :

$$(57) \qquad \beta_1\ y_t + \gamma_1\ x_t = u_{1t}$$

où β_1 désigne la première ligne de B , γ_1 la première ligne de Γ . Prémultiplions la forme réduite (24) par β_1 :

$$(58) \qquad \beta_1\ y_t = \beta_1\ \Pi\ x_t + \beta_1\ v_t$$

$$\beta_1\ v_t = \beta_1\ B^{-1} u_t = \left[\beta_{11} \cdots \beta_{1G}\right] \cdot \frac{1}{|B|} \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{G1} \\ b_{12} & b_{22} & \cdots & b_{G2} \\ & & & \\ b_{1G} & b_{2G} & \cdots & b_{GG} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ \\ u_{Gt} \end{bmatrix}$$

où b_{ij} désigne le cofacteur de β_{ij} .

$$\beta_1 v_t = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ \\ u_{Gt} \end{bmatrix} \quad (*)$$

$$= u_{1t}$$

Les coefficients de x_t des relations (57) et (58) doivent être identiques :

(59) $\quad -\gamma_1 = \beta_1 \Pi$

Soit G^{Δ} le nombre de variables endogènes et K^* le nombre de variables prédéterminées qui apparaissent dans la relation avec des coefficients différents de zéro. $G^{\Delta\Delta} = G - G^{\Delta}$ et $K^{**} = K - K^*$ sont donc le nombre de variables endogènes et prédéterminées respectivement qui sont exclues de la relation. Les variables peuvent être numérotées de façon à ce que les variables qui ne sont pas exclues aient les indices les plus petits, on partitionne donc les vecteurs β_1 et γ_1 comme suit :

$$\beta_1 = \begin{bmatrix} \beta_1 & 0_{\Delta\Delta} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \dots & \beta_{1G^{\Delta}} & 0_{1,\, G+1^{\Delta}} & \dots & 0_{1G} \end{bmatrix}$$

$$\gamma_1 = \begin{bmatrix} \gamma_{1*} & 0_{**} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1K^*} & 0_{1,\, K+1^*} & \dots & 0_{1K} \end{bmatrix}$$

On partitionne les G lignes de Π en les G^{Δ} premières et les $G^{\Delta\Delta}$ restantes et les K colonnes en les K^* premières et les K^{**} dernières, ce qui permet d'écrire (59) sous la forme :

$$-\begin{bmatrix} \gamma_{1*} & 0_{**} \end{bmatrix} = \begin{bmatrix} \beta_{1\Delta} & 0_{\Delta\Delta} \end{bmatrix} \begin{bmatrix} \Pi_{\Delta *} & \Pi_{\Delta,**} \\ \Pi_{\Delta\Delta,*} & \Pi_{\Delta\Delta,**} \end{bmatrix}$$

d'où l'on tire :

(60) $\quad -\gamma_{1*} = \beta_{1\Delta} \quad \Pi_{\Delta *}$

(61) $\quad 0_{**} = \beta_{1\Delta} \quad \Pi_{\Delta,**}$

(*) Calculons :

$$\begin{bmatrix} \beta_{11} & \dots & \beta_{1G} \end{bmatrix} \begin{bmatrix} b_{11} & b_{21} & \dots & b_{G1} \\ b_{12} & b_{22} & \dots & b_{G2} \\ \\ b_{1G} & b_{2G} & \dots & b_{GG} \end{bmatrix}$$ le produit est un vecteur ligne

son premier élément est égal à : $\beta_{11} b_{11} + \beta_{12} b_{12} + \dots + \beta_{1G} b_{1G} = |B|$

l'élément i $(i \neq 1)$ est égal à : $\beta_{11} b_{i1} + \beta_{12} b_{i2} + \dots \beta_{1G} b_{i\,G} = 0$ car on multiplie les éléments de la 1ère ligne par les cofacteurs de la ième ligne, qui est l'expression que l'on obtiendrait en calculant le déterminant d'une matrice dont la 1ère ligne égale à la ième.

en partitionnant les lignes en la première et les G-1 dernières, les colonnes en quatre ensembles de G^{Δ}, $G^{\Delta\Delta}$, K* et K** colonnes respectivement. Multiplions à gauche par B^{-1} :

$$(66) \qquad B^{-1} A = \left[\dot{I} \;\; -\Pi\right] = \begin{bmatrix} \dot{I}_{\Delta\Delta} & O_{\Delta,\Delta\Delta} & -\Pi_{\Delta,*} & -\Pi_{\Delta,**} \\ O_{\Delta\Delta,\Delta} & \dot{I}_{\Delta\Delta,\Delta\Delta} & -\Pi_{\Delta\Delta,*} & -\Pi_{\Delta\Delta,**} \end{bmatrix}$$

après avoir partitionné les lignes et les colonnes de B^{-1}. On voit en comparant (65) et (66) que :

$$(67) \qquad B^{-1} \bar{A} = \begin{bmatrix} O_{\Delta,\Delta\Delta} & -\Pi_{\Delta,**} \\ \dot{I}_{\Delta\Delta,\Delta\Delta} & -\Pi_{\Delta\Delta,**} \end{bmatrix}$$

où $$(68) \qquad \bar{A} = \begin{bmatrix} O_{\Delta\Delta} & O_{**} \\ A_{\Delta\Delta} & A_{**} \end{bmatrix}$$

multiplions (67) à droite par $\bar{\Pi}$ definie par :

$$\bar{\Pi} = \begin{bmatrix} \dot{I}_{\Delta\Delta,\Delta\Delta} & -\Pi_{\Delta\Delta,**} \\ O_{**,\Delta\Delta} & -I_{**,**} \end{bmatrix}$$

$$(69) \qquad B^{-1} \bar{A} \bar{\Pi} = \begin{bmatrix} O_{\Delta,\Delta\Delta} & \Pi_{\Delta,**} \\ \dot{I}_{\Delta\Delta,\Delta\Delta} & O_{\Delta\Delta,**} \end{bmatrix}$$

Comme B^{-1} et $\bar{\Pi}$ sont régulières, $B^{-1} \bar{A} \bar{\Pi}$ et $\bar{A}$ ont des rangs égaux (*). Et d'après la définition de $\bar{A}$:

$$\text{rang } \bar{A} = \text{rang} \begin{bmatrix} A_{\Delta\Delta} & A_{**} \end{bmatrix}$$

d'après (69) :

$$\text{rang } (B^{-1} \bar{A} \bar{\Pi}) = G^{\Delta\Delta} + \text{rang } (\Pi_{\Delta,**})$$

$$(70) \qquad \text{rang } (\Pi_{\Delta,**}) = \text{rang } A_{\Delta\Delta} \;\; A_{**} = (G - G^{\Delta})$$

La condition (64) est donc <u>équivalente</u> à :

$$(71) \qquad \boxed{\text{rang} \begin{bmatrix} A_{\Delta\Delta} & A_{**} \end{bmatrix} = G - 1}$$

(*) Si A et C sont régulières :

$$\text{rang } (AB) = \text{rang } (BC) = \text{rang } (B)$$

La matrice qui apparait dans la relation (71) est la matrice des coefficients, après arrangements, des G-1 équations restantes formées des variables endogènes et prédéterminées exclues de la première relation. La matrice $[A_{\Delta\Delta}\ A_{**}]$ a G-1 lignes, donc son rang est au plus égal à G-1, on en déduit d'après (70) que le rang $\Pi_{\Delta,**}$ est au plus égal à $G^{\Delta}-1$. Ainsi même si K^{**} est supérieur à $G^{\Delta}-1$, ce qui entraine que Π_{Δ}^{**} est à G^{Δ} lignes et au moins à G^{Δ} colonnes, son rang est au plus égal à $G^{\Delta}-1$ et aux coefficients de la forme réduite ne correspond qu'un ensemble unique de valeurs des paramètres structurels.

Si $K^{**} \geq G^{\Delta}-1$, les paramètres d'une relation sont identifiables. Les résultats de l'estimation peuvent être différents selon que $K^{**} = G^{\Delta}-1$ ou que $K^{**} > G^{\Delta}-1$. Dans le premier cas, le rang de $\hat{\Pi}_{\Delta,**}$ sera, en général, égal à $G^{\Delta}-1$, ce qui permet d'utiliser la méthode indirecte des moindres carrés. En remplaçant dans (61) les vraies valeurs par les estimations :

$$(72) \qquad 0_{**} = \hat{\beta}_{1\Delta}\ \hat{\Pi}_{\Delta,**}$$

on obtient $\hat{\beta}_{1\Delta}$, et $\hat{\gamma}_{1*}$ est tiré de :

$$(73) \qquad \hat{\gamma}_{1,*} = -\ \hat{\beta}_{1\Delta}\ \hat{\Pi}_{\Delta,*}$$

Si $K^{**} > G^{\Delta}-1$, il faut soit modifier la méthode indirecte des moindres carrés de telle sorte que le rang de $\hat{\Pi}_{\Delta,**}$ soit égal à $G^{\Delta}-1$, soit utiliser une méthode qui n'estime pas les paramètres structurels à partir des paramètres de la forme réduite.

L'identification : problème fondamental.

Pourquoi dira-t-on peut-être, se préoccuper de caractéristiques qui n'affectent pas les manifestations observables du phénomène considéré ? Ne suffit-il pas toujours de connaître la loi de probabilité des variables endogènes pour chaque ensemble de valeurs des variables exogènes ? L'identification n'est-elle pas un faux problème ?

Ces doutes seraient parfaitement fondés si le modèle offrait une représentation valable intégralement et universellement, pour les inductions statistiques considérées comme pour les utilisations auxquelles leurs résultats donneront lieu. Mais il arrive souvent que le modèle retenu comme base des investigations économétriques, sur une collectivité ou pendant une période, ne convienne pas tel quel pour une autre collectivité ou une autre période. Certaines lois du modèle s'appliquent bien sans changement, mais d'autres doivent être modifiées. Seule alors une connaissance exacte des structures peut servir de base aux révisions nécéssaires. Ce point a été développé de façon particulièrement convaincante par J. Marschak (*) .

3 - Méthodes d'estimations.

Il existe plusieurs méthodes d'estimation utilisables, on doit distinguer entre celles qui s'appliquent à une équation et celles qui s'appliquent à toutes les équations simultanément.

Moindres carrés ordinaires.

On applique la méthode simple des moindres carrés à chaque équation. Les estimateurs sont biaisés et ne sont pas corrects. Cependant, du fait de sa simplicité, cette méthode n'est pas nécéssairement inutilisable

Moindres carrés indirects

Cette méthode ne peut s'appliquer qu'à une relation exactement identifiée. On applique les moindres carrés ordinaires au modèle réduit et on déduit les estimateurs pour les paramètres structurels à partir de ceux des paramètres des formes réduites. Les estimateurs des paramètres structurels sont biaisés, mais ce sont ceux du maximum de vraisemblance si les erreurs suivent une distribution normale.

On dispose de deux autres méthodes qui s'appliquent à une équation à la fois :

- la méthode du maximum de vraisemblance avec information limitée ou méthode du rapport de la plus petite variance.
- la méthode des moindres carrés en deux étapes

(*) J. MARSCHAK (1953) - "Economic measurement for policy and prediction dans W.C. Hood et T.C. Koopmans (ed.) (1953). Studies in econometric method, Cowles commission monograph, n° 14, Wiley, New-York.

Pour estimer toutes les équations simultanément, on verra :

- la méthode du maximum de vraisemblance avec information totale,
- la méthode des moindres carrés en trois étapes.

4 - Méthode du maximum de vraisemblance avec information limitée ou méthode du rapport de la plus petite variance.

Il s'agit de deux méthodes qui conduisent aux mêmes formules d'estimation comme nous allons le voir. Considérons la première équation du modèle $B\,y_t + \Gamma\,x_t = u_t$, c'est-à-dire :

$$(74) \qquad \beta_1\, y_t + \gamma_1\, x_t = u_t \qquad t = 1, 2, \ldots, n$$

et supposons qu'elle soit suridentifiée donc que : $K^{**} > G^{\Delta} - 1$.
L'équation (74) s'écrit :

$$(75) \qquad \beta_{11}\, y_{1t} + \ldots + \beta_{1G^{\Delta}}\, y_{G^{\Delta}t} + \gamma_{11}\, x_{1t} + \ldots + \gamma_{1K^*}\, x_{K^*t} = u_{1t}$$

Soit la forme réduite : $y_t = \Pi z_t + v_t$. Si les vecteurs u suivent une distribution normale et sont indépendants, il en est de même des vecteurs v. Considérons les G^{Δ} équations réduites qui correspondent aux variables endogènes à coefficients différents de zéro ; et déterminons leur fonction de vraisemblance. Elle s'exprime en fonction de $\left[\Pi_{\Delta *}, \Pi_{\Delta, **}\right]$ qui correspond aux G^{Δ} premières lignes de la matrice Π de la forme réduite. Sa maximisation conduit à l'estimateur $\left[\hat{\Pi}_{\Delta, *}\ \hat{\Pi}_{\Delta, **}\right]$, mais $\hat{\Pi}_{**}$ a au moins G^{Δ} colonnes, donc un rang qui peut être égal à G^{Δ} , d'après (72) on a :

$$0_{**} = \hat{\beta}_{1\Delta}\, \hat{\Pi}_{\Delta,**}$$

ce qui permet pas de calculer $\hat{\beta}_{1\Delta}$.

La méthode du maximum de vraisemblance avec information limitée, revient à maximiser la fonction de vraisemblance sous la contrainte :

$$\text{rang}\ (\hat{\Pi}_{\Delta,**}) = G^{\Delta} - 1$$

ce qui permet de calculer $\hat{\beta}_{1\Delta}$.

On montre alors que le problème se ramène à rechercher un $\beta_{1\Delta}$ dont les éléments maximisent :

$$(76) \qquad L = -\frac{1}{2} \operatorname{Log} \frac{\beta_{1\Delta} \; W^*_{\Delta\Delta} \beta'_{1\Delta}}{\beta_{1\Delta} \; W_{\Delta\Delta} \; \beta'_{1\Delta}} \qquad (*)$$

$W^*_{\Delta\Delta}$ et $W_{\Delta\Delta}$ désignent des matrices de résidus que nous expliciterons. Pour siplifier l'écriture, écrivons (75) sous la forme :

$$(77) \qquad \tilde{y}_t = \beta_{11} y_{1t} + \beta_{12} y_{2t} + \dots + \beta_{1G}{}^{\Delta} y_{G^{\Delta}t} \qquad t = 1, 2, \dots n$$

et définissons les matrices des observations suivantes :

$$Y_\Delta = \begin{bmatrix} y_{11} & \dots & y_{G^{\Delta}1} \\ y_{1n} & \dots & y_{G^{\Delta}n} \end{bmatrix} \qquad X_* = \begin{bmatrix} x_{11} & \dots & x^*_{K1} \\ x_{1n} & \dots & x^*_{Kn} \end{bmatrix}$$

$$X_{**} = \begin{bmatrix} x^*_{K+1,\,1} & \dots & x_{K1} \\ x^*_{K+1,\,n} & \dots & x_{Kn} \end{bmatrix}$$

$$X = \begin{bmatrix} X_* & X_{**} \end{bmatrix}$$

On peut alors écrire :

$$(79) \qquad \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \\ \hat{y}_n \end{bmatrix} = Y_\Delta \, \beta'_{1\Delta}$$

La relation (75) spécifie que $\hat{y}$ depende de X_* et non pas de X_{**}, nous allons cependant utiliser cette matrice pour estimer (75). Le principe de la méthode du rapport de la plus petite variance pose que les β qui inter-

---

(*) Ce résultat est démontré dans :
T, W ANDERSON , H. RUBIN "Estimation of parameters of a single equation in a complete system of stochastic equations " AMS, 20, 46-63, 1949.

viennent dans (75) doivent être choisis de telle sorte que le rapport de la variance résiduelle de la régression de $\tilde{y}$ par rapport à X_* à la variance résiduelle de la régression de $\tilde{y}$ par rapport à X soit minimum. On utilise le fait que l'introduction des observations sur les variables exclues permet la réduction de la variance.

La somme des carrés des $\tilde{y}$ est :

$$\tilde{y}\ ,\ \tilde{y} = \beta_{1\Delta}\, Y'_{\Delta}\, Y_{\Delta}\, \beta'_{1\Delta}$$

dans la régression de $\tilde{y}$ sur X_*, le vecteur des estimateurs est $(X'_* X_*)^{-1} X'_* \tilde{y}$ et la somme des carrés expliquée est $\tilde{y}' X_* (X'_* X_*)^{-1} X'_* \tilde{y}$, d'où la somme des carrés des résidus :

$$\beta_{1\Delta}\, Y'_{\Delta}\, Y_{\Delta} \beta'_{1\Delta} - \beta_{1\Delta} Y'_{\Delta}\, X_* (X'_* X_*)^{-1} X'_* Y_{\Delta}\, \beta'_{1\Delta} = \beta_{1\Delta} W^*_{\Delta\Delta}\, \beta'_{1\Delta}$$

$$(80) \qquad W^*_{\Delta\Delta} = Y'_{\Delta}\, Y_{\Delta} - Y'_{\Delta}\, X\, (X'_* X_*)^{-1} X'_* Y_{\Delta}$$

De la même façon, on calcule la somme des carrés des résidus de la régression de $\tilde{y}$ sur X, c'est-à-dire :

$$\beta_{1\Delta}\, W_{\Delta\Delta} \beta'_{1\Delta}$$

$$(81) \qquad W_{\Delta\Delta} = Y'_{\Delta}\, Y_{\Delta} - Y'_{\Delta}\, X\, (X'X)^{-1} X'Y_{\Delta}$$

On cherche alors le $\beta_{1\Delta}$ qui minimise :

$$(82) \qquad l = \frac{\beta_{1\Delta}\, W^*_{\Delta\Delta}\, \beta'_{1\Delta}}{\beta_{1\Delta}\, W_{\Delta\Delta}\, \beta'_{\Delta\Delta}}$$

En comparant (82) et (76) on constate alors, que la minimisation de l est équivalente à la maximisation de L.

Calculons l'estimateur ; on calcule les dérivées partielles de l par rapport aux β_{1i} $(i = 1, \ldots G^{\Delta})$ que l'on annulle.

$$(83) \qquad w^*_i\, \beta'_{1\Delta} - l\, w_i\, \beta'_{1\Delta} = 0 \qquad i = 1, \ldots G^{\Delta}$$

où w^*_i et w_i désignent les lignes i de W^* et W, le système (83) se met sous la forme :

$$(84) \qquad (w^*_{\Delta\Delta} - l\, w_{\Delta\Delta})\, \beta'_{1\Delta} = 0$$

qui admet une solution non triviale si :

$$(85) \qquad |w^*_{\Delta\Delta} - l\, w_{\Delta\Delta}| = 0$$

Cette équation s'exprime en fonction des puissances de l et des observations. Soit $\hat{l}$ sa plus petite racine, on calcule alors $\hat{\beta}_{\Delta\Delta}$ en résolvant :

$$(86) \qquad (w^*_{\Delta\Delta} - \hat{l}\, w_{\Delta\Delta})\, \hat{\beta}'_{1\Delta} = 0$$

On estime les coefficients des variables prédéterminées, on obtient en calculant la régression de $\tilde{y}$ par rapport à X_* on obtient $(X'_* X_*)^{-1} X'_* \tilde{y}$. D'où :

$$(87) \qquad \hat{\gamma}_{1*} = - \hat{\beta}_{1\Delta}\, Y'_\Delta X_* (X'_* X_*)^{-1}$$

C'est exactement ce que l'on obtiendrait en utilisant (73) c'est-à-dire :

$$\hat{\gamma}_{1*} = - \hat{\beta}_{1\Delta}\, \hat{\Pi}_{\Delta *}$$

$\hat{\Pi}_{\Delta *}$ étant obtenu en appliquant la méthode simple des moindres carrés à chacune des G^Δ variables endogènes, par rapport aux K^* variables prédéterminées de la relation.

En conclusion, le calcul se déroule de la façon suivante :

(i) déterminer les matrices des observations Y_Δ, X_* et X,

(ii) calculer $w^*_{\Delta\Delta}$ et $w_{\Delta\Delta}$,

(iii) déterminer la plus petite racine $\hat{l}$ de $|w^*_{\Delta\Delta} - l\, w_{\Delta\Delta}| = 0$

(iv) Puis résoudre :

$$(w^*_{\Delta\Delta} - \hat{l}\, w_{\Delta\Delta})\, \hat{\beta}'_{1\Delta} = 0$$

$$\text{et } \hat{\gamma}_{1*} = - \hat{\beta}_{1\Delta}\, Y'_\Delta X_* (X'_* X_*)^{-1}$$

5 - <u>Méthode des moindres carrés en deux étapes</u>

Dans l'équation (75) posons $\beta_{11} = 1$:

$$(88) \qquad y_{1t} = -\beta_{12}\, y_{2t} \cdots -\beta_{1G^\Delta} y_{G^\Delta t} - \gamma_{11}\, x_{1t} \cdots - \gamma_{1K^*}\, x^*_{K^* t} + u_{1t}$$

$$t = 1, \ldots n$$

Le système d'équations s'écrit sous la forme matricielle suivante :

$$(89) \qquad y_1 = - Y_2 \beta'_2 - X_* \gamma'_{1*} + u_1$$

avec :

$$y_1 = \begin{bmatrix} y_{11} \\ \\ y_{1n} \end{bmatrix} \qquad Y_2 = \begin{bmatrix} y_{21} & \cdots & y_{G^{\Delta}1} \\ & & \\ y_{2n} & \cdots & y_{G^{\Delta}n} \end{bmatrix} \qquad \beta'_2 = \begin{bmatrix} \beta_{12} \\ \\ \beta_{1G^{\Delta}} \end{bmatrix}$$

(90)

$$X_* = \begin{bmatrix} x_{11} & \cdots & x_{K^*1} \\ & & \\ x_{1n} & & x_{K^*n} \end{bmatrix} \qquad \gamma'_{1*} = \begin{bmatrix} \gamma_{11} \\ \\ \gamma_{1K^*} \end{bmatrix}$$

Comme nous l'avons vu u_1 est en général corrélé avec les variables explicatives Y_2. L'idée de base de la méthode des moindres carrés en deux étapes consiste à remplacer dans (89) Y_2 par un estimateur $\hat{Y}_2$ obtenu en faisant une régression de Y_2 par rapport à toutes les variables prédéterminées. Enfin d'appliquer les moindres carrés simples à la relation entre y_1, $\hat{Y}_2$ et X_*. Il y a entre cette méthode et celle du maximum avec information limitée une légère similitude. L'une et l'autre utilise pour estimer les paramètres d'une relation donnée toutes les variables prédéterminées, mais sans introduire les autres relations du modèle.

Estimons tout d'abord Y_2. La régression de y_2 par rapport à X donne : $\hat{y}_2 = X(X'X)^{-1} X'y_2$. De la même façon celle de y_3 par rapport à X conduit à $\hat{y}_3 = X(X'X)^{-1} X'y_3$. D'où :

$$\hat{Y}_2 = \left[\hat{y}_2, \hat{y}_3, \ldots \hat{y}_{G\Delta} \right]$$

$$(91) \qquad \hat{Y}_2 = X\,(X'X)^{-1}\, X'Y_2$$

ou (92) $\qquad Y_2 = X\,(X'X)^{-1}\, X'Y_2 + V$

où V désigne la matrice des résidus pour les $G^{\Delta} - 1$ variables endogènes qui apparaissent dans le membre de droite de (89).

(89) peut s'écrire sous la forme :

$$(93) \qquad y_1 = -(Y_2 - V)\,\beta'_2 - X_* \gamma'_{1*} + (u - V\,\beta'_2)$$

$$(94) \qquad y_1 = -\left[(Y_2 - V), X_*\right]\begin{bmatrix}\beta'_2 \\ \gamma'_{1*}\end{bmatrix} + (u - V\,\beta'_2)$$

En appliquant les moindres carrés à cette relation, on obtient :

$$(95) \qquad \begin{bmatrix}\hat{\beta}'_2 \\ \hat{\gamma}'_{1*}\end{bmatrix} = -(A'A)^{-1} A'y_1$$

$$(96) \qquad A = \left[(Y_2 - V)\ X_*\right]$$

$$(Y_2 - V)'\,(Y_2 - V) = Y'_2\,Y_2 - V'Y_2 - Y'_2V + V'V$$

$$V'Y_2 = V'\,(\hat{Y}_2 + V) \text{ d'après (91) et (92)}$$
$$= V'V$$

Or, on sait que si on applique la méthode des moindres carrés les résidus et les valeurs ajustées ne sont pas correlées, donc : $V'\,\hat{Y}_2 = 0$
De la même façon :

$$Y'_2\,V = V'V$$

d'où $\qquad (Y_2 - V)'(Y_2 - V) = Y'_2\,Y_2 - V'V$

$$X'\,v = X'\left[Y_2 - X\,(X'X)^{-1}\,X'Y_2\right]$$
$$= 0$$

ce qui illustre la propriété selon laquelle les résidus ne sont pas corrélés aux variables explicatives. Quant $X'_*\,V$ qui est une sous-matrice de $X'V$, elle est aussi identiquement nulle. L'estimateur des moindres carrés en deux étapes (95) peut être écrit sous la forme :

$$(97) \begin{bmatrix}\hat{\beta}'_2 \\ \hat{\gamma}'_{1*}\end{bmatrix} = -\begin{bmatrix} Y'_2\,Y_2 - V'V & Y'_2\,X_* \\ X'_*\,Y_2 & X'_*X_*\end{bmatrix}^{-1}\begin{bmatrix} Y'_2 - V' \\ X'_*\end{bmatrix} y_1$$

il y a identification possible si cette matrice inverse existe, c'est-à-dire si $K^{**} \geqslant G^{\Delta} - 1$. $A = \left[(Y_2 - V\ X_*\right]$ est une matrice à n lignes et $G^{\Delta} - 1 + K^*$ colonnes. $A'A$ est donc une matrice carrée d'ordre $G^{\Delta} - 1 + K^*$ et rang $(A'A)$ = rang (A).

$$A = \left[(Y_2 - V)\ X_* \right] = X \left[(X'X)^{-1}\ X'Y_2 \qquad \begin{matrix} I \\ 0 \end{matrix} \right] \text{ d'après (92)}$$

où I est la matrice unité d'ordre K^* et 0 la matrice nulle à K^{**} lignes et K^* colonnes. Le rang de A est donc au plus égal au rang de X, qui est égal à K. Si le rang de A est inférieur à $G^\Delta - 1 + K^*$ alors A'A est singulière. Ceci se produit si :

$$K < G^\Delta - 1 + K^*$$

c'est-à-dire si $K^{**} < G^\Delta - 1$

et (88) est sous identifiée.

6 - Estimateurs de classe k

Theil (*) a introduit les estimateurs de classes k de (97) à partir de :

$$(98) \qquad \begin{bmatrix} \hat{\beta}'_2 \\ \\ \hat{v}'_{1*} \end{bmatrix}_k = - \begin{bmatrix} Y'_2\ Y_2 - kV'V & Y'_2\ X_* \\ \\ X'_*\ Y_2 & X'_*X_* \end{bmatrix}^{-1} \begin{bmatrix} Y'_2 - kV' \\ \\ X'_* \end{bmatrix} y_1$$

trois des estimateurs que nous avons introduits jusqu'ici appartiennent à cette catégorie. A k=0 correspondent les estimateurs des moindres carrés ordinaires, car (98) revient à une application directe des moindres carrés simples à la relation (88), avec $\left[Y_2\ X_* \right]$ comme matrice des observations sur les variables explicatives. Les moindres carrés en deux étapes correspondent à k=1, et le maximum de vraisemblance avec information limitée à $k = \hat{l}$, où $\hat{l}$ est défini par (85). Démontrons ce dernier résultat. Les estimateurs du maximum de vraisemblance avec information limitée sont obtenus à partir de (86) :

$$(w^*_{\Delta\Delta} - \hat{l}\ w_{\Delta\Delta}\ )\ \hat{\beta}'_{1\Delta} = 0$$

Dans l'estimateur de classe 1, nous posons :

$$\beta_{1\Delta} = \left[1\ \beta_2 \right]$$

(*) H. THEIL "Economic forecasts and policy" 2d ed, North-Holland Publishing Company, Amsterdam, 1961, chapitre 6.

en posant $\beta_{11} = 1$. D'après les définitions de $w^*_{\Delta\Delta}$ et de $w_{\Delta\Delta}$, nous pouvons écrire :

$$(99) \qquad w^*_{\Delta\Delta} = Y'_\Delta B_* Y_\Delta = \begin{bmatrix} y'_1 & B_* Y_\Delta \\ Y'_2 & B_* Y_\Delta \end{bmatrix}$$

avec

$$(100) \qquad B_* = I - X_* (X'_* X_*)^{-1} X'_*$$

$$(101) \qquad Y_\Delta = \begin{bmatrix} y_1 & Y_2 \end{bmatrix}$$

de la même façon :

$$(102) \qquad w_{\Delta\Delta} = Y'_\Delta B Y_\Delta = \begin{bmatrix} y'_1 B Y_\Delta \\ y'_2 B Y_\Delta \end{bmatrix}$$

$$(103) \qquad B = I - X (X'X)^{-1} X'$$

Considérons alors le système (86), il s'écrit (en laissant de côté la première équation) :

$$(104) \qquad (Y'_2 B_* Y_\Delta - \hat{1} Y'_2 B Y_\Delta) \hat{\beta}'_{1\Delta} = 0$$

Les estimateurs de classe k peuvent être introduits sous la forme :

$$(105) \qquad \begin{bmatrix} Y'_2 Y_2 - k V'V & Y'_2 X_* \\ X'_* Y_2 & X'_* X_* \end{bmatrix} \begin{bmatrix} \hat{\beta}'_2 \\ \hat{\gamma}'_{1*} \end{bmatrix}_k = - \begin{bmatrix} Y'_2 - kV' \\ X'_* \end{bmatrix} y_1$$

ce qui conduit à :

$$(106) \qquad (Y'_2 Y_2 - kV'V) (\hat{\beta}'_2)_k + Y'_2 X_* (\hat{\gamma}'_{1*})_k = - (Y_2 - kV') y_1$$

A partir de (92) on montre que :

$$V'V = Y'_2 B Y_2$$

car B est symétrique et idempotente. De la même façon :

$$V'y_1 = Y'_2 B y_1$$

En substituant dans (106) on obtient :

$$(107) \qquad (Y'_2 Y_2 - k Y'_2 B Y_2) (\hat{\beta}'_2)_k + Y'_2 y_1 - kY'_2 B y_1 + Y'_2 X_* (\hat{\gamma}'_{1*})_k = 0$$

De (105) on tire aussi :

$$(108) \qquad (\hat{\gamma}'_{1*})_k = - (X'_* X_*)^{-1} \left[X'_* Y_2 (\hat{\beta}'_2)_k + X'_* y_1 \right]$$

que l'on substitue dans (107), cette dernière relation s'écrit alors sous la forme :

$$(109) \quad (Y'_2 B_* Y_2 - kY'_2 BY_2)(\hat{\beta}'_2)_k + (Y'_2 B_* y_1 - kY'_2 By_1) = 0$$

qui s'écrit encore :

$$(110) \quad (Y'_2 B_* Y_\Delta - k Y'_2 BY_\Delta)(\hat{\beta}'_{1\Delta})_k = 0$$

estimateur identique à celui du maximum de vraisemblance avec information limitée, ce que l'on voit en posant $k = \hat{l}$.

Revenons sur la signification des moindres carrés en deux étapes. Et considérons la relation (92) :

$$l = \beta_{1\Delta} W^*_{\Delta\Delta} \beta'_{1\Delta} / \beta_{1\Delta} W_{\Delta\Delta} \beta'_{1\Delta}$$

et définissons :

$$\varphi = \beta_{1\Delta} W^*_{\Delta\Delta} \beta'_{1\Delta} - \beta_{1\Delta} W_{\Delta\Delta} \beta'_{1\Delta}$$

on peut montrer que les estimateurs des moindres carrés en deux étapes minimisent φ. Calculons les dérivées de φ par rapport à $\beta_{1\Delta}$; on obtient :

$$(W^*_{\Delta\Delta} - W_{\Delta\Delta}) b'_{1\Delta} = 0$$

Comme $\beta_{11} = 1$, on peut utiliser les relations (99) à (104) et ce système d'équations s'écrit alors :

$$(Y'_2 B_* Y_\Delta - Y'_2 BY_\Delta) b'_{1\Delta} = 0$$

qui est la relation (110) pour $k = 1$. Ainsi les moindres carrés en deux étapes consistent à minimiser la différence des carrés des résidus, alors que le maximum de vraisemblance avec information limitée en minimise le rapport.

Le choix d'un estimateur particulier de classe k dépend de ses propriétés et de considérations sur les calculs à effectuer. On montre (*) qu'ils sont corrects si $\rho \lim_{n \to \infty} k=1$. Ce n'est pas le cas des estimateurs des moindres carrés ordinaires pour lesquels nous avons k=0, par contre les moindres carrés en deux étapes donnent des estimateurs corrects (k=1) ainsi que le maximum de vraisemblance avec information limitée car $\rho \lim_{n \to \infty} \hat{l} = 1$

Les distributions des estimateurs de classe k, n'ont pas été obtenus pour de petits échantillons, par contre on dispose de quelques résultats

---

(*) THEIL op. cit. p. 232.

asymptotiques. Définissons les résidus des estimateurs de classe k :

$$e_k = \begin{bmatrix} \hat{\beta}'_2 \\ \hat{\gamma}'_{1*} \end{bmatrix}_k - \begin{bmatrix} \beta'_2 \\ \gamma'_1 \end{bmatrix}$$

alors si $\underset{n \to \infty}{\text{plim}} \sqrt{n}\,(k-1) = 0$. On a :

$$(111) \quad \lim_{n \to \infty} E\,(n e_k\, e'_k) = \sigma^2 \underset{n \to \infty}{\text{plim}} \begin{bmatrix} Y'_2\, Y_2 - V'V & Y'_2\, X_* \\ X'_*\, Y_2 & X'_*\, X_* \end{bmatrix}^{-1} \qquad (*)$$

où σ^2 désigne la variance du résidu de l'équation que l'on estime. On estime σ^2 par la somme des carrés des résidus empiriques divisée par le nombre de degré de liberté, et la limite en probabilité de la matrice par la matrice elle-même. Les estimateurs des moindres carrés en deux étapes et du maximum de vraisemblance avec information limitée ont la même matrice asymptotique des variances covariances.

Dans le cas de petits échantillons le choix de la méthode d'estimation est délicat, il existe quelques résultats (**), les méthode de Monte-Carlo semblent devoir être très fructueuses.

7 - Méthode du maximum de vraisemblance à information complète.

Considérons le modèle complet :

$$By_t + \Gamma x_t = u_t$$

d'après (32), la fonction de vraisemblance des variables endogènes sachant les variables exogènes est égale à :

$$L = p(y_1, \dots y_n / x_1, \dots x_n) = |\det B|^n\, p(u_1)\, p(u_2) \dots p(u_n)$$

Si u_t suit la distribution multinormale $N(0, \Phi)$;alors :

$$(112) \quad L^* = \text{Log } L = K + n \text{ Log } |\det B| + \det\left(-\frac{n}{2} \text{Log } \det\Phi\right) - \frac{1}{2} \sum_{t=1}^{n} u'_t\, \Phi^{-1}\, u_t$$

(*) cf. THEIL op. cit.

(**) A L NAGAR "The bias and moment matrix of the general k - class estimators of the parameters in simultaneous equations" Econometrica 27, PP. 575-595, 1959.

La maximisation de L^* conduit à annuler les dérivées partielles de L^* par rapport aux éléments de B, Γ et Φ. La résolution du système obtenu est en général très difficile. Par exemple :

$$\frac{\partial L^*}{\partial \beta_{ij}} = 0 \iff \frac{n}{|\det B|} \frac{\partial |\det B|}{\partial \beta_{ij}} - \frac{1}{2} \frac{\partial \mathcal{S}}{\partial \beta_{ij}} = 0$$

où $\quad \mathcal{S} = \sum_{t=1}^{n} u'_t \Phi^{-1} u_t.$

cette équation est non linéaire.

Un cas particulier où la résolution est possible est celui des systèmes recussifs. Un exemple de système recussif est fourni par :

$$\begin{cases} y_{1t} + \sum_{j=1}^{3} \gamma_{1j} x_{jt} = u_{1t} \\ \beta_{21} y_{1t} + y_{2t} + \sum_{j=1}^{3} \gamma_{2j} x_{jt} = u_{2t} \\ \beta_{31} y_{1t} + \beta_{32} y_{2t} + y_{3t} + \sum_{j=1}^{3} \gamma_{3j} x_{jt} = u_{3t} \end{cases}$$

dans ce cas B est triangulaire, det B = 1 - A ces systèmes, on associe généralement le nom de Wold (*) qui a prétendu que ces chaines de relations causales fournissent une représentation valable des mecanimes de l'économie. Cependant, on peut considérer (**) que l'utilisation de données partant sur d'assez longues périodes induit une interdépendance des équations. Dans le cas de systèmes récussifs si Φ est diagonal, l'estimation est facile. Les estimateurs obtenus sont les mêmes que ceux donnés par l'application des moindres carrés ordinaires à chaque équation.

β - Moindres carrés en trois étapes

Il s'agit d'une méthode assez récente(***) , soit l'équation (89) écrite sous la forme :

(113) $\quad y_1 = Y_2 \beta'_2 + X_* \gamma'_{1\Delta} + u_1$

(*) H. WOLD L. JUREEN "Demand analysis" Wiley, New-York, 1953.

(**) R. Dentzel B. Hansen "On recussiveness and independency in economic models" Rev. Economic Studies, vol 22, pp. 153-168, 1954-1955.

(***) A. Zellen, H. Theil "Three stage least squares : simultaneous estimation of simultaneous equations" Econometrica, vol 30, pp. 54-78, 1962.

On pose :

(114) $y_1 = Z_1 \delta_1 + u_1$

(115) $Z_1 = \left[Y_2 \; X_* \right] \quad \delta_1 = \begin{bmatrix} \beta'_2 \\ \gamma'_{1*} \end{bmatrix}$

Multiplions les deux membres de (114) à gauche par X', où $X = \left[X_* \; X_{**} \right]$ désigne la matrice n x k des observations sur les variables prédéterminées, nous obtenons :

(116) $X'y_1 = X' Z_1 \delta_1 + X'u_1$

système à K équations et $G^{\Delta} - 1 + K^*$ inconnus qui sont les éléments de δ_1. Si (114) est exactement identifiée $K^{**} = G^{\Delta} - 1$ et le nombre d'équations est égal au nombre d'inconnus. On peut alors estimer δ_1 en remplaçant le vecteur des résidus $X'u_1$ par son espérance c'est-à-dire zéro, l'estimateur s'écrit :

(117) $\hat{\delta}_1 = (X'Z_1)^{-1} X'y_1$

On peut vérifier que c'est aussi l'estimateur obtenu par la méthode indirecte des moindres carrés.

Lorsqu'il y a sur identification ($K^{**} > G^{\Delta} - 1$) le système (116) compte plus d'équations que d'inconnus, cependant les moindres carrés ordinaires ne conviennent pas.

Si nous supposons que les variables prédéterminées sont toutes exogènes, on a :

(118) $\text{var}\,(X'u_1) = E\,(X'u_1 \, u'_1 X) = \sigma_{11} X'X$

en supposant que $E\,(u_1 \, u'_1) = \sigma_{11} I$. Ce résultat suggère d'appliquer les moindres carrés simples à (116) d'où :

(119) $\hat{\delta}_1 = \left[Z'_1 X(X'X)^{-1} X'Z_1 \right]^{-1} Z'_1 X(X'X)^{-1} X'y_1$

qui est l'estimateur des moindres carrés en **deux** étapes développé dans (97).

L'idée de base des moindres carrés en trois étapes est d'écrire toutes les équations sous la forme (116) :

(120) $$\begin{bmatrix} X'y_1 \\ X'y_2 \\ \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & & & \\ & X'Z_2 & & 0 \\ & 0 & \ddots & \\ & & & X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \\ \delta_G \end{bmatrix} + \begin{bmatrix} X'u_1 \\ X'u_2 \\ \\ X'u_G \end{bmatrix}$$

ou encore :

(121) $A = B\delta + v$

qui est un système à K G équations. La matrice des variances - covariances des résidus s'écrit :

$$(122) \qquad V = \mathrm{var} \begin{bmatrix} X'u_1 \\ X'u_2 \\ \\ X'u_G \end{bmatrix} = X'X \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1G} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2G} \\ \\ \delta_{G1} & \delta_{G2} & \cdots & \delta_{GG} \end{bmatrix}$$

;

avec $\delta_{ij} = E(u_i\, u'_j) = \sigma_{ij}\, I$

L'application des moindres carrés à (121) donne :

$$(123) \qquad \hat{\delta} = (B'V^{-1}\, B)^{-1}\, B'V^{-1}\, A$$

Pour estimer V, les auteurs recommandent de remplacer les σ_{ij} par leurs estimés S_{ij} obtenus par la méthode des moindres carrés en deux étapes. Et on a :

$$(124) \qquad \mathrm{Var}\,(\hat{\delta}) = \begin{bmatrix} S^{11}\, Z'_1\, X(X'X)^{-1}X'Z_1 & \cdots & S^{1G}\, Z'_1\, X(X'X)^{-1}X'Z_G \\ \\ S^{G1}Z'_G\, X(X'X)^{-1}\, X'Z_1 & \cdots & S^{GG}\, Z'_G\, X(X'X)^{-1}\, X'Z_G \end{bmatrix}$$

où $\left[S^{ij}\right] = \left[S_{ij}\right]^{-1}$

Les auteurs montrent que les estimateurs obtenus sont meilleurs que ceux des moindres carrés en deux étapes. On sait que les estimateurs du maximum de vraisemblance avec information complète sont peu robustes relativement aux erreurs de spécification, en est-il de même pour ceux des moindres carrés en trois étapes ?

Chapitre 5

ETUDE DES PROPRIETES DES METHODES D'ESTIMATION PAR LA METHODE DE MONTE-CARLO

1 - Généralités

Pour l'essentiel les méthodes d'estimation ont été justifiées par leurs propriétés asymptotiques. Nous avons vu que les moindres carrés ordinaires donnaient des estimateurs non corrects et biaisés. Les autres méthodes d'estimation conduisent à des estimateurs biaisés mais par contre corrects. Sous l'hypothèse de normalité et d'autocorrélation nulle des erreurs, les estimateurs du maximum de vraisemblance à information limitée sont asymptotiquement normaux et efficaces (ce sont dans la classe des estimateurs définis par les mêmes hypothèses ceux qui ont la plus petite variance). Les moindres carrés en deux étapes possèdent les mêmes propriétés. Afin d'apprécier plus précisément ces méthodes et les estimateurs auxquels elles conduisent, il conviendrait de mieux connaitre leurs propriétés pour les échantillons de taille finie.

Ceci serait utile à deux points de vue. D'une part, les procédures statistiques associées à chaque méthode d'estimation seraient mieux définies. D'autre part, nos idées sur les avantages comparés des diverses méthodes pourraient changer si nous connaissions les distributions pour les échantillons finis. Telle procédure, qui semble inadequate asymptotiquement, pourrait avoir de l'interêt pour le traitement de petits échantillons de la taille de ceux employés en économétrie. On a ainsi parfois avancé que l'ajustement direct des équations structurelles pourrait bien constituer la meilleure methode d'estimation en pratique. Sans doute implique-t-il des biais. Mais si les dispersions sont faibles, les erreurs d'estimation peuvent encore être plus réduites qu'avec d'autres méthodes.

Deux approches sont concevables pour l'étude des distributions valables avec les petits échantillons :

(i) recherche de la forme analytique exacte des lois de probabilité (*).

(ii) méthode dite de "Monte-Carlo".

Nous ne nous intéressons qu'à la deuxième approche. Elle consiste à partir d'un modèle donné, dont les paramètres sont déterminés, à travailler sur des échantillons artificiels. R. Summers (**) semble avoir été un des premiers a développé cette approche. Avant de présenter quelques uns de ces résultats, donnons des définitions utiles.

Soit Θ la vraie valeur du paramètre que l'on cherche à estimer. Supposons que l'on ait engendré N échantillons artificiels et que l'on étudie une méthode donnée d'estimation. Désignons par $\hat{\Theta}_i$ la valeur prise par l'estimation pour l'échantillon i; et par $\bar{\Theta}$ la moyenne des $\hat{\Theta}_i$. On définit :

- le biais par : biais $= \bar{\hat{\Theta}} - \Theta$

- la variance par : $\text{var} = \dfrac{\sum_{i=1}^{N} (\hat{\Theta}_i - \bar{\hat{\Theta}})}{N}$

- l'erreur quadratique moyenne par :

$$\text{EQM} = \frac{\sum_{i=1}^{N} (\hat{\Theta}_i - \Theta)^2}{N}$$

On peut vérifier que :

$$\text{EQM} = \text{var} + (\text{biais})^2$$

(*) R. L. BASMANN "A note on the exact finite sample frequency functions of generalized classical linear estimators in two leading over identified cases" J. Am. Statist. Assoc. vol 56, pp. 619-636, 1961.

A. L. NAGAR "The bias and moment matrix of the general k - class estimators of the parameters in simultaneous equations". Econometrica, vol. 27, pp. 575 - 595, 1959

(**) R. SUMMERS "A capiyal intensive approach to the small sample properties of various simultaneous equation estimators". Econometrica, janvier 1965.

2 - Quelques résultats de Summers

Soit le modèle :

$$\begin{cases} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{12} z_{2t} + \gamma_{10} = u_{1t} \\ y_1 + \beta_{22} y_{2t} + \gamma_{23} z_{3t} + \gamma_{24} z_{24} + \gamma_{20} = u_{2t} \end{cases}$$

La variable aléatoire (u_1, u_2) soit une distribution normale à deux dimensions. Summers a répété ses expériences N = 50 fois, T désigne la taille des échantillons. La table ci-dessous indique les valeurs choisies pour les paramètres dans les différentes expériences.

Table - Valeur des paramètres

expérience	β_{12}	γ_{11}	γ_{12}	γ_{10}	β_{22}	γ_{23}	γ_{24}	γ_{20}	σ_{11}	σ_{12}	σ_{22}	T
1A et 1B	-0,7	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20
2A et 2B	-0,7	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20
3A et 3B	-0,1	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20
4A et 4B	-1,3	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20
5A et 5B	-0,7	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20
6A et 6B	-0,7	0,8	0,7	-149,5	0,4	0,6	-0,4	-149,6	400	200	400	20

Il suppose qu'il y a une erreur de spécification dans les expériences 5 et 6, et que z_1 apparait dans la deuxième équation avec un coefficient non nul. Il a pris :

$\gamma_{21} = 0,5$ pour 5A et 5B

$\gamma_{21} = -0,5$ pour 6A et 6B.

Dans les expériences A les variables exogènes sont faiblement corrélées, la corrélation est par contre forte dans les expériences.

Summers a étudié quatre méthodes différentes pour estimer les paramètres structurels :

- maximum de vraisemblance à information limitée (MVIL)
- moindres carrés en deux étapes (MC2)
- moindres carrés ordinaires (MCO)
- maximum de vraisemblance à information totale (MVIT)

Considerons les résultats concernant les expériences 1A à 4A. Nous estimons à chaque fois 8 paramètres, nous avons donc au total 32 estimés. Pour chaque critère, le tableau ci-dessous donne le nombre de fois où la méthode correspondante a donné le meilleur résultat (c'est-à-dire la plus petite valeur du critère).

Tableau : Fréquences des meilleurs résultats (*)
expériences 1 à 4

méthode	fréquences		
	Biais	écart type	REQM
MVIL	12,5	0	2,5
MC2	5	0,5	6,5
MCO	1	20	7,5
MVIT	13,5	11,5	15,5
Total	32	32	32

On constate que pour le critère REQM la méthode MVIT donne les meilleurs résultats, tandis que pour le critère écart-type la méthode MCO lui est supérieure. Le critère écart-type est moins bon que REQM car l'écart-type est calculé autour d'une valeur fausse de l'expérience. Si l'on considère les expériences 5A et 6A, les résultats sont très différents et MVIT ne s'impose plus.

Tableau - Fréquence des meilleurs résultats (expériences 5A et 6A)

méthode	fréquences		
	biais	écart type	REQM
MVIL	5	0	4 5/6
MC2	5	0	3 5/6
MCO	5	14	5
MVIT	1	2	2 1/3
Total	16	16	16

(*) REQM désigne la racine carré de EQM
Lorsque deux méthodes donnent le meilleur résultat, on a donné à chacune la fréquence 0,5.

Nous ne citerons pas d'autres résultats et nous terminerons par l'énoncé de quelques conclusions.

Pour estimer les paramètres structurels, la méthode du maximum de vraisemblance à information totale est sans doute la meilleure. Les calculs sont cependant lourds et la valeur des résultats est très sensible aux erreurs de spécification.

Pour estimer une seule équation, les autres méthodes peuvent classer ainsi (en partant de la meilleure) :

- moindres carrés en deux étapes,
- maximum de vraisemblance à information limitée,
- moindres carrés ordinaires.

En règle générale, l'application directe des moindres carrés aux formes réduites est déconseillée, mais il est difficile de tirer des conclusions générales.

Bien des problèmes ne sont pas résolus, mais les recherches semblent devoir être menées dans l'étude de la statistique des processus stochastiques. Car que deviennent ces méthodes et ces résultats lorsque les erreurs sont autocorrelées.

- Table des matières -

MODELES A UNE EQUATION

pages

Chapitre 1 - Modèles linéaires à erreurs sur les variables.... 1

Chapitre 2 - Autocorrélation..... 21

Chapitre 3 - Quelques points particuliers de l'étude des modèles à une équation..... 43

Chapitre 4 - L'estimation dans les modèles à équations multiples 65

Chapitre 5 - Etude des propriétés des méthodes d'estimation par la méthode de Monte-Carlo..... 100