Journal de la société statistique de Paris

J.-J. Boulanger

Remarques sur la détermination des régressions linéaires multiples en économie

Journal de la société statistique de Paris, tome 120, nº 2 (1979), p. 114-120
<http://www.numdam.org/item?id=JSFS_1979__120_2_114_0>

REMARQUES SUR LA DÉTERMINATION DES RÉGRESSIONS LINÉAIRES MULTIPLES EN ÉCONOMIE

J.-J. Boulanger

Dans l'étude qui suit l'auteur examine une variante de la méthode classique permettant de déterminer avec rapidité la régression linéaire multiple, la plupart du temps, la plus satisfaisante.

In the following study, the author examines a variant of the classical method which allows to determine with rapidity the multiple linear regression, which is, most of the time, the most satisfactory.

In der vorliegenden Studie untersucht der Verfasser eine Abweichung von der klassischen Methode zur Bestimmung mit Schnelligkeit der linearen und vielfachen Regression. In den meisten Fällen gibt diese Methode zufriedenstellende Resultate.

Diverses questions se posent lors de l'analyse des régressions multiples linéaires en économie. Mais elles sont principalement centrées sur les deux problèmes suivants : « Quelles variables choisir? » et « Quelle incidence ont leurs variations sur la variance de la variable analysée? ».

A la question : « Quelles variables choisir? », on peut répondre que les variables qui auront le plus de chance de faire partie de l'hyperplan définitif : $x_1 = ax_2 + bx_3 + cx_4 + ...$, sont en général celles qui ont avec x_1 un coefficient de corrélation (simple) élevé, mais ceci n'est pas obligatoire, une variable, présentant un coefficient de corrélation important avec x_1, pouvant être délaissée au profit de variables présentant avec x_1 des coefficients de corrélation encore plus importants.

Supposons que l'on cherche une régression de la forme :

$$x_1 = ax_2 + bx_3 + cx_4 + d$$

Posons :

$$X_1 = x_1 - \bar{x}_1,\ X_2 = x_2 - \bar{x}_2,\ X_3 = x_3 - \bar{x}_3 \text{ et } X_4 = x_4 - \bar{x}_4,\ \bar{x}_1, \bar{x}_2, \bar{x}_3 \text{ et } \bar{x}_4$$

étant les valeurs moyennes de x_1, x_2, x_3 et x_4 pour la période considérée; la régression devient : $X_1 = aX_2 + bX_3 + cX_4$. L'application de la méthode des moindres carrés conduit au système d'équations :

$$\begin{aligned} A_{12} &= aA_{22} + bA_{23} + cA_{24} \\ A_{13} &= aA_{23} + bA_{33} + cA_{34} \\ A_{14} &= aA_{24} + bA_{34} + cA_{44} \end{aligned}$$

avec
$$A_{ii} = \Sigma X_i^2 \quad \text{et} \quad A_{ij} = \Sigma X_i X_j.$$

Le système d'équations précédent permet de déterminer a, b et c (d résultant de la relation :

$$\bar{x}_1 = a\bar{x}_2 + b\bar{x}_3 + c\bar{x}_4 + d).$$

Si l'on cherche à déterminer a, b et c par la méthode de substitution on tire :

$$a = \frac{A_{12}}{A_{22}} - b\frac{A_{23}}{A_{22}} - c\frac{A_{24}}{A_{22}}$$

que l'on reporte dans les 2e et 3e équations, d'où

$$A_{12} = aA_{22} + bA_{23} + cA_{24}$$

$$A_{13} - \frac{A_{12}A_{23}}{A_{22}} = b\left(A_{33} - \frac{A_{23}^2}{A_{22}}\right) + c\left(A_{34} - \frac{A_{23}A_{24}}{A_{22}}\right)$$

$$A_{14} - \frac{A_{12}A_{24}}{A_{22}} = b\left(A_{34} - \frac{A_{23}A_{24}}{A_{22}}\right) + c\left(A_{44} - \frac{A_{24}^2}{A_{22}}\right)$$

d'où, en posant $A'_{13} = A_{13} - \dfrac{A_{12}A_{23}}{A_{22}}$, $A'_{33} = A_{33} - \dfrac{A_{23}^2}{A_{22}}, \ldots$

$$\left|\begin{array}{lcl} A_{12} & = & aA_{22} + bA_{23} + cA_{24} \\ A'_{13} & = & \phantom{aA_{22} +{}} bA'_{33} + cA'_{34} \\ A''_{14} & = & \phantom{aA_{22} +{}} bA'_{34} + cA''_{44} \end{array}\right.$$

De même, en tirant $b = \dfrac{A'_{13}}{A'_{33}} - c\dfrac{A'_{34}}{A'_{33}}$ et en reportant dans la 3e équation, on a ;

$$A'_{14} - \frac{A'_{13}A'_{34}}{A'_{33}} = c\left(A'_{44} - \frac{{A'_{34}}^2}{A'_{33}}\right)$$

d'où, en posant $A'_{14} - \dfrac{A'_{13}A'_{34}}{A'_{33}} = A''_{14}$ et $A'_{44} - \dfrac{{A'_{34}}^2}{A'_{33}} = A''_{44}$, $A''_{14} = cA''_{44}$

d'où le système :

$$\left|\begin{array}{lcl} A_{12} & = & aA_{22} + bA_{23} + cA_{24} \\ A'_{13} & = & \phantom{aA_{22} +{}} bA'_{33} + cA'_{34} \\ A''_{14} & = & \phantom{aA_{22} + bA'_{33} +{}} cA''_{44} \end{array}\right.$$

d'où : c, b et a.

D'autre part, la part de la variance de x_1 expliquée globalement par les variations de x_2, x_3 et x_4 est égale à $R^2 = \dfrac{aA_{12} + bA_{13} + cA_{14}}{A_{11}}$ (R étant le coefficient de corrélation multiple entre x_1 d'une part, x_2, x_3, x_4 d'autre part). Mais rien ne permet de déterminer quelle partie de cette part peut être attribuée à x_2 ou x_3 ou x_4, sinon une analyse un peu compliquée.

En effet posons :

$$R^2_{1,2} = \frac{a_1 A_{12}}{A_{11}} \text{ (Régression : } x_1 = a_1x_2 + b_1\text{)},$$

$$R^2_{1,23} = \frac{a_2 A_{12} + b_2 A_{13}}{A_{11}} \text{ (Régression : } x_1 = a_2x_2 + b_2x_3 + c_2\text{)},$$

$$R^2_{1,234} = \frac{a_3 A_{12} + b_3 A_{13} + c_3 A_{14}}{A_{11}} \text{ (Régression : } x_1 = a_3x_2 + b_3x_3 + c_3x_4 + d_3\text{)}$$

On peut dire que la part de la variance de x_1 expliquée par x_2 est $R^2_{1,2}$ et que celle expliquée par x_3, une fois éliminées les incidences de x_2, est égale à $R^2_{1,23} - R^2_{1,2}$. Appelons $R^2_{1,3(2=c^{te})}$

cette différence. On peut, de même, dire que la part de la variance de x_1 expliquée par x_4, une fois éliminées les incidences de x_2 et x_3, est égale à $R^2_{1,234} - R^2_{1,23}$. Appelons $R^2_{1,4\,(2\text{ et }3=C^{te})}$ cette différence. On peut écrire

$$R^2_{1,234} = R^2_{1,2} + R^2_{1,3\,(2=C^{te})} + R^2_{1,4\,(2\text{ et }3=C^{te})}$$

Mais on pourrait également écrire

$$R^2_{1,234} = R^2_{1,2} + R^2_{1,4\,(2=C^{te})} + R^2_{1,3\,(2\text{ et }4=C^{te})}$$

ou

$$R^2_{1,234} = R^2_{1,3} + R^2_{1,2\,(3=C^{te})} + R^2_{1,4\,(2\text{ et }3=C^{te})} \cdots$$

Évidemment, si x_2, x_3 et x_4 étaient indépendants, il n'y aurait pas de problème et le système analysé plus haut se ramènerait à

$$A_{12} = aA_{22}$$
$$A_{13} = bA_{33}$$
$$A_{14} = cA_{44}$$

d'où a, b, c, et les parts de la variance de x_1 à attribuer à x_2, x_3 et x_4 seraient respectivement

de $\frac{aA_{12}}{A_{11}}$, $\frac{bA_{13}}{A_{11}}$ et $\frac{cA_{14}}{A_{11}}$.

En pratique des variables x_2, x_3 et x_4 totalement indépendantes sont extrêmement rares et l'on est amené à recommencer le calcul des a, b, c... chaque fois que l'on ajoute ou retire un x_i dans la formule de régression. On ne sait jamais, à priori, si l'introduction d'une nouvelle variable x_k entraînera un ajustement « suffisamment meilleur pour être appréciable », c'est-à-dire un R^2 relativement plus élevé, et, il peut arriver que l'introduction d'une variable x_k entraîne une valeur de R^2 notablement plus élevée et que l'élimination de la variable x_j n'entraîne qu'une faible diminution de la valeur du R^2 nouvellement calculé etc... Pour tous ces essais les valeurs de a, b, c,... doivent être chaque fois recalculées.

La méthode, qui suit, consiste à utiliser de nouvelles variables indépendantes entr'elles. Supposons connus les A_{11} et A_{ij} comme précédemment et considérons une régression de la forme $\quad = aX_2 + bX_3 + cX_4$, en utilisant les variables centrées.

Posons : $X_1 = aX_2 + bX_3 + cX_4 + u$

Soient
$$\left|\begin{array}{l} \widehat{X}_1 = \lambda_1 X_2 \\ \widehat{X}_3 = \lambda_3 X_2 \\ \widehat{X}_4 = \lambda_4 X_2 \end{array}\right.$$

les droites de régression de X_1, X_3 et X_4 par rapport à X_2 et posons :

$$\left|\begin{array}{l} X_1 = \lambda_1 X_2 + X'_1 \\ X_3 = \lambda_3 X_2 + X'_3 \\ X_4 = \lambda_4 X_2 + X'_4 \end{array}\right.$$

X'_1, X'_3 et X'_4 sont indépendants de X_2.

De $X_1 = \lambda_1 X_2 + X'_1$ on tire $X_1X_2 = \lambda_1 X_2^2 + X_2X'_1$ d'où

$$\sum[X_1X_2] = \lambda_1\sum[X_2^2] + \sum[X_2X'_1] = \lambda_1\sum[X_2^2],$$

d'où $\lambda_1 = \dfrac{\Sigma[X_1\,X_2]}{\Sigma[X_2^2]} = \dfrac{A_{12}}{A_{22}}$

De même on a $\lambda_3 = \dfrac{A_{23}}{A_{22}}$ et $\lambda_4 = \dfrac{A_{24}}{A_{22}}$

Portons les expressions de X_1, X_3, X_4 dans la relation $X_1 = aX_2 + bX_3 + cX_4 + u$. On a

$$\lambda_1 X_2 + X'_1 = aX_2 + b(\lambda_3 X_2 + X'_3) + c(\lambda_4 X_2 + X'_4) + u,$$

d'où
$$\left|\begin{array}{l} \lambda_1 = a + b\lambda_3 + c\lambda_4 \\ X'_1 = bX'_3 + c'X_4 + u \end{array}\right.$$

De même soient $\left|\begin{array}{l} \widehat{X}'_1 = \lambda'_1 X'_3 \\ \widehat{X}'_4 = \lambda'_4 X'_3 \end{array}\right.$

les droites de régression de X'_1 et X'_4 par rapport à X'_3 et posons $\left|\begin{array}{l} X'_1 = \lambda'_1 X'_3 + X''_1 \\ X'_4 = \lambda'_4 X'_3 + X''_4 \end{array}\right.$

On a $\lambda'_1 = \dfrac{\Sigma[X'_1 X'_3]}{\Sigma[X'_3]^2}$. Montrons que l'on a $\lambda'_1 = \dfrac{A'_{13}}{A'_{33}}$. On a

$$\begin{aligned} \textstyle\sum[X'_1 X'_3] &= \textstyle\sum(X_1 - \lambda_1 X_2)(X_3 - \lambda_3 X_2) \\ &= \textstyle\sum[X_1 X_3] - \lambda_3 \sum[X_1 X_2] - \lambda_1 \sum[X_3 X_3] + \lambda_1 \lambda_3 \sum[X_2^2] \\ &= A_{13} - \frac{A_{23}}{A_{22}} A_{12} - \frac{A_{12}}{A_{22}} A_{23} + \frac{A_{12}}{A_{22}} \times \frac{A_{23}}{A_{22}} A_{22} = A_{13} - \frac{A_{12} A_{23}}{A_{22}} = A'_{13}. \end{aligned}$$

De même on a $\Sigma[X'^2_3] = \Sigma(X_3 - \lambda_3 X_2)^2 = \Sigma[X_3^2] - 2\lambda_3 \Sigma[X_2 X_3] + \lambda_3^2 \Sigma[X_2^2]$

$$= A_{33} - 2\frac{A_{23}}{A_{22}} A_{23} + \left(\frac{A_{23}}{A_{22}}\right)^2 A_{22} = A_{33} - \frac{A_{23}^2}{A_{22}} = A'_{33}$$

d'où $\lambda'_1 = \dfrac{A'_{13}}{A'_{33}}$

De même on a $\lambda'_4 = \dfrac{A'_{34}}{A'_{33}}$

Portons les expressions de $X'_1 = \lambda'_1 X'_3 + X''_1$ et $X'_4 = \lambda'_4 X'_3 + X''_4$ dans $X'_1 = bX'_3 + cX'_4 + u$. On en déduit $\lambda'_1 X'_3 + X''_1 = bX'_3 + c(\lambda'_4 X'_3 + X''_4) + u$

d'où
$$\left|\begin{array}{l} \lambda'_1 = b + c\lambda'_4 \\ X''_1 = cX''_4 + u \end{array}\right.$$

Soit $\widehat{X}''_1 = \lambda''_1 X''_4$ la droite de régression de X''_1 par rapport à X''_4 et posons $X''_1 = \lambda''_1 X''_4 + X''_1$. La relation $X''_1 = cX''_4 + u$ devient $\lambda''_1 X''_4 + X''_1 = cX''_4 + u$, d'où $\lambda''_1 = c$ et $X''_1 = u$,

d'où $c = \lambda''_1 = \dfrac{\Sigma[X''_1 X''_4]}{\Sigma[X''^2_4]} = \dfrac{A''_{14}}{A''_{44}}$

(pour la même raison que précédemment), d'où a et b par les formules : $\lambda'_1 = b + c\lambda'_4$ et $\lambda_1 = a + b\lambda_3 + c\lambda_4$.

On voit que, si l'on remplace les λ par leurs expressions, on retrouve les équations du système (1), à savoir :

$$(1) \quad \left|\begin{array}{lllll} A_{12} & = aA_{22} & + bA_{23} & + cA_{24} \\ A'_{13} & = & bA'_{33} & + cA'_{34} \\ A''_{14} & = & & cA''_{44} \end{array}\right.$$

Mais à l'opposé de la méthode classique cette méthode permet de calculer le R^2 (carré du coefficient de corrélation multiple) sans que l'on ait besoin de calculer les a, b, c,..., et surtout permet de remplacer la dernière variable considérée par une autre variable, sans qu'il soit nécessaire de recommencer tous les calculs. (Si l'on remplace X_4 par une autre variable X_5 les λ et λ' ne changent pas de valeurs.)

On déduit des relations précédentes :

$$X_1 = \lambda_1 X_2 + X_1' = \lambda_1 X_2 + (\lambda_1' X_3' + X_1'') = \lambda_1 X_2 + \lambda_1' X_3' + \lambda_1'' X_4'' + u,$$

d'où la régression : $\widehat{X}_1 = \lambda_1 X_2 + \lambda_1' X_3' + \lambda_1'' X_4'' = \frac{A_{12}}{A_{22}} X_2 + \frac{A_{13}'}{A_{33}'} X_3' + \frac{A_{14}''}{A_{44}''} X_{44}''$

régression que l'on peut prolonger en fonction des besoins.

On peut vérifier facilement que X_2, X_3', X_4'' et u sont indépendants. De l'expression

$$X_1 = \lambda_1 X_2 + \lambda_1' X_3' + \lambda_1'' X_4'' + u = \frac{A_{12}}{A_{22}} X_2 + \frac{A_{13}'}{A_{33}'} X_3' + \frac{A_{14}''}{A_{44}''} X_4 + u$$

on déduit $\Sigma[X_1^2] = \frac{A_{12}^2}{A_{22}^2} \Sigma[X_2^2] + \frac{A_{13}'^2}{A_{33}'^2} \Sigma[X_3'^2] + \frac{A_{14}''^2}{A_{44}''^2} \Sigma[X_4''^2] + \Sigma[u^2]$

ou $A_{11} = \frac{A_{12}^2}{A_{22}} + \frac{A_{13}'^2}{A_{33}'} + \frac{A_{14}''^2}{A_{44}''} + \Sigma[u^2]$

et le carré du coefficient de corrélation multiple $R^2_{1,234}$, qui représente la part de la variance de X_1 expliquée par la régression $\widehat{X}_1 = \lambda_1 X_2 + \lambda_1' X_3' + \lambda_1'' X_4''$, est égal à $1 - \frac{\Sigma[u^2]}{\Sigma[X_1^2]}$,

soit $R^2_{1,234} = \frac{1}{A_{11}} \left[\frac{A_{12}^2}{A_{22}} + \frac{A_{13}'^2}{A_{33}'} + \frac{A_{14}''^2}{A_{44}''} \right]$

D'autre part u et X_1''' étant égaux, le passage de l'une des méthodes à l'autre n'entraîne qu'une réorganisation interne, ce qui est normal puisque dans les deux cas on utilise la même quantité d'information. Et on en déduit que le $R^2_{1,234}$ calculé dans les 2 cas est le même.

Signalons que l'on peut encore écrire :

$$R^2_{1,234} = r_{12}^2 + (1 - r_{12}^2)\, r^2_{1,3\,(2=\mathrm{C^{te}})} + (1 - r_{12}^2)\left(1 - r^2_{13\,(2=\mathrm{C^{te}})}\right) r^2_{14\ (2\text{ et }3=\mathrm{C^{te}})}$$

avec r_{12} = coefficient de corrélation entre x_1 et x_2, $r_{13(2=\mathrm{C^{te}})}$ = coefficient de corrélation entre x_1 et x_3, x_2 étant constant, soit

$$r_{13\,(2=\mathrm{C^{te}})} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$r_{14(2\text{ et }3=\mathrm{C^{te}})}$ = coefficient de corrélation entre x_1 et x_4, x_2 et x_3 étant constants.

On a, en effet $R^2_{1,234} = \frac{1}{A_{11}} \left[\frac{A_{12}^2}{A_{22}} + \frac{A_{13}'^2}{A_{33}'} + \frac{A_{14}''^2}{A_{44}''} \right]$

Or on a $\frac{A_{12}^2}{A_{11} A_{22}} = r_{12}^2$

et comme on a $A_{11}' = A_{11} - \frac{A_{12}^2}{A_{22}} = A_{11}(1 - r_{12}^2)$,

on en déduit $\frac{A_{13}'^2}{A_{11} A_{33}'} = \frac{A_{11}'}{A_{11}} \cdot \frac{A_{13}'^2}{A_{11}' A_{33}'} = (1 - r_{12}^2)\, r^2_{13\,(2=\mathrm{C^{te}})}$

De même on a $\frac{A_{14}''^2}{A_{11} A_{44}''} = \frac{A_{11}'}{A_{11}} \cdot \frac{A_{11}''}{A'_{11}} \frac{A_{14}''^2}{A_{11}'' A_{44}''} = (1 - r_{12}^2)\left(1 - r^2_{13\,(2=\mathrm{C^{te}})}\right) r^2_{14\,(2\text{ et }3=\mathrm{C^{te}})}$

Nous retrouvons ainsi la formule énoncée plus haut :

$$R^2_{1,234} = R^2_{1,2} + R^2_{1,3\,(2=\mathrm{C^{te}})} + R^2_{1,4\ (2\text{ et }3=\mathrm{C^{te}})}$$

avec $R_{1,3\,(2=\mathrm{C^{te}})} = \sqrt{1 - r_{12}^2} \cdot r_{13\,(2=\mathrm{C^{te}})}$ et $R_{1,4\,(2\text{ et }3=\mathrm{C^{te}})} = \sqrt{(1 - r_{12}^2)(1 - r^2_{13\,(2=\mathrm{C^{te}})})} \cdot r_{1,4\,(2\text{ et }3=\mathrm{C^{te}})}$

Supposons maintenant que l'on a dénombré k-1 variables $x_2, x_3, \dots, x_k$ ayant avec x_1 des corrélations simples importantes. Supposons connu le stock de variances et covariances correspondant à ces k variables. On va chercher une régression de la forme $X_1 = \lambda_1 X_2 + \lambda'_1 X'_3 + \dots\dots + \lambda_1^{(k-2)} X_k^{(k-2)}$, les $X_2, X'_3, \dots, X_k^{(k-2)}$ étant définis comme précédemment.

On conçoit qu'on s'approchera le plus rapidement de la solution finale en choisissant pour X_2 le X_i dont les variations influent le plus sur X_1, c'est-à-dire le X_i pour lequel on a

$$R^2_{1,i} = \frac{A^2_{1i}}{A_{11} A_{ii}} \text{ maximum donc } \frac{A^2_{1i}}{A_{ii}} \text{ maximum,}$$

puisque tous les dénominateurs contiennent A_{11}. On cherchera ensuite à avoir $R^2_{1,ij}$ maximum, donc $R^2_{1,j(i=\mathrm{C^{te}})}$ maximum puisque $R^2_{1,j(i=\mathrm{C^{te}})}$ est égal à l'écart $R^2_{1,ij} - R^2_{1,i}$,

$$\text{donc } \frac{A'^2_{1j}}{A_{11} A'_{jj}} \text{ maximum, donc } \frac{A'^2_{1j}}{A'_{jj}} \text{ maximum}$$

et ainsi de suite.

On pourrait penser à priori que plus la régression contient de variables (Quand on ajoute une variable, R_2 croît ou reste stationnaire), plus l'ajustement a de valeur. Ceci est inexact. En effet le critière le plus intéressant pour apprécier la qualité de la régression est celui de la variance résiduelle minima. Or pour une régression de la forme

$$x_1 = a_0 + a_2 x_2 + \dots + a_k x_k,$$

si l'on pose $A_{ii} = \Sigma(x_i - \bar{x}_i)^2 = \Sigma X_i^2$ et $A_{ij} = \Sigma(x_i - \bar{x}_i)(x_j - \bar{x}_j) = \Sigma X_i X_j$, la part de la variance de x_1 expliquée par la régression :

$$x_1 = a_0 + a_2 x_2 + \dots + a_k x_k \text{ est égale à } \quad R^2 = \frac{a_2 A_{12} + a_{13} A_{13} + \dots + a_k A_{1k}}{A_{11}}$$

et la variance résiduelle de x_1 est égale à $\frac{A_{11}(1 - R^2)}{n - k}$, n étant le nombre d'années sur lesquelles porte l'analyse et k le nombre des variables du second membre augmenté de l. L'addition d'une variable x_{k+1} ne sera intéressante que si la variance résiduelle diminue, c'est à-dire si l'on a :

$$\frac{A_{11}(1 - R^2_{1,23\dots k(k+1)})}{n - (k+1)} < \frac{A_{11}(1 - R^2_{1,23\dots k})}{n - k} \quad \text{ou} \quad \frac{1 - R^2_{1,23\dots k(k+1)}}{n - k - 1} < \frac{1 - R^2_{1,23\dots k}}{n - k},$$

ce qui élimine les variables dont l'incidence sur R^2 est trop faible.

En résumé une méthode efficace pour obtenir rapidement une régression linéaire intéressante consiste à choisir, pour la première variable de la régression, la variable x_i, telle que $\frac{A^2_{1i}}{A_{ii}}$ est maximum, puis comme seconde variable la variable x_j telle que $\frac{A'^2_{1j}}{A'_{jj}}$ est maximum et de continuer jusqu'à ce qu'on obtienne pour x_1 la variance résiduelle minima. Mais il n'est pas prouvé que l'on obtiendra ainsi la variance résiduelle « minima minimarum ». On peut simplement penser que, la plupart du temps, ce sera également la variance résiduelle « minima minimarum ».

Remarque. Jusqu'à présent dans la régression $x_1 = ax_2 + bx_3 + \dots\dots$ on a supposé que les variables $x_1, x_2, x_3 \dots$ correspondaient toutes à la période d'observation t. Or par suite de la viscosité qui existe dans la transmission, de l'une à l'autre, des variations des variables économiques, il existe un certain déphasage entre les variations de $x_1, x_2, x_3, \dots$ [1]. Soient

1. Voir : *Journal de la Société de statistique de Paris*, n° 2, 1975, « Les corrélations déphasées et la méthode des rotations » par J.-J. Boulanger.

θ_2, θ_3, ... les retards moyens des variations de x_1 sur celles de x_2, x_3, ..., la ligne de régression prendra la forme : $(x_1)_t = a\,(x_2)_{t-\theta_2} + b\,(x_3)_{t-\theta_3} + \ldots$

ou si l'on préfère $x_1 = a\,(x_2)_{-\theta_2} + b\,(x_3)_{-\theta_3} + c\,(x_4)_{-\theta_4} + d$

(en se limitant à 3 variables incidentes)

En conclusion nous voudrions insister sur le fait que l'équation

$$x_1 = a\,(x_2)_{-\theta_2} + b\,(x_3)_{-\theta_3} + c\,(x_4)_{-\theta_4} + d \text{ (par exemple)}$$

doit être considérée comme l'équation fondamentale de l'évolution d'une économie, et ceci en regard des 2 objectifs suivants :

1° l'explication du passé;

2° la prévision de l'avenir. En effet, si les θ_2, θ_3, ... sont tous de l'ordre de 1 an ou supérieurs à 1 an, il sera facile de prévoir avec 1 an d'avance la valeur la plus probable de x_1.

D'autre part, comme on ne sait a priori quelles seront les variables x_i, x_j, ... qui conduiront au meilleur ajustement, plus on considérera au départ de variables susceptibles d'entrer dans la régression multiple, plus on aura de chances d'obtenir un meilleur ajustement; travail sans doute fastidieux, mais grandement facilité par les moyens actuels de mécanisation des calculs.