

REVUE DE STATISTIQUE APPLIQUÉE

E. MALINVAUD

Théorie de la régression simple

Revue de statistique appliquée, tome 11, n° 4 (1963), p. 49-75

http://www.numdam.org/item?id=RSA_1963__11_4_49_0

© Société française de statistique, 1963, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*
<http://www.numdam.org/>

THÉORIE DE LA RÉGRESSION SIMPLE ⁽¹⁾

E. MALINVAUD

Directeur de l'École Nationale de la Statistique
et de l'Administration Economique

Nous étudierons ici la méthode statistique appropriée au traitement d'un modèle dans lequel une variable endogène dépend linéairement d'une variable exogène et d'une perturbation aléatoire non observable. L'ajustement des moindres carrés de la première variable par rapport à la seconde apparaîtra comme le procédé convenant à l'estimation des paramètres.

La théorie qui va être présentée intéresse de nombreux domaines de la statistique appliquée. Nous nous référerons uniquement aux applications économétriques pour lesquelles elle a une importance particulière. En effet, parmi les nombreux modèles que l'on est amené à introduire pour l'analyse des données économiques, beaucoup constituent des généralisations plus ou moins simples de celui examiné ici.

Les résultats dont nous ferons état sont maintenant classiques.⁽²⁾ Seule notre présentation peut être originale.

I - LE MODELE

Une certaine grandeur x est considérée comme déterminée à partir d'une autre grandeur z . La dépendance supposée peut être bien représentée par une relation du type :

$$x = az + b + \varepsilon \quad (1)$$

dans laquelle a et b sont des coefficients inconnus, et ε un terme aléatoire non observable.

On dispose de T observations sur les valeurs prises par le couple des grandeurs x et z . Soit (x_t, z_t) une observation ($t = 1, 2, \dots, T$). Les données disponibles sont constituées par les $2T$ nombres :

- (1) A peu de choses près, cet article constitue le texte du chapitre III de l'ouvrage qui paraîtra très prochainement sous le titre "METHODES STATISTIQUES DE L'ECONOMETRIE". (Voir E. Malinvaud (1964) dans les références figurant in fine). La maison Dunod a bien voulu autoriser la publication anticipée de ces pages. Je l'en remercie vivement.
- (2) Le lecteur qui désire des références bibliographiques peut se reporter à E. Malinvaud (1964) notamment chapitre V et VI.

$$\begin{array}{cccc} x_1 & x_2 \dots & x_t \dots & x_T \\ z_1 & z_2 \dots & z_t \dots & z_T \end{array}$$

A partir de ces données, on veut estimer les coefficients a , b et certaines caractéristiques de la distribution des ε .

Par exemple x peut être l'épargne annuelle d'un ménage, z le revenu de ce ménage pendant la même année. Le modèle spécifie alors que l'épargne est déterminée à partir du revenu et d'autres éléments non observables, suivant une formule linéaire valable pour tous les ménages. Les données disponibles portent sur l'épargne x_t et le revenu z_t de T ménages repérés par un indice t .

Pour justifier le modèle, on admet souvent que la grandeur z observable et la grandeur ε non observable sont les "causes" qui déterminent la valeur prise par la grandeur x . Une formule linéaire du type (1) est souvent retenue à cause de sa simplicité, et faute d'une théorie qui préciserait la nature de la dépendance entre z et x . Le terme ε représente l'effet de tous les facteurs que nous ne pouvons ou ne savons pas identifier. Il est considéré comme aléatoire puisque notre ignorance sur ses modes de détermination lui donne la nature d'une variable aléatoire.

Par suite de la relation (1), la grandeur x est aléatoire comme ε . Pour toute valeur fixée de z , elle obéit à une loi de probabilité. Chaque valeur x_t est considérée comme une observation sur la variable aléatoire x , c'est-à-dire comme le résultat d'un tirage effectué suivant la distribution de probabilité de cette variable pour la valeur de z correspondante, soit z_t . De même, l'ensemble des valeurs x_1, x_2, \dots, x_T est traité comme un échantillon obtenu à partir des distributions correspondant respectivement à z_1, z_2, \dots, z_T .

L'espace échantillon est alors l'espace euclidien à T dimensions, dans lequel le point observé⁽¹⁾ a pour coordonnées x_1, x_2, \dots, x_T . La distribution de probabilité de ce point est déduite directement de celle de la variable x pour les T valeurs z_1, z_2, \dots, z_T de z . Elle dépend des paramètres inconnus : a , b et de certaines caractéristiques de la loi des ε . Le problème statistique consiste dans l'estimation de ces paramètres à partir de l'échantillon.

On peut encore présenter le modèle d'une autre manière, qui apparaît quelquefois plus suggestive, et dire : *le modèle spécifie la loi de probabilité conditionnelle de la variable endogène x pour toute valeur fixée de la variable exogène z* . L'échantillon permet l'estimation de caractéristiques de la loi de probabilité conditionnelle de x_1, x_2, \dots, x_T pour les valeurs z_1, z_2, \dots, z_T de z .

Cette formulation peut donner à penser que la variable exogène z est considérée comme aléatoire, et qu'il existe une loi de probabilité du couple (x, z) . En fait, les méthodes et propriétés que nous allons considérer s'appliquent aussi bien que la variable exogène soit certaine ou aléatoire. Elles ne font jamais intervenir que la loi de probabilité conditionnelle des x pour des valeurs fixées des z .

1) Le symbole x est utilisé ici pour désigner aussi bien la variable aléatoire, que l'ensemble des valeurs observées de cette variable. Puisque le lecteur en est averti, cela ne devrait pas provoquer de confusion.

On rencontre d'ailleurs des variables exogènes qui ne sauraient être considérées comme aléatoires. Ainsi, z_t désigne parfois le temps, mois ou année, de l'observation x_t . Quand z représente une grandeur économique, on peut souvent admettre que la valeur atteinte par cette grandeur résulte d'un phénomène en partie aléatoire. Mais ce phénomène fait alors lui-même intervenir d'autres grandeurs, qui y jouent le rôle de variables exogènes. Ainsi, même quand elle pourrait être considérée comme aléatoire, la variable z ne suivrait généralement pas une loi de distribution simple. De même, la loi de probabilité du couple (x, z) serait complexe. En la faisant intervenir dans la spécification du modèle, on compliquerait inutilement les choses.

II - HYPOTHESE FONDAMENTALE

Il nous faut maintenant préciser le modèle en formulant les hypothèses que nous ferons intervenir. Par souci de clarté, ces hypothèses ont été détaillées le plus possible. La liste complète en figure ci-dessous, bien que la plupart des résultats ne reposent que sur certaines d'entre elles. Ces hypothèses intéressent principalement la distribution des ε ; les deux dernières portent sur les valeurs des variables exogènes et sur les paramètres.

Hypothèse 1 - Les variables x_t et z_t (avec $t = 1, 2, \dots, T$) représentent des grandeurs numériques observées sans erreur. La variable x_t est aléatoire ; elle satisfait :

$$x_t = az_t + b + \varepsilon_t \quad \text{pour } t = 1, 2, \dots, T \quad (2)$$

égalité dans laquelle a et b sont deux coefficients numériques, et ε_t une variable aléatoire non observable dont l'espérance mathématique est nulle, quels que soient z_1, z_2, \dots, z_T .

Cette hypothèse reprend les indications générales sur le modèle, telles qu'elles ont été déjà données dans le paragraphe précédent. Notons cependant deux additions.

D'une part, il est précisé que les variables x et z sont observées sans erreur. S'il existait des écarts appréciables entre les valeurs observées et les valeurs vraies qui satisfont le modèle, il faudrait faire appel à d'autres méthodes statistiques, et à une autre analyse théorique. Bien que les données économiques soient indiscutablement entachées d'erreurs de mesure plus ou moins grandes, on estime souvent que ces erreurs restent négligeables en comparaison des éléments non observables influant la détermination des variables endogènes (les ε dans le modèle actuel).

Ainsi, les résultats relatifs à l'épargne et au revenu d'un groupe de ménages résultent généralement d'une enquête qui ne pouvait prétendre à une précision parfaite malgré le soin avec lequel elle fut menée. Le revenu et l'épargne enregistrés pour chaque ménage diffèrent quelque peu des grandeurs que l'on cherchait à mesurer. Mais les écarts sont le plus souvent très faibles vis-à-vis du terme ε qui, dans le modèle (1), représente l'effet sur l'épargne annuelle de tous les éléments autres que le revenu de l'année. C'est pourquoi, on convient de négliger les erreurs de mesure dans l'analyse économétrique de ces données.

D'autre part, il est spécifié dans l'hypothèse 1 que l'espérance mathématique des ε est nulle quelles que soient les valeurs prises par les

z. Faute d'une hypothèse de cette nature, les paramètres a et b du dèle ne seraient pas identifiables. Par exemple, une structure dans laquelle a = 1, b = 0, E(ε) = 0 conduirait à la même loi de probabilité conditionnelle des x qu'une autre structure dans laquelle a = 0, b = 1, E(ε) = z - 1, les ε suivant dans les deux cas une même distribution par rapport à leurs moyennes respectives. La nullité de l'espérance mathématique des ε est en somme indispensable si le modèle doit fournir le support d'analyses économétriques. C'est pourquoi elle a été imposée dès le départ dans l'hypothèse 1.

Concrètement, cette condition signifie que le terme additif ε du modèle prend des valeurs tantôt positives tantôt négatives, et cela pour toute valeur de la variable exogène. Elle n'est pas satisfaite si l'effet des facteurs non observables, ou non identifiés, est corrélé avec l'effet de la grandeur explicative x. C'est donc là une condition restrictive en pratique qu'il faudra garder présente à l'esprit dans toute application. On a souvent en économétrie, l'occasion de voir des cas importants dans lesquels elle ne convient pas, et d'envisager alors des méthodes statistiques différentes de celles du présent article.

On pourrait donner à l'hypothèse 1 une formulation qui permettrait d'éviter la difficulté d'identification signalée ici. Il suffirait d'énoncer directement que la distribution de probabilité conditionnelle de x_t pour des valeurs fixées de z_1, z_2, \dots, z_T a l'espérance mathématique $\alpha z_t + \beta$, soit(1)

$$E(x_t/z_1, z_2, \dots, z_T) = \alpha z_t + \beta \quad (3)$$

Les paramètres α et β seraient alors des caractéristiques de la loi de probabilité conditionnelle des x. Si nous connaissons les valeurs de ces paramètres et de z_t , nous obtenons directement la moyenne de x_t par $\alpha z_t + \beta$. Peu importe alors que αz_t soit en fait l'effet de la variable exogène (cas de la structure a = 1, b = 0, E(ε) = 0) ou celui d'autres facteurs non observables ou non identifiés (cas de la structure a = 0, b = 1, E(ε) = z - 1). La connaissance de z_t (et des paramètres α et β) suffit, dans un cas comme dans l'autre, à qui veut connaître la moyenne conditionnelle de x_t .

Cette manière de voir, préconisée par H. Wold (1960), est plus générale que celle retenue dans l'énoncé de l'hypothèse 1, puisqu'elle n'élimine pas les structures du type a = 0, b = 1, E(ε) = z - 1. Elle montre que les résultats établis ci-dessous peuvent recevoir une interprétation un peu différente de celle à laquelle nous nous en tiendrons.

Cette discussion se rattache d'ailleurs directement à l'étude de l'"identification". La connaissance de la loi de probabilité conditionnelle des variables endogènes suffit tant que le modèle ne subit aucun changement. Il n'est nécessaire d'identifier individuellement les paramètres de la loi conditionnelle que si on cherche à connaître la structure qui l'explique. Cette identification s'impose si on considère que certaines modifications doivent être apportées au modèle estimé avant son utilisation à des fins prévisionnelles.

De plus, il importe de formuler les hypothèses de base de telle façon qu'elles s'appliquent le plus directement possible aux modèles élaborés par la théorie économique, et qu'elles se prêtent à une discussion

(1) Dans le cas où l'hypothèse 1, dans sa formulation première, serait satisfaite, les paramètres a et b seraient évidemment égaux à α et β.

reposant sur l'examen direct des phénomènes représentés. Cela explique l'énoncé retenu pour l'hypothèse 1.

III - AUTRES HYPOTHESES

Considérons maintenant des limitations plus étroites portant sur la distribution des termes non observables ε_t , termes que nous appellerons "erreurs" quand cela ne prêterait pas à confusion.

Hypothèse 2 (Homoscédasticité) - L'erreur ε_t suit une distribution indépendante de t et des z_θ (pour $\theta = 1, 2, \dots, T$). Elle a une variance σ^2 .

Cette hypothèse précise la condition déjà incluse dans l'hypothèse 1 selon laquelle l'espérance mathématique de ε_t était la même, zéro, quelles que soient les valeurs prises par les z_θ . Ici, ε_t est supposé indépendant des z_θ ; c'est-à-dire que toute sa distribution de probabilité reste la même quelles que soient les valeurs prises par les z_θ . Les remarques présentées ci-dessus s'appliquent donc avec des exigences plus précises.

Dans l'hypothèse 1, il était implicitement admis que la distribution de ε_t avait une espérance mathématique. L'hypothèse 2 impose encore que la distribution ait une variance. C'est une condition peut-être contestable. Mais elle est considérée comme satisfaite dans la très grande majorité des applications pratiques.

Enfin, l'hypothèse 2 stipule encore que les erreurs ε_t suivent toutes la même distribution. On dit que les erreurs sont homoscédastiques quand leur distribution est indépendante de t et des z_θ . L'homoscédasticité peut être mise en défaut de diverses manières; on dit alors que les erreurs sont hétéroscédastiques. Parfois, des transformations simples sur les variables permettent d'éliminer les effets les plus importants de cette hétéroscédasticité.

Hypothèse 3 (Indépendance des diverses observations) - Les erreurs ε_t et ε_θ relatives à deux observations différentes quelconques t et θ sont indépendantes entre elles.

Cette hypothèse prend une importance particulière quand les observations se suivent dans le temps avec l'ordre 1, 2, ..., T. Il existe alors souvent une corrélation positive entre ε_t et l'erreur suivante ε_{t+1} , du fait que les facteurs non identifiés du phénomène agissent avec une certaine continuité et affectent souvent de manière analogue deux valeurs successives de la variable endogène. De même, l'indépendance des erreurs est douteuse si, comme il arrive souvent, les observations ont des origines diverses, et si on a des raisons de penser que les observations de même provenance sont affectées par certains facteurs communs non observables.

Hypothèse 4 (Normalité) - Les erreurs ε_t suivent une loi normale, ou de Laplace-Gauss.

Contrairement à ce que beaucoup croient, la normalité des ε_t ne constitue pas une hypothèse fondamentale pour la théorie présentée ici. Pour beaucoup de résultats l'existence de la variance seule importe. La normalité est cependant nécessaire à la justification complète de certaines des procédures proposées.

Hypothèse 5 (Variables exogènes) - Lorsque T augmente indéfiniment, la suite des valeurs prises par la variable exogène z_t (pour $t = 1, 2, \dots, \dots$ ad infinitum) est telle que sa moyenne

$$\bar{z} = \frac{1}{T} \sum_{t=1}^T z_t$$

et sa variance

$$\frac{1}{T} \sum_{t=1}^T (z_t - \bar{z})^2$$

tendent vers deux limites finies x_0 et s^2 . La limite s^2 est positive.

Cette hypothèse n'intervient évidemment que dans l'étude des propriétés asymptotiques des méthodes proposées, lorsque l'on suppose que le nombre T des observations augmente indéfiniment. Elle n'est pas toujours satisfaite, notamment si z_t est une fonction continuellement croissante, ou une fonction sinusoidale amortie de t . Les résultats asymptotiques cités ici pourraient être établis à partir d'hypothèses un peu moins restrictives. Mais ils n'ont pas grande importance pour la validité des méthodes étudiées. Mieux vaut donc se contenter d'une hypothèse simple.

Notons ici que nous admettrons toujours, sans l'écrire explicitement, que les T valeurs de z_t ne sont pas toutes égales à un même nombre z_0 . S'il en était ainsi, on ne pourrait identifier les deux paramètres a et b du modèle, puisque tous les x_t auraient alors une même distribution dont seule l'espérance mathématique $az_0 + b$ dépendrait des paramètres en question, et que même une connaissance parfaite de cette distribution et de cette espérance mathématique ne permettrait de déterminer que $az_0 + b$, et non a et b individuellement.

Hypothèse 6 - On ne dispose d'aucune information sur les coefficients numériques a et b , lesquels peuvent prendre a priori n'importe quelles valeurs positives, négatives ou nulles.

La formulation même du modèle, c'est-à-dire l'hypothèse 1, pourrait être considérée comme impliquant automatiquement cette nouvelle hypothèse. Nous l'écrivons ici séparément, car elle n'est pas satisfaite dans certains cas. L'étude théorique du modèle conduit parfois à faire apparaître que les coefficients doivent respecter certaines contraintes s'exprimant sous forme d'égalités ou d'inégalités. Il se peut par exemple que a et b doivent satisfaire $a + b = 1$, ou $a \geq 0$. Nous éliminons les cas de cette nature.

IV - L'AJUSTEMENT DES MOINDRES CARRES

Dans le cadre de notre modèle, les meilleures estimations de a et b sont fournies par les coefficients de la régression simple de x par rapport à z . Nous allons le montrer maintenant en étudiant les propriétés de ces estimations.

La régression de x par rapport à z est définie comme la relation linéaire $x^* = a^* z + b^*$, dans laquelle a^* et b^* sont les valeurs de a et b qui minimisent :

$$U = \sum_{t=1}^T (x_t - az_t - b)^2 \quad (4)$$

Des calculs simples conduisent aux expressions suivantes pour a^* et b^* :

$$a^* = \frac{\sum_{t=1}^T (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad (5)$$

$$b^* = \bar{x} - a^* \bar{z} \quad (6)$$

avec

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t \quad \bar{z} = \frac{1}{T} \sum_{t=1}^T z_t$$

Ces expressions sont linéaires par rapport aux valeurs x_t de la variable endogène. Désignons en effet par x^1, x^2, x^3 et x^4 quatre échantillons obtenus à partir de la même structure vraie et des mêmes valeurs z_1, z_2, \dots, z_T , conformément aux principes généraux exposés ci-dessus. Soient x_t^1, x_t^2, x_t^3 et x_t^4 les valeurs de x_t dans x^1, x^2, x^3 et x^4 , et, de même $a^{*1}, b^{*1}, a^{*2}, b^{*2}, a^{*3}, b^{*3}, a^{*4}, b^{*4}$ les valeurs des estimations (5) et (6). Le fait que a^* et b^* sont linéaires découle des deux observations suivantes :

- Si $x_t^3 = kx_t^1$ pour tout t avec un même nombre k convenable, alors aussi $a^{*3} = ka^{*1}$ et $b^{*3} = kb^{*1}$.

- Si $x_t^4 = x_t^1 + x_t^2$ pour tout t , alors aussi $a^{*4} = a^{*1} + a^{*2}$ et $b^{*4} = b^{*1} + b^{*2}$.

Grâce à la relation (2), on peut exprimer a^* et b^* en fonction des coefficients vrais a et b , des valeurs des z_t et des erreurs ε_t . On obtient directement d'après (2) :

$$x_t - \bar{x} = a(z_t - \bar{z}) + (\varepsilon_t - \bar{\varepsilon})$$

d'où, en reportant dans (5) :

$$a^* = a + \frac{\sum_{t=1}^T (\varepsilon_t - \bar{\varepsilon})(z_t - \bar{z})}{\sum_{t=1}^T (z_t - \bar{z})^2} = a + \frac{\sum_{t=1}^T \varepsilon_t (z_t - \bar{z})}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad (7)$$

ou encore :

$$a^* = a + \sum_{t=1}^T w_t \varepsilon_t \quad (8)$$

$$b^* = b + \bar{\varepsilon} - (a^* - a) \bar{z} \quad (9)$$

avec :

$$w_t = \frac{z_t - \bar{z}}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad (10)$$

En prenant, dans (8) et (9), les espérances mathématiques par rapport à la loi de probabilité de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$, pour des valeurs données de z_1, z_2, \dots, z_T , on obtient directement le résultat suivant.

Proposition 1 - L'hypothèse 1 étant vérifiée, les estimations a^* et b^* sont "centrées" puisqu'elles ont des espérances mathématiques égales aux valeurs vraies a et b .

De même, l'expression (8) permet d'écrire :

$$E (a^* - a)^2 = \sum_{t, \theta} w_t w_\theta E (\varepsilon_t \varepsilon_\theta)$$

Si les hypothèses 2 et 3 sont satisfaites (homoscédasticité et indépendance des observations), $E (\varepsilon_t \varepsilon_\theta)$ est nul quand $t \neq \theta$, égal à σ^2 quand $t = \theta$. Donc :

$$E (a^* - a)^2 = \sigma^2 \sum_t w_t^2 = \frac{\sigma^2}{\sum_{t=1}^T (z_t - \bar{z})^2}$$

Des calculs analogues s'appliquent pour $E (b^* - b)^2 = \sigma_{b^*}^2$ et

$$E [(a^* - a) (b^* - b)] = \text{Cov} (a^*, b^*)$$

D'où :

Proposition 2 - Si les hypothèses 1, 2 et 3 sont vérifiées, les moments du second ordre de a^* et b^* sont égaux à :

$$\sigma_{a^*}^2 = \frac{\sigma^2}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad \sigma_{b^*}^2 = \frac{\sigma^2}{T} \left[1 + \frac{T \bar{z}^2}{\sum_{t=1}^T (z_t - \bar{z})^2} \right] \quad (11)$$

$$\text{Cov} (a^*, b^*) = -\bar{z} \sigma_{a^*}^2 \quad (12)$$

Considérons maintenant le comportement de $a^* - a$ et $b^* - b$ quand le nombre des observations augmente indéfiniment avec une suite de valeurs z_t satisfaisant l'hypothèse 5. Les formules (11) montrent qu'alors les variances de a^* et de b^* tendent vers zéro. Ces estimations tendent vers les valeurs vraies a et b , en moyenne quadratique, et donc aussi en probabilité. D'où :

Proposition 3 - Si les hypothèses 1, 2, 3 et 5 sont vérifiées, les estimations a^* et b^* sont "convergentes".

Dénommons "résidu", et désignons par ε_t^* , l'écart entre la valeur observée x_t et la valeur déduite de la régression $a^* z_t + b^*$. Soit :

$$\varepsilon_t^* = x_t - a^* z_t - b^* = (a - a^*) z_t + (b - b^*) + \varepsilon_t \quad (13)$$

Ce résidu diffère de ε_t par un terme dépendant des "erreurs d'estimation" $a^* - a$ et $b^* - b$. Il a une espérance mathématique nulle du moment que l'hypothèse 1 est satisfaite. Il tend en probabilité vers ε_t si les hypothèses 2, 3 et 5 le sont aussi. La distribution empirique des ε_t^* tend alors en probabilité vers la distribution de probabilité des ε_t ; et les moments empiriques des ε_t^* vers les moments théoriques des ε_t .

En pratique, la variance σ^2 des erreurs est généralement inconnue, de même que la plupart des caractéristiques de distribution de ces erreurs. On peut penser à estimer σ^2 à partir de la variance empirique des ε_t^* , puisque celle-ci tend justement vers σ^2 dans les grands échantillons.

Pour étudier les petits échantillons, écrivons d'après (13) et (6) :

$$\varepsilon_t^* = (a - a^*) (z_t - \bar{z}) + (\varepsilon_t - \bar{\varepsilon})$$

La somme des carrés des résidus est alors égale à :

$$\sum_{t=1}^T \epsilon_t^{*2} = \sum_{t=1}^T (\epsilon_t - \bar{\epsilon})^2 + (a - a^*)^2 \sum_{t=1}^T (z_t - \bar{z})^2 + 2(a - a^*) \sum_{t=1}^T (z_t - \bar{z})(\epsilon_t - \bar{\epsilon})$$

ou, d'après (7) :

$$\sum_{t=1}^T \epsilon_t^{*2} = \sum_{t=1}^T (\epsilon_t - \bar{\epsilon})^2 - (a - a^*)^2 \sum_{t=1}^T (z_t - \bar{z})^2 \quad (14)$$

Dans le second membre, le premier terme a pour espérance mathématique $(T - 1) \sigma^2$ dès lors que les hypothèses 2 et 3 sont vérifiées ; la proposition 2 ci-dessus établit que le second terme a pour espérance mathématique σ^2 . Ainsi :

Proposition 4 - Si les hypothèses 1, 2 et 3 sont satisfaites, l'espérance mathématique de la somme des carrés des résidus est égale à $(T - 2) \sigma^2$. Si, de plus, l'hypothèse 5 est satisfaite, les résidus ϵ_t^* tendent en probabilité vers les valeurs vraies ϵ_t des erreurs.

La proposition 4 montre qu'un estimateur centré de σ^2 sera fournie par la "variance résiduelle", égale par définition à :

$$\sigma_\epsilon^{*2} = \frac{1}{T - 2} \sum_{t=1}^T \epsilon_t^{*2} \quad (15)$$

Pour obtenir la valeur de cette estimation, il n'est pas nécessaire de calculer les ϵ_t^* individuellement. La relation (6) montre que ϵ_t^* peut s'écrire :

$$\epsilon_t^* = (x_t - \bar{x}) - a^* (z_t - \bar{z}) \quad (16)$$

Et la relation (5) montre que la somme des carrés de cette expression s'écrit :

$$(T - 2) \sigma_\epsilon^{*2} = \sum_{t=1}^T (x_t - \bar{x})^2 - a^{*2} \sum_{t=1}^T (z_t - \bar{z})^2 \quad (17)$$

formule généralement utilisée pour le calcul de σ_ϵ^* .

Avec les hypothèses retenues dans la proposition 4, des estimateurs centrés $\sigma_{a^*}^{*2}$ et $\sigma_{b^*}^{*2}$ de σ_a^2 et σ_b^2 sont obtenus par simple substitution de σ_ϵ^{*2} à σ^2 dans les formules (11). Par exemple :

$$\sigma_{a^*}^{*2} = \frac{\sigma_\epsilon^{*2}}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad (18)$$

Les propositions précédentes justifient les calculs habituellement effectués à propos de l'estimation de a et de b .

Le résultat de ces calculs est généralement présenté sous la forme suivante :

$$x_t = a^* z_t + b^* \quad (19)$$

$(\sigma_{a^*}^*) \quad (\sigma_{b^*}^*)$

l'écart-type estimé de a^* et de b^* figurant entre parenthèses au-dessous des coefficients correspondants de la régression.

Bien entendu, ces résultats peuvent n'être guère valables si les hypothèses du modèle s'écartent trop de la réalité. Notamment, si l'hypothèse 3 n'est pas admissible, σ_a^* et σ_b^* constituent de mauvais estimateurs des écarts-types de a^* et b^* .

Il arrive souvent que l'on fasse figurer la valeur R^2 du carré du coefficient de corrélation à droite de la régression obtenue. Avec nos notations, R^2 peut s'écrire soit :

$$R^2 = \frac{\left[\sum (x_t - \bar{x}) (z_t - \bar{z}) \right]^2}{\sum (x_t - \bar{x})^2 \cdot \sum (z_t - \bar{z})^2}$$

soit :

$$R^2 = 1 - \frac{(T - 2) \sigma_e^{*2}}{\sum_{t=1}^T (x_t - \bar{x})^2} \quad (20)$$

Il faut cependant remarquer que, dans le cadre conceptuel retenu, R^2 n'a pas de signification particulièrement intéressante. Il ne renseigne pas directement sur la précision de l'ajustement, mais uniquement sur la forme du nuage de points utilisé. Avec les petits échantillons servant souvent aux analyses économétriques, R^2 peut être élevé sans que σ_a^* et σ_b^* soient faibles. Au contraire, un échantillon important conduit parfois à une estimation précise de a et de b , bien que R^2 reste modeste. (1)

V - UNE PROPRIÉTÉ IMPORTANTE DE LA MÉTHODE DES MOINDRES CARRÉS

Jusqu'à présent, nous avons décrit la méthode des moindres carrés et certaines propriétés des estimations auxquelles elle conduit. Mais nous n'avons pas montré en quoi cette méthode serait préférable à d'autres auxquelles on pourrait également penser. Les propositions 1 et 3 établissent bien que a^* et b^* sont centrés et convergents. Mais d'autres estimateurs auraient les mêmes propriétés. Il nous faut examiner maintenant des caractéristiques plus remarquables de la méthode.

Désignons un estimateur quelconque de a par $\hat{a} = f(x)$, x représentant le vecteur des T valeurs observées de la variable endogène x_t , et f une fonction numérique des T variables x_t . La fonction f peut dépendre des variables exogènes observables z_t , mais doit être indépendante des valeurs vraies des paramètres a et b puisque celles-ci sont inconnues par hypothèse. Nous dirons que l'estimateur est linéaire si :

$$f(kx) = k f(x) \quad \text{et} \quad f(x^1 + x^2) = f(x^1) + f(x^2) \quad (22)$$

quels que soient les échantillons x , x^1 , x^2 et le nombre k . Nous avons déjà observé que a^* et b^* étaient des estimateurs linéaires centrés. Nous allons maintenant démontrer qu'ils ont les variances les plus faibles de tous les estimateurs linéaires centrés de a et b .

(1) Nous verrons toutefois ci-dessous que R^2 permet l'application d'un test simple parfois intéressant (cf paragraphe 6).

Proposition 5⁽¹⁾ - Si les hypothèses 1, 2, 3 et 6 sont satisfaites, les estimateurs a^* et b^* ont des variances minimales parmi tous les estimateurs linéaires centrés de a et b .

Pour établir cette proposition, nous allons déterminer directement les estimateurs linéaires centrés à variance minimales, et vérifier que ce sont bien a^* et b^* . La forme la plus générale d'estimateurs linéaires est :

$$\hat{a} = \sum_{t=1}^T r_t x_t \quad \hat{b} = \sum_{t=1}^T s_t x_t \quad (23)$$

expressions dans lesquelles les r_t et s_t peuvent dépendre des valeurs prises par les variables exogènes z_1, z_2, \dots, z_T , mais non des valeurs des x_t .

L'espérance mathématique de x_t est égale à $az_t + b$. Par suite :

$$E(\hat{a}) = a \sum_{t=1}^T r_t z_t + b \sum_{t=1}^T r_t \quad E(\hat{b}) = a \sum_{t=1}^T s_t z_t + b \sum_{t=1}^T s_t$$

Pour que \hat{a} et \hat{b} soient des estimateurs centrés, il faut que ces deux expressions soient égales respectivement à a et b , et cela identiquement par rapport à a et b pour toutes valeurs numériques données à ces paramètres (cf. hypothèse 6). Comme r_t et s_t ne dépendent pas de a et b , il faut évidemment :

$$\left\{ \begin{array}{l} \sum_{t=1}^T r_t z_t = 1 \\ \sum_{t=1}^T r_t = 0 \end{array} \right. \quad \left\{ \begin{array}{l} \sum_{t=1}^T s_t z_t = 0 \\ \sum_{t=1}^T s_t = 1 \end{array} \right. \quad (24)$$

En explicitant x_t suivant la relation (2) et en tenant compte des relations (24), nous voyons que les équations (23) deviennent

$$\hat{a} - a = \sum_{t=1}^T r_t \epsilon_t \quad \hat{b} - b = \sum_{t=1}^T s_t \epsilon_t$$

d'où, en tenant compte des hypothèses 2 et 3 :

$$E(\hat{a} - a)^2 = \sigma^2 \sum_{t=1}^T r_t^2 \quad E(\hat{b} - b)^2 = \sigma^2 \sum_{t=1}^T s_t^2$$

Les estimateurs cherchés \hat{a} et \hat{b} sont donc définis à partir des coefficients r_t et s_t qui minimisent⁽²⁾ $\sum_{t=1}^T r_t^2$ et $\sum_{t=1}^T s_t^2$ sous les conditions (24).

(1) Nous pourrions également établir ici la propriété un peu plus forte suivante :

Proposition 5' - Si les hypothèses 1, 2, 3 et 6 sont satisfaites, toute forme linéaire $\lambda a^* + \mu b^*$ a une variance minimum dans l'ensemble des estimateurs linéaires dont l'espérance mathématique est égale à $\lambda a + \mu b$.

(2) Il est remarquable que les quantités à minimiser ne dépendent pas des paramètres inconnus a et b , de sorte que les estimateurs à variance minimum ne dépendront pas non plus de ces paramètres. C'est justement cette particularité qui constitue le fondement de la proposition 5.

Puisque la forme à minimiser est la somme des carrés des r_t (ou de s_t) et que les conditions de liaison sont linéaires en r_t (ou en s_t), nous obtenons le minimum unique en égalant à zéro les dérivées par rapport aux r_t (ou aux s_t) de la forme

$$\sum_{t=1}^T r_t^2 - 2 \alpha \sum_{t=1}^T r_t z_t - 2 \beta \sum_{t=1}^T r_t$$

(ou de la suivante) :

$$\sum_{t=1}^T s_t^2 - 2 \gamma \sum_{t=1}^T s_t z_t - 2 \delta \sum_{t=1}^T s_t$$

Dans ces expressions, α , β , γ et δ sont évidemment des multiplicateurs de Lagrange.

La dérivation conduit à :

$$r_t = \alpha z_t + \beta \qquad s_t = \gamma z_t + \delta$$

et les conditions (24) impliquent que α , β , γ et δ satisfassent :

$$\left\{ \begin{array}{l} \alpha \sum_{t=1}^T z_t^2 + \beta \sum_{t=1}^T z_t = 1 \\ \alpha \sum_{t=1}^T z_t + \beta T = 0 \end{array} \right. \qquad \left\{ \begin{array}{l} \gamma \sum_{t=1}^T z_t^2 + \delta \sum_{t=1}^T z_t = 0 \\ \gamma \sum_{t=1}^T z_t + \delta T = 1 \end{array} \right.$$

D'où :

$$\alpha = \frac{1}{\sum_{t=1}^T (z_t - \bar{z})^2} \qquad \gamma = \frac{-\bar{z}}{\sum_{t=1}^T (z_t - \bar{z})^2}$$

$$\beta = -\alpha \bar{z} \qquad \delta = -\gamma \bar{z} + \frac{1}{T}$$

$$r_t = \frac{z_t - \bar{z}}{\sum_{t=1}^T (z_t - \bar{z})^2} \qquad s_t = \frac{1}{T} - r_t \bar{z}$$

$$\hat{a} = \frac{\sum_{t=1}^T (z_t - \bar{z})(x_t - \bar{x})}{\sum_{t=1}^T (z_t - \bar{z})^2} \qquad \hat{b} = \bar{x} - \hat{a} \bar{z}$$

Ces deux dernières expressions établissent l'identité des estimateurs cherchés avec a^* et b^* respectivement, et fournissent la démonstration de la proposition 5.

Cette proposition montre que, si les ε_t sont homoscédastiques et indépendants entre eux, la méthode des moindres carrés conduit à des estimations dont la dispersion est aussi faible que possible, du moins pour autant qu'on envisage seulement les estimateurs linéaires centrés. Elle présente un intérêt particulier du fait qu'elle ne suppose rien sur la forme de la distribution de ε_t , si ce n'est l'existence d'une variance σ^2 . Elle est donc bien venue dans les applications économétriques, pour lesquelles il serait généralement difficile de préciser la nature des lois de distribution.

Convient-il de ne considérer que les estimateurs linéaires centrés ? Des estimateurs non linéaires ou non centrés seraient préférables à a^* et b^* s'ils conduisaient à des erreurs plus faibles que $a^* - a$ et $b^* - b$. En fait, la théorie des décisions statistiques montre que les estimations des moindres carrés sont douées de certaines propriétés optimales non seulement par rapport à l'ensemble des estimateurs linéaires, mais aussi par rapport à l'ensemble de tous les estimateurs possibles. (1)

Quoi qu'il en soit, le lecteur constate qu'il est faux de dire, comme on le fait souvent, que l'estimation par les moindres carrés suppose une distribution normale des erreurs ϵ_t . En fait la normalité est nécessaire seulement pour la justification de certains tests associés à la méthode des moindres carrés, et non pour celle des formules d'estimation (5) et (6).

VI - ROLE DE L'HYPOTHESE DE NORMALITE

Avec les hypothèses 1 à 4, la loi de distribution des T erreurs ϵ_t (pour $t = 1, 2, \dots, T$) a pour densité(2) :

(1) On a suggéré que le principe appliqué dans ce paragraphe pourrait être employé à la détermination d'estimateurs à variance minimum, parmi tous les estimateurs linéaires (cf Theil (1958) p. 496-497 et 532-534). On n'exigerait plus que \hat{a} et \hat{b} aient des espérances mathématiques égales aux valeurs vraies de a et b et on procéderait à la minimisation sans faire intervenir les conditions(24). Cette manière de poser le problème n'est guère fructueuse, car elle oblige à minimiser des expressions dans lesquelles a et b interviennent ; elle conduit en fait à des expressions qui ne sont pas de véritables estimateurs mais qui dépendent des paramètres inconnus a et b , et pas seulement des grandeurs observables.

Notons d'ailleurs que la démonstration présentée ci-dessus fait intervenir le fait que les r_t et les s_t ne dépendent pas de a et b . Si les r_t et s_t pouvaient être des fonctions de a et b , les conditions (24) ne s'imposeraient plus. La condition $E(\hat{a}) = a$ par exemple se traduit seulement par :

$$a \sum r_t z_t + b \sum r_t = a$$

et le pseudo-estimateur linéaire centré à variance minimum serait :

$$\hat{a} = \frac{\sum \left(z_t + \frac{b}{a} \right) x_t}{\sum \left(z_t + \frac{b}{a} \right)^2}$$

De même, le pseudo-estimateur linéaire à variance minimum (et non centré) peut s'écrire :

$$\hat{a} = \frac{\sum \left(z_t + \frac{b}{a} \right) x_t}{\sum \left(z_t + \frac{b}{a} \right)^2 + \frac{\sigma^2}{\sigma_a^2}}$$

2) L'expression $\exp \{x\}$ désigne la fonction exponentielle e^x .

$$\frac{1}{(2\pi)^{\frac{T}{2}}\sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2 \right\} \quad (25)$$

La distribution de l'échantillon des x_t a alors pour densité :

$$C \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (x_t - az_t - b)^2 \right\} \quad (26)$$

où C désigne la constante figurant devant l'exponentielle dans l'expression (25).

On peut écrire

$$(x_t - az_t - b)^2 = (x_t - a^*z_t - b^*)^2 + [(a^* - a)z_t + (b^* - b)]^2 \\ + 2(x_t - a^*z_t - b^*) [(a^* - a)z_t + (b^* - b)]$$

dont le dernier terme s'écrit aussi :

$$2[(x_t - \bar{x}) - a^*(z_t - \bar{z})][(a^* - a)(z_t - \bar{z}) + (b^* - b) + (a^* - a)\bar{z}]$$

expression dont la somme par rapport à t est nulle d'après les formules de définition de \bar{x} , \bar{z} et a^* . Après substitution dans (26), nous voyons que la densité de probabilité des x_t s'exprime aussi par :

$$C \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (x_t - a^*z_t - b^*)^2 \right\} \quad (27) \\ \exp \left\{ -\frac{1}{2\sigma^2} \left[(a^* - a)^2 \sum_{t=1}^T z_t^2 + 2T\bar{z}(a^* - a)(b^* - b) + T(b^* - b)^2 \right] \right\}$$

Cette densité de probabilité est le produit de deux facteurs dont le premier fait intervenir uniquement les variables aléatoires x_t et les estimateurs a^* et b^* , et le second uniquement de a^* et b^* et les paramètres vrais. Désignons la densité de probabilité de l'échantillon par

$f(x_1, x_2, \dots, x_T; a, b)$, le premier facteur de (27) par $H(x_1, x_2, \dots, x_T)$ et le second par $g(a^*, b^*; a, b)$.

Nous montrerons ci-dessous que $g(a^*, b^*; a, b)$ est égal à la densité de probabilité de a^* et b^* , du moins à une constante près. Il en résulte que a^* et b^* sont des estimateurs exhaustifs de a et b, c'est-à-dire qu'ils résument toute l'information que l'échantillon peut apporter sur les valeurs des paramètres a et b.

En effet, on peut toujours, grâce à un changement de variables, remplacer x_1, x_2, \dots, x_T par $a^*, b^*, y_3, y_4, \dots, y_T$, les y_t étant de nouvelles variables convenablement choisies. Si |J| désigne le jacobien de la transformation, la densité de probabilité de $a^*, b^*, y_3, y_4, \dots, y_T$ est égale à $f(x_1, x_2, \dots, x_T; a, b) |J|$. Cette densité peut aussi s'écrire

$$g(a^*, b^*; a, b) h(y_3, y_4, \dots, y_T/a^*, b^*; a, b)$$

expression dans laquelle h représente la densité de probabilité conditionnelle de y_3, y_4, \dots, y_T pour des valeurs données de a^* et b^* .

Or, nous voyons, d'après la forme de (27), que :

$$h(y_3, y_4, \dots, y_T/a^*, b^*; a, b) = \frac{1}{|J|} H(x_1, x_2, \dots, x_T)$$

Comme $|J|$ ne dépend évidemment pas des paramètres inconnus a^* et b^* , il résulte de l'identité ci-dessus que h ne dépend pas non plus de a et de b .

Ainsi, une fois a et b connus, la densité de probabilité des autres variables ne dépend plus des paramètres inconnus a et b . L'observation de ces autres variables ne saurait donc renseigner, en quelque manière que ce soit, sur les valeurs vraies de ces paramètres. C'est pourquoi on dit que a^* et b^* sont des estimateurs exhaustifs⁽¹⁾ de a et b .

Notons encore que, d'après la forme de (27),

$$\begin{bmatrix} \frac{\partial \text{Log } f}{\partial a} \\ \frac{\partial \text{Log } f}{\partial b} \end{bmatrix} = \begin{bmatrix} \frac{\partial \text{Log } g}{\partial a} \\ \frac{\partial \text{Log } g}{\partial b} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \sum_t z_t^2 & T\bar{z} \\ T\bar{z} & T \end{bmatrix} \begin{bmatrix} a^* - a \\ b^* - b \end{bmatrix}$$

La matrice qui figure dans le dernier membre peut être considérée comme celle donnant les dérivées seconde de la fonction $\rho(a, b) = a^2 \sum_t z_t^2 + 2 abT\bar{z} + b^2T$, fonction indépendante des x_1, x_2, \dots, x_r . Il en résulte que les estimateurs a^* et b^* exhaustifs et centrés, ont des variances inférieures, ou au plus égales, à celles de tous autres estimateurs réguliers centrés de a et b . (Pour établir ce résultat, se reporter à G. Darrois (1945) ou à H. Cramer (1946), Chapitre 32, § 3 et 6).

Il nous reste à montrer que le second facteur de (27) donne bien la densité de probabilité de a^* et b^* , et à compléter l'étude du cas dans lequel les ϵ_t suivent une distribution normale.

Effectuons dans l'espace échantillon un changement de coordonnées par rotation des axes, de telle façon que les deux premiers axes aient, dans l'ancien système, les coordonnées suivantes :

$$\frac{1}{\sqrt{T}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \quad \frac{1}{\left[\sum (z_t - \bar{z})^2 \right]^{1/2}} \begin{bmatrix} z_1 - \bar{z} \\ z_2 - \bar{z} \\ \vdots \\ \vdots \\ \vdots \\ z_r - \bar{z} \end{bmatrix}$$

Soient y_1, y_2, \dots, y_r les nouvelles coordonnées du point échantillon. On vérifie immédiatement que :

$$\begin{aligned} y_1 &= \bar{x} \sqrt{T} = (a^* \bar{z} + b^*) \sqrt{T} \\ y_2 &= a^* \left[\sum (z_t - \bar{z})^2 \right]^{1/2} \end{aligned}$$

Le point échantillon peut alors être considéré comme l'extrémité de la somme de deux vecteurs orthogonaux : x^* dont les coordonnées sont $(y_1, y_2, 0, \dots, 0)$ dans le nouveau système, ou encore les $a^*z_t + b^*$ dans

(1) "sufficient estimators" en anglais.

l'ancien ($t = 1, 2, \dots, T$), et ε^* dont les coordonnées sont $(0, 0, y_3, \dots, y_T)$ dans le nouveau système, ou les $\varepsilon_t^* = x_t - a^* z_t - b^*$ dans l'ancien.

Or, le Jacobien de la transformation des (x_1, x_2, \dots, x_T) dans les (y_1, y_2, \dots, y_T) est égal à 1, puisqu'une rotation est une transformation linéaire définie par une matrice orthogonale. L'expression (27) correspond aussi bien à la densité de probabilité des y_1, y_2, \dots, y_T qu'à celle des x_1, x_2, \dots, x_T . Le premier facteur dans (27) fait intervenir seulement les ε_t^* , c'est-à-dire les y_3, y_4, \dots, y_T ; le second facteur seulement a^* et b^* , c'est-à-dire, après changement de coordonnées, y_1 et y_2 (puisque z_t sont ici des éléments non aléatoires). Ainsi, la densité de probabilité du point échantillon est le produit de deux facteurs dont l'un dépend seulement de ε^* , l'autre seulement de x^* .

Nous voyons que ε^* et x^* sont indépendants et que chacun d'eux suit une loi normale, le premier dans un espace à $T - 2$ dimensions, le second dans un espace à 2 dimensions. Puisque x suit une loi normale sphérique, ce résultat aurait pu être déduit directement de l'observation selon laquelle x^* et ε^* sont les projections orthogonales de x sur deux sous-espaces orthogonaux entre eux : le sous-espace L engendré par les deux vecteurs e et z de coordonnées $e_t = 1$, et $z - \bar{z}$ de coordonnées $z_t - \bar{z}$ (donc aussi celui engendré par e et z), et le sous-espace complémentaire K . Le sous-espace L contient le vecteur y espérance mathématique du vecteur x , et de coordonnées $y_t = az_t + b$.

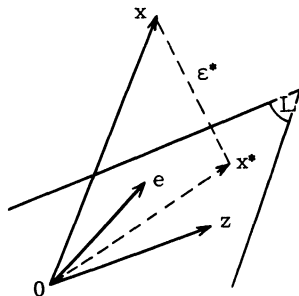


Fig. 1.

Les deux facteurs de (27) définissent les densités de probabilité de ε^* et x^* . Ainsi, on voit que les $T - 2$ coordonnées indépendantes non nulles de ε^* (y_3, y_4, \dots, y_T) suivent $T - 2$ lois normales indépendantes de moyennes nulles et d'écart-types égaux à σ . La quantité

$$\frac{1}{\sigma^2} \sum_{t=1}^T \varepsilon_t^{*2} \quad (28)$$

égale au rapport à σ^2 de la somme des carrés des résidus suit donc une loi de χ^2 à $T - 2$ degrés de liberté.

De même, l'expression du second facteur montre que a^* et b^* ont pour moyenne les valeurs vraies a et b , et pour matrice des covariances l'inverse de :

$$\frac{1}{\sigma^2} \begin{bmatrix} \sum_{t=1}^T z_t^2 & T\bar{z} \\ T\bar{z} & T \end{bmatrix} \quad (29)$$

inverse dont les éléments sont justement égaux aux expressions données par les formules (11) et (12).

Nous pouvons dès lors énoncer :

Proposition 6 - Si les hypothèses 1 à 4 et 6 sont satisfaites, les estimateurs centrés a^* et b^* obtenus par la régression des moindres carrés constituent des estimateurs exhaustifs de a et de b . Ils ont les variances les plus faibles possibles pour des estimateurs réguliers centrés⁽¹⁾. Leur distribution est normale, et indépendante de celle des résidus. La somme des carrés des résidus, divisée par σ^2 , suit une loi de χ^2 à $T - 2$ degrés de liberté.

Cette proposition, qui définit entièrement la loi de probabilité des estimations a^* et b^* , fournit la justification de tests et d'intervalles de confiance établis fréquemment.

Elle montre en particulier que $\frac{a^* - a}{\sigma_{a^*}}$ suit une loi normale centrée réduite, et que $(T - 2) \frac{\sigma_{a^*}^2}{\sigma^2} = (T - 2) \frac{R^2}{1 - R^2}$ suit une loi de χ^2 à $T - 2$ degrés de liberté, loi indépendante de celle de $\frac{a^* - a}{\sigma_{a^*}}$. Il en résulte que $\frac{a^* - a}{\sigma_{a^*}}$ suit une loi de Student à $T - 2$ degrés de liberté. Pour tester l'égalité de la valeur vraie a à une valeur fixée à l'avance a_0 , on peut donc comparer $\frac{a^* - a_0}{\sigma_{a^*}}$ à la valeur t_α donnée par la table de Student pour $T - 2$ degrés de liberté et la valeur α du seuil de signification. De même, un intervalle de confiance à 90 chances sur 100 par exemple, est donnée par l'ensemble des valeurs de a satisfaisant :

$$|a - a^*| \leq t_{0,05} \sigma_{a^*} \quad (30)$$

où $t_{0,05}$ est la valeur donnée par la table de Student pour le seuil de signification 5 %.

A l'aide des formules (18) et (20), on vérifie aisément que

$$\frac{a^*}{\sigma_{a^*}^2} = (T - 2) \frac{R^2}{1 - R^2}$$

(1) De même, toute forme linéaire $\lambda a^* + \mu b^*$ a une variance minimum dans l'ensemble des estimateurs dont l'espérance mathématique est égale à $\lambda a + \mu b$.

Notons encore que, dans la démonstration donnée ci-dessus, nous avons montré que a^* , b^* constituent des estimateurs exhaustifs, centrés et de variance minimum, en raisonnant comme si σ^2 était connu. Nous avons préféré ne pas alourdir exagérément l'exposé. En fait, quand σ^2 doit être estimé lui aussi, on peut établir que a^* , b^* et σ_c^{*2} constituent des estimateurs exhaustifs centrés de a , b et σ^2 . Les variances de a^* et b^* sont encore minimum (dans la classe de tous les estimateurs centrés), mais celle de σ_c^{*2} n'est plus minimum qu'asymptotiquement. (Sur ce sujet, le lecteur peut transposer la démonstration donnée dans H. Cramer (1946), chapitre 32, § 6, exercice 1).

Le terme de droite dépend uniquement de T et du coefficient R^2 . Il permet un test de l'hypothèse $a = 0$. En ce sens, la valeur de R^2 renseigne sur le caractère significatif de la dépendance entre z et x .

Les mêmes procédés s'appliqueraient pour la construction de tests ou de régions de confiance qui porteraient sur la valeur vraie du coefficient b indépendamment de la valeur de a . Mais il est également possible de tester simultanément les deux coefficients a et b .

En effet, comme le couple (a^*, b^*) obéit à une loi normale à deux dimensions de moyenne (a, b) on sait que :

$$\delta = \begin{bmatrix} a^* - a \\ b^* - b \end{bmatrix} \begin{bmatrix} \sigma_{a^*}^2 & \text{cov}(a^*, b^*) \\ \text{cov}(a^*, b^*) & \sigma_{b^*}^2 \end{bmatrix}^{-1} \begin{bmatrix} a^* - a \\ b^* - b \end{bmatrix}$$

suit une loi de χ^2 à deux degrés de liberté. L'inverse de la matrice des covariances est donné ici par la matrice (29), de sorte que δ prend la forme :

$$\delta = \frac{1}{\sigma^2} \left[\sum_{t=1}^T z_t^2 (a^* - a)^2 + 2 T \bar{z} (a^* - a) (b^* - b) + T (b^* - b)^2 \right] \quad (31)$$

La proposition 6 montre que $\frac{\sigma_{\epsilon}^{*2}}{\sigma^2}$ suit une loi de χ^2 à $T - 2$ degrés de liberté, loi indépendante de celle de σ . Ainsi :

$$F = \frac{\delta}{2} : \frac{1}{T - 2} \cdot \frac{\sigma_{\epsilon}^{*2}}{\sigma^2} = \frac{T - 2}{2} \cdot \frac{\sigma^2 \delta}{\sigma_{\epsilon}^{*2}} \quad (32)$$

suit une loi de Fisher avec les degrés de liberté 2 et $T - 2$.

Pour tester l'égalité de a et b à deux valeurs a_0 et b_0 fixées à l'avance, on peut calculer la valeur que prend F lorsqu'on substitue, dans δ , a_0 et b_0 à a et b , et comparer cette valeur à celle F_α que donne la table de Fisher pour le seuil de signification α , et les degrés de liberté 2 et $T - 2$. De même, sur un graphique à deux dimensions, portant a en abscisse et b en ordonnées, une région de confiance au seuil de signification α pour le couple (a, b) est constituée par le domaine intérieur à l'ellipse E d'équation $F = F_\alpha$.

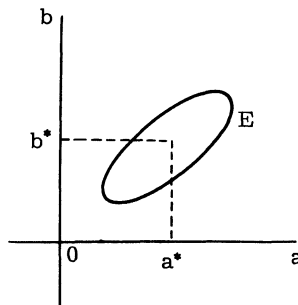


Fig. 2.

VII - PROPRIETES DE LA METHODE DES MOINDRES CARRES

Reprenons ici rapidement les résultats qui viennent d'être établis et complétons les par quelques indications annexes. Nous ferons ainsi mieux apparaître les avantages et les limitations que présentent les estimations des moindres carrés pour le traitement statistique du modèle linéaire.

Si les hypothèses 1 à 4 sont satisfaites, toutes les procédures statistiques associées habituellement à l'ajustement des moindres carrés sont entièrement justifiées.

Les estimateurs a^* et b^* apparaissent comme exhaustifs pour l'estimation des paramètres a et b . Leur variance est alors la plus faible possible pour des estimateurs centrés. La variance σ^2 des erreurs est estimée sans biais par la "variance résiduelle" σ_a^{*2} .

Dans ce cas, on connaît suffisamment bien la distribution des estimations pour pouvoir appliquer des tests rigoureux ou construire des régions de confiance exactes. Le couple (a^*, b^*) obéit à une loi normale dont la moyenne est (a, b) et dont la matrice des covariances peut être estimée sans erreur systématique (formule (18) par exemple). Des expressions telles que $\frac{(a^* - a)}{\sigma_a^*}$ sont distribuées suivant des lois de Student et permettent l'application de tests classiques.

Si seules les hypothèses 1, 2 et 3 sont satisfaites, l'ajustement des moindres carrés fournit encore en pratique les meilleures estimations dont nous puissions disposer ; a^* et b^* ont des variances inférieures ou au plus égales à celles de tous autres estimateurs linéaires centrés. On a également pu démontrer qu'avec une fonction de perte quadratique, ces estimateurs sont "minimax" et "admissibles" dans la classe de tous les estimateurs possibles.

En pratique, les tests et régions de confiance définis ci-dessus peuvent encore être valablement appliqués. Bien qu'ils ne soient pas rigoureusement exacts, ils reposent sur des approximations qui suffisent toujours et qui sont même excellentes quand le nombre des observations est grand.

D'une part, on a pu montrer que ces tests étaient peu sensibles aux écarts par rapport à la normalité. Si les erreurs suivent une loi non normale, mais douée d'une variance (hypothèse 2), on ne commet que des erreurs généralement faibles, sur les puissances ou les seuils de signification, en retenant les formules valables rigoureusement dans le cas d'erreurs normales. En somme, ces tests et ces régions de confiance sont "robustes" vis-à-vis de l'hypothèse 4.

D'autre part, on a pu démontrer que, moyennant l'hypothèse additionnelle 5, les distributions asymptotiques de $a^* \sqrt{T}$ et $b^* \sqrt{T}$ sont normales. Quand le nombre des observations augmente indéfiniment, les tests et régions de confiance tendent donc à devenir exacts, même si les erreurs ne suivent pas une loi normale.

Si seule l'hypothèse 1 est satisfaite, et si de plus la distribution des erreurs a des moments du second ordre, a^* et b^* ont bien toujours des espérances mathématiques égales aux valeurs vraies a et b ; mais aucune des autres propriétés ne s'applique en général. Il peut arriver que l'on connaisse la matrice des covariances de l'ensemble des T erreurs $\epsilon_1, \epsilon_2, \dots, \epsilon_T$ (matrice carrée d'ordre T). Dans ce cas, on peut trouver

une méthode d'estimation linéaire qui diffère en général de la méthode des moindres carrés, mais jouit des propriétés intéressantes examinées ci-dessus et se prête à des tests analogues. Il peut arriver encore que l'on connaisse imparfaitement la matrice des covariances des $\varepsilon_1 \dots \varepsilon_T$, mais que l'on puisse formuler sur elle des hypothèses telles qu'une estimation des caractéristiques inconnues devienne possible et permette une meilleure détermination de a et de b. En particulier, l'hypothèse 3 n'est pas satisfaite quand les ε_t émanent d'un processus aléatoire à liaison temporelle, c'est-à-dire quand l'erreur affectant l'observation de la période t est liée à celle affectant l'observation de la période t - 1.

L'hypothèse 1 reste évidemment fondamentale pour l'application de la méthode des moindres carrés.

Les propriétés que nous venons d'énoncer permettent d'apprécier les qualités de la méthode des moindres carrés. Pour l'étude économétrique, l'avantage principal de cette méthode réside dans le fait qu'elle donne de bons résultats sans exiger d'hypothèses trop strictes sur la loi de probabilité des variables. Son domaine d'application est ainsi assez vaste. L'économètre qui dispose rarement d'informations sur la nature exacte des lois de probabilité pourra y avoir généralement recours sans craindre de commettre de trop graves erreurs.

VIII - REGRESSIONS DANS LES LOIS A DEUX DIMENSIONS

Les propriétés de la régression simple viennent d'être étudiées dans le cadre du modèle linéaire défini au début de ce chapitre. Bien entendu, une étude similaire serait possible avec d'autres modèles. Nous aurons notamment l'occasion de nous demander ce que deviennent certaines des propriétés dégagées ci-dessus quand l'hypothèse 1 n'est plus entièrement satisfaite.

Il arrive souvent, en statistique mathématique, que la régression simple soit envisagée et justifiée d'une manière différente de celle retenue ici, mais à partir de modèles ou d'hypothèses assez voisines. Pour éviter les confusions, nous allons consacrer quelques lignes à ces autres points de vue.

La variable z est alors considérée comme aléatoire au même titre que x. Plus précisément, le couple (x, z) suit une loi de probabilité à deux dimensions. Les données $(x_1, z_1), (x_2, z_2) \dots (x_T, z_T)$ apparaissent comme T observations indépendantes du vecteur aléatoire (x, z).

Nous avons déjà indiqué que ce point de vue n'était pas nécessairement contradictoire avec celui adopté dans notre modèle. Peu importe que z soit aléatoire si nos hypothèses s'appliquent à la probabilité conditionnelle des x pour des valeurs données des z_t et si nous nous intéressons uniquement à la distribution conditionnelle de a^* et de b^* pour ces mêmes valeurs des z_t .

Mais, avec les modèles dans lesquels le couple (x, z) suit une même loi de probabilité valable pour toutes les observations, on s'intéresse toujours à la distribution non conditionnelle des coefficients de régression a^* et b^* .

C'est une attitude bien naturelle, mais c'est aussi une source supplémentaire de divergence avec la théorie présentée ci-dessus.

Considérons alors trois modèles de plus en plus restrictifs, et examinons les propriétés de a^* et b^* dans chacun d'eux.

(i) Les données (x_t, z_t) sont T observations indépendantes d'une même variable aléatoire (x, z) possédant des moments du second ordre⁽¹⁾.

A partir de la loi de probabilité à deux dimensions du couple (x, z) , on peut définir la "régression linéaire théorique de x par rapport à z " comme la droite $x = az + b$ dont les coefficients a et b rendent minimum l'expression :

$$E(x - az - b)^2 \quad (33)$$

Cette régression linéaire doit être soigneusement distinguée de la "régression théorique de x par rapport à z ", laquelle est définie comme :

$$x = E(x/z) \quad (34)$$

et n'a évidemment a priori aucune raison d'être linéaire.

Si ξ et ζ désignent les espérances mathématiques de x et z , les coefficients a et b de la régression linéaire théorique satisfont :

$$a = \frac{E[(x - \xi)(z - \zeta)]}{E[(z - \zeta)^2]} \quad b = \xi - a\zeta \quad (35)$$

relations que l'on établit sans peine.

Si le nombre des observations T augmente indéfiniment, les moyennes \bar{x} et \bar{z} tendent en probabilité vers les espérances mathématiques ξ et ζ les moments empiriques

$$\frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})(z_t - \bar{z}) \quad \text{et} \quad \frac{1}{T} \sum_{t=1}^T (z_t - \bar{z})^2$$

tendent en probabilité vers les moments théoriques

$$E[(x - \xi)(z - \zeta)] \quad \text{et} \quad E[(z - \zeta)^2]$$

Il en résulte que les coefficients a^* et b^* de la régression empirique tendent en probabilité vers les coefficients a et b de la régression linéaire théorique.

(ii) Les conditions (i) sont satisfaites, et, de plus, la régression théorique de x par rapport à z est linéaire.

Ecrivons cette régression théorique :

$$x = E(x/z) = az + b \quad (36)$$

et vérifions que les coefficients a et b minimisent bien l'expression (33). Soient en effet deux autres nombres quelconques α et β . Nous pouvons écrire

 (1) Herman Wold (1952) a proposé les expressions "spécification de Galton-Yule", et "spécification de Gauss-Fisher" pour désigner d'une part le modèle défini ici en (i), d'autre part le modèle linéaire étudié dans le reste de cet article.

$$E (x - \alpha z - \beta)^2 = E (x - az - b)^2 + E [(a - \alpha)z + (\beta - b)]^2 + E (H)$$

avec

$$H = 2 (x - az - b) [(a - \alpha)z + (\beta - b)]$$

Or, d'après (36), $E (H/z) = 0$, donc aussi $E (H) = 0$. Le minimum de $E (x - \alpha z - \beta)^2$ est alors bien atteint quand $\alpha = a$ et $\beta = b$.

Pour traiter ce cas, nous pouvons utiliser les résultats établis avec le modèle linéaire $x = az + b + \varepsilon$. Les hypothèses 1 et 3 sont automatiquement satisfaites. L'hypothèse 2 le serait aussi, si nous imposions quelques autres restrictions sur la distribution de (x, z) .

En particulier, nous savons que a^* et b^* sont des estimateurs centrés de a et b , conditionnellement pour des valeurs fixées de z_1, z_2, \dots, z_T , donc aussi inconditionnellement :

$$E (a^*) = E [E (a^*/z_1, z_2, \dots, z_T)] = E (a) = a$$

Quand l'hypothèse 2 est satisfaite, a^* et b^* ont des variances plus faibles que tous autres estimateurs linéaires centrés, \hat{a} et \hat{b} par exemple. En effet, nous savons que : $\text{Var} (a^*/z_1, \dots, z_T) \leq \text{Var} (\hat{a}/z_1, \dots, z_T)$ quels que soient $z_1 ; z_2, \dots, z_T$. Il en résulte que :

$\text{Var} (a^*) = E [\text{Var} (a^*/z_1, \dots, z_T)] \leq E [\text{Var} (\hat{a}/z_1, \dots, z_T)] = \text{Var} (\hat{a})$ et de même pour b^* et \hat{b} .

(iii) Les conditions (i) sont satisfaites, et, de plus, la loi de (x, z) est normale.

On sait qu'alors la régression théorique de x par rapport à z est linéaire, de sorte que (36) s'applique ici encore. De plus, la distribution du couple (x, z) dépend seulement de cinq paramètres, les deux espérances mathématiques et les trois moments théoriques du second ordre. On peut démontrer que, pour l'estimation de ces paramètres, et donc aussi des coefficients de la régression théorique, les deux moyennes et les trois moments empiriques du second ordre constituent un résumé exhaustif.

Les estimateurs a et b reçoivent alors des justifications analogues à celles présentées dans le cadre du modèle linéaire (cf paragraphe 7 ci-dessus). Nous pouvons même déduire leurs propriétés de celles obtenues avec ce modèle. Il est en effet facile de vérifier (1) que les hypothèses 1 à 4 sont satisfaites avec la spécification (iii).

Nous savons que la distribution conditionnelle de a^* et b^* , pour des valeurs données des z_1, z_2, \dots, z_T est normale. Comme les z obéissent à une loi normale, la distribution inconditionnelle (ou marginale) de a^* et b^* est normale elle aussi. De même, $\frac{a^* - a}{\sigma_{a^*}}$ a pour distribution condi-

 (1) Considérons le couple (ε, z) avec $\varepsilon = x - az - b$. Sa distribution est normale. La corrélation entre ε et z est nulle car

$$E [\varepsilon(z - \zeta)] = E [(x - \xi) (z - \zeta)] - a E [z - \zeta]^2 = 0$$

La variable aléatoire normale est donc indépendante de z . Les T valeurs $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ sont évidemment indépendantes entre elles.

tionnelle la loi de Student à $T - 2$ degrés de liberté. Comme cette distribution ne dépend pas des valeurs des z_1, z_2, \dots, z_T c'est aussi la distribution de $\frac{a^* - a}{\sigma_{a^*}^2}$ marginale. Des tests et des intervalles de confiance sur a sont donc établis ici exactement de la même façon que si nous adoptions le point de vue du paragraphe 7.

Les moments du second ordre de a^* et b^* s'obtiennent sans difficulté à la suite de déductions analogues. Ainsi :

$$\text{Var} (a^*) = E [\text{Var} (a^*/z_1, z_2, \dots, z_T)]$$

$$\text{Var} (a^*/z_1, z_2, \dots, z_T) = \frac{\sigma^2}{\sum_t (z_t - \bar{z})^2}$$

De plus, $\sigma^2 = E (\varepsilon^2 / z)$ ne dépend pas de z (cf renvoi (1) ci-dessus)
Donc :

$$\text{Var} (a^*) = \sigma^2 E \left[\frac{1}{\sum_t (z_t - \bar{z})^2} \right]$$

et, en première approximation :

$$\text{Var} (a^*) \sim \frac{\sigma^2}{T \sigma_z^2}$$

où σ_z^2 désigne évidemment la variance de z .

Notons pour terminer que l'approche examinée ici présente peu d'intérêt pour l'analyse économétrique ; car il est rare que l'on puisse considérer les observations (x_t, z_t) comme des tirages indépendants d'une même variable aléatoire à deux dimensions. Nous n'en disons pas plus sur ce point de vue. Le lecteur qui désirerait l'étudier pourra se reporter par exemple au mémoire de R. Féron (1956).

IX - PREVISION

Avec le modèle de ce chapitre, le problème prévisionnel peut être ainsi posé : connaissant la valeur z_θ de la variable exogène pendant une période future θ ou pour une unité non observée θ déterminer la valeur correspondante x_θ de la variable endogène. Il se peut qu'en réalité la valeur z_θ ne soit pas connue parfaitement, mais résulte elle-même d'une prévision. Nous raisonnerons cependant ici comme si z_θ n'était entachée d'aucune erreur ; ou, plus précisément, nous nous intéresserons à la prévision conditionnelle des valeurs prises par la variable endogène, pour des valeurs données de la variable exogène.

Nous retiendrons le modèle défini dans les premiers paragraphes de cet article, et admettrons que les paramètres a et b ont été estimés à l'aide d'un échantillon de T observations ($t = 1, 2, \dots, T$), échantillon ne contenant évidemment pas θ . Nous distinguerons la valeur prévue x_θ^p de la valeur x_θ qui se réalisera effectivement, et analyserons les propriétés de l'erreur de prévision $x_\theta^p - x_\theta$.

Cette erreur est évidemment aléatoire puisque x_θ l'est d'après la formulation même du modèle, et que x_θ^p dépend d'estimations aléatoires

des paramètres a et b . Nous caractériserons la distribution de l'erreur de prévision par une loi de probabilité qui sera conditionnelle à la fois par rapport à z_θ et par rapport à z_1, z_2, \dots, z_T .

Il serait également possible d'étudier les prévisions avec le point de vue présenté dans le paragraphe précédent. Il faudrait alors étudier pour $x_\theta^p - x_\theta$ une distribution qui ne serait plus conditionnelle que par rapport à z_θ . Mais, comme nous l'avons vu, les phénomènes économiques sont rarement justiciables des modèles qu'il faudrait alors retenir.

La formule la plus simple et la plus habituelle pour le calcul de la prévision x_θ^p repose sur l'application directe de la régression. Elle conduit à :

$$x_\theta^p = a^* z_\theta + b^* \quad (37)$$

L'erreur de prévision s'exprime alors par :

$$x_\theta^p - x_\theta = (a^* - a)z_\theta + (b^* - b) - \varepsilon_\theta \quad (38)$$

La formulation du modèle (hypothèse 1) et le fait que $E(a^*) = a$, $E(b^*) = b$ (proposition 1) justifient :

$$E(x_\theta^p - x_\theta) = 0 \quad (39)$$

L'erreur de prévision a ainsi une espérance mathématique nulle.

Si de plus les erreurs ε_t sont homoscédastiques et indépendantes entre elles (hypothèses 2 et 3), la variance de $x_\theta^p - x_\theta$ s'exprime par :

$$E(x_\theta^p - x_\theta)^2 = z_\theta^2 \text{Var}(a^*) + 2z_\theta \text{Cov}(a^*, b^*) + \text{Var}(b^*) + \sigma^2$$

et les formules (11) et (12) pour les moments du second ordre de a^* et b^* conduisent alors à :

$$E(x_\theta^p - x_\theta)^2 = (z_\theta - \bar{z})^2 \text{Var}(a^*) + \left(1 + \frac{1}{T}\right) \sigma^2 \quad (40)$$

ou

$$E(x_\theta^p - x_\theta)^2 = \sigma^2 \left[1 + \frac{1}{T} + \frac{(z_\theta - \bar{z})^2}{\sum_{t=1}^T (z_t - \bar{z})^2} \right] \quad (41)$$

Si les erreurs ε_t obéissent à une loi normale (hypothèse 4), alors a^* et b^* suivent une distribution normale (proposition 6) et la formule (38) montre que $x_\theta^p - x_\theta$ est aussi distribué normalement. La même formule permet de constater que $x_\theta^p - x_\theta$ est indépendant de σ_ε^{*2} , puisque a^* , b^* et ε_θ le sont.

Nous pouvons donc énoncer :

Proposition 7

Si les hypothèses 1, 2 et 3 sont satisfaites, une prévision calculée d'après la formule (37) comporte une erreur dont l'espérance mathématique est nulle et la variance est donnée par (41). Quand de plus, les ε_t suivent une loi normale, l'erreur de prévision est indépendante de σ_ε^{*2} et distribuée normalement.

La variance de l'erreur de prévision contient un premier terme irréductible σ^2 , terme qui traduit l'existence de l'erreur non observable ε_θ relative à la période ou à l'unité sur laquelle porte la prévision. Les deux termes suivants sont d'autant plus faibles que le nombre T des observations utilisées pour estimer a et b est plus élevé. Le dernier terme dépend de la valeur de z_θ . Il est nul quand z_θ égale la moyenne des valeurs de z_t dans l'échantillon, et d'autant plus grand que z_θ diffère davantage de cette moyenne.

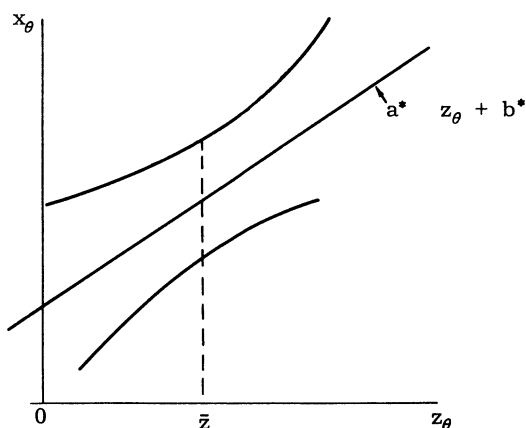


Fig. 3.

En général, on peut figurer sur un graphique, de part et d'autre de la régression, deux courbes qui se situent respectivement au-dessus et au-dessous de cette régression, à une distance égale à k fois l'écart-type de l'erreur de prévision. On fait ainsi bien apparaître l'importance que revêt cette erreur pour les diverses valeurs de z_θ (cf fig. 3).

On peut/même utiliser ces deux courbes pour donner la prévision sous la forme d'une "fourchette", c'est-à-dire d'un intervalle de confiance qui aura α chances sur cent de recouvrir la valeur vraie. Cette "prévision par intervalle" tient compte des aléas affectant l'estimation de a et de b comme de ceux affectant la détermination de la variable endogène pendant la période de prévision.

En général σ^2 n'est pas connu ; dans (41) on lui substitue donc σ_e^{*2} . Dès lors, le facteur k par lequel il convient de multiplier l'écart-type estimé des erreurs de prévision, s'obtient dans la table de Student pour $T - 2$ degrés de liberté, et le niveau de signification α retenu.

Demandons-nous maintenant quelles justifications peuvent être données pour la formule simple (37). Il importe évidemment que la méthode de prévision conduite aux erreurs les plus faibles possibles, c'est-à-dire aux erreurs dont les conséquences seront en moyenne le moins dommageables possibles.

Pour de nombreux problèmes, la perte ou le manque à gagner, résultant d'une prévision inexacte est approximativement proportionnelle au carré de l'erreur commise. Il est alors naturel de rechercher une formule qui rende minimum l'espérance mathématique du carré de l'erreur de prévision.

Ce critère ne conviendrait pas si les pertes résultant d'erreurs de même importance, mais de signes opposés, différaient notablement l'une de l'autre ; par exemple si une prévision trop élevée avait peu de conséquences dommageables, et une prévision trop faible de graves inconvénients. La recherche du minimum de la perte moyenne conduirait alors à des formules différentes de celles envisagées ici.

Le critère de l'erreur quadratique moyenne présente l'avantage de la simplicité. Nous allons voir qu'il peut être appliqué sans que la nature exacte de la loi de distribution des ε_t soit connue. Les autres critères exigent généralement une connaissance beaucoup plus précise de cette loi.

L'erreur de prévision s'écrit :

$$x_\theta^p - x_\theta = x_\theta^p - az_\theta - b - \varepsilon_\theta$$

Comme ε_θ est inconnu et indépendant des x_t pour $t = 1, 2, \dots, T$, il est aussi indépendant de x_θ^p . Ainsi :

$$E (x_\theta^p - x_\theta)^2 = \sigma^2 + E (x_\theta^p - az_\theta - b)^2 \quad (42)$$

La minimisation de l'erreur quadratique moyenne est donc équivalente à celle de

$$W = E (x_\theta^p - az_\theta - b)^2 \quad (43)$$

Si les coefficients a et b étaient connus exactement, la meilleure formule de prévision serait évidemment

$$x_\theta^p = az_\theta + b \quad (44)$$

qui annulerait W . Par analogie, il est naturel de retenir la formule générale

$$x_\theta^p = a(x) z_\theta + b(x) \quad (45)$$

dans laquelle $a(x)$ et $b(x)$ désignent des fonctions de l'échantillon observé. (Nous pourrions également écrire $a(x_1, x_2, \dots, x_T)$ et $b(x_1, x_2, \dots, x_T)$; mais cela alourdirait nos notations). La grandeur W devient alors :

$$W = E \{ [a(x) - a] z_\theta + [b(x) - b] \}^2 \quad (46)$$

Ainsi, la détermination de la meilleure prévision se ramène à celle des estimateurs $a(x)$ et $b(x)$ qui minimisent l'expression (46). La considération des pertes et manque-à-gagner associés à une prévision justifie la recherche d'estimateurs conduisant à des décisions statistiques optimales du point de vue de la fonction de risque " W ". Pour compléter cette étude, nous devrions donc étudier les qualités des estimateurs a^* et b^* dans la théorie des décisions statistiques. Contentons-nous ici de deux remarques.

Si les hypothèses 1 à 4 et 6 sont satisfaites, et si on se limite à priori aux estimateurs centrés, a^* et b^* apparaissent comme optimaux avec la fonction de risque (46). Cela résulte de la proposition 6 établie précédemment, et notamment du renvoi (1) annexé à cette proposition. Si la distribution des ε_t n'est pas normale, mais que l'on se limite à priori aux estimateurs linéaires centrés, a^* et b^* sont encore optimaux,

comme l'indique le renvoi de la proposition 5. Les propositions 5 et 6 reçoivent ainsi une signification intéressante du point de vue des utilisations du modèle.

REFERENCES

- H. CRAMER (1946) - *Mathematical Methods of Statistics* - (Princeton University Press).
- G. DARMOIS (1945) - "Sur les limites de dispersion de certaines lois" - (Revue de l'Institut International de Statistique - 13ème année).
- R. FERON (1956) - "Information, régression, corrélation" - (Publications de l'Institut de Statistique de l'Université de Paris - Vol. 5, fascicule 3-4).
- E. MALINVAUD (1964) - *Méthodes statistiques de l'économétrie* - (Dunod).
- H. THEIL (1958) - *Economic Forecasts and Policy* - (North - Holland Publishing Company - Amsterdam).
- H. WOLD (1952) - *Demand Analysis, A Study in Econometrics* - (John Wiley, New York).
- H. WOLD (1960) - "Ends and Means of Econometric Model Building" - dans U. Grenander ed., *Probability and Statistics* (Stockholm).