

# JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

JEAN-JACQUES BOULANGER

**Technique statistique et prévision économique**

*Journal de la société statistique de Paris*, tome 121, n° 4 (1980), p. 211-215

[http://www.numdam.org/item?id=JSFS\\_1980\\_\\_121\\_4\\_211\\_0](http://www.numdam.org/item?id=JSFS_1980__121_4_211_0)

© Société de statistique de Paris, 1980, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

# TECHNIQUE STATISTIQUE ET PRÉVISION ÉCONOMIQUE

Jean-Jacques BOULANGER

*Dans les pages qui suivent, l'auteur détermine des régressions linéaires multiples avec des variables macroéconomiques décalées dans le temps, qui permettent des prévisions économiques ayant une précision acceptable.*

*In the following pages, the author determines multiple linear regressions with time shifted macroeconomic variables, allowing sufficiently precised economic estimates.*

*In der vorliegenden Arbeit bestimmt der Verfasser lineare mehrfache Regressionen an Hand makroökonomischer Variablen für bestimmte Zeitabstände, die ökonomische Voraussagen gestatten, da sie eine annehmbare Genauigkeit besitzen.*

La technique statistique a toujours été associée depuis son développement à la prévision économique qu'elle autorisait et à la décision économique qu'elle encourageait.

En principe la méthode est simple. Soient  $\alpha_{\text{PIB}_m}$  le taux d'accroissement annuel en volume du Produit intérieur brut marchand et  $\alpha_{\text{CM}}$  le taux d'accroissement annuel en volume de la Consommation des ménages. La régression linéaire simple de  $\alpha_{\text{PIB}_m}$  par rapport à  $\alpha_{\text{CM}}$  pour la période 1969-78 est représentée par la formule :

$$(1) \quad \widehat{\alpha_{\text{PIB}_m}} = 1,239 \alpha_{\text{CM}} - 1,25$$

Donc, si le taux d'accroissement en volume de la Consommation des ménages augmente de 1 point, le taux d'accroissement en volume du Produit intérieur brut marchand augmentera de 1,2 point.

En réalité il n'en est rien. En effet il existe des relations très souvent importantes entre les variations des principales variables macroéconomiques déphasées dans le temps, et, comme au début de l'année  $t$  on connaît généralement avec une précision suffisante la plupart des variables macroéconomiques relatives aux années précédentes :  $t-1$ ,  $t-2$ , ..., la variance résiduelle de  $\alpha_{\text{PIB}_m}$  en est de beaucoup diminuée.

Soient  $t$  l'année considérée,  $(\beta_{\text{Exp}})_{-1}$  le taux d'accroissement des prix à l'exportation pour l'année précédente,  $(\alpha_{\text{PIB USA}})_{-1}$  le taux d'accroissement en volume du Produit intérieur brut américain pour l'année précédente et  $(x_2)_{-3}$  le taux d'accroissement de la mesure de la masse monétaire entre la fin des années  $t-4$  et  $t-3$ . La régression linéaire multiple de  $\alpha_{\text{PIB}_m}$  par rapport à ces variables pour la période 1969-78 s'écrit :

$$(2) \quad \widehat{\alpha_{\text{PIB}_m}} = 9,15 - 0,105 t - 0,242 (\beta_{\text{Exp}})_{-1} - 0,098 (\alpha_{\text{PIB USA}})_{-1} - 0,145 (x_2)_{-3}$$

Il lui correspond un  $R^2 = 98,4$  %, c'est-à-dire que plus de 98 % de la variance de  $\alpha_{\text{PIB}_m}$  sont expliqués par les variations de  $t$ ,  $(\beta_{\text{Exp}})_{-1}$ ,  $(\alpha_{\text{PIB USA}})_{-1}$ ,  $(x_2)_{-3}$ . Or au début de l'année  $t$  ces variables sont pratiquement connues. Il en découle des variations possibles de  $\alpha_{\text{PIB}_m}$  autour de  $\widehat{\alpha_{\text{PIB}_m}}$  beaucoup plus faibles que ce qu'on aurait supposé sans la connaissance

de ces 4 variables. La valeur de  $R^2 = 98,4 \%$  est en effet très importante. Elle ne saurait en outre être comparée aux valeurs de  $R^2$  que l'on obtient quand on calcule les coefficients de corrélation entre niveaux (PIB<sub>m</sub>, Consommation des ménages, Investissements productifs...). Un exemple permet de le préciser :

Pour la période 1964-74 on obtient un coefficient de corrélation  $r = 0,9975$  entre le logarithme de la Masse monétaire et celui du Produit intérieur brut total, soit un  $R^2 (= r^2)$  de 0,995. Or on sait la difficulté qu'il y a à contrôler l'évolution du PIB à l'aide de la masse monétaire. Au contraire si l'on considère les taux d'accroissement de la Masse monétaire et du PIB total en valeur, on obtient un coefficient de corrélation de 0,312, soit un  $R^2$  de 0,097; donc en gros 10 % de la variance du taux d'accroissement du PIB total s'expliquent par les variations du taux d'accroissement de la Masse monétaire (et inversement), ce qui correspond davantage à la réalité.

Reprenons la régression (2).  $\widehat{\alpha_{\text{PIB}_m}}$  est connu puisque le deuxième membre est connu et la variance résiduelle est égale à :

$$\frac{\Sigma (\alpha_{\text{PIB}_m} - \widehat{\alpha_{\text{PIB}_m}})^2 \times (1 - R^2)}{n-p-1} = 0,132,$$

d'où une dispersion résiduelle de 0,36, d'où, si l'on suppose que la variable résiduelle  $\widehat{\alpha_{\text{PIB}_m}} = \alpha_{\text{PIB}_m} - \widehat{\alpha_{\text{PIB}_m}}$  suit une loi de type gaussien, un écart probable de  $0,36 \cdot 0,675 \approx 0,25$ . Il y a donc une chance sur deux pour que l'on ait :

$$\widehat{\alpha_{\text{PIB}_m}} - 0,25 < \alpha_{\text{PIB}_m} < \widehat{\alpha_{\text{PIB}_m}} + 0,25$$

Donc si le calcul donne  $\widehat{\alpha_{\text{PIB}_m}} = 2,9 \%$ , on peut prévoir qu'il y a une chance sur deux pour que l'on ait :  $2,65 < \alpha_{\text{PIB}_m} < 3,15$ .

La régression (2) ayant été déterminée, on peut rechercher l'incidence sur l'évolution du PIB marchand pour l'année  $t$ , de la Consommation des ménages, des Investissements productifs et des Exportations que l'on envisage pour l'année  $t$ . Si l'on ajoute aux 4 variables de la partie droite de la régression (2) le taux d'accroissement annuel de la Consommation des ménages en volume, on obtient comme régression multiple la relation :

(3)  $\widehat{\alpha_{\text{PIB}_m}} = 7,91 - 0,101 t - 0,225 (\beta_{\text{EXP}})_{-1} - 0,064 (\alpha_{\text{PIB USA}})_{-1} - 0,122 (x_2)_{-3} + 0,143 \alpha_{\text{CM}}$ , et un  $R^2$  qui a peu évolué puisque égal à 98,6 %. On voit qu'une augmentation de  $\alpha_{\text{CM}}$  de 1 point n'entraîne plus qu'une augmentation de  $\widehat{\alpha_{\text{PIB}_m}}$  de 0,14 point, ce qui est peu.

Une variation de  $\alpha_{\text{CM}}$  de 1 point est d'ailleurs optimiste. En effet  $\alpha_{\text{CM}}$  est relié aux 4 variables du second membre de la régression (2) par une régression :

$$(4) \quad \widehat{\alpha_{\text{CM}}} = at + b(\beta_{\text{EXP}})_{-1} + c(\alpha_{\text{PIB USA}})_{-1} + d(x_2)_{-3} + e$$

avec un  $R^2 = 87,7 \%$ , donc 87,7 % de la variance de  $\alpha_{\text{CM}}$  résultent des variations des 4 variables  $t$ ,  $(\beta_{\text{EXP}})_{-1}$ ,  $(\alpha_{\text{PIB USA}})_{-1}$ ,  $(x_2)_{-3}$ . Or ces variables sont fixées et connues au début de l'année  $t$ . Dans ces conditions la variance résiduelle de  $\alpha_{\text{CM}}$  est donc égale à :

$$\Sigma (\alpha_{\text{CM}} - \widehat{\alpha_{\text{CM}}})^2 \times \frac{(1 - R^2)}{n-p-1} = 0,475$$

d'où une dispersion de 0,689, d'où, en supposant que la variable résiduelle  $\widehat{\alpha_{\text{CM}}} = \alpha_{\text{CM}} - \widehat{\alpha_{\text{CM}}}$  suit une loi de type gaussien, un écart probable de  $0,689 \cdot 0,675 = 0,465$ , d'où une incidence sur  $\widehat{\alpha_{\text{PIB}_m}}$  de  $0,465 \cdot 0,143 = 0,0665 \approx 0,07$ . Donc une fois sur deux (une chance sur deux)

l'incidence des variations de  $\alpha_{CM}$  sur  $\widehat{\alpha_{PIB_m}}$  n'atteindra pas un écart de 0,1. On est donc obligé de conclure que les variations de  $\alpha_{CM}$  n'influencent que très légèrement  $\widehat{\alpha_{PIB_m}}$  et que l'importante corrélation entre  $\alpha_{PIB_m}$  et  $\alpha_{CM}$  ( $r = 0,84$ ) n'est pas une relation de cause à effet, mais une covariation entre 2 variables étroitement dépendantes de  $t$ ,  $(\beta_{Exp})_{-1}$ ,  $(\alpha_{PIB USA})_{-1}$  et  $(x_2)_{-3}$ .

De même si l'on remplace  $\alpha_{CM}$  par  $\alpha_{IP}$  (taux d'accroissement des Investissements productifs en volume) on obtient la régression suivante :

$$(5) \widehat{\alpha_{PIB_m}} = 6,83 - 0,090 t - 0,141 (\beta_{Exp})_{-1} + 0,031 (\alpha_{PIB USA})_{-1} - 0,108 (x_2)_{-3} + 0,142 \alpha_{IP}$$

Mais, comme  $\alpha_{CM}$ ,  $\alpha_{IP}$  n'est pas susceptible de prendre n'importe quelle valeur. En effet  $\alpha_{IP}$  est relié à des variables connues par une régression du type :

$$(6) \widehat{\alpha_{IP}} = at + b (\beta_{Exp})_{-1} + c (\alpha_{CM})_{-1} + d (x_2)_{-3} + e$$

qui conduit à un  $R^2 = 0,926$ , d'où une variance résiduelle de 4,025 d'où une dispersion résiduelle de 2,01 et, si l'on suppose que la variable résiduelle  $\alpha_{IP} = \alpha_{IP} - \widehat{\alpha_{IP}}$  suit une loi de type gaussien, un écart probable de  $2,01 \cdot 0,675 = 1,35$ , d'où une incidence sur  $\widehat{\alpha_{PIB_m}}$  de  $1,35 \cdot 0,142 = 0,19$ . Donc une fois sur deux l'incidence des variations de  $\alpha_{IP}$  sur  $\widehat{\alpha_{PIB_m}}$  n'atteindra pas un écart de 0,2.

Enfin, si l'on remplace  $\alpha_{IP}$  par  $\alpha_{Exp}$  (taux d'accroissement en volume des exportations), on obtient la régression :

$$(7) \widehat{\alpha_{PIB_m}} = 8,77 - 0,094 t - 0,230 (\beta_{Exp})_{-1} - 0,101 (\alpha_{PIB USA})_{-1} - 0,143 (x_2)_{-3} + 0,020 \alpha_{Exp}$$

Mais  $\alpha_{Exp}$  est lié aux variables de base par une régression du type :

$$(8) \widehat{\alpha_{Exp}} = at + b (\beta_{Exp})_{-1} + c (\alpha_{PIB USA})_{-1} + d (x_2)_{-3} + e$$

qui conduit à un  $R^2 = 0,942$ . La variance résiduelle est de 3,61, soit une dispersion résiduelle de 1,90 et un écart probable de  $0,675 \cdot 1,90 = 1,28$ . L'incidence de cet écart sur  $\alpha_{PIB_m}$  est de  $0,02 \cdot 1,28 = 0,025$ , donc insignifiant (1).

Nous sommes donc amenés à conclure que les 3 locomotives classiques : Consommation des ménages, Formation de capital productif, Exportations, n'entraînent que fort peu la croissance du PIB marchand, et qu'il existe simplement une forte covariation entre les taux d'accroissement de ces 3 variables et celui du PIB marchand. Ce qui ne veut pas dire qu'il faille négliger, par exemple, la relance de telle variable macroéconomique, lorsque celle-ci devient nécessaire, ce qui pourrait entraîner par la suite une autre perturbation, mais que la nécessité de la relance en question peut être décelée un certain temps à l'avance.

1. Dans le cas où l'on s'intéresse plus particulièrement aux valeurs d'exception, on peut dire que, si la variable résiduelle  $\alpha$  suit une loi de Gauss, il y a seulement 5 chances sur 100 que l'on ait :  $\alpha > 1,645 \sigma_{Res}$ ,  $\sigma_{Res}$  étant la dispersion résiduelle de  $\alpha$ . Il y a donc seulement 5 chances sur 100 que l'on ait :

$$\begin{aligned} \alpha_{CM} &> 1,645 \cdot 0,689 = 1,13 \\ \alpha_{IP} &> 1,645 \cdot 2,01 = 3,31 \\ \alpha_{Exp} &> 1,645 \cdot 1,90 = 3,13 \end{aligned}$$

soient des incidences sur  $\widehat{\alpha_{PIB_m}}$  de  $1,13 \cdot 0,143 = 0,16$  pour  $\alpha_{CM}$ , de  $3,31 \cdot 0,142 = 0,47$  pour  $\alpha_{IP}$ , de  $3,13 \cdot 0,02 = 0,06$  pour  $\alpha_{Exp}$ . (Bien entendu ceci suppose que les décisions économiques prises au cours de l'année  $t$  ou durant la dernière partie de l'année  $t-1$  sont comparables à certaines des décisions prises au cours de la période de base, pour laquelle la régression a été calculée; sinon on ne peut rien dire.)

Seule l'incidence de  $\alpha_{IP}$  peut être importante, mais elle suppose un optimisme certain de la part des entrepreneurs et un fort encouragement financier de la part de l'État. Sans oublier que 5 chances sur 100 correspondent en moyenne à une année sur vingt et que, une année sur deux, l'incidence de  $\alpha_{IP}$  sur  $\widehat{\alpha_{PIB_m}}$  ne dépasse pas 0,2 point.

En ce qui concerne les prix, on obtient pour le taux d'accroissement des prix à la consommation des ménages  $\beta_{CM}$  une régression du même genre que les régressions précédentes, mais on a estimé préférable de considérer non le taux d'accroissement annuel  $\beta_{CM}$ , mais son écart au taux d'accroissement relatif à l'année précédente, soit  $\Delta\beta_{CM} = \beta_{CM} - (\beta_{CM})_{-1}$ , ce qui permet d'éliminer en grande partie la corrélation existant entre les valeurs de  $\beta_{CM}$  relatives à deux années consécutives.

Remarquons qu'il n'en a pas toujours été ainsi. En effet le taux d'accroissement des prix de la PIB s'est élevé à 16,6 % en 1951 et 12,4 % en 1952, mais est tombé à -0,2 % en 1953, 0,2 % en 1954 et 0,9 % en 1955. De même (mais le ralentissement fut plus lent) il s'est élevé à 11,6 % en 1958, puis est tombé à 6 % en 1959, 3 % en 1960 et 2,8 % en 1961. Au contraire à l'époque actuelle les prix des matières premières, énergétiques ou non, et les salaires sont répercutés rapidement sur le prix de la PIB et celui de la consommation des ménages. De même le prix à la consommation des ménages est rapidement répercuté sur les salaires. Et l'on est en présence d'un phénomène général d'entraînement de plus en plus effectif (1).

Deux régressions ont été retenues pour  $\Delta\beta_{CM}$ . Ce sont :

$$(9) \widehat{\Delta\beta_{CM}} = -0,941 \Delta(x_2)_{-5} - 0,110 \Delta\beta_{-1}^1 + 3,690 \Delta R_{-1}^0 - 0,59 \quad \text{et}$$

$$(10) \widehat{\Delta\beta_{CM}} = -0,602 \Delta(x_2)_{-5} + 1,299 \Delta R^0 - 0,075 \Delta\beta_{-1}^1 + 2,881 \Delta R_{-1}^0 + 0,138 \Delta\gamma_{Sn} - 0,76.$$

$\Delta$  représente l'écart entre les valeurs des différentes variables pour deux années consécutives,  $\beta_{CM}$  le taux d'accroissement annuel du prix à la consommation des ménages entre les années  $t-1$  et  $t$ ,  $(x_2)_{-5}$  le taux d'accroissement annuel de la masse monétaire entre la fin de l'année  $t-6$  et la fin de l'année  $t-5$ ,  $\beta_{-1}^1$  le taux d'accroissement des prix à l'importation des biens intermédiaires non pétroliers entre l'année  $t-2$  et l'année  $t-1$ ,  $R^0$  le rapport du montant des importations de produits pétroliers pour l'année  $t$  au PIB marchand de la même année,  $R_{-1}^0$  le même rapport pour l'année  $t-1$ , et  $\gamma_{Sn}$  le taux d'accroissement des salaires nominaux totaux entre les années  $t-1$  et  $t$ .

$\Delta(x_2)_{-5}$  a été choisi parce que, à lui seul, il explique 50 % de la variance de  $\Delta\beta_{CM}$  (coefficient de corrélation  $r = -0,705$ ); l'incidence de  $\Delta\gamma_{Sn}$  est encore plus importante ( $r = 0,79$ ) mais il s'agit là d'une covariation (pas de décalage dans le temps) et non d'une relation de cause à effet (alors que la corrélation entre  $\Delta(x_2)_{-5}$  et  $\Delta\gamma_{Sn}$  est de  $-0,68$ ).

Dans la régression (9),  $\Delta\beta_{CM}$  est déterminé puisque l'on connaît au début de l'année  $t$  les valeurs des 3 variables du second membre. Par contre le  $R^2$  est assez faible : 0,732, la variance résiduelle de  $\Delta\beta_{CM}$  est égale à 2,560, la dispersion résiduelle à 1,600 et l'écart probable à 1,1 dans l'hypothèse d'une variable résiduelle  $\widehat{\Delta\beta_{CM}}$  ( $= \Delta\beta_{CM} - \widehat{\Delta\beta_{CM}}$ ) gaussienne. On a une chance sur deux d'avoir :

$$\widehat{\Delta\beta_{CM}} - 1,1 < \Delta\beta_{CM} < \widehat{\Delta\beta_{CM}} + 1,1$$

et le champ d'évolution de  $\Delta\beta_{CM}$  est très étendu même pour un seuil de probabilité de 50 %.

Dans la régression (10),  $\Delta\beta_{CM}$  n'est pas déterminé puisque l'on ne connaît ni  $\Delta R^0$  ni  $\Delta\gamma_{Sn}$ . Pour chaque valeur hypothétique de  $\Delta R^0$  et  $\Delta\gamma_{Sn}$  on a une valeur de  $\Delta\beta_{CM}$  différente, par contre on a un  $R^2$  beaucoup plus élevé : 0,982. Il y a une chance sur deux pour que l'on ait :

$$\widehat{\Delta\beta_{CM}} - 0,35 < \Delta\beta_{CM} < \widehat{\Delta\beta_{CM}} + 0,35$$

et l'on obtient dans ce cas une marge beaucoup plus restreinte pour l'évolution de  $\Delta\beta_{CM}$ .

1. En fait il ne s'agit là que d'une apparence comme on le verra plus loin.

Soulignons la faible influence de  $\widehat{\Delta\gamma_{sn}}$  sur  $\widehat{\Delta\beta_{cm}}$ . Un écart entre 2 valeurs de  $\Delta\gamma_{sn}$  de 2 points entraîne une variation de  $\widehat{\Delta\beta_{cm}}$  inférieure à 0,3 point (régression 10). Là encore  $\Delta\beta_{cm}$  et  $\Delta\gamma_{sn}$  sont tous deux dépendants dans une large mesure de  $\Delta(x_2)_{-5}$ ,  $\Delta\beta_{-1}^1$  et  $\Delta R_{-1}^0$ . On peut en effet comparer à la régression (9) la régression :

$$(11) \quad \widehat{\Delta\gamma_{sn}} = a \Delta(x_2)_{-5} + b \Delta\beta_{-1}^1 + c \Delta R_{-1}^0 + d$$

pour laquelle on a un  $R^2 = 0,655$  contre un  $R^2 = 0,732$  pour la régression (9).  $\Delta\beta_{cm}$  et  $\Delta\gamma_{sn}$ , dépendant tous les deux de  $\Delta(x_2)_{-5}$ ,  $\Delta\beta_{-1}^1$  et  $\Delta R_{-1}^0$ , sont fortement corrélés (coefficient de corrélation de 0,79), mais une fois que  $\Delta(x_2)_{-5}$ ,  $\Delta\beta_{-1}^1$  et  $\Delta R_{-1}^0$  sont fixés,  $\Delta\gamma_{sn}$  n'aura plus qu'une influence beaucoup plus faible sur  $\widehat{\Delta\beta_{cm}}$ .

Pour conclure nous voudrions souligner l'intérêt des régressions linéaires multiples avec des variables décalées dans le temps comme instrument de prévision économique. La prévision sera d'autant plus précise que les variables explicatives seront connues au moment de la prévision et, pour l'intervalle de variation, que la variance résiduelle sera faible, donc que le coefficient de corrélation multiple sera élevé, sans que le nombre des variables explicatives soit trop important.

Les régressions peuvent être déterminées pour d'autres variables que  $\alpha_{PIBm}$  et  $\Delta\beta_{cm}$ . Une étude systématique permettrait d'en déduire un modèle expliquant l'enchaînement des variables économiques. Enfin, les coefficients des formules de régression multiple dépendent évidemment de la période pour laquelle la régression a été calculée. Si la période change, les coefficients de la formule de régression risquent de changer aussi. On ne peut que supposer que la formule de régression multiple ne se déformera pas trop rapidement dans le temps.