Journal de la société statistique de Paris

Pierre Thionet

Un indice statistique destiné à la comparaison des prévisions et des réalisations

Journal de la société statistique de Paris, tome 105 (1964), p. 181-189
<http://www.numdam.org/item?id=JSFS_1964__105__181_0>

VI

UN INDICE STATISTIQUE DESTINÉ A LA COMPARAISON DES PRÉVISIONS ET DES RÉALISATIONS

OBJET DE L'ARTICLE

La présente note est destinée à faire connaître les propriétés d'un indice statistique employé par certains économistes pour comparer un ensemble de prévisions et de réalisations. Cet indice est indiqué notamment par H. Theil dans *Economic Forecast ans Policy* (Amsterdam, 2e éd. 1961), et est employé par la Communauté Économique Européenne.

DÉFINITION DE L'INDICE

Soit une variable économique X dont les valeurs à des dates également espacées

$$t = 0, 1, 2, 3 \ldots . T$$

font l'objet de prévisions

$$P_0\ P_1\ P_2\ P_3 \ldots\ldots P_T$$

et qui prennent ensuite les valeurs (appelées *réalisations*) :

$$R_0\ R_1\ R_2\ R_3 \ldots R_T$$

Posons
$$\begin{aligned} b^2 &= P_0^2 + P_1^2 + P_2^2 + \ldots + P_T^2 \\ c^2 &= R_0^2 + R_1^2 + R_2^2 + \ldots + R_T^2 \\ a^2 &= (R_0 - P_0)^2 + (R_1 - P_1)^2 + \ldots + (R_T - P_T)^2 \end{aligned}$$

autrement dit

$$b^2 = \Sigma P^2,\ c^2 = \Sigma R^2;\ b \geqslant o,\ c \geqslant o$$
$$a^2 = \Sigma(R - P)^2;\ a \geqslant o$$

L'indice employé pour apprécier la concordance des prévisions et des réalisations est

$$Q = \frac{a}{b + c}$$

autrement dit

$$Q = \frac{\sqrt{\Sigma (R - P)^2}}{\sqrt{\Sigma P^2} + \sqrt{\Sigma R^2}}$$

Exemples : Empruntons au S. E. E. F. les données suivantes (pour la France) (1):
Prévision P : un budget économique.
Réalisation R : les Comptes de la Nation.

a) *Exemples* (en milliards de francs) choisis pour avoir Q très grand.

	Année	P	R	R — P	Observations
1) Déficit des Administrations	1956	303,0	171,0	— 132,0	
	1957	206,0	94,3	— 111,7	budget II
Q = 17,4 %	1958	— 424,0	— 391,0	+ 33,0	budget A
	1959	— 72,0	— 108,0	— 36,0	budget N 3
	1960	114,2	100,5	— 13,7	
2) Emprunt net des Administrations	1956	337	221	— 116	
	1957	240	38	— 202	
	1958	— 554	— 563	— 9	idem
Q = 18,3 %	1959	— 51	— 57	— 6	
	1960	2	51	49	
3) Prêt net de l'extérieur	1956	215	630	415	
	1957	102	101	— 1	
	1958	— 102	— 135	— 33	idem
Q = 56,9 %	1959	— 17	— 527	— 510	
	1960	223	78	— 145	

(1) Communiquées spécialement par la division des Comptes et Budgets du S. E. E. F.

b) *Exemples* où P, T sont des proportions pour cent (ratios d'une année sur l'autre).

	Annee	P	R	R — P	Observations
1) Produit intérieur brut (en valeur)	1956/55	106,2	108,6	2,4	
	1957/56	105,4	111,8	6,4	
	1958/57	115,7	113,9	— 1,8	idem
Q = 1,56 %	1959/58	105,3	107,8	2,5	
	1960/59	107,6	109,0	1,4	
2) Produit interieur brut (en volume)	1956/55	105,0	104,4	— 0,6	
	1957/56	103,5	106,2	2,7	
	1958/57	103,5	102,2	— 1,3	idem
Q = 0,69 %	1959/58	101,5	102,2	0,7	
	1960/59	105,0	105,7	0,7	
3) Consommation des menages (en valeur)	1956/55	106,7	110,2	3,5	
	1957/56	104,7	110,3	5,6	
	1958/57	114,0	112,6	— 2,2	idem
Q = 1,52 %	1959/58	104,7	107,1	2,4	
	1960/59	107,3	108,0	0,7	
4) Consommation des ménages (en volume)	1956/55	105,8	105,7	— 0,1	
	1957/56	103,6	105,2	1,6	
	1958/57	102,0	100,0	— 2,0	idem
Q = 0,62 %	1959/58	100,1	101,3	1,2	
	1960/59	104,3	104,0	— 0,3	

PROPRIÉTÉS DE L'INDICE Q

1. *L'indice Q est nul* si (et seulement si) chaque réalisation coïncide avec la prévision :

En effet $P_0 = R_0;\ P_1 = R_1;\ \ldots\ldots\ P_T = R_T$

entraîne $(R_t - P_t)^2 = 0$ quelle que soit la date t; d'où

$$a^2 = \Sigma\,(R - P)^2 = 0$$

2. Dans tous les autres cas, *Q est positif.*

3. *L'indice Q prend la valeur 1* si (cas général) toutes ses prévisions sont en sens contraire des réalisations, la proportion d'erreur étant rigoureusement la même chaque fois — ainsi que dans les cas singuliers suivants :

si toutes les prévisions sont nulles (auquel cas les réalisations peuvent être quelconques) — ou encore si toutes les réalisations sont nulles (et les prévisions quelconques).

Vérification : $P = 0;\ Q = \dfrac{\sqrt{\Sigma R^2}}{\sqrt{\Sigma R^2}} = 1$ c'est l'un des cas singuliers.

$$R = -\,nP$$

$$a^2 = \Sigma(R - P)^2 = (n + 1)^2\,\Sigma P^2 = (n + 1)^2\, b^2$$

$$b^2 = \Sigma P^2;\ c^2 = n^2 \Sigma P^2 = n^2\, b^2$$

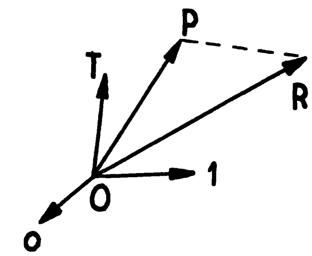
$$Q = \frac{a}{b + c} = \frac{(n + 1)\, b}{b + nb} = 1$$

c'est le cas général

4. *Réciproquement*

$$\text{on a} \begin{cases} Q = 1 \text{ dans les seuls cas précédents} \\ Q < 1 \text{ dans tous les autres cas} \end{cases}$$

En effet, dans l'espace à $(T + 1)$ dimensions (rapporté à des axes orthonormés) considérons les points : O de coordonnées 0
P de coordonnées P_t ou P_t
R de coordonnées R_t ou R_t



$\sqrt{\Sigma P^2} = b$ est la distance OP

$\sqrt{\Sigma R^2} = c$ est la distance OR

$\sqrt{\Sigma (R - P)^2} = a$ est la distance PR

$Q = 1$ équivaut à $\underline{PR = OP + OR}$

relation qui est vérifiée si et seulement si POR sont alignés (dans l'ordre POR).

L'inégalité $|OP - OR| \leqslant PR \leqslant OP + OR$ est *l'inégalité triangulaire* bien connue (tout côté d'un triangle est compris entre la somme et la différence des deux autres); elle s'étend de 2 ou 3 dimensions, à un nombre quelconque de dimensions.

5. Quand la série des comparaisons : Prévisions-Réalisations s'allonge d'une unité :

5. 1. Si le dernier couple prévision-réalisation indique une coïncidence parfaite, *l'indice* Q *diminue* (s'améliore) :

Car $$Q_o = \frac{a}{b + c}$$ est manifestement plus grand que

$$Q_x = \frac{\sqrt{a^2 + (x - x)^2}}{\sqrt{b^2 + x^2} + \sqrt{c^2 + x^2}} = \frac{a}{b\left(1 + \frac{x^2}{b^2}\right)^{1/2} + c\left(1 + \frac{x^2}{c^2}\right)^{1/2}}$$

5. 2. Si le dernier couple indique une prévision aussi mauvaise que la moyenne des précédentes, *l'indice* Q *diminue* (s'améliore)

Car $Q = \dfrac{a}{b + c}$ est supérieur à

$$q = \frac{\sqrt{a^2 + (x - y)^2}}{\sqrt{b^2 + x^2} + \sqrt{c^2 + y^2}} \text{ avec } x = bz,\ y = cz$$

d'ou $$q = Q\sqrt{\frac{1 + d^2 z^2}{1. + z^2}} \qquad \text{avec } d^2 = \left(\frac{b - c}{a}\right)^2$$

Or $$|b - c| < a \qquad \text{d'où } \left(\frac{b - c}{a}\right)^2 < 1,$$

est l'inégalité triangulaire rappelée plus haut; d'où :

$$\frac{1 + d^2 z^2}{1 + z^2} < 1 \Rightarrow \boxed{q < Q}$$

5. 3. Par conséquent Q peut rester constant ou même s'améliorer, alors que la dernière en date des comparaisons est un peu moins bonne que les précédentes.

6° Le rapport $Q = \dfrac{a}{b + c} = \dfrac{RP}{OP + OR}$ serait constant si les comparaisons successives étaient invariantes :

$$P_0 = P_1 = \ldots\ldots = P_T = P$$
$$R_0 = R_1 = \ldots\ldots = R_T = R$$

$$\Rightarrow \qquad Q = \frac{\sqrt{n}\,|P - R|}{\sqrt{n}\,|P| + \sqrt{n}\,|R|} = \frac{|P - R|}{|P + R|}$$

P et R de même sens

$$R = f\,P,\; f < 1 \Rightarrow Q = \frac{1 - f}{1 + f}$$

COMPARAISON ENTRE L'INDICE Q ET LA CORRÉLATION

On sait que, si les réalisations R_t étaient fonctions du premier degré des prévisions Q_t (en particulier proportionnelles à celles-ci) la corrélation entre prévisions et réalisations serait égale à 1 ou à — 1.

Ainsi il n'y a pas coïncidence entre l'indice Q et le coefficient de corrélation usuel ρ.

Pour obtenir la relation entre les deux indices, on peut procéder en deux étapes,en introduisant l'indice intermédiaire :

$$\rho^* = \frac{\Sigma P_t R_t}{\sqrt{\Sigma P_t^2}\,\sqrt{\Sigma R_t^2}} = \frac{\Sigma P_t R_t}{bc}$$

ρ^* coïnciderait avec ρ si les P_t et R_t avaient leurs moyennes nulles.

Les relations sont alors les suivantes :

1°
$$\begin{aligned} a^2 = \Sigma(P_t - R_t)^2 &= b^2 + c^2 - 2\,\Sigma P_t\,R_t \\ &= b^2 + c^2 - 2\,bc\,\rho^* \end{aligned}$$

autrement dit ρ^* est le *cosinus* de l'angle $\widehat{POR}$ du vecteur des prévisions P et du vecteur des réalisations R. D'où :

$$\boxed{1 - Q^2 = \frac{2bc}{(b + c)^2}\,(1 + \rho^*)}$$

Le facteur $m = 2\,bc/(b + c)^2$ passe par son maximum 1/2 quand $b = c$ et reste voisin de ce maximum quand c n'est pas foncièrement différent de b :

Exemple	$c = 7\,b/10$	$m = 140/289 = 0{,}485$
	$c = 6\,b/10$	$m = 120/256 = 0{,}47$
	$c = 5\,b/10$	$m = 100/225 = 0{,}44$

Nous supposerons donc $\underline{m = 0{,}5}$

2° D'autre part, si
- $p = \Sigma\,P_t/n$, $r = \Sigma\,R_t/n$; $\quad n = T + 1$;
- σ, τ sont les écarts types des P_t, R_t,
- γ, δ les coefficients de variation correspondants,

on a
$$\rho^* = \frac{\rho\sigma\tau + pr}{\sqrt{\sigma^2 + p^2}\,\sqrt{\tau^2 + r^2}}$$

$$\boxed{\rho^* = \frac{\rho + \gamma^{-1}\,\delta^{-1}}{(1 + \gamma^{-2})^{\frac{1}{2}}\,(1 + \delta^{-2})^{\frac{1}{2}}}}$$

Si γ et δ sont
- de l'ordre de 0,5 → $\rho^* = \dfrac{\rho + 4}{5}$
- de l'ordre de 1 → $\rho^* = \dfrac{\rho + 1}{2}$

3. *Correspondance entre Q et* ρ : $m = 0{,}5$ et $\gamma = \delta = 0{,}5$ ou $1{,}0$.

	Q =	0	0,10	0,15	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,80
	ρ^* =	1	0,98	0,96	0,92	0,88	0,82	0,68	0,50	0,28	0,02	— 0,28
$\gamma = \delta = 0{,}5$	ρ =	1	0,90	0,80	0,60	0,40	0,10	— 0,60	///	///	///	///////
$\gamma = \delta = 1{,}0$	ρ =	1	0,96	0,92	0,84	0,76	0,64	0,36	0	— 0,44	— 0,96	///////

$\gamma = \delta = 0{,}5$ | $\rho = 0$ correspond à $\rho^* = 0{,}80$ et $Q = 0{,}33$
$\gamma = \delta = 1$ | $\rho = 0$ correspond à $\rho^* = 0{,}50 = Q$

Certains utilisateurs de Q ont formulé le jugement que les prévisions pouvaient être considérées comme satisfaisantes lorsque le coefficient était inférieur à 0,3. On voit que pareille valeur de Q est susceptible de correspondre à des cas fort variés;

par exemple à : $\rho = \rho^* = 1$, mais $m = 0{,}455$, c'est-à-dire b/c de l'ordre de 3/2.

RECHERCHE D'UNE ZONE DU PLAN (P, R) OU LA PRÉSENCE DES POINTS $P_i R_i$ EST ACCEPTABLE

Supposons Q donné

$$\frac{a}{b+c} = Q$$

et ajoutons une nouvelle comparaison x, y. On désire que le point (xy) se trouve dans une région du plan telle que le nouvel indice n'ait pas augmenté; d'où la condition :

$$\frac{\sqrt{a^2 + (x-y)^2}}{\sqrt{b^2 + x^2} + \sqrt{c^2 + y^2}} \leqslant \frac{a}{b+c} = Q$$

Cette inéquation définit un *domaine* dont le contour vérifie l'équation

$$\frac{a^2 + (x-y)^2}{b^2 + x^2 + c^2 + y^2 + 2\sqrt{b^2 + x^2}\sqrt{c^2 + y^2}} = Q^2$$

autrement dit

$$\left[\frac{a^2}{Q^2} + \left(\frac{x-y}{Q}\right)^2 - x^2 - y^2 - b^2 - c^2\right]^2 = 4\,(b^2 + x^2)\,(c^2 + y^2)$$

En développant on constate que la *constante* disparaît :

$$\frac{a^2}{Q^2} - b^2 - c^2 \equiv 2\,bc$$

Posons $\Phi = x^2 + y^2 - \left(\frac{x-y}{Q}\right)^2$; $(\Phi - 2\,bc)^2 = \Phi^2 - 4\,bc\,\Phi + 4\,b^2 c^2$

l'équation s'écrit : $\Phi^2 - 4\,x^2 y^2 = 4\,bc\,\Phi + 4\,(c^2 x^2 + b^2 y^2)$
où le 1^er^ membre est homogène du 4^e^ degré et le second membre quadratique.

1) L'origine est un point double, l'équation des tangentes étant

$$bc\,\Phi + c^2 x^2 + b^2 y^2 = 0$$

c'est-à-dire :

$$\left(bc + c^2 - \frac{bc}{Q^2}\right) x^2 + \frac{2\,bc}{Q^2}\,xy + \left(bc + b^2 - \frac{bc}{Q^2}\right) y^2 = 0$$

Elle admet pour déterminant :

$$b^2c^2 - [Q^2\, c\,(b + c) - bc]\,[Q^2\, b\,(b + c) - bc] = bc\,(b + c)^2\,(Q^2\, 1 - Q^2)$$

toujours positifs. Il y a 2 tangentes réelles T_1 T_2.

Le produit des racines est du signe de :

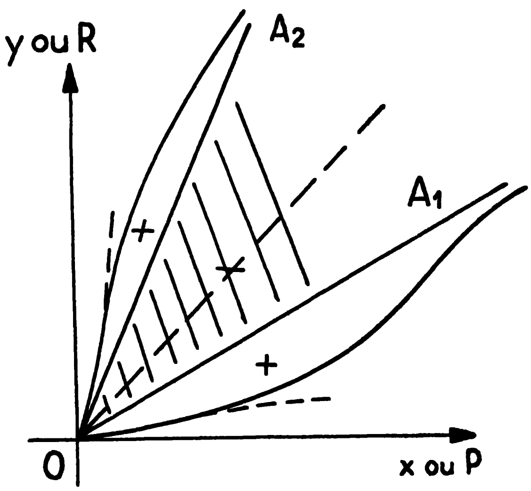
$$bc\,(b + c)^2 \left(Q^2 - \frac{b}{b + c}\right) \left(Q^2 - \frac{c}{b + c}\right)$$

d'où $\frac{b}{b + c} \gtrless Q^2 \gtrless \frac{c}{b + c}$: Racines de signes contraires

Autrement : 2 racines de même signe.

2) Le premier membre se décompose comme suit :

$$(\Phi - 2\,xy)\,(\Phi + 2\,xy) = (x - y)^2 \left(1 - \frac{1}{Q^2}\right) \left(x + y - \frac{x - y}{Q}\right) \left(x + y + \frac{x - y}{Q}\right)$$



du signe de :

$$[x\,(1 - Q) - y\,(1 + Q)]\,[x\,(1 + Q) - y\,(1 - Q)]$$

Pour $x = y$, on a

$$\Phi = 2\,x^2$$
$$\Phi^2 - 4\,x^2\,y^2 = 0$$
$$bc\Phi + c^2\,x^2 + by^2 = x^2\,(b + c)^2 > 0$$

La 1re bissectrice ne rencontre donc le contour qu'à l'origine 0. Pour ce qui de est l'inégalité, elle est forcément vérifiée. Le domaine cherché comprend donc la 1re bissectrice.

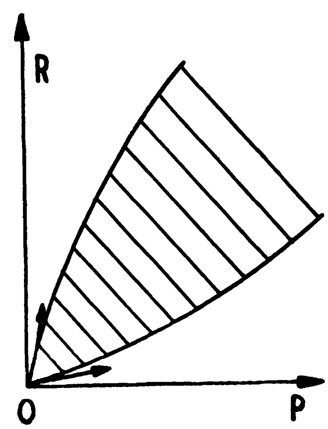
Deux des directions asymptotiques admettent des asymptotes passant par l'origine :

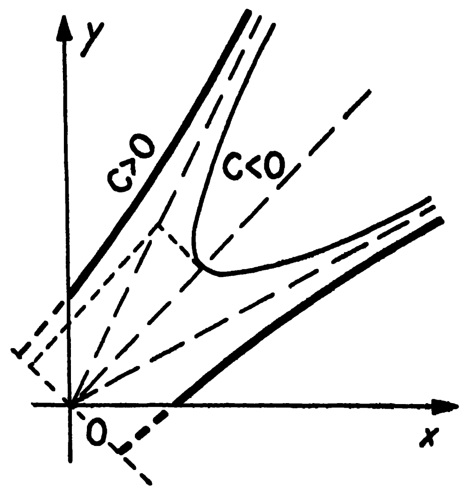
(A_1) $x(1 - Q) = y\,(1 + Q)$;
(A_2) $x(1 + Q) = y\,(1 - Q)$

La direction $(y - x)^2 = 0$ ne correspond pas à une branche réelle.

La méthode des régions indique que les branches sont à l'extérieur de l'angle aigu A_1 A^2., et à l'intérieur de l'angle des tangentes T_1 T_2.

Le domaine admissible a l'aspect de la zone hachurée ci-dessous.





RECHERCHE D'UNE HYPERBOLE EQUILATÈRE LIEU POSSIBLE DE POINTS (P, R)

Partons d'une hyperbole d'équation :

$$A\,x^2 - 2\,B\,xy + Ay^2 - C = 0$$

avec $B > A > 0$

Considérons n points (P, R) situés *sur* cette hyperbole, $P > 0$, $R > 0$.
Il vient (en sommant les n premiers membres)

$$A\,\Sigma\,P^2 - 2\,B\,\Sigma\,PR + A\,\Sigma\,R^2 - n\,C = 0$$

ou $$B\,\Sigma\,(P^2 - 2\,PR + R^2) = nC + (B - A)\,(\Sigma P^2 + \Sigma R^2)$$

ou $$\Sigma\,(P - R)^2 = \frac{nC}{B} + \frac{B - A}{B}\,(\Sigma P^2 + \Sigma R^2)$$

Rapprochons cette formule de :

$$\Sigma\,(P - R)^2 = Q^2\,[\Sigma\,P^2 + \Sigma\,R^2 + 2\,(\Sigma\,P^2.\,\Sigma\,R^2)^{1/2}]$$

Posons $$Q^2 = (B - A)/B = 1 - (A/B)$$

$$2Q^2\sqrt{\Sigma P^2.\Sigma R^2} = nC/B$$

ou $$s^2 = \frac{\sqrt{\Sigma P^2.\Sigma R^2}}{n} = \frac{C}{2\,BQ^2} = \frac{C}{2(B - A)}$$

Les points (P, R) sont sur une hyperbole sans point commun avec la bissectrice $P = R$ (: $C > 0$). Leur moment (généralisé) $s^2 = C/(2\,B - 2\,A)$ et leur indice $Q = 1 - (A/B)$ sont déterminés.

En particulier pour $C = 0$, on retrouve $Q = \dfrac{|1 - t|}{1 + t}$, $t = \dfrac{R}{P}$; et l'hyperbole dégénère en 2 droites passant par l'origine.

Réciproquement si l'on se donne Q et s^2, l'hyperbole est déterminée. Ces arcs d'hyperbole figurent en quelque sorte les 2 distributions (P, R) possibles, (l'arc supérieur correspondant à $P < R$).

Régionnement à l'aide du coefficient de corrélation

Soit un ensemble de couples (P_i R_i) dont le nuage de points est rapporté au point moyen G. Si les points sont assez nombreux l'adjonction d'une nouvelle comparaison ($x\ y$) ne modifie ni le point moyen G, ni les écarts types. La corrélation n'est pas diminuée si on a :

$$n\,\rho\sigma\tau + xy \geqslant \rho\,\sqrt{(n\sigma^2 + x^2)\,(n\tau^2 + y^2)}$$

où les ($x\ y$) sont les *variables PR centrées* (en G).
On suppose $n\,\rho\sigma\tau + xy > 0$

La frontière du domaine ainsi défini satisfait à :

$(n\,\rho\sigma\tau + xy)^2 = \rho^2\,(n\sigma^2 + x^2)\,(n\tau^2 + y^2)$
ou $(1 - \rho^2)\,x^2\,y^2 = n\rho^2\,(\sigma^2\,y^2 + \tau^2\,x^2 - 2\,\sigma\tau xy)$
ou $(1 - \rho^2)\,x^2\,y^2 = n\rho^2(\sigma y - \tau x)^2$

Cette courbe représente 2 hyperboles équilatères passant par l'origine; ces 4 arcs découpent le plan en 6 régions (dont 2 n'ont que l'origine comme point commun).

On observe que les points de la droite d'équation $\sigma y = \tau x$ vérifient l'inéquation, de sorte que la région hachurée appartient au domaine. Une discussion spéciale serait nécessaire pour les régions quadrillées.