

COEFFICIENT DE THEIL (J9)

(15 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **coefficient de THEIL** est destiné à apprécier la **qualité d'une prévision**. Il est distinct du **coefficient de concentration** du même nom.

(i) Soit $(\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times \mathbf{R}$ un **vecteur aléatoire** et :

$$(1) \quad \eta = f(\xi, b) + \varepsilon, \quad \text{avec } E \varepsilon = 0, b \in \mathbf{R}^Q,$$

l'écriture d'un **modèle de régression** non linéaire dans l'**espace des variables** (cf **régression non linéaire**). On suppose ce modèle estimé selon :

$$(2) \quad \tilde{y} = f(x, \tilde{b}),$$

où $\tilde{b}$ est un **estimateur** de b obtenu à partir du vecteur des **observations** y de l'**endogène** η et de la **matrice d'observation** X des K variables ξ_k ($\forall k \in N_k^*$). On note :

(a) $\xi = (\xi_1, \dots, \xi_K)$ les valeurs rétrospectives du vecteur ξ ;

(b) $\xi_{(N+1)}, \dots, \xi_{(N+H)}$ des valeurs (prospectives) données du vecteur **exogène** ξ , où N désigne le nombre d'observations (lignes de X) qui ont servi à estimer b et H désigne l'**horizon** de la prévision.

Le « **cheminement exogène** » prévu est décrit par l'équation :

$$(3) \quad \eta_{N+H} \sim = f(\xi_{(N+H)}, \tilde{b}), \quad \forall h \in N_H^* = \{1, \dots, H\}.$$

On note $y_{N+1}, \dots, y_{N+H}$ la **suite** des valeurs effectivement observées pour η sur les **indices** (d'**observation** ou de **période**) $N+1, \dots, N+H$.

On appelle alors **coefficient de H. THEIL** la **statistique** suivante :

$$(4) \quad T_N^2(H) = H^{-1} \cdot \sum_{h=1}^H (y_{N+h} \sim - y_{N+h})^2 / \{H^{-1} \cdot \sum_{h=1}^H (y_{N+h} \sim)^2 + N^{-1} \cdot \sum_{h=1}^H (y_{N+h})^2\}$$

(rapport entre la somme des carrés des écarts « prévus - observés » et la somme des carrés totale, « observés » et « prévus »).

En posant $\tilde{\eta}(H) = (\eta_{N+1} \sim, \dots, \eta_{N+H} \sim)'$ et $y(H) = (y_{N+1}, \dots, y_{N+H})'$, l'expression (4) définit une fonction symétrique en $(\tilde{\eta}(H), y(H))$:

$$(5) \quad T_N^2(H) = \|\tilde{\eta}(H) - y(H)\|^2 / (\|\tilde{\eta}(H)\|^2 + \|y(H)\|^2).$$

(ii) Le coefficient de THEIL vérifie $T_N^2(H) \in [0, 1]$. On peut considérer $T_N^2(H)$ comme un **coefficient** de **corrélacion** (empirique) :

(a) si la prévision est parfaite (ie $\eta_{N+h} \sim = y_{N+h}$, $\forall h \in N_H^*$), on a $T_N^2(H) = 0$;

(b) s'il existe une relation **linéaire** décroissante $\eta^{\sim} = -\gamma \cdot \eta$ (avec $\gamma > 0$), ie si $\eta_{N+h}^{\sim} = -\gamma \cdot y_{N+h}$, $\forall h \in N_H^*$, alors $T_N^2(H) = 1$.

Des **tests d'hypothèses** peuvent être déduit des propriétés du coefficient.

(iii) On appelle parfois aussi **coefficient de H. THEIL** le coefficient :

$$(5)' \quad T_N^2(N) = \|y(H)\|^{-2} \cdot \|\eta^{\sim}(H) - y(H)\|^2.$$