Économétrie

Contents

1 Introduction

1.1 Définition

L'économétrie est le domaine de l'économie qui s'occupe de l'application de la statistique mathématique et des outils de l'inférence statistique à la mesure empirique des relations postulées par la théorie économique

1.2 Prix Nobels

1980, Laurence Klein : Modèles macro-économétriques et leurs applications a l'analyse des fluctuations économiques
1989, Trygve Haavelmo : Approche probabiliste et les modèles a équations simultanées
2000, James Heckman : Travaux sur les théories et méthodes d'analyse des échantillons sélectifs
2000, Mac Fadden : Économétrie des choix discrets
2003, Robert F. Engle : Volatilité des séries temporelles pour les modèles ARCH 2003, Clive Granger : Théorie de la cointégration

2 Histoire de l'économétrie

2.1 Des Origines de l'économétrie à l'âge d'or de la modélisation Macro-économique

2.1.1 Premières études [17eme 18eme siècle]

L'autorité de la loi naturelle se dégage de celle de la religion et du prince
Apparition des statistiques

Statistique Allemande	**Arithmétique Politique** (Angleterre)
La Statistique est un moyen de classer les savoirs hétéroclites	Utilisation de techniques statistiques pour le dépouillement des registres paroissiaux
Conring, Herman	Petty, William : études natalité/mortalité
	King, Gregory : formalisation loi de demande

2.1.2 Genèse de l'économétrie [$19^{\text{ème}}$ siècle]

La statistique mathématique et l'évolution probabiliste de multiples champs permet de donner une dimension statistique à la représentation de la société

Galton	Statistique mathématique et Analyse mathématique de la régression (corrélation)
Edgeworth	Fonction de densité de la loi normale multivariée. Détermine les expressions de coefficient de régression multiple
Pearson	Coefficient de corrélation multiple : Analyse de la relation entre variables
Yule	Paupérisme, relation avec les mesures d'assistance
Hocker	Utilisation de variables retardées
Lenoir	Première estimation des lois d'offre et demande
Moore	Problème de détermination des salaires, utilisation des corrélations multiples, auto-corrélation et corrélogrammes

Conjoncturistes Américains :
Juglar, Kitchin , Kondratieff

Analyse des cycles économiques $\neq$ économie mathématique (Moore)

- Création d'instituts de conjecture (Russie 1920)
- Création du NBER (national bureau of economics research)

2.1.3 Imbrication de l'économie, des mathématiques et de la statistique [20$^{\text{ème}}$ siècle]

A la fin des années 20, l'économétrie gagne de l'intérêt en Europe
Création de la société d'économétrie le 29/12/1930 à Cleveland avec Frish, Timbergen. Elle officialise l'économétrie comme discipline et nomme Fisher président. La société d'économétrie recrute Cowles, qui, en contre-partie, demande la création d'une revue (Econometrica, RC : Frish) et la création d'un organisme de recherche Cowles Foundation
Roos est nommé premier président de la Cowles foundation

- Avant 1930 la théorie des probabilités est pratiquement inutilisée jusqu'à ce que Haavelmo propose d'y apporter une approche stochastique (aléatoire)
- Apparition d'une chaire d'économétrie
- Diversification des sources de financement par l'obtention de subventions (Rockfeller, NBER)
- Le terme d'identification apparaît après les travaux de Haavelmo

Tous ces travaux marquent le début de la modélisation macro-économique qui connaitra son age d'or dans les 60's - 70's avec Keynes et les comptes nationaux
Âge d'or de la modélisation macro-économique et essort des modèles dynamiques

1. Introduction de mécanismes dynamiques dans les modèles économétriques
2. Modèles a retard échelonnés Koyck
3. Développmeent des prévisions à a court terme Box & Jenkins

2.2 De la crise de la modélisation macro-économique à l'âge d'or de l'économétrie

2.2.1 Années 70 et la fin de l'âge d'or de la modélisation marco-économétrique selon la tradition de la Cowles Foundation

- Premier choc pétrolier → Remise en cause des modélisations
- Sims → Modèles VAR (*vector auto regressive*)
- Fonctions de réponse impulsionelle
- Etude de la causalité
- Test de la racine unitaire (Si les séries sont stationnaires ou non)
- Apparition de la modélisation ARCH, instauration d'une économétrie plurielle (modèle logit, probit)
- Introduction des principes Baillisiens

3 Le modèle linéaire simple à deux variables et généralisation a k variables

Modèle économétrique : Représentation simplifiée mais la plus exhaustive possible d'une entité économique donnée sous sa forme la plus courante représentée par un système d'équation (souvent linéaire) et relie des types de variables similaires

Variables explicatives → Variables exogènes
Variables expliquées → Variables endogènes

$$Y = f(X_1, X_2, .., X_n) \quad \text{une seule variable explicative } Y \text{ et une expliquée } X$$

$Y = \alpha + \beta X$ (α et β sont des paramètres inconnus). Cette relation est exacte, or cela est impossible en économie, on doit donc introduire un terme aléatoire (aléa ou erreur). cette variable a pour rôle de synthétiserl'ensemble des influences sur Y que X ne peut expliquer.

3 Types de modèles :

- Modèle en série chronologique (évolution au cours du temps)

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

- Modèle en coupe instantanée (à un moment donné dans le temps)

$$Y_i = \alpha + \beta X_i + U_i$$

- Modèle de panel

$$Y_{ti} = \alpha + \beta X_{ti} + U_{it}$$

Le fait d'introduire un aléa permet de faire des tests sur α et β.

Tout modèle non linéaire peut se ramener à un modèle linéaire (transformation par anamorphose).

Exemples (transformation)

$Y = \alpha e^{\beta X}$	$Y = \alpha X^{\beta}$	$Y = \frac{\alpha}{X^{\beta}}$
$\ln Y = \ln \alpha + \beta X$	$\ln Y = \ln \alpha + \beta \ln X$	$\ln Y = \ln \alpha - \beta \ln X$
$Y' = \alpha' + \beta X$	$Y' = \alpha' + \beta X'$	$Y' = \alpha' - \beta X'$

$$Y_t = \frac{Y_0}{1+\beta X^t} \Leftrightarrow \frac{Y_0}{Y_t} - 1 = \beta X^t$$

$$\ln\left(\frac{Y_0}{Y_t} - 1\right) = \ln\big(\beta X^t\big) \Leftrightarrow \ln Y_0 - \ln Y_t = \ln \beta + t \ln X$$

$$Y' = \beta' + tX'$$

3.1 Les hypothèses de base du modèle

Il existe deux méthodes qui permettent d'estimer α et β

- Méthode des MCO (moindres carrés ordinaires)
- Méthode du maximum de vraisemblance

Hypothèses

* X est une variable contrôlée (indépendante de l'aléa). $Cov(X, \varepsilon) = 0$

* $\varepsilon \to$ une hypothèse de normalité. $E(\varepsilon_t) = 0$
En moyenne l'ensemble des facteurs non expliqués par la régression (qui se retrouvent dans l'aléa) tendent a se compenser.

* Hypothèse d'homosédasticité :
la variance de ε_t est constante quel que soit le sous échantillon prélevé dans l'intervalle $\{1; n\}$
$E(\varepsilon_t^2) = \sigma_t^2$

* Hypothèse de non autocorrélation de l'aléa :
La distribution de ε_t qui correspond à X_t est indépendante de celle de $\varepsilon_{t'}$ qui correspond à $X_{t'}$
$Cov(\varepsilon_t, \varepsilon_{t'}) = E(\varepsilon_t \varepsilon_{t'}) = 0 \quad (\forall t \neq t')$

3.2 Les estimateurs des MCO (moindres carrés ordinaire)

La méthode des MCO consiste à minimiser la somme des carrés des écarts. Écarts entre les valeurs observées de la variable Y_t et la valeur calculée de cette même variable $\hat{Y}_t$.
Écart mesuré respectivement par les projections parallèlement à l'axe des ordonnées des points sur la droite de régression.

Programme de Minimisation :

$$\min \sum_t (Y_t - \hat{Y}_t)^2$$
$$\min \sum_t \varepsilon_t^2 = \min \sum_t (Y_t - \hat{Y}_t)^2 = \min \sum_t (Y_t - (\hat{\alpha} + \hat{\beta} X_t))^2$$

Conditions de premier ordre :

$$\begin{aligned} \frac{\partial \varphi}{\partial \hat{\alpha}} &= -2 \sum_t (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0 \\ \Leftrightarrow \quad & n\overline{Y} - n\hat{\alpha} - \hat{\beta} n \overline{X} = 0 \\ \Leftrightarrow \quad & \hat{\alpha} = \overline{Y} - \hat{\beta} - \overline{X} \end{aligned}$$

$$\begin{aligned} \frac{\partial \varphi}{\partial \hat{\beta}} &= -2 \sum_t (Y_t - \hat{\alpha} - \hat{\beta} X_t) X_t \\ \Leftrightarrow \quad & \sum_t (Y_t - (\overline{Y} - \hat{\beta}\overline{X} - \hat{\beta} X_t) X_t = 0 \\ \Leftrightarrow \quad & \sum_t X_t Y_t - \overline{Y} \sum_t X_t + \hat{\beta}\overline{X} \sum_t X_t - \hat{\beta} \sum_t X_t^2 = 0 \\ \Leftrightarrow \quad & \sum_t X_t Y_t - \overline{Y} n \overline{X} + \hat{\beta} \overline{X} n \overline{X} - \hat{\beta} \sum_t X_t^2 = 0 \\ \Leftrightarrow \quad & \sum_t X_t Y_t - n\overline{XY} - \hat{\beta} (\sum_t X_t^2 - n\overline{X^2}) = 0 \\ \Leftrightarrow \quad & \hat{\beta} = \frac{\frac{1}{n}\sum_t X_t Y_t - n\overline{XY}}{\sum_t X_t^2 - n\overline{X^2}} = \frac{nCov(X,Y)}{nV(X)} \end{aligned}$$

D'où

$$\hat{\beta} = \frac{\frac{1}{n}\sum_t X_t Y_t - \overline{XY}}{\frac{1}{n}\sum_t X_t^2 - \overline{X^2}} = \frac{Cov(X,Y)}{V(X)}$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta} - \overline{X}$$

Conditions de second ordre :

$$\frac{\partial^2 \varphi}{\partial \hat{\alpha}^2} = 2 > 0$$
$$\frac{\partial^2 \varphi}{\partial \hat{\beta}^2} = 2 > 0$$

La fonction est convexe, on a donc bien un minimum

La droite de régression passe par le point moyen qui est le centre de gravité (couple $\bar{X}, Y$).
Données centrées :

$$x_t = X_t - \bar{X} \qquad \hat{y_t} = \hat{Y_t} - \bar{Y}$$
$$y_t = Y_t - \bar{Y} \qquad \hat{x_t} = \hat{X_t} - \bar{Y}$$

On cherche $\hat{y}$:

$$\hat{Y} = \hat{y_t} + \bar{Y} = \hat{\alpha} + \hat{\beta} X_t = \hat{\alpha} + \hat{\beta}(x_t + \bar{X})$$

$$\begin{aligned}
\hat{y_t} &= -\bar{Y} + \hat{\alpha} + \hat{\beta} x_t + \hat{\beta}\bar{X} \\
&= -\bar{Y} + \hat{\beta} x_t + \hat{\beta}\bar{X} + (\bar{Y} - \hat{\beta}\bar{X} \\
\hat{y_t} &= \hat{\beta} x_t
\end{aligned}$$

4 Lois des estimateurs, intervalles de confiance et tests d'hypothèse

Comme les estimateurs sont linéaires de Y_t et comme Y_t dépend de ε_t et que ε_t est aléatoire et qu'il obéi a une loi normale, alors les estimateurs sont donc aléatoires et obéissent a une loi normale.

$$\hat{\alpha} \sim N\left(\alpha\ , \sigma_\varepsilon \sqrt{\frac{\sum X_t^2}{n \sum x_t^2}}\right) \qquad \hat{\beta} \sim N\left(\beta\ , \sigma_\varepsilon \sqrt{\frac{1}{\sum\limits_{t=1}^{n} x_t^2}}\right)$$

$$(n-2)\frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$$

4.1 Estimation par intervalle de confiance

4.1.1 Intervalle de confiance de β

$$\exists\ \beta_1, \beta_2 \ /\ 1 - \alpha = Prob[\beta_1 < \beta < \beta_2]$$

$$\frac{\frac{\hat{\beta} - \beta}{\sigma_\varepsilon / \sqrt{\sum\limits_{t=1}^{n} x_t^2}}}{\sqrt{(n-2) \cdot \frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2} \cdot \frac{1}{(n-2)}}} \sim T(n-2) \qquad \Leftrightarrow \qquad \frac{\hat{\beta} - \beta}{\hat{\sigma_\varepsilon}} \cdot \sqrt{\sum x_t^2} \sim T(n-2)$$

d'où :

$$
\begin{aligned}
1-\alpha &= Prob\left[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} < \frac{\hat{\beta}-\beta}{\hat{\sigma_\varepsilon}} \cdot \sqrt{\sum x_t^2} < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} \cdot \frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}} < \hat{\beta}-\beta < t_{1-\alpha/2} \cdot \frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}}\right] \\
&= Prob\left[\hat{\beta} - t_{1-\alpha/2} \cdot \frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}} < \beta < \hat{\beta} - t_{\alpha/2} \cdot \frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}}\right] \\
&\boxed{= Prob\left[\beta \in \left[\hat{\beta} \pm \left\{ \begin{array}{l} t_{\alpha/2} \\ t_{1-\alpha/2} \end{array} \right. \frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}}\right]\right]}
\end{aligned}
$$

4.1.2 Intervalle de confiance de $\hat{\alpha}$

$$\exists\, \alpha_1, \alpha_2 \,/\, 1-\alpha = Prob[\alpha_1 < \alpha < \alpha_2]$$

$$\frac{\frac{\hat{\alpha}-\alpha}{\sigma_\varepsilon / \sqrt{\frac{\sum X_t^2}{n \sum x_t^2}}}}{\sqrt{\frac{(n-2)}{n-2} \cdot \frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2}}} \sim T(n-2) \qquad \Leftrightarrow \qquad \frac{\hat{\alpha}-\alpha}{\hat{\sigma_\varepsilon}} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} \sim T(n-2)$$

d'où :

$$
\begin{aligned}
1-\alpha &= Prob\left[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} < \frac{\hat{\alpha}-\alpha}{\hat{\sigma_\varepsilon}} \cdot \frac{\sqrt{n \sum x_t^2}}{\sqrt{\sum X_t^2}} < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} \cdot \hat{\sigma_\varepsilon} \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} < \hat{\alpha}-\alpha < t_{1-\alpha/2} \cdot \hat{\sigma_\varepsilon} \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}}\right] \\
&= Prob\left[\hat{\alpha} - t_{1-\alpha/2} \cdot \hat{\sigma_\varepsilon} \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} < \alpha < \hat{\alpha} - t_{\alpha/2} \cdot \hat{\sigma_\varepsilon} \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}}\right] \\
&\boxed{= Prob\left[\alpha \in \left[\hat{\alpha} \pm \left\{ \begin{array}{l} t_{\alpha/2} \\ t_{1-\alpha/2} \end{array} \right. \hat{\sigma_\varepsilon} \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}}\right]\right]}
\end{aligned}
$$

4.1.3 Intervalle de confiance de σ_ε^2

$$\exists\, \sigma_1^2, \sigma_2^2 \,/\, 1-\alpha = Prob\left[\sigma_1^2 < \sigma_\varepsilon^2 < \sigma_2^2\right]$$

$$(n-2) \cdot \frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$$

d'où

$$\begin{aligned}
1-\alpha &= Prob\left[\chi^2_{\alpha/2} < \chi^2(n-2) < \chi^2_{1-\alpha/2}\right] \\
&= Prob\left[\chi^2_{\alpha/2} < (n-2)\cdot\frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2} < \chi^2_{1-\alpha/2}\right] \\
&= Prob\left[\frac{\chi^2_{\alpha/2}}{(n-2)\hat{\sigma_\varepsilon^2}} < \frac{1}{\sigma_\varepsilon^2} < \frac{\chi^2_{1-\alpha/2}}{(n-2)\hat{\sigma_\varepsilon^2}}\right] \\
&\boxed{= Prob\left[\frac{(n-2)\hat{\sigma_\varepsilon^2}}{\chi^2_{1-\alpha/2}} < \sigma_\varepsilon^2 < \frac{(n-2)\hat{\sigma_\varepsilon^2}}{\chi^2_{\alpha/2}}\right]}
\end{aligned}$$

4.2 Tests d'hypothèse

4.2.1 Test de β

Spécification du test

$$H_0 : \beta = 0 \quad H_1 : \beta \neq 0$$

Statistique de test

$$\frac{\hat{\beta}-\beta}{\hat{\sigma_\varepsilon}}\cdot\sqrt{\sum x_t^2} \sim T(n-2)$$

Intervalle d'acceptation

$$\begin{aligned}
1-\alpha &= Prob\left[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}\right] \\
&= Prob\left[t_{\alpha/2} < \frac{\hat{\beta}}{\hat{\sigma_\varepsilon}}\cdot\sqrt{\sum x_t^2} < t_{1-\alpha/2}\right] \text{ si } H_0 \text{ vrai} \\
&= Prob\left[t_{\alpha/2}\cdot\frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}} < \hat{\beta} < t_{1-\alpha/2}\cdot\frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}}\right] \\
&\boxed{= Prob\left[\hat{\beta} \in \left[\pm\, t_{\alpha/2}\frac{\hat{\sigma_\varepsilon}}{\sqrt{\sum x_t^2}}\right]\right] \text{ IA}}
\end{aligned}$$

RDD
Si $\hat{\beta} \in$ IA on accepte H_0 au risque de première espèce α
Si $\hat{\beta} \notin$ IA on rejette H_0 au risque de première espèce α (validité du modèle)

Autre façon de procéder

$$1-\alpha = Prob\left[t_{\alpha/2} < \frac{\hat{\beta}}{\hat{\sigma_\varepsilon} \,/\, \sqrt{\sum x_t^2}} < t_{1-\alpha/2}\right] = Prob\left[\left|\frac{\hat{\beta}}{\hat{\sigma_\varepsilon} \,/\, \sqrt{\sum x_t^2}}\right| < t_{1-\alpha/2}\right]$$

$$= Prob\left[\left|\frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}\right| < t_{1-\alpha/2}\right]$$

$$\boxed{= Prob\left[|t_c| < t_{1-\alpha/2}\right]} \qquad \text{Avec} \quad t_c = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \sim T(n-2)$$

RDD

Si $|t_c| < t_{1-\alpha/2}$ on accepte H_0 au risque de α
Si $|t_c| \geq t_{1-\alpha/2}$ on rejette H_0 au risque de α

4.2.2 Test de α

Spécification du test

$$H_0 : \alpha = 0 \quad H_1 : \alpha \neq 0$$

Statistique de test

$$\frac{\hat{\alpha}-\alpha}{\hat{\sigma_\varepsilon}} \cdot \frac{\sqrt{n\sum x_t^2}}{\sqrt{\sum X_t^2}} \sim T(n-2)$$

Intervalle d'acceptation

$$1-\alpha = Prob\left[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}\right]$$

$$= Prob\left[t_{\alpha/2} < \frac{\hat{\alpha}}{\hat{\sigma_\varepsilon}} \cdot \frac{\sqrt{n\sum x_t^2}}{\sqrt{\sum X_t^2}} < t_{1-\alpha/2}\right] \text{ si } H_0 \text{ vrai}$$

$$= Prob\left[t_{\alpha/2} \cdot \hat{\sigma_\varepsilon}\frac{\sqrt{\sum X_t^2}}{\sqrt{n\sum x_t^2}} < \hat{\alpha} < t_{1-\alpha/2} \cdot \hat{\sigma_\varepsilon}\frac{\sqrt{\sum X_t^2}}{\sqrt{n\sum x_t^2}}\right]$$

$$\boxed{= Prob\left[\hat{\alpha} \in \left[\pm\, t_{\alpha/2} \cdot \hat{\sigma_\varepsilon}\frac{\sqrt{\sum X_t^2}}{\sqrt{n\sum x_t^2}}\right]\right] \text{ IA}}$$

RDD

Si $\hat{\alpha} \in$ IA on accepte H_0 au risque de première espèce α
Si $\hat{\alpha} \notin$ IA on rejette H_0 au risque de première espèce α

Autre façon de procéder

$$t_c = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} \sim T(n-2)$$

RDD

Si $|t_c| < t_{1-\alpha/2}$ on accepte H_0 au risque de α
Si $|t_c| \geq t_{1-\alpha/2}$ on rejette H_0 au risque de α

4.2.3 Test de σ_ε^2

Spécification du test

$$H_0 : \sigma_\varepsilon^2 = \sigma_0^2 \quad H_1 : \sigma_\varepsilon^2 \neq \sigma_0^2$$

Statistique de test

$$(n-2)\frac{\hat{\sigma_\varepsilon^2}}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$$

Intervalle d'acceptation

$$\begin{aligned} 1-\alpha &= Prob\left[\chi^2_{\alpha/2} < \chi^2(n-2) < \chi^2_{1-\alpha/2}\right] \\ &= Prob\left[\chi^2_{\alpha/2} < (n-2)\cdot\frac{\hat{\sigma_\varepsilon^2}}{\sigma_0^2} < \chi^2_{1-\alpha/2}\right] \text{ si } H_0 \text{ vrai} \\ &\boxed{= Prob\left[\chi^2_{\alpha/2}\cdot\frac{\sigma_0^2}{n-2} < \sigma_\varepsilon^2 < \chi^2_{1-\alpha/2}\cdot\frac{\sigma_0^2}{n-2}\right]} \end{aligned}$$

RDD
Si $\sigma_\varepsilon^2 \in$ IA on accepte H_0 au risque de première espèce α
Si $\sigma_\varepsilon^2 \notin$ IA on rejette H_0 au risque de première espèce α

4.3 Étude de la corrélation

Le **coefficient de corrélation** est un coefficient qui mesure le degrés de covariation linéaire, c'est à dire la manière dont varient ensemble les variables entre elles.

$$r_{XY} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_y} = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$$

4.3.1 Propriétés

- $-1 \leq r \leq 1$
- r est sans dimension
- r est symétrique : $r_{XY} = r_{YX}$
- r n'est pas affecté par un changement de variable $rXY = r_{xy}$ avec $r_{xy} = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$
- Relation entre $\hat{\beta}$ et r : $\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2}$

$$r = \hat{\beta}\cdot\frac{\sqrt{\sum x_t^2}}{\sqrt{\sum y_t^2}} \quad\Leftrightarrow\quad r = \hat{\beta}\frac{\sqrt{\frac{1}{n}\sum\left(X_t-\overline{X}\right)^2}}{\sqrt{\frac{1}{n}\sum\left(Y_t-\overline{Y}\right)^2}} \quad\Leftrightarrow\quad r = \hat{\beta}\cdot\frac{s_X}{s_Y}$$

4.3.2 Analyse de la Variance

$$e_t = y_t - \hat{y}_t \qquad \text{On sait que : } \hat{y}_t = \hat{\beta} x_t$$

$$\sum_{t=1}^{n} y_t^2 = \sum_{t=1}^{n} (e_t + \hat{y}_t)^2 = \sum_{t=1}^{n} \hat{y}_t^2 + \sum_{t=1}^{n} e_t^2 + 2 \sum_{t=1}^{n} e_t \hat{y}_t$$

$$\begin{aligned}
\sum_{t=1}^{n} e_t \hat{y}_t &= \sum_{t=1}^{n} (y_t - \hat{y}_t)\, \hat{\beta} x_t \\
&= \hat{\beta} \left[\sum_{t=1}^{n} (y_t - \hat{y}_t)\, x_t \right] \\
&= \hat{\beta} \left[\sum_{t=1}^{n} \left(\hat{y}_t - \hat{\beta} x_t \right) x_t \right] \\
&= \hat{\beta} \left[\sum_{t=1}^{n} y_t - x_t - \hat{\beta} x_t^2 \right] \\
&= \hat{\beta} \underbrace{\left[\sum_{t=1}^{n} y_t x_t - \hat{\beta} \sum_{t=1}^{n} x_t^2 \right]}_{=0}
\end{aligned}$$

D'où $$\boxed{\sum_{t=1}^{n} y_t^2 = \sum_{t=1}^{n} \hat{y}_t^2 + \sum_{t=1}^{n} e_t^2} \qquad \text{EQ de l'analyse de la variance}$$

$$\Leftrightarrow \quad \underbrace{\sum_{t=1}^{n} \left(Y_t - \bar{Y}\right)^2}_{\text{Variance totale}} = \underbrace{\sum_{t=1}^{n} \left(\hat{Y}_t - \bar{Y}\right)^2}_{\text{Variance expliquée}} + \underbrace{\sum_{t=1}^{n} e_t^2}_{\text{Variance residuelle}}$$

La fluctuation totale des valeurs de Y_t autour de la moyenne de l'échantiollon peut etre décomposée en deux éléments qui sont :

- **La variance expliquée** : Variation des valeurs de $\hat{Y}$ autour de la moyenne. Somme des carrés expliquée par l'influence linéaire de X

- **La variance residuelle** : Variation résiduelle des valeurs de Y autour de la droite des moindres carrés

4.3.3 Calcul du coefficient de détermination R^2

$$\sum_{t=1}^{n} y_t^2 = \sum_{t=1}^{n} \hat{y}_t^2 + \sum_{t=1}^{n} e_t^2 \quad \Leftrightarrow \quad \frac{\sum y_t^2}{\sum y_t^2} = \frac{\sum \hat{y}_t^2}{\sum y_t^2} + \frac{\sum e_t^2}{\sum y_t^2} \quad \Leftrightarrow \quad 1 = \frac{\hat{\beta}^2 \sum x_t^2}{\sum y_t^2} + \frac{\sum e_t^2}{\sum y_t^2}$$

Or : $r = \hat{\beta}\frac{\sqrt{\sum x_t^2}}{\sqrt{\sum y_t^2}}$ D'où :

$$\Leftrightarrow 1 = r^2 + \frac{\sum e_t^2}{\sum y_t^2}$$

$$\Leftrightarrow r^2 = 1 - \frac{\sum e_t^2}{\sum y_t^2}$$

$$\boxed{\Leftrightarrow r^2 = 1 - \frac{\mathrm{VR}}{\mathrm{VT}} = \frac{\mathrm{VE}}{\mathrm{VT}}}$$

Avec : $0 \leq r^2 \leq 1$

$$r^2 = \hat{\beta}^2 \frac{\sum x_t^2}{\sum y_t^2} \quad \text{Dans le cas d'une regression lineaire : } r_{xy} = \sqrt{r^2}$$

4.3.4 Test du coefficient de corrélation linéaire r

Statistique de test :

$$T(n-2) \sim \frac{\hat{\beta} - \beta}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2}$$

Rappel :

$$\hat{\beta} = r \cdot \frac{\sqrt{\sum y_t^2}}{\sqrt{\sum x_t^2}} \qquad \text{et} \qquad \hat{\sigma}_\varepsilon^2 = \frac{\sum e_t^2}{n-2}$$

D'où :

$$1 = r^2 + \frac{\sum e_t^2}{\sum y_t^2} \quad \Leftrightarrow \quad \sum_{t=1}^{n} e_t^2 = (1 - r^2) \sum_{t=1}^{n} y_t^2$$

Donc :

$$\hat{\sigma}_\varepsilon^2 = \frac{(1 - r^2) \sum y_t^2}{n-2}$$

Posons $H_0 : \beta = 0$, sous H_0 :

$$T(n-2) \sim \frac{\hat{\beta}}{\hat{\sigma}_\varepsilon} \cdot \sqrt{\sum x_t^2} = r \cdot \frac{\cancel{\sqrt{\sum y_t^2}}}{\cancel{\sqrt{\sum x_t^2}}} \cdot \frac{\cancel{\sqrt{\sum x_t^2}}}{\sqrt{(1 - r^2)}\cancel{\sqrt{\sum y_t^2}}} \cdot \sqrt{(n-2)}$$

$$\boxed{T(n-2) \sim r \cdot \frac{\sqrt{n-2}}{\sqrt{1 - r^2}}}$$

Spécification du test

$$H_0 : \begin{matrix} \rho = 0 \\ \hat{\beta} = 0 \end{matrix} \qquad H_1 : \rho \neq 0$$

Statistique de test

$$T(n-2) \sim r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

Intervalle d'acceptation

$$\begin{aligned} 1-\alpha &= Prob\left[t_{\alpha/2} < T(n-2) < t_{1-\alpha/2}\right] \\ &= Prob\left[t_{\alpha/2} < r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}} < t_{1-\alpha/2}\right] \\ &\boxed{= Prob\left[\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| < t_{1-\alpha/2}\right]} \end{aligned}$$

RDD
Si $\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| < t_{1-\alpha/2}$ on accepte H_0 au risque de α
Si $\left|r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}\right| \geq t_{1-\alpha/2}$ on rejette H_0 au risque de α

4.3.5 Tableau de l'analyse de la variance

Origine des variations	$\sum$ des carrés des écarts	DDL	Carrés moyens
VE	$Q_1 = \sum \hat{y}^2 = \hat{\beta}^2 \sum x_t^2$	1	$Q_1/1 = \sum \hat{y}^2$
VR	$Q_2 = \sum e_t^2$	$n-2$	$Q_2/n-2 = \hat{\sigma}_\varepsilon^2$
VT	$Q_3 = \sum y_t^2$	$n-1$	$\times$

Remarque : $Q_3 \simeq Q_1 + Q_2$

5 Test du coefficient de détermination R^2

Spécification du test

$$H_0 : \rho^2 = 0 \quad H_1 : \rho^2 \neq 0$$

Statistique de test

$$T(n-2) \sim \frac{r}{\sqrt{1-r^2}}$$

$$F_c = \frac{r^2}{1-r^2}(n-2) \sim F(1, n-2)$$

Intervalle d'acceptation

$$1-\alpha = Prob\left[\frac{r^2}{1-r^2}(n-2) < F_{1-\alpha}(1, n-2)\right]$$

RDD
Si $F_c < F_{1-\alpha}(1, n-2)$ On accepte H_0 au risque de α
Si $F_c \geq F_{1-\alpha}(1, n-2)$ On rejette H_0 au risque de α

6 Utilisation du modèle de régression en prévision

6.1 Intervalle de confiance : de la valeur moyenne de Y connaissant une valeur donnée de X

X_0 : valeur donnée de X
$Y_0 = \alpha + \beta X_0 + \varepsilon_0 \qquad \hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t \qquad \Rightarrow \qquad \hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$
On appelle valeur moyenne de Y connaissant une valeur donnée de X :

$$E\left(Y_0/X_0\right) = \alpha + \beta X_0$$

$\hat{Y}_0$ est un estimateur linéaire sans biais de $\alpha + \beta X_0$ et de $E\left(Y_0/X_0\right)$

$$E(\hat{Y}_0) = \alpha + \beta X_0 \qquad E(\hat{Y}_0) = E(\hat{\alpha}) + E(\hat{\beta} X_0) = \alpha + \beta X_0$$

$$\hat{Y}_0 \sim N\left(\underbrace{E\left(\hat{Y}_0\right)}_{\alpha + X_0 \beta}, \underbrace{\sqrt{V\left(\hat{Y}_0\right)}}_{?}\right)$$

Calcul de $V(\hat{Y}_0)$

$$\begin{aligned} V(\hat{Y}_0) &= V(\hat{\alpha} + \hat{\beta} X_0) = V(\hat{\alpha}) + V(\hat{\beta} X_0) + 2\,\mathrm{Cov}(\hat{\alpha}, \hat{\beta} X_0) \\ &= V(\hat{\alpha}) + X_0^2 V(\hat{\beta}) + 2\,\mathrm{Cov}(\hat{\alpha}, \hat{\beta} X_0) \end{aligned}$$

Avec

$$\begin{aligned} V(\hat{\beta}) &= \frac{\sigma_\varepsilon^2}{\sqrt{\sum x_t^2}} \\ V(\hat{\alpha}) &= \sigma_\varepsilon^2 \cdot \frac{\sqrt{\sum X_t^2}}{\sqrt{n \sum x_t^2}} = \sigma_\varepsilon^2 \cdot \left(\frac{\sum x_t^2 + n\overline{X}^2}{n \sum x_t^2}\right) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x_t^2}\right) \\ \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) &= -\frac{\overline{X}}{\sum x_t^2} \cdot \sigma_\varepsilon^2 \end{aligned}$$

d'où

$$\begin{aligned} V(\hat{Y}_0) &= \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x_t^2}\right) + X_0^2 \left(\frac{\sigma_\varepsilon^2}{\sqrt{\sum x_t^2}}\right) + 2X_0 \left(-\frac{\overline{X}}{\sum x_t^2} \cdot \sigma_\varepsilon^2\right) \\ &= \sigma \left[\frac{1}{n} + \frac{\overline{X} + X_0^2 - 2X_0\overline{X}}{\sum x_t^2}\right] \\ &= \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(X_0 - \overline{X}^2)}{\sum x_t^2}\right] \end{aligned}$$

$$\boxed{\hat{Y}_0 \sim N\left(\alpha + \beta X_0; \sigma_\varepsilon \sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X}^2)}{\sum x_t^2}}\right)}$$

σ_ε Inconnu, d'où studentisation :

$$(n-2)\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(n-2) \qquad \hat{Y}_0 \sim N\left(\alpha + \beta X_0; \sigma_\varepsilon\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X}^2)}{\sum x_t^2}}\right)$$

$$t_c = \frac{\frac{\hat{Y}_0 - \alpha + \beta X_0}{\sigma_\varepsilon\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X}^2)}{\sum x_t^2}}}}{\sqrt{\frac{n-2}{n-2} \cdot \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \sim T(n-2)$$

$$t_c = \frac{\hat{Y}_0 - \alpha + \beta X_0}{\hat{\sigma}_\varepsilon\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X}^2)}{\sum x_t^2}}} \sim T(n-2)$$

Intervalle de confiance :

$$\begin{aligned} 1 - \alpha &= Prob\left[t_{\alpha/2} \leq T(n-2) \leq t_{1-\alpha/2}\right] \\ &= \end{aligned}$$

6.2 Tests d'hypothèse : comparaison d'une prévision ponctuelle à la droite des moindres carrés

7 Bibliographie

- Bourbonnais - Econométrie - Dunod
- Johnston Dinardo - Econométrie
- Greene - Econométrie - Pearson