Econométrie

Olivier Donni

1 Introduction

1.1 Qu'est-ce que l'économétrie?

C'est de la statistique appliqué à l'économie, qui permet de tester des théories économiques, prédire des comportement économiques et évaluer des politiques économiques. Les caractéristiques de l'économétrie sont:

(1) Les données utilisées ne sont pas expérimentales;

(2) Les modèles à estimer sont structurels (et non descriptifs).

1.2 Les étapes d'une analyses empirique

(1) D'abord, un modèle économique est construit. Exemples:
   - (a) le modèle de demande du consommateur;
   - (b) le modèle de capital humain et salaire (Mincer);
   - (c) le modèle de criminalité (Becker);
   - (d) le modèle de rationalité des gardiens de but (Chiappori & Levitt);
   - (e) le modèle des relations extra-conjugales (Fair).

(2) Ensuite, un modèle économétrique est construit en définissant les variables, en choisissant une forme fonctionnelle et en introduisant un terme aléatoire. Ce modèle est estimé.

1.3 Structure des données

Il existe plusieurs types de données. Entre autres,

(1) Des données transversales (*cross-section data*): en général, un échantillon

aléatoire issu d'une population; souvent des données microéconomiques.

Definition 1 *Un échantillon aléatoire est un ensemble de variables aléatoires indépendantes et de même distribution.*

(2) Des séries temporelles: journalières, hebdomadaires, mensuelles, trimestrielles, ... Ces données sont souvent macroéconomiques et ne constituent pas, en général, un échantillon aléatoire.

(3) Des séries temporelle de données transversales.

(4) Des données de panel.

1.4 Principal objectif d'une étude économétrique

Le principal objectif d'une étude empirique est généralement de mesurer une relation 'causale' d'une variable sur une autre. Dans ce cas, la notion de ceteris paribus joue un rôle important. Et les techniques économétriques simulent un effet cétéris paribus.

Exemples:

(1) Effets d'un engrais;

(2) Taux de rendement de l'éducation.

2 Modèle de régression linéaire simple

2.1 Définition du modèle de régression simple

Le modèle de régression simple s'écrit:

$$y = \beta_0 + \beta_1 x + u$$

où β_0 est la constante, β_1 est la pente et u le terme aléatoire. On parle de la régression de y sur x, où les variables y et x sont appelées:

$$y = \begin{cases} \text{variable dépendante} \\ \text{variable expliquée} \\ \text{variable de réponse} \\ \text{variable prédite} \\ \text{régressant} \end{cases}, \qquad x = \begin{cases} \text{variable indépendante} \\ \text{variable explicative} \\ \text{variable de contrôle} \\ \text{variable prédictrice} \\ \text{régresseur} \end{cases}$$

Ce modèle est linéaire car l'effet 'ceteris paribus' de x sur y est linéaire:

$$\Delta y = \beta_1 \cdot \Delta x \qquad \text{si} \qquad \Delta u = 0$$

où β_1 est souvent le paramètre d'intérêt.

Exemple 1. La relation entre le rendement de parcelles de terre et la quantité d'engrais utilisée s'écrit:

$$\text{REND} = \beta_0 + \beta_1 \cdot \text{ENG} + u$$

Exemple 2: La relation entre le salaire et le niveau d'éducation (mesuré en années) s'écrit:

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + u$$

2.2 Dérivation des estimateurs des MCO

2.2.1 Calcul des estimateurs

Soit un échantillon $\{(x_i, y_i) : i = 1, ..., N\}$. Les estimateurs des MCO, $\hat{\beta}_0$ et $\hat{\beta}_1$, sont obtenus par la minimisation du carré des résidus:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

où le résidu des MCO pour l'observation i est défini par

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i.$$

Donc, les estimateurs des MCO sont ceux qui minimisent le carré des résidus. Les conditions de premier ordre sont:

$$\begin{aligned} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \sum_{i=1}^{N} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \end{aligned}$$

La première équation devient:

$$\begin{aligned} \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \text{avec} \qquad \bar{y} &= \frac{1}{N} \sum_{i=1}^{N} y_i \text{ et } \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \end{aligned}$$

et donne $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

La deuxième équation devient:

$$\begin{aligned} \sum_{i=1}^{N} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \sum_{i=1}^{N} x_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) &= 0 \end{aligned}$$

Elle devient ensuite:

$$\begin{aligned}\sum_{i=1}^{N} x_i(y_i - \bar{y}) &= \sum_{i=1}^{N} x_i(\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x}) \\ \sum_{i=1}^{N} x_i(y_i - \bar{y}) &= \hat{\beta}_1 \sum_{i=1}^{N} x_i(x_i - \bar{x})\end{aligned}$$

Donc, si

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 \neq 0,$$

alors

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\widehat{\mathrm{cov}(x,y)}}{\widehat{\mathrm{var}(x)}} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}\end{aligned}$$

2.2.2 Quelques définitions

La valeur prédite des MCO de y_i conditionnellement à x_i est:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Le résidu des MCO est:

$$\hat{u}_i = y_i - \hat{y}_i = (\beta_0 + \beta_1 x_i + u_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

La droite de régression des MCO est:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Celle-ci est une estimation de la fonction de régression de la population:

$$E(y|x) = \beta_0 + \beta_1 x.$$

L'estimateur de la pente est:

$$\frac{\Delta \hat{y}}{\Delta x} = \hat{\beta}_1.$$

Ce dernier est généralement le paramètre le plus intéressant.

Exemple 4 (Wooldridge, 2003): La droite de régression du salaire des P.-D.G. sur le rendement de l'action de leur entreprise est égale à:

$$\widehat{\text{SAL}} = 963.191 + 18.501 \cdot \text{REND}$$

Exemple 5 (Wooldridge, 2003): La droite de régression du salaire sur le niveau d'éducation est égale à:

$$\widehat{\text{SAL}} = -0.90 + 0.54 \cdot \text{EDUC}$$

Exemple 6 (Wooldridge, 2003): La droite de régression du pourcentage de voix obtenues d'un candidat et de la part des dépenses (dans les dépenses totales) de ce candidat est égale à:

$$\widehat{\text{VOTE\_A}} = 40.90 + 0.306 \cdot \text{PART\_A}$$

2.3 Propriétés algébriques des MCO

Les résidus satisfont un certain nombre de propriétés algébriques. Celles-ci découlent directement de la manière dont les estimateurs des MCO sont construits, et ne nécessitent généralement pas de démonstration.

Propriété 1: La moyenne des résidus est nulle:

$$\sum_{i=1}^{N} \hat{u}_i = 0$$

Propriété 2: La covariance entre les résidus et les valeurs de la variable explicative est nulle:

$$\sum_{i=1}^{N} x_i \hat{u}_i = 0$$

Propriété 3: La régression passe par le point moyen de l'échantillon:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Propriété 4: La covariance entre les résidus et les valeurs prédites est nulle:

$$\sum_{i=1}^{N} \hat{y}_i \hat{u}_i = \sum_{i=1}^{N} (\beta_0 + \beta_1 \hat{x}_i) \hat{u}_i = \hat{\beta}_0 \sum_{i=1}^{N} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{N} x_i \hat{u}_i = 0.$$

La dérivation de la propriété 5 nécessite de définir les concepts suivants:

$$\begin{aligned} \text{SST} &= \sum_{i=1}^{N} (y_i - \bar{y})^2 \\ \text{SSE} &= \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 \\ \text{SSR} &= \sum_{i=1}^{N} \hat{u}_i^2 \end{aligned}$$

Propriété 5: La somme des carrés totaux est égale à la somme des carrés expliqués et la somme des carrés résiduels:

$$\text{SST} = \text{SSE} + \text{SSR}$$

Cette propriété permet de calculer le coefficient de détermination (R^2):

$$\begin{aligned} R^2 &= \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}} \\ &= \frac{\left(\widehat{\text{cov}(x,y)}\right)^2}{\widehat{\text{var}(x)}\widehat{\text{var}(y)}} \end{aligned}$$

Ce coefficient est égal au carré du coefficient de corrélation.

Démonstration de la Proposition 5:

$$
\begin{aligned}
\text{SST} &= \sum_{i=1}^{N}(y_i-\bar{y})^2 \\
&= \sum_{i=1}^{N}((y_i-\hat{y}_i)+(\hat{y}_i-\bar{y}))^2 \\
&= \sum_{i=1}^{N}(\hat{u}_i+(\hat{y}_i-\bar{y}))^2 \\
&= \sum_{i=1}^{N}(\hat{u}_i^2+2\hat{u}_i(\hat{y}_i-\bar{y})+(\hat{y}_i-\bar{y})^2) \\
&= \sum_{i=1}^{N}\hat{u}_i^2+2\sum_{i=1}^{N}\hat{u}_i(\hat{y}_i-\bar{y})+\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2 \\
&= \text{SSR}+2\sum_{i=1}^{N}\hat{u}_i(\hat{y}_i-\bar{y})+\text{SSE}
\end{aligned}
$$

Or

$$\sum_{i=1}^{N}u_i(\hat{y}_i-\bar{y})=\sum_{i=1}^{N}\hat{u}_i\hat{y}_i-\bar{y}\sum_{i=1}^{N}\hat{u}_i=0 \text{ en vertu de la Propriété 4}$$

Exemple 7 (Wooldridge, 2003): Le R^2 de la régression du salaire des P.-D.G. sur le rendement des actions:

$$
\begin{aligned}
\widehat{\text{salary}} &= 963.191+18.501\cdot\text{roe} \\
R^2 &= 0.0132
\end{aligned}
$$

Exemple 8 (Wooldridge, 2003): Le R^2 de la régression du pourcentage de voix obtenues sur la part des dépenses:

$$
\begin{aligned}
\widehat{\text{VOTE\_A}} &= 40.90+0.306\cdot\text{PART\_A} \\
R^2 &= 0.505
\end{aligned}
$$

2.4 Unités de mesure et non linéarité

2.4.1 Changement d'unités de mesure:

La variable expliquée est exprimée en miliers d'euros, et la variable explicative en pourcentages:

$$\widehat{\text{SAL}} = 963.191 + 18.501 \cdot \text{REND}.$$

La variable expliquée est exprimée en euros, et la variable explicative en pourcentages:

$$\widehat{\text{SAL\_DOLL}} = 963,191 + 18,501 \cdot \text{REND}.$$

La variable expliquée en expliquée en milliers d'euros, et la variable explicative en décimales:

$$\widehat{\text{SAL}} = 963.191 + 1850.1 \cdot \text{REND\_DEC}.$$

Dans tous les cas, le R^2 ne se modifie pas.

2.4.2 Forme fonctionnelle non linéaire:

Le modèle de régression simple est linéaire dans les paramètres. Cependant, certaines relations non linéaires entre les variables peuvent être modélisées.

Exemple 9 (Wooldridge, 2003): La droite de régression du logarithme du salaire sur le nibeau d'éducation est égale à:

$$\begin{aligned} \widehat{\log(\text{SAL})} &= 0.584 + 0.083 \cdot \text{EDUC} \\ R^2 &= 0.186 \end{aligned}$$

Exemple 10 (Wooldridge, 2003): La droite de régression du salaire des P.-D.G. sur le logarithme du chiffre d'affaire de leur entreprise est égale à:

$$\begin{aligned} \widehat{\log(\text{SAL})} &= 4.822 + 0.257 \cdot \log(\text{VENTES}) \\ R^2 &= 0.211 \end{aligned}$$

2.5 Propriétés statistiques des MCO

Les bonnes propriétés statistiques des estimateurs des MCO nécessitent qu'un ensemble d'hypothèses soient satisfaites.

Hypothèse 1′ (linéarité dans les paramètres): Le modèle dans la population peut se décrire par une relation linéaire à une seule variable explicative telle que:

$$y = \beta_0 + \beta_1 x + u$$

où β_0, β_1 sont des paramètres, et u est un terme aléatoire.

Hypothèse 2′ (échantillonnage aléatoire): Un échantillon aléatoire de N observations, $\{(y_i, x_i) : i = 1, \ldots, N\}$ issu du modèle de population décrit en H1.

Hypothèse 3′ (moyenne conditionnelle nulle): Le terme u a une espérance de zéro pour toute valeur des variables indépendantes. En d'autres termes,

$$E(u|x) = 0$$

Hypothèse 4′ (variation dans le régresseur): Dans l'échantillon (et donc dans la population), le régresseur n'est pas une constante:

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 \neq 0$$

Les démonstrations qui suivent sont basées sur la transformation suivante:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{N}(x_i-\bar{x})y_i-\sum_{i=1}^{N}(x_i-\bar{x})\bar{y}}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)y_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)(\beta_0+\beta_1 x_i+u_i)}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \beta_0\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)}{\sum_{i=1}^{N}(x_i-\bar{x})^2}+\beta_1\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)x_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2}+\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)u_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \beta_1\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)x_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2}+\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)u_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \\
&= \beta_1+\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)u_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2}
\end{aligned}$$

L'estimateur $\hat{\beta}_1$ est donc égal à la somme de la vraie valeur de β_1 dans la population et une combinaison de termes aléatoires. L'estimateur $\hat{\beta}_1$ est donc une variable aléatoire.

Théorème 1 (absence de biais des MCO): Sous les hypothèses $1'$ à $4'$, les estimateurs des MCO sont non biaisés:

$$E(\hat{\beta}_0)=\beta_0 \text{ et } E(\hat{\beta}_1)=\beta_1$$

Démonstration: (les espérances sont conditionnelles aux valeurs de l'échantillon; donc, s_x^2 et $(x_i-\bar{x})$ sont non-aléatoires)

Partie 1: De ce qui précède, on a:

$$\hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{N}(x_i-\bar{x}_i)u_i}{\sum_{i=1}^{N}(x_i-\bar{x})^2}.$$

Si l'on prend les espérances, on a:

$$
\begin{aligned}
E(\hat{\beta}_1) &= \beta_1 + E\left(\frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right) \\
&= \beta_1 + \frac{\sum_{i=1}^{N} E((x_i - \bar{x})u_i)}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})E(u_i)}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \times 0}{\sum_{i=1}^{N}(x_i - \bar{x})^2}
\end{aligned}
$$

Partie 2: La définition de $\hat{\beta}_0$ est:

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \\
&= \beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x} \\
&= \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}
\end{aligned}
$$

Si l'on prend les espérances, on a:

$$
\begin{aligned}
E(\hat{\beta}_0) &= \beta_0 + E((\beta_1 - \hat{\beta}_1)\bar{x}) + E(\bar{u}) \\
&= \beta_0 + \bar{x}E(\beta_1 - \hat{\beta}_1)
\end{aligned}
$$

Exemple 11 (score et subvention des repas): Considérons la régression du pourcentage de réussite au MEAP test (Michigan Educational Assessment Program) sur la proportion d'élèves qui bénéficient d'une subvention pour les repas:

$$
\begin{aligned}
\widehat{\text{MEAP}} &= 32.14 - 0.319 \cdot \text{PROP\_SUBV}, \\
R^2 &= 0.171.
\end{aligned}
$$

Les estimateurs sont certainement biaisés car l'hypothèse d'espérance conditionnelle n'est pas satisfaite.

Pour calculer la variance des estimateurs, l'hypothèse suivante, qui assure que les termes aléatoires ont une variance constante, est nécessaire.

Hypothèse 5′ (homoscédasticité): La variance des termes aléatoires (conditionellement à x) est constante. En d'autres termes,

$$var(u|x) = \sigma^2.$$

Remarque: Les hypothèses 3 et 5 peuvent être mises sous la forme de moyenne et de variance conditionnelle:

$$\begin{aligned} E(y|x) &= \beta_0 + \beta_1 x \\ var(y|x) &= \sigma^2 \end{aligned}$$

On a alors le théorème suivant.

Théorème 2 (variance des MCO): Sous les hypothèses $1'$ à $5'$,

$$\mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^N x_i^2}{N \sum_{i=1}^N (x_i - \bar{x})^2}, \qquad \mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Démonstration:

Partie 1: On a:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

En prenant la variance des membres de droite et de gauche, on obtient:

$$\begin{aligned}
\mathrm{var}(\hat{\beta}_1) &= \mathrm{var}\left(\frac{\sum_{i=1}^N (x_i-\bar{x})u_i}{\sum_{i=1}^N (x_i-\bar{x})^2}\right) = \frac{\mathrm{var}\left(\sum_{i=1}^N (x_i-\bar{x})u_i\right)}{\left(\sum_{i=1}^N (x_i-\bar{x})^2\right)^2} \\
&= \frac{\sum_{i=1i}^N (x_i-\bar{x})^2 \mathrm{var}\,(u_i)}{\left(\sum_{i=1}^N (x_i-\bar{x})^2\right)^2} \text{ puisque les } u_i \text{ sont indépendants} \\
&= \frac{\sum_{i=1}^N (x_i-\bar{x})^2 \sigma^2}{\left(\sum_{i=1}^N (x_i-\bar{x})^2\right)^2} \\
&= \sigma^2 \frac{\sum_{i=1}^N (x_i-\bar{x})^2}{\left(\sum_{i=1}^N (x_i-\bar{x})^2\right)^2} \\
&= \frac{\sigma^2}{\left(\sum_{i=1}^N (x_i-\bar{x})^2\right)}
\end{aligned}$$

Partie 2:

$$\begin{aligned}
\hat{\beta}_0 &= \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u} \\
\mathrm{var}(\hat{\beta}_0) &= \mathrm{var}\left((\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}\right) \\
&= \bar{x}^2 \mathrm{var}\left(\hat{\beta}_1\right) + \mathrm{var}\,(\bar{u}) + 2\bar{x}\mathrm{cov}\left(\hat{\beta}_1, \bar{u}\right)
\end{aligned}$$

Or

$$
\begin{aligned}
\operatorname{cov}\left(\hat{\beta}_1, \bar{u}\right) &= E(\left(\hat{\beta}_1 - \beta_1\right)\bar{u}) \\
&= E\left(\left(\frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sum_{i=1}^{N}(x_i-\bar{x})u_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}u_i\right)\right) \\
&= \frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2}E\left(\frac{1}{N}\sum_{j=1}^{N}u_j\sum_{i=1}^{N}(x_i-\bar{x})u_i\right) \\
&= \frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2}E\left(\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{N}(x_i-\bar{x})u_iu_j\right) \\
&= 0
\end{aligned}
$$

Donc:

$$
\begin{aligned}
\operatorname{var}(\hat{\beta}_0) &= \bar{x}^2\operatorname{var}\left(\hat{\beta}_1\right) + \operatorname{var}(\bar{u}) \\
\operatorname{var}(\hat{\beta}_0) &= \bar{x}^2\frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} + \frac{\sigma^2}{N}
\end{aligned}
$$

Or

$$
\sum_{i=1}^{N}(x_i-\bar{x})^2 = \sum_{i=1}^{N}x_i^2 - N\bar{x}^2
$$

Donc:

$$
\begin{aligned}
\operatorname{var}(\hat{\beta}_0) &= \left(\frac{\sum_{i=1}^{N}x_i^2 - \sum_{i=1}^{N}(x_i-\bar{x})^2}{N}\right)\frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} + \frac{\sigma^2}{N} \\
&= \frac{\sigma^2}{N\sum_{i=1}^{N}(x_i-\bar{x})^2}\sum_{i=1}^{N}x_i^2
\end{aligned}
$$

Puisque les formules de la variance des estimateurs dépendent de σ^2, on a besoin d'un estimateur de la variance des termes aléatoires. Remarquons, préalablement, que les résidus sont des 'approximations' des termes aléatoires. En effet,

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + u_i && \text{(erreur)} \\ y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i && \text{(résidu)} \end{aligned}$$

et donc,

$$\begin{aligned} \hat{u}_i &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= (\beta_0 + \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i. \end{aligned}$$

On a deux estimateurs possibles, mais seul le second est non biaisé.

Estimateur 1: Un estimateur naturel est le suivant:

$$\tilde{\sigma}^2 = N^{-1} \sum_{i=1}^{N} \hat{u}_i^2 = SSR/N$$

mais cet estimateur sera biaisé car les résidus doivent satisfaire des contraintes:

$$\sum_{i=1}^{N} \hat{u}_i = 0, \sum_{i=1}^{N} x_i \hat{u}_i = 0$$

Estimateur 2: Un second estimateur est le suivant:

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^{N} \hat{u}_i^2 = \frac{SSR}{N-2}.$$

Celui-ci est non biaisé comme le montre le théorème suivant.

Théorème 3 (absence de biais de $\hat{\sigma}^2$): Sous les hypothèses $1'$ à $5'$,

$$E(\hat{\sigma}^2) = \sigma^2$$

Démonstration. Les résidus sont définis par

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i.$$

Or:

$$\sum_{i=1}^{n} \hat{u}_i = 0 = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}$$

en vertu des propriétés algébriques des MCO (la somme des résidus est nulle), où $\bar{u} = \sum_{i=1}^{n} u_i$ et $\bar{x} = \sum_{i=1}^{n} x_i$. Donc, en soustrayant la seconde expression à la première, on obtient:

$$\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x}).$$

En prenant le carré des membres de droite et gauche, on obtient:

$$\hat{u}_i^2 = (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2 (x_i - \bar{x}) - 2(\hat{\beta}_1 - \beta_1)(u_i - \bar{u})(x_i - \bar{x}).$$

En sommant sur i, cette expression devient:

$$\begin{aligned} \sum_{i=1}^{n} \hat{u}_i^2 &= \sum_{i=1}^{n} (u_i - \bar{u})^2 + \sum_{i=1}^{n} (\hat{\beta}_1 - \beta_1)^2 (x_i - \bar{x})^2 \\ &\quad -2 \sum_{i=1}^{n} (\hat{\beta}_1 - \beta_1)(u_i - \bar{u})(x_i - \bar{x}). \end{aligned}$$

Et en prenant l'espérance,

$$E\left(\sum_{i=1}^{n} \hat{u}_i^2\right) = A + B + C.$$

avec

$$\begin{aligned} A &= E\left(\sum_{i=1}^{n} (u_i - \bar{u})^2\right), \\ B &= E\left(\sum_{i=1}^{n} (\hat{\beta}_1 - \beta_1)^2 (x_i - \bar{x})^2\right), \\ C &= -2E\left(\sum_{i=1}^{n} (\hat{\beta}_1 - \beta_1)(u_i - \bar{u})(x_i - \bar{x})\right) \end{aligned}$$

Enfin, par un résultat bien connu de statistique, on a:

$$A = (n-1)\sigma^2.$$

De plus:

$$\begin{aligned}
B &= \sum_{i=1}^{n} (x_i - \bar{x})^2 E\left((\hat{\beta}_1 - \beta_1)^2\right) \\
&= \sum_{i=1}^{n} (x_i - \bar{x})^2 E\left(\left(\frac{\sum_{i=1}^{n} (x_i - \bar{x}) u_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)^2\right) \\
&= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^2} E\left(\left(\sum_{i=1}^{n} (x_i - \bar{x}) u_i\right)^2\right) \\
&= \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 E\left(u_i^2\right) \\
&= \sigma^2
\end{aligned}$$

où l'avant dernière ligne utilise le fait que $E\left((x_i - \bar{x})(x_j - \bar{x}) u_i u_j\right) = 0$.

Finalement,

$$\begin{aligned}
C &= -2E\left(\sum_{i=1}^{n} \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)(u_i - \bar{u})(x_i - \bar{x})\right) \\
&= \frac{-2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} E\left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - \bar{x})(u_i - \bar{u})(u_j - \bar{u})(x_j - \bar{x})\right) \\
&= \frac{-2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} E\left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - \bar{x})(u_i - \bar{u})(u_j - \bar{u})(x_j - \bar{x})\right) \\
&= -2\sigma^2
\end{aligned}$$

Et en fin de compte,

$$E\left(\sum_{i=1}^{n}\hat{u}_i^2\right) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2.$$

3 Modèle de régression linéaire multiple 1: Définition et calcul

3.1 Motivation

Exemples de modèles de régression multiple:

Exemple 1.

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + \beta_2 \cdot \text{EXPER} + u$$

Exemple 2.

$$\text{SCORE\_MOY} = \beta_0 + \beta_1 \cdot \text{DEP} + \beta_2 \cdot \text{REV\_MOY} + u$$

Exemple 3.

$$\text{CONS} = \beta_0 + \beta_1 \cdot \text{REV} + \beta_2 \cdot \text{REV}^2 + u$$

Dans ce cas,

$$\frac{\Delta\text{CONS}}{\Delta\text{REV}} = \beta_1 + 2\beta_2 \cdot \text{REV}$$

Exemple 4.

$$\ln(\text{SAL}) = \beta_0 + \beta_1 \cdot \ln(\text{VENTE}) + \beta_2 \cdot \text{ceoten} + \beta_2 \cdot \text{ceoten}^2 + u$$

(linéaire dans les paramètres)

3.2 Mécanique et interprétation des MCO

Les estimateurs des moindres carrés ordinaires sont obtenus par la minimisation du carré des résidus, c'est-à-dire:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} \ldots - \hat{\beta}_K x_{Ki})^2$$

Comme précédemment la valeur prédite est donnée par

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_K x_{Ki}$$

et le résidu par

$$\hat{u}_i = y_i - \hat{y}_i.$$

Dans le cas de deux variables explicatives, la droite de régression est donnée par:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

où $\hat{\beta}_0$ est l'estimation de y quand $x_1 = 0$ et $x_2 = 0$, $\Delta\hat{y} = \hat{\beta}_1 \Delta x_1$ pour x_2 fixé, $\Delta\hat{y} = \hat{\beta}_2 \Delta x_2$ pour x_1 fixé. L'intérêt des MCO est de fournir des interprétations ceteris paribus même si les données n'ont pas été collectées de manière adéquates.

Exemple: Considérons la droite de régression suivante:

$$\widehat{\log(\mathrm{SAL})} = 0.284 + 0.092 \cdot \mathrm{EDUC} + 0.0041 \cdot \mathrm{EXPER} + 0.022 \cdot \mathrm{ANC}$$

Une année d'éducation supplémentaire, toutes autres choses étant égales, représentera un accroissement de salaire de 9%. Si plusieurs variables indépendantes se modifient simultanément, les effets se cummulent. Par exemple, une année d'expérience et d'ancienneté impliquera un accroissement de salaire de 26%:

$$\Delta \widehat{\log(\mathrm{SAL})} = 0.0041 \cdot \Delta\mathrm{EXPER} + 0.022 \cdot \Delta\mathrm{ANC} = 0.261$$

3.3 Propriétés des résidus et mesures de l'ajustement

Les estimateurs des MCO et les résidus possèdent les propriétés algébriques suivantes:

(1) $\sum \hat{u}_i = 0$;

(2) $\sum \hat{u}_i x_{ki} = 0$ pour $k = 1, \ldots, K$;

(3) $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \ldots + \hat{\beta}_K \bar{x}_K$;

(4) $\sum \hat{u}_i \hat{y}_i = 0$;

(5) Si l'on définit

$$\begin{aligned} \text{SST} &= \sum_{i=1}^{N} (y_i - \bar{y})^2, \\ \text{SSE} &= \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2, \\ \text{SSR} &= \sum_{i=1}^{N} (\hat{u}_i)^2, \end{aligned}$$

alors

$$\text{SST} = \text{SSE} + \text{SSR}.$$

Le coefficient de détermination est alors défini par

$$\begin{aligned} R^2 &= \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}} \\ &= \frac{\left(\widehat{\text{cov}\,(y_i, \hat{y}_i)}\right)^2}{\widehat{\text{var}\,(y_i)}\widehat{\text{var}\,(\hat{y}_i)}} \end{aligned}$$

c'est-à-dire le carré du coefficient de corrélation entre y_i et $\bar{y}$.

Remarque Le R^2 ne décroit jamais lorsque une ou plusieurs variables sont ajoutées.

3.4 Calcul des estimateurs des MCO: une formule utile

Pour simplifier, supposons qu'il n'y a que deux variables explicatives (les résultats qui suivent sont valables dans le cas plus général de K variables explicatives):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

Dans ce cas, l'estimateur de β_1 est égal à:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \hat{r}_{1i} y_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}$$

où $\hat{r}_{1i}$ est le résidu de la régression de x_1 sur x_2, c'est-à-dire:

$$\hat{r}_{1i} = x_{1i} - \hat{x}_{1i} = x_{1i} - \hat{\delta}_0 - \hat{\delta}_1 x_{2i}.$$

Donc: l'estimateur de β_1 est obtenu par la régression de y sur $\hat{r}_1$. Cette formulation a une interprétation particulière. En effet, le résidu $\hat{r}_{1i}$ est la part de la variable x_{1i} qui n'est pas corrélée avec x_{2i} ou, en d'autres termes, le résidu $\hat{r}_{1i}$ correspond à la variable x_{1i} dont on aurait enlevé l'effet de x_{2i}.

Ce résultat est obtenu de la manière suivante. Soit la régression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Les conditions de premier ordre donnent:

$$\sum_{i=1}^{n} x_{1i}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0.$$

Si l'on remplace x_{1i} par $\hat{\delta}_0 + \hat{\delta}_1 x_{2i} + \hat{r}_{1i}$, l'on obtient:

$$\sum_{i=1}^{n} (\hat{\delta}_0 + \hat{\delta}_1 x_{2i} + \hat{r}_{1i})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0.$$

Puisque $y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} = \hat{u}_i$ est un résidu, l'expression ci-dessus se simplifie de la manière suivante:

$$\sum_{i=1}^{n} \hat{r}_{1i}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0.$$

En utilisant les propriétés des résidus $\sum_{i=1}^{n} \hat{r}_{1i} = 0$ et $\sum_{i=1}^{n} x_{2i}\hat{r}_{1i} = 0$, on obtient:

$$\sum_{i=1}^{n} \hat{r}_{1i}(y_i - \hat{\beta}_1 x_{1i}) = 0.$$

En remplaçant à nouveau x_{1i} par $\hat{x}_{1i} + \hat{r}_{1i}$, et en utilisant la propriété des résidus $\sum_{i=1}^n \hat{r}_{1i}\hat{x}_{1i} = 0$, cela devient:

$$\sum_{i=1}^{n} \hat{r}_{1i}(y_i - \hat{\beta}_1 (\hat{x}_{1i} + \hat{r}_{1i})) = 0,$$

ou encore,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{1i} y_i}{\sum_{i=1}^n \hat{r}_{1i}\hat{r}_{1i}}.$$

4 Modèle de régression linéaire multiple 2: Espérance et variance des estimateurs

4.1 L'espérance des estimateurs des MCO

Hypothèse 1 (linéarité dans les paramètres): Le modèle dans la population peut se décrire par une relation linéaire à K variables explicatives telle que:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + u$$

où $\beta_0, \beta_1, \ldots, \beta_K$ sont des paramètres, et u est un terme aléatoire.

Hypothèse 2 (échantillonnage aléatoire): Un échantillon aléatoire de N observations, $\{(y_i, x_{1i}, \ldots, x_{Ki}) : i = 1, \ldots, N\}$ issu du modèle de population décrit en H1.

Hypothèse 3 (espérance conditionnelle égale à zero): Le terme u a une espérance de zéro pour toute valeur des variables indépendantes. En d'autres termes,

$$E(u|x_1, \ldots, x_K) = 0.$$

Hypothèse 4 (absence de collinéarité parfaite): Dans l'échantillon (et donc dans la population), aucune des variables indépendantes n'est constante, et il n'y a pas de relations linéaires exactes entre les variables indépendantes.

Remarque:

(1) Selon une autre interprétation de cette hypothèse, si l'on régresse une variable explicative quelconque sur l'ensemble des autres variables explicatives, le R^2 doit être inférieur à 1.

(2) Les variables explicatives peuvent être corrélées mais elles ne peuvent pas

être 'parfaitement' corrélées. Par exemple:

$$\mathrm{SCORE\_MOY} = \beta_0 + \beta_1 \cdot \mathrm{DEP} + \beta_2 \cdot \mathrm{REV\_MOY} + u,$$

ou

$$\mathrm{CONS} = \beta_0 + \beta_1 \cdot \mathrm{REV} + \beta_2 \cdot \mathrm{REV}^2 + u.$$

(3) Cas particuliers où il y a une corrélation parfaite:
- Une variables est le multiple d'une autre (les unités de mesure sont différentes) ou bien:

$$\log(\mathrm{CONS}) = \beta_0 + \beta_1 \cdot \log(\mathrm{REV}) + \beta_2 \cdot \log(\mathrm{REV}^2) + u$$

- Une variable est la somme de deux autres:

$$\mathrm{VOTE\_A} = \beta_0 + \beta_1 \cdot \mathrm{DEP\_A} + \beta_2 \cdot \mathrm{DEP\_B} + \beta_3 \cdot \mathrm{DEP\_TOT} + u$$

- La taille de l'échantillon est trop petite:.$N < K + 1$.

Les démonstrations qui vont suivre reposent sur le résultat suivant:

$$\begin{aligned}
\hat{\beta}_j &= \frac{\sum_{i=1}^N \hat{r}_{ji} y_i}{\sum_{i=1}^N \hat{r}_{ji}^2} \\
&= \frac{\sum_{i=1}^N \hat{r}_{ji} \left(\beta_0 + \beta_j x_{ji} + u_i\right)}{\sum_{i=1}^N \hat{r}_{ji}^2} \\
&= \beta_0 \frac{\sum_{i=1}^N \hat{r}_{ji}}{\sum_{i=1}^N \hat{r}_{ji}^2} + \beta_j \frac{\sum_{i=1}^N \hat{r}_{ji} \left(\hat{x}_{ji} + \hat{r}_{ji}\right)}{\sum_{i=1}^N \hat{r}_{ji}^2} + \frac{\sum_{i=1}^N \hat{r}_{ji} u_i}{\sum_{i=1}^N \hat{r}_{ji}^2} \\
&= \beta_j \left(\frac{\sum_{i=1}^N \hat{r}_{ji} \hat{x}_{ji}}{\sum_{i=1}^N \hat{r}_{ji}^2} + \frac{\sum_{i=1}^N \hat{r}_{ji}^2}{\sum_{i=1}^N \hat{r}_{ji}^2} \right) + \frac{\sum_{i=1}^N \hat{r}_{ji} u_i}{\sum_{i=1}^N \hat{r}_{ji}^2} \\
&= \beta_j + \frac{\sum_{i=1}^N \hat{r}_{ji} u_i}{\sum_{i=1}^N \hat{r}_{ji}^2}
\end{aligned}$$

Cette formule exprime le lien entre les estimateurs d'une part et les paramètres et les résidus d'autre part.

Theoreme 1 (absence de biais des estimateurs): Sous les hypothèses

1 à 4, les estimateurs des MCO sont non biaisés:

$$E(\hat{\beta}_j) = \beta_j$$

Démonstration. Les estimateurs et les paramètres dans la population sont liés par la formule suivante:

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^{N} \hat{r}_{ji} u_i}{\sum_{i=1}^{N} \hat{r}_{ji}^2}$$

Si l'on prend l'espérance et que l'on simplifie, l'on obtient:

$$\begin{aligned} E(\hat{\beta}_j) &= \beta_j + E\left(\frac{\sum_{i=1}^{N} \hat{r}_{ji} u_i}{\sum_{i=1}^{N} \hat{r}_{ji}^2}\right) \\ &= \beta_j + \frac{\sum_{i=1}^{N} \hat{r}_{ji} E(u_i)}{\sum_{i=1}^{N} \hat{r}_{ji}^2} \\ &= \beta_j \end{aligned}$$

Remarques:

(1) (Inclusion de variables non-pertinentes) Si l'on estime:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 + \hat{\beta}_3 \cdot x_3$$

alors que le vrai modèle est:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u.$$

Cela n'a aucun effet en termes de biais mais cela peut avoir des effets en termes de variance.

(2) (Exclusion de variables pertinentes: le cas simple) Si l'on estime:

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot x_1$$

alors que le 'vrai' modèle est

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u$$

les estimateurs seront, en général, biaisés (mauvaise spécification).
Example:

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + \beta_2 \cdot \text{HABIL} + u$$

Puisque abil n'est pas observable, on estime:

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + v$$

où $v = (\beta_2 \cdot \text{abil} + u) - \beta_2 \cdot E(\text{abil})$. Le problème est que la condition sur les espérance conditionnelle risque de ne pas être satisfaite.
L'estimateur de β_1 est donné par:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x})y_i}{\sum_{i=1}^{N}(x_{1i} - \bar{x})^2}.$$

Or:

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + u_i$$

L'introduction de cette équation dans la précédente donne:

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)\left(\beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + u_i\right)}{\sum_{i=1}^{N}(x_{1i} - \bar{x})^2} \\
&= \beta_0 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2} + \beta_1 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1) \cdot x_{1i}}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2} \\
&\quad + \beta_2 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1) \cdot x_{2i}}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2} + \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)u_i}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2} \\
&= \beta_1 + \beta_2 \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot x_{2i}}{\sum_{i=1}^{N}(x_i - \bar{x})^2} + \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)u_i}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2}
\end{aligned}$$

Si on prend les espérances:

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1) \cdot x_{2i}}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2}$$

où le dernier terme est la pente de la régression de x_{2i} sur x_{1i}:

$$x_2 = \hat{\delta}_0 + \hat{\delta}_1 x_1$$

Donc:

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \hat{\delta}_1$$

Les cas où il y a absence de biais sont lorsque x_2 et x_1 sont non corrélés dans l'échantillon et lorsque β_2 est égal à zéro.
Le biais est positif ($E\left(\tilde{\beta}_1\right) - \beta_1 > 0$) si signe($\beta_2$) = signe($\tilde{\delta}_1$).
Le biais est négatif ($E\left(\tilde{\beta}_1\right) - \beta_1 < 0$) si signe($\beta_2$) $\neq$ signe($\tilde{\delta}_1$).
Exemple:

$$\widehat{\log(\mathrm{SAL})} = 0.584 + 0.083 \cdot \mathrm{EDUC}$$

où la variable abil a été omise.

(3) (Exclusion de variables pertinentes: le cas général)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

On estime:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Supposons que:

- x_1 et x_2 sont non corrélés
- x_2 et x_3 sont non corrélés

Alors:

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1) \cdot x_{3i}}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2}$$

4.2 La variance des estimateurs

Hypothèse 5 (homoscédasticité): La variance du terme aléatoire conditionnelle à $x_1, ..., x_K$ est constante. En d'autres termes,

$$var(u|x_1, ..., x_K) = \sigma^2$$

Remarque: Les hypothèses 1 à 5 sont amppelées 'hypothèses de Gauss-Markov'.

Theoreme 2 (variance des pentes des MCO): Sous les hypothèses 1 à 5,

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\left(1 - R_j^2\right) \sum_{i=1}^{N} (x_{ji} - \bar{x}_j)^2}$$

où R_j^2 est le coefficient de détermination de la régression de x_j sur toutes les autres variables.

Démonstration. Les estimateurs et les paramètres dans la population sont liés par la formule suivante:

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^{N} \hat{r}_{ji} u_i}{\sum_{i=1}^{N} \hat{r}_{ji}^2}$$

Si l'on prend la variance de cette expression et que l'on simplifie, l'on obtient:

$$\text{var}(\hat{\beta}_j) = \text{var}\left(\frac{\sum_{i=1}^{N} \hat{r}_{ji} u_i}{\sum_{i=1}^{N} \hat{r}_{ji}^2}\right) = \frac{\sum_{i=1}^{N} \hat{r}_{ji}^2 \text{var}(u_i)}{\left(\sum_{i=1}^{N} \hat{r}_{ji}^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{N} \hat{r}_{ji}^2}.$$

En utilisant la définition du R_j^2 =SSE/SST, on obtient l'expression du théorème.

Remarque:

(1) Les composantes de la variance des estimateurs des moindres carrés ordinaires sont les suivantes:
   - La variance du terme aléatoire σ^2;
   - La variance des variables explicatives et le nombre d'observations (le problème de micronumérosité): $\sum_{i=1}^{N} (x_{ji} - \bar{x}_j)^2$;
   - La collinéarité entre les variables explicatives (problème de multicollinéarité): $1 - R_j^2$.

(2) La variance de $\hat{\beta}_j$ peut également s'écrire:

$$\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{N} \hat{r}_{ji}^2}$$

où $\hat{r}_{ji}$ est le résidu de la régression de la variable x_i sur l'ensemble des autres variables explicatives. Cette formulation sera également utilisée dans les résultats qui suivent.

(3) Comme dans le cas du modèle de régression simple, la variance du terme aléatoire doit être estimé.

Theoreme 3 (absence de biais de l'estimateurs de σ^2): Sous les hypothèses 1 à 5,

$$E(\hat{\sigma}^2) = E\left(\frac{\sum \hat{u}_i^2}{N-K-1}\right) = \sigma^2$$

Ce théorème n'est pas démontré. Il implique que l'estimateur de l'écart-type de la régression est:

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{N-K-1}}$$

où $N-K-1$ est le nombre de degrés de liberté (nombre d'observations moins nombre de paramètres à estimer) et que l'écart-type de $\hat{\beta}_j$ est:

$$\frac{\hat{\sigma}}{\sqrt{\left(1-R_j^2\right)\sum_{i=1}^{N}(x_{ji}-\bar{x}_j)^2}}.$$

4.2.1 Efficiacité des estimateurs des MCO

Il existe une infinité d'estimateurs sans biais, mais l'estimateur des MCO a une propriété très attractive.

Theoreme 4 (Théorème de Gauss-Markov): Sous les hypothèses 1-5, les estimateurs des MCO sont les meilleurs estimateurs linéaires non biaisés de $\beta_0, \beta_1, \ldots, \beta_K$.

Remarque:

(1) Un estimateur est linéaire s'il peut s'écrire sous la forme d'une fonction linéaire des données sur la variable dépendante:

$$\tilde{\beta}_j = \sum_{i=1}^{N} w_{ij} y_i$$

où chaque w_{ij} peut être une fonction de toutes les variables indépendantes. Or, l'estimateur des MCO est linéaire puisque

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \hat{r}_{1i} y_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}$$

(2) Un estimateur est "meilleur" q'un autre estimateur si sa variance est plus petite que celle de cet autre.

(3) Les estimateurs, qui ont la propriété décrite dans le Théorème 4, sont dits 'BLUE' (= Best Linear Unbiased Estimator).

5 Modèle de régression linéaire multiple 3: Inférence

5.1 Echantillonnage des estimateurs des MCO

On ne connaît que deux moments de la distribution des estimateurs (espérance et variance). Pour connaître les autres moments, on a besoin de l'hypothèse suivante:

Hypothèse 6 (normalité): Les termes aléatoires u sont sont distribués selon une loi normale.

Remarque:

(1) La normalité est justifiée par le Théorème central-limite, selon lequel la somme d'un très grand nombre N de variables aléatoires (divisée par $\sqrt{N}$) suit approximativement une loi normale..

(2) Si l'on ajoute les hypothèses 3 et 5 à l'hypothèse 6, on obtient que

$$u \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

(3) Les hypothèses 1 à 6 constituent les **hypothèses classiques** du modèle linéaire. Elles impliquent une forme plus forte du théorème de Gauss-Markov.

Theoreme 5 (Théorème de Gauss-Markov 'non linéaire'): Sous les hypothèses 1-6, les estimateurs des MCO sont les meilleurs estimateurs non biaisés de $\beta_0, \beta_1, \ldots, \beta_K$.

Remarque: Les estimateurs des MCO sont donc les meilleurs y compris dans la classe des estimateurs non linéaires.

Theoreme 6 (normalité des estimateurs): Sous les hypothèses 1-6, la distribution des estimateurs des MCO est normale

$$\hat{\beta}_j \rightsquigarrow \mathcal{N}(\beta_j, \mathrm{var}\left(\beta_j\right))$$

où $\mathrm{var}\left(\beta_j\right)$ est donné par

$$\frac{\sigma^2}{\left(1-R_j^2\right)\sum_{i=1}^N (x_{ji}-\bar{x}_j)^2}.$$

En particulier, $\left(\hat{\beta}_j - \beta_j\right)/\sqrt{\mathrm{var}(\hat{\beta}_j)} \rightsquigarrow \mathcal{N}(0,1)$.

Démonstration (intuition): L'estimateur des MCO est linéaire car il peut s'écrire sous la forme:

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^N \hat{r}_{ji} u_i}{\sum_{i=1}^N \hat{r}_{ji}^2}$$

Donc, $\hat{\beta}_j$ est une combinaison linéaire des termes aléatoires qui suivent une loi normale.

5.2 Test d'une seule restriction: le t-test

Le test de Student repose sur le théorème suivant.

Theoreme 7 (t-distribution des estimateurs standardisés): Sous les hypothèses 1-6, les estimateurs standardisés suivent une loi de Student:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}}} \rightsquigarrow \mathcal{T}(N-K-1)$$

où $K+1$ est le nombre de paramètres estimés, et

$$\sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}} = \frac{\hat{\sigma}}{\sqrt{\left(1-R_j^2\right)\sum_{i=1}^N (x_{ji}-\bar{x}_j)^2}}.$$

5.2.1 Test unilatéral.

Le théorème 1 permet de tester l'hypothèse nulle:

$$\mathrm{H}_0 : \beta_j = 0$$

contre l'hypothèse alternative:

$$\mathrm{H}_1 : \beta_j > 0 \text{ (test unilatéral à droite).}$$

Par exemple, considérons le modèle:

$$\log(\mathrm{SAL}) = \beta_0 + \beta_1 \cdot \mathrm{EDUC} + \beta_2 \cdot \mathrm{EXPER} + \beta_3 \cdot \mathrm{ANC} + u.$$

On peut vouloir tester l'hypothèse nulle $\beta_2 = 0$ (l'expérience n'a pas d'effet sur le salaire) contre l'hypothèse alternative $\beta_2 > 0$ (l'expérience a un effet positif sur le salaire).
Dans ce cas, le test de student est basé sur la statistique t suivante:

$$t = \hat{\beta}_j / \sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}}.$$

De manière intuitive, on sera d'autant plus tenté de rejeter l'hypothèse nulle que la valeur de cette statistique est grande. Donc, pour un test unilatéral à droite, la règle de rejet est égale à

$$t > c$$

où c est la valeur critique. La valeur critique est déterminée par le seuil de signification choisi par l'économètre, c'est-à-dire la probabilité de rejeter l'hypothèse nulle alors que cette dernière est correcte. Si l'on choisit un seuil de signification de 5% — le choix le plus courant —. la valeur critique c est la valeur de la loi de Student telle que la probabilité d'obtenir une valeur supérieure à c est égale à 5%, c'est-à-dire

$$\Pr(t > c | \mathrm{H}_0 \text{ est vraie}) = 0.05$$

Donc, si la statistique de Student t est supérieure à c, cela signifie que la probabilité que la distribution de t soit une loi de Student est faible.

Exemple 1: Considérons la régression du logarithme du salaire sur l'éducation, l'expérience et l'ancienneté et testons l'hypothèse que l'ancienneté n'affecte pas la variable dépendante.

$$\begin{aligned} \widehat{\ln(\text{SAL})} &= \underset{(0.104)}{0.284} + \underset{(0.007)}{0.092} \cdot \text{EDUC} + \underset{(0.0017)}{0.0041} \cdot \text{EXPER} + \underset{(0.003)}{0.022} \cdot \text{ANC} \\ N &= 526,\ R^2 = 0.316,\ t_{\beta_{exper}} = 0.0041/0.0017 \approx 2.41 \end{aligned}$$

Exemple 2: Considérons la régression du pourcentage de réussite au test MEAP (Michigan Educational Assessment Program) sur la rémunération moyenne des enseignants, le nombre d'enseignements et le nombre d'étudiants et testons l'hypothèse que le nombre d'étudiants n'affecte sur la réussite moyenne au test MEAP.

$$\begin{aligned} \widehat{\text{MEAP}} &= \underset{(6.113)}{2.274} + \underset{(0.00010)}{0.00046} \cdot \text{COMP\_TOT} + \underset{(0.040)}{0.048} \cdot \text{NB\_PERS} \\ &\quad -\underset{(0.00022)}{0.00020} \cdot \text{NB\_ETUD} \\ N &= 408,\ R^2 = 0.0541,\ t_{\beta_{\text{enroll}}} = -0.0002/0.00022 \approx -0.91 \\ \widehat{\text{MEAP}} &= -\underset{(48.70)}{207.66} + \underset{(4.06)}{21.16} \ln(\text{COMP\_TOT}) + \underset{(4.19)}{3.98} \ln(\text{NB\_PERS}) \\ &\quad -\underset{(0.69)}{1.29} \ln(\text{NB\_ETUD}) \\ N &= 408,\ R^2 = 0.0651,\ t_{\beta_{\ln(\text{enroll})}} = -1.29/0.69 \approx 1.87 \end{aligned}$$

5.2.2 Test bilatéral.

Le théorème 1 permet également de tester l'hypothèse nulle:

$$\text{H}_0 : \beta_j = 0$$

contre l'hypothèse alternative:

$$\mathrm{H}_1 : \beta_j \neq 0 \text{ (test bilatéral).}$$

Pour un test bilatéral, la règle de rejet est égale à

$$|t| > c$$

où c est la valeur critique. Autrement dit, l'hypothèse nulle sera rejetée si t est très grand ou si t est très petit. La valeur critique c est également déterminée par le seuil de signification (choisi de manière arbitraire). Formellement,

$$\Pr(|t| > c|\mathrm{H}_0 \text{ est vraie}) = 0.05.$$

Soulignons que, puisque la loi de Student est symétrique, $\Pr(|t| > c|\mathrm{H}_0$ est vraie$) = 2 \times \Pr(t > c|\mathrm{H}_0$ est vraie$)$.

5.2.3 Tests généraux

La formule la plus générale de la statistique de Student,

$$t = \left(\hat{\beta}_j - \beta_j\right) / \sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}}$$

permet de tester d'autres hypothèses nulles telles que:

$$\mathrm{H}_0 : \beta_j = b_j$$

contre l'hypothèse alternative:

$$\mathrm{H}_1 : \beta_j \neq b_j \text{ (test bilatéral).}$$

La procédure est exactement la même que celle qui vient d'être décrite.

Exemple 3: Considérons la régression du logarithme du nombre de crimes et délits commis sur les campus universitaire sur le nombre d'étudiants et testons l'hypothèse que nombre de crimes augmente proportionnellement avec le nombre d'étudiants:

$$\begin{aligned}\widehat{\ln(\text{CRIME})} &= \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27}\ln(\text{NB\_ETUD}) \\ N &= 97,\ R^2 = 0.585,\ t_{\beta_{\ln(\text{enrol})}} = (1.27-1)/0.11 \approx 2.454\end{aligned}$$

5.2.4 Tests impliquant plusieurs paramètres.

Le test de Student permet également de tester un restriction portant sur une combinaison linéaire de paramètres. Pour cela, il faut reparametrer le modèle.

Exemple 4: On désire tester si le taux de rendement d'une année passée dans un "junior college" est différent de celui d'une annéepassée dans un "college" (ou "university"). Le modèle à estimer est

$$\ln(\text{SAL}) = \beta_0 + \beta_1 \cdot \text{COLL} + \beta_2 \cdot \text{UNIV} + \beta_3 \cdot \text{EXPER} + u$$

et il faut tester

$$\text{H}_0 : \beta_1 = \beta_2.$$

Le modèle estimé est:

$$\begin{aligned}\widehat{\ln(\text{SAL})} &= \underset{(0.27)}{1.43} + \underset{(0.031)}{0.098}\cdot\text{COLL} + \underset{(0.035)}{0.124}\cdot\text{UNIV} + \underset{(0.008)}{0.019}\cdot\text{EXPER} \\ N &= 285, \qquad R^2 = 0.243\end{aligned}$$

On peut procéder en calculant la différence $\hat{\beta}_1 - \hat{\beta}_2$ et la variance de $\widehat{\text{var}(\hat{\beta}_1 - \hat{\beta}_2)}$. Toutefois il est plus simple de procéder à en changement de paramétrisation. On définit:

$$\theta = \beta_1 - \beta_2$$

et en substituant ce nouveau paramètre dans le modèle, on obtient:

$$\ln(\text{SAL}) = \beta_0 + \theta \cdot \text{COLL} + \beta_2 \cdot (\text{COLL} + \text{UNIV}) + \beta_3 \cdot \text{EXPER} + u$$

L'estimation de ce modèle donne:

$$\widehat{\ln(\text{SAL})} = \underset{(0.27)}{1.43} - \underset{(0.018)}{0.026}\cdot\text{COLL} + \underset{(0.035)}{0.124}\cdot\text{SUP\_TOT} + \underset{(0.008)}{0.019}\cdot\text{EXPER}$$

Le test de student est effectué à l'aide la statistique $t = 0.026/0.018$.

Exemple 5: La théorie économique montre que les demandes doivent être homogène. Cela signifie que:

$$q_1 = f(p_1, p_2, y) = f(tp_1, tp_2, ty).$$

Si l'on veut tester cette propriété, il faut construire un modèle économétrique et dériver les implications de la contraite d'homogénéité. On choisit:

$$\ln q_1 = \beta_0 + \beta_1 \cdot \ln p_1 + \beta_2 \cdot \ln p_2 + \beta_3 \cdot \ln y + u$$

La propriété d'homogénéité implique une restriction sur les paramètres. En effet, l'équation

$$\beta_0 + \beta_1 \cdot \ln tp_1 + \beta_2 \cdot \ln tp_2 + \beta_3 \cdot \ln ty + u$$

sera égale à

$$\beta_0 + \beta_1 \cdot \ln p_1 + \beta_2 \cdot \ln p_2 + \beta_3 \cdot \ln y + u$$

pour toute valeur de t si et seulement si

$$\beta_1 + \beta_2 + \beta_3 = 0.$$

On définit donc: $\theta = \beta_1 + \beta_2 + \beta_3$ et on effectue le changement de parametrisation comme précédemment afin de faire le test.

5.2.5 Tests et intervalles de confiance

En vertu du Théorème 6, l'inégalité suivante

$$-c < \frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\mathrm{var}}(\hat{\beta}_j)}} < c$$

où c est le 97.5 percentile de la loi de Student à $N - K - 1$ degrés de liberté, est satisfaite dans 95% des cas. Donc, l'intervalle de confiance est donné par:

$$\hat{\beta}_j - c \times \sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}} < \beta_j < \hat{\beta}_j + c \times \sqrt{\widehat{\mathrm{var}(\hat{\beta}_j)}}.$$

Exemple 6: Considérons la régression du logarithme du prix des maisons sur le logarithme de la superficie de celles-ci, le nombre de chambres et le nombre de salles de bain:

$$\begin{aligned}\widehat{\log(\mathrm{PRIX})} &= \underset{(1.15)}{7.46} + \underset{(0.184)}{0.634} \cdot \log(\mathrm{PIEDS\_CARRE}) \\ &\quad -\underset{(0.059)}{0.066} \cdot \mathrm{CH\_COUCH} + \underset{(0.075)}{0.158} \cdot \mathrm{SAL\_BAINS} \\ N &= 19, \qquad R^2 = 0.806\end{aligned}$$

et calculons l'intervalle de confiance à 95% de l'estimation de l'effet de la superficie:

$$0.634 + 2.131 \times 0.184 < \beta_j < 0.634 - 2.131 \times 0.184$$

5.3 Tests d'une mutiplicité de restrictions: le F-test

Le test de Fisher s'appuie sur le résultat suivant.

Theoreme 7 (F-distribution des rapports de SSR): Sous les hypothèses 1-6,

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{nr})/Q}{(\mathrm{SSR}_{nr})/(N-K-1)} \rightsquigarrow \mathcal{F}(Q, N-K-1)$$

où SSR_r est la somme des carrés résiduels du modèle restreint, SSR_{nr} la somme des carrés résiduels du modèle non restreint, q le nombre de restrictions, N le nombre d'observations, et K le nombre de variables explicatives.

Remarques:

(1) Le test de Fisher appliqué à une restriction unique est parfaitement valable. Dans ce cas, il donne exactement le même résultat que le test de Student.

En effet, on peut monter que la statistique de Fisher pour une seule restriction est égale au carré de la statistique de Student. Or, la distribution de Fisher à $(1, N-K-1)$ degrés de liberté est égale à la distribution du carré d'une variable qui suit une loi de Student à $(N-K-1)$.

(2) Le test de Fisher sous la forme de R^2 est très pratique. Soulignons que SSR_r =SST$\times(1-R_r^2)$ et SSR$_{nr}$ =SST$\times(1-R_{nr}^2)$. Donc:

$$F=\frac{(\mathrm{SSR}_r-\mathrm{SSR}_{nr})/q}{(\mathrm{SSR}_{nr})/(N-K-1)}=\frac{(R_{ur}^2-R_r^2)/q}{(1-R_{ur}^2)/(N-K-1)}.$$

(3) Le test de Fisher appliqué à l'ensemble des paramètres. Dans ce cas, l'hypothèse nulle s'écrit:

$$H_0:\beta_1=0,\beta_2=0,\ldots,\beta_K=0$$

et la statistique de Fisher devient:

$$F=\frac{R^2/K}{(1-R^2)/(N-K-1)}.$$

Exemple 7: Le salaire des houeurs de base-ball s'explique par le nombre d'années passé dans la ligue, le nombre de matchs joués en moyenne, et le niveau du joueur représenté par des indicateurs batting_average ("career batting average") home_runs ("home runs per year") et runs_batted ("runs batted in per year"). On estime l'équation suivante:

$$\begin{aligned}
\widehat{\ln(\mathrm{SAL})} &= \underset{(0.29)}{11.10}+\underset{(0.0121)}{0.0689}\cdot\mathrm{ANNEES}+\underset{(0.0026)}{0.126}\cdot\mathrm{NB_MATCHS}\\
&\quad+\underset{(0.0110)}{0.00098}\cdot\mathrm{BAT_AV}+\underset{(0.0161)}{0.0144}\cdot\mathrm{HOM_RUN}\\
&\quad+\underset{(0.0072)}{0.0108}\cdot\mathrm{RUN_BAT}\\
N &= 353,\qquad SSR=183.16,\qquad R^2=0.627
\end{aligned}$$

On désire tester que les performances des joueurs influencent le salaire. Pour cela, on estime le modèle contraint suivant:

$$\begin{aligned}\widehat{\ln(\text{salaire})} &= \underset{(0.11)}{11.22} + \underset{(0.0125)}{0.0713} \cdot \text{ANNEES} + \underset{(0.0013)}{0.0202} \cdot \text{NB\_MATCHS} \\ N &= 353, \qquad SSR = 198.311, \qquad R^2 = 0.5971\end{aligned}$$

La statistique de student est égale à: 9.55.

Exemple 8: Le test de Fisher permet également de tester d'autres formes de restrictions. Considérons l'exemple suivant:

$$\begin{aligned}\log(\text{PRIX}) &= \beta_0 + \beta_1 \cdot \log(\text{EVAL}) + \beta_2 \cdot \log(\text{SUP\_TER}) \\ &\quad + \beta_3 \cdot \log(\text{SUP\_HABIT}) + \beta_4 \cdot \text{CHAM\_COU} + u \\ \text{SSR}_{nr} &= 1.822;\ N = 88\end{aligned}$$

L'hypothèse nulle est la suivante:

$$\text{H}_0 : \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0.$$

Pour tester cette hypothèse, le modèle restreint est le suivant:

$$\log(\text{PRIX}) - \log(\text{EVAL}) = \beta_0 + u$$

Le modèle non restreint est le suivant:Or le SSR_r du modèle restreint est égal:

$$\text{SSR}_r = 1.880$$

et la statistique de Fisher à

$$\left[(1.880 - 1.822)/(1.822)\right](83/4) = 0.661$$

6 Modèle de régression linéaire multiple 4: Propriétés asymptotiques des estimateurs

Ce chapitre traite des propriétés des estimateurs des MCO lorsque le nombre d'observations devient très grand (tend vers l'infini).

6.1 Convergence

If you can't get it right as n goes to infinity, you shouldn't be in this business. (C.W..J. Granger).

Intuition de la convergence. Soit $\hat{\beta}_j$ l'estimateur des MCO de β_j. Pour tout N, $\hat{\beta}_j$ a une distribution de probabilité. Si l'estimateur $\hat{\beta}_j$ est convergent, alors sa distribution est de plus en plus 'concentrée' autour de sa valeur de population β_j, et sa distribution dégénère en un seul point β_j lorsque N tend vers l'infini.

Hypothèse $3''$ (moyenne et corrélations nulles): $\mathrm{E}(u) = 0, \mathrm{cov}(x_j, u) = 0$, $\mathrm{var}(x_j) < \infty$.

Remarque: Cette hypothèse est plus faible que l'hypothèse 3 précédente.

Théorème 8 (convergence des MCO): Sous les hypothèses 1-2, $3''$ et 4, les estimateurs des MCO $\hat{\beta}_j$ convergent vers β_j pour tout j:

$$\mathrm{plim}\hat{\beta}_j = \beta_j.$$

Démonstration (cas simple): L'estimateur des MCO s'écrit de la manière suivante:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{N}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
&= \beta_1 + \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})\, u_i}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}
\end{aligned}$$

Si l'on passe en plim,

$$\begin{aligned}
\text{plim}\hat{\beta}_1 &= \beta_1 + \frac{\text{cov}(x,u)}{\text{var}(x)} \text{ en vertu de la loi des grands nombres} \\
&= \beta_1 \text{ en vertu de l'hypothèse } 3'.
\end{aligned}$$

Remarques:

(1) De la corrélation entre u et un x_j cause la non-convergence des estimateurs des MCO. Le terme

$$\text{plim}\hat{\beta}_1 - \beta_1 = \frac{\text{cov}(x,u)}{\text{var}(x)}$$

désigne le **biais asymptotique**. Ce terme est positif si x et u sont positivement corrélés.

(2) L'omission d'une variable cause généralement un biais asymptotique. Supposons que le 'vrai' modèle est

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

et que le modèle estimé est

$$y = \beta_0 + \beta_1 x_1 + u$$

où $u = \beta_2 x_2 + \varepsilon$. Dans ce cas,

$$\frac{\text{cov}(x_1,u)}{\text{var}(x_1)} = \frac{\text{cov}(x_1, \beta_2 x_2 + \varepsilon)}{\text{var}(x_1)} = \beta_2 \frac{\text{cov}(x_1,x_2)}{\text{var}(x_1)}.$$

Donc,

$$\text{plim}\hat{\beta}_1 = \beta_1 + \beta_2 \delta_1$$

où

$$\delta_1 = \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)} = \text{ plim } \hat{\delta}_1$$

6.2 Normalité asymptotique

En grand échantillon, l'hypothèse de normalité sur les termes aléatoires peut être abandonnée. Cela est important puisque souvent l'hypothèse de normalité n'est pas crédible.

Théorème 9 (normalité asymptotique): Sous les hypothèses 1-5, et si $\text{var}(x) < \infty$ et $\text{var}(u) < \infty$,

(i) $\sqrt{N} \times (\hat{\beta}_j - \beta_j) \overset{a}{\rightsquigarrow} \mathcal{N}(0, \sigma^2/a_j^2)$ où σ^2/a_j^2 est la variance asymptotique de $\sqrt{N} \times (\hat{\beta}_j - \beta_j)$; et $a_j^2 = \text{plim}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{r}_{ji}^2\right)$, où $\hat{r}_{ji}$ est le résidu de la régression de x_j sur l'ensemble des variables autres explicatives;

(ii) $\hat{\sigma}^2$ est un estimateur convergent de σ^2.

Démonstration (intuition dans le cas simple): De la définition de $\hat{\beta}_1$, nous avons:

$$\sqrt{N} \times \left(\hat{\beta}_1 - \beta_1\right) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)u_i}{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}$$

où le dénominateur converge vers $var(x)$ et le numérateur converge en probabilité vers une loi normale, en vertu du Théorème central limite.

Remarque:

(1) Selon les formules traditionnelles (en petit échantillon), la variance de

$\sqrt{N} \times (\hat{\beta}_j - \beta_j)$ est égale à

$$N \times \frac{\sigma^2}{\left(1 - R_j^2\right) \sum_{i=1}^{N} (x_{ji} - \bar{x}_j)} = \frac{\sigma^2}{\frac{1}{N} \sum_{i=1}^{N} \hat{r}_{ji}^2}.$$

Quand la taille de l'échantillon grandit, le dénominateur de cette expression converge vers a_j^2.

(2) Pour tout j, puisque plim $\hat{\sigma}^2 = \sigma^2$,

$$\frac{\sqrt{N} \times (\hat{\beta}_j - \beta_j)}{\sqrt{\frac{\hat{\sigma}}{\frac{1}{N} \sum_{i=1}^{N} \hat{r}_{ji}^2}}} = \frac{(\hat{\beta}_j - \beta_j)}{\sqrt{\widehat{var}(\hat{\beta})}} \overset{a}{\rightsquigarrow} \mathcal{N}(0, 1).$$

La loi de Student converge vers la loi Normale centrée réduite lorsque le nombre de degrés de liberté tend vers l'infini.

(3) Les tests de Student et de Fisher sont asymptotiquement valides. Si l'échantillon est grand, les procédures d'inférences décrites précédemment peuvent être appliquées telles quelles.

7 Modèle de régression linéaire multiple 5: Specification

7.1 Le choix de la forme fonctionnelle

7.1.1 Les terms logarithmiques

Lorsque les variables expliquées et explicatives sont en logarithmes, les paramètres sont des élasticités. Lorsque seule la variable expliquée est en logarithme, les paramètres sont des demi-élasticitées. Ce sont des approximations si la variable explicative varie de manière discrete.

Exemple 1 (équation de prix hédonique):

$$\begin{aligned} \widehat{\log(\text{PRIX})} &= \underset{(0.19)}{9.23} - \underset{(0.066)}{0.718} \cdot \log(\text{OX\_NIT}) + \underset{(0.019)}{0.306} \cdot \text{PIECES}, \\ N &= 506, R^2 = 0.514 \end{aligned}$$

où nox désigne la quantité d'oxyde nitreux dans l'air. L'effet exact en pourcentage de l'accroissement d'une pièce sur le prix est:

$$\begin{aligned} \Delta\widehat{\log(\text{PRIX})} &= 0.306 \cdot \Delta\text{PIECES} \\ &\Rightarrow \log\left(\frac{\text{PRIX}_{\text{rooms = r+1}}}{\text{PRIX}_{\text{rooms = r}}}\right) = 0.306 \\ &\Rightarrow \frac{\text{PRIX}_{\text{rooms = r+1}}}{\text{PRIX}_{\text{rooms = r}}} = \exp(0.306) = 1.358 \end{aligned}$$

Remarque: Les variables monétaires, ou plus généralement celles dont la valeur est large, sont souvent en logarithmes. Les variables mesurées en années sont souvent en niveaux. Les taux sont souveaux en niveaux.

7.1.2 Les termes quadratiques

Les variables explicatives peuvent être en exprivées sous forme quadratique.

La forme fonctionnelle est alors caractérisée par un poit d'inflexion.

Exemple 2 (équation de salaire):

$$\begin{aligned}\widehat{\text{SAL}} &= \underset{(0.35)}{3.73} + \underset{(0.041)}{0.298} \cdot \text{EXPER} - \underset{(0.0009)}{0.0061} \cdot \text{EXPER}^2. \\ N &= 526, R^2 = 0.093\end{aligned}$$

L'effet de l'expérience est donné par

$$\Delta\widehat{\text{SAL}} = (0.298 - 2 \times 0.0061 \cdot \text{EXPER}) \times \Delta\text{EXPER}.$$

La première année d'expérience rapporte environ 29.8 centimes, la seconde année environ 28.6 centimes, la onzième année environ 17.6 centimes. L'année à partir de laquelle l'effet d'exper est négatif est donnée par

$$(0.298 - 2 \times 0.0061 \cdot \text{EXPER}^*) = 0 \Rightarrow \text{EXPER}^* = \frac{0.298}{2 \times 0.0061} \approx 24.4.$$

Remarque: Les termes quadratiques peuvent être mélangés avec des termes logarithmiques. Ils peuvent également être associés à des termes d'interaction.

Exemple 3 (équation de salaire, Wooldridge, p. 214):

$$\begin{aligned}\widehat{\log(\text{SAL})} &= \underset{(0.24)}{5.95} + \underset{(0.0174)}{0.0440} \cdot \text{EDUC} - \underset{(0.0200)}{0.0215} \cdot \text{EXPER} + \underset{(0.00153)}{0.00320} \cdot \text{EDUC} \times \text{EXPER} \\ n &= 935, R^2 = 0.135\end{aligned}$$

L'effet de l'expérience est donné par:

$$\frac{\Delta\widehat{\log(\text{SAL})}}{\Delta\text{EXPER}} = -0.0215 + 0.00320 \cdot \text{EDUC}.$$

et donc

$$-0.0215 + 0.00320 \cdot \text{EDUC}^* = 0 \Leftrightarrow \text{EDUC}^* = \frac{0.0215}{0.00320} \approx 6.71$$

L'effet de l'expérience sur le salaire est négatif pour les personnes ayant 6 années et moins d'éducation, et est positif pour les personnes ayant 7 années et plus d'éducation.

7.2 Les variables qualitatives

Les variables qualitatives sont des variables sous la forme binaire. Exemple: homme ou femme, français ou étranger, etc. On parle de variable 'dichotomique' ou de variable 'dummy'.

Personne	SAL	EDUC	EXPER	FEM	MAR
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	8	1	0

Exemple 4 (une seule variable binaire): Considérons la régression du salaire sur l'éducation et une variable binaire qui prend la valeur 1 lorsque le travailleur est une femme et zéro sinon:

$$\text{SAL} = \beta_0 + \delta_0 \cdot \text{FEM} + \beta_1 \cdot \text{EDUC} + u.$$

Ce modèle incorpore deux modèles particuliers:

- Si FEM = 1,

$$\text{SAL} = (\beta_0 + \delta_0) + \beta_1 \cdot \text{EDUC} + u$$

- Si FEM = 0,

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + u$$

Une variable explicative dichotomique implique un déplacement de la constante de la droite de régresion. Dans le ca présent, la catégore 'male' est la catégorie de référence, par rapport à laquelle l'effet d'être une femme est mesuré. Le choix de la catégorie de référence (male ou female) est arbitraire.

$$\begin{aligned}
\widehat{\text{SAL}} &= \underset{(0.72)}{-1.57} - \underset{(0.26)}{1.81} \cdot \text{FEM} \\
&\quad + \underset{(0.049)}{0.572} \cdot \text{EDUC} + \underset{(0.012)}{0.025} \cdot \text{EXPER} + \underset{(0.021)}{0.141} \cdot \text{ANC} \\
N &= 526, \qquad R^2 = 0.364 \\
&\Rightarrow \text{mesure de discrimination} \\
\widehat{\text{SAL}} &= \underset{(0.21)}{7.10} - \underset{(0.30)}{2.51} \cdot \text{FEM} \\
N &= 526, \qquad R^2 = 0.116
\end{aligned}$$

Exemple 5 (une variable binaire dans une régression logarithmique):

$$\begin{aligned}
\widehat{\log(\text{PRIX})} &= \underset{(0.65)}{5.56} + \underset{(0.038)}{0.168} \cdot \log(\text{SURF_TER}) + \underset{(0.093)}{0.707} \cdot \log(\text{SURF_HAB}) \\
&\quad + \underset{(0.029)}{0.027} \cdot \text{CH_COUCH} + \underset{(0.045)}{0.054} \cdot \text{COLONIAL} \\
N &= 88, \qquad R^2 = 0.649
\end{aligned}$$

Exemple 6 (croisement de variables binaires):

$$\begin{aligned}
\widehat{\log(\text{SAL})} &= \underset{(0.100)}{0.321} + \underset{(0.055)}{0.213} \cdot \text{HOM} \cdot \text{MAR} - \underset{(0.058)}{0.198} \cdot \text{FEM} \cdot \text{MAR} - \underset{(0.056)}{0.110} \cdot \text{FEM} \cdot \text{CEL} \\
&\quad + \underset{(0.007)}{0.079} \cdot \text{EDUC} + \underset{(0.005)}{0.027} \cdot \text{EXPER} \\
&\quad - \underset{(0.00011)}{0.00054} \cdot \text{EXPER}^2 + \underset{(0.007)}{0.029} \cdot \text{ANC} \\
&\quad - \underset{(0.00023)}{0.00053} \cdot \text{ANC}^2 \\
N &= 526, \qquad R^2 = 0.461
\end{aligned}$$

Exemple 7 (variables binaires et variables ordonnées, Hamermesh et Bidle, 1994): Une enquête est réalisée sur les salaires; les enquêteurs récoltent également de l'information sur la beauté physique des personnes interrogées. Hamermesh et Bidle classent les individus entre beauté inférieure à la maoyenne, moyenne, supérieure à la moyenne. Ils effectuent les estimations suivantes:

$$\begin{aligned}
\log \widehat{\text{SAL_HOM}} &= \beta_0 - \underset{(0.046)}{0.164} \cdot \text{SUP_MOY} + \underset{(0.033)}{0.016} \cdot \text{INF_MOY} + \text{AUTRES} \\
N &= 700 \\
\log \widehat{\text{SAL_FEM}} &= \beta_0 - \underset{(0.066)}{0.124} \cdot \text{SUP_MOY} + \underset{(0.049)}{0.035} \cdot \text{INF_MOY} + \text{AUTRES} \\
N &= 409
\end{aligned}$$

Exemple 8 (salaire des joueurs de base-ball et discrimination): Effet de la race sur les salaires des joueurs de baseball

$$\begin{aligned}
\widehat{\log(\text{SAL})} = \; & \underset{(2.18)}{10.34} + \underset{(0.0129)}{0.0673} \cdot \text{ANNEES} + \underset{(0.0034)}{0.0089} \cdot \text{NB_MATCHS} \\
& + \underset{(0.00151)}{0.00095} \cdot \text{BAT_AV} + \underset{(0.0164)}{0.0146} \cdot \text{HOM_RUN} + \underset{(0.0076)}{0.0045}\text{RUN_BAT} \\
& + \underset{(0.0046)}{0.0072}\text{RUN} + \underset{(0.0021)}{0.0011}\text{FLDPERC} + \underset{(0.0029)}{0.0075}\text{ALLSTAR} \\
& - \underset{(0.125)}{0.198}\text{NOIR} - \underset{(0.153)}{0.190}\text{HISP} + \underset{(0.0050)}{0.0125}\text{NOIRk} \times \text{PERC_NOIR} \\
& + \underset{(0.0098)}{0.0201}\text{HISPAN} \times \text{PERC_HISP}
\end{aligned}$$

7.3 Test de Chow

Le test de Chow permet de tester que des populations différentes sont caractérisées par les même valeurs de paramètres. L'estimation de

$$\text{SAL} = \beta_0 + \delta_0 \cdot \text{FEM} + \beta_1 \cdot \text{EDUC} + \delta_1 \cdot \text{FEM} \times \text{EDUC} + u$$

est équivalente aux estimations de

$$\text{SAL} = \beta_0 + \beta_1 \cdot \text{EDUC} + \beta_2 + u$$

pour les hommes, et de

$$\begin{aligned}\mathrm{SAL} &= (\beta_0+\delta_0)+(\beta_1+\delta_1)\cdot \mathrm{EDUC}+u \\ &= \beta_0^*+\beta_1^*\cdot \mathrm{EDUC}+u\end{aligned}$$

pour les femmes. Si l'on veut tester que la même droite de régression s'applique aux hommes et aux femmes, l'hypothèse nulle s'écrit::

$$\mathrm{H}_0 : \delta_0 = 0, \delta_1 = 0.$$

Le test de Fisher de cette hypothèse nulle s'écrit:

$$F = \frac{(\mathrm{SSR}_r - (\mathrm{SSR}_1 + \mathrm{SSR}_2))}{(\mathrm{SSR}_1 + \mathrm{SSR}_2)} \frac{N - 2\,(K+1)}{K+1}$$

7.4 Le R^2 ajusté et la sélection de modèles

Le R^2 s'écrit:

$$R^2 = 1 - \frac{\mathrm{SSR}/N}{\mathrm{SST}/N}.$$

Dans ce cas, R^2 est vu comme un estimateur de:

$$R^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}.$$

Mais, si le nombre de paramètres à estimer tend vers le nombre d'observation, le R^2 tend vers 1 (la SSR tend vers 0). Celà est dû au fait que les estimateurs de σ_u^2 et de σ_y^2 sont biaisés. Donc, le $\bar{R}^2$ est défini par:

$$\bar{R}^2 = 1 - \frac{\mathrm{SSR}/(N-K-1)}{\mathrm{SST}/(N-1)}$$

L'ajout d'une variable explicative fait augmenter sa valeur si et seulement si le t-test de la variable ajoutée est supérieur à 1. Le $\bar{R}^2$ peut être négatif. Le $\bar{R}^2$ peut être utilisé pour sélectionner des modèles non emboîtés.

Exemple. Soit deux modèles qui expliquent la recherche en développement en fonction du chiffre d'affaire:

$$\begin{aligned}\text{INTENS\_RD} &= \beta_0 + \beta_1 \cdot \log(\text{VENTE}) + u \\ \text{INTENS\_RD} &= \beta_0 + \beta_1 \cdot \text{VENTE} + \beta_2 \cdot \text{VENTEs}^2 + u\end{aligned}$$

Remarque. Si l'on insiste excessivement sur la minimisation des résidus, on peut contrôler pour trop de variables. Le choix des variables explicatives dépend de ce que l'on veut expliquer.

(1) Si l'on veut expliquer l'effet d'une taxation de l'alcohol sur le nombre d'accidents de la route,

$$\begin{aligned}\text{ACC} &= \beta_0 + \beta_1 \cdot \text{TAX} + \beta_2 \cdot \text{MILES} + \beta_3 \cdot \text{PERC\_HOM} + \ldots + u \\ \text{ACC} &= \beta_0 + \beta_1 \cdot \text{TAX} + \beta_2 \cdot \text{CONS\_BIERE} + \ldots + u\end{aligned}$$

le second modèle est inadapté.

(2) Si l'on veut calculer des prix hédoniques pour les maisons,

$$\begin{aligned}\log(\text{PRIX}) &= \beta_0 + \beta_1 \cdot \log(\text{SURF\_HAB}) + ... + u \\ \log(\text{PRIX}) &= \beta_0 + \beta_1 \cdot \log(\text{EVAL}) + \beta_1 \cdot \log(\text{SURF\_HAB}) + ... + u\end{aligned}$$

le second modèle est inadapté.

7.5 Prédiction et analyse de résidus

7.5.1 Variance de la moyenne conditionnelle

On effectue une régression de y sur un ensemble de variables explicatives:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots + \beta_K \cdot x_K + u,$$

et on obtient les estimateurs des paramètres $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_K$. Soit x_1^o, x_2^o, $\ldots, x_K^o$ des valeur particulière pour chacune des variables explicatives. On veut prédire la moyenne conditionnelle de y, c'est-à-dire:

$$\begin{aligned}\psi^o &= \beta_0 + \beta_1 \cdot x_1^o + \beta_2 \cdot x_2^o + \ldots + \beta_K \cdot x_K^o \\ &= E(y|x_1 = x_1^o, x_2 = x_2^o, \ldots, x_K = x_K^o).\end{aligned}$$

La prédiction de la moyenne conditionnelle est:

$$\hat{y}^o = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1^o + \hat{\beta}_2 \cdot x_2^o + \ldots + \hat{\beta}_K \cdot x_K^o.$$

Pour calculer l'écart-type de ψ_0, on procède en définissant:

$$\beta_0 = \psi^o - \beta_1 \cdot x_1^o - \beta_2 \cdot x_2^o - \ldots - \beta_K \cdot x_K^o$$

que l'on remplace dans l'équation estimée. On obtient la régression:

$$y = \psi^o - \beta_1 \cdot (x_1 - x_1^o) + \beta_2 \cdot (x_2 - x_2^o) + \ldots + \beta_K \cdot (x_K - x_K^o) + u.$$

L'estimation de cette relation donne la variance de l'estimateur de ψ^o, à savoir, $\text{var}(\hat{y}^o)$. Cette procédure permet de calculer un intervalle de confiance pour l'espérance conditionnelle.

7.5.2 Variance de la prédiction

Pour calculer un intervalle de confiance de la prédiction, il faut également tenir compte de la variance du terme aléatoire. Soit u^o une valeur (non observée) pour le terme aléatoire:

$$y^o = \beta_0 + \beta_1 \cdot x_1^o + \beta_2 \cdot x_2^o + \ldots + \beta_K \cdot x_K^o + u^o.$$

L'erreur de prédiction en utilisant $\hat{y}^o$ pour prédire y^o est:

$$\hat{e}^o = y^o - \hat{y}^o = \beta_0 + \beta_1 \cdot x_1^o + \beta_2 \cdot x_2^o + \ldots + \beta_K \cdot x_K^o + u^o - \hat{y}^o.$$

Or,

$$E(\hat{y}^o) = y^o \Rightarrow E(\hat{e}^o) = 0$$

Donc, l'espérance de l'erreur de prédiction est nulle. La variance de l'erreur de prédiction est:

$$\mathrm{var}(\hat{e}^o) = \mathrm{var}(u^o - \hat{y}^o) = \mathrm{var}(\hat{y}^o) + \sigma_u^2$$

car u^o et $\hat{y}^o$ sont non corrélés. Donc, il y a deux sources de la variance de $\hat{e}^o$, mais la première devient négligeable lorsque l'échantillon est grand.

7.6 Analyse de résidus

Exemples L'analyse des résidus peut être utile dans diverses circonstances:

(1) Pour l'achat d'une maison;
(2) Pour mesurer l'efficacité d'une firme;
(3) Pour classer les écoles de droit.

7.7 Problèmes de spécification

7.7.1 Mauvaise spécification de la forme fonctionnelle

Exemple 1. Le vrai modèle est:

$$\log(\mathrm{SAL}) = \beta_0 + \beta_1 \cdot \mathrm{EDUC} + \beta_2 \cdot \mathrm{EXPER} + \beta_3 \cdot \mathrm{EXPER}^2 + u,$$

et le modèle estimé est:

$$\log(\mathrm{SAL}) = \beta_0 + \beta_1 \cdot \mathrm{EDUC} + \beta_2 \cdot \mathrm{EXPER} + u.$$

pu encore

$$\mathrm{SAL} = \beta_0 + \beta_1 \cdot \mathrm{EDUC} + \beta_2 \cdot \mathrm{EXPER} + \beta_3 \cdot \mathrm{EXPER}^2 + u,$$

Remarque. Si l'on spécifie mal, la relation entre une relation entre la variable y et l'une des variables x, l'on parle de mauvaise sépcification. Cela mène généralement à un biais des estimateurs.

7.7.2 Test RESET.

Le test RESET (pour Regression Specification Error Test) de Ramsey permet de tester le forme fonctionnelle. Le modèle dont on veut tester la forme fonctionnelle est:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + u.$$

Dans ce cas, on estime:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u.$$

où $\hat{y}$ est la valeur prédite de la relation initiale. Le test consiste à tester $\mathrm{H}_0 : \delta_1 = \delta_2 = 0$ à l'aide d'un test de Fisher. Le test RESET est surtout valable pour faire un test contre un modèle plus général (qui englobe le modèle particulier).

7.7.3 Test de modèles non emboîtés.

Dans le cas de modèles non emboîtés, d'autres tests ont été développés. Considérons les modèles suivants:

$$\mathrm{I}: \qquad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

et

$$\mathrm{II}: \qquad y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u$$

Le modèle I n'est pas un cas particulier du modèle II; le modèle II n'est pas un cas particulier du modèle I. Le test de Mizon–Richard consiste à formuler un modèle général qui englobe les deux modèles particuliers:

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \log(x_1) + \gamma_4 \log(x_2) + u.$$

Le test du modèle I contre le modèle II (H_0 : le modèle I est vrai; H_1 : le modèle II est vrai) consiste à tester $\gamma_3 = \gamma_4 = 0$ par un test de Fisher. Le test du modèle II contre le modèle I (H_0 : le modèle II est vrai; H_1 : le modèle I est vrai) consiste à tester $\gamma_1 = \gamma_2 = 0$. Le test de Davidson–MacKinnon consiste à formuler les modèles:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y} + u$$

et

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_2 \hat{y} + u$$

où les valeurs prédites $\hat{y}$ sont obtenues par le modèle alternatif, et à tester soit $\theta_1 = 0$, soit $\theta_2 = 0$, par un test de Student.

Remarque. Les test de modèles emboîtés peuvent donner quatre résultats: le modèle I est 'accepté' et le modèle II est rejeté, le modèle I est rejeté et le modèle II est 'accepté', le modèle I et le modèle II sont rejetés, le modèle I et le modèle II sont 'acceptés'.

7.8 Utilisation de variables 'proxy'

Une variable 'proxy' peut dans certaines conditions remplacer une variable non observée. Soit le modèle:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u$$

où x_3^* n'est pas observé. Or,

$$x_3^* = \delta_0 + \delta_1 x_3 + \nu.$$

où x_3 est observé. Si l'on remplace cette expression dans le modèle initial, on obtient:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (\delta_0 + \delta_1 x_3 + \nu) + u \\ &= (\beta_0 + \beta_3 \delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_1 x_3 + (\beta_3 \nu + u) \end{aligned}$$

ou encore:

$$y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e$$

L'approche consiste donc à régresser y sur x_1, x_2, x_3 et les estimateurs des MCO seront convergents si (conditions suffisantes)

(1) Le terme aléatoire u est non corrélé avec x_1, x_2 et x_3^* (ce sont les hypothèses habituelles des MCO).
(2) Le terme aléatoire u est non corrélé avec x_3 (cette hypothèse naturelle puisque sinon la variable x_3 aurait été introduite dans le modèle).
(3) Le terme aléatoire ν est non corrélé avec x_1, x_2 et x_3 (c'est ce qui définit une bonne variable 'proxy').

Exemple. Le vrai modèle est:

$$\log(\text{SAL}) = \beta_0 + \beta_1 \cdot \log(\text{EDUC}) + \beta_2 \cdot \log(\text{EXPER}) + \beta_3 \cdot \log(\text{HABIL}) + u$$

Or, la variable abil n'est pas observée. La solution est d'utiliser la variable QI comme proxy, avec

$$\log(\text{HABIL}) = \delta_0 + \delta_1 \cdot \text{QI} + \nu.$$

Cela permet d'obtenir des estimateurs convergents de β_1 et de β_2 sous les conditions précitées.

7.9 Propriétés des MCO avec erreurs de mesure

7.9.1 Si les erreurs de mesure portent sur la variable dépendante:

Soit le modèle suivant:

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

où la variable y^* est affectée d'une erreur de mesure:

$$e_0 = y - y^*$$

Le modèle estimé est:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + (u - e_0)$$

Si l'erreur de mesure n'est pas corrélée avec x_1, x_2, x_3 les estimateurs seront convergents.

7.9.2 Si les erreurs de mesure portent sur la variable indépendante:

Soit le modèle suivant:

$$y = \beta_0 + \beta_1 x_1^* + u$$

où la variable x_1^* est affecté d'une erreur de mesure:

$$e_1 = x_1 - x_1^*$$

Le modèle estimé est:

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$$

Si l'erreur de mesure n'est pas corrélée avec x_1 les estimateurs seront convergents. Cependant, cette hypothèse est peu réaliste. Généralement, on aura:

$$\text{cov}(e_1, x_1^*) = 0$$

Dans ce cas, il y aura une corrélation entre la variable explicative et le terme aléatoire:

$$\begin{aligned}
\text{cov}(e_1, x_1) &= \text{cov}(e_1, x_1^* + e_1) \\
&= E(e_1 (x_1^* + e_1)) \\
&= E(e_1 x_1^* + e_1^2) \\
&= E(e_1 x_1^*) + E(e_1^2) \\
&= E(e_1^2) \\
&= \sigma_{e_1}^2
\end{aligned}$$

Pour cela, remarquons:

$$\begin{aligned}
\operatorname{cov}(x_1, u - \beta_1 e_1) &= E(x_1 (u - \beta_1 e_1)) \\
&= E(x_1 u - \beta_1 e_1 x_1) \\
&= E(x_1 u) - \beta_1 E(e_1 x_1) \\
&= -\beta_1 cov(e_1, x_1) \\
&= -\beta_1 \sigma_{e_1}^2
\end{aligned}$$

Cela permet de mesurer le biais asymptotique. En effet,

$$\begin{aligned}
\operatorname{plim}\left(\hat{\beta}_1\right) &= \beta_1 + \frac{cov\,(x_1, u - \beta_1 e_1)}{\operatorname{var}(x_1)} \\
&= \beta_1 - \frac{\beta_1 \sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} \\
&= \beta_1 \left(1 - \frac{\sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right) \\
&= \beta_1 \left(\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right)
\end{aligned}$$

C'est le biais d'atténuation.

7.10 Données manquantes, échantillons non aléatoires, et observations abérantes

La notion de sélection endogène et de sélection exogène.

8 Heteroscédasticité

L'hétéroscédasticité invalide les estimateurs de la variance des $\hat{\beta}_j$ et fausse les procédure classiques de tests; le théorème de Gauss-Markov n'est plus correct.

8.1 Tests de l'hétéroscédasticité

Les tests d'hétéroscédasticité sont intéressants car, en cas d'hétéroscédasticité, les estimateurs des MCO ne sont plus les meilleurs estimateurs. Soit le modèle de régression multiple:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

L'hypothèse nulle est:

$$\mathrm{H}_0 : \mathrm{var}\,(u|x_1, x_2) = \sigma^2$$

ou de manière équivalente:

$$\mathrm{H}_0 : \mathrm{E}\left(u^2|x_1, x_2\right) = \sigma^2$$

Le test d'homoscédasticité consiste à tester si u^2 est corrélé avec certaines variables explicatives. Une simple approche consiste à supposer une relation entre u^2 et x_1, x_2. Le test de Breusch-Pagan suppose une relation est linéaire:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + v^*$$

et la procédure de test se réduit à

$$\mathrm{H}_0 : \delta_1 = \delta_2 = 0.$$

Cependant, comme u^2 n'est pas observé, il doit être remplacé par $\hat{u}^2$:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + v$$

où $v = v^* + \hat{u}^2 - u^2$. La statistique de Fisher de la significativité globale de la

régression est:
$$F = \frac{R_u^2/K}{(1-R_u^2)/(n-K-1)}$$
Le test de White suppose une relation est quadratique:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 + \delta_5 x_1 x_2 + v^*$$

et le test de Breusch-Pagan se réduit à

$$\mathrm{H}_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.$$

La statistique de Fisher est la même que la précédente. Le test de White se justifie pour des raisons de convergence.

8.2 Inférence robuste à l'hétéroscédasticité

Les écarts-type des paramètres peuvent assez facilement être modifiés afin de tenir compte de l'hétéroscédasticité et de rendre possible des tests de student. Considérons le modèle de régression simple:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

avec
$$\mathrm{var}(u_i|x_i) = \sigma_i^2$$
Dans ce cas, l'estimateur des MCO s'écrit:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}_i) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

et donc:

$$\mathrm{var}(\hat{\beta}_1) = \mathrm{var}\left(\frac{\sum_{i=1}^N (x_i - \bar{x}_i) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}\right) = \frac{\sum_{i=1}^N (x_i - \bar{x}_i)^2 \sigma_i^2}{\left[\sum_{i=1}^N (x_i - \bar{x})^2\right]^2}$$

Sous certaines conditions, l'estimateur de la variance,

$$\widehat{\mathrm{var}(\hat{\beta}_1)} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2 \hat{u}_i^2}{\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^2}$$

est "satisfaisant", car

$$\mathrm{plim}\left(\frac{\frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x}_i)^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^2}\right) = n \times \mathrm{var}(\hat{\beta}_1)$$

Dans le cas général,

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_K x_{Ki} + u_i$$

$$\widehat{\mathrm{var}(\hat{\beta}_j)} = \frac{\sum_{i=1}^{N} \hat{r}_{ji}^2 \hat{u}_i^2}{\left[\sum_{i=1}^{N} \hat{r}_{ji}^2\right]^2}$$

C'est l'estimateur de la variance robuste à l'hétéroscédasticité (dit estimateur de White ou de Huber).

Exemple.

$$\begin{aligned}\widehat{\log(\mathrm{SAL})} \quad = \quad & \underset{\substack{(0.100)\\ [0.109]}}{0.321} + \underset{\substack{(0.055)\\ [0.057]}}{0.213} \cdot \mathrm{HOM\_MAR} - \underset{\substack{(0.058)\\ [0.058]}}{0.198} \cdot \mathrm{FEM\_MAR} - \underset{\substack{(0.056)\\ [0.057]}}{0.110} \cdot \mathrm{FEM\_CEL} \\ & + \underset{\substack{(0.0067)\\ [0.0074]}}{0.0789} \cdot \mathrm{EDUC} + \underset{\substack{(0.0055)\\ [0.0051]}}{0.0268} \cdot \mathrm{EXPERr} - \underset{\substack{(0.00011)\\ [0.00011]}}{0.00054} \cdot \mathrm{EXPER}^2 \\ & + \underset{\substack{(0.0068)\\ [0.0069]}}{0.0291} \cdot \mathrm{ANC} - \underset{\substack{(0.00023)\\ [0.00024]}}{0.00053} \cdot \mathrm{ANC}^2\end{aligned}$$

8.3 Estimation par les Moindres Carrés Pondérés

8.3.1 Cas 1: observation de $h(\boldsymbol{x})$.

Soit le modèle suivant:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + u_i$$

où le terme aléatoire est hétéroscédastique. L'hétéroscédasticité est connue à une fonction multiplicative près, c'est-à-dire

$$var(u|\boldsymbol{x}) = \sigma^2 \times h(\boldsymbol{x})$$

où $h(\boldsymbol{x})$ est une fonction positive connue des variables explicatives. Puisque $var(u_i|\boldsymbol{x}_i) = E(u_i^2|\boldsymbol{x}_i) = \sigma^2 \times h(\boldsymbol{x}_i)$, la variance de $u_i/h(\boldsymbol{x}_i)$ est égale à

$$E\left(\left(\frac{u_i}{\sqrt{h(\boldsymbol{x}_i)}}\right)^2 \middle| \boldsymbol{x}_i\right) = \frac{1}{h(\boldsymbol{x}_i)} E\left(u_i^2 \middle| \boldsymbol{x}_i\right) = \sigma^2$$

Si nous divisons les variables de l'équation de régression, nous obtenons:

$$\frac{y_i}{\sqrt{h(\boldsymbol{x}_i)}} = \beta_0 \frac{1}{\sqrt{h(\boldsymbol{x}_i)}} + \beta_1 \frac{x_{i1}}{\sqrt{h(\boldsymbol{x}_i)}} + \beta_2 \frac{x_{i2}}{\sqrt{h(\boldsymbol{x}_i)}} + \ldots + \beta_K \frac{x_{iK}}{\sqrt{h(\boldsymbol{x}_i)}} + \frac{u_i}{\sqrt{h(\boldsymbol{x}_i)}}.$$

ou

$$y_i^* = \beta_0 x_{0i}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \ldots + \beta_K x_{iK}^* + u_i^*$$

Cet estimateur est une variante des moindres carrés généralisés, appelée "moindres carrés pondérés". En effet,

$$\sum_{i=1}^{N} \frac{(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_K x_{iK})^2}{h(\boldsymbol{x}_i)}$$

donne:

$$\sum_{i=1}^{N} (y_i^* - \beta_0 x_{i0}^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^* - \ldots - \beta_K x_{iK}^*)^2 .$$

Les tests de Student et de Fisher peuvent être utilisés sur le modèle transformé. Le Théorème de Gauss-Markov est également apllicable.

Exemple. Le modèle initial est:

$$\text{EPAR}_i = \beta_0 + \beta_1 \text{REV}_i + u_i \text{ avec } \text{var}(u_i|\text{REV}_i) = \sigma^2 \times \text{REV}_i$$

Le modèle transformé est:

$$\frac{\text{EPAR}_i}{\sqrt{\text{REV}_i}} = \beta_0 \frac{1}{\sqrt{\text{REV}_i}} + \beta_1 \sqrt{\text{REV}_i} + u_i^*$$

Estimations				
Ind. Var	(1) OLS	(2) WLS	(3) OLS	(4) WLS
REV	0.147 (0.058)	0.172 (0.057)	0.109 (0.071)	0.101 (0.077)
TAILLE	–	–	67.6 (222.9)	−6.8 (168.4)
EDUC	–	–	151.8 (117.2)	139.4 (100.5)
AGE	–	–	0.28 (50.03)	21.7 (41.3)
NOIR	–	–	518.3 (1308.0)	137.2 (844.5)
CONST	124.8 (655.3)	−124.9 (480.8)	−1605.4 (2830.7)	−1854.8 (2351.8)
Obs.	100	100	100	100
R^2	0.0621	0.0853	0.0828	0.1042

8.3.2 Cas 2: non observation de $h(\boldsymbol{x})$.

En général, le fonction n'est pas observée mais doit être estimée. Pour cela, il faut choisir une forme fonctionnelle pour la variance:

$$\operatorname{var}(u|\boldsymbol{x}) = \sigma^2 \times \exp(\delta_1 x_1 + \ldots + \delta_K x_K).$$

L'objectif est d'estimer cette relation. De manière équivalente, écrivons:

$$u^2 = \sigma^2 \times \exp(\delta_1 x_1 + \ldots + \delta_K x_K) \times \upsilon,$$

où $E(\upsilon) = 1$. Si υ est indépendant de $x_1, \ldots, x_K$,

$$\log\left(u^2\right) = \log\left(\sigma^2\right) + \delta_1 x_1 + \ldots + \delta_K x_K + \log(\upsilon),$$

où $E(\log(\upsilon)) \neq 0$ en général. Donc:

$$\log\left(u^2\right) = \alpha_0 + \delta_1 x_1 + \ldots + \delta_K x_K + e,$$

où $\alpha_0 = \log(\sigma^2) + E(\log(\upsilon))$ et $e = \log(\upsilon) - E(\log(\upsilon))$. Les paramètres peuvent donc être estimés par une régression de:

$$\log\left(\hat{u}^2\right) \text{ sur } x_1, \ldots, x_K.$$

Cela donne les valeurs prédites:

$$\widehat{\log(\hat{u}^2)}_i = \hat{\alpha}_0 + \hat{\delta}_1 x_{1i} + \ldots + \hat{\delta}_K x_{Ki}.$$

Si l'on prend l'exponentielle de cette expression, l'on obtient une estimation de la fonction $h(\boldsymbol{x})$ à une constante multiplicative près:

$$\begin{aligned} \exp\left(\widehat{\log(\hat{u}^2)}_i\right) &= \exp\hat{\alpha}_0 \times \exp\left(\hat{\delta}_1 x_{1i} + \ldots + \hat{\delta}_K x_{Ki}\right) \\ &= \exp\hat{\alpha}_0 \times \widehat{h(\boldsymbol{x}_i)} \end{aligned}$$

Exemple.

$$
\begin{aligned}
\widehat{\text{CIGS}} &= \underset{(24.08)}{-3.64} + \underset{(0.728)}{0.880} \cdot \log(\text{REV}) - \underset{(5.773)}{0.751} \cdot \log(\text{PRIX\_CIGS}) - \underset{(0.167)}{0.501} \cdot \text{EDUC} \\
&\quad + \underset{(0.167)}{0.771} \cdot \text{AGE} - \underset{(0.0017)}{0.0990} \cdot \text{AGE}^2 - \underset{(1.11)}{2.83} \cdot \text{RESTAURN}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\text{CIGS}} &= \underset{(17.80)}{5.64} + \underset{(0.44)}{1.30} \cdot \log(\text{REV}) - \underset{(4.46)}{2.94} \cdot \log(\text{PRIX\_CIGS}) - \underset{(0.120)}{0.463} \cdot \text{EDUC} \\
&\quad + \underset{(0.097)}{0.482} \cdot \text{AGE} - \underset{(0.0009)}{0.0056} \cdot \text{AGE}^2 - \underset{(0.80)}{3.46} \cdot \text{RESTAURN}
\end{aligned}
$$

9 Régression avec séries temporelles: Bases

9.1 Caractéristiques des séries temporelles:

- Les observations sont ordonnées
- Les observations ne sont pas issues d'un échantillon aléatoire
  - Rappel: un échantillon de variables (y_i) de la population représentée par la densité $f(y;\theta)$ est aléatoire si les variables y_i sont indépendantes et ont une même densité $f(y;\theta)$.
- L'ensemble des observations est appelé "processus stochastique" (au lieu de "échantillon aléatoire").

9.2 Exemples de modèles estimés avec séries temporelles

Exemple 1 (modèles statiques):

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

Courbe de Phillips statique:

$$\text{INF}_t = \beta_0 + \beta_1 \cdot \text{UNEM}_t + u_t$$

Modèle de criminalité:

$$\text{MRDRTE}_t = \beta_0 + \beta_1 \cdot \text{CONRTE}_t + \beta_2 \cdot \text{UNEM}_t + \beta_3 \cdot \text{YNGMLE}_t + u_t$$

Exemple 2 (modèles avec retards échelonnés)

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t$$

Modèle de fertilité:

$$\text{GFR}_t = \beta_0 + \beta_1 \cdot \text{PE}_t + \beta_2 \cdot \text{PE}_{t-1} + \beta_3 \cdot \text{PE}_{t-2} + u_t$$

9.3 Propriétés en échantillon fini des MCO sous les hypothèses classiques

L'hypothèse d'échantillon aléatoire est abandonnée.

9.3.1 Absence de biais

Hypothèse 1 (linéarité dans les paramètres): Le processus stochastique $\{(x_{t1}, x_{t2}, \ldots, x_{tK}, y_t) : t = 1, \ldots, T\}$ se décrit par un modèle linéaire:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K z_{Kt} + u_t$$

où $\{u_t : t = 1, \ldots, T\}$ est une suite de variables aléatoires.

Remarque: Le modèle avec retard distribués présenté précédemment satisfait l'hypothèse 1, si $z_t = x_{1t}, z_{t-1} = x_{2t}, \ldots$

Hypothèse 2 (moyenne conditionnelle nulle): Pour tout t,

$$E(u_t|\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T) = 0.$$

Remarques:

(1) Cette hypothèse est plus forte que l'hypothèse alternative: $E(u_t|\mathbf{x}_t) = 0$, que l'on qualifie d'exogénéité contemporaine. L'hypothèse 2 est celle d'exogénéité stricte.

(2) L'hypothèse 2 remplace l'hypothèse d'échantillon aléatoire. Dans le cas d'un échantillon aléatoire, l'exogénéité "contemporaine" implique l'exogénéité structe.

(3) L'hypothèse 2 exclut la possibilité que des changements aujourd'hui dans

u ne causent des changements futurs dans z. Exemple:

$$\text{TXCRIM} = \beta_0 + \beta_1 \cdot \text{NB\_POL} + u_t$$

Hypothèse 3 (absence de colinéarité parfaite): Dans l'échantillon, les variables indépendantes ne doivent être ni constantes, ni des combinaisons linéaires des autres.

Théorème 1 (absence de biais des MCO): Sous les hypothèses 1-3, les estimateurs des MCO sont non biaisés: $E(\hat{\beta}_j) = \beta_j$.

9.3.1.1 La variance des estimateurs des MCO et le Théorème de Gauss-Markov

Deux hypothèses supplémentaires sont nécessaires.

Hypothèse 4 (homoscédasticité): Pour tout t: $var(u_t|\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T) = \sigma_u^2$

Hypothèse 5 (absence d'autocorrélation): Pour tout $t \neq s$,

$$\text{corr}(u_t, u_s|\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T)$$

Remarque: Dans un échantillon aléatoire, cette hypothèse est automatiquement satisfaite.

Théorème 2 (variance des estimateurs MCO): Sous les hypothèses 1-5, la variance des estimateurs des MCO $\hat{\beta}_j$ est:

$$\text{var}(\hat{\beta}_j) = \frac{\sigma_u^2}{(1 - R_j^2)\sum_{t=1}^{T}(x_{jt} - \bar{x}_j)},$$

où R_j^2 est R-carré obtenu de la régression de x_j sur l'ensemble des autres variables explicatives.

Théorème 3 (absence de biais de l'estimateur de σ_u^2): Sous les hypothèses 1-5, $\hat{\sigma}_u^2$ est un estimateur sans biais de σ_u^2: $E(\hat{\sigma}_u^2) = \sigma_u^2$.

Théorème 4 (Théorème de Gauss-Markov): Sous les hypothèses 1-5, les estimateurs des MCO sont les meilleurs estimateurs linéaire sans biais.

9.4 Inférence sous les hypothèses classiques

Hypothèse 6 (normalité des termes aléatoires): Les termes aléatoires u_t sont distribués selon une loi normale.

Remarque: Cette hypothèse, jointe aux hypothèses 2, 4 et 5, implique que les u_t sont indépendants de $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T$, indépendemment et identiquement selon une loi $\mathcal{N}(0, \sigma_u^2)$.

Théorème 5 (normalité des estimateurs MCO): Sous les hypothèses 1-6, les estimateurs des MCO sont distribués selon une loi normale . De plus, sous les hypothèseses nulles, les statistiques de Student ont une distribution de Student, et les statistiques de Fisher ont une distribution de Fisher.

9.4.1 Tendance et saisonalité

9.4.1.1 Caractérisation d'une série tendancielle

De nombreuses séries temporelles sont caractérisées par une tendance. Ignorer ce fait peut amener à mettre en évidence des relations de causalité qui n'existent pas.

Modèle de tendance linéaire:

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

Modèle de tendance exponentielle:

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t$$

Dans ce cas:

$$y_t = \exp\left(\alpha_0 + \alpha_1 t + e_t\right)$$

et

$$\alpha_1 = \Delta \log(y_t) \approx (\frac{y_t - y_{t-1}}{y_{t-1}}) = \text{ le taux de croissance de } y_t.$$

9.4.1.2 Utilisation de variables tendancielles dans une régression

Cela ne viole pas nécessairement les hypothèses 1-6. Cependant, on doit prendre en compte le fait que y_t peut également être influencé par une tendance (sinon problème de variable manquante et biais).

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 t + u_t$$

9.4.1.3 Détendancialisation

Définissons la tédencialisation par:

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$$

Considérons l'estimation du modèle précédent:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2 t$$

Alors, l'estimateur $\hat{\beta}_1$ peut être obtenu de la manière suivante:

(1) Régresser y_t et x_t sur t et une constante, et récupérer les résidus: $\ddot{y}_t$ et $\ddot{x}_t$. C'est-à-dire: $\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$. En un sens, $\ddot{y}_t$ est une série dont on a enlevé la tendance. En effet, $\ddot{y}_t$ est non corrélé avec t.

(2) Régresser $\ddot{y}_t$ sur $\ddot{x}_t$. Cela donne: $\hat{\beta}_1$. Donc, les estimateurs peuvent être obtenus par une régressions sur des données détendancialisées.

9.4.2 Calcul du R^2 quand la variable dépendante est tendancielle

En général, les R^2 obtenus sur séries temporelles sont très élevés. La raison est la suivante. La formule du R^2 ajusté est:

$$\bar{R}^2 = 1 - \frac{\sigma_u^2}{\hat{\sigma}_y^2}.$$

L'estimateur $\hat{\sigma}_u^2$ est non biaisé (si la tendance est introduite dans la régression). Mais, l'estimateur $\hat{\sigma}_y^2 = \frac{1}{N-1}\sum_{t=1}^{T}(y_t - \bar{y})^2$ est biaisé si y est tendanciel. Par exemple, si

$$y_t = \delta_0 + \delta_1 t + e_t.$$

Le principe est alors de calculer le R^2 à partir d'une série qui aura été préalablement détendancialisée, et donc de régresser $\ddot{y}_t$ sur x_t et t.

9.4.3 Utilisation de variables saisonnières dans une régression et désaisonnalisation

Si y_t est vraisemblablement influencé par la saison, le modèle suivant peut être estimée:

$$\ddot{y}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t - \hat{\delta}_1 \cdot \mathrm{jan}_t - \hat{\delta}_2 \cdot \mathrm{fév}_t - ... - \hat{\delta}_{11} \cdot \mathrm{nov}_t$$

Définissons la désaisonnalisation par:

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$$

L'estimateur β_1 peut être obtenu de la manière suivante:

(1) Régresser y_t et x_t sur une constante et les variables binaires mensuelles. On obtient: $\ddot{y}_t$ sur $\ddot{x}_t$, qui sont les résidus de ces régressions.

$$\ddot{y}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t - \hat{\delta}_1 \cdot \mathrm{jan}_t - \hat{\delta}_2 \cdot \mathrm{fév}_t - ... - \hat{\delta}_{11} \cdot \mathrm{nov}_t$$

Ces résidus sont des variables désaisonnalisées.

(2) Régresser $\ddot{y}_t$ sur $\ddot{x}_t$ donne l'estimateur $\hat{\beta}_1$.

10 Régression avec séries temporelles: Eléments de théorie asymptotique

10.1 Stationarité et dépendance faible

10.1.1 Séries stationaires et non stationaires

Définition 1 (stationnarité): Le processus stochastique $\{x_t : t = 1, 2, ...\}$ est stationnaire si pour toute collection d'indices $1 \leqslant t_1 < t_2 < ... < t_m$, la distribution jointe de $\{x_{t_1}, x_{t_2}, ..., x_{t_m}\}$ est la même que la distribution jointe de $\{x_{t_1+h}, x_{t_2+h}, ..., x_{t_m+h}\}$ pour tous les entiers $h \geqslant 1$.

Définition 2 (stationnarité en covariance): Le processus stochastique $\{x_t : t = 1, 2, ...\}$ avec des moments de second ordre finis ($E(x_t^2) < \infty$) est stationnaire en covariance si (a) $E(x_t)$ est constant; (b) $var(x_t)$ est constant; et (c) pour tout $t, h \geqslant 1$, $\text{cov}(x_t, x_{t+h})$ dépend seulement de h.

La dépendance faible peut maintenant être définie de manière intuitive de la manière suivante: un processus stochastique stationnaire en covariance est caractérisée par les corrélations entre x_t et x_{t+h}. Un processus stochastique stationnaire en covariance est faiblement dépendant si

$$\lim_{h \to \infty} \text{cov}(x_t, x_{t+h}) = 0.$$

Cette hypothèse remplace l'hypothèse d'échantillon aléatoire.

10.2 Propriétés asymptotiques des MCO

Hypothèse 1' (linéarité dans les paramètres): Le processus stochastique $\{(x_{t1}, x_{t2}, \ldots, x_{tK}, y_t) : t = 1, \ldots, T\}$ se décrit par un modèle linéaire:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K z_{Kt} + u_t$$

où $\{u_t : t = 1, \ldots, T\}$ est une suite de variables aléatoires. De plus, $\{(x_{t1}, x_{t2}, \ldots, x_{tK}, y_t) : t = 1, \ldots, T\}$ est un processus stochastique stationnaire en covariance et faiblement dépendant.

Hypothèse 2' (moyenne conditionnelle nulle): Pour tout t, $E(u_t|\mathbf{x}_t) = 0$.

Hypothèse 3' (absence de colinéarité parfaite): Dans l'échantillon, les variables indépendantes ne doivent être ni constantes, ni des combinaisons linéaires des autres.

Théorème 6 (convergence des MCO): Sous les hypothèses $1' - 3'$, les estimateurs des MCO sont non convergents: $\text{plim}\hat{\beta}_j = \beta_j$.

Hypothèse 4' (homoscédasticité): Pour tout t: $var(u_t|\mathbf{x}_t) = \sigma^2$

Hypothèse 5' (absence d'autocorrélation): Pour tout $t \neq s$, $corr(u_t, u_s|\mathbf{x}_t, \mathbf{x}_s) = 0$

Théorème 7 (normalité des estimateurs MCO): Sous les hypothèses $1' - 5'$, les estimateurs des MCO sont asymptotiquement distribués selon une loi normale . De plus, sous les hypothèseses nulles, les tests de Student, et les tests de Fisher sont asymptotiquement valides.

11 Autocorrélation

11.1 Propriétés des MCO en présence d'autocorrélation

Les estimateurs ne sont plus BLUE en présence d'autocorrélation. Considérons:

$$u_t = \rho u_{t-1} + e_t$$

avec $|\rho| < 1$ et

$$E(e_t|u_{t-1}, u_{t-2}, \ldots) = 0$$
$$var(e_t|u_{t-1}, u_{t-2}, \ldots) = \sigma_e^2$$

On suppose que:

$$E(u_t) = E(u_{t-1}) = 0.$$

Un processus autorégressif d'ordre 1.

$$var(u_t) = \rho^2 var(u_{t-1}) + var(e_t)$$

Donc

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_e^2$$
$$\sigma_u^2 = \frac{\sigma_e^2}{1-\rho^2}$$
$$E(u_t u_{t-1}) = \rho E(u_{t-1}^2) + E(e_t u_{t-1})$$
$$cov(u_t, u_{t-1}) = \rho \sigma_u^2$$
$$cov(u_t, u_{t-2}) = \rho^2 \sigma_u^2$$

Dans le modèle:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

Pour simplifier, on suppose que $\bar{x} = 0$. Les estimateurs des MCO sont:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum x_t u_t}{\sum x_t^2}$$

$$
\begin{aligned}
var\left(\hat{\beta}_1\right) &= var\left(\frac{\sum x_t u_t}{\sum x_t^2}\right) \\
&= \frac{var\left(\sum x_t u_t\right)}{\left(\sum x_t^2\right)^2} \\
&= \frac{\sum x_t^2 var\left(u_t\right) + 2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} x_t x_{t+j} cov\left(u_t, u_{t+j}\right)}{\left(\sum x_t^2\right)^2} \\
&= \frac{\sigma_u^2}{\sum x_t^2} + 2\frac{\sigma_u^2 \sum_{t=1}^{T-1}\sum_{j=1}^{T-t} x_t x_{t+j}\rho^j}{\left(\sum x_t^2\right)^2}
\end{aligned}
$$

La variance sera généralement biaisée. Si $\rho > 0$, et si les x sont autocorrélés, on sous-estimera la variance. Si $\rho < 0$, on ne peut pas tirer de conclusion générale.

11.2 Tests d'autocorrélation

$$
\begin{gathered}
u_t = \rho u_{t-1} + e_t \\
E(e_t | u_{t-1}, u_{t-2}, \ldots) = 0 \\
var(e_t | u_{t-1}, u_{t-2}, \ldots) = \sigma_e^2
\end{gathered}
$$

Hypothèse:

$$H_0 \equiv \rho = 0$$

11.2.0.1 Durbin-Watson

$$
\begin{aligned}
DW &= \frac{\sum_{t=2}^{T}\left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=1}^{T}\hat{u}_t^2} \\
&= \frac{\sum_{t=2}^{T}\hat{u}_t^2 + \sum_{t=2}^{T}\hat{u}_{t-1}^2 - 2\sum_{t=2}^{T}\hat{u}_t^2\hat{u}_{t-1}^2}{\sum_{t=1}^{T}\hat{u}_t^2} \\
&\simeq 2 \times (1 - \hat{\rho})
\end{aligned}
$$

If $DW < d_L$, rejet.

If $DW > d_U$, non rejet.

If $d_U > DW > d_L$, test non concluant.

Par exemple, pour un test à 5%, n=45, k= 4,$d_U = 1.720; d_L = 1.336$

11.3 Correction de l'autocorrélation

L'autocorrélation est connue.

On adopte les hypothèses traditionnelles de Gauss-Markov sauf l'hypothèse 5 sur l'autocorrélation.

$$u_t = \rho u_{t-1} + e_t$$
$$var(u_t) = \frac{\sigma_e^2}{(1-\rho^2)}$$
$$y_t = \beta_0 + \beta_1 x_t + u_t$$

On écrit:

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1}$$
$$y_t = \beta_0 + \beta_1 x_t + u_t$$

D'où,

$$y_t - \rho y_{t-1} = (1-\rho)\beta_0 + \beta_1 (x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1})$$

ou

$$\tilde{y}_t = (1-\rho)\beta_0 + \beta_1 \tilde{x}_t + e_t$$

Cet estimateur n'est pas BLUE. Pour cela, on use:

$$y_1 = \beta_0 + \beta_1 x_1 + u_1$$

où u_1 est non corrélé avec e. mis

$$var(u_1) = \frac{\sigma_e^2}{(1-\rho^2)}$$

$$\sqrt{(1-\rho^2}y_1 = \beta_0\sqrt{(1-\rho^2} + \beta_1\sqrt{(1-\rho^2}x_1 + \sqrt{(1-\rho^2}u_1$$

où

$$\tilde{y}_1 = \beta_0\sqrt{(1-\rho^2} + \beta_1\sqrt{(1-\rho^2}\tilde{x}_1 + \tilde{u}_1$$
$$var(\tilde{u}_1) = (1-\rho^2)var(u_1) = \sigma_e^2$$

Estimateur GLS

FGLS: Cochrane-Orcutt, Prais-Winsten + itération

11.4 Inférence robuste à l'autocorrélation

11.5 Annexe

Les formules suivantes sont régulièrement utilisées dans le cours:

$$\begin{aligned}
\sum_{i=1}^{N} x_i(x_i-\bar{x}) &= \sum_{i=1}^{N} x_i(x_i-\bar{x}) - \bar{x}\left(\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i\right) \\
&= \sum_{i=1}^{N} x_i(x_i-\bar{x}) - \bar{x}\left(\sum_{i=1}^{N} x_i - N\bar{x}\right) \\
&= \sum_{i=1}^{N} x_i(x_i-\bar{x}) - \bar{x}\left(\sum_{i=1}^{N} (x_i-\bar{x})\right) \\
&= \sum_{i=1}^{N} \left(x_i(x_i-\bar{x}) - \bar{x}\,(x_i-\bar{x})\right) \\
&= \sum_{i=1}^{N} (x_i-\bar{x})^2 \\
\sum_{i=1}^{N} x_i(y_i-\bar{y}) &= \sum_{i=1}^{N} x_i(y_i-\bar{y}) - \bar{x}\left(\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} y_i\right) \\
&= \sum_{i=1}^{N} x_i(y_i-\bar{y}) - \bar{x}\left(\sum_{i=1}^{N} y_i - N\bar{y}\right) \\
&= \sum_{i=1}^{N} x_i(y_i-\bar{y}) - \bar{x}\left(\sum_{i=1}^{N} (y_i-\bar{y})\right) \\
&= \sum_{i=1}^{N} \left(x_i(y_i-\bar{y}) - \bar{x}\,(y_i-\bar{y})\right) \\
&= \sum_{i=1}^{N} (x_i-\bar{x})\,(y_i-\bar{y})
\end{aligned}$$