Méthodes d'analyse empirique

Partie Quantitative

Michel Beine (suppl. S. Laurent)

michel.beine@uni.lu

Université du Luxembourg

http://www.michelbeine.be

Manuel

- Ce cours est basé sur le livre Introductory Econometrics, a Modern Approach (second or third editions) de *Jeffrey M. Wooldridge*, édition Thomson South-Western.

Manuel

- Ce cours est basé sur le livre Introductory Econometrics, a Modern Approach (second or third editions) de *Jeffrey M. Wooldridge*, édition Thomson South-Western.
- Logiciel économétrique : Ox et PcGive.

Manuel

- Ce cours est basé sur le livre Introductory Econometrics, a Modern Approach (second or third editions) de *Jeffrey M. Wooldridge*, édition Thomson South-Western.
- Logiciel économétrique : Ox et PcGive.
- Evaluation: examen à livre fermé.

Plan du cours

- Chapitre 17. Modèles de choix discret et maximum de vraissemblance.

Plan du cours

- Chapitre 17. Modèles de choix discret et maximum de vraissemblance.
- Chapitre 15. Variables instrumentales et MCO en 2 étapes.

Plan du cours

- Chapitre 17. Modèles de choix discret et maximum de vraissemblance.
- Chapitre 15. Variables instrumentales et MCO en 2 étapes.
- Chapitre 16. Modèles à équations simultanées.

Plan du cours

- Chapitre 17. Modèles de choix discret et maximum de vraissemblance.
- Chapitre 15. Variables instrumentales et MCO en 2 étapes.
- Chapitre 16. Modèles à équations simultanées.
- Les processus modérateurs et médiateurs, ANOVA, ...

Maximum de vraissemblance et modèles à choix discrets

Motivation

- Le maximum de vraisemblance (MV) est une méthode d'estimation utile quand les paramètres d'un modèle ne peuvent pas être estimés par MCO. Des tests de restrictions sont également disponibles (principe du test de rapport des vraisemblances).

Motivation

- Le maximum de vraisemblance (MV) est une méthode d'estimation utile quand les paramètres d'un modèle ne peuvent pas être estimés par MCO. Des tests de restrictions sont également disponibles (principe du test de rapport des vraisemblances).
- La méthode nécessite la spécification entière de la distribution de probabilité de l'échantillon observé, pas uniquement la moyenne conditionnelle de l'échantillon (tel le modèle de régression linéaire estimé par MCO).

Motivation

- Le maximum de vraisemblance (MV) est une méthode d'estimation utile quand les paramètres d'un modèle ne peuvent pas être estimés par MCO. Des tests de restrictions sont également disponibles (principe du test de rapport des vraisemblances).
- La méthode nécessite la spécification entière de la distribution de probabilité de l'échantillon observé, pas uniquement la moyenne conditionnelle de l'échantillon (tel le modèle de régression linéaire estimé par MCO).
- L'estimation par MV du modèle de régression linéaire demande de spécifier aussi la structure de la dépendance de l'échantillon via les erreurs (i.i.d. par exemple) et la distribution du terme d'erreurs (comme $\epsilon_t \sim N(0, \sigma^2)$, mais une autre distribution est possible).

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.
- Ceci est un exemple d'expérience.

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.
- Ceci est un exemple d'expérience.
- Avant de lancer la pièce, nous savons que le nombre de piles sera un nombre de 0 à 10.

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.
- Ceci est un exemple d'expérience.
- Avant de lancer la pièce, nous savons que le nombre de piles sera un nombre de 0 à 10.
- Une variable aléatoire est une variable qui prend des valeurs numériques et a un résultat qui est déterminé par une expérience.

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.
- Ceci est un exemple d'expérience.
- Avant de lancer la pièce, nous savons que le nombre de piles sera un nombre de 0 à 10.
- Une variable aléatoire est une variable qui prend des valeurs numériques et a un résultat qui est déterminé par une expérience.
- $\longrightarrow$ le nombre de fois que l'on obtient pile est un exemple de variable aléatoire.

Variable aléatoire

- Supposons que nous lançons une pièce 10 fois et comptons le nombre de fois que nous obtenons pile.
- Ceci est un exemple d'expérience.
- Avant de lancer la pièce, nous savons que le nombre de piles sera un nombre de 0 à 10.
- Une variable aléatoire est une variable qui prend des valeurs numériques et a un résultat qui est déterminé par une expérience.
- $\rightarrow$ le nombre de fois que l'on obtient pile est un exemple de variable aléatoire.
- $\rightarrow$ 7 piles sur 10 est un exemple de réalisation de cette variable aléatoire pour une expérience donnée.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.
- Nous notons une valeur particulière que peut prendre cette variable aléatoire par une minuscule: x. Exemple: $x = 6$.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.
- Nous notons une valeur particulière que peut prendre cette variable aléatoire par une minuscule: x. Exemple: $x = 6$.
- Notons que le lancer d'une pièce de monnaie peut être considéré comme une variable aléatoire suivant une distribution de Bernoulli où $X = 1$ est appelé succes et $X = 0$ échec.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.
- Nous notons une valeur particulière que peut prendre cette variable aléatoire par une minuscule: x. Exemple: $x = 6$.
- Notons que le lancer d'une pièce de monnaie peut être considéré comme une variable aléatoire suivant une distribution de Bernoulli où $X = 1$ est appelé succes et $X = 0$ échec.
- Une variable aléatoire de Bernoulli est le plus simple exemple de variable aléatoire discrète.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.
- Nous notons une valeur particulière que peut prendre cette variable aléatoire par une minuscule: x. Exemple: $x = 6$.
- Notons que le lancer d'une pièce de monnaie peut être considéré comme une variable aléatoire suivant une distribution de Bernoulli où $X = 1$ est appelé succes et $X = 0$ échec.
- Une variable aléatoire de Bernoulli est le plus simple exemple de variable aléatoire discrète.
- La seule chose que nous devons décrire pour une telle variable est la probabilité de succes: $P(X = 1)$.

Variable aléatoire

- Nous notons une variable aléatoire par une lettre majuscule: X.
- Nous notons une valeur particulière que peut prendre cette variable aléatoire par une minuscule: x. Exemple: $x = 6$.
- Notons que le lancer d'une pièce de monnaie peut être considéré comme une variable aléatoire suivant une distribution de Bernoulli où $X = 1$ est appelé succes et $X = 0$ échec.
- Une variable aléatoire de Bernoulli est le plus simple exemple de variable aléatoire discrète.
- La seule chose que nous devons décrire pour une telle variable est la probabilité de succes: $P(X = 1)$.
- Dans le cas d'une pièce équilibrée, $P(X = 1) = 0.5$ et comme $P(X = 1) + P(X = 0) = 1, P(X = 0) = 0.5$.

Variable aléatoire discrère

- Il n’y a pas de raison de penser que la probabilité de succes de toute variable binaire (0-1) est 0.5.

Variable aléatoire discrère

- Il n'y a pas de raison de penser que la probabilité de succes de toute variable binaire (0-1) est 0.5.
- De manière générale disons que $P(X=1)=\theta$ et comme $P(X=1)+P(X=0)=1, P(X=0)=1-\theta$.

Variable aléatoire discrère

- Il n'y a pas de raison de penser que la probabilité de succes de toute variable binaire (0-1) est 0.5.
- De manière générale disons que $P(X = 1) = \theta$ et comme $P(X = 1) + P(X = 0) = 1, P(X = 0) = 1 - \theta$.
- Comment estimer θ sur base d'un échantillon, c-à-d une expérience ?

Variable aléatoire discrère

- Il n'y a pas de raison de penser que la probabilité de succes de toute variable binaire (0-1) est 0.5.
- De manière générale disons que $P(X = 1) = \theta$ et comme $P(X = 1) + P(X = 0) = 1, P(X = 0) = 1 - \theta$.
- Comment estimer θ sur base d'un échantillon, c-à-d une expérience ?
- Plus généralement encore si X peut prendre des valeurs $\{x_1, x_2, \ldots, x_k\}$, alors les probabilités $p_1, p_2, \ldots, p_k$ sont définies par

  $$p_j = P(X = x_j), j = 1, 2, \ldots, k$$

  avec $p_1 + p_2 + \ldots + p_k = 1$.

Fonction de densité de probabilité

- La fonction de densité de probabilité (pdf) de X résume l'information concernant les différents résultats possibles de X et ses probabilités correspondantes:

  $$f(x_j) = p_j, j = 1, 2, \ldots, k$$

  avec $f(x) = 0$ pour toutes les valeurs de x hors de $\{x_1, x_2, \ldots, x_k\}$.

Fonction de densité de probabilité

- La fonction de densité de probabilité (pdf) de X résume l'information concernant les différents résultats possibles de X et ses probabilités correspondantes:

$$f(x_j) = p_j, j = 1, 2, \ldots, k$$

avec $f(x) = 0$ pour toutes les valeurs de x hors de $\{x_1, x_2, \ldots, x_k\}$.

- La fonction de distribution jointe de n variables aléatoires discrètes $X_j, j = 1, 2, \ldots, n$ est notée $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$.

Fonction de densité de probabilité

- La fonction de densité de probabilité (pdf) de X résume l'information concernant les différents résultats possibles de X et ses probabilités correspondantes:

$$f(x_j) = p_j, j = 1, 2, \ldots, k$$

avec $f(x) = 0$ pour toutes les valeurs de x hors de $\{x_1, x_2, \ldots, x_k\}$.

- La fonction de distribution jointe de n variables aléatoires discrètes $X_j, j = 1, 2, \ldots, n$ est notée $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$.
- Ces n variables aléatoires sont indépendantes si $f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$.

Fonction de densité de probabilité

- La fonction de densité de probabilité (pdf) de X résume l'information concernant les différents résultats possibles de X et ses probabilités correspondantes:

$$f(x_j) = p_j, j = 1, 2, \ldots, k$$

avec $f(x) = 0$ pour toutes les valeurs de x hors de $\{x_1, x_2, \ldots, x_k\}$.

- La fonction de distribution jointe de n variables aléatoires discrètes $X_j, j = 1, 2, \ldots, n$ est notée $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$.
- Ces n variables aléatoires sont indépendantes si $f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$.
- Ceci s'étend évidemement aux variables aléatoires continues.

Fonction de densité de probabilité

- La fonction de densité de probabilité (pdf) de X résume l'information concernant les différents résultats possibles de X et ses probabilités correspondantes:

$$f(x_j) = p_j, j = 1, 2, \ldots, k$$

avec $f(x) = 0$ pour toutes les valeurs de x hors de $\{x_1, x_2, \ldots, x_k\}$.

- La fonction de distribution jointe de n variables aléatoires discrètes $X_j, j = 1, 2, \ldots, n$ est notée $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$.
- Ces n variables aléatoires sont indépendantes si $f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$.
- Ceci s'étend évidemement aux variables aléatoires continues.
- Un exemple courant est la distribution normale.

MV pour une régression linéaire

- Posons $y_t = \beta' x_t + \epsilon_t, \ (t = 1, \ldots, n)$ avec $\epsilon_t \sim$ i.N$(0, \sigma^2)$.
- Par l'indépendance, la distribution jointe des ϵ_t est

$$f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n | \beta, \sigma^2) = \prod_{t=1}^{n} f(\epsilon_t | \beta, \sigma^2), \tag{1}$$

où $f(\epsilon_t|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\epsilon_t^2}{2\sigma^2})$.

- ϵ_t n'étant pas observé, on peut le remplacer par $y_t - \beta' x_t$. Autrement dit, $y_t | x_t \sim N(\beta' x_t, \sigma^2)$, et la distribution de probabilité jointe de l'échantillon est :

$$f(y_1, y_2, \ldots, y_n | \beta, \sigma^2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(y_t - \beta' x_t)^2}{2\sigma^2}\Big). \tag{2}$$

Fonction de vraisemblance : idée

- La fonction de vraisemblance (FV) d'un échantillon $y_1, y_2, \ldots, y_n$ est la distribution de probabilité jointe de cet échantillon, vue comme fonction des paramètres du modèle (β et σ^2).

Fonction de vraisemblance : idée

- La fonction de vraisemblance (FV) d'un échantillon $y_1, y_2, \ldots, y_n$ est la distribution de probabilité jointe de cet échantillon, vue comme fonction des paramètres du modèle (β et σ^2).
- Pour une valeur donnée de β et σ^2, on peut calculer (2) : le résultat nous donne la "probabilité" d'observer l'échantillon en question $y_1, y_2, \ldots, y_n$ pour une valeur donnée des paramètres.

Fonction de vraisemblance : idée

- La fonction de vraisemblance (FV) d'un échantillon $y_1, y_2, \ldots, y_n$ est la distribution de probabilité jointe de cet échantillon, vue comme fonction des paramètres du modèle (β et σ^2).
- Pour une valeur donnée de β et σ^2, on peut calculer (2) : le résultat nous donne la "probabilité" d'observer l'échantillon en question $y_1, y_2, \ldots, y_n$ pour une valeur donnée des paramètres.
- Pour une valeur différente de β et σ^2, on obtiendra une probabilité différente.

Fonction de vraisemblance : idée

- La fonction de vraisemblance (FV) d'un échantillon $y_1, y_2, \ldots, y_n$ est la distribution de probabilité jointe de cet échantillon, vue comme fonction des paramètres du modèle (β et σ^2).
- Pour une valeur donnée de β et σ^2, on peut calculer (2) : le résultat nous donne la "probabilité" d'observer l'échantillon en question $y_1, y_2, \ldots, y_n$ pour une valeur donnée des paramètres.
- Pour une valeur différente de β et σ^2, on obtiendra une probabilité différente.
- L'**ESTIMATEUR MV (EMV)** est la valeur des paramètres qui maximise la FV; càd que ce sont les valeurs de β et σ^2 pour lesquelles l'échantillon à la plus grande probabilité d'être observé.

Fonction de vraisemblance : notations

- L'expression $L(\beta, \sigma^2 | y_1, y_2, \ldots, y_n)$ désigne la FV, où les arguments de la fonction sont β et σ^2, et non les données observées $y_1, y_2, \ldots, y_n$. Formellement :

$$L(\beta, \sigma^2 | y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n | \beta, \sigma^2). \tag{3}$$

  $L(\beta, \sigma^2 | y_1, y_2, \ldots, y_n) : \mathbb{R}^k \times (0, \infty) \to \mathbb{R}_+$, alors que $f(y_1, y_2, \ldots, y_n | \beta, \sigma^2) : \mathbb{R}^n \to \mathbb{R}_+$.

- Quand on peut identifier la nature exacte des données, on écrit même $L(\beta, \sigma^2)$ au lieu de $L(\beta, \sigma^2 | y_1, y_2, \ldots, y_n)$.
- On utilise aussi les notations plus synthétiques et génériques $L(\theta)$ et $L(\theta | y)$: θ pour les paramètres (donc $\theta = (\beta', \ \sigma^2)$ dans une régression), et y pour l'échantillon de taille n.

Etapes de l’estimation MV

Estimer les paramètres d’un modèle par MV demande deux étapes :

- étape 1 : obtenir l’expression de la FV à partir de la formulation du modèle et des hypothèses.

Etapes de l'estimation MV

Estimer les paramètres d'un modèle par MV demande deux étapes :

- étape 1 : obtenir l'expression de la FV à partir de la formulation du modèle et des hypothèses.
- étape 2 : trouver la valeur des paramètres maximisant cette fonction.

  Il faut donc solutionner le problème de maximisation :

$$\max_{\theta \in \Theta} L(\theta | y),$$

  où Θ est l'ensemble des valeurs des paramètres acceptables θ.

Solutionner la maximisation

- Il est beaucoup plus aisé, et équivalent de résoudre

$$\max_{\theta \in \Theta} l(\theta|y),$$

où $l(\theta|y) = \log L(\theta|y)$ (logarithme népérien). Puisque $L(\theta|y) = \prod_{t=1}^{n} f(y_t|\theta)$, l'expression donne en logarithme $l(\theta|y) = \sum_{t=1}^{n} \log f(y_t|\theta) = \sum_{t=1}^{n} l_t(\theta)$.

Solutionner la maximisation

- Il est beaucoup plus aisé, et équivalent de résoudre

$$\max_{\theta \in \Theta} l(\theta|y),$$

où $l(\theta|y) = \log L(\theta|y)$ (logarithme népérien). Puisque $L(\theta|y) = \prod_{t=1}^{n} f(y_t|\theta)$, l'expression donne en logarithme $l(\theta|y) = \sum_{t=1}^{n} \log f(y_t|\theta) = \sum_{t=1}^{n} l_t(\theta)$.

- Pour l'exemple de la régression linéaire, on doit résoudre

$$\max_{\beta \in \mathbb{R}^k, \sigma^2 > 0} \left(-\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - \beta' x_t)^2 \right),$$

où le terme $-n \log(\sqrt{2\pi})$ fut négligé car il ne dépend ni de β ni de σ^2.

Conditions de premier ordre

- Si $l(\theta)$ est différentiable, θ, l'EMV, est la solution du système des conditions de premier ordre (CPO)

$$q(\theta) = 0 \quad \text{où} \quad q(\theta) = \frac{\partial l(\theta)}{\partial \theta},$$

  pour autant que les conditions de second ordre sont satisfaites, càd que la matrice des dérivées secondes, évaluée à la solution des CPO, est définie négative.
  NB: $q(\theta)$ est appelé la fonction score.

- Appelons $\hat{\theta}$ la solution des CPO : alors, $\hat{\theta}$ est l'EMV si

$$q(\hat{\theta}) = 0 \quad \text{et} \quad H(\hat{\theta}) < 0.$$

  où $H(\hat{\theta})$ est la matrice Hessienne $\frac{\partial l(\theta)}{\partial \theta \partial \theta'}$ évaluée à $\hat{\theta}$.

CPO pour β

- Dans l'exemple de la régression linéaire, les CPO sont

$$\frac{1}{\sigma^2}\sum_{t=1}^{n} x_t(y_t - \beta' x_t) = 0 \quad (k \text{ équations}) \tag{4}$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{t=1}^{n}(y_t - \beta' x_t)^2 = 0 \quad (1 \text{ equation}) \tag{5}$$

- La solution de (4) ne dépend pas de σ^2 et est la solution de $\sum_{t=1}^{n} x_t(y_t - \beta' x_t) = 0$: c'est le système des "équations normales" de l'estimation MCO. Donc, $\hat{\beta} = b$: l'EMV, en supposant la normalité du terme d'erreurs, est identique à celui des MCO.

CPO pour σ^2

- Ayant résolu (4), il reste à solutionner

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{t=1}^{n} (y_t - \hat{\beta}' x_t)^2 = 0,$$

qui donne

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{\beta}' x_t)^2 = \frac{1}{n} \sum_{t=1}^{n} e_t^2. \tag{6}$$

- L'EMV de σ^2 n'est pas l'estimateur habituel

$$s^2 = \tfrac{1}{n-k} \textstyle\sum_{t=1}^{n} e_t^2,$$

mais si n est grand, la différence est petite.

Conditions de second ordre

- Les calculs donnent

$$H(\hat{\beta}, \hat{\sigma}^2) = \begin{pmatrix} -\frac{\sum_{t=1}^{n} x_t x_t'}{\hat{\sigma}^2} & 0 \\ 0' & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix} = \begin{pmatrix} -\frac{X'X}{\hat{\sigma}^2} & 0 \\ 0' & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix},$$

qui est définie négative puisque $X'X$ est définie positive (si X est de rang k).

Conditions de second ordre

- Les calculs donnent

$$H(\hat{\beta},\hat{\sigma}^2) = \begin{pmatrix} -\frac{\sum_{t=1}^{n} x_t x_t'}{\hat{\sigma}^2} & 0 \\ 0' & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix} = \begin{pmatrix} -\frac{X'X}{\hat{\sigma}^2} & 0 \\ 0' & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix},$$

  qui est définie négative puisque $X'X$ est définie positive (si X est de rang k).

- b et $\hat{\sigma}^2$ sont les EMV de β et σ^2.

Propriétés de l'EMV

- On note l'EMV de θ par $\hat{\theta}_n$ pour insister sur sa dépendance à la taille de l'échantillon. Les propriétés générales sont vérifiées asymptotiquement (càd qu'on laisse n tendre vers ∞). Pour rappel, $\hat{\theta}_n$ à une distribution d'échantillon puisqu'il dépend d'un échantillon de v.a. qui suivent un processus stochastique (le PGD).
NB : les propriétés exactes en échantillon fini dépendent du modèle considéré et sont obtenues sur une base de cas par cas quand c'est possible.

Propriétés de l'EMV

- On note l'EMV de θ par $\hat{\theta}_n$ pour insister sur sa dépendance à la taille de l'échantillon. Les propriétés générales sont vérifiées asymptotiquement (càd qu'on laisse n tendre vers ∞). Pour rappel, $\hat{\theta}_n$ à une distribution d'échantillon puisqu'il dépend d'un échantillon de v.a. qui suivent un processus stochastique (le PGD).
  NB : les propriétés exactes en échantillon fini dépendent du modèle considéré et sont obtenues sur une base de cas par cas quand c'est possible.
- Les propriétés asymptotiques générales sont satisfaites sous un ensemble de "conditions de régularité". Elles assurent soit l'existence de l'estimateur, soit la convergence et la normalité asymptotique. Pour les séries temporelles, ces conditions sont du type : le DGP est stationnaire et faiblement dépendant.

Propriétés de l'EMV

- Convergence : $\hat{\theta}_n \xrightarrow{p} \theta$ où θ est la valeur du paramètre dans le PGD (appelé "vraie valeur").

Propriétés de l'EMV

- Convergence : $\hat{\theta}_n \xrightarrow{p} \theta$ où θ est la valeur du paramètre dans le PGD (appelé "vraie valeur").
- Normalité asymptotique : on procède comme si $\hat{\theta}_n \sim N(\theta, \hat{V})$ si n est grand, sachant que plus n est grand, meilleure est l'approximation.

Propriétés de l'EMV

- Convergence : $\hat{\theta}_n \xrightarrow{p} \theta$ où θ est la valeur du paramètre dans le PGD (appelé "vraie valeur").
- Normalité asymptotique : on procède comme si $\hat{\theta}_n \sim N(\theta, \hat{V})$ si n est grand, sachant que plus n est grand, meilleure est l'approximation.
- $\hat{V}$ est un estimateur de la matrice de variance-covariances exacte V car cette dernière dépend généralement de paramètres inconnus (en particulier, mais pas exclusivement, du vrai θ). Il n'existe pas de choix unique pour $\hat{V}$ mais les choix habituels (slide suivant) sont fournis dans les logiciels économétriques (comme OxMetrics, TSP, EVIEWS, RATS, SAS, STATA, etc.).

Formules pour $\hat{V}$

- Basée sur la matrice Hessienne :

$$\hat{V}_H = -[H(\hat{\theta})]^{-1} = -[\sum_{t=1}^{n} H_t(\hat{\theta})]^{-1},$$

où $H_t(\hat{\theta}) = \frac{\partial l_t(\theta)}{\partial\theta\partial\theta'}$ est évalué à $\hat{\theta}$.

Formules pour $\hat{V}$

- Basée sur la matrice Hessienne :

$$\hat{V}_H = -[H(\hat{\theta})]^{-1} = -[\sum_{t=1}^{n} H_t(\hat{\theta})]^{-1},$$

où $H_t(\hat{\theta}) = \frac{\partial l_t(\theta)}{\partial\theta\partial\theta'}$ est évalué à $\hat{\theta}$.

- Basée sur la contribution aux scores :

$$\hat{V}_G = [\sum_{t=1}^{n} q_t(\hat{\theta})q_t(\hat{\theta})']^{-1},$$

où $q_t(\hat{\theta}) = \frac{\partial l_t(\theta)}{\partial\theta}$ est évalué à $\hat{\theta}$ et constitue la contribution de l'observation t à la fonction score.
NB: $q(\theta) = \sum_{t=1}^{n} q_t(\theta)$.

Test de Wald avec l'EMV

- Etant donné que $\hat{\theta}_n \sim N(\theta, \hat{V})$, on peut tester des restrictions (linéaires ou non) par des tests de Wald en χ^2. Par exemple :
  $H_0 : R\theta = r$ est testé par la statistique
  $(R\hat{\theta} - r)'(R\hat{V}R')^{-1}(R\hat{\theta} - r) \sim \chi^2(m)$ sous H_0.

Test de Wald avec l'EMV

- Etant donné que $\hat{\theta}_n \sim N(\theta, \hat{V})$, on peut tester des restrictions (linéaires ou non) par des tests de Wald en χ^2. Par exemple :
  $H_0 : R\theta = r$ est testé par la statistique
  $(R\hat{\theta} - r)'(R\hat{V}R')^{-1}(R\hat{\theta} - r) \sim \chi^2(m)$ sous H_0.
- La statistique de test pour $H_0 : \theta_3 = 0$ est $\hat{\theta}_3^2/\hat{v}_{33} \sim \chi^2(1)$. Elle peut être remplacée par $\hat{\theta}_3/\sqrt{\hat{v}_{33}} \sim N(0,1)$ si on désire un test unilatéral.

Test du rapport des vraisemblances

Sous les conditions de régularité, une manière alternative de tester les restrictions, disons $H_0 : R\theta = r$ existe :

- Etape 1 : estimer le modèle par MV sans imposer les contraintes. Sauver $l(\hat{\theta})$, la valeur de la FV à l'optimum.

Test du rapport des vraisemblances

Sous les conditions de régularité, une manière alternative de tester les restrictions, disons $H_0 : R\theta = r$ existe :

- Etape 1 : estimer le modèle par MV sans imposer les contraintes. Sauver $l(\hat{\theta})$, la valeur de la FV à l'optimum.
- Etape 2 : estimer le modèle (par MV) en imposant les restrictions. Sauver $l(\tilde{\theta})$, la valeur de la FV à l'optimum contraint. Donc, $\tilde{\theta} = \arg\max_\theta l(\theta)$ sous contraintes $R\theta = r$.

Test du rapport des vraisemblances

Sous les conditions de régularité, une manière alternative de tester les restrictions, disons $H_0 : R\theta = r$ existe :

- Etape 1 : estimer le modèle par MV sans imposer les contraintes. Sauver $l(\hat{\theta})$, la valeur de la FV à l'optimum.
- Etape 2 : estimer le modèle (par MV) en imposant les restrictions. Sauver $l(\tilde{\theta})$, la valeur de la FV à l'optimum contraint. Donc, $\tilde{\theta} = \arg\max_\theta l(\theta)$ sous contraintes $R\theta = r$.
- Calculer le rapport des vraisemblances (RV)

$$LR = 2[l(\hat{\theta}) - l(\tilde{\theta})].$$

Rejeter H_0 au seuil α si la p-valeur du RV est inférieure à α, où la p-valeur est obtenue en utilisant la $\chi^2(m)$. Par construction, $RV \geq 0$.

Problèmes possibles avec MV

- Résoudre analytiquement les CPO, comme dans l'exemple de la régression linéaire, n'est généralement pas possible. Des méthodes numériques sont implémentées dans les logiciels.

Problèmes possibles avec MV

- Résoudre analytiquement les CPO, comme dans l'exemple de la régression linéaire, n'est généralement pas possible. Des méthodes numériques sont implémentées dans les logiciels.
- L'EMV n'est pas toujours unique : le système des CPO peut avoir plusieurs solutions. Si ce problème est suspecté, on doit chercher numériquement plusieurs maxima et choisir le global.

Problèmes possibles avec MV

- Résoudre analytiquement les CPO, comme dans l'exemple de la régression linéaire, n'est généralement pas possible. Des méthodes numériques sont implémentées dans les logiciels.
- L'EMV n'est pas toujours unique : le système des CPO peut avoir plusieurs solutions. Si ce problème est suspecté, on doit chercher numériquement plusieurs maxima et choisir le global.
- Cas extrême : l'EMV n'existe pas (ou infinité de solutions : problème d'identification). Exemple : la régression linéaire quand X n'est pas de rang plein.

Problèmes possibles avec MV

- Résoudre analytiquement les CPO, comme dans l'exemple de la régression linéaire, n'est généralement pas possible. Des méthodes numériques sont implémentées dans les logiciels.
- L'EMV n'est pas toujours unique : le système des CPO peut avoir plusieurs solutions. Si ce problème est suspecté, on doit chercher numériquement plusieurs maxima et choisir le global.
- Cas extrême : l'EMV n'existe pas (ou infinité de solutions : problème d'identification). Exemple : la régression linéaire quand X n'est pas de rang plein.
- L'EMV peut être à la frontière de l'espace des paramètres. $\Rightarrow$ Normalité asymptotique non valable $\Rightarrow$ RV non asymptotiquement distribué selon une χ^2.

Forces et faiblesses du MV

- Forces : si les hypothèses sont correctes, l'EMV est l'estimateur le plus efficient asymptotiquement, au sens qu'il est non-biaisé asymptotiquement et possède la plus petite variance. Intuitivement, MV utilise au mieux l'information des données pour estimer les paramètres.

Forces et faiblesses du MV

- Forces : si les hypothèses sont correctes, l'EMV est l'estimateur le plus efficient asymptotiquement, au sens qu'il est non-biaisé asymptotiquement et possède la plus petite variance. Intuitivement, MV utilise au mieux l'information des données pour estimer les paramètres.
- Faiblesses : MV nécessite de choisir la distribution de probabilité. Quelles sont donc les conséquences d'une hypothèse erronée sur la distribution de probabilité des y_t?

Forces et faiblesses du MV

- Forces : si les hypothèses sont correctes, l'EMV est l'estimateur le plus efficient asymptotiquement, au sens qu'il est non-biaisé asymptotiquement et possède la plus petite variance. Intuitivement, MV utilise au mieux l'information des données pour estimer les paramètres.
- Faiblesses : MV nécessite de choisir la distribution de probabilité. Quelles sont donc les conséquences d'une hypothèse erronée sur la distribution de probabilité des y_t?
- La réponse à cette question est :
  - dans certains cas, l'EMV n'est même pas convergent,
  - dans d'autres cas, il est toujours convergent et asymptotiquement normal mais la matrice $\hat{V}$ est différente de celle fournie précédemment.

Modèles à choix discrets

Modèles Logit et Probit

- $y_i = 0$ si le ménage *i* ne possède pas de voiture,
  $y_i = 1$ sinon.

Modèles Logit et Probit

- $$\begin{array}{ll} y_i = 0 & \text{si le ménage } i \text{ ne possède pas de voiture,} \\ y_i = 1 & \text{sinon.} \end{array}$$

- Dans ce cas un modèle linéaire n'est pas approprié pour étudier les déterminants de y.

Modèles Logit et Probit

- $$\begin{array}{ll} y_i = 0 & \text{si le ménage } i \text{ ne possède pas de voiture,} \\ y_i = 1 & \text{sinon.} \end{array}$$
- Dans ce cas un modèle linéaire n'est pas approprié pour étudier les déterminants de y.
- Dans un modèle de choix discret, l'intérêt se porte sur l'estimation de

$$P(y = 1|\boldsymbol{x}) = P(y = 1|x_1, x_2, \ldots, x_k),$$

où par exemple x_1 est le niveau d'éducation, x_2 le nombre d'enfants, etc.

Modèles Logit et Probit

- Pour s'assurer que cette probabilité est toujours entre 0 et 1, on considère le plus souvent une classe de modèles du type suivant:

  $$P(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) = G(\beta_0 + \boldsymbol{xB}),$$

  où G est une fonction prenant des valeurs entre 0 et 1, c-à-d $0 < G(z) < 1$ pour tout nombre réel z.

Modèles Logit et Probit

- Pour s'assurer que cette probabilité est toujours entre 0 et 1, on considère le plus souvent une classe de modèles du type suivant:

  $$P(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) = G(\beta_0 + \boldsymbol{xB}),$$

  où G est une fonction prenant des valeurs entre 0 et 1, c-à-d $0 < G(z) < 1$ pour tout nombre réel z.

- Modèle logit: $G(z) = \frac{exp(z)}{1+exp(z)} = \Lambda(z)$, c-à-d la fonction cumulative d'une variable aléatoire logistique de variance unitaire.

Modèles Logit et Probit

- Pour s'assurer que cette probabilité est toujours entre 0 et 1, on considère le plus souvent une classe de modèles du type suivant:

  $$P(y=1|\boldsymbol{x}) = G(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k) = G(\beta_0+\boldsymbol{xB}),$$

  où G est une fonction prenant des valeurs entre 0 et 1, c-à-d $0 < G(z) < 1$ pour tout nombre réel z.
- Modèle logit: $G(z) = \frac{exp(z)}{1+exp(z)} = \Lambda(z)$, c-à-d la fonction cumulative d'une variable aléatoire logistique de variance unitaire.
- Modèle Probit: $G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)dv$, où $\phi(z)$ est la densité normale standard $\phi(z) = (2\pi)^{-1/2} exp(-z^2/2)$.

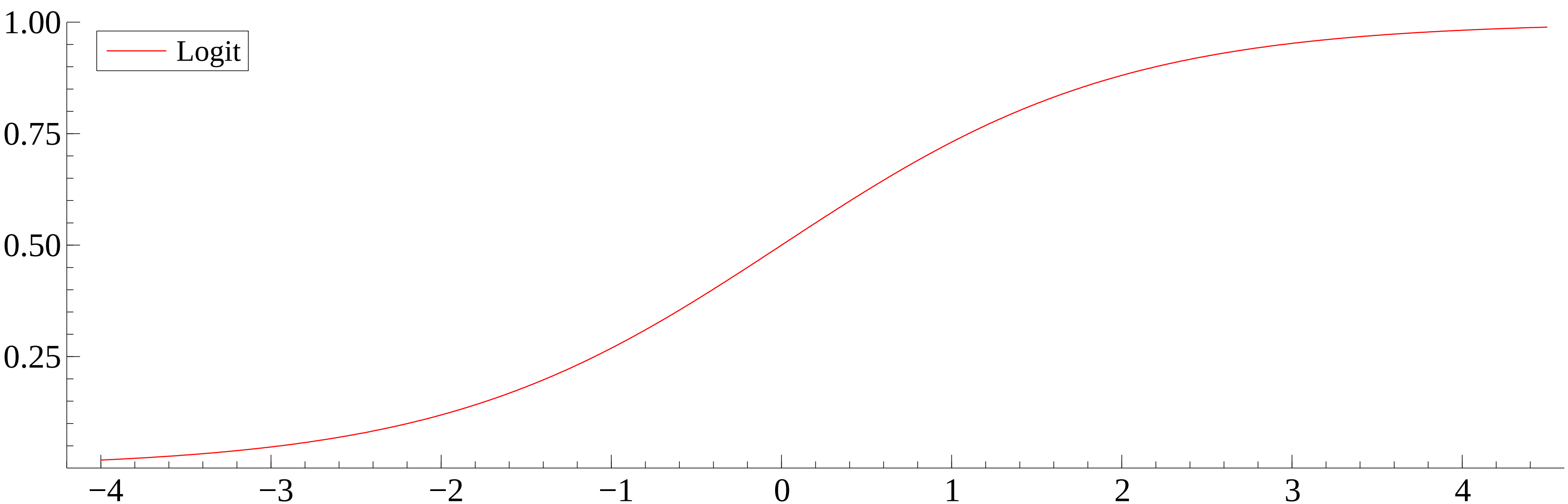
Modèles Logit et Probit

- Pour s'assurer que cette probabilité est toujours entre 0 et 1, on considère le plus souvent une classe de modèles du type suivant:

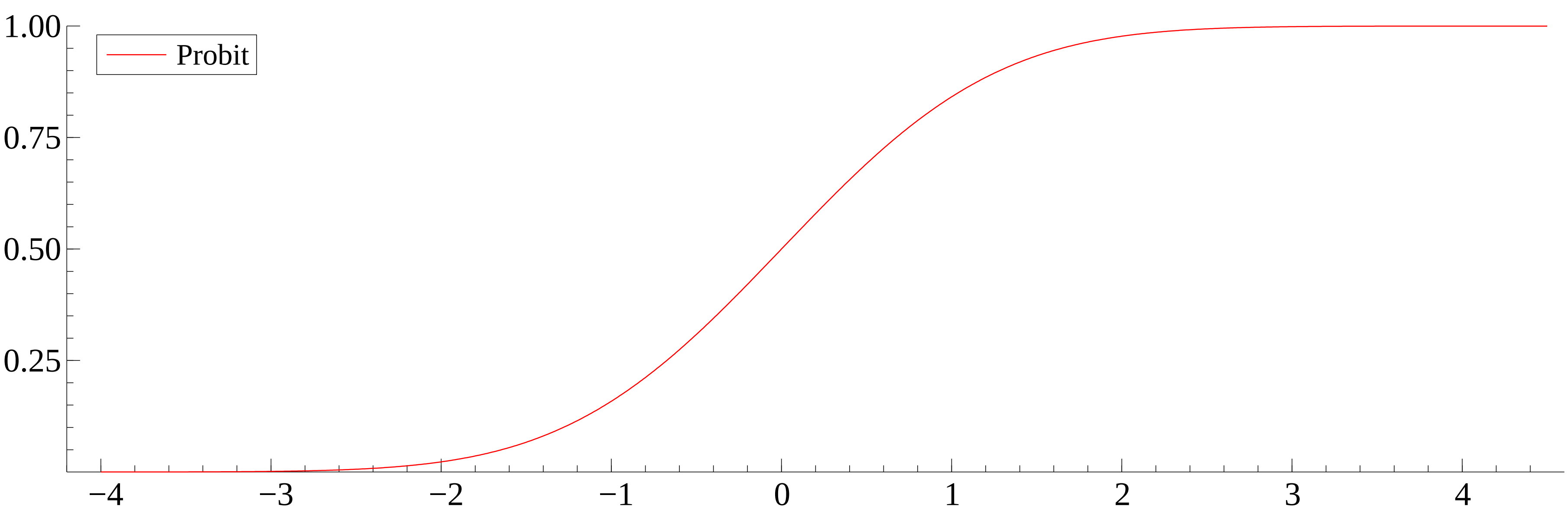
  $$P(y=1|\boldsymbol{x}) = G(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k) = G(\beta_0+\boldsymbol{xB}),$$

  où G est une fonction prenant des valeurs entre 0 et 1, c-à-d $0 < G(z) < 1$ pour tout nombre réel z.
- Modèle logit: $G(z) = \frac{exp(z)}{1+exp(z)} = \Lambda(z)$, c-à-d la fonction cumulative d'une variable aléatoire logistique de variance unitaire.
- Modèle Probit: $G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)dv$, où $\phi(z)$ est la densité normale standard $\phi(z) = (2\pi)^{-1/2} exp(-z^2/2)$.
- Pour les deux modèles, $G(z) \to 0$ quand $z \to -\infty$ et $G(z) \to 1$ quand $z \to \infty$.

$G(z)$



Logit
1.00
0.75
0.50
0.25
−4
−3
−2
−1
0
1
2
3
4



Probit
1.00
0.75
0.50
0.25
−4
−3
−2
−1
0
1
2
3
4

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$
- La fonction $1[.]$ est appelée fonction indicatrice.

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$
- La fonction $1[.]$ est appelée fonction indicatrice.
- $y = 1$ si $y^* > 0$, 0 sinon.

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$
- La fonction $1[.]$ est appelée fonction indicatrice.
- $y = 1$ si $y^* > 0$, 0 sinon.
- Nous supposons que e est indépendant de ***x*** et que e a une distribution logistique standard ou $N(0, 1)$.

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$
- La fonction $1[.]$ est appelée fonction indicatrice.
- $y = 1$ si $y^* > 0$, 0 sinon.
- Nous supposons que e est indépendant de ***x*** et que e a une distribution logistique standard ou $N(0, 1)$.
- $\rightarrow e$ est symmétriquement distribué autour de 0.

Modèles Logit et Probit

- Les modèles Logit et Probit peuvent être dérivés d'un modèles soujacent à variable latente.
- Soit y^* une variable latente (ou non observable) définie par
$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$
- La fonction $1[.]$ est appelée fonction indicatrice.
- $y = 1$ si $y^* > 0$, 0 sinon.
- Nous supposons que e est indépendant de $\boldsymbol{x}$ et que e a une distribution logistique standard ou $N(0, 1)$.
- $\rightarrow e$ est symmétriquement distribué autour de 0.
- $\rightarrow 1 - G(-z) = G(z)$.

$$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$$

-

$$\begin{aligned} P(y=1|\boldsymbol{x}) &= P(y^* > 0|\boldsymbol{x}) = P[e > -(\beta_0 + \boldsymbol{xB})|\boldsymbol{x}] \\ &= 1 - G[-(\beta_0 + \boldsymbol{xB})] = G(\beta_0 + \boldsymbol{xB}). \end{aligned}$$

$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$

- $$\begin{aligned} P(y=1|\boldsymbol{x}) &= P(y^*>0|\boldsymbol{x}) = P[e > -(\beta_0 + \boldsymbol{xB})|\boldsymbol{x}] \\ &= 1 - G[-(\beta_0 + \boldsymbol{xB})] = G(\beta_0 + \boldsymbol{xB}). \end{aligned}$$

- Dans la plupart des analyses impliquant des modèles Logit ou Probit, l'objet premier est d'expliquer l'effet d'une variation de x_j sur $P(y=1|\boldsymbol{x})$.

$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$

- $$\begin{aligned} P(y=1|\boldsymbol{x}) &= P(y^*>0|\boldsymbol{x}) = P[e > -(\beta_0 + \boldsymbol{xB})|\boldsymbol{x}] \\ &= 1 - G[-(\beta_0 + \boldsymbol{xB})] = G(\beta_0 + \boldsymbol{xB}). \end{aligned}$$

- Dans la plupart des analyses impliquant des modèles Logit ou Probit, l'objet premier est d'expliquer l'effet d'une variation de x_j sur $P(y=1|\boldsymbol{x})$.

- Il est important de noter que la direction de l'effet de x_j sur $E(y^*|\boldsymbol{x}) = \beta_0 + \boldsymbol{xB}$ et sur $E(y|\boldsymbol{x}) = P(y=1|\boldsymbol{x}) = G(\beta_0 + \boldsymbol{xB})$ est toujours la même.

$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$

- $$\begin{aligned} P(y = 1|\boldsymbol{x}) &= P(y^* > 0|\boldsymbol{x}) = P[e > -(\beta_0 + \boldsymbol{xB})|\boldsymbol{x}] \\ &= 1 - G[-(\beta_0 + \boldsymbol{xB})] = G(\beta_0 + \boldsymbol{xB}). \end{aligned}$$
- Dans la plupart des analyses impliquant des modèles Logit ou Probit, l'objet premier est d'expliquer l'effet d'une variation de x_j sur $P(y = 1|\boldsymbol{x})$.
- Il est important de noter que la direction de l'effet de x_j sur $E(y^*|\boldsymbol{x}) = \beta_0 + \boldsymbol{xB}$ et sur $E(y|\boldsymbol{x}) = P(y = 1|\boldsymbol{x}) = G(\beta_0 + \boldsymbol{xB})$ est toujours la même.
- Par contre l'amplitude de β_j n'est pas vraiment parlante.

$y^* = \beta_0 + \boldsymbol{xB} + e, y = 1[y^* > 0]$

- $$\begin{aligned} P(y=1|\boldsymbol{x}) &= P(y^* > 0|\boldsymbol{x}) = P[e > -(\beta_0 + \boldsymbol{xB})|\boldsymbol{x}] \\ &= 1 - G[-(\beta_0 + \boldsymbol{xB})] = G(\beta_0 + \boldsymbol{xB}). \end{aligned}$$
- Dans la plupart des analyses impliquant des modèles Logit ou Probit, l'objet premier est d'expliquer l'effet d'une variation de x_j sur $P(y=1|\boldsymbol{x})$.
- Il est important de noter que la direction de l'effet de x_j sur $E(y^*|\boldsymbol{x}) = \beta_0 + \boldsymbol{xB}$ et sur $E(y|\boldsymbol{x}) = P(y=1|\boldsymbol{x}) = G(\beta_0 + \boldsymbol{xB})$ est toujours la même.
- Par contre l'amplitude de β_j n'est pas vraiment parlante.
- Pour trouver l'effet partiel d'une variation d'une unité de x_j sur la probabilité de succes, c-à-d $P(y=1|\boldsymbol{x})$, il faut effectuer quelques calculs.

Effet partiel

- Si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$

où $g(z) \equiv \frac{dG}{dz}(z)$.

Effet partiel

- Si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$

  où $g(z) \equiv \frac{dG}{dz}(z)$.

- Comme G est la cdf d'une variable aléatoire continue, g est une fonction de densité de probabilité (pdf).

Effet partiel

- Si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:
$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$
où $g(z) \equiv \frac{dG}{dz}(z)$.
- Comme G est la cdf d'une variable aléatoire continue, g est une fonction de densité de probabilité (pdf).
- Comme $G(.)$ est strictement croissante pour les modèles Logit et Probit, $g(z) > 0 \ \forall z$.

Effet partiel

- Si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:
$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$
où $g(z) \equiv \frac{dG}{dz}(z)$.
- Comme G est la cdf d'une variable aléatoire continue, g est une fonction de densité de probabilité (pdf).
- Comme $G(.)$ est strictement croissante pour les modèles Logit et Probit, $g(z) > 0 \; \forall z$.
- Donc l'effet partiel de x_j sur $p(\boldsymbol{x})$ dépend de $\boldsymbol{x}$ de part la quantité $g(\beta_0 + \boldsymbol{xB})$.

Effet partiel

- Si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:
$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$
où $g(z) \equiv \frac{dG}{dz}(z)$.
- Comme G est la cdf d'une variable aléatoire continue, g est une fonction de densité de probabilité (pdf).
- Comme $G(.)$ est strictement croissante pour les modèles Logit et Probit, $g(z) > 0 \; \forall z$.
- Donc l'effet partiel de x_j sur $p(\boldsymbol{x})$ dépend de $\boldsymbol{x}$ de part la quantité $g(\beta_0 + \boldsymbol{xB})$.
- Par contre l'effet a toujours de même signe que β_j.

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j$$

- La conséquence est que l'effet <u>relatif</u> de deux variables explicatives continues ne dépend pas de ***x***.

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j$$

- La conséquence est que l'effet <u>relatif</u> de deux variables explicatives continues ne dépend pas de ***x***.
- En effet le ratio des effets partiels pour x_j et x_h est $\frac{\beta_j}{\beta_h}$.

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j$$

- La conséquence est que l'effet relatif de deux variables explicatives continues ne dépend pas de ***x***.
- En effet le ratio des effets partiels pour x_j et x_h est $\frac{\beta_j}{\beta_h}$.
- L'effet partiel est maximum lorsque $\beta_0 + \boldsymbol{xB} = 0$. Pour le modèle Probit, $g(z) = \phi(z)$ et donc $g(0) = \phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0.40$.

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j$$

- La conséquence est que l'effet relatif de deux variables explicatives continues ne dépend pas de ***x***.
- En effet le ratio des effets partiels pour x_j et x_h est $\frac{\beta_j}{\beta_h}$.
- L'effet partiel est maximum lorsque $\beta_0 + \boldsymbol{xB} = 0$. Pour le modèle Probit, $g(z) = \phi(z)$ et donc $g(0) = \phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0.40$.
- Pour le Logit, $g(z) = \frac{exp(z)}{[1+exp(z)]^2}$ et donc $g(0) = 0.25$.

Variables binaires x_j

- Le calcul de l'effet partiel pour une variable binaire est différent.

Variables binaires x_j

- Le calcul de l'effet partiel pour une variable binaire est différent.
- Si x_1 est une variable explicative binaire, l'effet partiel d'un changement de x_1 de 0 à 1, toutes choses restant égales par ailleurs est simplemement

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k x_k).$$

Variables binaires x_j

- Le calcul de l'effet partiel pour une variable binaire est différent.
- Si x_1 est une variable explicative binaire, l'effet partiel d'un changement de x_1 de 0 à 1, toutes choses restant égales par ailleurs est simplemement

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k x_k).$$

- Encore une fois, l'effet dépend de ***x*** mais connaitre le signe de β_1 donne une indication parfaire sur le signe de l'effet (et sa significativité).

Variables binaires x_j

- Le calcul de l'effet partiel pour une variable binaire est différent.
- Si x_1 est une variable explicative binaire, l'effet partiel d'un changement de x_1 de 0 à 1, toutes choses restant égales par ailleurs est simplement

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k x_k).$$

- Encore une fois, l'effet dépend de ***x*** mais connaitre le signe de β_1 donne une indication parfaire sur le signe de l'effet (et sa significativité).
- Egalement, si x_1 est une variable discrète, comme le nombre d'enfants, le passage de la valeur c_1 à $1 + c_1$ (exemple 1 à 2) à un effet de

$$G(\beta_0 + \beta_1(1 + c_1) + \beta_2 x_2 + \ldots + \beta_k x_k) - G(\beta_0 + \beta_1 c_1 + \beta_2 x_2 + \ldots + \beta_k x_k).$$

Variables du type $log(x_j)$ et x_j^2

- Soit le modèle
  $P(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 log(x_3))$.

Variables du type $log(x_j)$ et x_j^2

- Soit le modèle
  $P(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 log(x_3))$.
- L'effet partiel de x_1 sur $P(y = 1|\boldsymbol{x})$ est

$$\frac{\partial P(y = 1|\boldsymbol{x})}{\partial x_1} = g(\beta_0 + \boldsymbol{xB})(\beta_1 + 2\beta_2 x_1).$$

Variables du type $log(x_j)$ et x_j^2

- Soit le modèle $P(y=1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 log(x_3))$.
- L'effet partiel de x_1 sur $P(y=1|\boldsymbol{x})$ est

$$\frac{\partial P(y=1|\boldsymbol{x})}{\partial x_1} = g(\beta_0 + \boldsymbol{xB})(\beta_1 + 2\beta_2 x_1).$$

- L'effet partiel de x_3 sur $P(y=1|\boldsymbol{x})$ est

$$\frac{\partial P(y=1|\boldsymbol{x})}{\partial x_3} = g(\beta_0 + \boldsymbol{xB})\left(\frac{\beta_3}{x_3}\right).$$

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.
- Supposons que nous disposons de n observations.

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.
- Supposons que nous disposons de n observations.
- La distribution de y_i conditionnelle à $\boldsymbol{x}_i$ est:

$$f(y|\boldsymbol{x}_i;\boldsymbol{B}) = [G(\boldsymbol{x}_i\boldsymbol{B})]^y[1 - G(\boldsymbol{x}_i\boldsymbol{B})]^{1-y} \text{ avec } y = 0, 1.$$

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.
- Supposons que nous disposons de n observations.
- La distribution de y_i conditionnelle à $\boldsymbol{x}_i$ est:

$$f(y|\boldsymbol{x}_i;\boldsymbol{B}) = [G(\boldsymbol{x}_i\boldsymbol{B})]^y[1 - G(\boldsymbol{x}_i\boldsymbol{B})]^{1-y} \text{ avec } y = 0, 1.$$

- Pour simplifier, l'intercept est inclus dans $\boldsymbol{x}$.

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.
- Supposons que nous disposons de n observations.
- La distribution de y_i conditionnelle à $\boldsymbol{x}_i$ est:

$$f(y|\boldsymbol{x}_i;\boldsymbol{B}) = [G(\boldsymbol{x}_i\boldsymbol{B})]^y[1 - G(\boldsymbol{x}_i\boldsymbol{B})]^{1-y} \text{ avec } y = 0, 1.$$

- Pour simplifier, l'intercept est inclus dans $\boldsymbol{x}$.
- Quand $y = 1$ $f(y|\boldsymbol{x}_i;\boldsymbol{B}) = G(\boldsymbol{x}_i\boldsymbol{B})$ et quand $y = 0$ $f(y|\boldsymbol{x}_i;\boldsymbol{B}) = 1 - G(\boldsymbol{x}_i\boldsymbol{B})$.

Estimation

- Les modèles Logit et Probit sont généralement estimés par MV.
- Supposons que nous disposons de n observations.
- La distribution de y_i conditionnelle à $\boldsymbol{x}_i$ est:

$$f(y|\boldsymbol{x}_i;\boldsymbol{B}) = [G(\boldsymbol{x}_i\boldsymbol{B})]^y[1 - G(\boldsymbol{x}_i\boldsymbol{B})]^{1-y} \text{ avec } y = 0, 1.$$

- Pour simplifier, l'intercept est inclus dans $\boldsymbol{x}$.
- Quand $y = 1$ $f(y|\boldsymbol{x}_i;\boldsymbol{B}) = G(\boldsymbol{x}_i\boldsymbol{B})$ et quand $y = 0$ $f(y|\boldsymbol{x}_i;\boldsymbol{B}) = 1 - G(\boldsymbol{x}_i\boldsymbol{B})$.
- La fonction de log-vraissemblance pour l'observation i est une fonction des parametres $\boldsymbol{B}$ est des données $\boldsymbol{x}_i$, c-à-d

$$l_i(\boldsymbol{B}) = y_i log[G(\boldsymbol{x}_i\boldsymbol{B})] + (1 - y_i) log[1 - G(\boldsymbol{x}_i\boldsymbol{B})].$$

Estimation

- La log-vraissemblance pour l'échantillon de taille n est simplement $L(\boldsymbol{B}) = \sum_{i=1}^{n} l_i(\boldsymbol{B})$.

Estimation

- La log-vraissemblance pour l'échantillon de taille n est simplement $L(\boldsymbol{B}) = \sum_{i=1}^{n} l_i(\boldsymbol{B})$.
- L'estimateur du MV est celui qui maximise $L(\boldsymbol{B})$.

Estimation

- La log-vraissemblance pour l'échantillon de taille n est simplement $L(\boldsymbol{B}) = \sum_{i=1}^{n} l_i(\boldsymbol{B})$.
- L'estimateur du MV est celui qui maximise $L(\boldsymbol{B})$.
- De part la non-linéarité de $G(.)$, il n'est pas possible de dériver des formules donnant la solution à ce problème de maximisation. On utilisera donc des outils d'optimisation numérique.

Estimation

- La log-vraissemblance pour l'échantillon de taille n est simplement $L(\boldsymbol{B}) = \sum_{i=1}^{n} l_i(\boldsymbol{B})$.
- L'estimateur du MV est celui qui maximise $L(\boldsymbol{B})$.
- De part la non-linéarité de $G(.)$, il n'est pas possible de dériver des formules donnant la solution à ce problème de maximisation. On utilisera donc des outils d'optimisation numérique.
- Chaque $\hat{\beta}_j$ est reporté avec un écart-type (asymptotique).

Estimation

- Nous cherchons à savoir si une nouvelle méthode d'enseignement de l'économie, le système personnalisé d'instruction (PSI), influait significativement sur les performances dans les cours de gestion.

Estimation

- Nous cherchons à savoir si une nouvelle méthode d'enseignement de l'économie, le système personnalisé d'instruction (PSI), influait significativement sur les performances dans les cours de gestion.
- Variable dépendante: GRADE = 1 si le résultat de l'étudiant dans un cours intermédiaire de gestion financière est plus élevé que dans un cours de base.

Estimation

- Nous cherchons à savoir si une nouvelle méthode d'enseignement de l'économie, le système personnalisé d'instruction (PSI), influait significativement sur les performances dans les cours de gestion.
- Variable dépendante: GRADE = 1 si le résultat de l'étudiant dans un cours intermédiaire de gestion financière est plus élevé que dans un cours de base.
- Autres variables: GPA = note moyenne de l'étudiant, TUCE = score à un pré-test qui renseigne sur la connaissance de l'étudiant et PSI = 1 si l'étudiant a suivi la nouvelle méthode d'apprentissage.

Output LOGIT

```
CS( 1) Modelling GRADE by LOGIT
       The estimation sample is 1 - 32

                  Coefficient  Std.Error  t-value  t-prob
Constant             -13.0213      4.931    -2.64   0.013
GPA                   2.82611      1.263     2.24   0.033
TUCE                0.0951577     0.1416    0.672   0.507
PSI                   2.37869      1.065     2.23   0.034

log-likelihood    -12.8896342  no. of states              2
no. of observations        32  no. of parameters          4
baseline log-lik    -20.59173  Test: Chi^2(  3)      15.404 [0.0015]**
AIC                33.7792684  AIC/n             1.05560214
mean(GRADE)           0.34375  var(GRADE)          0.225586
Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence

              Count   Frequency Probability    loglik
State 0          21     0.65625     0.65625    -6.015
State 1          11     0.34375     0.34375    -6.875
Total            32     1.00000     1.00000    -12.89
```

Output PROBIT

```
CS( 2) Modelling GRADE by PROBIT
       The estimation sample is 1 - 32

                  Coefficient  Std.Error  t-value  t-prob
Constant             -7.45231      2.572    -2.90   0.007
GPA                   1.62582     0.6897     2.36   0.026
TUCE                0.0517280    0.08119    0.637   0.529
PSI                   1.42633     0.5870     2.43   0.022

log-likelihood    -12.8188041  no. of states                2
no. of observations        32  no. of parameters            4
zeroline log-lik    -22.18071  Test: Chi^2(  4)        18.724 [0.0009]**
AIC                33.6376081  AIC/n               1.05117525
mean(GRADE)           0.34375  var(GRADE)            0.225586
Newton estimation (eps1=0.0001; eps2=0.005): Strong convergence

              Count   Frequency Probability    loglik
State 0          21     0.65625     0.65728    -6.010
State 1          11     0.34375     0.34272    -6.808
Total            32     1.00000     1.00000    -12.82
```

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.
- C'est normal car l'effet partiel est fonction de $g(.)\beta_j$ avec $g(.)$ la fonction cdf de la logistique ou de la normale.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.
- C'est normal car l'effet partiel est fonction de $g(.)\beta_j$ avec $g(.)$ la fonction cdf de la logistique ou de la normale.
- Notons que nous avons vu que $g(0) = \phi(0) =\approx 0.4$ pour la normale et $\frac{exp(0)}{[1+exp(0)]^2} = 0.25$ pour la logistique.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.
- C'est normal car l'effet partiel est fonction de $g(.)\beta_j$ avec $g(.)$ la fonction cdf de la logistique ou de la normale.
- Notons que nous avons vu que $g(0) = \phi(0) = \approx 0.4$ pour la normale et $\frac{exp(0)}{[1+exp(0)]^2} = 0.25$ pour la logistique.
- Donc pour avoir des effet marginaux equivalents il faut $\frac{\phi(0)\beta_j^p}{\frac{exp(0)}{[1+exp(0)]^2}\beta_j^l} = 1$.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.
- C'est normal car l'effet partiel est fonction de $g(.)\beta_j$ avec $g(.)$ la fonction cdf de la logistique ou de la normale.
- Notons que nous avons vu que $g(0) = \phi(0) =\approx 0.4$ pour la normale et $\frac{exp(0)}{[1+exp(0)]^2} = 0.25$ pour la logistique.
- Donc pour avoir des effet marginaux equivalents il faut $\frac{\phi(0)\beta_j^p}{\frac{exp(0)}{[1+exp(0)]^2}\beta_j^l} = 1$.
- C-à-d $\frac{0.4\beta_j^p}{0.25\beta_j^l} = 1$ ou encore $\beta_j^l = 1.6\beta_j^p$.

Comparaison des résultats

- Les niveaux de significativités des 2 modèles sont très proches.
- Les coefficients estimés sont différents.
- C'est normal car l'effet partiel est fonction de $g(.)\beta_j$ avec $g(.)$ la fonction cdf de la logistique ou de la normale.
- Notons que nous avons vu que $g(0) = \phi(0) =\approx 0.4$ pour la normale et $\frac{exp(0)}{[1+exp(0)]^2} = 0.25$ pour la logistique.
- Donc pour avoir des effet marginaux equivalents il faut $\frac{\phi(0)\beta_j^p}{\frac{exp(0)}{[1+exp(0)]^2}\beta_j^l} = 1$.
- C-à-d $\frac{0.4\beta_j^p}{0.25\beta_j^l} = 1$ ou encore $\beta_j^l = 1.6\beta_j^p$.
- Exemple pour PSI: $\frac{2.37869}{1.42633} = 1.6677$.

Exemple de LRT sur PSI

```
CS( 2) Modelling GRADE by PROBIT
       The estimation sample is 1 - 32

                  Coefficient  Std.Error  t-value  t-prob
Constant             -6.03435      2.166    -2.79   0.009
GPA                   1.40958     0.6510     2.17   0.039
TUCE                0.0526670    0.07433    0.709   0.484

log-likelihood    -16.1521573  no. of states               2
no. of observations        32  no. of parameters           3
zeroline log-lik    -22.18071  Test: Chi^2(  3)       12.057 [0.0072]**
AIC                38.3043147  AIC/n              1.19700983
mean(GRADE)           0.34375  var(GRADE)           0.225586
```

$\rightarrow LRT = 2[-12.8188041 - (-16.1521573)] = 2 \times 3.33 = 6.66 >> 3.86(\chi^2_{5\%}(1))$.
Mais $\chi^2_{1\%}(1) = 6.63$.
Attention ce test est asymptotique alors que $n = 32$.

Exemple de test de Wald

```
Test for linear restrictions (Rb=r):
R matrix
     Constant          GPA         TUCE          PSI
      0.00000      0.00000      0.00000       1.0000
r vector
      0.00000
LinRes Chi^2(1) = 5.90506 [0.0151] *
```

Attention ce test est également asymptotique alors que $n = 32$.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le pourcentage de prédictions correctes.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le pourcentage de prédictions correctes.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le <u>pourcentage de prédictions correctes</u>.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.
- Soit $\tilde{y}_i = 1$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) \geq 0.5$ et $\tilde{y}_i = 0$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) < 0.5$.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le <u>pourcentage de prédictions correctes</u>.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.
- Soit $\tilde{y}_i = 1$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i \hat{\boldsymbol{B}}) \geq 0.5$ et $\tilde{y}_i = 0$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i \hat{\boldsymbol{B}}) < 0.5$.
- Etant donné $\{\tilde{y}_i : i = 1, 2, \ldots, n\}$, on peut voir la qualité de la prévision de y_i.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le <u>pourcentage de prédictions correctes</u>.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.
- Soit $\tilde{y}_i = 1$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) \geq 0.5$ et $\tilde{y}_i = 0$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) < 0.5$.
- Etant donné $\{\tilde{y}_i : i = 1, 2, \ldots, n\}$, on peut voir la qualité de la prévision de y_i.
- Pour chaque i, la paire $(\tilde{y}_i, y_i)$ peut prendre quatres valeurs: $(0,0), (0,1), (1,0)$ et $(1,1)$.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le pourcentage de prédictions correctes.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.
- Soit $\tilde{y}_i = 1$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) \geq 0.5$ et $\tilde{y}_i = 0$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) < 0.5$.
- Etant donné $\{\tilde{y}_i : i = 1, 2, \ldots, n\}$, on peut voir la qualité de la prévision de y_i.
- Pour chaque i, la paire $(\tilde{y}_i, y_i)$ peut prendre quatres valeurs: $(0,0), (0,1), (1,0)$ et $(1,1)$.
- Pour les paires $(0,0)$ et $(1,1)$, la prévision est correcte.

Goodness-of-fit

- Une mesure fréquemment reportée lors de l'estimation de modèles Logit et Probit est le pourcentage de prédictions correctes.
- Cette mesure s'apparente au R^2 d'un modèle linéaire.
- Soit $\tilde{y}_i = 1$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) \geq 0.5$ et $\tilde{y}_i = 0$ si $G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}}) < 0.5$.
- Etant donné $\{\tilde{y}_i : i = 1, 2, \ldots, n\}$, on peut voir la qualité de la prévision de y_i.
- Pour chaque i, la paire $(\tilde{y}_i, y_i)$ peut prendre quatres valeurs: $(0,0), (0,1), (1,0)$ et $(1,1)$.
- Pour les paires $(0,0)$ et $(1,1)$, la prévision est correcte.
- Le pourcentage de prédictions correctes est le %age de valeurs $\tilde{y}_i = y_i$.

Output PROBIT and LOGIT

Table of actual and predicted

	State 0	State 1	Sum actual
State 0	18	3	21
State 1	3	8	11
Sum pred	21	11	32

$\rightarrow \frac{18+8}{32} = 0.8125$

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.
- En effet, il est possible d'obtenir des pourcentages de prédictions correctes élevés même lorsque l'évènement le moins probable est très mal prédit.

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.
- En effet, il est possible d'obtenir des pourcentages de prédictions correctes élevés même lorsque l'évènement le moins probable est très mal prédit.
- Exemple: $n = 200$, 160 observations ont $y_i = 0$ et hors de ces 160 observations 140 ont $\tilde{y}_i = 0$. Donc on prédit correctement dans 87.5% des cas les valeurs de $y_i = 0$.

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.
- En effet, il est possible d'obtenir des pourcentages de prédictions correctes élevés même lorsque l'évènement le moins probable est très mal prédit.
- Exemple: $n = 200$, 160 observations ont $y_i = 0$ et hors de ces 160 observations 140 ont $\tilde{y}_i = 0$. Donc on prédit correctement dans 87.5% des cas les valeurs de $y_i = 0$.
- Même si aucune prévision n'est correcte pour $y_i = 1$, on prédit toujours correctement 70% des observations ($140/200$).

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.
- En effet, il est possible d'obtenir des pourcentages de prédictions correctes élevés même lorsque l'évènement le moins probable est très mal prédit.
- Exemple: $n = 200$, 160 observations ont $y_i = 0$ et hors de ces 160 observations 140 ont $\tilde{y}_i = 0$. Donc on prédit correctement dans 87.5% des cas les valeurs de $y_i = 0$.
- Même si aucune prévision n'est correcte pour $y_i = 1$, on prédit toujours correctement 70% des observations ($140/200$).
- Il est souvant utile de pouvoir également prédire les évènements les moins probables.

Goodness-of-fit

- Même si cette mesure est utile elle peut être trompeuse.
- En effet, il est possible d'obtenir des pourcentages de prédictions correctes élevés même lorsque l'évènement le moins probable est très mal prédit.
- Exemple: $n = 200$, 160 observations ont $y_i = 0$ et hors de ces 160 observations 140 ont $\tilde{y}_i = 0$. Donc on prédit correctement dans 87.5% des cas les valeurs de $y_i = 0$.
- Même si aucune prévision n'est correcte pour $y_i = 1$, on prédit toujours correctement 70% des observations ($140/200$).
- Il est souvant utile de pouvoir également prédire les évènements les moins probables.
- Il est donc important de calculer ce %age pour chaque valeur de y_i, c-à-d 0 et 1.

Goodness-of-fit

- Une autre mesure reportée par la plupart des logicles économériques est le pseudo-R^2 de McFadden.

Goodness-of-fit

- Une autre mesure reportée par la plupart des logicles économériques est le pseudo-R^2 de McFadden.
- Ce pseudo-$R^2 = 1 - \frac{L_{nc}}{L_0}$, où L_{nc} est la log-vraissemblance du modèle (non contraint) et L_0 est celle du modèle contraint ne comportant qu'une constante.

Goodness-of-fit

- Une autre mesure reportée par la plupart des logicles économériques est le pseudo-R^2 de McFadden.
- Ce pseudo-$R^2 = 1 - \frac{L_{nc}}{L_0}$, où L_{nc} est la log-vraissemblance du modèle (non contraint) et L_0 est celle du modèle contraint ne comportant qu'une constante.
- Pourqui est-ce que cette mesure a du sens ?

Goodness-of-fit

- Une autre mesure reportée par la plupart des logicles économériques est le pseudo-R^2 de McFadden.
- Ce pseudo-$R^2 = 1 - \frac{L_{nc}}{L_0}$, où L_{nc} est la log-vraissemblance du modèle (non contraint) et L_0 est celle du modèle contraint ne comportant qu'une constante.
- Pourqui est-ce que cette mesure a du sens ?
- $L < 0$ donc $\frac{L_{nc}}{L_0} = \frac{|L_{nc}|}{|L_0|}$.

Goodness-of-fit

- Une autre mesure reportée par la plupart des logicles économériques est le pseudo-R^2 de McFadden.
- Ce pseudo-$R^2 = 1 - \frac{L_{nc}}{L_0}$, où L_{nc} est la log-vraissemblance du modèle (non contraint) et L_0 est celle du modèle contraint ne comportant qu'une constante.
- Pourqui est-ce que cette mesure a du sens ?
- $L < 0$ donc $\frac{L_{nc}}{L_0} = \frac{|L_{nc}|}{|L_0|}$.
- De plus $|L_{nc}| \leq |L_0|$.

Goodness-of-fit

- Si les variables explicatives n'ont pas de pouvoir explicatif $\frac{L_{nc}}{L_0} = 1$ et donc pseudo-$R^2 = 0$.

Goodness-of-fit

- Si les variables explicatives n'ont pas de pouvoir explicatif $\frac{L_{nc}}{L_0} = 1$ et donc pseudo-$R^2 = 0$.
- Généralement $|L_{nc}| < |L_0|$ et donc $1 - \frac{L_{nc}}{L_0} > 0$

Goodness-of-fit

- Si les variables explicatives n'ont pas de pouvoir explicatif $\frac{L_{nc}}{L_0} = 1$ et donc pseudo-$R^2 = 0$.
- Généralement $|L_{nc}| < |L_0|$ et donc $1 - \frac{L_{nc}}{L_0} > 0$
- Si $L_{nc} = 0$ le pseudo-$R^2 = 1$.

Goodness-of-fit: Alternative

- Soit $\hat{y}_i = G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$ les prévisions de probabilité du modèle Logit ou Probit.

Goodness-of-fit: Alternative

- Soit $\hat{y}_i = G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$ les prévisions de probabilité du modèle Logit ou Probit.
- Comme ces prévisions sont aussi des prévisions de $E(y_i|\boldsymbol{x}_i)$, on peut calculer un genre de R^2 en comparant $\hat{y}_i$ et y_i.

Goodness-of-fit: Alternative

- Soit $\hat{y}_i = G(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$ les prévisions de probabilité du modèle Logit ou Probit.
- Comme ces prévisions sont aussi des prévisions de $E(y_i|\boldsymbol{x}_i)$, on peut calculer un genre de R^2 en comparant $\hat{y}_i$ et y_i.
- $\rightarrow$ reporter $corr(\hat{y}_i, y_i)^2$.

Output Descriptive Statistics

```
Means, standard deviations and correlations
The sample is: 1 - 32
Means
   P(State 1)        GRADE
      0.34375      0.34375
Standard deviations (using T-1)
   P(State 1)        GRADE
      0.31690      0.48256
Correlation matrix:
                P(State 1)        GRADE
P(State 1)          1.0000      0.65263
GRADE              0.65263       1.0000
```

$\longrightarrow$ Exportez d'abord les probabilités estimées.

$\longrightarrow$ $corr(GRADE, P(State_1)) = 0.65263$

Reporter des effets partiels

- On a vu que si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{x}\boldsymbol{B})\beta_j,$$

où $g(z) \equiv \frac{dG}{dz}(z)$.

Reporter des effets partiels

- On a vu que si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$

  où $g(z) \equiv \frac{dG}{dz}(z)$.

- Par conséquent

$$\nabla \hat{p(\boldsymbol{x})} \approx g(\hat{\beta}_0 + \boldsymbol{x}\hat{\boldsymbol{B}})\hat{\beta}_j \nabla x_j,$$

  pour des petites variations de x_j.

Reporter des effets partiels

- On a vu que si x_j est une variable continue, son effet partiel sur $p(\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$ est obtenu depuis la dérivée partielle suivante:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{xB})\beta_j,$$

  où $g(z) \equiv \frac{dG}{dz}(z)$.

- Par conséquent

$$\nabla p(\hat{\boldsymbol{x}}) \approx g(\hat{\beta}_0 + \boldsymbol{x\hat{B}})\hat{\beta}_j \nabla x_j,$$

  pour des petites variations de x_j.

- Mais l'effet partiel dépend de $\boldsymbol{x}$, où $g(\hat{\beta}_0 + \boldsymbol{x\hat{B}})$ est appelé facteur d'échelle. Que reporter ?

Reporter des effets partiels

- Les logiciels économétriques reportent souvant des effets partiels ou le facteur d'échelle est évalué en $\bar{\boldsymbol{x}}$, c-à-d

$$g(\hat{\beta}_0 + \bar{\boldsymbol{x}}\hat{\boldsymbol{B}}) = g(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_k \bar{x}_k).$$

Reporter des effets partiels

- Les logiciels économétriques reportent souvant des effets partiels ou le facteur d'échelle est évalué en $\bar{\boldsymbol{x}}$, c-à-d

$$g(\hat{\beta}_0 + \bar{\boldsymbol{x}}\hat{\boldsymbol{B}}) = g(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_k\bar{x}_k).$$

- En multipliant ce facteur d'échelle par β_j on obtient l'effet partiel de x_j pour la personne *moyenne*.

Reporter des effets partiels

- Les logiciels économétriques reportent souvant des effets partiels ou le facteur d'échelle est évalué en $\bar{\boldsymbol{x}}$, c-à-d

$$g(\hat{\beta}_0 + \bar{\boldsymbol{x}}\hat{\boldsymbol{B}}) = g(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_k \bar{x}_k).$$

- En multipliant ce facteur d'échelle par β_j on obtient l'effet partiel de x_j pour la personne *moyenne*.
- Qu'est-ce que l'individu moyen si par exemple $x_1 = Male$ (1 si homme et 0 pour femme).

Reporter des effets partiels

- Les logiciels économétriques reportent souvant des effets partiels ou le facteur d'échelle est évalué en $\bar{\boldsymbol{x}}$, c-à-d

$$g(\hat{\beta}_0 + \bar{\boldsymbol{x}}\hat{\boldsymbol{B}}) = g(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_k\bar{x}_k).$$

- En multipliant ce facteur d'échelle par β_j on obtient l'effet partiel de x_j pour la personne *moyenne*.
- Qu'est-ce que l'individu moyen si par exemple $x_1 = Male$ (1 si homme et 0 pour femme).
- Si on a 52.5% d'hommes $\bar{x}_1 = 0.525$. En quoi cela représente la presonne moyenne ?

Reporter des effets partiels

- De plus si x_1 est par exemple $log(sales)$. Doit-on évaluer l'impact moyen en $\overline{log(sales)}$ ou $log(\overline{sales})$? Les logiciels econométriques reportent le premier.

Reporter des effets partiels

- De plus si x_1 est par exemple $log(sales)$. Doit-on évaluer l'impact moyen en $\overline{log(sales)}$ ou $log(\overline{sales})$? Les logiciels econométriques reportent le premier.
- Une approche différente pour calculer le facteur d'échelle qui premettrait de régler ces problèmes est la suivante.

Reporter des effets partiels

- De plus si x_1 est par exemple $log(sales)$. Doit-on évaluer l'impact moyen en $\overline{log(sales)}$ ou $log(\overline{sales})$? Les logiciels econométriques reportent le premier.
- Une approche différente pour calculer le facteur d'échelle qui premettrait de régler ces problèmes est la suivante.
- $$n^{-1}\sum_{i=1}^{n} g(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$$

Reporter des effets partiels

- De plus si x_1 est par exemple $log(sales)$. Doit-on évaluer l'impact moyen en $\overline{log(sales)}$ ou $log(\overline{sales})$? Les logiciels econométriques reportent le premier.
- Une approche différente pour calculer le facteur d'échelle qui premettrait de régler ces problèmes est la suivante.
- $$n^{-1}\sum_{i=1}^{n} g(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$$
- $\rightarrow$ on prend la moyenne de la fonction linéaire et non la fonction linéaire évaluée en la moyenne.

Reporter des effets partiels

- De plus si x_1 est par exemple $log(sales)$. Doit-on évaluer l'impact moyen en $\overline{log(sales)}$ ou $log(\overline{sales})$? Les logiciels econométriques reportent le premier.
- Une approche différente pour calculer le facteur d'échelle qui premettrait de régler ces problèmes est la suivante.
- $$n^{-1}\sum_{i=1}^{n} g(\hat{\beta}_0 + \boldsymbol{x}_i\hat{\boldsymbol{B}})$$
- $\rightarrow$ on prend la moyenne de la fonction linéaire et non la fonction linéaire évaluée en la moyenne.
- Les deux méthodes puvent donner des facteurs d'échelle assez différents.

Reporter des effets partiels

- Si les variables explicatives comprennent des variables discrètes, ces formules ne sont pas valables.

Reporter des effets partiels

- Si les variables explicatives comprennent des variables discrètes, ces formules ne sont pas valables.
- Pour un changement de x_1 de c_1 à $1 + c_1$, l'analogue discret de l'effet partiel est

$$G(\hat{\beta}_0 + \hat{\beta}_1(1 + c_1) + \ldots + \hat{\beta}_k \bar{x}_k) - G(\hat{\beta}_0 + \hat{\beta}_1 c_1 + \ldots + \hat{\beta}_k \bar{x}_k).$$

Reporter des effets partiels

- Si les variables explicatives comprennent des variables discrètes, ces formules ne sont pas valables.
- Pour un changement de x_1 de c_1 à $1+c_1$, l'analogue discret de l'effet partiel est

$$G(\hat{\beta}_0 + \hat{\beta}_1(1+c_1) + \ldots + \hat{\beta}_k \bar{x}_k) - G(\hat{\beta}_0 + \hat{\beta}_1 c_1 + \ldots + \hat{\beta}_k \bar{x}_k).$$

- La moyenne des effets partiels est quant à elle donnée par

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} \{ \quad & G(\hat{\beta}_0 + \hat{\beta}_1(1+c_1) + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}) \\ - \quad & G(\hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}) \}. \end{aligned}$$

Exemple: GRADE et PROBIT

Summary statistics for explanatory variables

State 0:

	#obs	min	mean	max	std.dev
Constant	21	1	1	1	0
GPA	21	2.06	2.9519	3.57	0.34861
TUCE	21	12	21.095	26	3.6892
PSI	21	0	0.28571	1	0.45175

State 1:

	#obs	min	mean	max	std.dev
Constant	11	1	1	1	0
GPA	11	2.39	3.4327	4	0.47972
TUCE	11	17	23.545	29	3.6021
PSI	11	0	0.72727	1	0.44536

All States:

	#obs	min	mean	max	std.dev
Constant	32	1	1	1	0
GPA	32	2.06	3.1172	4	0.45936
TUCE	32	12	21.938	29	3.8401
PSI	32	0	0.4375	1	0.49608

$\rightarrow \overline{GPA} = 3.1172$ et $\overline{TUCE} = 21.938$.

Exemple: GRADE et PROBIT

- $$\begin{aligned} & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE} + 1.426) \\ - & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE}) \\ = & 0.57273563 - 0.106997165 = 0.465738465. \end{aligned}$$

Exemple: GRADE et PROBIT

- $$\begin{aligned} & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE} + 1.426) \\ - \quad & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE}) \\ = \quad & 0.57273563 - 0.106997165 = 0.465738465. \end{aligned}$$
- Comment évalue ce différentiel en fonction de GPA ?

Exemple: GRADE et PROBIT

- $$\begin{aligned} & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE} + 1.426) \\ - \quad & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE}) \\ = \quad & 0.57273563 - 0.106997165 = 0.465738465. \end{aligned}$$
- Comment évalue ce différentiel en fonction de GPA ?
- $$\begin{aligned} & G(-7.452 + 1.626GPA + 0.052\overline{TUCE} + 1.426) \\ - \quad & G(-7.452 + 1.626GPA + 0.052\overline{TUCE}) \end{aligned}$$

Exemple: GRADE et PROBIT

- $$\begin{aligned} & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE} + 1.426) \\ - \quad & G(-7.452 + 1.626\overline{GPA} + 0.052\overline{TUCE}) \\ = \quad & 0.57273563 - 0.106997165 = 0.465738465. \end{aligned}$$
- Comment évalue ce différentiel en fonction de GPA ?
- $$\begin{aligned} & G(-7.452 + 1.626GPA + 0.052\overline{TUCE} + 1.426) \\ - \quad & G(-7.452 + 1.626GPA + 0.052\overline{TUCE}) \end{aligned}$$
- GRADE_prob.xls

Output

```
Strong convergence using numerical derivatives
 Function value = -401.302
                   Coefficient  Std.Error  t-value  t-prob
Constant              0.270077    0.50860   0.5310  0.5956
NWIFEINC             -0.012024  0.0048399   -2.484  0.0132
EDUC                  0.130905   0.025254    5.183  0.0000
EXPER                 0.123347   0.018725    6.587  0.0000
EXPERSQ              -0.001887 0.00060040   -3.143  0.0017
AGE                  -0.052853  0.0084772   -6.235  0.0000
KIDSLT6              -0.868327    0.11852   -7.326  0.0000
KIDSGE6               0.036005   0.043477   0.8281  0.4079
```

PcGive

```
CS( 2) Modelling INLF by Probit
       The estimation sample is 1 - 753

                    Coefficient  Std.Error  t-value  t-prob
Constant               0.270077     0.5081    0.532   0.595
NWIFEINC             -0.0120237   0.004939    -2.43   0.015
EDUC                   0.130905    0.02540     5.15   0.000
EXPER                  0.123348    0.01876     6.58   0.000
EXPERSQ             -0.00188708  0.0005999    -3.15   0.002
AGE                  -0.0528527   0.008463    -6.25   0.000
KIDSLT6               -0.868329     0.1184    -7.33   0.000
KIDSGE6               0.0360050    0.04403    0.818   0.414

log-likelihood      -401.302193  no. of states                 2
no. of observations         753  no. of parameters             8
zeroline log-lik      -521.9398  Test: Chi^2(  8)       241.28 [0.0000]**
AIC                  818.604386  AIC/n                1.08712402
mean(INLF)             0.568393  var(INLF)              0.245322
BFGS estimation (eps1=0.0001; eps2=0.005): Strong convergence
```

LR test

$H_0 : KIDSLT6 = 0, KIDSGE6 = 0$

```
Strong convergence using numerical derivatives
 Function value = -432.809
                     Coefficient  Std.Error  t-value  t-prob
Constant               -0.619011    0.41656   -1.486  0.1377
NWIFEINC               -0.011281  0.0046216   -2.441  0.0149
EDUC                    0.105111   0.023991    4.381  0.0000
EXPER                   0.125348   0.018194    6.890  0.0000
EXPERSQ                -0.002041 0.00059101   -3.453  0.0006
AGE                    -0.028674  0.0068451   -4.189  0.0000
```

$LRT = 2 \times (-401.302193 + 432.809) = 63.018 >> \chi^2(2)$.

Indeed,

```
 Chi^2(2) = 63.018 [0.0000] **
```

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.
- 2 marques: Heintz et Hunts.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.
- 2 marques: Heintz et Hunts.
- 2226 achats de Heintz contre 265 pour Hunts.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.
- 2 marques: Heintz et Hunts.
- 2226 achats de Heintz contre 265 pour Hunts.
- Variable endogène: Heintz $= 1$ si achat de Heintz, 0 sinon.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.
- 2 marques: Heintz et Hunts.
- 2226 achats de Heintz contre 265 pour Hunts.
- Variable endogène: Heintz $= 1$ si achat de Heintz, 0 sinon.
- Variables explicatives: Display (tête de gondole) only, feature (folder) only, feature and display et la difference du log du prix de Heintz et Hunts.

Une application en marketing

- Quels sont les déterminants du choix entre deux marques de ketchup ?
- 2798 observations concernant 300 individus.
- On ne selectionne pas la dernière observation de chaque individu (laissée pour des prévisions out-of-sample) $\rightarrow 2498$ observations dans l'estimation.
- 2 marques: Heintz et Hunts.
- 2226 achats de Heintz contre 265 pour Hunts.
- Variable endogène: Heintz $= 1$ si achat de Heintz, 0 sinon.
- Variables explicatives: Display (tête de gondole) only, feature (folder) only, feature and display et la difference du log du prix de Heintz et Hunts.
- Heintz1.in7

Output Logit sur HEINZ

	Coefficient	Std.Error	t-value	t-prob
Constant	3.29016	0.1505	21.9	0.000
DISPLHEINZ	0.525596	0.2545	2.07	0.039
DISPLHUNTS	-0.650595	0.2539	-2.56	0.010
FEATHEINZ	0.474176	0.3196	1.48	0.138
FEATHUNTS	-1.03337	0.3605	-2.87	0.004
FEATDISPLHEINZ	0.472853	0.4889	0.967	0.334
FEATDISPLHUNTS	-1.98104	0.4791	-4.13	0.000
LPRICEHEINZ	-5.98750	0.4005	-14.9	0.000

log-likelihood	-601.238259	no. of states	2
no. of observations	2498	no. of parameters	8
baseline log-lik	-859.7672	Test: Chi^2(7)	517.06 [0.0000]**
mean(HEINZ)	0.891113	var(HEINZ)	0.0970307

Output Logit sur HEINZ

Table of actual and predicted

	State 0	State 1	Sum actual
State 0	71	201	272
State 1	28	2198	2226
Sum pred	99	2399	2498

Output Probit sur HEINZ

	Coefficient	Std.Error	t-value	t-prob
Constant	1.84643	0.07503	24.6	0.000
DISPLHEINZ	0.270800	0.1288	2.10	0.036
DISPLHUNTS	-0.376042	0.1499	-2.51	0.012
FEATHEINZ	0.187592	0.1575	1.19	0.234
FEATHUNTS	-0.572908	0.1995	-2.87	0.004
FEATDISPLHEINZ	0.254583	0.2484	1.02	0.306
FEATDISPLHUNTS	-1.09385	0.2747	-3.98	0.000
LPRICEHEINZ	-3.27430	0.2119	-15.4	0.000

log-likelihood	-598.528353	no. of states	2	
no. of observations	2498	no. of parameters	8	
zeroline log-lik	-1731.482	Test: Chiˆ2(8)	2265.9	[0.0000]**
mean(HEINZ)	0.891113	var(HEINZ)	0.0970307	

Output Probit sur HEINZ

Table of actual and predicted

	State 0	State 1	Sum actual
State 0	70	202	272
State 1	28	2198	2226
Sum pred	98	2400	2498

Une application en marketing

- L'intercept est positif. Notons que $\frac{exp(3.29)}{1+exp(3.29)} = 0.964$.

Une application en marketing

- L'intercept est positif. Notons que $\frac{exp(3.29)}{1+exp(3.29)} = 0.964$.
- Donc la probabilité d'achat de Heintz si les variables explicatives sont 0 est de 96.4%.

Une application en marketing

- L'intercept est positif. Notons que $\frac{exp(3.29)}{1+exp(3.29)} = 0.964$.
- Donc la probabilité d'achat de Heintz si les variables explicatives sont 0 est de 96.4%.
- Notons que la probabilité inconditionnelle d'acaht de Heintz est de 89.1%.

Une application en marketing

- L'intercept est positif. Notons que $\frac{exp(3.29)}{1+exp(3.29)} = 0.964$.
- Donc la probabilité d'achat de Heintz si les variables explicatives sont 0 est de 96.4%.
- Notons que la probabilité inconditionnelle d'acaht de Heintz est de 89.1%.
- Les trois variables relatives a Hunts sont d'ampleur plus importantes que celles de Heintz, sont négatives et largement significatives.

Une application en marketing

- L'intercept est positif. Notons que $\frac{exp(3.29)}{1+exp(3.29)} = 0.964$.
- Donc la probabilité d'achat de Heintz si les variables explicatives sont 0 est de 96.4%.
- Notons que la probabilité inconditionnelle d'acaht de Heintz est de 89.1%.
- Les trois variables relatives a Hunts sont d'ampleur plus importantes que celles de Heintz, sont négatives et largement significatives.
- Une politique de marketing aggressive de la part de Hunts peut faire gagner des parts de marché même si Heintz reste le leader dans le domaine.

Une application en marketing

- L’effet est le plus fort pour une promotion couplant la mise en évidence du produit Hunts (tête de gondole) et un affichage dans le folder promotionel.

Une application en marketing

- L'effet est le plus fort pour une promotion couplant la mise en évidence du produit Hunts (tête de gondole) et un affichage dans le folder promotionel.
- Les variables promotion de Heintz ne sont pas très significatives (uniquement display mais au seuil de 4%).

Une application en marketing

- L'effet est le plus fort pour une promotion couplant la mise en évidence du produit Hunts (tête de gondole) et un affichage dans le folder promotionel.
- Les variables promotion de Heintz ne sont pas très significatives (uniquement display mais au seuil de 4%).
- Le coefficient de $LPRICEHEINZ$ est significativement négatif et donc une augmentation de prix de Heintz réduit les ventes et donc une diminution du prix de Hunts permet d'augmenter les ventes de Hunts.

$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\hat{\beta}_0 + \bar{\boldsymbol{x}}\hat{\boldsymbol{B}})\hat{\beta}_j$ pour le Logit

```
Derivatives of probabilities at regressor means
Probabilities:
State 0                  0.036095
State 1                   0.96390
Derivatives:
                              mean          State 0          State 1
Constant                    1.0000         -0.11447          0.11447
DISPLHEINZ                 0.16053        -0.018287         0.018287
DISPLHUNTS                0.037230         0.022636        -0.022636
FEATHEINZ                  0.13090        -0.016498         0.016498
FEATHUNTS                 0.038431         0.035954        -0.035954
FEATDISPLHEINZ            0.035228        -0.016452         0.016452
FEATDISPLHUNTS            0.010408         0.068925        -0.068925
LPRICEHEINZ               0.014010          0.20832         -0.20832
```

Prévisions (Logit) avec 1 si $\hat{p}(x) > 0.5$

```
Predictions:


Table of actual and predicted
              State 0  State 1  Sum actual
State 0             4       31          35
State 1             1      264         265
Sum pred            5      295         300
```

$\rightarrow 88\%$ de bonnes prévisions de $1, 1.3\%$ de bonnes prévisions de 0.

$\rightarrow 10.3\%$ de mauvaises prévisions de 0 et 0.33% de mauvaises prévisions de 1.

Pour $\hat{p}(x) > c$ avec $c \neq 0.5$ (par exemple $E(Heintz) = 0.891$) exportez les prévisions et faites le calcul en excel.

Estimation par variables instrumentales et doubles moindres carrés : Chapitre 15

Utilisation des variables instrumentales

L'utilisation de la méthode des variables instrumentales (VI) se justifie dans différents contextes :

- Variables explicatives endogènes.

Utilisation des variables instrumentales

L'utilisation de la méthode des variables instrumentales (VI) se justifie dans différents contextes :

- Variables explicatives endogènes.
- Problème d'erreurs de mesures sur certaines variables. Exemple : revenu déclaré.

Utilisation des variables instrumentales

L'utilisation de la méthode des variables instrumentales (VI) se justifie dans différents contextes :

- Variables explicatives endogènes.
- Problème d'erreurs de mesures sur certaines variables. Exemple : revenu déclaré.
- Problème d'omission de variables explicatives appropriées.

Variables explicatives endogènes

- Dans ce contexte, l'utilisation de VI donne lieu à une nouvelle méthode d'estimation : les doubles moindres carrés.

Variables explicatives endogènes

- Dans ce contexte, l'utilisation de VI donne lieu à une nouvelle méthode d'estimation : les doubles moindres carrés.
- Méthode très populaire, notamment pour estimer des modèles macroéconomiques à plusieurs équations.

Variables explicatives endogènes

- Dans ce contexte, l'utilisation de VI donne lieu à une nouvelle méthode d'estimation : les doubles moindres carrés.
- Méthode très populaire, notamment pour estimer des modèles macroéconomiques à plusieurs équations.
- Des exemples spécifiques seront développés dans le Chapitre 16 : modèles à équations simultanées.

Variables explicatives endogènes

- Dans ce contexte, l'utilisation de VI donne lieu à une nouvelle méthode d'estimation : les doubles moindres carrés.
- Méthode très populaire, notamment pour estimer des modèles macroéconomiques à plusieurs équations.
- Des exemples spécifiques seront développés dans le Chapitre 16 : modèles à équations simultanées.
- Exemple : Relation entre le taux de criminalité (meurtres) et la taille des forces de police. Le taux de criminalité dépend de la taille des forces de polices et la taille des forces de police dépend du taux de criminalité. Problème de simultanéité.

Erreurs de mesure

- Problème d'erreurs de mesures sur certaines variables explicatives.

Erreurs de mesure

- Problème d'erreurs de mesures sur certaines variables explicatives.
- Exemple : revenu déclaré.

Omission de variables explicatives

- L'omission de variables explicatives appropriées mène à un biais d'estimation des coefficients de régression.

Omission de variables explicatives

- L'omission de variables explicatives appropriées mène à un biais d'estimation des coefficients de régression.
- Ce biais peut être éliminé si l'on dispose de variables proxy. Quid si pas de proxy.

Omission de variables explicatives

- L’omission de variables explicatives appropriées mène à un biais d’estimation des coefficients de régression.
- Ce biais peut être éliminé si l’on dispose de variables proxy. Quid si pas de proxy.
- Dans un contexte de données de panel, élimination du biais par différenciation si variable omise est indépendante du temps.

Omission de variables explicatives

- L'omission de variables explicatives appropriées mène à un biais d'estimation des coefficients de régression.
- Ce biais peut être éliminé si l'on dispose de variables proxy. Quid si pas de proxy.
- Dans un contexte de données de panel, élimination du biais par différenciation si variable omise est indépendante du temps.
- Quid si la variable omise dépend du temps, si l'intérêt se porte sur des variables qui ne varient pas dans le temps (éliminées par différenciation) ou si on ne dispose pas de données de panel ?

Illustration : variables omises

Exemple : relation salaire-éducation

- Le vrai modèle de régression est :

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e.$$

Exemple : relation salaire-éducation

- Le vrai modèle de régression est :

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e.$$

- Problème : on n'observe pas $abil$ (compétence).

Exemple : relation salaire-éducation

- Le vrai modèle de régression est :

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e.$$

- Problème : on n'observe pas $abil$ (compétence).
- On peut utiliser le QI comme variable proxy de $abil$.
  $\rightarrow$ Quid si indisponible ?

Exemple : relation salaire-éducation

- Le vrai modèle de régression est :

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e.$$

- Problème : on n'observe pas $abil$ (compétence).
- On peut utiliser le QI comme variable proxy de $abil$.
$\longrightarrow$ Quid si indisponible ?
- Si on ignore $abil$, le modèle de régression devient :

$$wage = \beta_0 + \beta_1 educ + u.$$

Exemple : relation salaire-éducation

- Le vrai modèle de régression est :

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e.$$

- Problème : on n'observe pas $abil$ (compétence).
- On peut utiliser le QI comme variable proxy de $abil$.
$\rightarrow$ Quid si indisponible ?
- Si on ignore $abil$, le modèle de régression devient :

$$wage = \beta_0 + \beta_1 educ + u.$$

- $\rightarrow u = \beta_2 abil + e$.
$\rightarrow abil \in u$.
$\rightarrow$ Si $abil$ est corrélé avec $educ$, $E(u|x) \neq 0$.
$\rightarrow$ Estimateur de β_1 par MCO biaisé car violation de l'hypothèse **SLR.3**.

Modèle de régression simple

- Le modèle de régression s'écrit :

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

Modèle de régression simple

- Le modèle de régression s'écrit :

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

- A cause de l'omission de certaines variables :

$$Cov(x, u) \neq 0.$$

$\rightarrow$ Les MCO donnent des estimateurs biaisés.

Modèle de régression simple

- Le modèle de régression s'écrit :

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

- A cause de l'omission de certaines variables :

$$Cov(x, u) \neq 0.$$

$\rightarrow$ Les MCO donnent des estimateurs biaisés.

- Idée des VI $\rightarrow$ utiliser une information additionnelle sous la forme d'une variable observable z qui remplit 2 conditions.

Modèle de régression simple

- Le modèle de régression s'écrit :

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

- A cause de l'omission de certaines variables :

$$Cov(x, u) \neq 0.$$

  $\rightarrow$ Les MCO donnent des estimateurs biaisés.

- Idée des VI $\rightarrow$ utiliser une information additionnelle sous la forme d'une variable observable z qui remplit 2 conditions.
- Condition 1 : $Cov(z, u) = 0$.

Modèle de régression simple

- Le modèle de régression s'écrit :

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

- A cause de l'omission de certaines variables :

$$Cov(x, u) \neq 0.$$

$\rightarrow$ Les MCO donnent des estimateurs biaisés.

- Idée des VI $\rightarrow$ utiliser une information additionnelle sous la forme d'une variable observable z qui remplit 2 conditions.
- Condition 1 : $Cov(z, u) = 0$.
- Condition 2 : $Cov(z, x) \neq 0$.

Conditions 1 des VI

- **Condition 1**. Dans la population,

$$Cov(z, u) = 0.$$

Conditions 1 des VI

- **Condition 1**. Dans la population,

$$Cov(z, u) = 0.$$

- z est dite exogène dans le modèle de régression.

Conditions 1 des VI

- **Condition 1**. Dans la population,

$$Cov(z, u) = 0.$$

- z est dite exogène dans le modèle de régression.
- Il est difficile de tester cette hypothèse.
  $\rightarrow$ Faire d'abord appel à la théorique économique.

Conditions 2 des VI

- **Condition 2**.

$$Cov(z, x) \neq 0.$$

Conditions 2 des VI

- **Condition 2**.

$$Cov(z, x) \neq 0.$$

- z est corrélée avec la variable explicative x.

Conditions 2 des VI

- **Condition 2**.

$$Cov(z, x) \neq 0.$$

- z est corrélée avec la variable explicative x.
- On peut tester cette hypothèse :

$$x = \pi_0 + \pi_1 z + \nu.$$

Conditions 2 des VI

- **Condition 2**.

$$Cov(z, x) \neq 0.$$

- z est corrélée avec la variable explicative x.
- On peut tester cette hypothèse :

$$x = \pi_0 + \pi_1 z + \nu.$$

- Comme $\pi_1 = Cov(z, x)/Var(z)$, la Condition 2 tient ssi on rejette l'hypothèse $H_0 : \pi_1 = 0$ contre l'hypothèse alternative $H_1 : \pi_1 \neq 0$.

Exemples et contre-exemples de VI

Revenons à la relation salaire-éducation.

- Regardons quelques candidats de VI pour $educ$.
  $\rightarrow$ trouver z tel que $Cov(z, abil) = 0$ et $Cov(z, educ) \neq 0$.

Exemples et contre-exemples de VI

Revenons à la relation salaire-éducation.

- Regardons quelques candidats de VI pour $educ$.
  $\rightarrow$ trouver z tel que $Cov(z, abil) = 0$ et $Cov(z, educ) \neq 0$.
- Derniers numéros de sécurité sociale de l'individu :
  Condition 1 OK; Condition 2 non remplie.

Exemples et contre-exemples de VI

Revenons à la relation salaire-éducation.

- Regardons quelques candidats de VI pour $educ$.
$\rightarrow$ trouver z tel que $Cov(z, abil) = 0$ et $Cov(z, educ) \neq 0$.
- Derniers numéros de sécurité sociale de l'individu : Condition 1 OK; Condition 2 non remplie.
- QI : Condition 2 OK mais Condition 1 violée car QI corrélé avec $abil$ et donc u.

Exemples et contre-exemples de VI

Condition 1 : $Cov(z, abil) = 0$ et Condition 2 : $Cov(z, educ) \neq 0$.

- Education de la mère : Condition 2 OK (les parents éduqués ont plus de chance d'avoir des enfants éduqués); Condition 1 moins évidente car corrélation compétence-éducation de la mère pas claire.
  $\rightarrow$ Voir la théorie économique.

Exemples et contre-exemples de VI

Condition 1 : $Cov(z, abil) = 0$ et Condition 2 : $Cov(z, educ) \neq 0$.

- Education de la mère : Condition 2 OK (les parents éduqués ont plus de chance d'avoir des enfants éduqués); Condition 1 moins évidente car corrélation compétence-éducation de la mère pas claire.
  $\rightarrow$ Voir la théorie économique.
- Nombre de frères et soeurs pendant l'enfance.
  $\rightarrow$ Condition 1 vraisemblable.
  $\rightarrow$ Condition 2 aussi car certaines études ont montré que les familles nombreuses avaient un niveau d'éducation significativement plus faible.

Estimation et Inférence par la méthode des variables instrumentales

Estimateurs des VI

Si l'on dispose d'une bonne VI, on peut estimer de manière consistante l'équation $y = \beta_0 + \beta_1 x + u$.

- Les 2 conditions des VI permettent d'identifier β_1, c'est-à-dire d'écrire β_1 en termes de moments des variables observables :
  $Cov(z, y) = Cov(z, \beta_0 + \beta_1 x + u) = \beta_1 Cov(z, x) + Cov(z, u)$.

Estimateurs des VI

Si l'on dispose d'une bonne VI, on peut estimer de manière consistante l'équation $y = \beta_0 + \beta_1 x + u$.

- Les 2 conditions des VI permettent d'identifier β_1, c'est-à-dire d'écrire β_1 en termes de moments des variables observables : $Cov(z, y) = Cov(z, \beta_0 + \beta_1 x + u) = \beta_1 Cov(z, x) + Cov(z, u)$.
- Voir pages 712-713 pour les propriétés de calcul des covariances.

Estimateurs des VI

Si l'on dispose d'une bonne VI, on peut estimer de manière consistante l'équation $y = \beta_0 + \beta_1 x + u$.

- Les 2 conditions des VI permettent d'identifier β_1, c'est-à-dire d'écrire β_1 en termes de moments des variables observables : $Cov(z, y) = Cov(z, \beta_0 + \beta_1 x + u) = \beta_1 Cov(z, x) + Cov(z, u)$.
- Voir pages 712-713 pour les propriétés de calcul des covariances.
- Si $Cov(z, u) = 0$, alors $\beta_1 = \frac{Cov(z,y)}{Cov(z,x)}$ qui existe ssi $Cov(z, x) \neq 0$.

Estimateurs des VI

- L'estimateur par VI de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}.$$

Estimateurs des VI

- L'estimateur par VI de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}.$$

- Remarque : si $x = z$ (cas de x exogène), alors on retrouve l'estimateur des MCO. Dans ce cas, x est son propre instrument.

Estimateurs des VI

- L'estimateur par VI de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}.$$

- Remarque : si $x = z$ (cas de x exogène), alors on retrouve l'estimateur des MCO. Dans ce cas, x est son propre instrument.
- L'estimateur de l'intercept est simplement : $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.

Estimateurs des VI

- L'estimateur par VI de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}.$$

- Remarque : si $x = z$ (cas de x exogène), alors on retrouve l'estimateur des MCO. Dans ce cas, x est son propre instrument.
- L'estimateur de l'intercept est simplement : $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$.
- Sous les conditions 1 et 2, $plim(\hat{\beta}_1) = \beta_1 \rightarrow \hat{\beta}_1$ est convergent.

Inférence avec les estimateurs des VI

- Les estimateurs VI sont, comme les estimateurs MCO, asymptotiquement distribués normalement. Pour faire de l'inférence, on a donc besoin d'un estimateur de la variance de $\hat{\beta}_0$ et $\hat{\beta}_1$.

Inférence avec les estimateurs des VI

- Les estimateurs VI sont, comme les estimateurs MCO, asymptotiquement distribués normalement. Pour faire de l'inférence, on a donc besoin d'un estimateur de la variance de $\hat{\beta}_0$ et $\hat{\beta}_1$.
- On fait l'hypothèse d'homoscédasticité : $E(u^2|z) = Var(u) = \sigma^2$.

Inférence avec les estimateurs des VI

- Les estimateurs VI sont, comme les estimateurs MCO, asymptotiquement distribués normalement. Pour faire de l'inférence, on a donc besoin d'un estimateur de la variance de $\hat{\beta}_0$ et $\hat{\beta}_1$.
- On fait l'hypothèse d'homoscédasticité : $E(u^2|z) = Var(u) = \sigma^2$.
- Sous les conditions 1 et 2 des VI et cette hypothèse, alors la variance asymptotique de $\hat{\beta}_1$ est égale à $\frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2}$ où ρ_{xz} est la corrélation entre x et z.

Inférence avec les estimateurs des VI

- Les estimateurs VI sont, comme les estimateurs MCO, asymptotiquement distribués normalement. Pour faire de l'inférence, on a donc besoin d'un estimateur de la variance de $\hat{\beta}_0$ et $\hat{\beta}_1$.
- On fait l'hypothèse d'homoscédasticité : $E(u^2|z) = Var(u) = \sigma^2$.
- Sous les conditions 1 et 2 des VI et cette hypothèse, alors la variance asymptotique de $\hat{\beta}_1$ est égale à $\frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2}$ où ρ_{xz} est la corrélation entre x et z.
- Un estimateur consistent de σ^2 peut s'obtenir comme pour les MCO, c-à-d $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$.

Inférence avec les Estimateurs des VI

- La variance asymptotique de β_1 peut également s'écrire : $\frac{\hat{\sigma}^2}{SCT_x R^2_{x,z}}$, où $SCT_x = \sum_{i=1}^n x_i^2$.

Inférence avec les Estimateurs des VI

- La variance asymptotique de β_1 peut également s'écrire : $\frac{\hat{\sigma}^2}{SCT_x R^2_{x,z}}$, où $SCT_x = \sum_{i=1}^n x_i^2$.
- Comparons avec la variance asymptotique de β_1 estimé par MCO : $\frac{\hat{\sigma}^2}{SCT_x}$.

Inférence avec les Estimateurs des VI

- La variance asymptotique de β_1 peut également s'écrire : $\frac{\hat{\sigma}^2}{SCT_x R^2_{x,z}}$, où $SCT_x = \sum_{i=1}^n x_i^2$.
- Comparons avec la variance asymptotique de β_1 estimé par MCO : $\frac{\hat{\sigma}^2}{SCT_x}$.
- On voit que plus $R^2_{x,z}$ (R^2 de la régression de X_i sur Z_i) est faible (instruments faibles), plus la précision de l'estimateur des VI est faible.

Inférence avec les Estimateurs des VI

- La variance asymptotique de β_1 peut également s'écrire : $\frac{\hat{\sigma}^2}{SCT_x R^2_{x,z}}$, où $SCT_x = \sum_{i=1}^n x_i^2$.
- Comparons avec la variance asymptotique de β_1 estimé par MCO : $\frac{\hat{\sigma}^2}{SCT_x}$.
- On voit que plus $R^2_{x,z}$ (R^2 de la régression de X_i sur Z_i) est faible (instruments faibles), plus la précision de l'estimateur des VI est faible.
- Comme $R^2_{x,z} < 1$, la variance des estimateurs VI et des DMC est toujours plus élevée que celle des MCO : le coût de l'estimation des VI se paye en termes de précision de l'estimateur.

Illustration : mroz.in7

```
EQ(1) Modelling LWAGE by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant            -0.185197     0.1852    -1.00   0.318
EDUC                 0.108649    0.01440     7.55   0.000

sigma                0.680032  RSS             197.001022
R^2                  0.117883  F(1,426) = 56.93 [0.000]**
log-likelihood        -441.26  DW                    1.98
no. of observations       428  no. of parameters        2
mean(LWAGE)           1.19017  var(LWAGE)        0.521793
```

$\rightarrow \log(wage) = \beta_0 + \beta_1 educ + u$.

$\rightarrow$ Hors si $abil \in u$ et $Cov(abil, educ) \neq 0 \rightarrow$ Rejet **SLR.3**.

$\rightarrow$ VI : Education du père $(fathereduc)$?

Tester la Condition 2

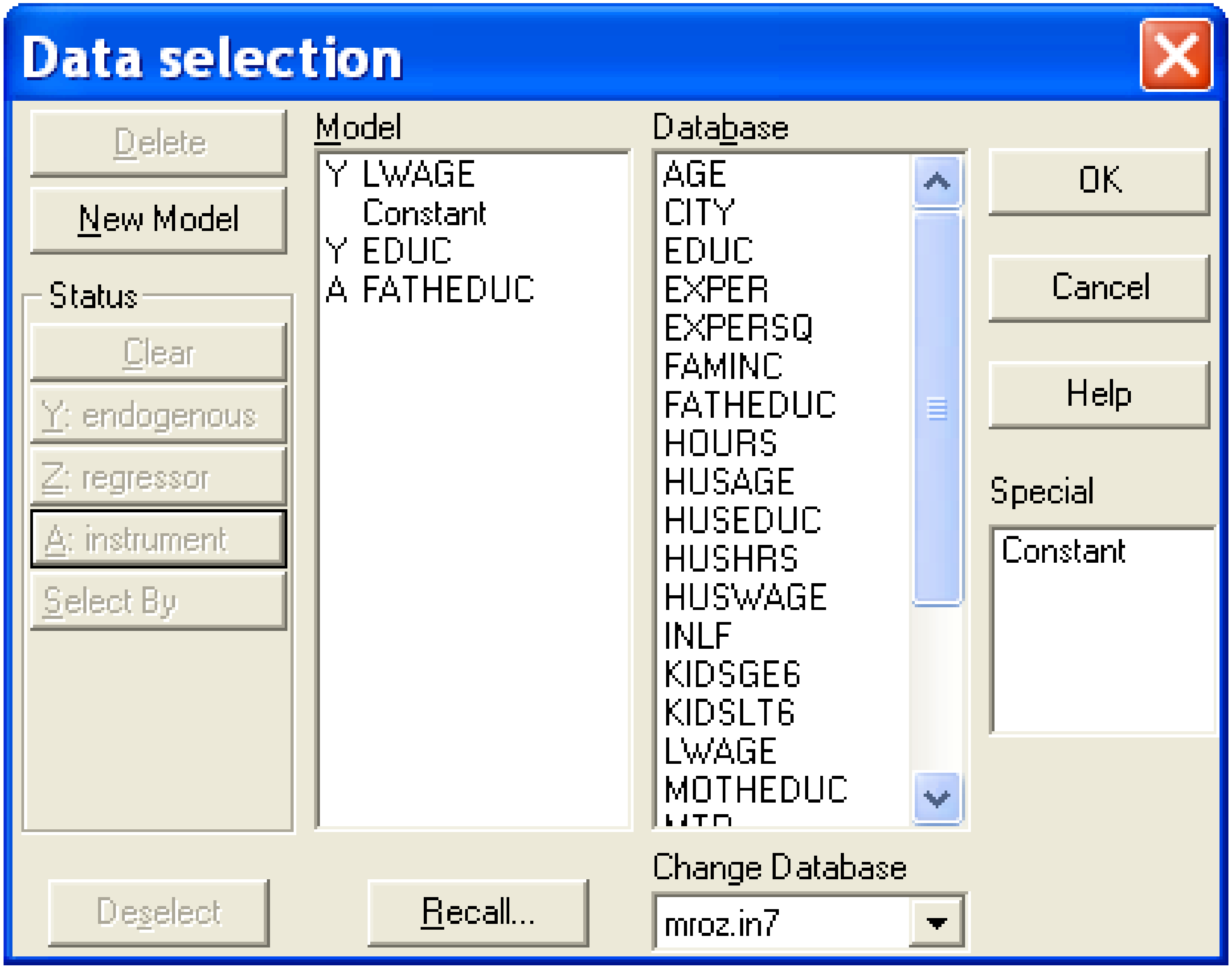
```
EQ( 2) Modelling EDUC by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              10.2371     0.2759     37.1   0.000   0.7636
FATHEDUC             0.269442    0.02859     9.43   0.000   0.1726

sigma                  2.0813  RSS                  1845.35428
R^2                   0.17256  F(1,426) =    88.84 [0.000]**
log-likelihood       -920.025  DW                         1.92
no. of observations       428  no. of parameters             2
mean(EDUC)            12.6589  var(EDUC)               5.21074
```

$\longrightarrow x = \pi_0 + \pi_1 z + \nu$.

$\longrightarrow R^2 = 0.17256$.

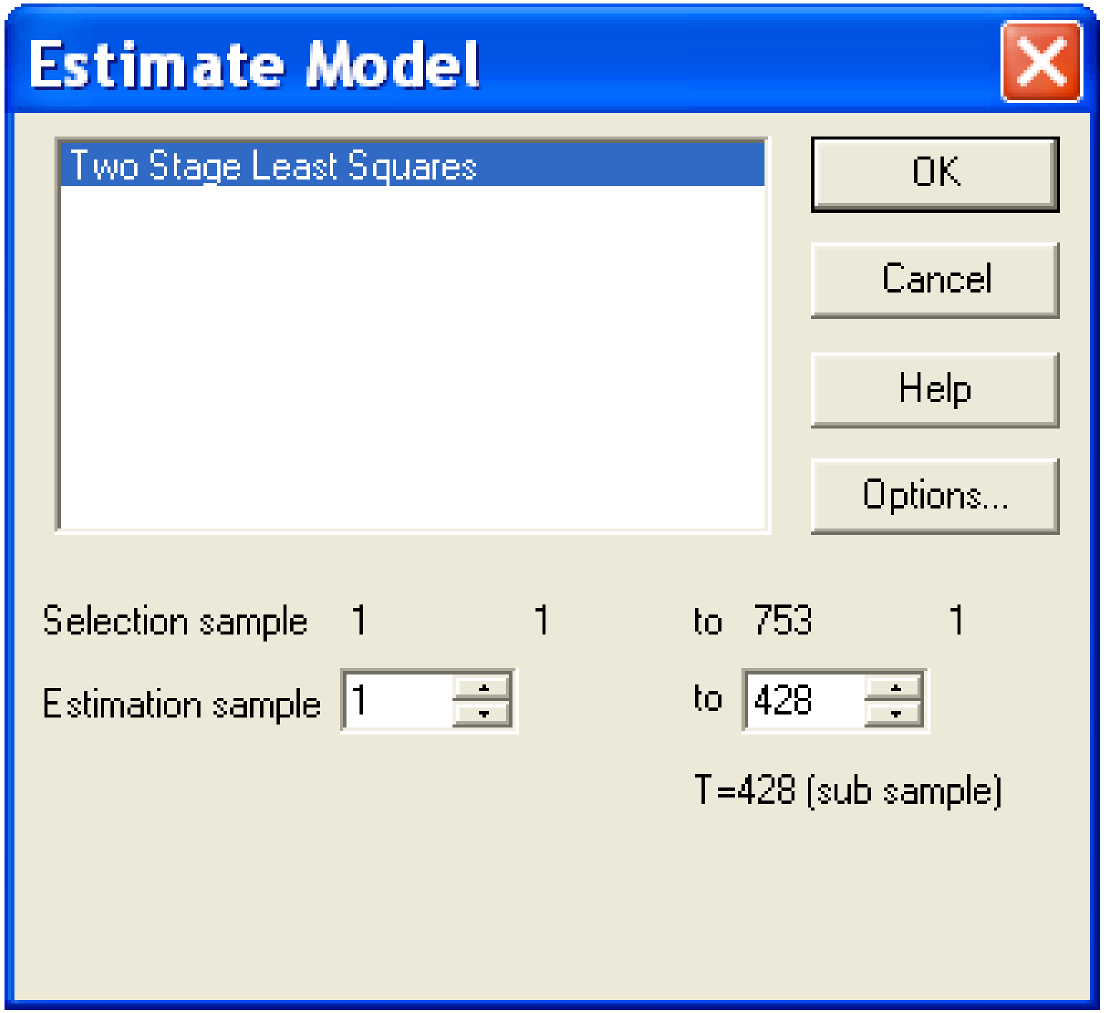
$\longrightarrow$ t-stat de $\hat{\pi}_1 = 9.43 \longrightarrow$ rejet de $H_0 : \pi_1 = 0$.

PcGive : Sélection des variables



→ Sélectionner y en premier lieu. $EDUC : Y$.

PcGive : Méthode



Output PcGive : VI

```
EQ(3) Modelling LWAGE by IVE-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
EDUC           Y    0.0591735    0.03514     1.68   0.093
Constant             0.441103     0.4461    0.989   0.323

sigma                 0.68939  RSS              202.46008
Reduced form sigma    0.72186
no. of observations       428  no. of parameters        2
no. endogenous variables    2  no. of instruments       2
mean(LWAGE)           1.19017  var(LWAGE)        0.521793
Additional instruments:
[0] = FATHEDUC
Testing beta = 0:   Chiˆ2(1) =   2.8354 [0.0922]
```

$\longrightarrow$ L'estimation par VI donne un impact plus faible (0.059 vs. 0.109) et des écart-types plus élevés que ceux obtenus par MCO (0.035 vs. 0.014).

$\longrightarrow$ `Testing beta` correspond au F-test $H_0 : \beta_1 = \ldots = \beta_k = 0$ et suit une $\chi^2(k)$ sous H_0.

Propriétés des VI avec VI faibles

- Instruments faibles = VI pour lesquelles la corrélation entre x et z est faible $\rightarrow$ conséquences non désirables.

Propriétés des VI avec VI faibles

- Instruments faibles = VI pour lesquelles la corrélation entre x et z est faible $\rightarrow$ conséquences non désirables.
- Rappel pour les MCO :

$$
\begin{aligned}
plim\ \hat{\beta}_1 &= \beta_1 + \frac{Cov(x,u)}{Var(x)} \\
&= \beta_1 + Corr(x,u)\frac{\sigma_u}{\sigma_x} \\
&= \beta_1, \text{si } Corr(x,u) = 0.
\end{aligned}
$$

Propriétés des VI avec VI faibles

- Instruments faibles = VI pour lesquelles la corrélation entre x et z est faible $\rightarrow$ conséquences non désirables.
- Rappel pour les MCO :

$$\begin{aligned} plim\ \hat{\beta}_1 &= \beta_1 + \frac{Cov(x,u)}{Var(x)} \\ &= \beta_1 + Corr(x,u)\frac{\sigma_u}{\sigma_x} \\ &= \beta_1, \text{si } Corr(x,u) = 0. \end{aligned}$$

- Pour les VI :

$$plim\ \hat{\beta}_1 = \beta_1 + \left[\frac{Corr(z,u)}{Corr(z,x)}\left(\frac{\sigma_u}{\sigma_x}\right)\right].$$

Propriétés des VI avec VI faibles

- Si $Corr(z,x)$ est faible $\rightarrow$ estimateur largement inconsistent.

Propriétés des VI avec VI faibles

- Si $Corr(z, x)$ est faible $\rightarrow$ estimateur largement inconsistent.
- Et ce, même si $Corr(z, u)$ est faible.

Propriétés des VI avec VI faibles

- Si $Corr(z, x)$ est faible $\rightarrow$ estimateur largement inconsistent.
- Et ce, même si $Corr(z, u)$ est faible.
- En termes de consistence, les VI sont préférables aux MCO ssi $\frac{Corr(z,u)}{Corr(z,x)} < Corr(x, u)$.

Propriétés des VI avec VI faibles

- Si $Corr(z,x)$ est faible $\rightarrow$ estimateur largement inconsistent.
- Et ce, même si $Corr(z,u)$ est faible.
- En termes de consistence, les VI sont préférables aux MCO ssi $\frac{Corr(z,u)}{Corr(z,x)} < Corr(x,u)$.
- $\rightarrow$ Même en cas d'endogénéité, les MCO peuvent être préférables : tout dépend de la corrélation entre la VI et x.

Illustration

- Relation entre le poids du bébé à la naissance et le nombre de cigarettes fumées par la mère.

Illustration

- Relation entre le poids du bébé à la naissance et le nombre de cigarettes fumées par la mère.
- Exemple : bwght.in7.

Illustration

- Relation entre le poids du bébé à la naissance et le nombre de cigarettes fumées par la mère.
- Exemple : bwght.in7.
- Problème : le nombre de cigarettes peut être corrélé avec des facteurs de santé non observés dans la régression $\rightarrow$ corrélation entre x et u.

Illustration

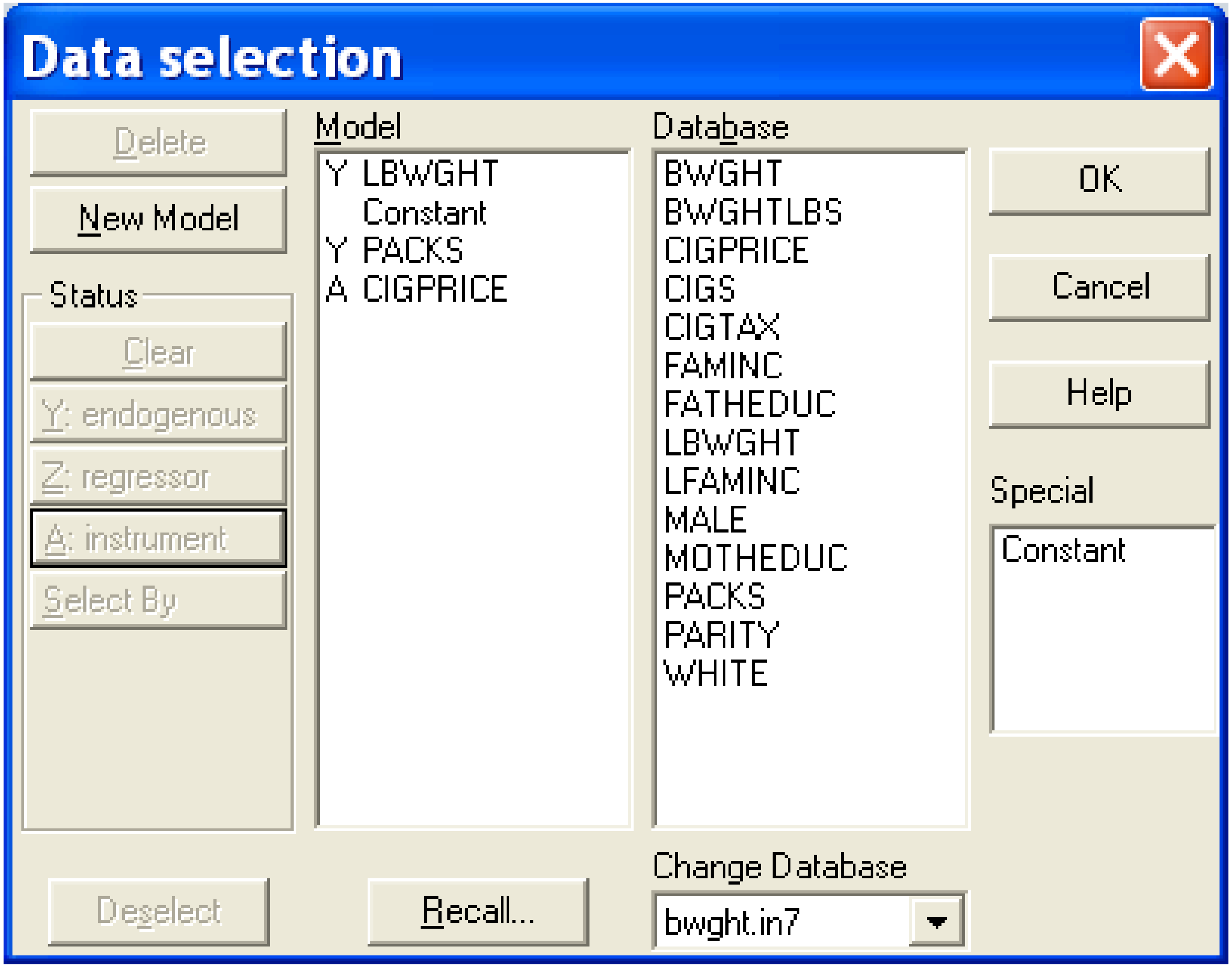
- Relation entre le poids du bébé à la naissance et le nombre de cigarettes fumées par la mère.
- Exemple : bwght.in7.
- Problème : le nombre de cigarettes peut être corrélé avec des facteurs de santé non observés dans la régression $\rightarrow$ corrélation entre x et u.
- Idée de VI : prix des cigarettes dans la région d'habitation.

Output PcGive : MCO

```
EQ(1) Modelling LBWGHT by OLS-CS (using bwght.in7)
       The estimation sample is: 1 to 1388

                  Coefficient  Std.Error  t-value   t-prob
Constant              4.76940   0.005369     888.    0.000
PACKS              -0.0898131    0.01698    -5.29    0.000

sigma                0.188834  RSS              49.4225599
R^2                 0.0197893  F(1,1386) = 27.98 [0.000]**
log-likelihood        345.151  DW                     1.93
no. of observations      1388  no. of parameters         2
mean(LBWGHT)          4.76003  var(LBWGHT)       0.0363259
```

$\longrightarrow$ Fumer à un effet négatif sur le poids des bébés.

PcGive : Sélection des variables



Output PcGive : VI $= CIGPRICE$

```
EQ(2) Modelling LBWGHT by IVE-CS (using bwght.in7)
       The estimation sample is: 1 to 1388

                  Coefficient  Std.Error  t-value  t-prob
PACKS           Y     2.98868      8.699    0.344   0.731
Constant              4.44814     0.9082     4.90   0.000

sigma                0.938861  RSS             1221.70415
Reduced form sigma    0.19053
no. of observations      1388  no. of parameters        2
no. endogenous variables    2  no. of instruments       2
mean(LBWGHT)          4.76003  var(LBWGHT)      0.0363259
Additional instruments:
[0] = CIGPRICE
Testing beta = 0:   Chi^2(1) =  0.11804 [0.7312]
```

$\longrightarrow$ Fumer à un effet positif mais non significatif sur le poids des bébés.

Tester la Condition 2

```
EQ(3) Modelling CIGPRICE by OLS-CS (using bwght.in7)
       The estimation sample is: 1 to 1388

                  Coefficient  Std.Error  t-value  t-prob
Constant              130.524     0.2914     448.   0.000
PACKS                0.332897     0.9214    0.361   0.718

sigma                 10.2477  RSS              145551.189
R^2               9.41712e-005  F(1,1386) =  0.1305 [0.718]
log-likelihood       -5198.44  DW                     0.102
no. of observations      1388  no. of parameters          2
mean(CIGPRICE)        130.559  var(CIGPRICE)        104.874
```

$\longrightarrow x = \pi_0 + \pi_1 z + \nu$.

$\longrightarrow R^2 = 9.41712e - 005$.

$\longrightarrow$ t-stat de $\hat{\pi}_1 = 0.361 \longrightarrow$ non-rejet de $H_0 : \pi_1 = 0$.

Autre exemple

- Estimation d'une fonction d'offre et de demande de poisson sur des données du 'Fulton Fish Market' à N-Y City.

Autre exemple

- Estimation d'une fonction d'offre et de demande de poisson sur des données du 'Fulton Fish Market' à N-Y City.
- Fish: Whiting (merlan en français).

$$\text{S: QTY} = \underset{(1.03)}{+10.29} - \underset{(0.193)}{0.3986}\ \text{PRICE}_t - \underset{(0.354)}{0.6429}\ \text{WINDSPD}_t$$

$$\text{D: QTY} = \underset{(0.0762)}{8.419} - \underset{(0.179)}{0.5409}\ \text{PRICE}_t$$

Autre exemple

- Estimation d'une fonction d'offre et de demande de poisson sur des données du 'Fulton Fish Market' à N-Y City.
- Fish: Whiting (merlan en français).

$$\text{S: QTY} = \underset{(1.03)}{+10.29} - \underset{(0.193)}{0.3986}\ \text{PRICE}_t - \underset{(0.354)}{0.6429}\ \text{WINDSPD}_t$$

$$\text{D: QTY} = \underset{(0.0762)}{8.419} - \underset{(0.179)}{0.5409}\ \text{PRICE}_t$$

- Les estimateurs MCO de l'équation d'offre sont incompatibles avec la théorie économique qui appelle une pente positive ou nulle de la fonction d'offre.

Autre exemple

- Estimation d'une fonction d'offre et de demande de poisson sur des données du 'Fulton Fish Market' à N-Y City.
- Fish: Whiting (merlan en français).

$$\text{S: QTY} = \underset{(1.03)}{+10.29} - \underset{(0.193)}{0.3986}\ \text{PRICE}_t - \underset{(0.354)}{0.6429}\ \text{WINDSPD}_t$$

$$\text{D: QTY} = \underset{(0.0762)}{8.419} - \underset{(0.179)}{0.5409}\ \text{PRICE}_t$$

- Les estimateurs MCO de l'équation d'offre sont incompatibles avec la théorie économique qui appelle une pente positive ou
- nulle de la fonction d'offre.

$$\text{D (IV avec WINDSPD): QTY} = -\underset{(0.473)}{1.265}\ \text{PRICE}_t + \underset{(0.117)}{8.278}$$

Estimation par VI en régression multiple

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u$.
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.
- y_1 : variable **dépendante ou endogène** $\rightarrow log(wage)$.

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u$.
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.
- y_1 : variable **dépendante ou endogène** $\rightarrow log(wage)$.
- y_2 : variable **explicative endogène** (corrélée avec u_1)
  $\rightarrow educ$.

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.
- y_1 : variable **dépendante ou endogène** $\rightarrow log(wage)$.
- y_2 : variable **explicative endogène** (corrélée avec u_1)
  $\rightarrow educ$.
- z_1 : variable explicative **exogène** (non corrélée avec u_1)
  $\rightarrow exper$.

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.
- y_1 : variable **dépendante ou endogène** $\rightarrow log(wage)$.
- y_2 : variable **explicative endogène** (corrélée avec u_1)
  $\rightarrow educ$.
- z_1 : variable explicative **exogène** (non corrélée avec u_1)
  $\rightarrow exper$.
- $E(u_1) = 0, Corr(u_1, z_1) = 0$ mais $Corr(u_1, y_2) \neq 0$.

Régression multiple

- Exemple : $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u$.
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$.
  $\rightarrow$ Equation structurelle.
  $\rightarrow$ Objectif = estimation des β_j.
- y_1 : variable **dépendante ou endogène** $\rightarrow log(wage)$.
- y_2 : variable **explicative endogène** (corrélée avec u_1)
  $\rightarrow educ$.
- z_1 : variable explicative **exogène** (non corrélée avec u_1)
  $\rightarrow exper$.
- $E(u_1) = 0, Corr(u_1, z_1) = 0$ mais $Corr(u_1, y_2) \neq 0$.
- Si estimation par MCO : tous les $\hat{\beta}_j$ seront biaisés.

Régression multiple

- z_1 ne peut pas servir d'instrument pour y_2 (car elle apparaît comme variable explicative) $\rightarrow$ recourt à z_2, une VI telle que $Cov(z_2, u_1) = 0$ et $Cov(z_2, y_2) \neq 0$.

Régression multiple

- z_1 ne peut pas servir d'instrument pour y_2 (car elle apparaît comme variable explicative) $\rightarrow$ recourt à z_2, une VI telle que $Cov(z_2, u_1) = 0$ et $Cov(z_2, y_2) \neq 0$.
- Hypothèses cruciales : $E(u_1) = 0$, $Cov(z_1, u_1) = 0$, $Cov(z_2, u_1) = 0 \Rightarrow E(u_1) = E(z_1 u_1) = E(z_2 u_1) = 0$.

Régression multiple

- z_1 ne peut pas servir d'instrument pour y_2 (car elle apparaît comme variable explicative) $\rightarrow$ recourt à z_2, une VI telle que $Cov(z_2, u_1) = 0$ et $Cov(z_2, y_2) \neq 0$.
- Hypothèses cruciales : $E(u_1) = 0$, $Cov(z_1, u_1) = 0$, $Cov(z_2, u_1) = 0 \Rightarrow E(u_1) = E(z_1 u_1) = E(z_2 u_1) = 0$.
- Approche par la méthode des moments :

$$\begin{aligned}
\sum_{i=1}^{n}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) &= 0 \\
\sum_{i=1}^{n} z_{i1}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) &= 0 \\
\sum_{i=1}^{n} z_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) &= 0.
\end{aligned}$$

Régression multiple

- 3 équations linéaires pour 3 inconnues $(\beta_0, \beta_1, \beta_2)$.

Régression multiple

- 3 équations linéaires pour 3 inconnues $(\beta_0,\beta_1,\beta_2)$.
- Si $z_2 = y_2$ (c-à-d y_2 exogène), on retrouve les conditions du premier ordre des MCO.

Régression multiple

- 3 équations linéaires pour 3 inconnues $(\beta_0,\beta_1,\beta_2)$.
- Si $z_2 = y_2$ (c-à-d y_2 exogène), on retrouve les conditions du premier ordre des MCO.
- La variable instrumentale z_2 doit être conditionnellement corrélée avec y_2 :

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \nu_2.$$

Il s'agit d'un exemple d'**équation en forme réduite** : la variable endogène est exprimée exclusivement en termes de variables exogènes.

Régression multiple

- 3 équations linéaires pour 3 inconnues $(\beta_0,\beta_1,\beta_2)$.
- Si $z_2 = y_2$ (c-à-d y_2 exogène), on retrouve les conditions du premier ordre des MCO.
- La variable instrumentale z_2 doit être conditionnellement corrélée avec y_2 :

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \nu_2.$$

  Il s'agit d'un exemple d'**équation en forme réduite** : la variable endogène est exprimée exclusivement en termes de variables exogènes.
- La condition d'identification (généralisation de la Condition 2 vue précédemment) est : $\pi_2 \neq 0$.

Exemple de régression multiple : card.in7

- Lien éducation-salaire : 14 variables exogènes $\rightarrow$ expérience $(EXPER)$, expérience au carré $(EXPERSQ)$, une dummy pour noir $(BLACK)$, une dummy pour zone urbaine $(SMSA)$, une dummy pour zone urbaine en 1966 $(SMSA66)$, une dummy pour état du sud $(SOUTH)$ et 5 variables binaires régionales $(REG662 - REG669)$.

Exemple de régression multiple : card.in7

- Lien éducation-salaire : 14 variables exogènes $\rightarrow$ expérience $(EXPER)$, expérience au carré $(EXPERSQ)$, une dummy pour noir $(BLACK)$, une dummy pour zone urbaine $(SMSA)$, une dummy pour zone urbaine en 1966 $(SMSA66)$, une dummy pour état du sud $(SOUTH)$ et 5 variables binaires régionales $(REG662 - REG669)$.
- Une variable explicative endogène : $EDUC$.

Exemple de régression multiple : card.in7

- Lien éducation-salaire : 14 variables exogènes $\rightarrow$ expérience $(EXPER)$, expérience au carré $(EXPERSQ)$, une dummy pour noir $(BLACK)$, une dummy pour zone urbaine $(SMSA)$, une dummy pour zone urbaine en 1966 $(SMSA66)$, une dummy pour état du sud $(SOUTH)$ et 5 variables binaires régionales $(REG662 - REG669)$.
- Une variable explicative endogène : $EDUC$.
- Une VI candidate pour $EDUC$: dummy pour habitation proche d'une école $(NEARC4)$.

Output PcGive : MCO

```
EQ(1) Modelling LWAGE by OLS-CS (using card.in7)
       The estimation sample is: 1 to 3010

                  Coefficient  Std.Error  t-value    t-prob
Constant              4.62081    0.07423     62.2     0.000
EDUC                0.0746933   0.003498     21.4     0.000
BLACK               -0.199012    0.01825    -10.9     0.000
EXPER               0.0848320   0.006624     12.8     0.000
EXPERSQ           -0.00228704  0.0003166    -7.22     0.000
SMSA                 0.136385    0.02010     6.79     0.000
SOUTH               -0.147955    0.02598    -5.69     0.000
REG662              0.0963672    0.03590     2.68     0.007
...
SMSA66              0.0262417    0.01945     1.35     0.177

sigma                 0.37228  RSS                 414.946049
R^2                  0.299836  F(15,2994) =  85.48 [0.000]**
log-likelihood       -1288.78  DW                        1.88
no. of observations      3010  no. of parameters           16
```

Tester la Condition 2

```
EQ(2) Modelling EDUC by OLS-CS (using card.in7)
       The estimation sample is: 1 to 3010

                    Coefficient  Std.Error  t-value    t-prob
Constant                16.6383     0.2406     69.1     0.000
BLACK                 -0.935529    0.09373    -9.98     0.000
EXPER                 -0.412533    0.03370    -12.2     0.000
EXPERSQ             0.000868574   0.001650    0.526     0.599
SMSA                   0.402182     0.1048     3.84     0.000
SOUTH                -0.0516126     0.1354   -0.381     0.703
REG662               -0.0786363     0.1871   -0.420     0.674
...
SMSA66                0.0254805     0.1058    0.241     0.810
NEARC4                 0.319899    0.08786     3.64     0.000
```

$\longrightarrow y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_{14} z_{14} + \pi_{15} z_{15} + \nu.$

$\longrightarrow$ t-stat de $\hat{\pi}_{15} = 3.64 \longrightarrow$ rejet de $H_0 : \pi_{15} = 0.$

Estimation par VI

```
EQ(3) Modelling LWAGE by IVE-CS (using card.in7)
       The estimation sample is: 1 to 3010

                  Coefficient  Std.Error  t-value  t-prob
EDUC           Y     0.131504    0.05496     2.39   0.017
Constant              3.66615     0.9248     3.96   0.000
BLACK               -0.146776    0.05390    -2.72   0.007
EXPER                0.108271    0.02366     4.58   0.000
...
SMSA66              0.0185311    0.02161    0.858   0.391

Additional instruments:
[0] = NEARC4
Testing beta = 0:   Chi^2(15)=   765.11 [0.0000]**
```

$\longrightarrow$ L'estimation par VI donne un rendement de l'éducation plus élevé que l'estimation par MCO (0.132 vs. 0.075) mais une plus grande incertitude (0.055 vs. 0.003).

Doubles Moindres Carrés

Variable explicative endogène unique

- Soit le modèle de régression : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$.

Variable explicative endogène unique

- Soit le modèle de régression : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$.
- On dispose de 2 variables exogènes exclues : z_2 et z_3 $\rightarrow$ 2 restrictions d'exclusion.

Variable explicative endogène unique

- Soit le modèle de régression : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$.
- On dispose de 2 variables exogènes exclues : z_2 et z_3 $\rightarrow$ 2 restrictions d'exclusion.
- Si z_2 et z_3 sont corrélées avec y_2, on dispose de 2 VI potentielles.

Variable explicative endogène unique

- Soit le modèle de régression : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$.
- On dispose de 2 variables exogènes exclues : z_2 et z_3 $\rightarrow$ 2 restrictions d'exclusion.
- Si z_2 et z_3 sont corrélées avec y_2, on dispose de 2 VI potentielles.
- Dans ce cas, on va choisir la combinaison linéaire des variables exogènes z_1,z_2,z_3 la plus corrélée avec $y_2 \rightarrow$ y_2^*.

Variable explicative endogène unique

- Soit l'équation en forme réduite :
  $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2$.

Variable explicative endogène unique

- Soit l'équation en forme réduite : $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2$.
- Les hypothèses sont les suivantes : $E(\nu_2) = 0$, $Cov(z_1, \nu_2) = 0$, $Cov(z_2, \nu_2) = 0$ et $Cov(z_3, \nu_2) = 0$.

Variable explicative endogène unique

- Soit l'équation en forme réduite :
  $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2$.
- Les hypothèses sont les suivantes : $E(\nu_2) = 0$, $Cov(z_1, \nu_2) = 0$, $Cov(z_2, \nu_2) = 0$ et $Cov(z_3, \nu_2) = 0$.
- $y_2^* = E(y_2 | z_1, z_2, z_3) = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3$.

Variable explicative endogène unique

- Soit l'équation en forme réduite :
  $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2$.
- Les hypothèses sont les suivantes : $E(\nu_2) = 0$, $Cov(z_1, \nu_2) = 0$, $Cov(z_2, \nu_2) = 0$ et $Cov(z_3, \nu_2) = 0$.
- $y_2^* = E(y_2 | z_1, z_2, z_3) = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3$.
- La condition d'identification devient : $\pi_2 \neq 0$ ou $\pi_3 \neq 0$. Cette condition peut s'évaluer en testant H_0 : $\pi_2 = 0$ et $\pi_3 = 0$ à l'aide d'un F-test.

Les doubles moindres carrés (DMC)

La méthode des DMC met en oeuvre ce qu'on vient de voir à travers 2 étapes $(y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2)$.

- Étape 1 : On estime la meilleure combinaison linéaire des VI en régressant par MCO y_2 sur z_1,z_2 et z_3 $\rightarrow \hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3 = y_2 - \hat{\nu}_2$.

Les doubles moindres carrés (DMC)

La méthode des DMC met en oeuvre ce qu'on vient de voir à travers 2 étapes $(y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2)$.

- Étape 1 : On estime la meilleure combinaison linéaire des VI en régressant par MCO y_2 sur z_1,z_2 et z_3 $\rightarrow \hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3 = y_2 - \hat{\nu}_2$.
- Étape 2 : on utilise $\hat{y}_2$ comme instrument de y_2. La troisième condition des moments devient :

$$\sum_{i=1}^{n} \hat{y}_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0.$$

Les doubles moindres carrés (DMC)

La méthode des DMC met en oeuvre ce qu'on vient de voir à travers 2 étapes $(y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \nu_2)$.

- Étape 1 : On estime la meilleure combinaison linéaire des VI en régressant par MCO y_2 sur z_1,z_2 et z_3 $\rightarrow \hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3 = y_2 - \hat{\nu}_2$.
- Étape 2 : on utilise $\hat{y}_2$ comme instrument de y_2. La troisième condition des moments devient :

$$\sum_{i=1}^{n} \hat{y}_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0.$$

- L'étape 2 implique donc simplement d'utiliser $\hat{y}_2$ obtenu en première étape comme régresseur : $y_1 = \beta_0 + \beta_1 \hat{y}_2 + \beta_2 z_1 + \nu_2$.

Illustration des DMC

Reprenons l'exemple de la relation salaire-éducation (mroz.in7).

- Equation structurelle :

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1.$$

$exper$ et $exper^2$ sont non corrélés avec u_1.

Illustration des DMC

Reprenons l'exemple de la relation salaire-éducation (mroz.in7).

- Equation structurelle :

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1.$$

  $exper$ et $exper^2$ sont non corrélés avec u_1.

- On va utiliser l'éducation des parents comme VI : $fatheduc$ et $motheduc$.

Illustration des DMC

Reprenons l'exemple de la relation salaire-éducation (mroz.in7).

- Equation structurelle :

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1.$$

  $exper$ et $exper^2$ sont non corrélés avec u_1.
- On va utiliser l'éducation des parents comme VI : $fatheduc$ et $motheduc$.
- L'equation en forme réduite s'écrit donc : $educ = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 motheduc + \pi_4 fatheduc + \nu_2$.

Illustration des DMC

Reprenons l'exemple de la relation salaire-éducation (mroz.in7).

- Equation structurelle :

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1.$$

  $exper$ et $exper^2$ sont non corrélés avec u_1.

- On va utiliser l'éducation des parents comme VI : $fatheduc$ et $motheduc$.
- L'equation en forme réduite s'écrit donc : $educ = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 motheduc + \pi_4 fatheduc + \nu_2$.
- La condition d'identification requiert donc : $\pi_3 \neq 0$ ou $\pi_4 \neq 0$, ce qui peut se tester par un test classique en F.

Output PcGive : MCO

```
EQ(1) Modelling LWAGE by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant            -0.522041     0.1986    -2.63   0.009
EDUC                 0.107490    0.01415     7.60   0.000
EXPER               0.0415665    0.01318     3.15   0.002
EXPERSQ          -0.000811193  0.0003932    -2.06   0.040

sigma                 0.66642  RSS                188.305144
R^2                   0.15682  F(3,424) = 26.29 [0.000]**
log-likelihood       -431.599  DW                       1.96
no. of observations       428  no. of parameters           4
mean(LWAGE)           1.19017  var(LWAGE)           0.521793
```

Condition d'indentification

```
EQ(2) Modelling EDUC by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 753

                   Coefficient  Std.Error  t-value  t-prob
Constant               8.36672     0.2667     31.4   0.000
EXPER                0.0853780    0.02555     3.34   0.001
EXPERSQ            -0.00185645  0.0008276    -2.24   0.025
FATHEDUC              0.184575    0.02450     7.53   0.000
MOTHEDUC              0.185617    0.02599     7.14   0.000

sigma                   1.9636  RSS              2884.0966
R^2                   0.262387  F(4,748) = 66.52 [0.000]**
log-likelihood        -1574.06  DW                       2
no. of observations        753  no. of parameters        5
mean(EDUC)             12.2869  var(EDUC)          5.19262
```

$\longrightarrow$ Equation en forme réduite.

Condition d'indentification : F-test

```
Test for linear restrictions (Rb=r):
R matrix
Constant      EXPER      EXPERSQ     FATHEDUC     MOTHEDUC
 0.00000   0.00000      0.00000       1.0000      0.00000
 0.00000   0.00000      0.00000      0.00000       1.0000
r vector
      0.00000      0.00000
LinRes F(2,748) = 124.76 [0.0000]**
```

$\longrightarrow$ Rejet de $H_0 : \pi_3 = \pi_4 = 0$.

Output PcGive : DMC

```
EQ(3) Modelling LWAGE by IVE-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
EDUC           Y    0.0613966    0.03144     1.95   0.051
Constant            0.0481003     0.4003    0.120   0.904
EXPER               0.0441704    0.01343     3.29   0.001
EXPERSQ          -0.000898970  0.0004017    -2.24   0.026

Additional instruments:
[0] = FATHEDUC
[1] = MOTHEDUC

Specification test: Chi^2(1) =  0.37807 [0.5386]
Testing beta = 0:   Chi^2(3) =   24.422 [0.0000]**
```

$\longrightarrow$ La régression dans la forme structurelle par DMC donne un rendement de l'éducation des femmes (coefficient $\hat{\beta}_1$) de 0.061 et à peine significatif (gonflement de l'écart-type 0.031).

Extensions liées aux VI

Variables endogènes multiples

- Cas de plusieurs variables endogènes. Exemple : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + u_1$, où z_1 et z_2 sont des variables exogènes et y_2 et y_3 des variables endogènes.

Variables endogènes multiples

- Cas de plusieurs variables endogènes. Exemple : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + u_1$, où z_1 et z_2 sont des variables exogènes et y_2 et y_3 des variables endogènes.
- Condition nécessaire d'identification = Condition d'ordre ou de rang. $\longrightarrow$ On a besoin d'au moins autant de variables exogènes exclues (de l'équation structurelle) que de variables endogènes présentes (dans l'équation structurelle).

Variables endogènes multiples

- Cas de plusieurs variables endogènes. Exemple : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + u_1$, où z_1 et z_2 sont des variables exogènes et y_2 et y_3 des variables endogènes.
- Condition nécessaire d'identification = Condition d'ordre ou de rang. $\longrightarrow$ On a besoin d'au moins autant de variables exogènes exclues (de l'équation structurelle) que de variables endogènes présentes (dans l'équation structurelle).
- Dans l'exemple, l'identification requiert 2 variables exogènes z_3 et z_4 qui doivent être significatives dans l'équation en forme réduite.

Erreurs de mesure

- Vrai modèle de régression : $y_1 = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u$, où x_1^* est observé avec erreur.
  $\rightarrow$ on n'observe que x_1 : $x_1^* = x_1 + e_1$.

Erreurs de mesure

- Vrai modèle de régression : $y_1 = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u$, où x_1^* est observé avec erreur.
  $\rightarrow$ on n'observe que x_1 : $x_1^* = x_1 + e_1$.
- $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u - \beta_1 e_1)$.
  $\rightarrow$ Ceci crée un problème d'endogénéité : corrélation entre x_1 et le terme d'erreur $(u - \beta_1 e_1) \rightarrow$ MCO biaisés.

Erreurs de mesure

- Vrai modèle de régression : $y_1 = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u$, où x_1^* est observé avec erreur.
  $\longrightarrow$ on n'observe que x_1 : $x_1^* = x_1 + e_1$.
- $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u - \beta_1 e_1)$.
  $\longrightarrow$ Ceci crée un problème d'endogénéité : corrélation entre x_1 et le terme d'erreur $(u - \beta_1 e_1) \rightarrow$ MCO biaisés.
- Solution : trouver une VI pour x_1. Dans le cas des problèmes de mesure, idée est de trouver une variable z_1 corrélée avec x_1^* mais non corrélée avec e_1.

Erreurs de mesure

Exemples :

- Salaire annuel reporté par travailleurs (erreurs de mesure) : VI = salaire reporté par l'employeur.

Erreurs de mesure

Exemples :

- Salaire annuel reporté par travailleurs (erreurs de mesure) : VI = salaire reporté par l'employeur.
- Revenu annuel d'un ménage : VI = niveau annuel de l'épargne du ménage.

Erreurs de mesure

Exemples :

- Salaire annuel reporté par travailleurs (erreurs de mesure) : VI = salaire reporté par l'employeur.
- Revenu annuel d'un ménage : VI = niveau annuel de l'épargne du ménage.
- Niveau d'éducation reporté par les travailleurs : VI = nombre d'années d'éducation reporté par le frère jumeau ou la soeur jumelle.

Test d'endogénéité

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \longrightarrow$ Supposons que z_1 et z_2 sont des variables exogènes et z_3 et z_4 des variables exclues.

Test d’endogénéité

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \longrightarrow$ Supposons que z_1 et z_2 sont des variables exogènes et z_3 et z_4 des variables exclues.
- Comment tester si y_2 est endogène ?

Test d'endogénéité

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \longrightarrow$ Supposons que z_1 et z_2 sont des variables exogènes et z_3 et z_4 des variables exclues.
- Comment tester si y_2 est endogène ?
- Test d'Hausman (1978) basé sur la comparaison entre valeurs estimées par MCO et par DMC.

Test d'endogénéité

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \longrightarrow$ Supposons que z_1 et z_2 sont des variables exogènes et z_3 et z_4 des variables exclues.
- Comment tester si y_2 est endogène ?
- Test d'Hausman (1978) basé sur la comparaison entre valeurs estimées par MCO et par DMC.
- La statistique du test d'Hausman peut être calculée comme suit :

$$\zeta_H = \left(\hat{\beta}_{VI} - \hat{\beta}_{MCO}\right)' \left[Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MCO})\right]^{-1} \left(\hat{\beta}_{VI} - \hat{\beta}_{MCO}\right).$$

Test d'endogénéité

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \longrightarrow$ Supposons que z_1 et z_2 sont des variables exogènes et z_3 et z_4 des variables exclues.
- Comment tester si y_2 est endogène ?
- Test d'Hausman (1978) basé sur la comparaison entre valeurs estimées par MCO et par DMC.
- La statistique du test d'Hausman peut être calculée comme suit :

$$\zeta_H = \left(\hat{\beta}_{VI} - \hat{\beta}_{MCO}\right)' \left[Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MCO})\right]^{-1} \left(\hat{\beta}_{VI} - \hat{\beta}_{MCO}\right).$$

- Sous $H_0 : E(y_2, u_1) = 0, \zeta_H \sim \chi^2(K)$, où K est le nombre de coefficients estimés (c-à-d 4 dans l'exemple).

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.
- Idée : reformuler l'hypothèse $H_0 : E(y_2, u_1) = 0$ en terme d'un test de restriction de paramètres.

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.
- Idée : reformuler l'hypothèse $H_0 : E(y_2, u_1) = 0$ en terme d'un test de restriction de paramètres.
- **Étape 1**. Estimation du modèle en forme réduite : $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + \nu_2$.

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.
- Idée : reformuler l'hypothèse $H_0 : E(y_2, u_1) = 0$ en terme d'un test de restriction de paramètres.
- **Étape 1**. Estimation du modèle en forme réduite : $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + \nu_2$.
- Comme les variables z sont exogènes, $E(z_i u_1) = 0$, $\forall i = 1, \ldots, 4$.

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.
- Idée : reformuler l'hypothèse $H_0 : E(y_2, u_1) = 0$ en terme d'un test de restriction de paramètres.
- **Étape 1**. Estimation du modèle en forme réduite : $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + \nu_2$.
- Comme les variables z sont exogènes, $E(z_i u_1) = 0$, $\forall i = 1, \ldots, 4$.
- Par conséquent, si y_2 est également exogène (H_0), ν_2 est non corrélé avec u_1.

Test LM d'endogénéité d'Hausman

- Test de Durbin, Wu et Hausman en 3 étapes, connu sous le nom de test LM d'Hausman.
- Idée : reformuler l'hypothèse $H_0 : E(y_2, u_1) = 0$ en terme d'un test de restriction de paramètres.
- **Étape 1**. Estimation du modèle en forme réduite : $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + \nu_2$.
- Comme les variables z sont exogènes, $E(z_i u_1) = 0$, $\forall i = 1, \ldots, 4$.
- Par conséquent, si y_2 est également exogène (H_0), ν_2 est non corrélé avec u_1.
- Pour tester l'exogénéité de y_2, écrivons $u_1 = \delta_1 \nu_2 + e_1$, où e_1 est non corrélé à ν_2 et $E(e_1) = 0$. Si $\delta_1 = 0, E(\nu_2 u_1) = 0$. Pour tester H_0 il suffit donc d'ajouter ν_2 au modèle structurel et tester $H_0 : \delta_1 = 0$.

Test LM d'endogénéité d'Hausman

- **Étape 2**. On récupère les résidus $\hat{\nu}_2$, contrepartie observable de ν_2 dans l'étape 1.

Test LM d'endogénéité d'Hausman

- **Étape 2**. On récupère les résidus $\hat{\nu}_2$, contrepartie observable de ν_2 dans l'étape 1.
- **Étape 3**. On estime par MCO $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \delta_1 \hat{\nu}_2 + erreur$ et on teste par un $t - test\ H_0 : \delta_1 = 0. \rightarrow$ On conclut à l'endogénéité de y_2 si on rejette H_0.

Test LM d'endogénéité d'Hausman

- Pourquoi test LM ?

Test LM d'endogénéité d'Hausman

- Pourquoi test LM ?
- Extension: 2 variables explicatives endogènes y_2 et y_3.

Test LM d'endogénéité d'Hausman

- Pourquoi test LM ?
- Extension: 2 variables explicatives endogènes y_2 et y_3.
- **Étape 1**. Estimation des 2 modèles en forme réduite :
  $y_2 = \pi_{0,2} + \pi_{1,2} z_1 + \pi_{2,2} z_2 + \pi_{3,2} z_3 + \pi_{4,2} z_4 + \nu_2$ et
  $y_3 = \pi_{0,3} + \pi_{1,3} z_1 + \pi_{2,3} z_2 + \pi_{3,3} z_3 + \pi_{4,3} z_4 + \nu_3$.

Test LM d'endogénéité d'Hausman

- Pourquoi test LM ?
- Extension: 2 variables explicatives endogènes y_2 et y_3.
- **Étape 1**. Estimation des 2 modèles en forme réduite :
  $y_2 = \pi_{0,2} + \pi_{1,2}z_1 + \pi_{2,2}z_2 + \pi_{3,2}z_3 + \pi_{4,2}z_4 + \nu_2$ et
  $y_3 = \pi_{0,3} + \pi_{1,3}z_1 + \pi_{2,3}z_2 + \pi_{3,3}z_3 + \pi_{4,3}z_4 + \nu_3$.
- Si il y a au moins une variable endogène y_2 et/ou y_3, ν_2 et/ou ν_3 sont/est corrélé(s) avec u_1 et y_1.
  $\rightarrow$ **Étape 2**. On récupère les résidus $\hat{\nu}_2$ et $\hat{\nu}_3$, contrepartie observable de ν_2 et ν_3.

Test LM d'endogénéité d'Hausman

- Pourquoi test LM ?
- Extension: 2 variables explicatives endogènes y_2 et y_3.
- **Étape 1**. Estimation des 2 modèles en forme réduite :
  $y_2 = \pi_{0,2} + \pi_{1,2} z_1 + \pi_{2,2} z_2 + \pi_{3,2} z_3 + \pi_{4,2} z_4 + \nu_2$ et
  $y_3 = \pi_{0,3} + \pi_{1,3} z_1 + \pi_{2,3} z_2 + \pi_{3,3} z_3 + \pi_{4,3} z_4 + \nu_3$.
- Si il y a au moins une variable endogène y_2 et/ou y_3, ν_2 et/ou ν_3 sont/est corrélé(s) avec u_1 et y_1.
  $\rightarrow$ **Étape 2**. On récupère les résidus $\hat{\nu}_2$ et $\hat{\nu}_3$, contrepartie observable de ν_2 et ν_3.
- **Étape 3**. On estime par MCO
  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + \delta_1 \hat{\nu}_2 + \delta_2 \hat{\nu}_3 + erreur$
  et on teste $H_0 : \delta_1 = \delta_2 = 0$. $\rightarrow$ On conclut à l'endogénéité de y_2 ou y_3 si on rejette H_0.

Exemple de test d'endogénéité

On récupère les résidus $\hat{\nu}_2$ (res) après l'estimation par MCO de :
$educ = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 motheduc + \pi_4 fatheduc + \nu_2$.

```
EQ(4) Modelling LWAGE by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
EDUC                0.0639033    0.02921     2.19   0.029
Constant           -0.0114040     0.3592  -0.0317   0.975
EXPER               0.0463071    0.01344     3.45   0.001
EXPERSQ          -0.000944415  0.0004001    -2.36   0.019
res                 0.0558771    0.03279     1.70   0.089

sigma                0.664929  RSS              187.021076
R^2                   0.16257  F(4,423) = 20.53 [0.000]**
log-likelihood       -430.135  DW                     1.93
no. of observations       428  no. of parameters         5
```

$\rightarrow t - stat$ de $\hat{\delta}_1 = 1.7$ $[0.089]$ $\rightarrow$ Evidence limitée d'endogénéité de $educ$.
$\rightarrow$ Dans le doute il est préférable d'utiliser conjointement MCO et DMC.

Test de suridentification

- De manière générale, on ne peut pas tester la Condition 1 des VI, c-à-d $Cov(z, u) = 0$.

Test de suridentification

- De manière générale, on ne peut pas tester la Condition 1 des VI, c-à-d $Cov(z, u) = 0$.
- Néanmoins il y a une exception importante : lorsque l'on dispose d'au moins 1 restriction de suridentifcation.

Test de suridentification

- De manière générale, on ne peut pas tester la Condition 1 des VI, c-à-d $Cov(z, u) = 0$.
- Néanmoins il y a une exception importante : lorsque l'on dispose d'au moins 1 restriction de suridentifcation.
- Le nombre de restrictions de suridentification $=$ nombre d'instruments excédentaires par rapport au nombre de variables endogènes.

Test de suridentification

- De manière générale, on ne peut pas tester la Condition 1 des VI, c-à-d $Cov(z, u) = 0$.
- Néanmoins il y a une exception importante : lorsque l'on dispose d'au moins 1 restriction de suridentifcation.
- Le nombre de restrictions de suridentification $=$ nombre d'instruments excédentaires par rapport au nombre de variables endogènes.
- Exemple : $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1$. Si on dispose de 2 VI z_3 et z_4, on dispose d'1 condition de suridentification.

Test de suridentification

- Pourquoi est-ce possible ? Idée : on peut estimer les β_j avec **une** VI (disons z_3), récupérer les résidus et tester la corrélation par rapport à l'autre VI (c-à-d entre u_1 et z_4). $\rightarrow$ Pas possible si on ne dispose que d'une seule VI.

Test de suridentification

- Pourquoi est-ce possible ? Idée : on peut estimer les β_j avec **une** VI (disons z_3), récupérer les résidus et tester la corrélation par rapport à l'autre VI (c-à-d entre u_1 et z_4). $\rightarrow$ Pas possible si on ne dispose que d'une seule VI.
- 3 étapes.

Test de suridentification

- Pourquoi est-ce possible ? Idée : on peut estimer les β_j avec **une** VI (disons z_3), récupérer les résidus et tester la corrélation par rapport à l'autre VI (c-à-d entre u_1 et z_4). $\rightarrow$ Pas possible si on ne dispose que d'une seule VI.
- 3 étapes.
- **Étape 1**. Estimer l'équation structurelle par DMCO ($\rightarrow$ la VI $=$ combinaison linéaire de z_3 et z_4) et récupérer les résidus $\hat{u}_1$.

Test de suridentification

- Pourquoi est-ce possible ? Idée : on peut estimer les β_j avec **une** VI (disons z_3), récupérer les résidus et tester la corrélation par rapport à l'autre VI (c-à-d entre u_1 et z_4). $\rightarrow$ Pas possible si on ne dispose que d'une seule VI.
- 3 étapes.
- **Étape 1**. Estimer l'équation structurelle par DMCO ($\rightarrow$ la VI $=$ combinaison linéaire de z_3 et z_4) et récupérer les résidus $\hat{u}_1$.
- **Étape 2**. Régresser $\hat{u}_1$ sur toutes les variables exogènes $z_1, \ldots, z_4$ et récupérer le R-carré R_1^2.

Test de suridentification

- Tester H_0 : toutes les variables $z_1, \dots, z_4$ sont exogènes et donc non corrélées avec u_1.

Test de suridentification

- Tester H_0 : toutes les variables $z_1, \ldots, z_4$ sont exogènes et donc non corrélées avec u_1.
- **Étape 3**. Test LM. Sous H_0, $nR_1^2 \sim \chi_q^2$ (asymptotiquement), où q est le nombre de conditions de suridentification, c-à-d le nombre de VI hors du modèle $-$ le nombre de variables explicatives endogènes.

Test de suridentification

- Tester H_0 : toutes les variables $z_1, \ldots, z_4$ sont exogènes et donc non corrélées avec u_1.
- **Étape 3**. Test LM. Sous H_0, $nR_1^2 \sim \chi_q^2$ (asymptotiquement), où q est le nombre de conditions de suridentification, c-à-d le nombre de VI hors du modèle $-$ le nombre de variables explicatives endogènes.
- $\rightarrow$ Si nR_1^2 excède la valeur critique (à 5% par exemple), on rejette la validité de la Condition 2 et on conclut qu'au moins une des 2 VI n'est pas exogène.

Test de suridentification

- Tester H_0 : toutes les variables $z_1, \ldots, z_4$ sont exogènes et donc non corrélées avec u_1.
- **Étape 3**. Test LM. Sous H_0, $nR_1^2 \sim \chi_q^2$ (asymptotiquement), où q est le nombre de conditions de suridentification, c-à-d le nombre de VI hors du modèle $-$ le nombre de variables explicatives endogènes.
- $\rightarrow$ Si nR_1^2 excède la valeur critique (à 5% par exemple), on rejette la validité de la Condition 2 et on conclut qu'au moins une des 2 VI n'est pas exogène.
- Ce test est connu sous le nom de Test de Sargan de validité des instruments.

Exemple de test d’endogénéité

On récupère les résidus $\hat{u}_1$ (res) après l’estimation du modèle $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_1$ par DMC avec comme instruments $fatheduc$ et $motheduc$.

```
EQ(5) Modelling res by OLS-CS (using mroz)
       The estimation sample is: 1 to 428

                      Coefficient  Std.Error  t-value  t-prob
Constant                0.0109641     0.1413   0.0776   0.938
EXPER                -1.83348e-005   0.01333 -0.00138   0.999
EXPERSQ               7.34138e-007 0.0003985  0.00184   0.999
FATHEDUC               0.00578226    0.01118    0.517   0.605
MOTHEDUC              -0.00660653    0.01189   -0.556   0.579

R^2                   0.000883344  F(4,423) =  0.0935 [0.984]
no. of observations           428  no. of parameters         5
```

$\rightarrow$ Estimer par MCO : $\hat{u}_1 = \gamma_0 + \gamma_1 exper + \gamma_2 exper^2 + \gamma_3 fatheduc + \gamma_4 motheduc + v$.

$\rightarrow nR_1^2 = 428 \times 0.00088 = 0.3781[0.5386]$.

$\rightarrow$ Ces 2 instruments passent le test de suridentification.

Output PcGive : Specification test

```
EQ(3) Modelling LWAGE by IVE-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
EDUC            Y    0.0613966    0.03144     1.95   0.051
Constant             0.0481003     0.4003    0.120   0.904
EXPER                0.0441704    0.01343     3.29   0.001
EXPERSQ           -0.000898970  0.0004017    -2.24   0.026

Additional instruments:
[0] = FATHEDUC
[1] = MOTHEDUC

Specification test: Chi^2(1) =  0.37807 [0.5386]
Testing beta = 0:   Chi^2(3) =   24.422 [0.0000]**
```

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\longrightarrow$ Tester cette hypothèse.

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\longrightarrow$ Tester cette hypothèse.
- $\longrightarrow$ En cas de rejet de cette hypothèse, il faut utiliser des écarts-type robustes ou adapter la méthode des moindres carrés pondérés.

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\longrightarrow$ Tester cette hypothèse.
- $\longrightarrow$ En cas de rejet de cette hypothèse, il faut utiliser des écarts-type robustes ou adapter la méthode des moindres carrés pondérés.
- Comment tester cette hypothèse ?

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\rightarrow$ Tester cette hypothèse.
- $\rightarrow$ En cas de rejet de cette hypothèse, il faut utiliser des écarts-type robustes ou adapter la méthode des moindres carrés pondérés.
- Comment tester cette hypothèse ?
- Estimer le modèle par DMCO et sauver les résidus $\hat{u}$.

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\longrightarrow$ Tester cette hypothèse.
- $\longrightarrow$ En cas de rejet de cette hypothèse, il faut utiliser des écarts-type robustes ou adapter la méthode des moindres carrés pondérés.
- Comment tester cette hypothèse ?
- Estimer le modèle par DMCO et sauver les résidus $\hat{u}$.
- Estimer par MCO le modèle $\hat{u}^2 = \gamma_0 + \gamma_1 z_1 + \ldots + \gamma_m z_m + e$.

Hétéroscédasticité

- Tout comme pour les MCO, la méthode des DMCO pose l'hypothèse d'homoscédasticité pour que l'inférence classique soit valide.
- $\rightarrow$ Tester cette hypothèse.
- $\rightarrow$ En cas de rejet de cette hypothèse, il faut utiliser des écarts-type robustes ou adapter la méthode des moindres carrés pondérés.
- Comment tester cette hypothèse ?
- Estimer le modèle par DMCO et sauver les résidus $\hat{u}$.
- Estimer par MCO le modèle $\hat{u}^2 = \gamma_0 + \gamma_1 z_1 + \ldots + \gamma_m z_m + e$.
- Tester $H_0 : \gamma_1 = \ldots = \gamma_m = 0$ (c-à-d homoscédasticité) via un F-test.

Output PcGive : DMCO

```
EQ(1) Modelling LWAGE by IVE-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                        Coefficient  Std.Error  t-value  t-prob
EDUC               Y      0.0803918    0.02177     3.69   0.000
Constant                  -0.186857     0.2854   -0.655   0.513
EXPER                     0.0430973    0.01326     3.25   0.001
EXPERSQ                -0.000862797  0.0003962    -2.18   0.030

Additional instruments:
[0] = FATHEDUC
[1] = MOTHEDUC
[2] = HUSEDUC

Specification test: Chi^2(2) =   1.1150 [0.5726]
Testing beta = 0:   Chi^2(3) =   34.574 [0.0000]**
res [1 to 428] saved to mroz.in7
```

Output PcGive : Test d'homoscédasticité

```
EQ(2) Modelling res2 by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant             0.729764     0.2774     2.63   0.009
FATHEDUC           0.00879304    0.01781    0.494   0.622
EXPER              -0.0552427    0.02070    -2.67   0.008
EXPERSQ            0.00121010  0.0006190     1.95   0.051
MOTHEDUC            0.0119529    0.01857    0.644   0.520
HUSEDUC           -0.00337671    0.01788   -0.189   0.850

sigma                 1.04834  RSS              463.784598
R^2                 0.0290988  F(5,422) =   2.53 [0.028]*
log-likelihood       -624.489  DW                     1.58
no. of observations       428  no. of parameters         6
mean(res2)           0.443773  var(res2)           1.11609
```

$\longrightarrow$ Rejet de H_0 au seuil de 5%.

Modèles à équations simultanées : Chapitre 16

Simultanéité

Dans le chapitre précédent, nous avons vu 2 raisons principales d'utiliser la méthode des VI.

- 1) Problème d'erreurs de mesure.

Simultanéité

Dans le chapitre précédent, nous avons vu 2 raisons principales d'utiliser la méthode des VI.

- 1) Problème d'erreurs de mesure.
- 2) Problème d'omission de variables explicatives appropriées.

Simultanéité

Dans le chapitre précédent, nous avons vu 2 raisons principales d'utiliser la méthode des VI.

- 1) Problème d'erreurs de mesure.
- 2) Problème d'omission de variables explicatives appropriées.
- $\rightarrow$ Dans chaque cas, le problème donne lieu à un problème d'endogénéité.

Simultanéité

Dans le chapitre précédent, nous avons vu 2 raisons principales d'utiliser la méthode des VI.

- 1) Problème d'erreurs de mesure.
- 2) Problème d'omission de variables explicatives appropriées.
- $\rightarrow$ Dans chaque cas, le problème donne lieu à un problème d'endogénéité.
- Dans ce chapitre, on va étudier en détails une autre raison importante : le problème de simultanéité.

Simultanéité

- Une ou plusieurs variables explicatives sont déterminées conjointement avec la variable dépendante, par exemple à travers une relation d'équilibre.

Simultanéité

- Une ou plusieurs variables explicatives sont déterminées conjointement avec la variable dépendante, par exemple à travers une relation d'équilibre.
- $\longrightarrow$ Modèles à équations simultanées (MES).

Simultanéité

- Une ou plusieurs variables explicatives sont déterminées conjointement avec la variable dépendante, par exemple à travers une relation d'équilibre.
- $\longrightarrow$ Modèles à équations simultanées (MES).
- $\longrightarrow$ La méthode d'estimation des VI et des DMC s'applique pour estimer les MES alors que l'estimation par MCO génère des estimateurs biaisés.

Simultanéité

- Une ou plusieurs variables explicatives sont déterminées conjointement avec la variable dépendante, par exemple à travers une relation d'équilibre.
- $\longrightarrow$ Modèles à équations simultanées (MES).
- $\longrightarrow$ La méthode d'estimation des VI et des DMC s'applique pour estimer les MES alors que l'estimation par MCO génère des estimateurs biaisés.
- $\longrightarrow$ Biais de simultanéité.

Caractéristiques des MES

- Exemple de MES : Offre et demande de travail.

Caractéristiques des MES

- Exemple de MES : Offre et demande de travail.
- Equation d'offre de travail : $h_s = \alpha_1 w + \beta_1 z_1 + u_1$.

Caractéristiques des MES

- Exemple de MES : Offre et demande de travail.
- Equation d'offre de travail : $h_s = \alpha_1 w + \beta_1 z_1 + u_1$.
- Equation structurelle avec $h_s =$ offre de travail, $w =$ salaire et $z_1 =$ facteurs exogènes.

Caractéristiques des MES

- Exemple de MES : Offre et demande de travail.
- Equation d'offre de travail : $h_s = \alpha_1 w + \beta_1 z_1 + u_1$.
- Equation structurelle avec $h_s =$ offre de travail, $w =$ salaire et $z_1 =$ facteurs exogènes.
- Problème : la condition *Ceteris Paribus* n'est pas respectée.
  $\rightarrow$ Les variation de w ne sont pas exogènes car w_i est le salaire de l'individu i observé à l'équilibre.
  $\rightarrow$ $h_{id} = h_{is}$, c-à-d demande de travail $=$ offre de travail.

Caractéristiques des MES

- Equation de demande de travail : $h_d = \alpha_2 w + \beta_2 z_2 + u_2$.

Caractéristiques des MES

- Equation de demande de travail : $h_d = \alpha_2 w + \beta_2 z_2 + u_2$.
- Equation structurelle avec $h_d =$ demande de travail, $w =$ salaire et z_2 facteurs exogènes (de la demande).

Caractéristiques des MES

- Equation de demande de travail : $h_d = \alpha_2 w + \beta_2 z_2 + u_2$.
- Equation structurelle avec $h_d =$ demande de travail, $w =$ salaire et z_2 facteurs exogènes (de la demande).
- A l'équilibre, pour chaque individu i, on a : $h_i = h_{id} = h_{is}$, où h_i est le nombre d'heures travaillées (et donc observé).

Caractéristiques des MES

- Equation de demande de travail : $h_d = \alpha_2 w + \beta_2 z_2 + u_2$.
- Equation structurelle avec $h_d =$ demande de travail, $w =$ salaire et z_2 facteurs exogènes (de la demande).
- A l'équilibre, pour chaque individu i, on a : $h_i = h_{id} = h_{is}$, où h_i est le nombre d'heures travaillées (et donc observé).
- On obtient alors un MES :

$$\begin{aligned} h_i = h_{is} &= \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \\ h_i = h_{id} &= \alpha_2 w_i + \beta_2 z_{i2} + u_{i2}. \end{aligned}$$

Caractéristiques des MES

- Equation de demande de travail : $h_d = \alpha_2 w + \beta_2 z_2 + u_2$.
- Equation structurelle avec $h_d =$ demande de travail, $w =$ salaire et z_2 facteurs exogènes (de la demande).
- A l'équilibre, pour chaque individu i, on a : $h_i = h_{id} = h_{is}$, où h_i est le nombre d'heures travaillées (et donc observé).
- On obtient alors un MES :

$$\begin{aligned} h_i = h_{is} &= \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \\ h_i = h_{id} &= \alpha_2 w_i + \beta_2 z_{i2} + u_{i2}. \end{aligned}$$

- $\rightarrow$ MES à 2 équations dans leur forme structurelle.
$\rightarrow$ h_i et w_i sont des variables endogènes.
$\rightarrow$ z_{i1} et z_{i2} sont des variables exogènes telles que $E(u_1 z_1) = E(u_2 z_2) = 0$.

Autres exemples de MES

Relation taux de criminalité (meurtres) et taille des forces de polices.

- Equation explicative des meurtres par ville :
  $murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1$,

Autres exemples de MES

Relation taux de criminalité (meurtres) et taille des forces de polices.

- Equation explicative des meurtres par ville : $murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1$,
- où $murdpc =$ taux de criminalité par habitant, $polpc =$ nombre de policiers par habitant et $incpc =$ revenu par habitant.

Autres exemples de MES

Relation taux de criminalité (meurtres) et taille des forces de polices.

- Equation explicative des meurtres par ville : $murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1$,
- où $murdpc =$ taux de criminalité par habitant, $polpc =$ nombre de policiers par habitant et $incpc =$ revenu par habitant.
- Equation explicative du nombre de policiers : $polpc = \alpha_2 murdpc + \beta_{20} +$ d'autres facteurs $+ u_2$.

Autres exemples de MES

Relation taux de criminalité (meurtres) et taille des forces de polices.

- Equation explicative des meurtres par ville : $murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1$,
- où $murdpc =$ taux de criminalité par habitant, $polpc =$ nombre de policiers par habitant et $incpc =$ revenu par habitant.
- Equation explicative du nombre de policiers : $polpc = \alpha_2 murdpc + \beta_{20} +$ d'autres facteurs $+ u_2$.
- $\rightarrow polpc$ et $murdpc$ sont 2 variables endogènes d'un MES à 2 équations.

Biais de simultanéité

L'utilisation des MCO pour estimer les équations structurelles d'un MES donne lieu à un biais.

- Soit un MES à 2 équations :

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\ y_2 &= \alpha_2 y_1 + \beta_2 z_2 + u_2. \end{aligned}$$

Biais de simultanéité

L'utilisation des MCO pour estimer les équations structurelles d'un MES donne lieu à un biais.

- Soit un MES à 2 équations :

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\ y_2 &= \alpha_2 y_1 + \beta_2 z_2 + u_2. \end{aligned}$$

- Trouvons la forme réduite de y_2 :
  $y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$.

Biais de simultanéité

L'utilisation des MCO pour estimer les équations structurelles d'un MES donne lieu à un biais.

- Soit un MES à 2 équations :

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\ y_2 &= \alpha_2 y_1 + \beta_2 z_2 + u_2. \end{aligned}$$

- Trouvons la forme réduite de y_2 :
  $y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$.
- Ceci se réécrit comme ceci :

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + \nu_2,$$

où $\pi_{21} = \frac{\alpha_2 \beta_1}{1-\alpha_1\alpha_2}$, $\pi_{22} = \frac{\beta_2}{1-\alpha_1\alpha_2}$ et $\nu_2 = \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2}$.

Biais de simultanéité

- On peut estimer par MCO de manière consistante cette forme réduite car z_1 et z_2 sont exogènes et donc non corrélés avec u_1 et u_2 et donc ν_2.

Biais de simultanéité

- On peut estimer par MCO de manière consistante cette forme réduite car z_1 et z_2 sont exogènes et donc non corrélés avec u_1 et u_2 et donc ν_2.
- Par contre l'estimation de la forme structurelle par MCO donne lieu à des estimateur biaisé.

Biais de simultanéité

- On peut estimer par MCO de manière consistante cette forme réduite car z_1 et z_2 sont exogènes et donc non corrélés avec u_1 et u_2 et donc ν_2.
- Par contre l'estimation de la forme structurelle par MCO donne lieu à des estimateur biaisé.
- Exemple : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow y_2$ est corrélé avec u_1 car $y_2 = \pi_{21} z_1 + \pi_{22} z_2 + \nu_2$ et $u_1 \in \nu_2$.

Biais de simultanéité

- On peut estimer par MCO de manière consistante cette forme réduite car z_1 et z_2 sont exogènes et donc non corrélés avec u_1 et u_2 et donc ν_2.
- Par contre l'estimation de la forme structurelle par MCO donne lieu à des estimateur biaisé.
- Exemple : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow y_2$ est corrélé avec u_1 car $y_2 = \pi_{21} z_1 + \pi_{22} z_2 + \nu_2$ et $u_1 \in \nu_2$.
- Plus formellement : $Cov(y_2, u_1) = Cov(\nu_2, u_1) = \left(\frac{\alpha_2}{1-\alpha_1\alpha_2}\right) E(u_1^2) = \left(\frac{\alpha_2}{1-\alpha_1\alpha_2}\right) \sigma_1^2$.

Biais de simultanéité

- On peut estimer par MCO de manière consistante cette forme réduite car z_1 et z_2 sont exogènes et donc non corrélés avec u_1 et u_2 et donc ν_2.
- Par contre l'estimation de la forme structurelle par MCO donne lieu à des estimateur biaisé.
- Exemple : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow y_2$ est corrélé avec u_1 car $y_2 = \pi_{21} z_1 + \pi_{22} z_2 + \nu_2$ et $u_1 \in \nu_2$.
- Plus formellement : $Cov(y_2, u_1) = Cov(\nu_2, u_1) = \left(\frac{\alpha_2}{1-\alpha_1\alpha_2}\right) E(u_1^2) = \left(\frac{\alpha_2}{1-\alpha_1\alpha_2}\right) \sigma_1^2$.
- Si $\alpha_2 \neq 0$ (cas d'un MES), alors MCO biaisés car rejet de **MLR.3**.

Biais de simultanéité

- Rappels : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow plim\ \hat{\alpha}_1 = \alpha_1 + \frac{Cov(y_2,u_1)}{Var(y_2)}$.

Biais de simultanéité

- Rappels : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow plim\ \hat{\alpha}_1 = \alpha_1 + \frac{Cov(y_2, u_1)}{Var(y_2)}$.
- Le biais (asymptotique) de simultanéité aura le même signe que la covariance entre ν_2 et u_1.

Biais de simultanéité

- Rappels : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1.$
  $\rightarrow plim\ \hat{\alpha}_1 = \alpha_1 + \frac{Cov(y_2,u_1)}{Var(y_2)}$.
- Le biais (asymptotique) de simultanéité aura le même signe que la covariance entre ν_2 et u_1.
- Le biais de simultanéité dépend des paramètres α_2 et α_1.

Biais de simultanéité

- Rappels : $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$.
  $\rightarrow plim\ \hat{\alpha}_1 = \alpha_1 + \frac{Cov(y_2,u_1)}{Var(y_2)}$.
- Le biais (asymptotique) de simultanéité aura le même signe que la covariance entre ν_2 et u_1.
- Le biais de simultanéité dépend des paramètres α_2 et α_1.
- Exemple : si $\alpha_2 > 0$ et $\alpha_1 \alpha_2 < 1$, alors le biais asymptotique sera positif.
  $\rightarrow$ L'effet de y_2 sera surestimé.

Identification d'un MES à 2 équations

Considérons le cas d'un MES à 2 équations.

- $$\begin{aligned} y_1 &= \beta_{10} + \alpha_1 y_2 + z_1\beta_1 + u_1 \\ y_2 &= \beta_{20} + \alpha_2 y_1 + z_2\beta_2 + u_2. \end{aligned}$$

Identification d'un MES à 2 équations

Considérons le cas d'un MES à 2 équations.

- $$\begin{aligned} y_1 &= \beta_{10} + \alpha_1 y_2 + z_1\beta_1 + u_1 \\ y_2 &= \beta_{20} + \alpha_2 y_1 + z_2\beta_2 + u_2. \end{aligned}$$
- Les variables z_1 et z_2 représentent un ensemble de respectivement k_1 et k_2 variables exogènes (avec une possible intersection entre les 2) :

$$\begin{aligned} z_1\beta_1 &= \beta_{11}z_{11} + \beta_{12}z_{12} + \ldots + \beta_{1k_1}z_{1k_1} \\ z_2\beta_2 &= \beta_{21}z_{21} + \beta_{22}z_{22} + \ldots + \beta_{2k_2}z_{2k_2}. \end{aligned}$$

Identification d'un MES à 2 équations

- La présence ou non de ces variables z_i va permettre de définir des restrictions d'exclusion qui permettront de définir la condition d'identification de chaque équation.

Identification d'un MES à 2 équations

- La présence ou non de ces variables z_i va permettre de définir des restrictions d'exclusion qui permettront de définir la condition d'identification de chaque équation.
- Condition d'ordre pour l'identification d'une équation structurelle.
  $\longrightarrow$ La condition nécessaire pour identifier la première équation dans un MES à 2 équations est que la *seconde équation* contienne au moins une variable exogène exclue de la première équation.

Identification d'un MES à 2 équations

- La présence ou non de ces variables z_i va permettre de définir des restrictions d'exclusion qui permettront de définir la condition d'identification de chaque équation.
- Condition d'ordre pour l'identification d'une équation structurelle.
  $\longrightarrow$ La condition nécessaire pour identifier la première équation dans un MES à 2 équations est que la *seconde équation* contienne au moins une variable exogène exclue de la première équation.
- Condition de rang pour l'identification d'une équation structurelle.
  $\longrightarrow$ La première équation dans un MES à 2 équations est identifiée ssi la *seconde équation* contient au moins une variable exogène avec un coefficient β_{2j} non nul qui est exclue de la première équation.

Exemples d'identification

Modèle de demande et d'offre définit à l'équilibre $q = q_s = q_d$.

- q est la consommation de lait par habitant d'un pays et p est le prix moyen du lait par litre dans ce pays.

Exemples d'identification

Modèle de demande et d'offre définit à l'équilibre
$q = q_s = q_d$.

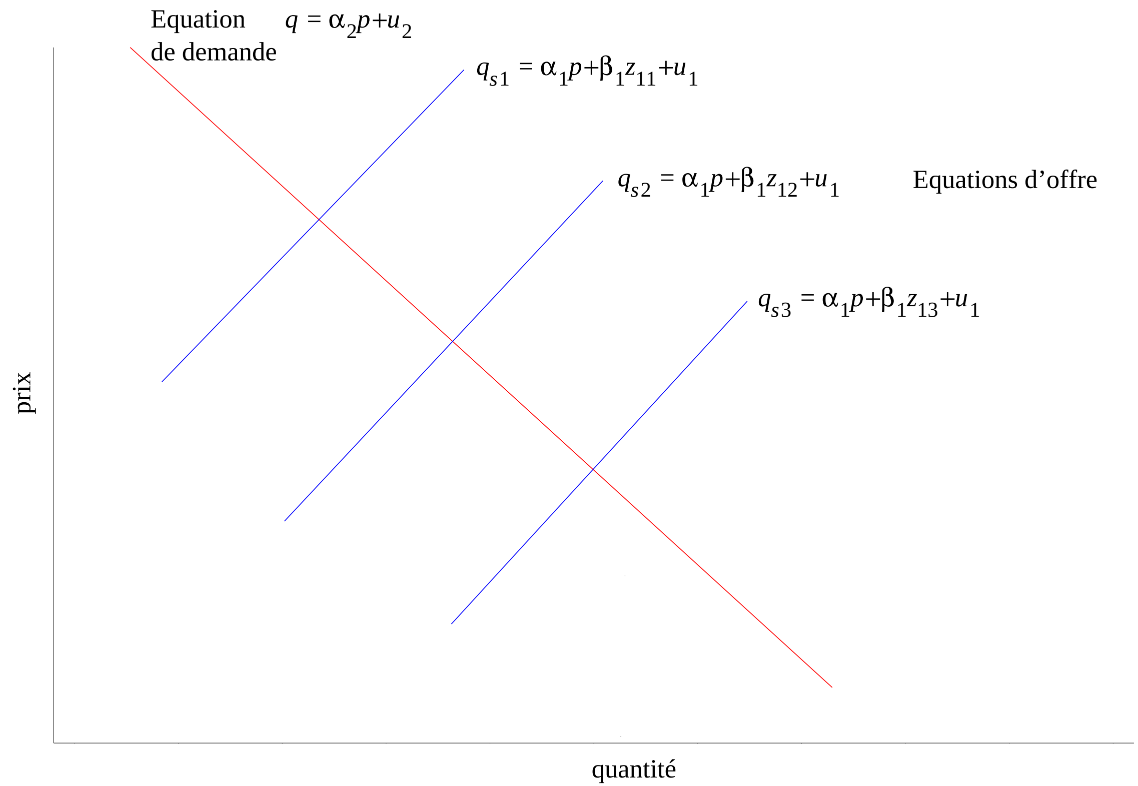
- q est la consommation de lait par habitant d'un pays et p est le prix moyen du lait par litre dans ce pays.
- Equation d'offre : $q_s = \alpha_1 p + \beta_1 z_1 + u_1$.
  $\rightarrow z_1$ est le prix de l'alimentation du bétail $=$ variable exogène.
  $\rightarrow \Delta z_1$ doit faire Δ l'offre de lait.

Exemples d'identification

Modèle de demande et d'offre définit à l'équilibre $q = q_s = q_d$.

- q est la consommation de lait par habitant d'un pays et p est le prix moyen du lait par litre dans ce pays.
- Equation d'offre : $q_s = \alpha_1 p + \beta_1 z_1 + u_1$.
  $\rightarrow z_1$ est le prix de l'alimentation du bétail $=$ variable exogène.
  $\rightarrow \Delta z_1$ doit faire Δ l'offre de lait.
- Equation de demande : $q_d = \alpha_2 p + u_2$.
  $\rightarrow$ Pas de variable exogène.
  $\rightarrow \Delta z_1$ ne doit pas faire Δ la demande de lait.

Déplacer l’équation d’offre



$\longrightarrow$ Dans ce MES, l’offre n’est pas identifiée (rejet de la condition d’ordre). Par contre, la condition d’ordre de la fonction de demande est remplie.

$\longrightarrow$ En effet, on peut utiliser z_1 comme VI pour q_d mais pas pour q_s.

$\longrightarrow$ Déplacer q_s permet d’identifier q_d.

Exemples d'identification

Modèle : offre de travail des femmes mariées.
$\rightarrow$ 2 équations structurelles.

- Equation d'offre de travail (heures travaillées) :

$$\begin{aligned} hours &= \alpha_1 log(wage) + \beta_{10} + \beta_{11} educ \\ &+ \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1, \end{aligned}$$

où $kidslt6 =$ nombre d'enfants < 6 ans et $nwifeinc =$ revenu non salarial de la femme mariée.

Exemples d'identification

Modèle : offre de travail des femmes mariées.
$\rightarrow$ 2 équations structurelles.

- Equation d'offre de travail (heures travaillées) :

$$\begin{aligned} hours &= \alpha_1 log(wage) + \beta_{10} + \beta_{11} educ \\ &+ \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1, \end{aligned}$$

  où $kidslt6 =$ nombre d'enfants < 6 ans et $nwifeinc =$ revenu non salarial de la femme mariée.

- Equation de salaire : $log(wage) = \alpha_2 hours + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + u_2$.

Exemples d'identification

Modèle : offre de travail des femmes mariées.
$\rightarrow$ 2 équations structurelles.

- Equation d'offre de travail (heures travaillées) :

$$\begin{aligned} hours &= \alpha_1 log(wage) + \beta_{10} + \beta_{11} educ \\ &+ \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1, \end{aligned}$$

  où $kidslt6 =$ nombre d'enfants < 6 ans et $nwifeinc =$ revenu non salarial de la femme mariée.

- Equation de salaire : $log(wage) = \alpha_2 hours + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + u_2$.
- La condition d'ordre de l'équation d'offre de travail est remplie $($$exper$ et $exper^2$ sont exclues$)$.

Exemples d'identification

- A partir des équations structurelles, la condition de rang pour l'identification de l'équation d'offre de travail est : $\beta_{22} \neq 0$ ou $\beta_{23} \neq 0$.

Exemples d'identification

- A partir des équations structurelles, la condition de rang pour l'identification de l'équation d'offre de travail est : $\beta_{22} \neq 0$ ou $\beta_{23} \neq 0$.
- On peut vérifier cette condition de rang via la forme réduite :

$$\begin{aligned} log(wage) &= \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt6 \\ &+ \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + \nu_2. \end{aligned}$$

Exemples d'identification

- A partir des équations structurelles, la condition de rang pour l'identification de l'équation d'offre de travail est : $\beta_{22} \neq 0$ ou $\beta_{23} \neq 0$.
- On peut vérifier cette condition de rang via la forme réduite :

$$\begin{aligned} log(wage) &= \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt6 \\ &+ \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + \nu_2. \end{aligned}$$

- $\rightarrow$ L'identification de l'équation de salaire requiert : $\pi_{25} \neq 0$ ou $\pi_{26} \neq 0$.

Exemples d'identification

- A partir des équations structurelles, la condition de rang pour l'identification de l'équation d'offre de travail est : $\beta_{22} \neq 0$ ou $\beta_{23} \neq 0$.
- On peut vérifier cette condition de rang via la forme réduite :

$$
\begin{aligned}
log(wage) &= \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt6 \\
&+ \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + \nu_2.
\end{aligned}
$$

- $\rightarrow$ L'identification de l'équation de salaire requiert : $\pi_{25} \neq 0$ ou $\pi_{26} \neq 0$.
- $\rightarrow$ Tester $H_0 : \pi_{25} = \pi_{26} = 0$ via un F-test.

Exemples d'identification

- La condition d'ordre de l'équation de salaire est également remplie $(age, kidslt6$ et $nwifeinc$ sont exclues$)$.

Exemples d'identification

- La condition d'ordre de l'équation de salaire est également remplie $(age, kidslt6$ et $nwifeinc$ sont exclues$)$.
- L'équation de salaire est identifiée si β_{12}, β_{13} ou $\beta_{14} \neq 0$. $\rightarrow$ Si π_{12}, π_{13} ou $\pi_{14} \neq 0$ (condition de rang) dans l'équation en forme réduite suivante :

$$\begin{aligned} hours &= \pi_{10} + \pi_{11} educ + \pi_{12} age + \pi_{13} kidslt6 \\ &+ \pi_{14} nwifeinc + \pi_{15} exper + \pi_{16} exper^2 + \nu_1. \end{aligned}$$

Estimation par DMC

- L'estimation par DMC peut se faire pour une équation identifiée.

Estimation par DMC

- L'estimation par DMC peut se faire pour une équation identifiée.
- Dans ce cas, les VI sont simplement les variables exogènes de l'autre équation.

Estimation par DMC

- L'estimation par DMC peut se faire pour une équation identifiée.
- Dans ce cas, les VI sont simplement les variables exogènes de l'autre équation.
- Exemple : Offre de travail des femmes mariées (mroz.in7).

Estimation par DMC

- L'estimation par DMC peut se faire pour une équation identifiée.
- Dans ce cas, les VI sont simplement les variables exogènes de l'autre équation.
- Exemple : Offre de travail des femmes mariées (mroz.in7).
- Fonction d'offre de travail :

$$\begin{aligned} hours &= \alpha_1 log(wage) + \beta_{10} + \beta_{11} educ \\ &+ \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1, \end{aligned}$$

Estimation par DMC

- L'estimation par DMC peut se faire pour une équation identifiée.
- Dans ce cas, les VI sont simplement les variables exogènes de l'autre équation.
- Exemple : Offre de travail des femmes mariées (mroz.in7).
- Fonction d'offre de travail :

$$\begin{aligned} hours &= \alpha_1 log(wage) + \beta_{10} + \beta_{11} educ \\ &+ \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1, \end{aligned}$$

- Equation de salaire : $log(wage) = \alpha_2 hours + \beta_{20} + \beta_{21} educ + \beta_{22} exper + \beta_{23} exper^2 + u_2$.

Estimation : fonction d'offre de travail

```
EQ(1) Modelling HOURS by IVE-CS (using mroz)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant              2225.66      574.6     3.87   0.000
LWAGE          Y      1639.56      470.6     3.48   0.001
EDUC                 -183.751      59.10    -3.11   0.002
AGE                  -7.80609      9.378   -0.832   0.406
KIDSLT6              -198.154      182.9    -1.08   0.279
NWIFEINC             -10.1696      6.615    -1.54   0.125

sigma                  1354.2  RSS               773893120
Reduced form sigma     729.76
no. endogenous variables    2  no. of instruments        7
Additional instruments:
[0] = EXPER
[1] = EXPERSQ

Specification test: Chi^2(1) =  0.86220 [0.3531]
Testing beta = 0:   Chi^2(5) =   17.205 [0.0041]**
```

Interprétation

- Pente positive de la fonction d'offre de travail. Le coefficient de $log(wage) = 1639.56$ et est très significatif.

Interprétation

- Pente positive de la fonction d'offre de travail. Le coefficient de $log(wage) = 1639.56$ et est très significatif.
- *Ceteris Paribus*, $\Delta \widehat{hours} \approx 16.4(\%\Delta wage)$.

Interprétation

- Pente positive de la fonction d'offre de travail. Le coefficient de $log(wage) = 1639.56$ et est très significatif.
- *Ceteris Paribus*, $\Delta \hat{hours} \approx 16.4(\%\Delta wage)$.
- L'élasticité de l'offre de travail $= 1640/hours$ car

$$
\begin{aligned}
100(\Delta \hat{h}ours/hours) &\approx (1640/hours)(\%\Delta wage) \\
\%\Delta \hat{h}ours &\approx (1640/hours)(\%\Delta wage)
\end{aligned}
$$

Interprétation

- Pente positive de la fonction d'offre de travail. Le coefficient de $log(wage) = 1639.56$ et est très significatif.
- *Ceteris Paribus*, $\Delta h\hat{ours} \approx 16.4(\%\Delta wage)$.
- L'élasticité de l'offre de travail $= 1640/hours$ car

$$
\begin{aligned}
100(\Delta \hat{h}ours/hours) &\approx (1640/hours)(\%\Delta wage) \\
\%\Delta \hat{h}ours &\approx (1640/hours)(\%\Delta wage)
\end{aligned}
$$

- Si $hours = \overline{hours} = 1303$, l'élasticité de l'offre de travail $\approx 1640/1303 = 1,26\% > 1\%$.

Interprétation

- Pente positive de la fonction d'offre de travail. Le coefficient de $log(wage) = 1639.56$ et est très significatif.
- *Ceteris Paribus*, $\Delta \hat{hours} \approx 16.4(\%\Delta wage)$.
- L'élasticité de l'offre de travail $= 1640/hours$ car

$$\begin{aligned} 100(\Delta \hat{hours}/hours) &\approx (1640/hours)(\%\Delta wage) \\ \%\Delta \hat{hours} &\approx (1640/hours)(\%\Delta wage) \end{aligned}$$

- Si $hours = \overline{hours} = 1303$, l'élasticité de l'offre de travail $\approx 1640/1303 = 1,26\% > 1\%$.
- Si $hours = 800$ l'élasticité de l'offre de travail $\approx 1640/800 = 2,05\%$.

Estimation par MCO

```
EQ(2) Modelling HOURS by OLS-CS (using mroz)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant              1523.77      305.6     4.99   0.000
EDUC                 -6.62187      18.12   -0.366   0.715
LWAGE                -2.04680      54.88  -0.0373   0.970
AGE                  0.562254      5.140    0.109   0.913
KIDSLT6              -328.858      101.5    -3.24   0.001
NWIFEINC             -5.91846      3.683    -1.61   0.109

sigma                 766.633  RSS                248020491
R^2                 0.0361062  F(5,422) = 3.162 [0.008]**
log-likelihood       -3447.06  DW                      2.03
no. of observations       428  no. of parameters          6
mean(HOURS)           1302.93  var(HOURS)            601194
```

$\rightarrow \hat{\alpha}_1 = -2,04680$ et est non significatif.

Condition de rang pour $hours$

```
EQ(3) Modelling LWAGE by OLS-CS (using mroz.in7)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant            -0.447161     0.2852    -1.57   0.118
EDUC                 0.101111    0.01496     6.76   0.000
AGE                -0.00255613   0.005192  -0.492   0.623
KIDSLT6             -0.0532185   0.08844   -0.602   0.548
NWIFEINC            0.00555999   0.003310    1.68   0.094
EXPER                0.0418643   0.01324     3.16   0.002
EXPERSQ          -0.000762482  0.0004008   -1.90   0.058

LinRes F(2,421) =   9.3293 [0.0001]**
```

$$\begin{aligned} log(wage) &= \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt6 \\ &+ \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + \nu_2. \end{aligned}$$

$\longrightarrow$ Rejet de $H_0 : \pi_{25} = \pi_{26} = 0$.

Estimation par DMCO

```
EQ(1) Modelling HOURS by IVE-CS (using mroz)
       The estimation sample is: 1 to 428

                  Coefficient  Std.Error  t-value  t-prob
Constant              2225.66      574.6     3.87   0.000
LWAGE           Y     1639.56      470.6     3.48   0.001
EDUC                 -183.751      59.10    -3.11   0.002
AGE                  -7.80609      9.378   -0.832   0.406
KIDSLT6              -198.154      182.9    -1.08   0.279
NWIFEINC             -10.1696      6.615    -1.54   0.125

Additional instruments:
[0] = EXPER
[1] = EXPERSQ

Specification test: Chi^2(1) =  0.86220 [0.3531]
Testing beta = 0:   Chi^2(5) =   17.205 [0.0041]**
```

$\longrightarrow$ Les instruments passent le test de suridentification de Sargan (Specification test).

Test d'endogénéité de $log(wage)$

```
EQ(4) Modelling LWAGE by OLS-CS (using mroz)
       The estimation sample is: 1 to 428
```

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	-0.447161	0.2852	-1.57	0.118	0.0058
EDUC	0.101111	0.01496	6.76	0.000	0.0979
AGE	-0.00255613	0.005192	-0.492	0.623	0.0006
KIDSLT6	-0.0532185	0.08844	-0.602	0.548	0.0009
NWIFEINC	0.00555999	0.003310	1.68	0.094	0.0067
EXPER	0.0418643	0.01324	3.16	0.002	0.0232
EXPERSQ	-0.000762482	0.0004008	-1.90	0.058	0.0085

```
res [1 to 428] saved to mroz
```

$\rightarrow$ Estimer par MCO le modèle en forme réduite pour $log(wage)$.

$\rightarrow$ $log(wage) = \pi_0 + \pi_1 educ + \pi_2 age + \pi_3 kidslt6 + \pi_4 nwifeinc + \pi_5 exper + \pi_6 exper^2 + v_2$.

$\rightarrow$ Sauver $\hat{v}_2(res)$.

Test d'endogénéité de $log(wage)$

```
EQ(5) Modelling HOURS by OLS-CS (using mroz)
       The estimation sample is: 1 to 428
```

	Coefficient	Std.Error	t-value	t-prob
Constant	2225.66	310.0	7.18	0.000
EDUC	-183.751	31.88	-5.76	0.000
LWAGE	1639.56	253.9	6.46	0.000
AGE	-7.80609	5.059	-1.54	0.124
KIDSLT6	-198.154	98.69	-2.01	0.045
NWIFEINC	-10.1696	3.569	-2.85	0.005
res	-1714.36	259.4	-6.61	0.000

$\longrightarrow$ Estimer par MCO le modèle

$hours = \beta_{10} + \beta_{11} educ + \beta_{12} age + \beta_{13} kidslt6 + \beta_{14} nwifeinc + \beta_{15} log(wage) + \delta_1 \hat{v}_2 + u_1.$

$\longrightarrow$ $\hat{\delta}_1 = -1714.36$ et très significatif.

$\longrightarrow$ Endogénéité de $log(wage)$.

Estimation de la fonction de salaire

$\longrightarrow$ $hours$ n'est pas significatif.

MES à plus de 2 équations

Considérons le cas d'un MES à 3 équations

- $$\begin{aligned} y_1 &= \alpha_{12} y_2 + \alpha_{13} y_3 + \beta_{11} z_1 + u_1 \\ y_2 &= \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2 \\ y_3 &= \alpha_{32} y_2 + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3 + \beta_{34} z_4 + u_3. \end{aligned}$$

MES à plus de 2 équations

Considérons le cas d'un MES à 3 équations

- $$\begin{aligned} y_1 &= \alpha_{12} y_2 + \alpha_{13} y_3 + \beta_{11} z_1 + u_1 \\ y_2 &= \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2 \\ y_3 &= \alpha_{32} y_2 + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3 + \beta_{34} z_4 + u_3. \end{aligned}$$

- Il n'est pas facile de déterminer les équations identifiées mais plus facile de déterminer les équations non identifiées.

MES à plus de 2 équations

Considérons le cas d'un MES à 3 équations

- $$\begin{aligned} y_1 &= \alpha_{12} y_2 + \alpha_{13} y_3 + \beta_{11} z_1 + u_1 \\ y_2 &= \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2 \\ y_3 &= \alpha_{32} y_2 + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3 + \beta_{34} z_4 + u_3. \end{aligned}$$
- Il n'est pas facile de déterminer les équations identifiées mais plus facile de déterminer les équations non identifiées.
- L'équation expliquant y_3 est clairement non identifiée car toutes les variables exogènes sont présentes.
  $\rightarrow$ Il ne reste plus de VI pour estimer α_{32}.

MES à plus de 2 équations

- La condition d'ordre de l'équation décrivant y_1 semble remplie car il y a 2 variables endogènes $(y_2$ et $y_3)$ et on dispose de 3 instruments $(z_2, z_3$ et $z_4)$.

MES à plus de 2 équations

- La condition d'ordre de l'équation décrivant y_1 semble remplie car il y a 2 variables endogènes $(y_2$ et $y_3)$ et on dispose de 3 instruments $(z_2, z_3$ et $z_4)$.
- Condition d'ordre pour l'identification.

MES à plus de 2 équations

- La condition d'ordre de l'équation décrivant y_1 semble remplie car il y a 2 variables endogènes $(y_2$ et $y_3)$ et on dispose de 3 instruments $(z_2, z_3$ et $z_4)$.
- Condition d'ordre pour l'identification.
- Une équation dans un MES respecte la condition d'ordre pour l'identification si le nombre de variables exogènes exclues de l'équation est plus grand ou égal au nombre de variables explicatives endogènes de l'équation.

MES à plus de 2 équations

- La condition d'ordre de l'équation décrivant y_1 semble remplie car il y a 2 variables endogènes $(y_2$ et $y_3)$ et on dispose de 3 instruments $(z_2, z_3$ et $z_4)$.
- Condition d'ordre pour l'identification.
- Une équation dans un MES respecte la condition d'ordre pour l'identification si le nombre de variables exogènes exclues de l'équation est plus grand ou égal au nombre de variables explicatives endogènes de l'équation.
- La condition d'ordre est une condition nécessaire mais non suffisante. $\rightarrow$ Les conditions suffisantes sont plus difficiles à déterminer.

MES à plus de 2 équations

$y_2 = \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2$.

- Pour identifier la seconde équation, on dispose d'un instrument (z_4) pour une variable endogène (y_1).
  $\longrightarrow$ La condition d'ordre est remplie. Mais il faut également que $\beta_{34} \neq 0$.

MES à plus de 2 équations

$y_2 = \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2$.

- Pour identifier la seconde équation, on dispose d'un instrument (z_4) pour une variable endogène (y_1).
  $\rightarrow$ La condition d'ordre est remplie. Mais il faut également que $\beta_{34} \neq 0$.
- La première équation est sur-identifiée car on dispose de 3 VI (z_1, z_2, z_3) pour 2 variables endogènes (y_2, y_3).