Problème d'endogénéité : Méthode de V. instrumentales

I/ hypothèses des MCO du modèle linéaire : Exogénéité

$Y = X\beta + \varepsilon \rightarrow$ H1 : ${}^tXX$ inv et rang $({}^tX.X) = K$

H2 : $E(\varepsilon/X) = 0$ exogénéité   H3 : $V(\varepsilon/X) = \sigma^2$ H4 : $E(\varepsilon_i \varepsilon_j / X) = 0$ $\forall i \neq j$

Les causes de l'endogénéité : $E(\varepsilon/X) \neq 0$

1/ Problème d'endogénéité et biais de variable omise :

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \rightarrow$ omettant la var x_2 $\rightarrow$ On estime

$y = \delta_0 + \delta_1 x_1 + u$ avec $u = \beta_2 x_2 + \varepsilon$

2/ Problème d'endogénéité et causalité : Actusport $\overset{x}{}\leftrightarrow\overset{y}{}$ poids

on estime $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ et il y a une causalité inv

tq $x_1 = \delta_0 + \delta_2 x_2 + \gamma Z + \alpha y + u \rightarrow Cov(x_1, \varepsilon) \neq 0$

Estimer $\ln(PIB_t) - \ln(PIB_{t-1}) = \beta_0 + \beta_1 INV_t + \beta_2 EXP_t + \varepsilon$

sachant $INV_t = \delta_0 + \delta_1 PIB_{t-1} + \delta_2 Z + v$ $EXP_t = \gamma_0 + \gamma_1 PIB_{t-1} + \gamma_2 Z + v$

3/ Erreur de mesures CEV sur la variable explicative
erreur classique

CEV est corrélés avec au moins 1 carac de l'indiv

$y = \beta_0 + \beta_1 x_1^* + \varepsilon$ variable non observée et on observe $e_1 = x_1 - x_1^*$

$\perp \varepsilon \leftrightarrow \perp x_1^*$

$\rightarrow y = \beta_0 + \beta_1 x_1 + (\varepsilon - \beta_1 e_1)$ tq

$Cov(x_1, \varepsilon - \beta_1 e_1) \neq 0 = \beta_1 \sigma^2_{e_1} \rightarrow$ corrélé avec $\varepsilon - \beta_1 e_1$

Estimation par l'approche des variables instrumentales

Estimation MCO biaisée $\rightarrow$ éliminer le biais par les VI ou

Par MC en 2 étapes "2SLS"

l'estimation VI : exige des variables supplémentaires.

Soit le modèle : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ avec $Cov(x_1, \varepsilon) \neq 0$

z : instrument de la var x_1 $\quad E(z\varepsilon) = 0$

z : var exogène $\Rightarrow Cov(z, \varepsilon) = 0$: Cond d'exogénéité des VI

nbr des var inst = nbr des var endogènes $\quad m(Z) \geq m(VEnd)$

z variable corrélée à x_1 : $Cov(z, x_1) \neq 0$: Condition de pertinence des IVs $\quad E(XZ) \neq 0$

Σ_{ZX} : matrice de covariance entre variables explicatives endogène $\Rightarrow$ inversible

$\Rightarrow$ purger l'effet de la variable omise sur x_k

Comment trouver un instrument valide ?

$\rightarrow$ La théorie économique : $Cov(z, \varepsilon) = 0$

$\rightarrow$ Tester formellement la prop $Cov(z, x_1) \neq 0$ par le wald-test

$x = a_0 + a_1 z + v \quad H_0 : a_1 = 0$ pour s'assurer qu'au moins un des coeff des instruments est significativement non nul

---

Variables Instrumentales - Regression simple :

$y_i = \beta x_i + \varepsilon_i \qquad \frac{1}{N} \sum (y_i - \hat{\beta}_{VI} x_i) z_i = 0$

$E(\varepsilon / z) = 0 \Rightarrow E(\hat{\varepsilon}_{VI} / z) = 0 \Rightarrow \hat{\beta}_{VI} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\sum z_i \varepsilon_i}{\sum z_i x_i}$

$V(\hat{\beta}_{VI}) = \sigma^2 . \frac{\sum z_i^2}{(\sum z_i x_i)^2} \qquad \hat{\varepsilon}_i = y_i - \hat{\beta}_{VI} x_i$

Soit $S^2 = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 \Rightarrow$ Estimateur consistant de σ^2

...s method:

$E(\varepsilon / z) = 0 \Rightarrow \hat{\beta}_{OLS} = \frac{\sum x_i y_i}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$

$V(\hat{\beta}_{OLS}) = \frac{\sigma^2}{\sum x_i^2} \quad \hat{\varepsilon}_i = y_i - \hat{\beta}_{OLS} x_i$

Variables Instrumentales - Regression multiples:

$Y = X\beta + \varepsilon \quad X(N \times K) \quad Z(N \times L)$: matrice des var exogènes (exp + inst)

$\Rightarrow$ Pb : $\min {}^t(Y - X\beta) P_Z (Y - X\beta)$ avec $P_Z = Z({}^tZZ)^{-1}\,{}^tZ$

${}^tP_Z . P_Z = P_Z$ (idempotent)

$\hookrightarrow \hat{\beta}_{VI} = ({}^tX P_Z X)^{-1}\,{}^tX P_Z Y$

Si $Z(N \times K)$: $\hat{\beta}_{VI} = ({}^tZX)^{-1}\,{}^tZY \quad V(\hat{\beta}_{VI}) = \frac{\hat{\varepsilon}_{VI} . {}^t\hat{\varepsilon}_{VI}}{N - K} ({}^tX P_Z X)^{-1}$

$\Rightarrow$ l'estimateur VI peut être aussi obtenu par MC en 2 étapes

- Estimer les param par les MCO et calculer les $\hat{x}_i$
- Effectuer une reg de y sur le vect $x_1, \dots, x_{K-1}, \hat{x}_K$ et obtenir $\hat{\beta}_{2SLS}$

Cas pratique : Estimation de l'équation d'offre :

$q_s = a_0 + a_1 P + \varepsilon_s \quad q_d = \beta_0 + \beta_1 P + \beta_2 r + \varepsilon_d \quad q_s \equiv q_d$

N = 20 marchés q : cons / hab P : prix r : Revenu

1/ Estimation classique du modèle :

reg q P : $q = 0{,}063 P + 94{,}68 \quad R^2 = 0{,}0096$

$\Rightarrow$ Résultat biaisé, incompatible et incohérent (l'hypothèse d'exog du P est non valide)

Skewness/Kurtosis test : sk = 0,1787 : dist inclinée vers la droite

Kurt = 0,88 : a plus de valeurs extrême que la loi normale.

Breush-Pegan: hétéroscédasticité : H_0 : variance const

$Pr = 0,81$: Accep H_0

Ramsey test: H_0: pas de ommitted variables $F=0,97$, $Pr=0,432$

2/ Estimation 2SLS:

1ere étape qui est basé sur la condition: $Cov(X,Z) = Cov(P,r) \neq 0$

reg P r: $P = 0,284\, r + 72,34$ test r : $r=0$
$P=0,009$ $P=0,0092 \Rightarrow$ sign $\neq 0$

→ l'instrument r est significativement $\neq 0 \Rightarrow$ r est bien en relation linéaire avec la variable endog explicative p

2ème étape : les var endog exp sont remplacées par les valeurs estimées : $Cov(\varepsilon, Z) = Cov(\varepsilon, r) = 0$: Cond d'exog des VI

$q = 0,863\, P + 14,597$

3/ Estimation VI : $\neq$ OLS

→ Définit un $R^2 \notin [0,1]$ car l'orthogonalité des résidus se réfère à l'inst et non aux var explicatives.

GR^2 = principe général de sélection de modèle

$GR^2 = y - \hat{x}\hat{\beta}_{2SLS}$

→ Les vraies erreurs standard sont ceux de la méthode VI.

IV (2SLS) regression: $q = 0,863\, P + 14,596$

First stage regression: $P = 0,284\, r + 72,34$

... Test d'exogénéité des régresseurs du modèle :

→ test de comparaison entre les deux estimateurs

Hausman-test : $H_0 : E(\varepsilon/X) = 0$ contre $H_1 : E(\varepsilon/X) \neq 0$

$\hat{\beta}_{E,OLS}$ consistant et efficient $\hat{\beta}_{E,OLS}$ non consistant

$\hat{\beta}_{C,VI}$ consistant et non efficient $\hat{\beta}_{C,VI}$ consistant

β_{OLS} → effet aléatoire RE, $\hat{\beta}_{VI}$ → Les données de Panel

Hausman test : tester si $\hat{\beta}_C - \hat{\beta}_E$ est stat significative

H_0 Rejetée ⇒ Rejet $E(\varepsilon/X) = 0$ ⇒ On doit estimer notre modèle par VI method

H_0 acceptée ⇒ les deux estimateurs sont consistants

La stat du test : ${}^t(\hat{\beta}_C - \hat{\beta}_E)(Var(\hat{\beta}_C - \hat{\beta}_E))^{-1}(\beta_C - \beta_E)$

où $Var(\hat{\beta}_C - \hat{\beta}_E)$ est définie positive càd $Var(\hat{\beta}_C) > Var(\hat{\beta}_E)$