Université de Grenoble Alpes
(L3 MIASH, S2)

ÉCONOMÉTRIE

Régression Linéaire, Endogénéité, et Variables Instrumentales (1)

(Cette version : 12 février 2024)

Michal W. Urdanivia [1]

1. Contact : michal.wong-urdanivia@univ-grenoble-alpes.fr, Université de Grenoble Alpes, Faculté d'Économie, GAEL.

Table des matières

1. Exogénéité(s) dans un modèle de régression linéaire

L'objectif d'un modèle de régression linéaire est de mesurer la relation entre une variable $Y \in \mathbb{R}$ appelée variable dépendante du modèle et un vecteur de $X \in \mathbb{R}^K$ appelé (vecteur des) régresseurs du modèle. Les variables (Y, X) sont supposées être des variables aléatoires car elles portent sur un phénomène imparfaitement observé par le chercheur. En outre, on se concentre sur l'espérance conditionnelle $\mathrm{E}(Y|X)$ et on suppose aussi que la relation entre Y et X obéit à :

$$Y = X^\intercal \beta + U. \tag{1}$$

Dans (1) U est une variable aléatoire réelle qui représente(résume) l'ensemble des facteurs inobservés induisant des variations de Y pour des valeurs de X données. $\beta \in \mathbb{R}^K$ est un vecteur de paramètres inconnus dont la connaissance nous permettrait si l'on pouvait fixer les valeurs de U, de mesurer les effets de X sur Y. Par exemple pour un élément X_k de X continu nous pourrions calculer $\partial Y/\partial X_k = \beta_k$, pour $k = 1, \ldots, K$.

U n'étant pas observé on se concentre sur un modèle de régression linéaire qui est un modèle pour $\mathrm{E}(Y|X)$ où l'on suppose que,

$$\mathrm{E}(Y|X) = X^\intercal \beta. \tag{2}$$

où il est aussi supposé que $\mathrm{E}(|Y|) < \infty)$ de sorte que $\mathrm{E}(Y|X)$ existe.

Il apparaît que pour avoir (2) tout en supposant (1) on doit aussi supposer dans (1) que U est en moyenne indépendant de X,

$$\mathrm{E}(U|X) = 0, \tag{3}$$

car alors,

$$\mathrm{E}(Y|X) = X^\intercal \beta + \mathrm{E}(U|X) = X^\intercal \beta.$$

Dans ce cas β permet de mesurer les effets en moyenne d'un régresseur X_k sur Y pour des valeurs données des autres régresseurs et sachant qu'en moyenne les déterminats inobservés de Y représentés par U sont indépendants des régresseurs X. Autrement dit, $\beta_k = \partial\, \mathrm{E}(Y|X)/\partial X_k$ est un effet causal car la variation de Y induite par celle de X_k ne résulte que de celle-ci et non d'autres facteurs que l'on ne peut pas contrôler car ils sont inobservés.

La condition (11) est qualifié d'**exogénéité forte**, et les régresseurs sont alors qualifiés de fortement exogènes. Cette condition implique que,

$$\mathrm{E}(U) = \mathrm{E}\left(\mathrm{E}(U|X)\right) = 0,$$

où nous avons utilisé la règle des espérance itérées. Cette dernière peut être appliqué de nouveau pour avoir,

$$\mathrm{E}(XU) = \mathrm{E}\left(X\,\mathrm{E}(U|X)\right) = 0. \tag{4}$$

Il convient ici de rappeler que (4) est une condition d'identification de β dans (1) en ce sens que sous cette condition et en supposant que $\mathrm{E}(XX^\intercal)$ soit de plein rang :

$$\beta = \left(\mathrm{E}(XX^\intercal)\right)^{-1} \mathrm{E}(XY) \tag{5}$$

La condition (4) est qualifié d'**exogénéité faible**, et les régresseurs sont alors qualifiés de faiblement exogènes. On note aussi que l'estimateur des MCO de β peut être motivé comme contrepartie empirique de (5) et/ou estimateur des moments en utilisant (4) :

$$\beta^{MCO} = \left(\sum_{i=1}^{n} X_i X_i^{\intercal}) \right)^{-1} \sum_{i=1}^{n} X_i Y_i, \tag{6}$$

où nous supposons avoir des observations i.i.d. de (Y, X), $\{(Y_i, X_i)\}_{i=1}^n$. Enfin c'est aussi en utilisant (4) qu'on peut montrer que β^{MCO} est convergent pour β.

Il est utile de résumer cette discussion :

- Nous avons deux types d'exogénéité :
  (1) L'**exogénéité forte** qui est l'indépendance en moyenne entre les déterminants inobservés U de la variable dépendante Y et les régresseurs : $\mathrm{E}(U|X) = 0$.
  (2) L'**exogénéité faible** qui est l'absence de corrélation (au sens d'une corrélation nulle) [2] entre les déterminants inobservés de la variable dépendante Y et les régresseurs : $\mathrm{E}(XU) = 0$.
- L'exogénéité permet d'identifier β sans avoir besoin de supposer l'exogénéité forte, et d'établir la convergence de l'estimateur des MCO.
- Pour pouvoir interpréter β comme les paramètres d'un modèle de régression on doit néanmoins supposer l'exogénéité forte.

2. ENDOGÉNÉITÉ

Dès lors (4) n'est pas vérifiée on parle d'endogénéité des régresseurs. Ceci ne signifie pas que $\mathrm{E}(X_k U) \neq 0$ pour tous les $k = 1, \ldots, K$. En fait cela peut seulement concerner certains d'entre eux(voire un seul) tout en posant un problème important pour l'inférence basée sur la méthode des MCO. Pour préciser ce **problème d'endogénéité** considérons la version suivante de (1),

$$Y = X_1^{\intercal}\beta_1 + X_2^{\intercal}\beta_2 + U =: X^{\intercal}\beta + U, \ \mathrm{E}(X_1 U) \neq 0, \ \mathrm{E}(X_2 U) = 0. \tag{7}$$

Dans (7) X_1 est un vecteur $(K_1 \times 1)$ de régresseurs endogènes, X_2 est un vecteur $K_2 \times 1$ de régresseurs exogènes, β_1 et β_2 sont les vecteurs de paramètres qui leur sont associés. L'ensemble des régresseurs est donc $X := (X_1^{\intercal}, X_2^{\intercal})^{\intercal}$ et celui des paramètres $\beta := (\beta_1^{\intercal}, \beta_2^{\intercal})^{\intercal}$, et $K_1 + K_2 =: K$ est le nombre total de régresseurs/paramètres.

2.1. MCO.

Considérons l'estimateur des MCO de β_1 dans (7) à partir d'observations i.i.d. de la variable dépendante Y, des régresseurs endogènes X_1, et des régresseurs exogènes X_2, $\{(Y_i, X_{1i}, X_{2i})\}_{i=1}^n$. Pour $i = 1, \ldots, n$ on considère donc,

$$Y = X_{1i}^{\intercal}\beta_1 + X_{2i}^{\intercal}\beta_2 + U_i =: X_i^{\intercal}\beta + U_i, \ \mathrm{E}(X_{1i} U_i) \neq 0, \ \mathrm{E}(X_{2i} U_i) = 0. \tag{8}$$

Utilisons la version matricielle de (8),

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{U},$$

2. Afin de parler ici d'une corrélation on suppose que $\mathrm{E}(U = 0$.

où $\mathbf{Y}$ est le vecteur $(n\times 1)$ ayant pour élément i Y_i, $\mathbf{X}_1$ est la matrice $(n\times K_1)$ de régresseurs endogènes ayant pour ligne i $X_{1i}^\intercal$, $\mathbf{X}_2$ est la matrice $(n\times K_2)$ de régresseurs exogènes ayant pour ligne i $X_{2i}^\intercal$, et $\mathbf{U}$ est le vecteur $(n\times 1)$ ayant pour élément i U_i. L'estimateur des MCO de β_1 est,

$$\begin{aligned}\widehat{\beta}_{1n} &= (\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{Y}\\ &= \beta_1 + (\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{U}\end{aligned}$$

où $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2^\intercal\mathbf{X}_2)^{-1}\mathbf{X}_2^\intercal$. Nous avons,

$$n^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{X}_1 = n^{-1}\sum_{i=1}^n X_{1i}X_{1i}^\intercal - n^{-1}\sum_{i=1}^n X_{1i}X_{2i}^\intercal\left(n^{-1}\sum_{i=1}^n X_{2i}X_{2i}^\intercal\right)^{-1} n^{-1}\sum_{i=1}^n X_{2i}X_{1i}^\intercal$$

$$n^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{U} = n^{-1}\sum_{i=1}^n X_{1i}U_i - n^{-1}\sum_{i=1}^n X_{1i}X_{2i}^\intercal\left(n^{-1}\sum_{i=1}^n X_{2i}X_{2i}^\intercal\right)^{-1} n^{-1}\sum_{i=1}^n X_{2i}U_i^\intercal$$

Supposons que,

(A.1) Les observations $\{(Y_i, X_i)\}_{i=1}^n$ sont i.i.d.

(A.2) $\mathrm{E}(X_{ik}^2) < \infty$ pour tout $k = 1, ..., K$.

(A.3) $\mathrm{E}(X_iX_i^\intercal)$ est définie positive.

(A.4) $\mathrm{E}(U_i^2) < \infty$.

Par la loi faible des grands nombre,

$$\begin{aligned}n^{-1}\sum_{i=1}^n X_{1i}X_{1i}^\intercal &\xrightarrow{p} \mathrm{E}(X_{1i}X_{1i}^\intercal)\\ n^{-1}\sum_{i=1}^n X_{1i}X_{2i}^\intercal &\xrightarrow{p} \mathrm{E}(X_{1i}X_{2i}^\intercal)\\ n^{-1}\sum_{i=1}^n X_{2i}X_{2i}^\intercal &\xrightarrow{p} \mathrm{E}(X_{2i}X_{2i}^\intercal)\\ n^{-1}\sum_{i=1}^n X_{2i}U_i^\intercal &\xrightarrow{p} 0\\ n^{-1}\sum_{i=1}^n X_{1i}U_i^\intercal &\xrightarrow{p} \mathrm{E}(X_{1i}U_i)\end{aligned}$$

Ainsi,

$$\begin{aligned}n^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{X}_1 &\xrightarrow{p} \mathrm{E}(X_{1i}X_{1i}^\intercal) - \mathrm{E}(X_{1i}X_{2i}^\intercal)\left(\mathrm{E}(X_{2i}X_{2i}^\intercal)\right)^{-1}\mathrm{E}(X_{2i}X_{1i}^\intercal)\\ n^{-1}\mathbf{X}_1^\intercal\mathbf{M}_2\mathbf{U} &\xrightarrow{p} \mathrm{E}(X_{1i}U_i) - \mathrm{E}(X_{1i}X_{2i}^\intercal)\left(\mathrm{E}(X_{2i}X_{2i}^\intercal)\right)^{-1}\mathrm{E}(X_{2i}U_i)\\ &= \mathrm{E}(X_{1i}U_i)\\ &\neq 0\end{aligned}$$

et nous concluons que $\hat{\beta}_{1n}$ n'est pas convergent,

$$\begin{aligned}\hat{\beta}_{1n} &\xrightarrow{p} \beta_1 + \left(\mathrm{E}(X_{1i}X_{1i}^\intercal) - \mathrm{E}(X_{1i}X_{2i}^\intercal)\left(\mathrm{E}(X_{2i}X_{2i}^\intercal)\right)^{-1}\mathrm{E}(X_{2i}X_{1i}^\intercal)\right)^{-1}\mathrm{E}(X_{1i}U_i)\\ &\neq \beta_1\end{aligned}$$

La non convergence de l'estimateur des MCO de β_2 peut être montré de manière similaire. Nous avons,

$$\hat{\beta}_{2n} = \beta_2 + (\mathbf{X}_2^\intercal \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^\intercal \mathbf{M}_1 \mathbf{U}$$

où $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1^\intercal \mathbf{X}_1)^{-1}\mathbf{X}_1^\intercal$. Et nous avons,

$$\begin{aligned}\hat{\beta}_{2n} &\xrightarrow{p} \beta_2 + \left(\mathrm{E}(X_{2i}X_{2i}^\intercal) - \mathrm{E}(X_{2i}X_{1i}^\intercal)\left(\mathrm{E}(X_{1i}X_{1i}^\intercal)\right)^{-1}\mathrm{E}(X_{1i}X_{2i}^\intercal)\right)^{-1}\mathrm{E}(X_{2i}X_{1i}^\intercal)\,\mathrm{E}(X_{1i}X_{1i}^\intercal)^{-1}\,\mathrm{E}(X_{1i}U_i) \\ &\neq \beta_2\end{aligned}$$

2.2. **Sources d'endogénéité.**

Variables omises. Considérons l'équation de salaire suivante,

$$\begin{aligned}\log Sal_i &= \alpha + \beta_1 Etudes + \gamma Genre + \delta Abilit + V_i \\ &= \alpha + \beta_1 Etudes + \gamma Genre + U_i\end{aligned}$$

Étant donné que *Abilit* est inobservable elle se retrouve dans le terme d'erreur du modèle $U_i = \delta Abilit + V_i$. Nous pouvons considérer que la variable *Genre* est exogène, mais *Abilit* est vraisemblablement corrélée avec le niveau d'études , et par conséquent *Etudes* est endogène.

Érreurs de mesure. Supposons que le vrai modèle soit,

$$Y_i = \tilde{X}_{i1}^\intercal \beta + X_{i2}^\intercal \beta_2 + V_i$$

où cependant $\tilde{X}_{i1}$ est inobservable. On observe à la place, $X_{i1} = \tilde{X}_{i1} + \epsilon_i$ où ϵ est un vecteur de bruits indépendant de $\tilde{X}_{i1}$, et X_{i2}. Substituons $\tilde{X}_{i1}$ dans l'équation précédente,

$$Y_i = X_{i1}^\intercal \beta + X_{i2}^\intercal \beta_2 - \epsilon_i^\intercal \beta + V_i$$

Posons $U_i = -\epsilon_i^\intercal \beta + V_i$. Alors que X_{i2} est exogène, X_{i1} est endogène car corrélé avec U_i par le biais de ϵ_i.

Simultanéité. Considérons l'équation suivante,

$$Heures_i = \beta_1 Enfants_i + X_{i2}^\intercal \beta_2 + U_i$$

où $Heures_i$ est le nombre d'heures travaillées par semaine, $Enfants_i$ est le nombre d'enfants dans une famille, et X_{i2} est un vecteur de variables exogènes. Alors que le nombre d'enfant affecte l'offre de travail, il est raisonnable de penser que les décision de carrière affectent la taille de la famille, i.e., on doit considérer une autre équation qui détermine le nombre d'enfants dans la famille,

$$Enfants_i = \gamma_1 Heures_i + Z_{i1}^\intercal \gamma_2 + + V_i$$

où Z_{i1} est un autre vecteur de variables exogènes. En substituant l'expression pour les heures dans l'équation pour le nombre d'enfants, nous obtenons(en supposant que $1 - \beta_1\gamma_1 \neq 0$),

$$Enfants_i = X_{i2}^\intercal \left(\frac{\beta_2\gamma_1}{1-\beta_1\gamma_1}\right) + Z_{i1}^\intercal \left(\frac{\gamma_2}{1-\beta_1\gamma_1}\right) + \left(\frac{\gamma_1}{1-\beta_1\gamma_1}\right) U_i + \left(\frac{1}{1-\beta_1\gamma_1}\right) V_i$$

En supposant que X_{i2}, Z_{i1}, V_i ne sont pas corrélés avec U_i, nous obtenons,

$$\begin{aligned}\mathrm{E}(U_i Enfants_i) &= \left(\frac{\gamma_1}{1-\beta_1\gamma_1}\right)\mathrm{E}(U_I^2) \\ &\neq 0\end{aligned}$$

3. MÉTHODE DES VARIABLES INSTRUMENTALES

3.1. **Variables instrumentales.** Soit Z_{1i} un vecteur $(K_1 \times 1)$ de variables exogènes [3],

$$\mathrm{E}(Z_{1i}U_i) = 0.$$

Il est important de noter que Z_{1i} est exclu des régresseurs dans (8), i.e., Z_{1i} ne contient aucun des éléments de X_{2i}. En fait toutes ces variables sont supposées exogènes en ce sens qu'elles sont supposées satisfaire,

$$\mathrm{E}(Z_iU_i) = 0, \tag{9}$$

avec $Z_i := (Z_{1i}\intercal, X_{2i}^\intercal)^\intercal$ qui est un vecteur de $K_1 + K_2 = K$ variables exogènes.

On appelle les appelle *variables instrumentales*(VIs) ou plus simplement *instruments* des variables qui vérifient des condition du type (9) ainsi que la *condition de rang* suivante,

$$\mathrm{Rang}\left(\mathrm{E}(Z_iX_i^\intercal)\right) = K \tag{10}$$

Cette condition concerne l'information apporté par les instruments pour identifier les paramètres du modèle. On dit aussi que les instruments sont *informatifs* par rapport aux régresseurs. Elle échouera si, par exemple, $\mathrm{E}(Z_{1i}X_i^\intercal) = 0$($Z_{1i}$ est exogène mais c'est un bruit aléatoire). La condition de rang échouera aussi si certains éléments de Z_{1i} sont des combinaisons linéaires des éléments dans les régresseurs exogènes inclus X_{2i}. Par exemple, pour le cas "Heures/Enfants", Angrist et Evans(1998) on suggéré d'utiliser la composition en termes de sexe des deux premier enfants comme instrument pour le nombre d'enfants dans une famille(l'échantillon utilisé est restreint aux femmes avec au moins deux enfants). Ceci est motivé par l'idée que si les deux premiers enfants sont du même sexe(fille-fille, ou garçon-garçon) la famille sera plus encline à avoir un troisième enfant que dans le cas où les deux premiers enfants sont de sexe différent. En conséquence, la variable indicatrice d'avoir deux premiers enfants du même sexe doit être positivement corrélée avec le nombre d'enfants. D'un autre côté , l'instrument est exogène car la composition en termes de sexe des deux premiers enfants est déterminée aléatoirement.

Avant de procéder à l'estimation des paramètres du modèle notons que les régresseurs exogènes apparaissent dans le vecteur des VIs, et que pour chaque variables endogène nous avons une variables exogène(une VI) qui est exclue du modèle $Y_i = X_i^\intercal\beta + U_i$. Lorsque tous les régresseurs sont endogènes nous n'avons plus aucun élément commun à X_i et à Z_i.

3.2. **Estimation.** L'application de la méthode des moments avec la condition (9) suggère un estimateur comme solution du système suivant de K équations,

$$n^{-1}\sum_{i=1}^{n} Z_i\left(Y_i - X_i^\intercal\hat{\beta}_n^{VI}\right) = 0$$

d'où,

$$\begin{aligned}\hat{\beta}_n^{VI} &= \left(\sum_{i=1}^{n} Z_iX_i^\intercal\right)^{-1}\sum_{i=1}^{n} Z_iY_i \\ &= (\mathbf{Z}^\intercal\mathbf{X})^{-1}\mathbf{Z}^\intercal\mathbf{Y}\end{aligned}$$

3. Dans cette partie par exogénéité nous entendons la version faible de celle-ci

où $\mathbf{X}$ est la matrice $(n \times K)$ ayant pour élément i $X_i^\intercal$, et $\mathbf{Z}$ est la matrice $(n \times K)$ ayant pour élément i $Z_i^\intercal$.
L'estimateur $\hat{\beta}_n^{VI}$ est appelé *estimateur des variables instrumentales* de β.

3.3. **Convergence et normalité asymptotique.** Pour établir la convergence de $\hat{\beta}_n^{VI}$ on utilise (9)-(10) auxquelles on doit ajouter des conditions suivantes supplémentaires. Ceci est résumé ainsi :

Condition C1. (**conditions pour la convergence**)

(C1.1) Les observations $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ sont i.i.d.

(C1.2) $Y_i = X_i^\intercal \beta + U_i$.

(C1.3) $\mathrm{E}(Z_i U_i) = 0$.

(C1.4) $\mathrm{Rang}\left(\mathrm{E}(Z_i X_i^\intercal)\right) = K$.

(C1.5) $\mathrm{E}(X_{ik}^2) < \infty$ pour tout $k = 1, ..., K$.

(C1.6) $\mathrm{E}(Z_{1ik}^2) < \infty$ pour tout $k = 1, ..., K_1$.

(C1.7) $\mathrm{E}(U_i^2 Z_i Z_i^\intercal)$ est définie positive.

Propriété P1. *(**Convergence**) Sous les conditions de* (C1)

$$\hat{\beta}_n^{VI} \xrightarrow{p} \beta$$

Démonstration. Écrivons,

$$\hat{\beta}_n^{VI} = \beta + \left(n^{-1} \sum_{i=1}^n Z_i X_i^\intercal \right)^{-1} n^{-1} \sum_{i=1}^n Z_i U_i \tag{11}$$

Notons que sous les hypothèses faites plus haut, par l'inégalité de Cauchy-Schwartz,

$$\begin{aligned} \mathrm{E}(\left|Z_{i,r} X_{i,s}\right|) &\leq \sqrt{\mathrm{E}(Z_{i,r}^2)\,\mathrm{E}(X_{i,s}^2)} \\ &< \infty \text{ pour tout } r, s = 1, ..., K. \end{aligned}$$

Par conséquent, par le théorème de Slutsky,

$$\begin{aligned} \hat{\beta}_n^{VI} &\xrightarrow{p} \beta + \mathrm{E}(Z_i X_i^\intercal)^{-1}\,\mathrm{E}(Z_i U_i) \\ &= \beta \end{aligned}$$

□

La normalité asymptotique est obtenue avec deux conditions supplémentaires. Ceci est résumé ainsi :

Condition C2. (**conditions pour la normalité asymptotique**)

(C2.1) $\mathrm{E}(Z_{ik}^4) < \infty$, pour tout $k = 1, ..., K$.

(C2.2) $\mathrm{E}(U_i^4) < \infty$.

Propriété P2. *(**Normalité asymptotique**) Sous les conditions dans (C1) - (C2),*

$$\begin{aligned} n^{1/2}(\hat{\beta}_n^{VI} - \beta) &\xrightarrow{d} (\mathrm{E}(Z_i X_i^\intercal))^{-1} \mathcal{N}\left(0, \mathrm{E}(U_i^2 Z_i Z_i^\intercal)\right) \\ &= \mathcal{N}(0, \mathrm{V}) \end{aligned}$$

où

$$\mathrm{V} = (\mathrm{E}(Z_i X_i^\intercal))^{-1} \mathrm{E}(U_i^2 Z_i Z_i^\intercal)(\mathrm{E}(X_i Z_i^\intercal))^{-1}$$

Démonstration. Écrivons (11) comme suit,

$$n^{1/2}(\hat{\beta}_n^{VI} - \beta) = \left(n^{-1} \sum_{i=1}^{n} Z_i X_i^\intercal\right)^{-1} n^{-1/2} \sum_{i=1}^{n} Z_i U_i$$

Notons que du fait des hypothèses précédentes, pour tout $r, s = 1, \ldots, K$,

$$\begin{aligned} \mathrm{E}\left(\left|U_i^2 Z_{i,r} Z_{i,s}\right|\right) &\leq \left(\mathrm{E}(U_i^4)\right)^{1/2} \left(\mathrm{E}(Z_{i,r}^4)\, \mathrm{E}(Z_{i,s}^4)\right)^{1/4} \\ &< \infty \end{aligned}$$

Par conséquent, par le théorème central-limite et le théorème de convergence de Cramer,

$$\begin{aligned} n^{1/2}(\hat{\beta}_n^{VI} - \beta) &\xrightarrow{d} (\mathrm{E}(Z_i X_i^\intercal))^{-1} \mathcal{N}\left(0, \mathrm{E}(U_i^2 Z_i Z_i^\intercal)\right) \\ &= \mathcal{N}\left(0, (\mathrm{E}(Z_i X_i^\intercal))^{-1} \mathrm{E}(U_i^2 Z_i Z_i^\intercal)(\mathrm{E}(X_i Z_i^\intercal))^{-1}\right) \end{aligned}$$

□

La matrice de variances-covariances asymptotique prend une forme en sandwich et peut être estimé de manière convergente par,

$$\left(n^{-1} \sum_{i=1}^{n} Z_i X_i^\intercal\right)^{-1} n^{-1} \sum_{i=1}^{n} (\hat{U}_i^2 Z_i Z_i^\intercal) \left(n^{-1} \sum_{i=1}^{n} X_i Z_i^\intercal\right)^{-1}$$

où $\hat{U}_i = Y_i - X_i^\intercal \hat{\beta}_n^{VI}$.