Université de Grenoble Alpes
(L3 MIASH, S2)

ÉCONOMÉTRIE

Régression Linéaire et Hétéroscédasticité :

Moindres Carrés Généralisés

(Cette version : 3 mars 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. Estimateur des moindres carrés généralisés 2
2. Propriétés asymptotiques 4
3. Moindres carrés quasi-généralisés 5

1. Estimateur des moindres carrés généralisés

On considère le modèle linéaire sans endogénéité étudié dans les cours précédents. Toutefois, plutôt que la condition $\mathrm{E}(X_iU_i) = 0$ nous allons utiliser la condition(plus forte) $\mathrm{E}(U_i|X_i) = 0$. En résumé, le conditions que l'on impose sont les suivantes ;

Condition C1. Les données $\{(Y_i, X_i), i = 1, \ldots, n\}$ sont un échantillon i.i.d.

Condition C2. Y_i et X_i vérifient,

$$Y_i = X_i^\intercal\beta + U_i \quad i = 1, \ldots, n$$

où U_i est une variable inobservée(ou terme d'erreur) vérifiant $\mathrm{E}(U_i) = 0$.

Condition C3. X_i est exogène par rapport à U_i,

$$\mathrm{E}(U_i|X_i) = 0$$

Condition C4. La matrice $\mathrm{E}(X_iX_i^\intercal)$ est finie et définie positive.

Condition C5. $\mathrm{E}(X_{i,k}^4) < \infty$, pour tout $k = 1, \ldots, K$.

Condition C6. $\mathrm{E}(U_i^4) < \infty$

Condition C7. $\mathrm{E}(U_i^2X_iX_i^\intercal)$ est définie positive.

Comme cela a été déjà noté la condition (C3) est plus forte que celle dont on a besoin pour établir la convergence et la normalité asymptotique de l'estimateur des MCO. Toutefois elle va nous permettre d'étudier le problème de l'efficacité dans le cas d'erreurs hétécédastiques où $\mathrm{E}(U_i^2|X_i) = \sigma_i^2$, et σ_i^2 étant une fonction de X_i : $\sigma_i^2 = \sigma^2(X_i)$.

Exemple 1. Supposons que $Y_{i,j} = X_{i,j}^\intercal\beta + U_{i,j}$ pour $i = 1, ..., n$(n secteurs industriels) et $j = 1, ..., m_i$(m_i entreprises dans le secteur i). Supposons les observations i.i.d. selon i et j, et que l'on observe seulement des valeurs moyennes des variables pour les n secteurs, soit : $\bar{Y}_i = m_i^{-1}\sum_{j=1}^{m_i} Y_{i,j}$, et $\bar{X}_i = m_i^{-1}\sum_{j=1}^{m_i} X_{i,j}$. Supposons aussi que des erreurs $U_{i,j}$ homoscédastiques, i.e.,$\mathrm{E}(U_{i,j}^2|X_{i,j}) = \sigma^2$ pour tout $i = 1, ..., n$ et $j = 1, ..., m_i$. Cependant, $\bar{U}_i = m_i^{-1}\sum_{j=1}^{m_i} U_{i,j}$, et $\mathrm{E}\left(\bar{U}_i^2|\bar{X}_i\right) = \sigma^2/m_i$.

En présence d'hétéroscédasticité, l'estimateur des MCO est convergent et asymptotiquement normal, mais il n'est plus efficace. Il existe en effet un estimateur avec une variance asymptotique plus petite. Sous (C1), et (C3), $\mathrm{E}\left(\hat{\beta}_n|\mathbf{X}\right) = \beta$(sans biais), et,

$$\mathrm{V}\left(\hat{\beta}_n|\mathbf{X}\right) = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{D}\mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1}$$

avec,

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & 0 & . & . & . & 0 \\ 0 & \sigma_2^2 & . & . & . & 0 \\ . & 0 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & \sigma_n^2 \end{pmatrix}$$

Supposons aussi que $\sigma_i^2 = \sigma^2(X_i)$ soit connu pour tout i. L'estimateur des moindres carrés généralisés(MCG) est défini comme suit,

$$\begin{aligned}\hat{\beta}_n^{MCG} &= \left(\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{Y}\\ &= \sum_{i=1}^{n}(\sigma_i^{-2}X_iX_i^\intercal)^{-1}\sum_{i=1}^{n}\sigma^{-2}X_iY_i\end{aligned} \tag{1}$$

Dans le cas de $\mathbf{D}$ diagonale, l'estimateur des MCG est aussi appelé *estimateur des moindres carrés pondérés* car il fait intervenir les moyennes pondérées des termes $X_iX_i^\intercal$ et X_iY_i pondérées par les termes σ_i^{-2}.
Sous les conditions (C1) et (C3), on établit que $\hat{\beta}_n^{MCG}$ est sans biais,

$$\begin{aligned}\mathrm{E}\left(\hat{\beta}_n^{MCG}|\mathbf{X}\right) &= \beta + \left(\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\mathbf{D}^{-1}\,\mathrm{E}(\mathbf{U}|\mathbf{X})\\ &= \beta\end{aligned}$$

Et sa variance est donnée par,

$$\begin{aligned}\mathrm{V}\left(\hat{\beta}_n^{MCG}|\mathbf{X}\right) &= \left(\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\mathbf{D}^{-1}\,\mathrm{E}\left(\mathbf{U}\mathbf{U}^\intercal|\mathbf{X}\right)\mathbf{D}^{-1}\mathbf{X}\left(\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\\ &= \left(\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\end{aligned}$$

Montrons à présent que $\mathrm{V}\left(\hat{\beta}_n^{MCG}|\mathbf{X}\right) \leq \mathrm{V}\left(\hat{\beta}_n|\mathbf{X}\right)$. Notons d'abord que,

$$\mathrm{V}\left(\hat{\beta}_n^{MCG}|\mathbf{X}\right) \leq \mathrm{V}\left(\hat{\beta}_n|\mathbf{X}\right) \Leftrightarrow \left(\mathrm{V}\left(\hat{\beta}_n^{MCG}|\mathbf{X}\right)\right)^{-1} \geq \left(\mathrm{V}\left(\hat{\beta}_n|\mathbf{X}\right)\right)^{-1}$$

Ensuite que,

$$\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X} - \mathbf{X}^\intercal\mathbf{X}\left(\mathbf{X}^\intercal\mathbf{D}\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\mathbf{X} = \mathbf{X}^\intercal\mathbf{D}^{-1/2}\left(\mathbf{I} - \mathbf{D}^{1/2}\mathbf{X}\left(\mathbf{X}^\intercal\mathbf{D}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{D}^{1/2}\right)\mathbf{D}^{-1/2}\mathbf{X}$$

où l'on note que $\mathbf{I} - \mathbf{D}^{1/2}\mathbf{X}\left(\mathbf{X}^\intercal\mathbf{D}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{D}^{1/2}$ est symmétrique et semi-définie positive, et par conséquent,

$$\mathbf{X}^\intercal\mathbf{D}^{-1}\mathbf{X} \geq \mathbf{X}^\intercal\mathbf{X}\left(\mathbf{X}^\intercal\mathbf{D}\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\mathbf{X}$$

L'efficacité de l'estimateur des MCG résulte de l'application du théorème de Gauss-Markov. En effet, avec l'hétéroscédasticité des erreurs une des conditions pour l'obtenir n'est plus satisfaite. Mais considèrons le modèle tranformé suivant,

$$\tilde{Y}_i = \tilde{X}_i^\intercal\beta + \tilde{U}_i$$

où $\tilde{Y}_i = Y_i/\sigma$, $\tilde{X}_i = X_i/\sigma$, et $\tilde{U}_i = U_i/\sigma$. Les erreurs transformés sont homoscédastiques,

$$\begin{aligned}\mathrm{E}\left(\tilde{U}_i^2|X_i\right) &= \sigma_i^2/\sigma_i^2\\ &= 1\end{aligned}$$

et par le théorème de Gauss-Markov l'estimateur BLUE est donné par,

$$\begin{aligned}\left(\sum_{i=1}^{n} \tilde{X}_i\tilde{X}_i^\intercal\right)^{-1} \sum_{i=1}^{n} \tilde{X}_i\tilde{Y}_i &= \left(\sum_{i=1}^{n} \sigma_i^{-2}X_iX_i^\intercal\right)^{-1} \sum_{i=1}^{n} \sigma_i^{-2}X_iY_i \\ &= \hat{\beta}_n^{MCG}\end{aligned}$$

2. PROPRIÉTÉS ASYMPTOTIQUES

On commence par discuter la convergence de l'estimateur. Pour cela écrivons,

$$\hat{\beta}_n^{MCG} = \beta + \left(\sum_{i=1}^{n}(\sigma_i^{-2}X_iX_i^\intercal\right)^{-1} \sum_{i=1}^{n} \sigma_i^{-2}X_iU_i$$

Nous supposons que la fonction $\sigma^2(X_i)$ présente une borne inférieure, i.e., avec une probabilité de 1, $\sigma^2(X_i) > \underline{\sigma}^2 > 0$. Ce qui garantit, que $E(\sigma_i^{-2}X_iX_i^\intercal)$ "n'explose pas". Pour $k, l = 1, ..., K$, nous avons $E\left(\left|\sigma_i^{-2}X_{i,k}X_{i,l}\right|\right) \leq \underline{\sigma}^2\, E\left(\left|X_{i,k}X_{i,l}\right|\right) < \infty$(en utilisant la condition (C4)). Par la loi faible des grands nombres et le théorème de Slutsky,

$$\left(\sum_{i=1}^{n} \sigma_i^{-2}X_iX_i^\intercal\right)^{-1} \xrightarrow{p} \left(E(\sigma_i^{-2}X_iX_i^\intercal)\right)^{-1}$$

Nous avons aussi que,

$$\begin{aligned}E\left(\sigma_i^{-2}X_iU_i\right) &= E\left(\sigma_i^{-2}X_i\, E(U_i|X_i)\right) \\ &= 0\end{aligned}$$

et ainsi,

$$\sum_{i=1}^{n} \sigma_i^{-2}X_iU_i \xrightarrow{p} 0$$

Par conséquent, $\hat{\beta}_n^{MCG} \xrightarrow{p} \beta$ lorsque $n \to \infty$. Notons qu'en général, $\hat{\beta}_n^{MCG}$ n'est pas convergent sous la condition $E(X_iU_i) = 0$, étant donné que comme σ_i^2 est une fonction de X_i on ne peut pas garantir que $E\left(\sigma_i^{-2}X_iU_i\right) = 0$ avec $E(X_iU_i) = 0$. Pour établir la normalité asymptotique de $\hat{\beta}_n^{MCG}$, écrivons,

$$n^{1/2}\left(\hat{\beta}_n^{MCG} - \beta\right) = \left(n^{-1}\sum_{i=1}^{n} \sigma_i^{-2}X_iX_i^\intercal\right)^{-1} n^{-1/2}\sum_{i=1}^{n} \sigma_i^{-2}X_iU_i$$

Nous avons,

$$\begin{aligned}V\left(\sigma_i^{-2}X_iU_i\right) &= E\left(\sigma_i^{-4}X_iX_i^\intercal U_i^2\right) \\ &= E\left(\sigma_i^{-4}X_iX_i^\intercal\, E(U_i^2|X_i)\right) \\ &= E\left(\sigma_i^{-2}X_iX_i^\intercal\right)\end{aligned}$$

et par conséquent,

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

3. Moindres carrés quasi-généralisés

L'estimateur des MCG n'est pas applicable dès lors que l'on ne connaît pas σ_i^2. Une solution naturelle à ce problème consiste à remplacer le terme inconnu σ_i^2 dans (1) par un estimateur $\hat{\sigma}_i^2$. Pour cela une pratique courante consiste à supposer que,

$$
\sigma_i^2 = Z_i^{\intercal}\alpha \tag{2}
$$

où Z_i est un vecteur $q\times 1$ de fonctions de X_i. Ils consistent en général en produits et produits croisés des éléments de X_i. Étant donné que $\sigma_i^2 = \mathrm{E}(U_i^2|X_i)$, on peut écrire,

$$
U_i^2 = Z_i^{\intercal}\alpha + \nu_i
$$

où $\mathrm{E}(\nu_i|Z_i) = 0$. Ce modèle est appelé *régression scédastique*. Comme les termes U_i ne sont pas observés on utilise les résidus de la régression estimée par MCO(de Y_i sur X_i), $\hat{U}_i$, et on estime α par,

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

Et l'on peut montrer que $\hat{\alpha}_n \xrightarrow{p} \alpha$ et $n^{1/2}(\hat{\alpha}_n-\alpha) \xrightarrow{d} \mathcal{N}(0,\mathrm{V}_\alpha)$ où V_α est la même variance que dans le cas où les termes U_i^2 étaient observés. L'estimateur des moindres carrés quasi-généralisés(MCQG) est alors donné par,

$$
\begin{aligned}
\hat{\beta}_n^{MCQG} &= \left(\mathbf{X}^{\intercal}\hat{\mathbf{D}}_n^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\intercal}\hat{\mathbf{D}}_n^{-1}\mathbf{Y}\\
&= \left(\sum_{i=1}^n \hat{\sigma}_i^{-2}X_iX_i^{\intercal}\right)^{-1}\sum_{i=1}^n \hat{\sigma}_i^{-2}X_iY_i
\end{aligned}
$$

où

$$
\hat{\mathbf{D}}_n = \begin{pmatrix}
\hat{\sigma}_1^2 & 0 & . & . & . & 0\\
0 & \hat{\sigma}_2^2 & . & . & . & 0\\
. & 0 & . & . & . & .\\
. & . & . & . & . & .\\
. & . & . & . & . & .\\
0 & . & . & . & 0 & \hat{\sigma}_n^2
\end{pmatrix}
$$

et,

$$
\hat{\sigma}_i^2 = Z_i^{\intercal}\hat{\alpha}_n
$$

De plus, on peut montrer que $\hat{\beta}_n^{MCQG} \xrightarrow{p} \beta$, et que,

$$n^{1/2}\left(\hat{\beta}_n^{MCQG} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left(\sigma_i^{-2} X_i X_i^\intercal\right)^{-1}\right) \tag{3}$$

autrement dit, la même distribution que pour l'estimateur des MCG, à condition que (2) soit correctement spécifié. Ci-après nous listons les étapes pour construire l'estimateur des MCQG.

(1) Obtenir $\hat{\beta}_n$, l'estimateur des MCO de β.

(2) Construire $\hat{U}_i = Y_i - X_i^\intercal \hat{\beta}_n$.

(3) Obtenir $\hat{\alpha}_n$, l'estimateur des MCO dans la régression de $\hat{U}_i^2$ sur Z_i.

(4) Construire $\hat{\sigma}_i^2 = Z_i^\intercal \hat{\alpha}_n$.

(5) Calculer $\hat{\beta}_n^{MCQG}$.

Un problème avec l'estimateur des MCQG dans l'approche ci-avant est que $\hat{\sigma}_i^2 = Z_i^\intercal \hat{\alpha}_n$ peut être très proche de zéro voire négatif. Plusieurs solutions sont possibles. La première consiste à tronquer la distribution des $\hat{\sigma}_i^2$ en choisissant $\underline{\sigma}^2 > 0$ et posant $\hat{\sigma}_i^2 = \max\left\{Z_i\hat{\alpha}_n, \underline{\sigma}^2\right\}$. Une alternative est de considérer une régression scédastique non-linéaire, par exemple,

$$\sigma_i^2 = \exp\left(Z_i^\intercal \alpha\right)$$

Dans ce cas à la troisième étape on doit régresser $\log\left(\hat{U}_i^2\right)$ sur Z_i et à la quatrième étape calculer $\hat{\sigma}_i^2 = \exp\left(Z_i\hat{\alpha}_n\right)$.

La procédure pour les MCQG repose sur deux hypothèses très fortes. D'abord que la régression scédastique soit correctement spécifiée. En cas de mauvaise spécification, $\hat{\sigma}_i^2$ fournit seulement une approxiamtion de σ_i^2, et la variance asymptotique dans (3) aura la forme en " sandwich",

$$\left(\mathrm{E}\left((Z_i^\intercal\alpha)^{-1} X_i X_i^\intercal\right)\right)^{-1} \mathrm{E}\left((Z_i^\intercal\alpha)^{-2}\sigma_i^2 X_i X_i^\intercal\right)\left(\mathrm{E}\left((Z_i^\intercal\alpha)^{-1} X_i X_i^\intercal\right)\right)^{-1}$$

et l'estimateur des MCQG aura une performance encore moin bonne que celle de l'estimateur des MCO. En outre, si la condition $\mathrm{E}(U_i|X_i) = 0$ n'est pas satisfaite les estimateurs des MCG et des MCQG ne seront pas convergents. Alors que l'estimateur des MCO est moins efficace que celui des MCQG sous certaines conditions, il fournit des estimateurs plus *robustes*.