

Moindres Carrés Généralisés

GSF-6053

Hiver 2025

Table des matières

1	Introduction	2
2	Multicolinéarité et Problèmes de Spécification	2
2.1	Multicolinéarité	2
2.1.1	Comment Détecter la Multicolinéarité ?	4
2.1.2	Quoi Faire en Présence de Multicolinéarité ?	5
2.2	Problèmes de Spécification	5
2.2.1	Omission d'une Variable Explicative Importante	6
2.2.2	Inclusion d'une Variable Non Pertinente	7
3	Erreurs Non-Sphériques	8
3.1	Moindres Carrés Généralisés	8
3.1.1	Introduction	8
3.1.2	Implications de $\Omega \neq I$	9
3.2	Les Moindres Carrés Généralisés	10
3.2.1	Espérance et Variance des Estimateurs GLS	12
4	Bibliographie	13
	Bibliographie	13

1 Introduction

[Gujarati et Porter](#) : Chapitres 10, 11, 12, 13 et appendice C.

[Version 7 de Greene](#) : 2, 3, 4, 5, 9, 14, 20 C et D.

[Wooldridge](#) : chapitres 3, 8, 12.

Nous avons maintenant dérivé les estimateurs moindres carrés ordinaires et maximum de vraisemblance pour le modèle linéaire classique. Il est important de réaliser que la performance de ces estimateurs dépend de la pertinence des hypothèses faites pour arriver à ce résultat.

Qu'arrive-t-il si les hypothèses ne sont plus appropriées ? Comment « sauver » nos estimateurs ? C'est ce que nous regarderons maintenant.

2 Multicolinéarité et Problèmes de Spécification

2.1 Multicolinéarité

[Gujarati et Porter](#) chapitre 10.

[Wooldridge](#) : chapitre 3. (dont 3.4a et MLR.3)

Une des hypothèses du modèle linéaire classique est qu'il n'y a pas de colinéarité parfaite entre les variables explicatives dans la matrice X . Si les régresseurs sont linéairement dépendants, nous aurons la relation suivante :

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_K X_K = 0$$

Alors la matrice $(X'X)$ n'est plus de plein rang colonne et n'est pas inversible, car une des colonnes peut être écrite en fonction linéaire des autres. Il survient alors un problème majeur au niveau de l'identification des paramètres.

De plus, la multicolinéarité « imparfaite » est aussi importante :



$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_K X_K + v_i = 0$$

Où les régresseurs ne sont pas parfaitement corrélés, mais fortement corrélés à un choc près.

Au chapitre 3, nous démontrons qu'en présence de plusieurs régresseurs, la variance des estimateurs $\hat{\beta}$ est la suivante :

$$\text{var}(\hat{\beta}) = \frac{\sigma^2}{\text{SST}_j} \frac{1}{1 - R^2}$$

$$\text{SST}_j = \sum_{i=1}^T (x_{ij} - \bar{x}_j)^2$$

(Où R^2 est le R^2 de la régression de X_j sur les $K - 2$ variables restantes et SST_j est la somme des variations totale dans la variable j .)

On voit donc que si l'on peut écrire le régresseur j parfaitement en fonction des autres régresseurs, la variance de son estimateur $\text{var}(\hat{\beta})$, est infinie puisque $R^2 = 1$. Ce problème ne viole pas l'hypothèse du modèle linéaire simple (puisque la collinéarité est imparfaite), mais le problème sera aussi important en cas de multicolinéarité imparfaite puisque le R^2 sera très élevé et la variance de l'estimateur le sera aussi. Le problème de multicolinéarité est donc sévère, car il influence la variance des estimateurs même si la corrélation n'est pas parfaite entre les régresseurs.

Notez que la multicolinéarité parfaite aura aussi un impact sur les estimateurs des MCO qui exploitent l'inversibilité de la matrice $(X'X)$. Si cette matrice ne s'inverse pas, vous n'aurez pas d'estimateurs MCO avec $\hat{\beta} = (X'X)^{-1}X'Y$.

Dans le cas de multicolinéarité imparfaite, il faut noter que les estimateurs OLS ne perdront pas la propriété BLUE, mais **CE LA NE VEUT PAS DIRE QUE LA VARIANCE DES ESTIMATEURS SERA ASSEZ PETITE POUR PERMETTRE UNE ANALYSE**. Les bonnes pro-



priétés du OLS BLUE ne vous seront pas d'une grande consolation si les écarts types sont tellement grands qu'ils ne permettent de rejeter aucune hypothèse nulle.

2.1.1 Comment Détecter la Multicolinéarité ?

Les causes de multicolinéarité sont multiples et souvent difficiles à identifier. Par exemple, avoir un échantillon qui varie peu en X ou qui est contraint en X . Ajouter des régresseurs en polynômes peut créer des problèmes. Un modèle comportant trop de régresseurs par rapport aux nombres d'observations sera problématique également. Une tendance commune dans les données peut également poser problème.

Des situations qui soulèvent des doutes sur la colinéarité des régresseurs et qui suggèrent un problème :

- Des petits changements dans les données produisent des changements importants dans les estimations.
- Les coefficients $\hat{\beta}_k$ ont des écarts types très élevés et peu de significativité individuelle, même si le test F et le R^2 de la régression sont bons.
- Les coefficients sont douteux économiquement, ils ont le mauvais signe ou une magnitude qui ne fait pas de sens économique.
- Des corrélations élevées entre les régresseurs.
- Regarder les VIF (Variance Inflation Factors) : $VIF_j = \frac{1}{1-R^2}$. Il faut que ce coefficient soit le plus petit possible. Par contre, il y a un flou sur ce que veut dire le plus petit possible.

Une façon de diagnostiquer la multicolinéarité est de pratiquer des régressions auxiliaires sur les régresseurs les uns sur les autres. Si vous êtes capable de rejeter l'hypothèse nulle de non-significativité conjointe des autres régresseurs sur un régresseur en particulier, cela suggère de la multicolinéarité. Une règle du pouce connue est de regarder les R^2 de la régression d'intérêt et des régressions auxiliaires. Le R^2 de la régression d'intérêt devrait être plus grand que ceux des régressions auxiliaires.



2.1.2 Quoi Faire en Présence de Multicolinéarité ?

- Rien... La multicolinéarité est un problème lié à votre échantillon de données. Vous n'avez pas de contrôle sur les valeurs observées... Si votre objectif est de faire de la prévision plutôt que des tests d'hypothèses, la multicolinéarité imparfaite ne vous en empêchera pas.
- Spécifier la forme de dépendance et ajuster le modèle en conséquence. Le problème ici est que nous voulons tester plutôt qu'imposer des contraintes.
- Utiliser des données et une analyse Panel plutôt que les séries chronologiques.
- Retirer des variables du modèle reste la solution la plus simple, mais cela peut introduire un biais de variables omises (voir plus bas).
- Transformer le modèle, ce qui peut entraîner des problèmes d'hétéroscédasticité.
- Trouver de nouvelles données.
- Enlever les régresseurs en polynômes.

2.2 Problèmes de Spécification

Gujarati et Porter chapitre 13. Wooldridge : Chapitre 3.

Le modèle des OLS postule que nous avons la forme « vraie » fonctionnelle qui définit la relation entre Y et des régresseurs X . Mais comment savoir si nous avons les bons X ? Quels sont les impacts de se tromper dans le choix de X ?

Nous étudions maintenant ces questions. Avant de commencer, il est important de comprendre qu'une grosse partie de la tâche d'économètre se trouve dans cette question. De façon générale, nous avons les guides suivants pour choisir les variables explicatives :

- Le modèle doit être admissible et possible à tester.
- Le modèle doit être cohérent avec la théorie économique et financière.
- Choisir des régresseurs exogènes au terme d'erreur.



- Les paramètres estimés doivent être stables empiriquement.
- Les résidus de régression. Sinon, on peut revisiter la spécification de X ou si elle est correcte corriger pour la non-sphéricité des erreurs.

2.2.1 Omission d'une Variable Explicative Importante

Wooldridge 3.3b et 3.3c

Que se passe-t-il si j'omets une variable explicative importante ? Considérons un modèle simple où le vrai modèle est :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

Je roulez la régression suivante à la place :

$$Y_t = \beta_1 X_{1t} + u_t$$

En sachant que $\hat{\beta} = (X'X)^{-1}X'Y$

Récrivons cette expression en remplaçant Y par son vrai modèle :

$$\begin{aligned}\hat{\beta}_1 &= \frac{X_1'Y}{X_1'X_1} \\ &= \frac{X_1'(\beta_1 X_1 + \beta_2 X_2 + u)}{X_1'X_1} \\ &= \beta_1 + \beta_2 \frac{X_1'X_2}{X_1'X_1} + \frac{X_1'u}{X_1'X_1}\end{aligned}$$

Sachant que $X_1'X_2$ est le coefficient de régression de X_2 sur X_1 , noté b , nous avons :

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 b$$

Conclusion : $\hat{\beta}_1$ sera un estimateur biaisé de β_1 . Le biais dépend du coefficient de la variable omise et du coefficient de la régression de la variable omise sur les variables incluses. Plus les régresseurs sont corrélés, plus ce biais sera important.



Par contre, de l'équation 1.1, on sait que la variance du modèle avec un seul régresseur sera plus petite. Par conséquent, il y a en apparence un gain dans la variance des estimateurs lorsque des coefficients sont omis. Le problème est que nous ne savons pas comment se comporte le biais. Omettre des variables va aussi faire augmenter σ^2 ce qui aura un effet pervers sur la variance.

Donc, l'inférence en présence d'une variable omise sera fallacieuse et les prédictions du modèle seront mauvaises, puisqu'elles seront basées sur un mauvais modèle.

2.2.2 Inclusion d'une Variable Non Pertinente

Wooldridge 3.3a

On suppose que le modèle suivant est vrai :

$$Y_t = \beta_1 X_{1t} + u_t$$

À la place, nous estimons :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

Quelles sont les conséquences de cette erreur de spécification sur $\hat{\beta}_{OLS}$?

Soient les régresseurs X_1 et X_2 . On peut réécrire les régresseurs du vrai modèle comme

$$X^* = \begin{bmatrix} X_1 \\ 0 \end{bmatrix}$$

Avec les

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'(\beta_1 X_1 + u)$$



$$\hat{\beta} = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix} + (X'X)^{-1}X'u$$

$$E(\hat{\beta}) = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

L'estimateur de $\hat{\beta}$ restera un estimateur sans biais. Pour cette raison, la littérature a souvent suggéré d'inclure plus de variables dans le doute pour ne pas en omettre.

Par contre, la formule 1.1 nous démontre encore que le VIF avec des variables redondantes sera plus grand et que cela fera grandir les variances.

Cette approche est aussi remise en question dans la littérature récente : on remarque que les variables redondantes causent beaucoup de problèmes d'identification plus sérieux encore que la multicollinéarité, car elles sont très difficiles à détecter (*Beaulieu, Dufour et Khalaf*, 2010 et application dans *Beaulieu, Gagnon et Khalaf* 2014) ainsi que des problèmes d'estimation et de convergence (*Kan et Zhang* 1999a et 1999b).

Il n'est donc pas recommandé d'ajouter des variables sans avoir un indice économique sérieux qu'elles doivent être incluses dans le modèle.

3 Erreurs Non-Sphériques

Gujarati et Porter Chapitres 11 et 12. *Wooldridge* chapitres 8 et 12.

3.1 Moindres Carrés Généralisés

3.1.1 Introduction

Jusqu'à maintenant, le modèle linéaire comporte des hypothèses très strictes sur la matrice de variance covariance des erreurs du modèle. Celle-ci est diagonale et de forme $E(uu') = V = \sigma^2 I_T$. Cela implique que les erreurs du modèle sont dites homoscédastiques, c'est-à-dire que sa variance est constante dans le temps, ne dépend pas de variables exogènes et n'est pas autocorrélée. C'est une hypothèse distributionnelle très forte surtout en coupe



transversale. Les raisons pour rencontrer des erreurs non stochastiques sont nombreuses : meilleure collection des données dans le temps (l'erreur devrait donc diminuer), présence d'observations adhérentes ou d'outliers, les agents apprennent avec le temps et les déviations au modèle deviennent plus petites, mauvaise spécification du modèle, mauvaise spécification des moments supérieurs de la distribution des erreurs, etc.

Nous allons maintenant dériver les estimateurs pour des formes de matrices variance covariance des erreurs plus générales.

Soit le modèle suivant :

$$Y = X\beta + u$$

$$\begin{aligned} E(uu') &= V \\ &= \sigma^2\Omega \end{aligned}$$

Où Ω est une matrice symétrique et inversible, mais n'est pas égale à la matrice d'identité. Cette formulation permet l'hétéroscédasticité des erreurs de plusieurs formes.

- Hétéroscédasticité groupée de différentes formes,
- Corrélation sérielle des erreurs,
- Variance qui varie en fonction de variables exogènes,
- Effets ARCH, etc.

3.1.2 Implications de $\Omega \neq I$

Vérifions si l'estimateur des OLS reste sans biais et convergent lorsque la variance du terme d'erreur n'est plus homoscedastique.

Soit le modèle :

$$Y = X\beta + u \quad E(uu') = V = \sigma^2\Omega$$

D'abord, est-ce que l'estimateur OLS garde sa propriété sans biais si l'hypothèse que $\Omega = I$ n'est pas respectée ?



$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u$$

On voit que comme la matrice de variance covariance des erreurs n'intervient pas dans le calcul de l'espérance de l'estimateur OLS, il est toujours sans biais.

Ensuite, est-ce que la variance de l'estimateur OLS sera toujours la même ?

$$V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'\Omega^{-1}X)^{-1}]$$

$$V(\hat{\beta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Cet estimateur est problématique, car on ne peut pas le comparer à l'estimateur des OLS trouvé plus haut. On ne peut pas dire s'il est plus grand, plus petit et on ne peut donc pas appliquer la preuve du théorème de Gauss-Markov ni regarder la borne de Cramér-Rao.

3.2 Les Moindres Carrés Généralisés

Nous allons donc essayer de transformer le modèle pour le ramener dans un contexte où les OLS sont optimaux.

Ω inversible, \exists une matrice P telle que

$$P'P = PP' = \Omega$$

$$\Leftrightarrow \Omega^{-1} = [PP']^{-1} = [P']^{-1}P^{-1}$$

où P n'est pas aléatoire.

Considérons le modèle transformé :

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u$$



La variance des erreurs sera donc :

$$\begin{aligned}
 E[(P^{-1}u)(P^{-1}u)'] &= E(P^{-1}uu'P^{-1'}) \\
 &= P^{-1}E(uu')P^{-1'} \\
 &= P^{-1}\sigma^2\Omega P^{-1'} \\
 &= \sigma^2 P^{-1}PP'P^{-1'} \\
 &= \sigma^2 I
 \end{aligned}$$

Donc, l'estimateur OLS appliqué au modèle transformé sera optimal !

Calculons cet estimateur :

$$\begin{aligned}
 \hat{\beta}_{GLS} &= [(P^{-1}X)'(P^{-1}X)]^{-1}(P^{-1}X)'(P^{-1}Y) \\
 &= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}Y
 \end{aligned}$$

Si Ω (ou P) n'est pas connu, on le remplace par un estimé convergent $\hat{\Omega}$ (\hat{P}). L'estimateur des moindres carrés faisables ainsi obtenus sera dénoté :

$$\hat{\beta}_{FGLS} = [X'\hat{\Omega}^{-1}X]^{-1}X'\hat{\Omega}^{-1}Y$$

$$\lim \hat{\beta}_{FGLS} = \beta$$

Ce résultat demande que $\hat{\Omega}$ soit convergent. Cet estimateur n'est pas optimal au sens de Cramér-Rao, il est optimal asymptotiquement.

Note : nous n'avons pas encore rencontré la propriété de convergence. Un estimateur est convergent (consistent en anglais) si lorsque la taille de l'échantillon augmente vers l'infini, l'estimateur se concentre (converge) sur la vraie valeur du paramètre. Les conditions suffisantes pour la convergence en probabilité sont :

$$\lim_{T \rightarrow \infty} E(\hat{\beta}_T) = \beta$$



$$\lim_{T \rightarrow \infty} \text{Var}(\hat{\beta}_T) = 0$$

Alors que l'efficacité est une propriété en échantillon fini (le nombre d'observations n'influence pas la propriété), la convergence est une propriété qui fait intervenir la loi des grands nombres et le théorème de limite centrale qui sont des résultats statistiques asymptotiques. C'est pour cette raison que c'est une propriété asymptotique des estimateurs. (*voir chapitre 5 de Wooldridge*)

3.2.1 Espérance et Variance des Estimateurs GLS

Si Ω est connu et non estimé,

$$\begin{aligned}\hat{\beta}_{GLS} &= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}Y \\ &= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}(X\beta + u) \\ &= \beta + [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}u\end{aligned}$$

$$E(\hat{\beta}_{GLS}) = \beta$$

$$\begin{aligned}V(\hat{\beta}_{GLS}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}uu'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}]\end{aligned}$$

$$V(\hat{\beta}_{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Note : il faut faire attention si on compare les variances des différents estimateurs. On a :

— La variance des OLS sur le modèle standard sans hétéroscédasticité :

$$V(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$$

— La variance des OLS avec hétéroscédasticité sur le modèle non trans-



formé :

$$V(\hat{\beta}_{OLS,HE}) = \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega X(X'\Omega^{-1}X)^{-1}$$

— La variance du modèle GLS transformé par la matrice P :

$$V(\hat{\beta}_{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Les éléments sur la diagonale de $V(\hat{\beta}_{OLS,HE})$ sont plus grands que ceux de $V(\hat{\beta}_{GLS}) \Rightarrow$ par contre, ce n'est pas le même modèle qui est estimé de même que les coefficients associés sont différents. Il n'est donc pas approprié de comparer les deux séries de résultats entre eux.

Cependant, lorsqu'il y a hétéroscédasticité, la variance des OLS standards $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$ est **FAUSSE** et elle ne peut pas être comparée aux deux autres. On ne sait pas si elle est plus petite ou plus grande, elle est simplement fausse.

Donc, si on connaît Ω , il est préférable de prendre l'estimateur GLS. Si nous n'avons pas Ω et qu'il faut l'estimer et utiliser les FGLS (ce qui est presque toujours le cas), ce résultat tient, mais en limite seulement.

4 Bibliographie

Références

- [1] Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill.
- [2] Greene, W. H. (2012). *Econometric Analysis*. Pearson.
- [3] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

