Extensions au modèle linéaire simple (Séance 4)

Simon-Pierre Boucher

Département de finance, assurance et immobilier
Faculté des sciences de l'administration
Université Laval

4 février 2025

Références

Obligatoires:

- Woolridge: chapitres 3, 8, 12.

Complémentaires:

- Gujarati et Porter: chapitres 10, 11, 12, 13 et appendice C.
- Greene: chapitres 2, 3, 4, 5, 9, 14, 20 C et D

Plan de la séance

- Multicolinéarité
- Problèmes de spécification
- Erreurs non-sphériques
- Les Moindres carrés généralisés

Multicolinéarité

- 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.
- Supposons le modèle suivant :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i$$

- Multicolinéarité parfaite (ou exacte) :

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0$$

- Multicolinéarité imparfaite :

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} + v_i = 0$$

Multicolinéarité (suite)

- Dans le cas de la multicolinéarité parfaite :
  - Alors la matrice $(X^T X)$ 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.
- Dans le cas de la multicolinéarité imparfaite :
  - Les régressseurs ne sont pas parfaitement corrélés, mais fortement corrélés à un choc près.

Multicolinéarité - Conséquences

- **Les conséquences de la multicolinéarité**
  - La multicolinéarité imparfaite ne viole pas les hypothèses du théorème de Gauss-Markov.
  - Les estimateurs des MCO en cas de multicolinéarité gardent la propriété BLUE.
- Les variances et les erreurs standard des estimations des coefficients de régression vont augmenter.
  - Cela signifie des t-statistiques plus faibles.
- L'ajustement global de l'équation de régression ne sera pas affecté par la multicolinéarité.
  - Cela signifie également que la prévision et la prédiction ne seront pas affectées.
- Les coefficients de régression seront sensibles aux spécifications.
  - Les coefficients de régression peuvent changer considérablement lorsque des variables sont ajoutées ou supprimées.

Multicolinéarité - Détection

- **La détection de la multicolinéarité**
  - Coefficients de corrélation élevés :
    - Les corrélations par paires entre les variables indépendantes peuvent être élevées (en valeur absolue).
    - Règle générale : si la corrélation est ¿ 0.8, il peut y avoir une forte multicolinéarité.
  - R^2 élevé avec des valeurs de t-statistiques faibles :
    - Il est possible que les coefficients de régression individuels ne soient pas significatifs mais que l'ajustement global de l'équation soit élevé.
  - Facteurs d'inflation de la variance (VIF) élevés :
    - VIF quantifie dans quelle mesure la multicolinéarité a augmenté la variance d'un coefficient estimé.
    - Il examine dans quelle mesure une variable explicative peut être expliquée par toutes les autres variables explicatives de l'équation.

Multicolinéarité - Remèdes

- **Remèdes contre la multicolinéarité**
  - Ne rien faire
  - Abandon d'une variable redondante :
    - Si une variable est redondante, elle n'aurait jamais dû être incluse dans le modèle en premier lieu. Ainsi, l'abandonner ne fait que corriger une erreur de spécification.
    - Utilisez la théorie économique pour guider votre choix de la variable à supprimer.
  - Transformer les variables multicolinéaires :
    - Vous pouvez réduire la multicolinéarité en respécifiant le modèle, par exemple, en créant une combinaison des variables multicolinéaires.
  - Augmenter la taille de l'échantillon :
    - L'augmentation de la taille de l'échantillon améliore la précision d'un estimateur et réduit les effets négatifs de la multicolinéarité.
    - En général, l'ajout de données n'est pas possible.

Problèmes de spécification

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

Problèmes de spécification - Omission d'une variable explicative importante

- On suppose le vrai modèle suivant :

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

- Cependant, vous avez effectué l'estimation d'un modèle en omettant la variable indépendante X_2.

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

- Étant donné que X_2 n'est pas inclus dans la régression, nous allons uniquement obtenir une solution pour l'estimateur β_1 :

$$\hat{\beta}_1 = X_1^T \left(X_1^T X_1 \right)^{-1} X_1^T Y = \left(X_1^T Y \right) \left(X_1^T X_1 \right)^{-1}$$

Problèmes de spécification - Omission d'une variable explicative importante (suite)

- Afin de voir les effets de l'omission de la variable X_2, nous allons substituer dans l'équation de l'estimateur $\hat{\beta}_1$, l'équation du vrai modèle.

$$\hat{\beta}_1 = \frac{X_1^T Y}{X_1^T X_1} = \frac{X_1^T(\beta_1 X_1 + \beta_2 X_2 + u)}{X_1^T X_1}$$

- Cela donne :

$$\hat{\beta}_1 = \frac{X_1^T X_1 \beta_1 + X_1^T X_2 \beta_2 + X_1^T u}{X_1^T X_1}$$

- Ce qui nous mène à :

$$\hat{\beta}_1 = \beta_1 + \frac{X_1^T X_2}{X_1^T X_1}\beta_2 + \frac{X_1^T u}{X_1^T X_1}$$

Problèmes de spécification - Omission d'une variable explicative importante (suite)

- Nous allons maintenant prendre l'espérance de chaque côté de l'équation de $\hat{\beta}_1$:

$$E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\beta_2 X_1^T X_2}{X_1^T X_1}\right) + E\left(\frac{X_1^T u}{X_1^T X_1}\right)$$

- Sachant que $E\left(\frac{X_1^T u}{X_1^T X_1}\right) = 0$ et que $\frac{X_1^T X_2}{X_1^T X_1}$ est le coefficient de régression de X_2 sur X_1, que nous représenterons par b_{21}, cela donne :

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 b_{21}$$

- $\hat{\beta}_1$ 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.

Problèmes de spécification - Inclusion d'une variable non pertinente

- On suppose le vrai modèle suivant :

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

- Cependant, vous avez effectué l'estimation d'un modèle en ajoutant la variable indépendante X_2, qui est non pertinente au modèle :

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

- Nous avons donc les régressseurs X_1 et X_2 dans notre modèle.
- On peut représenter les deux régressseurs dans la matrice X :

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

Problèmes de spécification - Inclusion d'une variable non pertinente (suite)

- Nous avons donc les régressseurs X_1 et X_2 dans notre modèle.
- On peut représenter les deux régressseurs dans la matrice X :

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$$

- Sachant que dans le vrai modèle nous avons uniquement 1 comme régressseur, on peut écrire le régressseur du vrai modèle comme suit :

$$X \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Problèmes de spécification - Inclusion d'une variable non pertinente (suite)

- Maintenant, nous allons substituer Y dans l'équation de l'estimateur $\hat{\beta}$, par le vrai modèle.

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

- En substituant $Y = \beta_1 X_1 + u$, nous obtenons :

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

- Cela donne :

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

- Ce qui peut être réécrit comme :

$$\hat{\beta} = \beta_1 \times \left((X^T X)^{-1} X^T X_1 \right) + (X^T X)^{-1} X^T u$$

Problèmes de spécification - Inclusion d'une variable non pertinente (suite)

- On applique l'espérance de chaque côté de l'équation de l'estimateur $\hat{\beta}$:

$$E(\hat{\beta}) = \beta_1 \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (X^T X)^{-1} X^T E(u)$$

- Sachant que $E(u) = 0$, cela donne :

$$E(\hat{\beta}) = \beta_1 \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \beta_1$$

Problèmes de spécification - Inclusion d'une variable non pertinente (suite)

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

Erreurs non-sphériques

- 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.
- Raisons pour rencontrer des erreurs non-sphériques sont nombreuses :
  - Meilleure collection des données dans le temps (l'erreur devrait donc diminuer)
  - Présence d'observations aberrantes ou d'outliers
  - Mauvaise spécification du modèle

Erreurs non-sphériques - Estimateurs pour des formes de matrices variance-covariance des erreurs plus générales

- Soit le modèle suivant :

$$Y = X\beta + u$$

- Avec :

$$E(uu') = V = \sigma^2\Omega$$

- Où Ω est une matrice symétrique et inversible, mais n'est pas égale à la matrice d'identité.

Erreurs non-sphériques - Matrice Ω

- Cette formulation permet l'hétéroscédasticité des erreurs de plusieurs formes :
  - Hétéroscédasticité groupée de différentes formes.
  - Corrélation sériale des erreurs.
  - Variance qui varie en fonction de variables exogènes.
  - Effets arch.

Erreurs non-sphériques - Implication si $\Omega \neq I$

- Vérifions si l'estimateur des OLS reste sans biais et convergent lorsque la variance du terme d'erreur n'est plus homoscédastique.
- Biais de l'estimateur OLS si l'hypothèse que $\Omega = I$ n'est pas respectée.
- L'estimateur OLS est donné par :

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

- En substituant $Y = X\beta + u$, cela donne :

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

- Ce qui se simplifie à :

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

Erreurs non-sphériques - Implication si $\Omega \neq I$ (suite)

- $E(\hat{\beta}_{OLS}) = \beta + (X^T X)^{-1} X^T E(u)$
- Sachant que $E(u) = 0$, cela donne :

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

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

Erreurs non-sphériques - Implication si $\Omega \neq I$ (suite)

- **Variance de l'estimateur OLS.**

  $$V(\hat{\beta}) = E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T\right] = E\left[(X^TX)^{-1}X^T\omega\omega^TX(X^TX)^{-1}\right]$$

  $$= (X^TX)^{-1}X^TE(\omega\omega^T)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}\Omega(X^TX)^{-1}$$

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

Les Moindres Carrés Généralisés

- Transformation du modèle contenant des erreurs sphériques, pour le rendre optimal et compatible avec les hypothèses des MCO.
- Pour tout Ω inversible, il existe une matrice P telle que :

$$P^T P = PP^T = \Omega$$

- Si on inverse les deux côtés de la dernière équation :

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

où P n'est pas aléatoire.

Les Moindres Carrés Généralisés - Modèle transformé

- Modèle transformé :

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

- **Variance des erreurs :**

$$E\left[(P^{-1}u)(P^{-1}u)^T\right] = E\left[P^{-1}uu^TP^{-1}\right]$$

$$= P^{-1}E(uu^T)P^{-1} = P^{-1}\sigma^2\Omega P^{-1} = \sigma^2 P^{-1}\Omega P^{-1} = \sigma^2 I$$

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

Les Moindres Carrés Généralisés - Estimateur GLS

- Modèle transformé :
- Estimateur GLS :

$$\hat{\beta}_{GLS} = \left((P^{-1}X)(P^{-1}X)^T\right)^{-1}(P^{-1}X)^T(P^{-1}Y)$$

$$= \left[X^T P^{-1} P^{-1}\right]^{-1} X^T P^{-1} P^{-1} Y$$

$$= \left[X^T \Omega^{-1} X\right]^{-1} X^T \Omega^{-1} Y$$

Les Moindres Carrés Généralisés - Estimateur FGLS

- Si Ω (ou P) n'est pas connu, on le remplace par un estimé convergent $\hat{\Omega}(\hat{P})$.
- Estimateur FGLS :

$$\hat{\beta}_{FGLS} = \left[(\hat{\beta}^{-1}X)(P^{-1}X)^{T}(\hat{\beta}^{-1}X)(P^{-1}Y)\right]$$

$$= \left[X^{T}\hat{\Omega}^{-1}X\right]^{-1}X^{T}\hat{\Omega}^{-1}Y$$

$$= \text{PLIM}\hat{\beta}_{FGLS} = \beta$$

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

Les Moindres Carrés Généralisés - Estimateur convergent

- 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 :

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

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

- 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.

Les Moindres Carrés Généralisés - Espérance estimateur GLS

- Si Ω est connu et non estimé.
- Estimateur GLS :

$$\hat{\beta}_{GLS} = \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}Y$$

$$= \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}(X\beta + u)$$

$$= \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}X\beta + \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}u$$

$$= \beta + \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}u$$

- En espérance :

$$E(\hat{\beta}_{GLS}) = \beta + \left[X^T\Omega^{-1}X\right]^{-1}X^T\Omega^{-1}E(u)$$

$$= \beta$$

Les Moindres Carrés Généralisés - Variance estimateur GLS

- Si Ω est connu et non estimé.
- Variance de l'estimateur GLS :

$$V(\hat{\beta}_{GLS}) = E\left[(\hat{\beta}_{GLS} - \beta)(\hat{\beta}_{GLS} - \beta)^T\right]$$

$$= E\left[(X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}uu^T\Omega^{-1}X(X^T\Omega^{-1}X)^{-1}\right]$$

$$= (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}E(uu^T)\Omega^{-1}X(X^T\Omega^{-1}X)^{-1}$$

$$= \sigma^2(X^T\Omega^{-1}X)^{-1}\Omega\Omega^{-1}X(X^T\Omega^{-1}X)^{-1}$$

$$= \sigma^2(X^T\Omega^{-1}X)^{-1}$$

Les Moindres Carrés Généralisés - Variances des différents estimateurs

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

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

- La variance des OLS avec hétéroscédasticité sur le modèle non transformé :

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

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

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

Les Moindres Carrés Généralisés - Variances des différents estimateurs (suite)

- Les éléments sur la diagonale de $V(\hat{\beta}_{OLS})_{HE}$ sont plus grands que ceux de $V(\hat{\beta}_{GLS})$.
- 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.
- Lorsqu'il y a hétéroscédasticité, la variance des OLS standards $V(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ est **fausse** et elle ne peut pas être comparée aux autres.
- On ne sait pas si elle est plus petite ou plus grande, elle est simplement fausse.

Les Moindres Carrés Généralisés - Si on connaît Ω ou non

- Si on connaît Ω :
  - Il est préférable de prendre l'estimateur GLS.
- Si on ne connaît pas Ω :
  - Il est préférable de prendre l'estimateur GLS et d'estimer Ω.
  - Le résultat de l'estimateur FGLS tient, mais en limite seulement.