

Section 02 : Régression et Moindres Carrés Ordinaires (Séance 4)

GSF-6053: Économétrie Financière

Simon-Pierre Boucher¹

¹Département de finance, assurance et immobilier
Faculté des sciences de l'administration
Université Laval

1 Février 2022

Références

Obligatoires:

- ▶ **Notes de cours:** Section 2 (Professeure: Marie-Hélène Gagnon)
- ▶ **Woolridge:** chapitres 2 à 7

Complémentaires:

- ▶ **Gujarati et Porter:** chapitres 1 à 9.
- ▶ **Greene:** chapitres 2, 3, 4, 5, 9, 14, 20, appendices C et D

Plan de la séance

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Analyse de Variance

- ▶ Variation dans la variable expliquée ($\mathbf{Y}$)

$$Y_t = \hat{Y}_t + \hat{u}_t$$

- ▶ Composantes de la variation de $\mathbf{Y}$
 - ▶ $\hat{Y}_t$ est la variation due à la partie expliquée
 - ▶ $\hat{u}_t$ est induite par la partie de $\mathbf{Y}$ non expliquée

Analyse de Variance

- Variation totale de $\mathbf{Y}$

$$SS_{tot} = \sum_{t=1}^T (Y_t - \bar{Y})^2$$

- Variation partie expliquée de $\mathbf{Y}$

$$SS_{reg} = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$$

- Variation non expliquée de $\mathbf{Y}$ ou Résidus

$$SS_{err} = \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^T \hat{u}_t^2$$

Analyse de Variance

- ▶ À l'aide des 3 dernières définitions, soit SS_{tot} , SS_{reg} et SS_{err} , nous allons maintenant prouver l'équation suivante:

$$SS_{tot} = SS_{reg} + SS_{err}$$

- ▶ On pose la définition de SS_{tot} qu'on sait déjà:

$$SS_{tot} = \sum_{t=1}^T (Y_t - \bar{Y})^2$$

Sachant $Y_t = \hat{Y}_t + \hat{u}_t$

$$SS_{tot} = \sum_{t=1}^T (\hat{Y}_t + \hat{u}_t - \bar{Y})^2$$

Analyse de Variance

- Afin d'éliminer et de transformer certains termes, il nous faut distribuer les termes de notre parenthèse élevé au carré.

$$SS_{tot} = \sum_{t=1}^T (\hat{Y}_t + \hat{u}_t - \bar{Y}) \times (\hat{Y}_t + \hat{u}_t - \bar{Y})$$

$$= \sum_{t=1}^T \left[\hat{Y}_t^2 + \hat{Y}_t \hat{u}_t - \hat{Y}_t \bar{Y} + \hat{u}_t \hat{Y}_t + \hat{u}_t^2 - \hat{u}_t \bar{Y} - \bar{Y} \hat{Y}_t - \bar{Y} \hat{u}_t + \bar{Y}^2 \right]$$

Analyse de Variance

- ▶ Nous allons maintenant simplifier notre équation SS_{tot} , en utilisant l'identité remarquable.
 - ▶ Rappel: Si nous avons $(a^2 + 2ab + b^2)$, alors on peut écrire les 3 termes sous une forme polynomial $(a + b)^2$
- ▶ Dans le cas qui nous interesse, notre identité remarquable sera composé de $\hat{Y}_t$ et $\bar{Y}$. On peut donc isoler, dans notre dernière équation les termes nécessaires:

$$SS_{tot} = \sum_{t=1}^T \left[(\hat{Y}_t^2 + \bar{Y}^2 - 2\hat{Y}_t\bar{Y}) + 2\hat{Y}_t\hat{u}_t + \hat{u}_t^2 - 2\hat{u}_t\bar{Y} \right]$$

Sachant que selon l'identité remarquable

$$(\hat{Y}_t^2 + \bar{Y}^2 - 2\hat{Y}_t\bar{Y}) = (\hat{Y}_t - \bar{Y})^2$$

$$SS_{tot} = \sum_{t=1}^T \left[(\hat{Y}_t - \bar{Y})^2 + 2\hat{Y}_t\hat{u}_t + \hat{u}_t^2 - 2\hat{u}_t\bar{Y} \right]$$

Analyse de Variance

- ▶ On doit maintenant distribuer notre opérateur sommation à tous les termes de l'équation exprimant SS_{tot} .

$$SS_{tot} = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^T \hat{u}_t^2 + 2 \sum_{t=1}^T \hat{Y}_t \hat{u}_t - 2\bar{Y} \sum_{t=1}^T \hat{u}_t$$

Sachant:

- ▶ $SS_{reg} = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2$
- ▶ $SS_{err} = \sum_{t=1}^T \hat{u}_t^2$
- ▶ Alors on peut écrire l'équation exprimant SS_{tot} comme suit:

$$SS_{tot} = SS_{reg} + SS_{err} + 2 \sum_{t=1}^T \hat{Y}_t \hat{u}_t - 2\bar{Y} \sum_{t=1}^T \hat{u}_t$$

Analyse de Variance

- ▶ Les résidus sont orthogonaux au sous-espace engendré par les colonnes de X

$$2 \sum_{t=1}^T \hat{Y}_t \hat{u}_t = 0$$

- ▶ La somme des résidus est nulle

$$2\bar{Y} \sum_{t=1}^T \hat{u}_t = 0$$

- ▶ Avec les deux dernières propriétés on très bien que dans l'équation du SS_{tot} , les deux derniers termes s'annulent:

$$SS_{tot} = SS_{reg} + SS_{err}$$

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Coefficient de détermination (R^2)

- ▶ Indique le pouvoir explicatif de la régression
- ▶ Proportion des variations de Y expliquées par les régresseurs du modèle par rapport aux variations total.
- ▶ On peut donc exprimer le R^2 comme étant le ratio entre la somme des résidus au carrés et somme des variations total au carrés.

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{err}}{SS_{tot}}$$

- ▶ On voit clairement que notre R^2 sera compris entre 0 et 1

$$0 \leq R^2 \leq 1$$

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Coefficient de détermination ajusté (R^2 ajusté)

- ▶ Contrairement au R^2 standard, le R^2 ajusté vient pénalisé l'ajout de variables explicatives inutiles.
- ▶ $R^2 \rightarrow$ Si on ajoute un variable explicative à notre modèle, il augmentera forcément
- ▶ R^2 ajusté $\rightarrow$ Si on ajoute un variable explicative à notre modèle, il peut augmenter, mais si l'ajout de cette variable est inutile il pourra également diminuer.

Coefficient de détermination ajusté (R^2 ajusté)

- ▶ L'équation exprimant le R^2 ajusté contient comme pour le R^2 , le SS_{err} et le SS_{tot} .
- ▶ Cependant afin de tenir compte du nombre de régresseurs ajoutés au modèle, on aura besoin de la constante k , étant le nombre de variable indépendantes.
- ▶ Nous aurons également besoin du nombres d'observation dans notre échantillons, représenté par n .

Coefficient de détermination ajusté (R^2 ajusté)

- Équation exprimant le R^2 ajusté:

$$\begin{aligned} R_{ADJ}^2 &= 1 - \frac{\left(\frac{SS_{err}}{n-k}\right)}{\left(\frac{SS_{tot}}{n-1}\right)} \\ &= 1 - \frac{SS_{err}(n-1)}{SS_{tot}(n-k)} \end{aligned}$$

Où k représente le nombre de régresseurs

Coefficient de détermination ajusté (R^2 ajusté)

- ▶ On peut également montrer qu'il est possible d'exprimer le R^2 ajusté en fonction du R^2 standard
- ▶ Si $\frac{SS_{reg}}{SS_{tot}} = R^2$ alors $\frac{SS_{err}}{SS_{tot}} = 1 - R^2$ Comme montré ci-haut, le R^2 ajusté peut être exprimé comme suit:

$$R_{ADJ}^2 = 1 - \frac{SS_{err}}{SS_{tot}} \times \frac{(n-1)}{n-k}$$

Sachant $\frac{SS_{err}}{SS_{tot}} = 1 - R^2$, on peut écrire à nouveau le R^2 ajusté en fonction du R^2

$$R_{ADJ}^2 = 1 - (1 - R^2) \times \frac{(n-1)}{n-k}$$

Coefficient de détermination ajusté (R^2 ajusté)

Coefficient de détermination ajusté:

- ▶ Si nous avons uniquement 1 variables explicatives dans notre modèle, alors le R^2 sera égale au R^2_{ADJ} .
 - ▶ Dans ce cas, nous avons 1 régresseur, soit $k = 1$

$$\begin{aligned} R^2_{ADJ} &= 1 - (1 - R^2) \times \frac{(n - 1)}{n - k} \\ &= 1 - (1 - R^2) \times \frac{(n - 1)}{n - 1} \\ &= 1 - (1 - R^2) \\ &= R^2 \end{aligned}$$

Coefficient de détermination ajusté (R^2 ajusté)

Coefficient de détermination ajusté:

- ▶ Si nous avons uniquement plusieurs variables explicatives dans notre modèle, alors le R^2 sera plus grand ou égale au R^2_{ADJ} .
 - ▶ Dans ce cas, nous avons 2 régresseur et plus, soit $k \geq 2$
- ▶ S'il y a plusieurs variables explicatives alors

$$n - 1 > n - k$$

De façon équivalente:

$$\frac{(n - 1)}{n - k} > 1$$

- ▶ Le terme $\frac{(n-1)}{n-k}$ supérieurs à 1 est multiplié à ce qui n'est pas expliqué par le modèle ($1 - R^2$)

Ce qui nous permet d'affirmer:

$$R^2_{ADJ} \leq R^2 \text{ et } R^2_{ADJ} \leq 1$$

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Test d'hypothèse

Un test d'hypothèse sur les paramètres d'un modèle économétrique requiert les étapes suivantes:

▶ **Étape 1:**

- ▶ Écrire l'hypothèse nulle (H_0) et l'alternative (H_1).
- ▶ La plupart du temps, H_0 est l'hypothèse que l'on veut rejeter.
- ▶ Par exemple, qu'un paramètre est non significatif ou égal à une certaine valeur.
- ▶ L'hypothèse alternative peut être unilatérale ($<$ et $>$) ou bilatérale ($\neq$).

Test d'hypothèse

► Étape 2:

- Définir la statistique de test (student, F, Wald, LR, LM etc..)
- Déterminer si possible, la distribution loi de cette statistique sous H_0 .

► Étape 3:

- Choisir le niveau de significativité du test α .
- On fixe donc la possibilité d'erreur de type 1 soit la probabilité de rejeter H_0 lorsque celle-ci est vraie.
- Généralement, on fixe le niveau à 5 % ou 1 % et plus rarement à 10 %.

Test d'hypothèse

▶ Étape 4:

- ▶ Déterminer la règle de décision du test avec la région critique de confiance CR_α .
- ▶ La plupart du temps, cela demande de savoir la distribution de la statistique sous H_0 .
- ▶ Lorsque la valeur calculée pour la statistique de test se trouve dans la région critique: on rejette H_0 en faveur de H_1 au seuil de confiance α .

▶ Étape 5:

- ▶ Utiliser les valeurs obtenues de la régression pour calculer la statistique de test.

▶ Étape 6:

- ▶ Appliquer la règle de décision vue dans l'étape 4.

Test d'hypothèse

Note sur la p-value

- ▶ Parfois, on regarde la p-value au lieu de comparer CR_α à la valeur de la statistique.
- ▶ P-value : C'est le plus petit niveau auquel on peut rejeter l'hypothèse nulle.
- ▶ En d'autres mots, c'est la probabilité d'avoir un événement aussi extrême que celui observé en assumant que H_0 est vraie.
- ▶ Plus la p-value est faible, plus la probabilité que l'événement observé soit faible, étant donné l'hypothèse nulle.
 - ▶ plus il y a de chances que l'hypothèse nulle est rejetée.

Test d'hypothèse

Test de significativité:

Significativité individuelle d'un paramètre

- ▶ On regarde si le régresseur est statistiquement non nul
- ▶ Utilise le test de student

Significativité conjointe des paramètres

- ▶ On regarde si au moins un des régresseurs est statistiquement non nul (un effet)
- ▶ Utilise le F-test

Test d'hypothèse

Test de Student:

Two-tailed test

- ▶ Hypothèse nulle est la non-significativité du coefficient de régression
- ▶ Hypothèses:
 - ▶ $H_0 : \beta_k = 0 \rightarrow$ **Hypothèse nulle**
 - ▶ $H_1 : \beta_k \neq 0 \rightarrow$ **Hypothèse alternative**
- ▶ Règle de décision: Rejeter H_0 si:

$$t = \frac{|\hat{\beta}_k - \beta_0|}{SE_{\hat{\beta}_k}} > t_{n-k, \alpha/2}$$

- ▶ Ou β_0 est la valeur du coefficient sous l'hypothèse nulle, soit 0
- ▶ $SE_{\hat{\beta}_k}$ est l'écart-type associé à l'estimation de $\hat{\beta}_k$

Test d'hypothèse

Test de Student:

Right-tailed test

- ▶ Hypothèses:
 - ▶ $H_0 : \beta_k \leq 0 \rightarrow$ **Hypothèse nulle**
 - ▶ $H_1 : \beta_k > 0 \rightarrow$ **Hypothèse alternative**
- ▶ Règle de décision: Rejeter H_0 si:

$$t = \frac{\hat{\beta}_k - \beta_0}{SE_{\hat{\beta}_k}} > t_{n-k, \alpha}$$

- ▶ Ou β_0 est la valeur du coefficient sous l'hypothèse nulle, soit 0
- ▶ $SE_{\hat{\beta}_k}$ est l'écart-type associé à l'estimation de $\hat{\beta}_k$

Test d'hypothèse

Test de Student:

Left-tailed test

- ▶ Hypothèses:
 - ▶ $H_0 : \beta_k \geq 0 \rightarrow$ **Hypothèse nule**
 - ▶ $H_1 : \beta_k < 0 \rightarrow$ **Hypothèse alternative**
- ▶ Règle de décision: Rejeter H_0 si:

$$t = \frac{\hat{\beta}_k - \beta_0}{SE_{\hat{\beta}_k}} < -t_{n-k,\alpha}$$

- ▶ Ou β_0 est la valeur du coefficient sous l'hypothèse nulle, soit 0
- ▶ $SE_{\hat{\beta}_k}$ est l'écart-type associé à l'estimation de $\hat{\beta}_k$

Test d'hypothèse

F-test (significativité conjointe)

- ▶ Comparer la différence entre les résidus au carrés d'un modèle contraint et d'un modèle non contraint.
 - ▶ Si la différence est grande
 - ▶ Plus les résidus du modèle contraint sont grand par rapport au modèle contraint
 - ▶ Plus la contrainte coûte cher à appliquer en terme de performance
 - ▶ Plus la statistique F est grande
 - ▶ Passé un certain coût critique $\rightarrow$ rejette H_0

Test d'hypothèse

F-test (significativité conjointe)

Hypothèses:

- ▶ $H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$
- ▶ $H_1 : \beta_2 \neq 0 \text{ et } \beta_3 \neq 0 \text{ et } \dots \text{ et/ou } \beta_k \neq 0$

Statistique de test:

- ▶ Modèle non contraint $\rightarrow (\hat{u}'\hat{u})$
- ▶ Modèle contraint $\rightarrow (\hat{u}'_0\hat{u}_0)$

$$F = \frac{\hat{u}'_0\hat{u}_0 - \hat{u}'\hat{u}}{\hat{u}'\hat{u}} \times \frac{t - k}{q} \sim (q, t - k)$$

Test d'hypothèse

F-test (significativité conjointe)

Décision:

On rejette H_0 à un seuil de α si

$$F = \frac{\hat{u}'_0 \hat{u}_0 - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} \times \frac{t - k}{q} > F(q, t - k; \alpha)$$

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Les contraintes linéaires

Format général:

$$H_0 : R\beta = r$$

- ▶ R = matrice de sélection contenant une ligne pour chaque contrainte non redondante q
- ▶ Vecteur de réel constants

Les contraintes linéaires

- ▶ On cherche à maximiser notre log-vraisemblance afin de trouver une solution pour β , tout en respectant une contrainte linéaire.
- ▶ Il s'agit essentiellement d'une optimization sous contrainte à l'aide d'un multiplicateur de lagrange.
- ▶ On représente l'estimateur β obtenus avec la méthode des moindres carrés ordinaire est représenté par β^{OLS} .
- ▶ L'estimateur β obtenus sous une contrainte linéaire est représenté par β^{LC}

Les contraintes linéaires

- ▶ Nous voulons minimiser la somme des carrés des résidus, mais cette fois, nous posons la contrainte : $R\beta = r$
- ▶ Cela conduit à la fonction de Lagrange suivante:

$$\begin{aligned} L(\beta, \lambda) &= (Y - X\beta)'(Y - X\beta) + 2\lambda'(R\beta - r) \\ &= Y'Y - 2Y'X\beta + \beta'X'X\beta + 2\lambda'R\beta - 2\lambda'r \end{aligned}$$

- ▶ On dérive ensuite la fonction $L(\beta, \lambda)$ par rapport à β et λ

Les contraintes linéaires

Dérivé par rapport à β

$$\frac{\partial L(\beta, \lambda)}{\partial \beta} = -2X'Y + 2X'X\beta + 2\lambda'R = 0$$

$$X'X\beta + R'\lambda = X'Y$$

Dérivé par rapport à λ

$$\frac{\partial L(\beta, \lambda)}{\partial \lambda} = 2R\beta - 2r = 0$$

$$R\beta = r$$

Les contraintes linéaires

Format matricielle

- ▶ Équation 1:

$$X'X \times \beta + R' \times \lambda = X'Y$$

- ▶ Équation 2:

$$R \times \beta + 0 \times \lambda = r$$

Ce qui nous permet d'obtenir la représentation suivante:

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'Y \\ r \end{bmatrix}$$

Les contraintes linéaires

- ▶ On peut ensuite obtenir une solution pour β et λ

$$\begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ r \end{bmatrix}$$

- ▶ Pour obtenir une solution de β^{LC} , il nous suffit de trouver la solution pour β dans l'équation 1.

$$X'X\beta^{LC} + R'\lambda = X'Y$$

$$X'X\beta^{LC} = X'Y - R'\lambda$$

$$\beta^{LC} = (X'X)^{-1}X'Y - (X'X)^{-1}R'\lambda$$

Les contraintes linéaires

Estimateur sous contrainte linéaire (β^{LC})

$$\beta^{LC} = (X'X)^{-1}X'Y - (X'X)^{-1}R'\lambda$$

Estimateur OLS (β^{OLS})

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

On voit donc facilement qu'il nous est possible d'exprimer l'estimateur β^{LC} en fonction de l'estimateur β^{OLS}

$$\beta^{LC} = \beta^{OLS} - (X'X)^{-1}R'\lambda$$

Les contraintes linéaires

- ▶ On doit également trouver une solution pour λ afin de pouvoir l'incorporer dans la solution de β^{LC}
- ▶ On commence par multiplier chaque côté l'équation de la solution de β^{LC} par R

$$R\beta^{LC} = R\beta^{OLS} - R(X'X)^{-1}R'\lambda$$

- ▶ On sait déjà que $R\beta^{LC} = r$, étant donné la contrainte linéaire posée

$$r = R\beta^{OLS} - R(X'X)^{-1}R'\lambda$$

- ▶ Pour finalement isoler β^{OLS}

$$\lambda = (R(X'X)^{-1}R')^{-1}(R\beta^{OLS} - r)$$

Les contraintes linéaires

- ▶ On substitue la solution de λ dans l'équation de la solution de β^{LC}

$$\beta^{LC} = \beta^{OLS} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\beta^{OLS} - r)$$

- ▶ On voit que β^{LC} (Contraint) est exprimé en fonction de β^{OLS} (Non-Contraint).
- ▶ Le test de Fisher (F-Test) est une spécification d'une contrainte linéaire.
 - ▶ Le modèle non-contraint est le modèle que nous souhaitons estimer et vérifier la significativité.
 - ▶ Le modèle contraint est simplement un modèle dans lequel nous allons contraindre tous les coefficients d'être égaux à 0.

Les contraintes linéaires

Représentation du F-test sous le format de contrainte linéaire

- ▶ On suppose un modèle avec un intercept β_0 et une pente β_1 .
- ▶ On cherche à tester une contrainte linéaire ayant le format $R\beta = r$
- ▶ La matrice R :

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- ▶ La matrice β :

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Les contraintes linéaires

Représentation du F-test sous le format de contrainte linéaire

- La matrice r

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Ce qui nous donne la contrainte linéaire ($R\beta = r$) dans le cas du F-Test:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Les contraintes linéaires

- ▶ Nous regardons maintenant trois classes de tests qui peuvent être utilisées dans des contextes plus généraux:
 - ▶ Test d'hypothèse de restrictions non linéaires sur les paramètres
 - ▶ Test d'hypothèse sur la matrice de variance-covariance
 - ▶ Tests dans des modèles sans normalités des erreurs
- ▶ Les tests présentés sont
 - ▶ Multiplicateur de Lagrange (LM)
 - ▶ Test de Wald
 - ▶ Ratio de vraisemblance (LR)

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Test de Wald

- ▶ Le test de Wald est une forme quadratique basé sur la distance entre $(R\beta - r)$ et zéro.
- ▶ C'est-à-dire si l'estimé non contraint vérifie la contrainte.
- ▶ On rejette l'hypothèse nulle si la statistique est suffisamment grande.

Hypothèses

- ▶ $H_0 : R\beta = r$
- ▶ $H_1 : R\beta \neq r$

Statistique W

$$W = (R\beta - r)'[V(R\beta - r)]^{-1}(R\beta - r)$$

Où $V(R\beta - r)$ est la variance entre les deux termes.

Test de Wald

Statistique W

- ▶ En décomposant la variance, on trouve une solution pour la statistique de Wald.

$$W = \frac{1}{\hat{\sigma}^2} (R\beta - r)' [R(X'X)^{-1}R']^{-1} (R\beta - r)$$

- ▶ Sous l'hypothèse nulle, W suit aussi une khi carré (q).
- ▶ Le Wald est intéressant, car tous les estimés nécessaires sont non contraints.
- ▶ On ne fait pas d'estimation contrainte, car on s'intéresse à savoir si le modèle non contraint rejette l'hypothèse nulle d'être « assez prêt » de la contrainte.

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Ratio de vraisemblance (LR)

- ▶ **Intuition:** comparer la valeur de la vraisemblance aux maximum contraint et non contraint.
- ▶ Si ces deux valeurs sont proches l'une de l'autre, cela implique que la contrainte est aisément satisfaite par les observations et non couteuse en termes de maximisation de la vraisemblance.
- ▶ La vraisemblance du modèle non-contraint est représenté par $\hat{L}$, alors que la vraisemblance du modèle contraint est représenté par $\hat{L}_c$
- ▶ L'estimateur $\hat{L}$ sera une fonction des estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ du modèle non contraint $\rightarrow \hat{L}(\hat{\beta}, \hat{\sigma}^2)$
- ▶ L'estimateur $\hat{L}_c$ sera une fonction des estimateurs $\hat{\beta}_c$ et $\hat{\sigma}_c^2$ du modèle non contraint $\rightarrow \hat{L}_c(\hat{\beta}_c, \hat{\sigma}_c^2)$

Ratio de vraisemblance (LR)

- Afin de pouvoir simplifier la vraisemblance, nous poserons l'hypothèse de normalité des résidus.

$$\begin{aligned}\hat{L} &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{1}{2} \times \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\hat{\sigma}^2} \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{1}{2} \times \frac{\hat{u}'\hat{u}}{\hat{\sigma}^2} \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{1}{2} \times \frac{T\hat{\sigma}^2}{\hat{\sigma}^2} \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{T}{2}\end{aligned}$$

Ratio de vraisemblance (LR)

Vraisemblance du modèle non-contraint

$$\hat{L} = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{T}{2}$$

Vraisemblance du modèle contraint

$$\hat{L}_c = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}_c^2) - \frac{T}{2}$$

Ratio de vraisemblance (LR)

- ▶ On veut comparer la valeur de la vraisemblance $\hat{L}$ et $\hat{L}_c$
- ▶ La statistique LR s'écrit donc comme la différence entre les estimés contraints et non-contraints de σ^2

$$\begin{aligned} LR &= 2[\hat{L} - \hat{L}_c] \\ &= 2 \left[-\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}^2) - \frac{T}{2} \right] \\ &\quad - 2 \left[-\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\hat{\sigma}_c^2) - \frac{T}{2} \right] \\ &= T \log(\hat{\sigma}_c^2) - T \log(\hat{\sigma}^2) \\ &= T \log \left(\frac{\hat{\sigma}_c^2}{\hat{\sigma}^2} \right) = T \log \left(\frac{\hat{u}'_c \hat{u}_c}{\hat{u}' \hat{u}} \right) \end{aligned}$$

- ▶ La statistique suit donc une loi Khi carré (q) asymptotiquement

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Multiplicateur de Lagrange

- ▶ Il s'agit d'une régression auxiliaire.
- ▶ La statistique de test sera : $T \times R^2$
- ▶ Cette statistique suit alors une khi-carré avec q degré de liberté.
- ▶ Intuition de ce test : vérifier si le score (la dérivé première du Lagrangien en fonction des paramètres) est proche de zéro évalué en $\hat{\beta}_C$.
- ▶ Si oui, la contrainte n'est pas très couteuse en termes d'optimisation et il est probable que les paramètres prennent les valeurs définies par l'hypothèse nulle.

Multiplicateur de Lagrange

Statistique LM

$$LM = \left[\frac{\partial L}{\partial \beta}(\hat{\beta}_c) \right]' \left[V \frac{\partial L}{\partial \beta}(\hat{\beta}_c) \right]^{-1} \left[\frac{\partial L}{\partial \beta}(\hat{\beta}_c) \right]$$

En effectuant les dérivées premières et quelques manipulations, on obtient la solution pour la statistique LM.

$$LM = \frac{1}{\sigma^2} (\hat{\beta} - \hat{\beta}_c)' R' C R (\hat{\beta} - \hat{\beta}_c)$$

Multiplicateur de Lagrange

Statistique LM

- ▶ On peut aussi réécrire cette statistique comme une fonction des résidus contraints et non contraints.

$$LM = \frac{\hat{u}'_c \hat{u}_c - \hat{u}' \hat{u}}{\frac{\hat{u}'_c \hat{u}_c}{T}}$$

- ▶ Cette statistique suit une loi Chi carré avec q degré de liberté asymptotiquement sous l'hypothèse nulle.
- ▶ Le LM exploite le fait que la maximisation du log vraisemblance sous la contrainte de l'hypothèse nulle revient à maximiser sans contrainte la fonction Lagrangienne associée.
- ▶ Le test est alors basé sur le fait que si la contrainte sous H_0 est respectée par les données, le vecteur de Lagrange devrait être nul.

Analyse de Variance

Coefficient de détermination (R^2)

Coefficient de détermination ajusté (R^2 ajusté)

Test d'hypothèse

Les contraintes linéaires

Test de Wald

Ratio de vraisemblance (LR)

Multiplicateur de Lagrange

Lien entre les Statistiques F, WALD, LR et LM

Lien entre les Statistiques F, WALD, LR et LM

- ▶ Toutes ces statistiques se calculent à partir de la somme des résidus au carré.
- ▶ Il est possible d'exprimer les trois statistiques présentées comme une transformation de la statistique F dans le cas de contraintes linéaires et du modèle linéaire simple.

Statistique F

$$F = \frac{\hat{u}_0' \hat{u}_0 - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} \times \frac{t - k}{q} > F(q, t - k; \alpha)$$

- ▶ On rejette H_0 si la statistique de test est plus grande que le point critique associé

Lien entre les Statistiques F, WALD, LR et LM

Statistique Wald

$$Wald = T \frac{\hat{u}'_c \hat{u}_c - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} = \frac{Tq}{T-K} F$$

Statistique LR

$$LR = T \log \left[\frac{\hat{u}'_c \hat{u}_c}{\hat{u}' \hat{u}} \right] = T \log \left[F \frac{q}{T-K} + 1 \right]$$

Statistique LM

$$LM = \frac{\hat{u}'_c \hat{u}_c - \hat{u}' \hat{u}}{\frac{\hat{u}'_c \hat{u}_c}{T}} = \frac{T}{\left[\frac{1}{F} \right] \left[\frac{T-K}{q} \right] + 1}$$

Lien entre les Statistiques F, WALD, LR et LM

- ▶ Rejette si la statistique de test est plus grande que le point critique d'une Chi-carré avec q degrés de liberté.
- ▶ Bien que toutes ses statistiques soient maintenant exprimées en fonction des résidus contraints et non contraints, elles ne donnent pas la même valeur.
- ▶ L'inférence pourrait donc potentiellement être différente.