Faculté
des sciences économiques et de gestion
Université de Strasbourg

Année universitaire 2022/2023

Double Licence 3ème année Mathématiques - Economie et Gestion
DUAS I

Semestre 6 – Session 1 / Contrôle terminal / Mai 2023

Matière: UE Econométrie - théorie et pratique (Bertrand Koebel)
Durée: 2H00
Les notations utilisées sont celles du cours
Documents autorisés: notes de cours et de TD, codes R, calculatrice non-programmable.

Exercice 1 (3 points, d'après Hansen, 2022, ex. 4.23).
Considérer le modèle de régression linéaire tel que $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}_0$. L'estimateur de type ridge est définit par:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{y},$$

où $\lambda > 0$ dénote un réel (une constante fixée à $\lambda = 0.05$ par exemple).

1. Déterminer $E\left[\widehat{\boldsymbol{\beta}}|\mathbf{X}\right]$, est-ce que $\widehat{\boldsymbol{\beta}}$ est sans bias pour $\boldsymbol{\beta}_0$?

2. Les auteurs de Wikipedia prétendent que $\widehat{\boldsymbol{\beta}}$ peut s'obtenir en résolvant

$$\min_{\boldsymbol{\beta}}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^\top\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) + \lambda\boldsymbol{\beta}^\top\boldsymbol{\beta}.$$

Démontrer ce résultat. Quelle interprétation donner à cet estimateur de type ridge?

Exercice 2. (4 points, d'après Ruud, exercice 9.6).

1. (1 pt) Donner l'expression de l'estimateur $\widehat{\beta}_1$ et $\widehat{\beta}_2$ des moindres carrés ordinaires dans la régression partitionnée $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$.

2. (1 pt) Calculer $\operatorname{cov}\left(\widehat{\beta}_1, \widehat{\beta}_2|\mathbf{X}\right)$.

3. (2 pts) Est-il possible d'affirmer que, si $\mathrm{V}(\mathbf{y}|\mathbf{X}) = \sigma^2 I_N$ et $\mathbf{X}_1^\top\mathbf{X}_2 = 0$, alors $\operatorname{cov}\left(\widehat{\beta}_1, \widehat{\beta}_2|\mathbf{X}\right) = 0$?

Exercice 3 (4 points). Etudier le code R suivant, puis répondre aux questions

```
N = 1000
x1 = runif(N)
x2 = 2*x1 + rnorm(N)
y = x1 + x2 + rnorm(N)
results = lm( y ~ x1 )
summary( results )
```

1. (1.5 pts.) Quelles seront les valeurs numériques (approximatives) des paramètres estimés dans la ligne 5 ?

2. (2.5 pts.) A partir des équations normales, discuter de la portée et des limites de l'estimateur $\widetilde{\beta}$ calculé en ligne 5.

Exercice 4 (9 points).

W. Greene analyse la demande de gazole aux Etats-Unis, et exprime la relation suivante entre les variables:

$$y_t = x_t^\top \beta + \varepsilon_t, \qquad (1)$$

avec des termes aléatoires ε_t tels que $\varepsilon|X \sim \mathcal{N}\left(0, \sigma^2 I_T\right)$. L'indice t représente l'année: $t = 1953, \ldots, 2004$, et T le nombre d'observations: $T = 52$. La variable expliquée et les variables explicatives sont définies ci-dessous:

y_t		consommation de gazole aux USA (en log)
x_t	$x_{1t} = 1$	constante
	x_{2t}	revenu total (en log)
	x_{3t}	population (en log)
	x_{4t}	prix du gazole (en log)
	x_{5t}	prix des voitures neuves (en log)
	x_{6t}	prix des voitures d'occasion (en log)
	x_{7t}	prix des transports publics (en log)

Les résultats d'estimation de plusieurs variantes de la spécification (1) sont résumés dans le tableau suivant:

	(1)	(2)	(3a)	(3b)
$\widehat{\beta}_1$	−24.84 (0.65)	−27.92 (3.78)	−26.50 (2.11)	−24.03 (1.52)
$\widehat{\beta}_2$	1.22 (0.20)	0.48 (0.07)	0.84 (0.17)	0.43 (0.10)
$\widehat{\beta}_3$	1.19 (0.36)	2.08 (0.31)	1.36 (0.13)	1.65 (0.18)
$\widehat{\beta}_4$	−0.02 (0.04)	0.02 (0.05)	0.02 (0.17)	−0.10 (0.02)
$\widehat{\beta}_5$	−0.16 (0.21)	−0.73 (0.17)	0.57 (0.20)	0.14 (0.07)
$\widehat{\beta}_6$	0.04 (0.11)	0.22 (0.11)	−0.25 (0.07)	0.06 (0.04)
$\widehat{\beta}_7$	−0.20 (0.13)	0.01 (0.13)	0.11 (0.09)	−0.21 (0.05)
N	52	52	21	31
SSR	0.114	0.153	0.002	0.004

Les erreurs standards sont indiquées entre parenthèses.

Le point est utilisé comme séparateur décimal.

1. (1 pt.) Interpréter avec soin 3 paramètres de votre choix, estimés par les MCO et reportés dans la colonne (1) du tableau
2. (1 pt.) Réaliser les tests d'hypothèse suivants pour le modèle estimé en colonne (1) du tableau
   (a) $H_0 : \beta_2 = 0$
   (b) $H_0 : \beta_2 = 1$
3. (2 pts.)
   (a) Encadrer les valeurs numériques estimées de $cov\left(\widehat{\beta}_2, \widehat{\beta}_3 | \mathbf{X}\right)$.
   (b) Expliquer comment tester l'hypothèse selon laquelle le niveau absolu de la population x_{3t} (en log) n'est pas fondamental, mais c'est le log du revenu par tête $x_{2t} - x_{3t}$ qui est pertinent pour expliquer le log de la consommation par tête $y_t - x_{3t}$. Cette hypothèse vous parait-elle satisfaite au vu des estimations de la colonne (1)? Réaliser le test correspondant, en allant aussi loin que vous pouvez.
4. (3 pts.) L'hypothèse microéconomique d'homogénéité de la demande de gazole de degré 0 en revenu et prix ont été imposées dans la régression (2).
   (a) Expliquer en quoi consiste cette restriction et ses conséquences pour l'estimation
   (b) Expliquer comment imposer cette restriction
   (c) Réaliser le test de la validité de cette restriction.
5. (2 pts.) Deux chocs pétrolier ont marqué les années 1974 et 1979, ces chocs ont peut-être fait changer les habitudes de consommation des ménages entre les périodes 1953-1973 et 1974-2004. Le ministère de l'énergie effectue une régression sur chacune des sous-périodes. Les résultats sont reportés en colonnes $(3a)$ et $(3b)$. La notation $\beta^{(j)}$ représente le vecteur des vraies valeurs des paramètres pour la période (j), avec $j = 1, 2$.
   (a) Effectuer le test d'hypothèse d'égalité de l'elasticité du gazole par rapport au revenu entre les périodes: $H_0 : \beta_2^{(1)} = \beta_2^{(2)}$, en expliquant votre démarche.
   (b) Effectuer le test d'hypothèse d'égalité de tous les paramètres entre les deux périodes, càd $H_0 : \beta^{(1)} = \beta^{(2)}$, en expliquant votre démarche.