CORRIGES des EXERCICES du CHAPITRE II

1 Corrigé de l'exercice II-1

1) La matrice ${}^tXX$ est la matrice diagonale suivante car les produits scalaires sont nuls

$$\begin{pmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

son inverse

$$({}^tXX)^{-1} = \begin{pmatrix} 1/7 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/10 \end{pmatrix}$$

$${}^tX\overrightarrow{y} = \begin{pmatrix} 35 \\ 10 \\ 20 \end{pmatrix} \Rightarrow \overrightarrow{\widehat{a}} = \begin{pmatrix} 1/7 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/10 \end{pmatrix} \begin{pmatrix} 35 \\ 10 \\ 20 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 2 \end{pmatrix}$$

2)

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (y_t - \overline{y})^2}$$

or $\sum (y_t - \overline{y})^2 = \sum y_t^2 - n\overline{y}^2 = 740 - 10(\frac{20}{10})^2 = 700$

$$\sum e_t^2 = (1 - R^2) \sum (y_t - \overline{y})^2 = 0.01x700 = 7$$

Comme le modèle a k=3 variables explicatives

$$s^2 = \frac{\sum e_t^2}{(n-k)} = \frac{7}{10-3} = 1$$

La matrice de variance-covariance de $\overrightarrow{\widehat{a}}$ est $\sigma^2({}^tXX)^{-1}$ estimé par $s^2({}^tXX)^{-1}$ qui est ici égal à $({}^tXX)^{-1}$ car $s^2 = 1$

On a donc $\mathrm{var}(\widehat{a_1}) = 1/7$, $\mathrm{var}(\widehat{a_2}) = 1/5$, $\mathrm{var}(\widehat{a_3}) = 1/10$ et les covariances entre les coefficients nulles.

Tout cela lorsque les hypothèses sur les erreurs sont vérifiées. En cas de normalité des erreurs, les covariances nulles entrainent l'indépendance des lois des coefficients, $\widehat{a_1}$ par exemple suit indépendamment des autres coefficients, une loi $\mathrm{N}(\mathrm{a}_1, 1/7)$

3) Dans le modèle avec seulement deux variables explicatives la nouvelle matrice s'écrit

$${}^tXX = \begin{pmatrix} 7 & 0 \\ 0 & 10 \end{pmatrix} \qquad ({}^tXX)^{-1} = \begin{pmatrix} 1/7 & 0 \\ 0 & 1/10 \end{pmatrix}$$

$${}^tX\overrightarrow{y} = \begin{pmatrix} 35 \\ 20 \end{pmatrix} \quad \Rightarrow \overrightarrow{\widehat{b}} = \begin{pmatrix} 1/7 & 0 \\ 0 & 1/10 \end{pmatrix} \begin{pmatrix} 35 \\ 20 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

On obtient les mêmes résultas que pour $\widehat{a_1}$ et $\widehat{a_3}$ parce que les produits scalaires des vecteurs étant nuls, les vecteurs des variables explicatives sont orthogonaux, et les estimateurs indépendants.

2 Corrigé de l'exercice II-2

1) Nous ne pouvons comparer ces deux modèles avec leur s car la variable endogène n'est pas la même , dans l'équation 1 la variable est IMP et dans l'équation 2 la variable est le LOG de IMP. Nous allons utiliser les résultats de la transformation de BOX-COX (voir section II de Choix de modèles): on prend le minimum entre le s_0 du modèle linéaire et le $s_1 e^{\overline{LIMP}}$ du modèle en LOG ($\overline{LIMP}$ étant la moyenne de la variable LIMP)
$s_0 = 7706.13$ $\qquad$ $s_1 e^{\overline{LIMP}} = 0.044 \exp(11.6634) = 5114.5$
$s_1 e^{\overline{LIMP}} < s_0$ on préfère donc le modèle en LOG au modèle linéaire.
Remarquons que pour faire cette comparaison il faut les mêmes variables et que ces variables soient toutes positives, sinon les deux modèles ne sont pas comparables.

2) Les modèle 3 a moins de variables, on a oublié les variables LDI et LTAXE et pourtant les deux s sont égaux (à 3 chiffres après la virgule). cela signifie que l'on a retiré deux variables qui n'étaient pas intéressantes dans le modèle. Le modèle 3 est donc plus intéressant que le modèle 2.
Nous verrons dans le chapitre suivant que les " t de student" des coefficients des variables non intéressantes (le s chiffres entre parenthèses sous les coefficients) montrent qu'en effet ces variables peuvent être supprimées car leur coefficient est significativement nul.
Comparaison des modèles 3 et 4:
Les deux modèles sont en LOG on peut donc les comparer, ils diffèrent d'une seule variable qui est une variable muette. Il s'est passé au troisième trimestre de 1982 un phénomène qui a créé un résidu très fort et qui ne s'est pas poursuivi après. (voir partie sur les résidus). Le fait de mettre cette variable muette a fait baissé le s qui est passé de 0.044 à 0.413, ce modèle 4 est donc le meilleur modèle des quatre modèles proposés. Cela ne veut pas dire qu'il n'existe pas de modèle meilleur que celui-ci.