Examen TISD - Master 1 Pro
Vendredi 14 Janvier 2011

Durée 2 h. Calculatrices autorisées. Notes de cours autorisées.
Tran Viet Chi, `chi.tran@univ-lille1.fr`, bureau 316 (bâtiment M3).

Exercice 1 (Modèle de marché, 4 points)
On considère le modèle simple suivant pour le rendement d'une action :

$$R_t = \alpha + \beta RM_t + \varepsilon_t \tag{1}$$

où R_t est le rendement de l'action au temps t, RM_t est le rendement du portefeuille de marché (un indicateur de santé de l'économie tel que le CAC40 pour la France). Si $\beta > 1$, les variations du portefeuille de marché sont amplifiées par l'action. Elle est considérée comme agressive. Si $\beta < 1$, l'action est défensive. On considère l'action d'Elf. Entre janvier 1990 et juillet 1999, nous avons 497 dates.

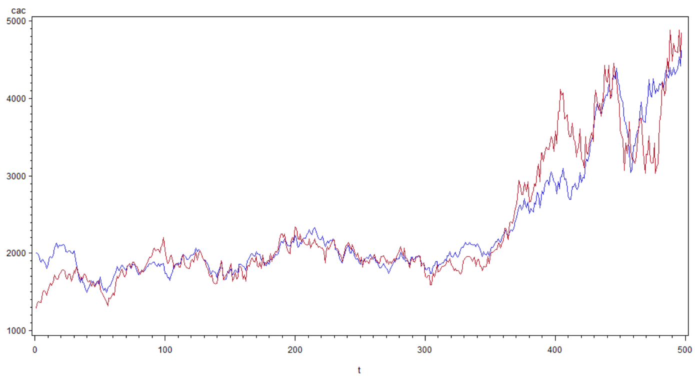

FIGURE 1 – *Evolution hebdomadaire du CAC40 et de l'action d'Elf (renormalisée par le facteur multiplicatif 2325/70) entre janvier 1990 et juillet 1999.*

1. Dans l'annexe, pour la régression 1, des informations ont été perdues. Recalculer le coefficient de détermination, les statistiques de Student pour le test de significativité des coefficients, la borne supérieure de l'intervalle de confiance pour le coefficient β.

2. Faire un test (à 5%) pour dire s'il y a une rupture de tendance à la date 300. Pour ce faire, on utilisera les régressions de l'annexe. On donne (en langage **R**) : `qf(0.95,2,493)=3.01`, `qf(0.975,298,195)=1.30`, `qf(0.025,298,195)=0.78`.

Exercice 2 (Vente de micro-ordinateurs)
On étudie l'évolution du volume des ventes de micro-ordinateurs q (en milliers) en fonction de l'indice des prix p, entre le 3ème trimestre 2002 et le second trimestre 2007. Cet indice p vaut 1 au second semestre 2002. On a $n = 20$ données, pour les 20 trimestres successifs d'observation : $(q_t, p_t)_{t\in\{1,\ldots,20\}}$. On s'intéresse au modèle économique suivant

$$q_t = b\ p_t^a. \tag{2}$$

Notations : dans la suite du problème, pour deux séries d'observations $(x_t)_{t\in\{1,\ldots,20\}}$ et $(y_t)_{t\in\{1,\ldots,20\}}$, on notera :

$$\bar{x} = \frac{1}{20}\sum_{t=1}^{20} x_t, \qquad \mathrm{Var}(x) = \frac{1}{20}\sum_{t=1}^{20}(x_t-\bar{x})^2, \qquad \mathrm{Cov}(x,y) = \frac{1}{20}\sum_{t=1}^{20}(x_t-\bar{x})(y_t-\bar{y}),$$

pour les moyennes, variances et covariances empiriques (ces deux dernières étant renormalisées par n).

1. Ecrire (2) sous forme d'une relation linéaire. Quelle est la variable expliquée ? la variable explicative ?

2. Sous **R**, les vecteurs $(q_t)_{t\in\{1,\ldots,20\}}$ et $(p_t)_{t\in\{1,\ldots,20\}}$ sont stockés dans les vecteurs `q` et `p`. On donne :

```
> sum(log(p))
[1] -5.171312
> sum(log(p)^2)
[1] 1.930369
> sum(log(q))
[1] 37.87757
> sum(log(q)^2)
[1] 72.49429
> sum(log(p)*log(q))
[1] -10.45751
```

Calculer $\overline{\log(p)}$, $\overline{\log(q)}$, $\mathrm{Var}(\log(p))$, $\mathrm{Var}(log(q))$, $\mathrm{Cov}(log(p), log(q))$.

3. En s'appuyant sur **1**, rappeler les expressions théoriques des estimateurs MCO des paramètres a et $\log(b)$. Faire l'application numérique avec les résultats de la question **2**.

On définit les résidus estimés par $(\widehat{\varepsilon}_t)_{t\in\{1,\ldots,20\}}$ où $\widehat{\varepsilon}_t = \log(q_t) - \log(\widehat{q}_t)$, où $\log(\widehat{q}_t) = \widehat{a}\log(p_t) + \widehat{\log(b)}$ est la prédiction de la variable expliquée connaissant les explicatives.

4. Justifier que $\overline{\log(\widehat{q})} = \overline{\log(q)}$.

5. En déduire que $\mathrm{Var}(\log(\widehat{q})) = \widehat{a}^2\mathrm{Var}(\log(p))$.

6. Exprimer $\widehat{\varepsilon}_t$ en fonction de $\log(q_t)$, $\overline{\log(q)}$, $\log(p_t)$, $\overline{\log(p)}$ et $\widehat{a}$.

7. En déduire que $\mathrm{Var}(\widehat{\varepsilon}) = \mathrm{Var}(\log(q)) - \mathrm{Var}(\log(\widehat{q}))$. En quoi ceci est-il une décomposition de la variance ?

8. Exprimer le coefficient de détermination R^2 en fonction de $\widehat{a}$, $\mathrm{Var}(log(q))$ et $\mathrm{Var}(log(p))$. Faire l'application numérique et commenter.

9. On désigne par $V_t = p_t q_t$ le chiffre d'affaire à la date t. Montrer que la relation linéaire établie à la question **1** est équivalente à :

$$\log(V) = c\log(p) + d, \tag{3}$$

où on exprimera c et d en fonction de a et b.

10. Exprimer (en justifiant) les estimateurs MCO $\widehat{c}$ et $\widehat{d}$ en fonction de $\widehat{a}$ et $\widehat{b}$.

11. En déduire que les deux régressions linéaires ont les mêmes résidus estimés.

12. Montrer que $\mathrm{Var}(\log(p) + \log(q)) = \mathrm{Var}(\log(q)) + (1 + 2\widehat{a})\mathrm{Var}(\log(p))$.

13. En déduire le coefficient de détermination de (3). Faire l'application numérique.

14. Comparer les résultats des questions **8** et **13**. Pourquoi peut-on dire qu'on ne peut pas apprécier la pertinence d'une modélisation en se basant sur la seule valeur d'un coefficient de détermination d'une régression ?

Exercice 3 (Régression à variables expliquée et explicative binaires)
Nous considérons un couple de variables aléatoire (Y, X) où X est une variable explicative prenant les valeurs 0 ou 1, et où les Y est une variable à expliquer, prenant elle aussi les valeurs 0 et 1, et pour laquelle on suppose qu'il existe une variable latente Y^* telle que

- conditionnellement à X, Y^* suit une loi $\mathcal{N}(aX + b, 1)$ avec a, $b \in \mathbb{R}$,
- $Y = \mathbf{1}_{Y^*>0}$.

1. Quelles sont les lois de Y et Y^* conditionnellement à X ? Montrer que $\mathbb{P}(Y = 1 \,|\, X = 1) = \Phi(a + b)$ où Φ, la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$. Calculer les probabilités $\mathbb{P}(Y = 1 \,|\, X = 0)$, $\mathbb{P}(Y = 0 \,|\, X = 1)$ et $\mathbb{P}(Y = 0 \,|\, X = 0)$.

Nous disposons d'un échantillon de 100 observations, $(y_i, x_i)_{i\in[\![1,100]\!]}$ réalisations des 100 couples $(Y_i, X_i)_{i\in[\![1,100]\!]}$ indépendants deux à deux et de même loi que (Y, X) étudié à la question 2.1. Ces observations sont résumées dans le Tableau 1.

		Y	
		0	1
X	0	$n_{00} = 46$	$n_{01} = 50$
	1	$n_{10} = 57$	$n_{11} = 63$

Table 1 – Tableau de contingence de X et Y

2. Ecrire la vraisemblance des observations $(Y_1, \cdots Y_{100})$ conditionnellement à $(X_1, \cdots X_{100})$ en utilisant les nouvelles variables $\alpha = \Phi(a+b)$ et $\beta = \Phi(b)$ ainsi que les effectifs n_{00}, n_{01}, n_{10} et n_{11} définis dans le Tableau 1.

3. Donner la log-vraisemblance. Puis, écrire les conditions du premier ordre du programme de maximisation de la log-vraisemblance par rapport à α et β.

4. En déduire les estimateurs du maximum de vraisemblance $\widehat{\alpha}$ et $\widehat{\beta}$ pour α et β.

5. En utilisant que $\Phi^{-1}(63/120) = 0,0627$ et $\Phi^{-1}(50/96) = 0,052$, déduire les valeurs numériques des EMV $\widehat{a}$ et $\widehat{b}$ pour les observations décrites au Tableau 1.