Econométrie et méthodes empiriques en finance

1 Généralités

L'examen se divise en **14/15 questions valant chacune un (1) point**. Les questions Vrai/Faux (4/5) ne nécessitent pas de justification et rapportent (1) point en cas de réponse correcte, (-0.5) point en cas de réponse **incorrecte** et (0) point en cas d'**absence** de réponse. Pour les questions ouvertes (10/11), une réponse non justifiée, peu claire ou imprécise joue en votre défaveur.

La durée de l'examen est de **une heure trente minutes (1h30)**.

N'oubliez pas d'indiquer votre **nom** et **prénom** sur chaque feuille.

Toute documentation et calculatrice sont interdits.

2 Questions Vrai/Faux

Question 1

Un gérant de fortune applique une méthode incrémentale pour calculer la Value at Risk (VaR) de son portefeuille P, prenant des allocations $a = (a_1, .., a_n)$ sur les actifs $X_1, .., X_n$:

$$VaR\,(a, 5\%) = \sum_{i=1}^{n} a_i \frac{\partial VaR\,(a, 5\%)}{\partial a_i} \quad .$$

Il souhaite former un nouveau portefeuille P' en ajoutant une quantité a_{n+1} d'un nouvel actif X_{n+1} dont il connaît la VaR individuelle, notée $VaR\,(a_{n+1}, 5\%)$.
Proposition: si le gérant ne change pas ses allocations $(a_1, .., a_n)$ dans les n premiers actifs et conserve sa méthode de calcul incrémentale, alors la VaR de son nouveau portefeuille est donnée par

$$VaR\,(a, 5\%) + a_{n+1} VaR\,(a_{n+1}, 5\%) \quad .$$

- Vrai
- Faux

Answer: False. In general, VaR is not subadditive, so the formula above may not hold in some cases.

Question 2

Soit $(x_1, .., x_T)$ des observations de la variable aléatoire X. On définit la fonction $K(u)$ sur l'ensemble des réels par

$$K(u) = \begin{cases} \frac{4}{3}u + \frac{4}{3} \text{ si } -1 \leq u < -\frac{1}{2} & , \\ \frac{2}{3} \text{ si } -\frac{1}{2} \leq u \leq \frac{1}{2} & , \\ -\frac{4}{3}u + \frac{4}{3} \text{ si } \frac{1}{2} < u \leq 1 & , \\ 0 \quad \text{sinon} & . \end{cases}$$

Proposition: l'estimateur $\widehat{f}(x) = \frac{1}{Th} . \sum_{t=1}^{T} K\left(\frac{x_t - x}{h}\right)$ est un estimateur par noyau de la densité de la distribution de X.

- Vrai
- Faux

Answer: True. You have to check if (i) K is greater than zero and (ii) if K integrates to 1. In this case, these two conditions hold. In addition, this kernel is also symmetric, so it could not be used with bounded data.

Question 3

Soit $(x_1, \ldots, x_T)$ et $(y_1, \ldots, y_T)$ des observations des variables aléatoires X et Y. On définit $\widehat{\sigma}_{XY}$ et $\overline{\sigma}_{XY}$ par

$$\widehat{\sigma}_{XY} = \frac{1}{T}\sum_{t=1}^{T}(x_t - \widehat{m}_X)(y_t - \widehat{m}_Y) \quad et \quad \overline{\sigma}_{XY} = \frac{1}{T-1}\sum_{t=1}^{T} x_t y_t - \frac{T}{T-1}\widehat{m}_X \widehat{m}_Y$$

$$où \quad \widehat{m}_X = \frac{1}{T}\sum_{t=1}^{T} x_t \quad et \quad \widehat{m}_Y = \frac{1}{T}\sum_{t=1}^{T} y_t \quad .$$

Proposition: l'estimateur $\widehat{\sigma}_{XY}$ est un estimateur asymptotiquement sans biais de la covariance entre X et Y, et $\overline{\sigma}_{XY}$ est un estimateur sans biais de cette même covariance.

- Vrai
- Faux

Answer: True. For $\widehat{\sigma}_{XY}$, we take the expectation of $\widehat{\sigma}_{XY}$ and a procedure similar to that performed in the Probability/Statistics Review shows that the estimator is biased, but as $T \to \infty$ the bias disappears. A similar procedure shows that $\overline{\sigma}_{XY}$ is unbiased.

Question 4

Soit $(x_1, .., x_T)$ les rendements de l'actif X observés sur une période de longueur T. Ces rendements sont de loi normale $N\left(\mu, \sigma^2\right)$ et indépendants.
Proposition: la Value at Risk (VaR) de l'actif X ne peut pas être calculée par la méthode des valeurs extrêmes car la queue de la distribution de X suit une loi normale.

- Vrai
- Faux

Answer: False. You can use extreme value theory with any tail distribution.

Question 5

Soit X une variable aléatoire normale centrée et de variance σ^2. Soit $Y = X^2$.

Proposition: Les variables X et Y sont corrélées.

- Vrai
- Faux

Answer: False. They are dependent, but not correlated. Correlation is a linear measure of dependence.

Question 6

Proposition: Dans le cas d'une régression linéaire, contrairement au test de Student, le test de Fisher permet de tester des relations non linéaires sur les coefficients de la régression.

- Vrai
- Faux

Answer: True. The Fisher test tests for white noise, regardless of whether the model is linear or not.

Question 7

Proposition: Si l'on utilise un noyau gaussien avec une largeur de fenêtre h, l'estimateur non paramétrique de la fonction de densité f d'une variable aléatoire Y à partir des réalisations $(Y_1, \ldots, Y_T)$ s'écrit:

$$\widehat{f}(y) = \frac{1}{T} \sum_{t=1}^{T} \Phi\left(\frac{Y_t - y}{h}\right),$$

où Φ est la fonction de répartition associée à la loi normale.

- Vrai
- Faux

Answer: False, Φ should be the probability density function (PDF, fonction de densité de probabilité) and NOT the cumulative density function (CDF, fonction de répartition).

Question 8

Proposition: Soit ε_t un bruit blanc. Le processus y_t suivant est un ARMA(1,2).

$$y_t - 0.2y_{t-2} = \varepsilon_t + 0.5\varepsilon_{t-1}$$

- Vrai
- Faux

Answer: False. It is an ARMA (2,1), by definition.

Question 9

Proposition: Pour tester la nulle hypothèse qui un processus AR(1) est intégré, on peut utiliser le test Student.

- Vrai
- Faux

Answer: False. See section and TP on non-stationary processes.

Question 10

Proposition: Le Théorème central limite est valable seulement pour des variables aléatoires suivant une loi normale.

- Vrai
- Faux

Answer: False. Only requirement is finite first and second order moments of the random variable.

3 Questions Ouvertes

Question 1

Soit $(\varepsilon_0, \varepsilon_1, .., \varepsilon_t, ..)$ des variables aléatoires indépendantes de loi normale centrée et de variance σ^2. Pour chacun des processus définis par les équations suivantes, précisez, en l'expliquant, si le processus est stationnaire au second ordre.

1. $y_t = \varepsilon_t + \varepsilon_{t-1}$.
2. $y_t = y_{t-1} + \varepsilon_t$, avec $y_0 = 0$.
3. $y_t = 0.6 + 0.2\, y_{t-1} + \varepsilon_t$, avec $y_0 = 3/4$.
4. $y_t = \sqrt{1-t^2}\, \varepsilon_t + t\, \varepsilon_{t-1}$.

Answer: Check the conditions for second-order stationarity. Processes 2 and 4 are not second-order stationary. The other two are second-order stationary. Need to check that Expectation and autocovariance of the processes are independent of t to test second-order stationarity.

Answer for process (3): The unconditional mean of y_t is $\frac{0.6}{1-0.2}$, since $E(y_t) = 0.6 + 0.2E(y_{t-1}) + 0$ and $E(y_t) = E(y_{t-1})$. As for the unconditional variance, we have $Var(y_t) = 0 + (0.2)^2 Var(y_{t-1}) + \sigma^2$ and since $Var(y_t) = Var(y_{t-1})$, we get $Var(y_t) = \frac{\sigma^2}{1-(0.2)^2}$. NOte that $E(y_t) = E(y_{t-1})$ and $Var(y_t) = Var(y_{t-1})$ because we are talking about unconditional properties. It is easy to check that the autocovariance is independent of t.

Question 2

A partir des variables X et Y, on considère la régression non-paramétrique

$$Y = m(X) + \varepsilon \quad .$$

On dispose des observations $(Y_1, .., Y_T)$ et $(X_1, .., X_T)$. Proposez en l'expliquant, un estimateur $\widehat{m}$ de la fonction m.

Answer: This is the Kernel Estimator of the conditional mean explained in Lecture A.IX.

Question 3

Dans le cadre de petits échantillons, veuillez donner une méthode d'évaluation de la précision d'un estimateur à distance finie, par opposition à une méthode fondée sur une approximation asymptotique. Expliquez brièvement le fonctionnement de cette méthode.

Answer: See Bootstrap lecture A.XIII

Question 4

Pour une série d'observations $(y_1, .., y_T)$, la statistique de Ljung-Box (Portmanteau) pour les 6 premiers retards est de 24.38. La valeur critique du test au seuil de 5% est de 32.76. Quelle est l'interprétation de ce test de Ljung-Box et que veut-il dire pour la série considérée?

Answer: Since $24.36 < 32.76$, the null of no autocorrelation is NOT rejected, so our model is captures the autocorrelation structure of the dataset with 5% confidence.

Question 5

Soit le modèle AR(1)-ARCH(1) définit par

$$\begin{aligned} y_t &= \varphi y_{t-1} + \varepsilon_t \quad , \\ \varepsilon_t &= \left(\gamma + \alpha \varepsilon_{t-1}^2\right)^{1/2} \eta_t \quad , \end{aligned}$$

où η_t est un bruit blanc gaussien de moyenne nulle et variance unitaire, et indépendant de $\underline{\varepsilon_{t-1}} = (\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$. On suppose que le processus joint $\begin{pmatrix} y_t \\ \varepsilon_t \end{pmatrix}$ est strictement stationnaire.

1. Quelle est la variance conditionnelle σ_t^2 de ε_t sachant $\underline{\varepsilon_{t-1}}$? Déduisez-en la variance inconditionnelle de ε_t.
2. Donnez les expressions de la moyenne conditionnelle m_t de y_t sachant $\underline{y_{t-1}}$, et la variance conditionnelle w_t^2 de y_t sachant $\underline{y_{t-1}}$.
3. Quelle est la variance inconditionnelle w^2 de y_t ?

Answer: 1. Conditional variance is just $\gamma + \alpha \varepsilon_{t-1}^2$. The unconditional variance can be computed from the conditional variance using the law of iterated variance (see Probability/Statistics Review)
2. Conditional mean is φy_{t-1}, conditional variance is $\gamma + \alpha \varepsilon_{t-1}^2$.
3. Using the results from part 2, use again the law of iterated variance to get the result.

Question 6

Les résultats de l'estimation de la régression linéaire ordinaire (OLS) des excès de rendement $(Y_t - r_f)$ du Fonds Fidelity, sur les excès de rendement $(Y_{m,t} - r_f)$ de l'indice S&P500, r_f représentant le taux sans risque valable sur la période,

$$Y_t - r_f = \alpha + \beta\,(Y_{m,t} - r_f) + \epsilon_t, \quad t = 1, .., T \tag{1}$$

sont les suivants:

Variable	Coefficient	Std. Error	t-Statistic	p-value
Intercept	$\hat{\alpha}$ =0.249329	0.082338	3.028125	0.0027
Exc. Ret.	$\hat{\beta}$ =0.649295	0.018719	34.68690	0.0000

1. A quel modèle théorique correspond cette équation empirique?
2. Quelle(s) conclusion(s) peut-on donner sur la validité de ce modèle?
3. Soit $\widetilde{\beta} = \frac{\widehat{cov}(Y_t, Y_{m,t})}{\widehat{var}(Y_{m,t})}$. Commentez la valeur prise par $\hat{\beta}$. Les estimateurs $\widetilde{\beta}$ et $\hat{\beta}$ sont-ils différents?
4. On met à votre disposition 99 autres séries de rendements de différents fonds, observés au cours de la même période. Proposez un test plus complet du modèle théorique considéré.

Answer: 1. CAPM 2. α seems to be significant at the 1% level, so CAPM is rejected.
3. $\widetilde{\beta}$ and $\hat{\beta}$ are different in general, unless r_f is constant.
4. See TP 4 for additional tests of the CAPM (Fama MacBeth Procedure, for example).

Question 7

Soit $(y_1, .., y_T)$ une série d'observations de la différence (spread) entre le taux long moyen des obligations d'entreprises américaines notées BBB et le taux des bons du trésor américain à 10 ans. Les observations ont été enregistrées mensuellement. Soit $(x_1, .., x_T)$ le pourcentage de changement dans l'indice américain de la production industrielle. L'estimation de la régression linéaire ordinaire (OLS) donnée par

$$y_t = a + bx_t + \varepsilon_t \quad , \tag{2}$$

où les ε_t sont indépendants et de loi normale, a montré que la production industrielle est un élément explicatif clé des spreads observés.

1. Donnez deux statistiques de test qui ont pu être utiliséee pour arriver à cette conclusion. Précisez la manière de les calculer ainsi que la manière de les interpréter.

2. Soit $\widehat{a}$ et $\widehat{b}$ les estimateurs (OLS) de a et b. Quelle(s) méthode(s) pourriez-vous utiliser pour vérifier la présence d'effets ARCH qui ne sont pas pris en compte dans l'équation (2)?

Answer: 1. F-test for overall fit and t-statistic of b for testing the null that b is not zero.
2. ARCH tests can be tested with a regression on squared fitted residuals, as stated in B.VI. Fitted residuals can be easily computed from $\widehat{a}$ and $\widehat{b}$.

Question 8

On conserve les spécifications de la question 7.

1. Ecrivez les équations et hypothèses complètes du modèle où l'équation (2) serait enrichie d'une spécification GARCH(2,1) en conservant l'hypothèse de normalité conditionnelle.

2. Soit $\widehat{\varepsilon}_t$ les résidus estimés par OLS à partir de l'équation (2). Les coefficients d'auto-corrélation ainsi que ceux d'autocorrélation partielle de la variable $\widehat{\varepsilon}_t^2$ sont significatifs pour les retards 1, 2 et 3. Les retards supérieurs ne sont pas significatifs au seuil de 95%. Quels modèles ARCH et/ou GARCH devriez-vous essayer d'implémenter? Expliquez pourquoi.

Answer: 1. Equation for ε_t^2 following the definition of GARCH process, with normal conditional disturbances.
2. Can start with GARCH(0,3). If this does not work, may try GARCH(n,3) with $n = 1, 2, 3)$ as done in the TP.

Question 9

Les deux graphes suivants tracent les coefficients d'autocorrélation et d'autocorrélation partielle d'une série simulée, ainsi que l'intervalle de confiance à 95% sous l'hypothèse nulle.

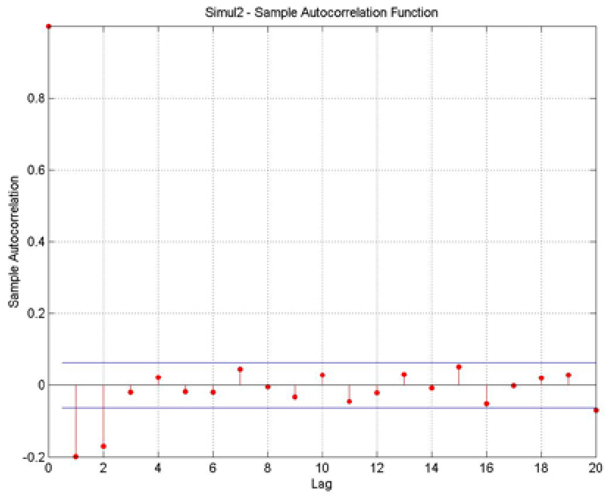

Figure 1: Autocorrélogramme de la série

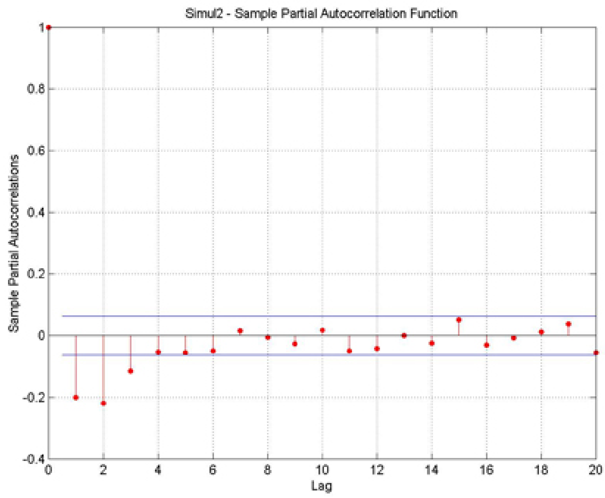

Figure 2: Autocorrélogramme Partiel de la série

Quel peut être un modèle utilisé pour simuler ces données? Donnez son équation théorique ainsi que toutes ses hypothèses.

Answer: ARMA(3,2), but other specifications such as ARMA(2,2) could work as well because the third lag in PACF is close to the confidence interval.

Question 10

On considère une série de données $Y = (Y_1, \ldots, Y_T)$ de longueur T. Le tableau suivant présente les estimations par maximum de vraisemblance des coefficients de trois modèles économétriques, ainsi que la statistique du test ARCH appliqué aux résidus des trois modèles (les p-valeurs sont indiquées entre crochets).

	AR(2)	AR(2)-ARCH(1)	AR(2)-GARCH(1,1)
$\widehat{\mu}$	0.531 [0.001]	0.542 [0.001]	0.567 [0.000]
$\widehat{\omega}_1$	0.074 [0.012]	0.072 [0.013]	0.067 [0.019]
$\widehat{\omega}_2$	0.013 [0.053]	0.013 [0.053]	0.012 [0.054]
$\widehat{c}$	-	0.039 [0.012]	0.04 [0.011]
$\widehat{a}_1$	-	-	0.38 [0.006]
$\widehat{b}_1$	-	0.49 [0.001]	0.34 [0.002]
ARCH test	53.2 [0.007]	37.1 [0.044]	25.8 [0.086]

1. Ecrivez les équations régissant les deux premiers modèles estimés ci-dessus.
2. Parmi ces trois modèles, quel est celui que vous conseillez et pourquoi?

Answer: 1. Just write down form of the model from definition of AR, AR-GARCH 2. It seems the last model is best since the ARCH test statistic is lowest and estimates are all significant.

Question 11

Soit le processus AR(1) suivant:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t,$$

où $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$. Le vecteur de paramètres à estimer est $\theta \equiv (c, \phi, \sigma^2)'$.

Indiquez la fonction de maximum de vraisemblance des paramètres θ en fonction des observations $y_1, y_2, \ldots, y_n$: $L(\theta) = f(y_1, y_2, \ldots, y_n|\theta)$, ainsi que la fonction de log-vraisemblance $l(\theta)$.

Answer: Since the error term is normally distributed, the likelihood and log-likelihood functions are the same as those explicitly computed in the TPs.

Question 12

Soit le modèle ARCH(2) défini par:

$$Y_t = \sqrt{h_t}\varepsilon_t \ ,$$

où

$$h_t = c + \sum_{i=1}^{2} a_i Y_{t-i}^2 \ , \quad c, a_1, a_2 > 0,$$

et $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$. On suppose que le processus Y_t est stationnaire.

Donnez l'expression de la variance conditionnelle et inconditionnelle de Y_t, en détaillant les calculs.

Answer: Similar answer to question 5.

Question 13

Soit $m = E[Y_t]$.

1. Que représente l'expression suivante:

$$\frac{E\left[(Y_t - m)^3\right]}{\left(\sqrt{E\left[(Y_t - m)^2\right]}\right)^3}.$$

2. Donnez son estimateur empirique et indiquez comment corriger le biais en échantillon fini.

Answer: Skewness. To correct the bias, see the estimators in A.I.

Question 14

Soit le processus MA(2) suivant:

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},$$

où $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$. On suppose que Y_t est stationnaire.

Donnez l'expression de l'espérance et des autocovariances de Y_t, en détaillant les calculs.

Answer: This is similar to exercise in Probability/Statistics review material on the course webpage.

Question 15

Les vecteurs colonne `CocaPrices` et `SP500Index` contiennent les valeurs hebdomadaires de l'indice S&P 500 et de l'action CocaCola entre le 12/11/1992 et le 14/11/2002. Commentez chaque ligne du code MATLAB ci-dessous. Expliquez en particulier le contenu du vecteur `B` après l'exécution du programme.

```
Rs = (CocaPrices(2:end)-CocaPrices(1:end-1))./CocaPrices(1:end-1);
Rm = (SP500Index(2:end)-SP500Index(1:end-1))./SP500Index(1:end-1);

B = zeros(4,1);
for i=0:3
    Rmt = Rm(52*i+1:52*(i+1));
    Rst = Rs(52*i+1:52*(i+1));
    COV = cov(Rst,Rmt);
    B(i+1) = COV(1,1)/var(Rmt);
end
B
```

Answer: Yearly correlations between Coca-Cola returns and the S&P 500 index.

Question 16

Veuillez rappeler les notions de stationnarité stricte et de stationnarité au second ordre (faible).

Question 17

Pour une série d'observations $(Y_1, \ldots, Y_T)$, expliquez ce que représentent les coefficients d'auto-corrélation et d'auto-corrélation partielle. Vous exposerez également comment les calculer.

Question 18

Soit ε_t un bruit blanc.

1. Vérifiez que les processus suivants sont stationnaires au second ordre:
   [i.]
   (a) $X_t = \varepsilon_t$,
   (b) $Y_t = (-1)^t \varepsilon_t$.
2. Montrez que leur somme $Z_t = X_t + Y_t$ n'est pas stationnaire.