Université d'Orléans - Licence Economie et Gestion
Statistique Appliquée à l'Economie et la Gestion

C. Hurlin. Examen Décembre 2011

Exercice 1 *Barème : 18 points. Test de validation d'intervalle de confiance*

Christoffersen (1998)[1] propose un test de validation des prévisions par intervalle de confiance. Soit X_t une série temporelle et $C_t(\alpha) = [LB_t(\alpha), UB_t(\alpha)]$ la prévision par intervalle de confiance à $1-\alpha\%$ associée à cette série, où LB (*lower bound*) et UB (*upper bound*) désignent respectivement les bornes inférieure et supérieure de cet intervalle. L'objectif du test est de vérifier si l'intervalle de confiance est correctement spécifié, c'est à dire si *ex-post* on vérifie bien que :

$$\Pr[X_t \notin C_t(\alpha)] = \alpha. \tag{1}$$

Pour ce faire, Christoffersen considère une variable binaire, dite violation, valant un si la réalisation de la variable X_t n'appartient pas à l'intervalle de confiance et zéro sinon :

$$I_t = \begin{cases} 1 & \text{si } X_t \notin C_t(\alpha) \\ 0 & \text{si } X_t \in C_t(\alpha) \end{cases}. \tag{2}$$

On admet que l'on dispose d'un échantillon $\{I_1, .., I_N\}$ *i.i.d.* de même loi que I_t issu d'une séquence d'intervalles de prévision $\{C_1(\alpha), .., C_N(\alpha)\}$. Sur la Figure 1, est reporté l'exemple d'une séquence de 10 intervalles de confiances (IC) à 90% (risque 10%) pour une série X_t.

Partie I : Test d'hypothèses simples (13 points)

On considère un test d'hypothèse simple contre hypothèse simple inspiré du test de Christoffersen (1998).

$$H_0 : \mathbb{E}(I_t) = \alpha, \tag{3}$$

$$H_1 : \mathbb{E}(I_t) = \beta, \tag{4}$$

avec $\beta > \alpha$.

Question 1 (1 point) Montrez que $\mathbb{E}(I_t) = \Pr[X_t \notin C_t(\alpha)]$ et commentez le choix de l'hypothèse nulle H_0.

Question 2 (1 point) Que signifie l'hypothèse alternative H_1 notamment par rapport à l'amplitude de l'intervalle de confiance $C_t(\alpha)$?

Question 3 (1 point) On admet que, de façon générale, la variable I_t suit une loi de Bernouilli de probabilité p, où p désigne la probabilité $\Pr(I_t = 1)$. Montrez que la vraisemblance de l'échantillon $\{I_1, .., I_N\}$ s'écrit

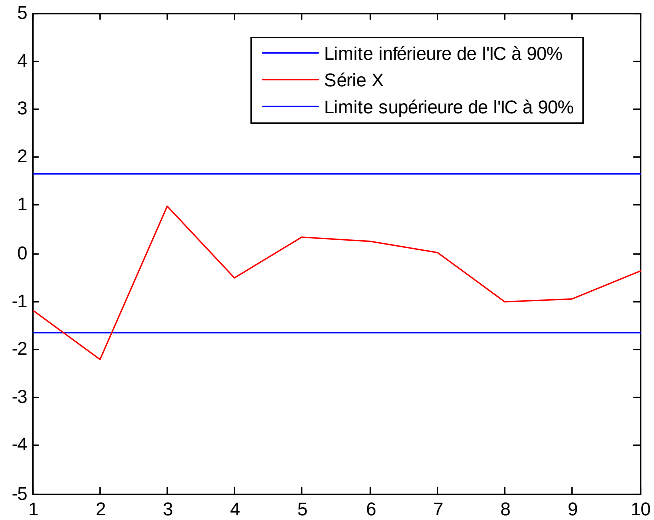
$$L(I_1, .., I_N; p) = p^H \times (1-p)^{N-H}, \tag{5}$$

où H désigne le nombre total de violations :

$$H = \sum_{t=1}^{N} I_t. \tag{6}$$

[1] Christoffersen (1998), "Evaluating interval forecasts", *International Economic Review* 39, p. 841-862.

Figure 1: Exemple d'une séquence de 10 prévisions par Intervalle de Confiance (IC) à 90% (risque 10%)



Question 4 (1 point) Montrez que le rapport des vraisemblances de l'échantillon sous les hypothèses nulle et alternative peut s'écrire sous la forme suivante :

$$\frac{L\left(I_1,..,I_N;\alpha\right)}{L\left(I_1,..,I_N;\beta\right)}=\left(\frac{\alpha}{\beta}\right)^H\left(\frac{1-\alpha}{1-\beta}\right)^{N-H}. \tag{7}$$

Question 5 (2 points) A partir du résultat de la question 4, démontrez que la région critique du test UPP au sens de Neyman Pearson est de la forme :

$$W=\left\{I_1,..,I_N \middle|\ \overline{I_N}>A\right\}, \tag{8}$$

où $\overline{I_N}=H/N=(1/N)\sum_{t=1}^N I_t$ désigne la fréquence empirique des violations et où A est une constante déterminée par le niveau de risque de première espèce. *Remarque : on admet que si $\alpha<\beta$, on a :*

$$\ln\left(\frac{\alpha}{\beta}\right)-\ln\left(\frac{1-\alpha}{1-\beta}\right)<0. \tag{9}$$

Question 6 (1 point) On admet que H suit une loi Binomiale qui peut être approximée par une loi normale. Déterminez la loi de de la variable $\overline{I_N}$ sous l'hypothèse nulle H_0, sachant que sous l'hypothèse nulle $\mathbb{E}\left(I_t\right)=\alpha$, on a $\mathbb{V}\left(I_t\right)=\alpha\left(1-\alpha\right)$.

Question 7 (2 points) Montrez que le seuil critique A du test UPP pour un niveau de risque de première espèce $\gamma\%$ est égal à :

$$A=\alpha+\Phi^{-1}\left(1-\gamma\right)\sqrt{\frac{\alpha\left(1-\alpha\right)}{N}}. \tag{10}$$

Remarque: ne pas confondre le risque de l'intervalle de confiance α et le risque de première espèce du test de validité de l'intervalle de confiance γ.

Question 8 (2 points) On cherche à tester la validité d'un intervalle de confiance à 90% (risque 10%) à partir d'une séquence de $N = 10$ prévisions. Sur la Figure 1, sont reportés les 10 réalisations de la variable X_t et les bornes inférieures et supérieures de l'intervalle de confiance. On teste $H_0 : \mathbb{E}(I_t) = 0.1$ contre $H_1 : \mathbb{E}(I_t) = 0.15$. Que pouvez vous conclure quant à la validité de l'intervalle de confiance reporté sur le Figure 1 pour un risque de première espèce γ de 5%.

Question 9 (2 points) On admet que sous l'hypothèse alternative,

$$\overline{I_N} \sim N\left(\beta, \frac{\beta(1-\beta)}{N}\right). \tag{11}$$

Calculez la puissance du test de la question 8.

Partie II : Test d'hypothèse simple contre hypothèse multiple (5 points)

On considère à présent le test suivant :

$$H_0 : \mathbb{E}(I_t) = \alpha, \tag{12}$$

$$H_1 : \mathbb{E}(I_t) > \alpha. \tag{13}$$

Question 1 (1 point) En reprenant vos résultats concernant la région critique du test $H_0 : \mathbb{E}(I_t) = \alpha$ contre $H_1 : \mathbb{E}(I_t) = \beta$, avec $\alpha < \beta$, justifiez l'existence d'un test UPP pour $H_0 : \mathbb{E}(I_t) = \alpha$ contre $H_1 : \mathbb{E}(I_t) > \alpha$. Montrez que la région critique du test UPP pour un niveau de risque de première espèce $\gamma = 5\%$ est :

$$W = \left\{ I_1, .., I_N | \ \overline{I_N} > \alpha + \Phi^{-1}(1-\gamma)\sqrt{\frac{\alpha(1-\alpha)}{N}} \right\}. \tag{14}$$

Question 2 (2 points) On considère un échantillon de 10 intervalles de confiance à 90%. On teste la validité de ces intervalles sous la forme $H_0 : \mathbb{E}(I_t) = 0.1$ contre $H_1 : \mathbb{E}(I_t) > 0.1$ pour un risque de première espèce $\gamma = 5\%$. Calculez les valeurs prises par la courbe de puissance ou courbe d'efficacité du test pour les valeurs suivantes $\mathbb{E}(I_t) = \beta > 0.1$:

$$\beta = 0.1 \quad \beta = 0.15 \quad \beta = 0.2 \quad \beta = 0.5. \tag{15}$$

Question 3 (2 points) On considère les mêmes hypothèse que dans la question8. On souhaite à présent tester l'hypothèse :

$$H_0 : \mathbb{E}(I_t) = 0.1, \tag{16}$$

$$H_1 : \mathbb{E}(I_t) \neq 0.1, \tag{17}$$

Construisez la région critique associée à ce test pour un niveau de risque de première espèce de 5%. Que pouvez conclure quand à la validité de l'intervalle si l'on observe une violation sur les 10 intervalles de confiances (cf. Figure 1) ?

Exercice 2 *Analyse de retour de campagne marketing. Barème : 4 points.*

Une entreprise souhaite le lien entre son canal de distribution et les résultats de ses ventes. Pour cela, elle dispose du tableau de contingence des achats moyens en euros pour 1500 clients. *Exemple : 140 clients passant par le canal A ont acheté en moyenne entre 150 et 300 euros de produits.* Testez l'indépendance entre le montant moyen des achats (variable X) et le canal de distribution (variable Y) au seuil de risque de 5%. *Vous détaillerez précisément votre démarche.*

$X \setminus Y$	canal A	canal B	canal C	Total
150-300 euros	140	200	350	690
300 - 600 euros	220	50	100	370
plus de 600 euros	140	250	50	440
Total	500	500	500	1500

Bon courage