

Cours de Formation par la Recherche, ENSAE-CREST.

Leçon 3 : Tests de détection de ruptures

Gilles Teyssière
stats@gillesteysiere.net

Mars–Avril 2007

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?
- Pour les séries financières, l'hypothèse de constance des paramètres du processus de volatilité est peu plausible,

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?
- Pour les séries financières, l'hypothèse de constance des paramètres du processus de volatilité est peu plausible,
- La détection de ruptures d'un processus de volatilité GARCH permet d'estimer les paramètres de ce processus sur le plus grand intervalle d'homogénéité, et donc d'avoir un estimation non biaisée de la volatilité actuelle

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?
- Pour les séries financières, l'hypothèse de constance des paramètres du processus de volatilité est peu plausible,
- La détection de ruptures d'un processus de volatilité GARCH permet d'estimer les paramètres de ce processus sur le plus grand intervalle d'homogénéité, et donc d'avoir un estimation non biaisée de la volatilité actuelle
- Ceci est primordial pour :

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?
- Pour les séries financières, l'hypothèse de constance des paramètres du processus de volatilité est peu plausible,
- La détection de ruptures d'un processus de volatilité GARCH permet d'estimer les paramètres de ce processus sur le plus grand intervalle d'homogénéité, et donc d'avoir un estimation non biaisée de la volatilité actuelle
- Ceci est primordial pour :
 - ① Les gestionnaires de risque qui utilisent des modèles GARCH pour calculer la VaR,

Problématique

- Considérons une série chronologique observée $\{Y_1, \dots, Y_T\}$,
- Cette série est elle caractérisée par un vecteur θ de paramètres constant ?
- Ou bien ce vecteur de paramètres θ évolue-t'il dans le temps ?
- Pour les séries financières, l'hypothèse de constance des paramètres du processus de volatilité est peu plausible,
- La détection de ruptures d'un processus de volatilité GARCH permet d'estimer les paramètres de ce processus sur le plus grand intervalle d'homogénéité, et donc d'avoir un estimation non biaisée de la volatilité actuelle
- Ceci est primordial pour :
 - ① Les gestionnaires de risque qui utilisent des modèles GARCH pour calculer la VaR,
 - ② Les traders qui utilisent les modèles GARCH pour corriger le biais de la formule de Black-Scholes.

Détection de ruptures : Un exemple, l'indice FTSE 100

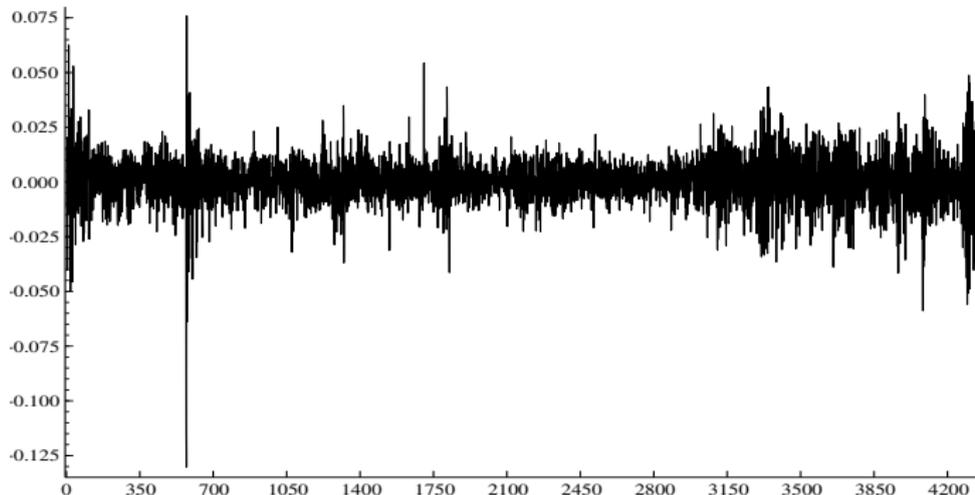


Figure: Rendements sur l'indice FTSE 100 $X_t = \log(P_t/P_{t-1})$ (1986–2002)

Nous observons un phénomène d' "intermittence" : des périodes de forte variabilité succèdent à des périodes de faible variabilité.

Un découpage de cette série en périodes de même variance

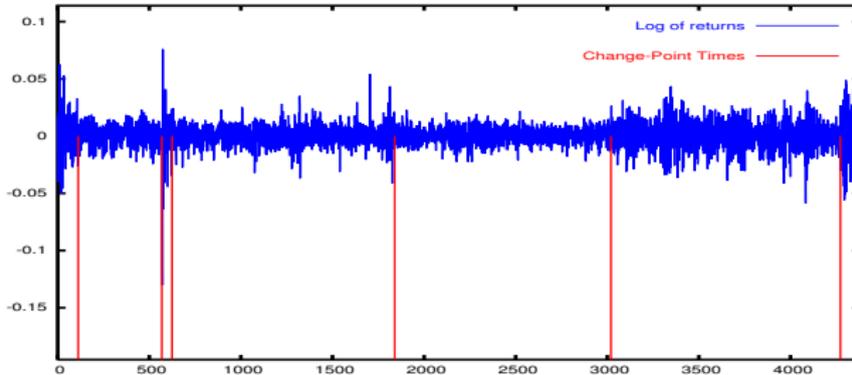


Figure: Rendements sur l'indice FTSE 100 $X_t = \log(P_t/P_{t-1})$ (1986–2002)

Note

La méthode adaptative Gaussienne utilisée pour obtenir ce découpage sera présentée plus loin dans la leçon.

Propriétés empiriques de “dépendance forte” de la volatilité

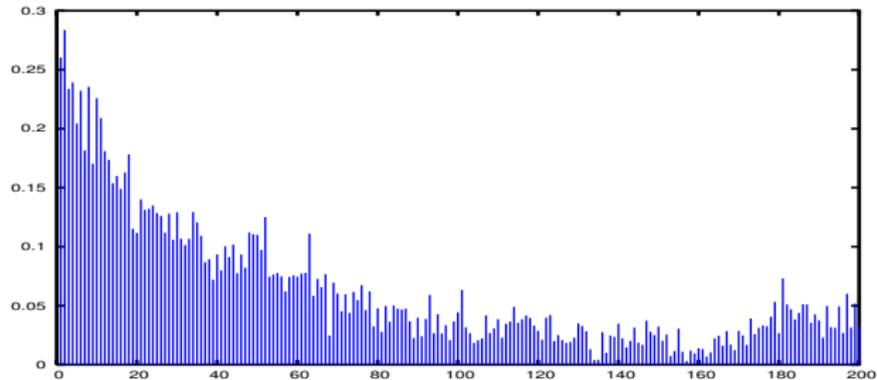


Figure: ACF de la valeur absolue des rendements $|X_t|$ sur cet indice boursier (1986–2002)

Remarque

L'ACF de cette série décroît hyperboliquement vers zéro, ce qui est typique d'un processus fortement dépendant.

Caractérisation de la dépendance d'un processus par l'ACF

- Pour un processus à mémoire courte (et faiblement dépendant) la fonction d'autocovariance décroît rapidement vers 0, le taux de décroissance est dit exponentiel,

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$$

Caractérisation de la dépendance d'un processus par l'ACF

- Pour un processus à mémoire courte (et faiblement dépendant) la fonction d'autocovariance décroît rapidement vers 0, le taux de décroissance est dit exponentiel,

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$$

Exemple

Processus AR(1) $Y_t = a_1 Y_{t-1} + \varepsilon$, qui admet une solution stationnaire si $|a_1| < 1$,

$$\gamma(k) = \frac{\sigma^2}{1 - a_1^2} a_1^k, \quad \sigma^2 = \text{Var}(\varepsilon)$$

Caractérisation de la dépendance d'un processus par l'ACF

- Pour un processus à mémoire courte (et faiblement dépendant) la fonction d'autocovariance décroît rapidement vers 0, le taux de décroissance est dit exponentiel,

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$$

Exemple

Processus AR(1) $Y_t = a_1 Y_{t-1} + \varepsilon$, qui admet une solution stationnaire si $|a_1| < 1$,

$$\gamma(k) = \frac{\sigma^2}{1 - a_1^2} a_1^k, \quad \sigma^2 = \text{Var}(\varepsilon)$$

- Pour un processus stationnaire du second ordre fortement dépendant, l'autocovariance décroît de façon hyperbolique comme suit :

$$\gamma(k) \sim k^{2d-1}, \quad d \in (0, 1/2),$$

Considérons l'ACF sur des sous-intervalles

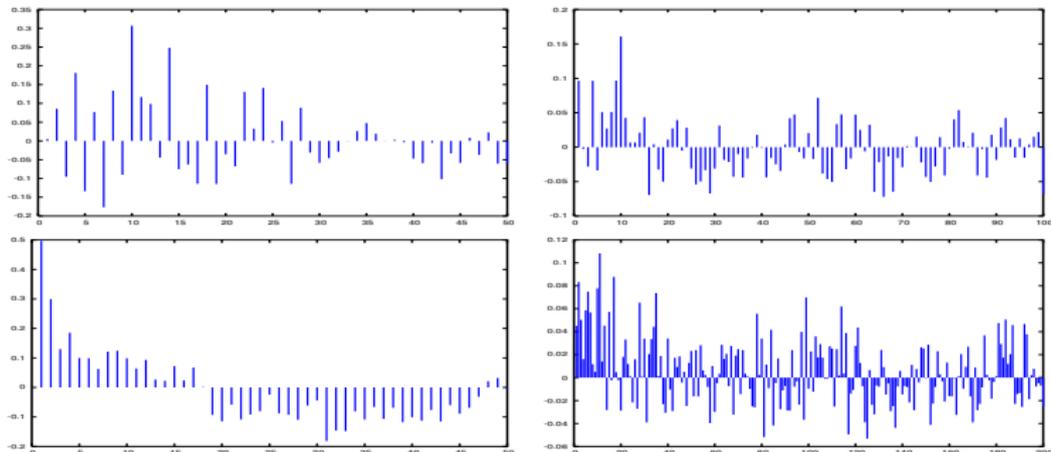


Figure: ACF de la valeur absolue $|X_t|$ sur les intervalles :

a: Haut, gauche: $[1, 112]$; b: Haut, droite: $[113, 568]$

c: Bas, gauche: $[569, 624]$; d: Bas, droite: $[625, 1840]$

Conclusions temporaires (voir aussi la leçon 1)

- 1 Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.

Conclusions temporaires (voir aussi la leçon 1)

- 1 Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- 2 En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.

Conclusions temporaires (voir aussi la leçon 1)

- 1 Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- 2 En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.
- 3 Dans ce cas particulier, la propriété de dépendance forte était déduite du comportement asymptotique de la fonction d'autocorrélation empirique

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \text{Var}(Y).$$

Conclusions temporaires (voir aussi la leçon 1)

- 1 Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- 2 En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.
- 3 Dans ce cas particulier, la propriété de dépendance forte était déduite du comportement asymptotique de la fonction d'autocorrélation empirique

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \text{Var}(Y).$$

- 4 On peut alors se poser la question de l'adéquation des méthodes statistiques simples (voire naïves) utilisées,

Conclusions temporaires (voir aussi la leçon 1)

- ❶ Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- ❷ En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.
- ❸ Dans ce cas particulier, la propriété de dépendance forte était déduite du comportement asymptotique de la fonction d'autocorrélation empirique

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \text{Var}(Y).$$

- ❹ On peut alors se poser la question de l'adéquation des méthodes statistiques simples (voire naïves) utilisées,
- ❺ D'abord, la fonction d'autocorrélation est peu informative si le processus n'est pas Gaussien (Samorodnitsky, 2002) ce qui est le cas des données financières.

Conclusions temporaires (voir aussi la leçon 1)

- ① Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- ② En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.
- ③ Dans ce cas particulier, la propriété de dépendance forte était déduite du comportement asymptotique de la fonction d'autocorrélation empirique

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \text{Var}(Y).$$

- ④ On peut alors se poser la question de l'adéquation des méthodes statistiques simples (voire naïves) utilisées,
- ⑤ D'abord, la fonction d'autocorrélation est peu informative si le processus n'est pas Gaussien (Samorodnitsky, 2002) ce qui est le cas des données financières.
- ⑥ Ensuite, si le processus n'est pas faiblement stationnaire (stationnarité du second ordre), les conclusions tirées du comportement asymptotique de la fonction d'autocorrélation empirique sont fausses.

Conclusions temporaires (voir aussi la leçon 1)

- ❶ Les séries financières, étudiées dans leur ensemble, présentent certaines caractéristiques des processus fortement dépendants.
- ❷ En revanche, lorsque ces séries sont étudiées sur des sous-intervalles, les propriétés de dépendance forte sont moins aisément mises en évidence.
- ❸ Dans ce cas particulier, la propriété de dépendance forte était déduite du comportement asymptotique de la fonction d'autocorrélation empirique

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \text{Var}(Y).$$

- ❹ On peut alors se poser la question de l'adéquation des méthodes statistiques simples (voire naïves) utilisées,
- ❺ D'abord, la fonction d'autocorrélation est peu informative si le processus n'est pas Gaussien (Samorodnitsky, 2002) ce qui est le cas des données financières.
- ❻ Ensuite, si le processus n'est pas faiblement stationnaire (stationnarité du second ordre), les conclusions tirées du comportement asymptotique de la fonction d'autocorrélation empirique sont fausses.
- ❼ Enfin, le processus de volatilité peut mixer de la dépendance forte et des ruptures.

Détection de ruptures : Tests paramétriques

- Ces tests sont basés sur l'hypothèse que le processus qui génère les données est caractérisé par un nombre fini de paramètres,

Détection de ruptures : Tests paramétriques

- Ces tests sont basés sur l'hypothèse que le processus qui génère les données est caractérisé par un nombre fini de paramètres,
- Nous faisons l'hypothèse que le processus est de type ARCH, c-à-d avec hétéroscédasticité dynamique conditionnelle,

Détection de ruptures : Tests paramétriques

- Ces tests sont basés sur l'hypothèse que le processus qui génère les données est caractérisé par un nombre fini de paramètres,
- Nous faisons l'hypothèse que le processus est de type ARCH, c-à-d avec hétéroscédasticité dynamique conditionnelle,
- Ces tests sont basés sur la comparaisons du processus des résidus du processus ARCH (GARCH) ou la valeur de la vraisemblance sur une fenêtre mobile séparant l'échantillon étudié en deux,

Détection de ruptures : Tests paramétriques

- Ces tests sont basés sur l'hypothèse que le processus qui génère les données est caractérisé par un nombre fini de paramètres,
- Nous faisons l'hypothèse que le processus est de type ARCH, c-à-d avec hétéroscédasticité dynamique conditionnelle,
- Ces tests sont basés sur la comparaisons du processus des résidus du processus ARCH (GARCH) ou la valeur de la vraisemblance sur une fenêtre mobile séparant l'échantillon étudié en deux,
- Intuitivement, ces tests ont une limitation (du moins dans l'état actuel des recherches) : ils ne considèrent que des ruptures uniques, (quoique cette méthode puisse s'étendre au cas de ruptures multiples avec plusieurs fenêtres).

Tests basés sur le processus empirique du carré des résidus de séquences ARCH

- Processus ARCH(p) défini par :

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = b_0 + \sum_{j=1}^p b_j X_{t-j}^2, \quad \varepsilon_t \text{ iid}, \quad E \varepsilon_0 = 0, \quad E \varepsilon_0^2 = 1,$$

$b_0 > 0$ et $b_i \geq 0$.

- Le processus empirique du carré des résidus $\hat{\varepsilon}_t^2 = \frac{X_t^2}{\hat{\sigma}_t^2}$

$$\hat{K}_T(s, t) = \frac{1}{\sqrt{T}} \sum_{1 \leq i \leq [Ts]} [\mathbf{I}\{\hat{\varepsilon}_i^2 \leq t\} - F(t)], \quad 0 < s \leq 1.$$

F fonction de répartition de ε_0^2

Tests basés sur le processus empirique du carré des résidus de séquences ARCH

- Tests basés sur le processus :

$$\hat{T}_T(s, t) = \sqrt{T} \cdot \frac{[Ts]}{T} \left(1 - \frac{[Ts]}{T}\right) \left(\hat{F}_{[Ts]}(t) - \hat{F}_{T-[Ts]}^*(t)\right),$$

avec

$$\hat{F}_{[Ts]}(t) = \frac{1}{[Ts]} \sum_{1 \leq i \leq [Ts]} \mathbf{1}_{\{\hat{\varepsilon}_i^2 \leq t\}}$$

- $F_{T-[Ts]}^*(t)$ est de façon analogue défini en utilisant les résidus d'indices plus grand que $[Ts]$.
- Ce test compare la fonction de repartition empirique de $\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_{[Ts]}^2$ à celle de $\hat{\varepsilon}_{[Ts]+1}^2, \dots, \hat{\varepsilon}_T^2$
- $\hat{T}_T(s, t)$ a la même limite que $\hat{K}_T(s, t) - \frac{[Ts]}{T} \hat{K}_T(1, t)$

Tests basés sur le processus empirique

Hypothèses

- ① La fonction de distribution F de ε_0^2 admet une dérivée $f(t) = F'(t)$ continue sur $(0, \infty)$,
- ② $\lim_{t \rightarrow 0} tf(t) = 0$ et $\lim_{t \rightarrow \infty} tf(t) = 0$,
- ③ le vecteur des paramètres b est estimé par un estimateur non-biaisé $\hat{b} = (\hat{b}_0, \dots, \hat{b}_p)$ qui admet la représentation :

$$\hat{b}_i - b_i = \frac{1}{n} \sum_{1 \leq j \leq n} l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \dots) + o_P(T^{-1/2}), \quad 0 \leq i \leq p,$$

(L'estimateur du Pseudo Maximum de Vraisemblance vérifie cette hypothèse.)

- ④ Les fonctions l_i sont régulières au sens suivant:

$$E l_i(\varepsilon_0^2) = 0, \quad E [l_i(\varepsilon_0^2)]^2 < \infty, \quad E [f_i(\varepsilon_0, \varepsilon_{-1}, \dots)]^2 < \infty, \quad 0 \leq i \leq p,$$

- ⑤ $E \varepsilon_0^4 < \infty$,
- ⑥ $(E \varepsilon_0^4)^{1/2} \sum_{1 \leq j \leq p} b_j < 1$.

Tests basés sur le processus empirique du carré des résidus de séquences ARCH

- Sous les hypothèses précédentes, la distribution asymptotique du processus empirique du carré des résidus d'un processus ARCH est

$$\hat{K}_T(s, t) \xrightarrow{d} K(s, t) + stf(t)\xi,$$

f est la densité de ε_i^2 , ξ est une VA Gaussienne corrélée avec $K(s, t)$.

- Comme

$$\hat{K}_T(s, t) \xrightarrow{d} K(s, t) + stf(t)\xi,$$

alors

$$\begin{aligned} \hat{T}_T(s, t) &= \hat{K}_T(s, t) - \frac{[Ts]}{T} \hat{K}_T(1, t) \\ &\sim (K(s, t) + stf(t)\xi) - s(K_T(1, t) + tf(t)\xi) \\ &= K(s, t) - sK(1, t), \end{aligned}$$

Tests basés sur le processus empirique

Test de Kolmogorov-Smirnov

- Pour $1 \leq k \leq T$ définissons (la date de rupture k est inconnue)

$$\hat{F}_k(t) = \frac{1}{k} \#\{i \leq k : \hat{\varepsilon}_i^2 \leq t\}, \quad \hat{F}_k^*(t) = \frac{1}{T-k} \#\{i > k : \hat{\varepsilon}_i^2 \leq t\},$$

$$\hat{T}_T(k, t) = \sqrt{T} \cdot \frac{k}{T} \left(1 - \frac{k}{T}\right) \left| \hat{F}_k(t) - \hat{F}_k^*(t) \right|;$$

$$\hat{M}_T = \sup_{0 \leq t \leq \infty} \max_{1 \leq k \leq T} |\hat{T}_T(k, t)| = \max_{1 \leq k \leq T} \max_{1 \leq j \leq T} |\hat{T}_T(k, \hat{\varepsilon}_j^2)|.$$

- La distribution asymptotique de \hat{M}_T est la même que celle de la statistique généralisée de Kolmogorov-Smirnov.

Tests basés sur le processus empirique

Test de Kolmogorov-Smirnov

Remarque

Estimateur de la date de rupture (supposée unique) :

$$\hat{k}_M = \max \left\{ k : \max_{1 \leq k \leq T} \max_{1 \leq j \leq T} | \hat{T}_T(k, \hat{\varepsilon}_j^2) | \right\} \quad \text{si} \quad \max_{1 \leq k \leq T} \max_{1 \leq j \leq T} | \hat{T}_T(k, \hat{\varepsilon}_j^2) | \geq \hat{c}_T(\alpha).$$

$\hat{c}_T(\alpha)$: *valeur critique asymptotique ou bootstrap au seuil α*

Tests basés sur le processus empirique

Test de Cramér–Von Mises

- La statistique de Cramér – von Mises est définie comme suit :

$$\hat{B} := \int_0^1 \left\{ \frac{1}{T} \sum_{i=1}^T [\hat{\mathcal{K}}([Ts], \hat{\varepsilon}_i^2)]^2 \right\} ds$$

qui a approximativement la distribution asymptotique suivante :

$$\hat{B} \sim B := \int_0^1 \int_0^1 \mathcal{K}^2(s, u) dud s,$$

où \mathcal{K} est le processus de Kiefer.

- Les valeurs critiques sont obtenues à partir de Blum, Kiefer et Rosenblatt (1961).

Tests basé sur le rapport de vraisemblance généralisé

- Sous l'hypothèse nulle, le processus est un GARCH(1,1) avec coefficients constants:

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2,$$

- Sous l' hypothèse alternative: pour $t > t_0$, le processus est ainsi défini :

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \omega^* + \beta^* \sigma_{t-1}^2 + \alpha^* \varepsilon_{t-1}^2,$$

où $\omega^* \neq \omega$, ou $\beta^* \neq \beta$, ou $\alpha^* \neq \alpha$.

Définition

La statistique du rapport de vraisemblance généralisé est ainsi définie :

$$\Lambda_{t_0} = \frac{\text{maximum de la fonction de vraisemblance sous l'hypothèse nulle}}{\text{maximum de la fonction de vraisemblance si rupture en } t_0}.$$

(Voir Csörgő et Horvath, 1997)

Tests basé sur le rapport de vraisemblance généralisé

Notations

- $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$: les valeurs estimées basées sur X_1, \dots, X_T ,
- $\tilde{\omega}$, $\tilde{\alpha}$, $\tilde{\beta}$: les valeurs estimées basées sur X_1, \dots, X_{t_0} ,
- $\bar{\omega}$, $\bar{\alpha}$, $\bar{\beta}$: les valeurs estimées basées sur X_{t_0+1}, \dots, X_T .
- Définissons $\hat{\sigma}_t^2 = \hat{\omega} + \hat{\beta}\hat{\sigma}_{t-1}^2 + \hat{\alpha}\hat{\varepsilon}_{t-1}^2$,
- Définissons $(\bar{\sigma}_t^2, \tilde{\sigma}_t^2)$ de façon similaire à $\hat{\sigma}_t^2$.

Puisque nous utilisons une fonction de vraisemblance Gaussienne :

$$\begin{aligned}
 -2 \ln \Lambda_{t_0} = & - \left[\sum_{t=1}^{t_0} (\ln \tilde{\sigma}_t^2 - \ln \hat{\sigma}_t^2) + \sum_{t=t_0+1}^T (\ln \bar{\sigma}_t^2 - \ln \hat{\sigma}_t^2) \right] \\
 & + \sum_{t=1}^n \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}_t^2} - \sum_{t=1}^{t_0} \frac{\hat{\varepsilon}_t^2}{\tilde{\sigma}_t^2} - \sum_{t=t_0+1}^T \frac{\hat{\varepsilon}_t^2}{\bar{\sigma}_t^2}
 \end{aligned}$$

Tests basé sur le rapport de vraisemblance généralisé

Puisque la date de rupture t_0 est inconnue, nous considérons la statistique:

$$\Lambda_T^* = \max_{1 \leq k \leq T} -2 \ln \Lambda_k$$

- Même si les observations sont iid, la statistique Λ_T^* vérifie un théorème limite de type Erdős, avec une distribution exponentielle comme limite.
- La valeur critique pour laquelle on rejette H_0 est :

$$c_T(\alpha) = \frac{[D_d(\log T) - \log[-\log(1 - \alpha)] + \log 2]^2}{2 \log \log T},$$

avec

$$D_d(x) = 2 \log x + \frac{d}{2} \log \log x - \log \Gamma \left(\frac{d}{2} \right).$$

Remarque

La vitesse de convergence vers cette limite est très lente, et les valeurs critiques asymptotiques sont bien plus grandes que les valeurs critiques pour des échantillons finis, obtenues par simulation; voir Gombay et Horváth (1996).

Tests basé sur le rapport de vraisemblance généralisé

- Statistique du rapport de vraisemblance pondérée basée sur le processus :

$$\{h(1-h)(-2 \ln \Lambda_{[Th]}), 0 < h < 1\}$$

- Si les observations sont indépendantes avec une densité dépendant de b paramètres, alors ce processus peut être approximé par $\left\{ \sum_{i=1}^b (W_i^0)^2(h), 0 < h < 1 \right\}$, où $W_i^0(\cdot), i = 1, \dots, b$ sont des ponts Browniens indépendants sur l'intervalle unité $[0, 1]$.
- Sous l'hypothèse nulle de constance des paramètres

$$\Delta_T^* := T^{-3} \sum_{k=1}^{T-1} k(T-k)(-2 \ln \Lambda_k) \xrightarrow{d} \int_0^1 \sum_{i=1}^b (W_i^0)^2(h) dh.$$

- Les valeurs critiques pour Δ_T^* sont obtenues à partir de Kiefer (1959)

Test d'Inclan et Tiao (1994)

- Basé sur le processus $\{D_T(h), h \in [0, 1]\}$

$$D_T(h) := \frac{\sum_{j=1}^{\lfloor Th \rfloor} X_j^2}{\sum_{j=1}^T X_j^2} - \frac{\lfloor Th \rfloor}{T}, \quad h \in [0, 1].$$

- Sous l'hypothèse nulle (H_0 : variance inconditionnelle constante), le processus $\{D_T(h), h \in [0, 1]\}$ converge vers un pont Brownien sur $[0, 1]$.
- Un test de constance de la variance inconditionnelle est basé sur la fonctionnelle suivante du processus $\{D_T(h)\}$, qui sous l'hypothèse nulle H_0 converge en distribution vers le supremum d'un pont Brownien sur l'intervalle $[0, 1]$

$$\sqrt{T/2} \sup_{0 \leq h \leq 1} |D_T(h)| \xrightarrow{d} \sup_{0 \leq h \leq 1} |W^0(h)|.$$

Test CUSUM de Kokoszka et Leipus (1999)

- On fait l'hypothèse que le processus $\{X_t\}$ est un processus ARCH(∞) défini par :

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{iid}, \quad E\varepsilon_0 = 0, \quad \text{Var } \varepsilon_0 = 1,$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{\infty} \alpha_j X_{t-j}^2, \quad t = 1, \dots, t_0,$$

$$\sigma_t^2 = \omega^* + \sum_{j=1}^{\infty} \alpha_j^* X_{t-j}^2, \quad t = t_0 + 1, \dots, T,$$

et on fait l'hypothèse que la variance inconditionnelle change à une date t_0 inconnue, c-à-d :

$$\Delta(n) = \frac{\omega}{1 - \sum_{j=1}^{\infty} \alpha_j} - \frac{\omega^*}{1 - \sum_{j=1}^{\infty} \alpha_j^*} \neq 0.$$

- L'hypothèse nulle est $H_0 : \omega = \omega^*, \alpha_j = \alpha_j^*$ pour tout j , tandis que l'hypothèse alternative est qu'il existe un j tel que $H_A : \omega \neq \omega^*$ ou $\alpha_j \neq \alpha_j^*$.

Test CUSUM de Kokoszka et Leipus (1999)

- Le test CUSUM est basé sur le processus $\{U_T(h), h \in [0, 1]\}$ défini par :

$$U_T(h) := \sqrt{T} \frac{[Th](T - [Th])}{T^2} \left(\frac{1}{[Th]} \sum_{j=1}^{[Th]} X_j^2 - \frac{1}{T - [Th]} \sum_{j=[Th]+1}^T X_j^2 \right),$$

- La statistique de test est basée sur la fonctionnelle suivante du processus $\{U_T(h), h \in [0, 1]\}$

$$\sup_{0 \leq h \leq 1} |U_T(h)| / \sigma \xrightarrow{d} \sup_{0 \leq h \leq 1} |W^0(h)|,$$

- La variance σ^2 est habituellement estimée par des méthodes non-paramétriques, on utilisera un estimateur spectral $s_T^2(q)$.

La méthode directe : l'algorithme de segmentation binaire

Algorithme proposé par Vostrikova (1981), qui permet d'utiliser les méthodes de détection d'une rupture unique au cas de détections multiples :

La méthode directe : l'algorithme de segmentation binaire

Algorithme proposé par Vostrikova (1981), qui permet d'utiliser les méthodes de détection d'une rupture unique au cas de détections multiples :

Algorithme

- 1 Appliquer la procédure de test sur l'ensemble de l'échantillon,

La méthode directe : l'algorithme de segmentation binaire

Algorithme proposé par Vostrikova (1981), qui permet d'utiliser les méthodes de détection d'une rupture unique au cas de détections multiples :

Algorithme

- 1 Appliquer la procédure de test sur l'ensemble de l'échantillon,
- 2 Si une rupture est détectée, diviser la série en deux,

La méthode directe : l'algorithme de segmentation binaire

Algorithme proposé par Vostrikova (1981), qui permet d'utiliser les méthodes de détection d'une rupture unique au cas de détections multiples :

Algorithme

- ➊ Appliquer la procédure de test sur l'ensemble de l'échantillon,
- ➋ Si une rupture est détectée, diviser la série en deux,
- ➌ Sur chacun de ces nouveaux segments ainsi définis, appliquer la procédure de détection de rupture,

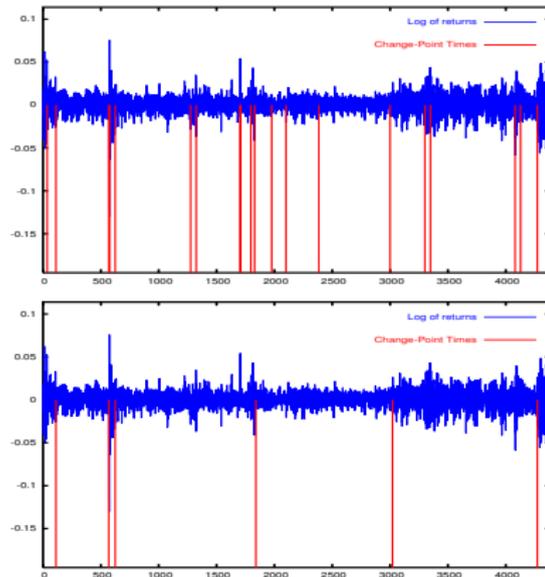
La méthode directe : l'algorithme de segmentation binaire

Algorithme proposé par Vostrikova (1981), qui permet d'utiliser les méthodes de détection d'une rupture unique au cas de détections multiples :

Algorithme

- ➊ Appliquer la procédure de test sur l'ensemble de l'échantillon,
- ➋ Si une rupture est détectée, diviser la série en deux,
- ➌ Sur chacun de ces nouveaux segments ainsi définis, appliquer la procédure de détection de rupture,
- ➍ Répéter la procédure de détection de rupture jusqu'à ce qu'aucune nouvelle rupture ne soit détectée sur les nouveaux segments.

Limitation de la méthode locale



- Haut : Procédure de segmentation binaire,
- Bas : Méthode adaptative

Question ouverte

Quelle segmentation est la bonne ?

La méthode globale

Nous proposons ici une méthode globale qui possède certaines propriétés intéressantes :

- 1 Cette méthode est capable de détecter globalement des ruptures multiples,

La méthode globale

Nous proposons ici une méthode globale qui possède certaines propriétés intéressantes :

- 1 Cette méthode est capable de détecter globalement des ruptures multiples,
- 2 Même si les données sont aussi fortement dépendantes,

La méthode globale

Nous proposons ici une méthode globale qui possède certaines propriétés intéressantes :

- 1 Cette méthode est capable de détecter globalement des ruptures multiples,
- 2 Même si les données sont aussi fortement dépendantes,
- 3 Cette méthode est basée sur une fonction de vraisemblance Gaussienne, et peut donc s'adapter facilement au cas multivarié,

La méthode globale

Nous proposons ici une méthode globale qui possède certaines propriétés intéressantes :

- 1 Cette méthode est capable de détecter globalement des ruptures multiples,
- 2 Même si les données sont aussi fortement dépendantes,
- 3 Cette méthode est basée sur une fonction de vraisemblance Gaussienne, et peut donc s'adapter facilement au cas multivarié,
- 4 Ce cas multivarié est pertinent si on considère que certaines séries peuvent avoir des dates de rupture communes,

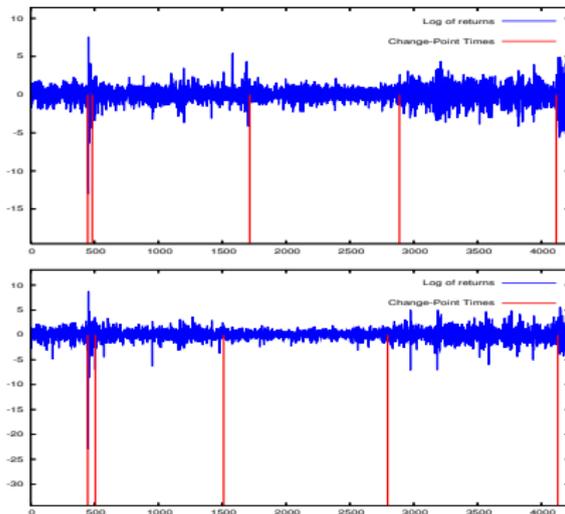
La méthode globale

Nous proposons ici une méthode globale qui possède certaines propriétés intéressantes :

- 1 Cette méthode est capable de détecter globalement des ruptures multiples,
- 2 Même si les données sont aussi fortement dépendantes,
- 3 Cette méthode est basée sur une fonction de vraisemblance Gaussienne, et peut donc s'adapter facilement au cas multivarié,
- 4 Ce cas multivarié est pertinent si on considère que certaines séries peuvent avoir des dates de rupture communes,
- 5 Enfin, même si les séries financières sont non Gaussiennes, cette méthode donne des résultats pertinents sur des données financières

Pertinence de la méthode globale avec dates de ruptures communes dans le cas multivarié

Un exemple empirique : les indices FTSE 100 et S&P 500 sur la période 1986–2002



- Détection adaptative de ruptures multiples dans le cas univarié :
- Haut : Les séries de rendements sur l'indice FTSE 100
- Bas : les séries de rendements sur l'indice S&P 500
- (les dates estimées de rupture sont représentées par des lignes verticales)

Remarque

Les dates de ruptures des deux séries semblent communes.

Présentation de la méthode globale

- Nous supposons que le processus m -dimensionnel $\{\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{m,t})'\}$ change de façon abrupte et est caractérisé par un paramètre $\theta \in \Theta$ qui reste constant entre deux ruptures.
- Soit K un entier et soit $\tau = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$ une suite ordonnée d'entiers vérifiant $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < T$.
- Pour tout $1 \leq k \leq K$, soit $U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta)$ une fonction de contraste pour estimer la vraie valeur inconnue du paramètre d'intérêt sur le segment k .
- L'estimateur du contraste minimum de $\hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})$, calculé sur le $k^{\text{ème}}$ segment de τ , est défini comme la solution du problème de minimisation suivant :

$$U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right) \leq U(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \theta), \forall \theta \in \Theta.$$

La méthode globale : Fonction de contraste

Pour tout $1 \leq k \leq K$, définissons G comme suit :

$$G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}) = U\left(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k})\right).$$

Définissons la fonction de contraste $J(\boldsymbol{\tau}, \mathbf{Y})$:

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K G(\mathbf{Y}_{\tau_{k-1}+1}, \dots, \mathbf{Y}_{\tau_k}),$$

avec $\tau_0 = 0$ et $\tau_K = T$.

La méthode globale

- Nous considérons des ruptures dans la matrice de covariance de la suite $\{\mathbf{Y}_t\}$,

La méthode globale

- Nous considérons des ruptures dans la matrice de covariance de la suite $\{\mathbf{Y}_t\}$,
- Plus précisément, nous supposons qu'il existe un entier K^* , une suite $\tau^* = \{\tau_1^*, \tau_2^*, \dots, \tau_{K^*}^*\}$ avec $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*-1}^* < \tau_{K^*}^* = T$ et K^* $(m \times m)$ matrices de covariance $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_{K^*}$ telles que $\text{Cov} \mathbf{Y}_t = \mathbb{E}(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))' = \boldsymbol{\Sigma}_k$ for $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$.

La méthode globale

- Nous considérons des ruptures dans la matrice de covariance de la suite $\{\mathbf{Y}_t\}$,
- Plus précisément, nous supposons qu'il existe un entier K^* , une suite $\tau^* = \{\tau_1^*, \tau_2^*, \dots, \tau_{K^*}^*\}$ avec $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*-1}^* < \tau_{K^*}^* = T$ et K^* $(m \times m)$ matrices de covariance $\Sigma_1, \Sigma_2, \dots, \Sigma_{K^*}$ telles que $\text{Cov} \mathbf{Y}_t = \mathbb{E}(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))' = \Sigma_k$ for $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$.
- Nous considérons deux configurations particulières :

La méthode globale

Modèle M1: Il existe un vecteur m -dimensionnel μ tel que $\mathbb{E}(\mathbf{Y}_t) = \mu$ pour $t = 1, 2, \dots, T$. De plus, $\Sigma_k \neq \Sigma_{k+1}$ pour $1 \leq k \leq K^* - 1$.

Dans ce cas simple de ruptures dans la matrice de covariance (sans rupture dans la moyenne) qui présente un intérêt pour les processus de volatilité multivariés, nous utilisons la fonction de contraste basée sur la fonction de log-vraisemblance gaussienne :

$$J(\tau, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K n_k \log |\hat{\Sigma}_{\tau_k}|, \quad (C_1)$$

où $n_k = \tau_k - \tau_{k-1}$ est la longueur du segment k , $\hat{\Sigma}_{\tau_k}$ est la matrice ($m \times m$) de covariance empirique calculée sur le segment k :

$$\hat{\Sigma}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})'$$

$\bar{\mathbf{Y}} = T^{-1} \sum_{t=1}^T \mathbf{Y}_t$: la moyenne empirique du vecteur m -dimensionnel \mathbf{Y}_t calculée sur la série complète.

La méthode globale

Modèle M2: Il existe K^* vecteurs m -dimensionnels μ_1, \dots, μ_{K^*} tels que $\mathbb{E}(\mathbf{Y}_t) = \mu_k$ pour $\tau_{k-1}^* + 1 \leq t \leq \tau_k^*$. De plus, $(\mu_k, \Sigma_k) \neq (\mu_{k+1}, \Sigma_{k+1})$ pour $1 \leq k \leq K^* - 1$.

Pour la détection de ruptures dans le vecteur des moyennes et/ou la matrice de covariance d'une suite multivariée de variables aléatoires, cette fonction de contraste se réduit à :

$$J(\tau, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K n_k \log |\hat{\Sigma}_{\tau_k}| \quad (C_2)$$

où la matrice $\hat{\Sigma}_{\tau_k}$ de covariance empirique de dimension $(m \times m)$ est calculée sur le segment k :

$$\hat{\Sigma}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})'$$

$\bar{\mathbf{Y}}_{\tau_k} = n_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$: moyenne empirique du vecteur m -dimensionnel \mathbf{Y}_t sur ce segment.

La méthode globale

Des résultats asymptotiques pour l'estimateur du contraste minimum de τ^* peuvent être obtenus dans la cadre asymptotique suivant :

A1 Pour tout $1 \leq i \leq m$ et tout $1 \leq t \leq T$, définissons $\eta_{t,i} = Y_{t,i} - \mathbb{E}(Y_{t,i})$. Il existe $C > 0$ et $1 \leq h < 2$ tels que pour tout $u \geq 0$ et tout $s \geq 1$,

$$\mathbb{E} \left(\sum_{t=u+1}^{u+s} \eta_{t,i} \right)^2 \leq C(\theta) s^h.$$

(**A1** est vérifiée avec $h = 1$ pour des suites faiblement dépendantes et $1 < h < 2$ pour des suites fortement dépendantes.)

A2 Il existe une suite $0 < a_1 < a_2 < \dots < a_{K^*-1} < a_{K^*} = 1$ telle que pour tout $T \geq 1$ et pour tout $1 \leq k \leq K^* - 1$, $\tau_k^* = [Ta_k]$.

La méthode globale

Quand le vrai nombre K^* de segments est connu, nous avons le résultat suivant sur la vitesse de convergence de l'estimateur de contraste minimum de τ^* :

Théorème

*Supposons que les conditions **A1-A2** soient vérifiées. Dans le cas du modèle **M1** (resp. modèle **M2**), soit $\hat{\tau}_T$ les dates qui minimisent le contraste empirique $J(\tau, \mathbf{Y})$ défini par l'équation (C₁) (resp. (C₂)). Alors, la suite $\{T \|\hat{\tau}_T - \tau^*\|_\infty\}$ est uniformément tendue en probabilité :*

$$\lim_{T \rightarrow \infty} \lim_{\delta \rightarrow \infty} \mathbb{P} \left(\max_{1 \leq k \leq K^* - 1} |\hat{\tau}_{T,k} - \tau_k^*| > \delta \right) = 0. \quad (C_3)$$

(Ici, $J(\tau, \mathbf{Y})$ est minimisé sur toutes les séquences possibles τ de longueur K^)*

La méthode globale

On utilise une fonction de contraste Gaussienne (cas de détection de changements de la variance)

- Cas univarié : $J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K T_k \log \sigma_k^2$, $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$ est une suite ordonnée d'entiers vérifiant $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < T$, $n_k = \tau_k - \tau_{k-1}$ est la longueur du segment k ,

La méthode globale

On utilise une fonction de contraste Gaussienne (cas de détection de changements de la variance)

- Cas univarié : $J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K T_k \log \sigma_k^2$, $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$ est une suite ordonnée d'entiers vérifiant $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < T$, $n_k = \tau_k - \tau_{k-1}$ est la longueur du segment k ,

- Cas multivarié : $J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^K T_k \log |\hat{\boldsymbol{\Sigma}}_{\tau_k}|$

La matrice ($m \times m$) de covariance empirique $\hat{\boldsymbol{\Sigma}}_{\tau_k}$ est calculée sur le segment k

$$\hat{\boldsymbol{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})'$$

$\bar{\mathbf{Y}}_{\tau_k} = T_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$: moyenne empirique de la série m -dimensionnelle \mathbf{Y}_t sur ce segment.

La méthode globale

Les dates de ruptures sont estimées en minimisant la fonction de contraste pénalisée

$$J(\boldsymbol{\tau}, \mathbf{y}) + \beta \text{pen}(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \mathbf{y}) + \beta_T K$$

où

- ① $\beta_T K$: terme de pénalité qui contrôle le niveau de résolution de la segmentation $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$.

La méthode globale

Les dates de ruptures sont estimées en minimisant la fonction de contraste pénalisée

$$J(\boldsymbol{\tau}, \mathbf{y}) + \beta \text{pen}(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \mathbf{y}) + \beta_T K$$

où

- 1 $\beta_T K$: terme de pénalité qui contrôle le niveau de résolution de la segmentation $\boldsymbol{\tau} = \{\tau_1, \tau_2, \dots, \tau_{K-1}\}$.
- 2 Si β est une fonction de T qui tend vers 0 à une vitesse appropriée quand T tend vers l'infini, le théorème suivant établit que le nombre de segments estimés converge en probabilité vers K^* et que l'équation (C_3) est toujours vérifiée.

Terme de pénalité

Théorème

Soit $\{\beta_T\}$ une suite positive de nombres réels telle que

$$\beta_T \xrightarrow{T \rightarrow \infty} 0 \text{ et } T^{2-h} \beta_T \xrightarrow{T \rightarrow \infty} \infty, \quad 1 \leq h < 2.$$

Alors, sous les hypothèses **A1-A2**, le nombre estimé de segments $K(\hat{\tau}_T)$, où $\hat{\tau}_T$ est l'estimateur du minimum de contraste pénalisé de τ^* , obtenu en minimisant $J(\tau, \mathbf{Y}) + \beta_T \text{pen}(\tau)$, converge en probabilité vers K^* .

(Ici, la fonction de contraste $J(\tau, \mathbf{Y})$ est minimisée sur l'ensemble de toutes les suites possibles τ et pour tous les K possibles, $1 \leq K \leq K_{\max}$, K_{\max} étant une borne supérieure finie de K^*)

Terme de pénalité

Choix standards pour β (qui surestiment le nombre de ruptures) :

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4 \log(T)/T^{1-2d}$ pour des séries fortement dépendantes.
- Un praticien étudiant des données réelles ne connaît pas d ,

Terme de pénalité

Choix standards pour β (qui surestiment le nombre de ruptures) :

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4 \log(T)/T^{1-2d}$ pour des séries fortement dépendantes.
- Un praticien étudiant des données réelles ne connaît pas d ,
- Comment estimer d à partir des données ?

Terme de pénalité

Choix standards pour β (qui surestiment le nombre de ruptures) :

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4 \log(T)/T^{1-2d}$ pour des séries fortement dépendantes.
- Un praticien étudiant des données réelles ne connaît pas d ,
- Comment estimer d à partir des données ?
- Si le processus n'est pas stationnaire, les méthodes d'estimation dans le domaine spectral donnent des résultats incorrects qui surestiment d et donc augmentent de façon artificielle β .

Terme de pénalité

Choix standards pour β (qui surestiment le nombre de ruptures) :

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4 \log(T)/T^{1-2d}$ pour des séries fortement dépendantes.
- Un praticien étudiant des données réelles ne connaît pas d ,
- Comment estimer d à partir des données ?
- Si le processus n'est pas stationnaire, les méthodes d'estimation dans le domaine spectral donnent des résultats incorrects qui surestiment d et donc augmentent de façon artificielle β .
- Une possibilité est d'utiliser des méthodes d'estimation dans le domaine des ondelettes, mais dans ce cas une grande taille d'échantillon est requise

Terme de pénalité

Choix standards pour β (qui surestiment le nombre de ruptures) :

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4 \log(T)/T^{1-2d}$ pour des séries fortement dépendantes.
- Un praticien étudiant des données réelles ne connaît pas d ,
- Comment estimer d à partir des données ?
- Si le processus n'est pas stationnaire, les méthodes d'estimation dans le domaine spectral donnent des résultats incorrects qui surestiment d et donc augmentent de façon artificielle β .
- Une possibilité est d'utiliser des méthodes d'estimation dans le domaine des ondelettes, mais dans ce cas une grande taille d'échantillon est requise
- Les meilleurs résultats sont obtenus avec des méthodes adaptatives, de telle façon que la segmentation ne dépende pas (trop) de β
(Détection d'une rupture dans la courbure de la courbe (K, J_K) : on choisit K de telle façon que J_K cesse de décroître significativement)

Choix adaptatif du paramètre de pénalité

Supposons que la pénalisation $\text{pen}(\tau)$ dépende seulement de la dimension du modèle, c-à-d de K le nombre de segments, posons

$$\begin{aligned} J_K &= J(\hat{\tau}_K, \mathbf{Y}), \\ p_K &= \text{pen}(\tau), \quad \forall \tau \in \mathcal{T}_K \\ \hat{p}_K &= \text{pen}(\hat{\tau}_K). \end{aligned}$$

Donc, pour tout paramètre de pénalité $\beta > 0$, la solution $\hat{\tau}(\beta)$ minimise le contraste pénalisé :

$$\begin{aligned} \hat{\tau}(\beta) &= \arg \min_{\tau} (J(\tau, \mathbf{Y}) + \beta \text{pen}(\tau)) \\ &= \hat{\tau}_{\hat{K}(\beta)} \end{aligned}$$

où

$$\hat{K}(\beta) = \arg \min_{K \geq 1} \{J_K + \beta p_K\}.$$

Choix adaptatif du paramètre de pénalité

La solution $\hat{K}(\beta)$ est une fonction constante par morceaux de β . Plus précisément, si $\hat{K}(\beta) = K$,

$$J_K + \beta p_K < \min_{L \neq K} (J_L + \beta p_L).$$

Donc, β vérifie

$$\max_{L > K} \frac{J_K - J_L}{p_L - p_K} < \beta < \min_{L < K} \frac{J_L - J_K}{p_K - p_L}.$$

Donc, il existe une suite $\{K_1 = 1 < K_2 < \dots\}$, et une suite $\{\beta_0 = \infty > \beta_1 > \dots\}$, avec

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{p_{K_{i+1}} - p_{K_i}}, \quad i \geq 1,$$

telle que $\hat{K}(\beta) = K_i, \forall \beta \in [\beta_i, \beta_{i-1})$. De plus, le sous-ensemble $\{(p_{K_i}, J_{K_i}), i \geq 1\}$ est l'enveloppe convexe de l'ensemble $\{(p_K, J_K), K \geq 1\}$.

Choix adaptatif du paramètre de pénalité

- La séquence estimée $\hat{\tau}(\beta)$ ne doit pas fortement dépendre du choix du coefficient de pénalisation β (une petite variation de β ne doit pas mener à une solution $\hat{\tau}$ radicalement différente.)
- Cette stabilité de la solution par rapport au choix de β sera assurée si nous retenons seulement les plus grands intervalles $[\beta_i, \beta_{i-1}]$, $i \geq 1$.

Nous proposons la procédure suivante :

- 1 pour $K = 1, 2, \dots, K_{MAX}$, calculer $\hat{\tau}_K$, $J_K = J(\hat{\tau}_K, \mathbf{Y})$ et $\hat{p}_K = \text{pen}(\hat{\tau}_K)$,
- 2 calculer les séries $\{K_i\}$ et $\{\beta_i\}$, et les longueurs $\{l_{K_i}\}$ des intervalles $[\beta_i, \beta_{i-1}]$,
- 3 garder la(les) plus grande(s) valeur(s) de K_i telle(s) que $l_{K_i} \gg l_{K_j}$, pour $j > i$.

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- 1 examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- 1 examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
- 2 sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- ① examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
 - ② sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.
- Approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) ,

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- ① examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
 - ② sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.
- Approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) ,
 - Dérivée seconde de cette courbe est reliée directement avec la longueur de l'intervalle $([\beta_i, \beta_{i-1}), i \geq 1)$,

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- 1 examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
 - 2 sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.
- Approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) ,
 - Dérivée seconde de cette courbe est reliée directement avec la longueur de l'intervalle $([\beta_i, \beta_{i-1}), i \geq 1)$,
 - Si nous représentons les points (p_K, J_K) , pour $1 \leq K \leq K_{MAX}$, β_i est la pente entre les points (p_{K_i}, J_{K_i}) et $(p_{K_{i+1}}, J_{K_{i+1}})$,

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- ① examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
 - ② sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.
- Approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) ,
 - Dérivée seconde de cette courbe est reliée directement avec la longueur de l'intervalle $([\beta_i, \beta_{i-1}], i \geq 1)$,
 - Si nous représentons les points (p_K, J_K) , pour $1 \leq K \leq K_{MAX}$, β_i est la pente entre les points (p_{K_i}, J_{K_i}) et $(p_{K_{i+1}}, J_{K_{i+1}})$,
 - Donc, chercher quand J_K cesse de décroître équivaut à chercher une rupture dans la pente de cette courbe,

Choix adaptatif du paramètre de pénalité: Approche heuristique

Une méthode graphique classique et naturelle pour sélectionner la dimension K peut être résumée ainsi :

- ① examiner comment le contraste J_K décroît quand K (c-à-d, p_K) augmente,
 - ② sélectionner la dimension K pour laquelle J_K cesse de décroître significativement.
- Approche heuristique cherche la courbure maximale de la courbe (p_K, J_K) ,
 - Dérivée seconde de cette courbe est reliée directement avec la longueur de l'intervalle $([\beta_i, \beta_{i-1}], i \geq 1)$,
 - Si nous représentons les points (p_K, J_K) , pour $1 \leq K \leq K_{MAX}$, β_i est la pente entre les points (p_{K_i}, J_{K_i}) et $(p_{K_{i+1}}, J_{K_{i+1}})$,
 - Donc, chercher quand J_K cesse de décroître équivaut à chercher une rupture dans la pente de cette courbe,
 - Enfin, la variation de la pente au point (p_K, J_K) est précisément la longueur l_{K_i} de cet intervalle $[\beta_i, \beta_{i-1})$.

Choix adaptatif et automatique du paramètre de pénalité

- La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{J_i\}$ mais est difficile à automatiser.

Choix adaptatif et automatique du paramètre de pénalité

- La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{J_i\}$ mais est difficile à automatiser.
- Considérons une autre approche pour sélectionner le modèle qui fournit de très bons résultats et qui est bien plus facile à automatiser pour des applications pratiques.

Choix adaptatif et automatique du paramètre de pénalité

- La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{J_i\}$ mais est difficile à automatiser.
- Considérons une autre approche pour sélectionner le modèle qui fournit de très bons résultats et qui est bien plus facile à automatiser pour des applications pratiques.
- Idée de la méthode : modéliser la décroissance de la suite $\{J_K\}$ quand il n'y a pas de rupture dans la série $\{\mathbf{Y}_t\}$ et de chercher pour quelle valeur de K ce modèle ajuste la série des contrastes observés.

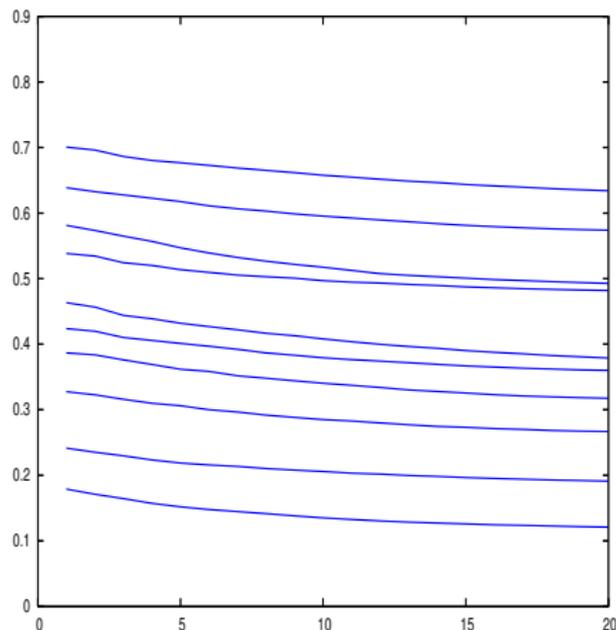
Choix adaptatif et automatique du paramètre de pénalité

- La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{J_i\}$ mais est difficile à automatiser.
- Considérons une autre approche pour sélectionner le modèle qui fournit de très bons résultats et qui est bien plus facile à automatiser pour des applications pratiques.
- Idée de la méthode : modéliser la décroissance de la suite $\{J_K\}$ quand il n'y a pas de rupture dans la série $\{\mathbf{Y}_t\}$ et de chercher pour quelle valeur de K ce modèle ajuste la série des contrastes observés.
- Sans rupture dans la variance, la distribution jointe de la séquence $\{J_K\}$ est très difficile à calculer de façon analytique,

Choix adaptatif et automatique du paramètre de pénalité

- La méthode proposée requiert une inspection soigneuse de la séquence des longueurs $\{J_i\}$ mais est difficile à automatiser.
- Considérons une autre approche pour sélectionner le modèle qui fournit de très bons résultats et qui est bien plus facile à automatiser pour des applications pratiques.
- Idée de la méthode : modéliser la décroissance de la suite $\{J_K\}$ quand il n'y a pas de rupture dans la série $\{\mathbf{Y}_t\}$ et de chercher pour quelle valeur de K ce modèle ajuste la série des contrastes observés.
- Sans rupture dans la variance, la distribution jointe de la séquence $\{J_K\}$ est très difficile à calculer de façon analytique,
- Mais quelques simulations de Monte-Carlo montrent que cette séquence décroît comme $c_1 K + c_2 K \log(K)$.

Choix adaptatif et automatique du paramètre de pénalité



- Dix séquences de fonctions de contraste $\{J_K\}$ calculées à partir de dix séquences de variables aléatoires gaussiennes i.i.d., dont le coefficient de corrélation est $\rho = 0.5$
- L'ajustement avec la fonction $c_1 K + c_2 K \log(K)$ est presque toujours parfait ($r^2 > 0.999$). (les coefficients estimés \hat{c}_1 et \hat{c}_2 sont différents pour chacune de ces séries).

Algorithme pour le choix adaptatif et automatique du paramètre de pénalité

Algorithme

Pour $i = 1, 2, \dots$,

- 1 ajuster le modèle :

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

à la série $\{J_K, K \geq K_i\}$, en supposant que $\{e_K\}$ est une suite de variables aléatoires gaussiennes centrées et i.i.d.,

Algorithme pour le choix adaptatif et automatique du paramètre de pénalité

Algorithme

Pour $i = 1, 2, \dots$,

- 1 ajuster le modèle :

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

à la série $\{J_K, K \geq K_i\}$, en supposant que $\{e_K\}$ est une suite de variables aléatoires gaussiennes centrées et i.i.d.,

- 2 évaluer la probabilité que $J_{K_{i-1}}$ suive aussi ce modèle, c-à-d., estimer la probabilité

$$\mathcal{P}_{K_i} = P(e_{K_{i-1}} \geq J_{K_{i-1}} - \hat{c}_1(K_i - 1) + \hat{c}_2(K_i - 1) \log(K_i - 1)),$$

sous le modèle estimé.

Algorithme pour le choix adaptatif et automatique du paramètre de pénalité

Algorithme

Pour $i = 1, 2, \dots$,

- ajuster le modèle :

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

à la série $\{J_K, K \geq K_i\}$, en supposant que $\{e_K\}$ est une suite de variables aléatoires gaussiennes centrées et i.i.d.,

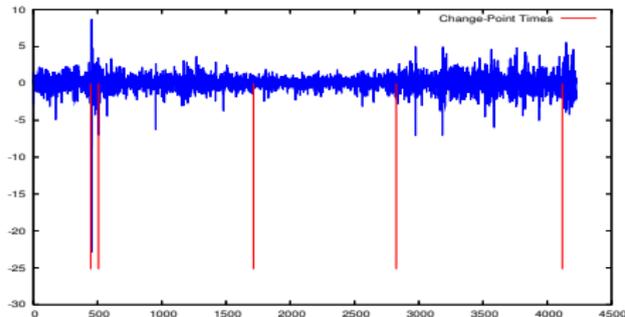
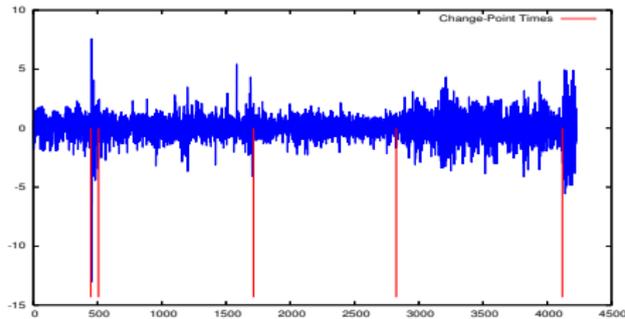
- évaluer la probabilité que $J_{K_{i-1}}$ suive aussi ce modèle, c-à-d., estimer la probabilité

$$\mathcal{P}_{K_i} = P(e_{K_{i-1}} \geq J_{K_{i-1}} - \hat{c}_1(K_i - 1) + \hat{c}_2(K_i - 1) \log(K_i - 1)),$$

sous le modèle estimé.

- Le nombre estimé de segments sera la plus grande valeur de K_i telle que la P -valeur \mathcal{P}_{K_i} soit plus petite qu'un seuil donné α . (Nous fixons $\alpha = 10^{-7}$ et $K_{MAX} = 20$)

Application à la série bivarée du FT100 et S&P 500



- Détection adaptative multivariée des ruptures
- Haut : la série des rendements sur le FTSE 100
- Bas : la série des rendements sur le S&P 500

Domaine de validité de la méthode

- Sur des données simulées (processus GARCH multivariés), nous observons qu'utiliser la méthode multivariée permet de détecter les ruptures avec bien plus de précision que dans le cas univarié,

Domaine de validité de la méthode

- Sur des données simulées (processus GARCH multivariés), nous observons qu'utiliser la méthode multivariée permet de détecter les ruptures avec bien plus de précision que dans le cas univarié,
- Sur des données simulées, nous observons que cette méthode semiparamétrique donne de biens meilleurs résultats qu'une méthode paramétrique, par exemple la méthode du rapport de vraisemblance généralisé étendue au cas multivarié,

Domaine de validité de la méthode

- Sur des données simulées (processus GARCH multivariés), nous observons qu'utiliser la méthode multivariée permet de détecter les ruptures avec bien plus de précision que dans le cas univarié,
- Sur des données simulées, nous observons que cette méthode semiparamétrique donne de biens meilleurs résultats qu'une méthode paramétrique, par exemple la méthode du rapport de vraisemblance généralisé étendue au cas multivarié,
- Sur des données réelles (non Gaussiennes), nous remarquons que la méthode automatique détecte les principales variations de la volatilité (krach boursiers)

Domaine de validité de la méthode

- Sur des données simulées (processus GARCH multivariés), nous observons qu'utiliser la méthode multivariée permet de détecter les ruptures avec bien plus de précision que dans le cas univarié,
- Sur des données simulées, nous observons que cette méthode semiparamétrique donne de biens meilleurs résultats qu'une méthode paramétrique, par exemple la méthode du rapport de vraisemblance généralisé étendue au cas multivarié,
- Sur des données réelles (non Gaussiennes), nous remarquons que la méthode automatique détecte les principales variations de la volatilité (krach boursiers)
- Néanmoins, l'utilisateur peut choisir une configuration qui lui semble plus réaliste en choisissant la P -valeur \mathcal{P}_{K_i} .

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d’un processus dont les données sont en cours de collecte.

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.
- Une fois processus X_1, \dots, X_m observé, on estime ses paramètres θ_m ,

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.
- Une fois processus X_1, \dots, X_m observé, on estime ses paramètres θ_m ,
- Sur le processus en cours d'observation après la date m , on détecte la “date d'alarme” la plus précoce κ , telle que le processus ne soit plus considéré comme homogène.

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.
- Une fois processus X_1, \dots, X_m observé, on estime ses paramètres θ_m ,
- Sur le processus en cours d'observation après la date m , on détecte la “date d'alarme” la plus précoce κ , telle que le processus ne soit plus considéré comme homogène.
- On utilise la fonction de décision suivante :

$$C(\kappa) := \left(1 + \frac{\kappa}{m}\right)^{-1} \sqrt{m}^{-1} \left| \left[\sum_{m < t \leq m + \kappa} \frac{\partial \mathcal{L}_t(\hat{\theta}_m)}{\partial \theta} \left(\hat{\mathcal{I}}(\hat{\theta}_m) \right)^{-1/2} \right] \right|, \quad 1 \leq \kappa < \infty,$$

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.
- Une fois processus X_1, \dots, X_m observé, on estime ses paramètres θ_m ,
- Sur le processus en cours d'observation après la date m , on détecte la “date d'alarme” la plus précoce κ , telle que le processus ne soit plus considéré comme homogène.
- On utilise la fonction de décision suivante :

$$C(\kappa) := \left(1 + \frac{\kappa}{m}\right)^{-1} \sqrt{m}^{-1} \left\| \left[\sum_{m < t \leq m + \kappa} \frac{\partial \mathcal{L}_t(\hat{\theta}_m)}{\partial \theta} \left(\hat{\mathcal{I}}(\hat{\theta}_m)\right)^{-1/2} \right] \right\|, \quad 1 \leq \kappa < \infty,$$

- où $\|\cdot\|$ est la norme maximale des vecteurs, $\mathcal{L}_t(\cdot)$ est la contribution de l'observation t à la fonction de vraisemblance $\hat{\mathcal{I}}(\cdot)$ est l'estimateur OPG de la matrice de covariance de l'estimateur.

Détection en ligne

- Approche alternative : détecter les ruptures “en ligne”, c-à-d détecter les changements de paramètres d'un processus dont les données sont en cours de collecte.
- Berkes *et al.* (2004) : Tests “en ligne” de la constance des paramètres d'un processus GARCH(p, q) basé sur le score de la vraisemblance :
- On suppose que le processus $\{X_t\}$ est supposé homogène sur l'intervalle $[X_1, X_m]$.
- Une fois processus X_1, \dots, X_m observé, on estime ses paramètres θ_m ,
- Sur le processus en cours d'observation après la date m , on détecte la “date d'alarme” la plus précoce κ , telle que le processus ne soit plus considéré comme homogène.
- On utilise la fonction de décision suivante :

$$C(\kappa) := \left(1 + \frac{\kappa}{m}\right)^{-1} \sqrt{m}^{-1} \left\| \left[\sum_{m < t \leq m + \kappa} \frac{\partial \mathcal{L}_t(\hat{\theta}_m)}{\partial \theta} \left(\hat{\mathcal{I}}(\hat{\theta}_m)\right)^{-1/2} \right] \right\|, \quad 1 \leq \kappa < \infty,$$

- où $\|\cdot\|$ est la norme maximale des vecteurs, $\mathcal{L}_t(\cdot)$ est la contribution de l'observation t à la fonction de vraisemblance $\hat{\mathcal{I}}(\cdot)$ est l'estimateur OPG de la matrice de covariance de l'estimateur.
- Les valeurs critiques de la statistique $C(\kappa)$ sont données dans Berkes *et al.* (2004).

Références I

-  Berkes, I., Gombay, E., Horváth, L., Kokoszka, P. (2004). Sequential change-point detection in GARCH(p, q) models. *Econometric Theory*, **20**, 1140–1167.
-  Blum, J. R., Kiefer, J., Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32**, 485–498.
-  Csörgö, M, Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley.
-  Gombay, E., Horváth, L. (1996). On the rate of approximations for maximum likelihood tests in change-point models. *Journal of Multivariate Analysis*, **56**, 120–152.
-  Horváth, L., Kokoszka, P.S., Teyssière, G. (2001). Empirical processes for the squared residuals of an ARCH sequence. *The Annals of Statistics*, **29**, 445–469.
-  Horváth, L., Kokoszka, P.S., Teyssière, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation*, **74**, 469–485.
-  Inclán, C., Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes in variance. *Journal of the American Statistical Association*, **89**, 913–923.
-  Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall: Signal Processing Series.
-  Kiefer, J. (1959). K -sample analogues of the Kolmogorov–Smirnov and Cramér-v. Mises tests. *Annals of Mathematical Statistics*, **30**, 420–447.

Références II



Kokoszka, P. S., Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39**, 231–247.



Kokoszka, P.S., Teysnière, G. (2002). Change–point detection in GARCH models: Asymptotic and bootstrap tests. *Prépublication*.



Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, **83**, 79–102.



Lavielle, M., Teysnière, G. (2006). Detection of multiple change–points in multivariate time Series. *Lithuanian Mathematical Journal*, **46**, 287–306.



Lavielle, M., Teysnière, G. (2007). Adaptive detection of multiple change–points in asset price volatility, dans *Long–Memory in Economics*, G. Teysnière et al. éditeurs, 129–156. Springer Verlag, Berlin.



Samorodnitsky, G. (2002). Long Range Dependence. *Notes de cours*.



Vostrikova, L. Yu. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.