

Cours de Formation par la Recherche, ENSAE-CREST. Leçon 4 : Méthodes statistiques robustes aux ruptures

Gilles Teyssière
stats@gillesteyssiere.net

Mars–Avril 2007

Exemple introductif : le problème du test de racine unité

- Considérons un processus X_t , nous voulons tester H_0 : X_t est $I(1)$ contre H_1 : X_t est $I(0)$.

Exemple introductif : le problème du test de racine unité

- Considérons un processus X_t , nous voulons tester H_0 : X_t est $I(1)$ contre H_1 : X_t est $I(0)$.
- Tests de racine unitaire peuvent être emboîtés dans le cadre autoregressif :

$$X_t = aX_{t-1} + \varepsilon_t$$

et nous testons $\hat{a} = 1$ (racine unité) contre $\hat{a} < 1$, c-à-d une alternative autoregressive en utilisant une t statistique sur le modèle de régression,

Exemple introductif : le problème du test de racine unité

- Considérons un processus X_t , nous voulons tester H_0 : X_t est $I(1)$ contre H_1 : X_t est $I(0)$.
- Tests de racine unitaire peuvent être emboîtés dans le cadre autoregressif :

$$X_t = aX_{t-1} + \varepsilon_t$$

et nous testons $\hat{a} = 1$ (racine unité) contre $\hat{a} < 1$, c-à-d une alternative autoregressive en utilisant une t statistique sur le modèle de régression,

- Une conséquence de cette spécification AR est que la t -statistique du paramètre estimé \hat{a} a une distribution nonstandard, donc la t -statistique doit être tabulée pour chaque spécification de la fonction de régression

$$X_t = \hat{a}X_{t-1} + \text{tendances diverses} + \text{sauts} + \text{etc} + \text{residus}$$

Exemple introductif : le problème du test de racine unité

- Considérons un processus X_t , nous voulons tester $H_0: X_t$ est $I(1)$ contre $H_1: X_t$ est $I(0)$.
- Tests de racine unitaire peuvent être emboîtés dans le cadre autoregressif :

$$X_t = aX_{t-1} + \varepsilon_t$$

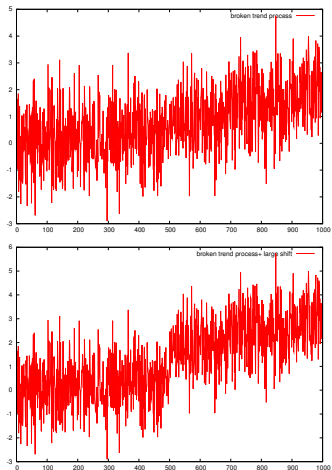
et nous testons $\hat{a} = 1$ (racine unité) contre $\hat{a} < 1$, c-à-d une alternative autoregressive en utilisant une t statistique sur le modèle de régression,

- Une conséquence de cette spécification AR est que la t -statistique du paramètre estimé \hat{a} a une distribution nonstandard, donc la t -statistique doit être tabulée pour chaque spécification de la fonction de régression

$$X_t = \hat{a}X_{t-1} + \text{tendances diverses} + \text{sauts} + \text{etc} + \text{residus}$$

- Ce qui soulève la question de la spécification correcte de cette fonction de régression.

Spécification de la fonction de régression

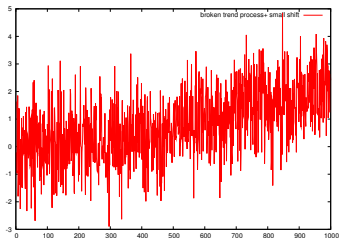


- Processus avec tendance :

$$X_t = X_0 + \mu_0 \mathbb{I}_{\{t > [\delta T]\}} c_0 t + c_1 \mathbb{I}_{\{t > [\delta T]\}} (t - [\delta T]) + \varepsilon_t, \\ \delta \in (0, 1), \quad \varepsilon_t \sim \text{iid } N(0, 1).$$

- $\delta = 0.5$: la rupture intervient au milieu de la série.
- Haut $\mu_0 = 0$
- Bas $\mu_0 = 1.0$
- La rupture est visible à l'œil nu, et on peut facilement spécifier la fonction de régression,

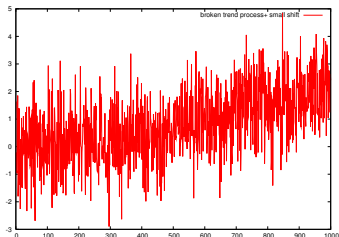
Cas difficile de spécification de la fonction de régression



- Même processus qu'avant mais ici $\mu_0 = 0.1$

$$\begin{aligned}
 X_t &= X_0 + \mu_0 \mathbb{I}_{\{t > [\delta T]\}} c_0 t \\
 &\quad + c_1 \mathbb{I}_{\{t > [\delta T]\}} (t - [\delta T]) + \varepsilon_t, \\
 \delta &= 0.5, \quad \varepsilon_t \sim \text{iid } N(0, 1).
 \end{aligned}$$

Cas difficile de spécification de la fonction de régression



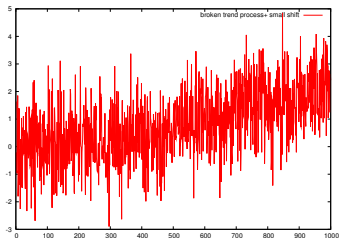
- Même processus qu'avant mais ici $\mu_0 = 0.1$

$$X_t = X_0 + \mu_0 \mathbb{I}_{\{t > [\delta T]\}} c_0 t + c_1 \mathbb{I}_{\{t > [\delta T]\}} (t - [\delta T]) + \varepsilon_t,$$

$$\delta = 0.5, \quad \varepsilon_t \sim \text{iid } N(0, 1).$$

- La rupture n'est pas visible à l'œil nu,

Cas difficile de spécification de la fonction de régression



- Même processus qu'avant mais ici $\mu_0 = 0.1$

$$X_t = X_0 + \mu_0 \mathbb{I}_{\{t > [\delta T]\}} c_0 t + c_1 \mathbb{I}_{\{t > [\delta T]\}} (t - [\delta T]) + \varepsilon_t, \\ \delta = 0.5, \quad \varepsilon_t \sim \text{iid } N(0, 1).$$

- La rupture n'est pas visible à l'œil nu,
- Une approche purement visuelle peut conduire à la négliger, ce qui conduit à conclure à la présence d'une racine unitaire.

Une approche alternative : les alternatives fractionnaires

Au lieu de considérer des processus avec des ordres d'intégration entiers, c-à-d $I(0)$, $I(1)$, $I(2)$, etc., nous considérons des processus avec un ordre d'intégration réel, notés par $I(d)$,

- ① si $d = 1$, c'est un processus avec une racine unitaire,

Une approche alternative : les alternatives fractionnaires

Au lieu de considérer des processus avec des ordres d'intégration entiers, c-à-d $I(0)$, $I(1)$, $I(2)$, etc., nous considérons des processus avec un ordre d'intégration réel, notés par $I(d)$,

- 1 si $d = 1$, c'est un processus avec une racine unitaire,
- 2 si $d = 0$, c'est un processus à courte portée,

Une approche alternative : les alternatives fractionnaires

Au lieu de considérer des processus avec des ordres d'intégration entiers, c-à-d $I(0)$, $I(1)$, $I(2)$, etc., nous considérons des processus avec un ordre d'intégration réel, notés par $I(d)$,

- 1 si $d = 1$, c'est un processus avec une racine unitaire,
- 2 si $d = 0$, c'est un processus à courte portée,
- 3 si $0 < d < 0.5$, le processus est stationnaire et fortement dépendant,

Une approche alternative : les alternatives fractionnaires

Au lieu de considérer des processus avec des ordres d'intégration entiers, c-à-d $I(0)$, $I(1)$, $I(2)$, etc., nous considérons des processus avec un ordre d'intégration réel, notés par $I(d)$,

- 1 si $d = 1$, c'est un processus avec une racine unitaire,
- 2 si $d = 0$, c'est un processus à courte portée,
- 3 si $0 < d < 0.5$, le processus est stationnaire et fortement dépendant,
- 4 si $0.5 < d < 1$, le processus est non stationnaire mais revient vers la moyenne.

Une approche alternative : les alternatives fractionnaires

Au lieu de considérer des processus avec des ordres d'intégration entiers, c-à-d $I(0)$, $I(1)$, $I(2)$, etc., nous considérons des processus avec un ordre d'intégration réel, notés par $I(d)$,

- 1 si $d = 1$, c'est un processus avec une racine unitaire,
- 2 si $d = 0$, c'est un processus à courte portée,
- 3 si $0 < d < 0.5$, le processus est stationnaire et fortement dépendant,
- 4 si $0.5 < d < 1$, le processus est non stationnaire mais revient vers la moyenne.

Remarque

Pour cette classe de test, la distribution limite sous l'hypothèse nulle est standard (à la différence des tests DF, ADF, etc, les valeurs critiques n'ont pas à être tabulées pour chaque spécification du modèle sous H_0)

Test paramétriques pour $d = d_0$ contre des alternatives fractionnaires

- Pour cette classe de tests, la distribution de la statistique de test sous H_0 est standard, (reste le problème de la spécification du modèle sous H_0)

Test paramétriques pour $d = d_0$ contre des alternatives fractionnaires

- Pour cette classe de tests, la distribution de la statistique de test sous H_0 est standard, (reste le problème de la spécification du modèle sous H_0)
- Robinson (1994) (voir leçon 2) considère des processus de la forme :

$$\phi(L)X_t = (1 - L)^d X_t = \xi_t,$$

ξ_t est une suite stationnaire du second ordre d'espérance zéro et avec une densité spectrale $f(\lambda) = (\sigma^2/2\pi)g(\lambda; \tau)$ dépendant de paramètres inconnus $\tau \in \mathbf{R}^k$ et σ^2 .

Test paramétriques pour $d = d_0$ contre des alternatives fractionnaires

- Pour cette classe de tests, la distribution de la statistique de test sous H_0 est standard, (reste le problème de la spécification du modèle sous H_0)
- Robinson (1994) (voir leçon 2) considère des processus de la forme :

$$\phi(L)X_t = (1 - L)^d X_t = \xi_t,$$

ξ_t est une suite stationnaire du second ordre d'espérance zéro et avec une densité spectrale $f(\lambda) = (\sigma^2/2\pi)g(\lambda; \tau)$ dépendant de paramètres inconnus $\tau \in \mathbf{R}^k$ et σ^2 .

- où $g(\lambda)$ doit être spécifiée :

Test paramétriques pour $d = d_0$ contre des alternatives fractionnaires

- Pour cette classe de tests, la distribution de la statistique de test sous H_0 est standard, (reste le problème de la spécification du modèle sous H_0)
- Robinson (1994) (voir leçon 2) considère des processus de la forme :

$$\phi(L)X_t = (1 - L)^d X_t = \xi_t,$$

ξ_t est une suite stationnaire du second ordre d'espérance zéro et avec une densité spectrale $f(\lambda) = (\sigma^2/2\pi)g(\lambda; \tau)$ dépendant de paramètres inconnus $\tau \in \mathbf{R}^k$ et σ^2 .

- où $g(\lambda)$ doit être spécifiée :

- Processus AR(k), dont le spectre est : $g(\lambda; \tau) = \left| 1 - \sum_{j=1}^k \tau_j e^{ij\lambda} \right|^{-2}$,

Test paramétriques pour $d = d_0$ contre des alternatives fractionnaires

- Pour cette classe de tests, la distribution de la statistique de test sous H_0 est standard, (reste le problème de la spécification du modèle sous H_0)
- Robinson (1994) (voir leçon 2) considère des processus de la forme :

$$\phi(L)X_t = (1 - L)^d X_t = \xi_t,$$

ξ_t est une suite stationnaire du second ordre d'espérance zéro et avec une densité spectrale $f(\lambda) = (\sigma^2/2\pi)g(\lambda; \tau)$ dépendant de paramètres inconnus $\tau \in \mathbf{R}^k$ et σ^2 .

- où $g(\lambda)$ doit être spécifiée :
 - Processus AR(k), dont le spectre est : $g(\lambda; \tau) = \left| 1 - \sum_{j=1}^k \tau_j e^{ij\lambda} \right|^{-2}$,
 - Processus de Bloomfield, dont le spectre est : $g(\lambda; \tau) = \exp\{2 \sum_{j=1}^k \tau_j \cos(j\lambda)\}$.
(Processus toujours stationnaire)

Alternatives fractionnaires stationnaires

Définition

Soit $\{Y_t, t \in \mathbb{R}\}$ un processus stationnaire du second ordre. Ce processus est un processus à longue portée si son spectre $f_Y(\lambda)$ est tel que dans un voisinage proche positif de la fréquence zéro,

$$f_Y(\lambda) \sim c_f \lambda^{-2d}, \quad \lambda \rightarrow 0_+, \quad c_f \in (0, \infty),$$

ou de façon équivalente, si sa fonction d'autocorrélation (ACF) $\rho_Y(k)$ décroît de façon hyperbolique comme suit :

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

Test du score $d = d_0$ contre des alternatives fractionnaires

- La statistique du score \hat{r} est définie comme suit : $\hat{r} = \frac{T^{1/2}}{\hat{\sigma}^2} \hat{A}^{-1/2} \hat{a}$.

Test du score $d = d_0$ contre des alternatives fractionnaires

- La statistique du score \hat{r} est définie comme suit : $\hat{r} = \frac{T^{1/2}}{\hat{\sigma}^2} \hat{A}^{-1/2} \hat{a}$.
- Sous l'hypothèse nulle, $H_0 : d = d_0$, $\hat{r} \rightarrow_D N(0, 1)$,

Test du score $d = d_0$ contre des alternatives fractionnaires

- La statistique du score \hat{r} est définie comme suit : $\hat{r} = \frac{T^{1/2}}{\hat{\sigma}^2} \hat{A}^{-1/2} \hat{a}$.
- Sous l'hypothèse nulle, $H_0 : d = d_0$, $\hat{r} \rightarrow_D N(0, 1)$,
- avec

$$\varphi(\lambda) = \operatorname{Re} \left\{ \log \left(\phi(e^{i\lambda}) \right)'_{d=d_0} \right\} = \log |2 \sin(\lambda/2)|, \quad \lambda \in [-\pi, \pi), \quad \tilde{\xi}_t = (1-L)^{d_0} X_t,$$

$$I_{\tilde{\xi}}(\lambda) = (1/2\pi T) \left| \sum_{t=1}^T \tilde{\xi}_t e^{it\lambda} \right|^2, \quad \lambda_j = 2\pi j/T, \quad \hat{\zeta}(\lambda) = (\partial/\partial\tau) \log g(\lambda; \hat{\tau}),$$

$$\sigma^2(\tau) = \frac{2\pi}{T} \sum_j \frac{I_{\tilde{\xi}}(\lambda_j)}{g(\lambda_j; \tau)}, \quad \hat{\sigma}^2 = \sigma^2(\hat{\tau}), \quad \hat{a} = -\frac{2\pi}{T} \sum_j' \varphi(\lambda_j) \frac{I_{\tilde{\xi}}(\lambda_j)}{g(\lambda_j; \hat{\tau})},$$

$$\hat{A} = \frac{2}{T} \left(\sum_j' |\varphi(\lambda_j)|^2 - \sum_j' \varphi(\lambda_j) \hat{\zeta}(\lambda_j)' \left\{ \sum_j' \hat{\zeta}(\lambda_j) \hat{\zeta}(\lambda_j)' \right\}^{-1} \sum_j' \hat{\zeta}(\lambda_j) \varphi(\lambda_j) \right)$$

Tests semiparamétriques pour $d = 0$ (voir leçon 2)

Tests basés sur le processus des sommes partielles (centré),

$$S_k^* = \sum_{1 \leq i \leq k} (X_i - \bar{X})$$

et son comportement asymptotique sous les hypothèses de courte et longue portée

① Amplitude standardisée S_k^* (Mandelbrot et Taqqu, 1979)

$$Q_T(q) = \frac{1}{\hat{s}_T(q)} \left[\max_{1 \leq k \leq T} S_k^* - \min_{1 \leq k \leq T} S_k^* \right] = \frac{\hat{R}_T}{\hat{s}_T(q)}.$$

Tests semiparamétriques pour $d = 0$ (voir leçon 2)

Tests basés sur le processus des sommes partielles (centré),

$$S_k^* = \sum_{1 \leq i \leq k} (X_i - \bar{X})$$

et son comportement asymptotique sous les hypothèses de courte et longue portée

- ❶ Amplitude standardisée S_k^* (Mandelbrot et Taqqu, 1979)

$$Q_T(q) = \frac{1}{\hat{s}_T(q)} \left[\max_{1 \leq k \leq T} S_k^* - \min_{1 \leq k \leq T} S_k^* \right] = \frac{\hat{R}_T}{\hat{s}_T(q)}.$$

- ❷ Variance de S_k^* : (Giraitis *et al.*, 2003)

$$M_T(q) = \frac{1}{T^2 \hat{s}_T^2(q)} \left[\sum_{k=1}^T (S_k^*)^2 - \frac{1}{T} \left(\sum_{k=1}^T S_k^* \right)^2 \right] = \frac{\hat{V}_T}{T \hat{s}_T^2(q)},$$

Estimateur consistant de la variance

- $\hat{\sigma}_T^2(q)$ est un estimateur consistant de la variance du processus :

$$\hat{\sigma}_T^2(q) = T^{-1} \sum_{i=1}^T (X_i - \bar{X})^2 + 2 \sum_{i=1}^q \omega_i(q) \hat{\gamma}_i \quad \text{with} \quad \omega_i(q) = 1 - \frac{1}{q+1},$$

Estimateur consistant de la variance

- $\hat{s}_T^2(q)$ est un estimateur consistant de la variance du processus :

$$\hat{s}_T^2(q) = T^{-1} \sum_{i=1}^T (X_i - \bar{X})^2 + 2 \sum_{i=1}^q \omega_i(q) \hat{\gamma}_i \quad \text{with} \quad \omega_i(q) = 1 - \frac{1}{q+1},$$

- Cet estimateur prend en compte les autocovariances $\hat{\gamma}_i$ du processus jusqu'à l'ordre q .

Estimateur consistant de la variance

- $\hat{\sigma}_T^2(q)$ est un estimateur consistant de la variance du processus :

$$\hat{\sigma}_T^2(q) = T^{-1} \sum_{i=1}^T (X_i - \bar{X})^2 + 2 \sum_{i=1}^q \omega_i(q) \hat{\gamma}_i \quad \text{with} \quad \omega_i(q) = 1 - \frac{1}{q+1},$$

- Cet estimateur prend en compte les autocovariances $\hat{\gamma}_i$ du processus jusqu'à l'ordre q .
- La fenêtre q est telle que

$$\frac{1}{T} + \frac{q}{T} \longrightarrow 0, \quad T \rightarrow \infty$$

Estimateur consistant de la variance

- $\hat{\sigma}_T^2(q)$ est un estimateur consistant de la variance du processus :

$$\hat{\sigma}_T^2(q) = T^{-1} \sum_{i=1}^T (X_i - \bar{X})^2 + 2 \sum_{i=1}^q \omega_i(q) \hat{\gamma}_i \quad \text{with} \quad \omega_i(q) = 1 - \frac{1}{q+1},$$

- Cet estimateur prend en compte les autocovariances $\hat{\gamma}_i$ du processus jusqu'à l'ordre q .
- La fenêtre q est telle que

$$\frac{1}{T} + \frac{q}{T} \longrightarrow 0, \quad T \rightarrow \infty$$

Remarque

Pas de règle pour choisir q : (on voudrait une statistique qui ne dépende pas trop de q ce qui exclut la statistique R/S)

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)
 - le périodogramme $I(\lambda)$

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)
 - le périodogramme $I(\lambda)$
- Le cas d'une tendance (variant faiblement) peut être considéré en utilisant une statistique basée sur le périodogramme "effilé" (par exemple l'effileur cosinus),

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)
 - le périodogramme $I(\lambda)$
- Le cas d'une tendance (variant faiblement) peut être considéré en utilisant une statistique basée sur le périodogramme "effilé" (par exemple l'effileur cosinus),
- On peut considérer le cas d'une rupture (unique) ou d'une tendance dans le processus, mais ceci implique une parfaite connaissance du type de ruptures pour bien spécifier le modèle sous l'hypothèse nulle,

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)
 - le périodogramme $I(\lambda)$
- Le cas d'une tendance (variant faiblement) peut être considéré en utilisant une statistique basée sur le périodogramme "effilé" (par exemple l'effileur cosinus),
- On peut considérer le cas d'une rupture (unique) ou d'une tendance dans le processus, mais ceci implique une parfaite connaissance du type de ruptures pour bien spécifier le modèle sous l'hypothèse nulle,
- Ce qui ne permet pas de traiter les données financières caractérisées par de brusques variations multiples (et peu évidentes à formaliser),

Limitations des tests paramétriques et semiparamétriques standards

- Les statistiques précédentes dépendent de statistiques globales telles que :
 - la moyenne du processus \bar{X} (supposée constante)
 - le périodogramme $I(\lambda)$
- Le cas d'une tendance (variant faiblement) peut être considéré en utilisant une statistique basée sur le périodogramme "effilé" (par exemple l'effileur cosinus),
- On peut considérer le cas d'une rupture (unique) ou d'une tendance dans le processus, mais ceci implique une parfaite connaissance du type de ruptures pour bien spécifier le modèle sous l'hypothèse nulle,
- Ce qui ne permet pas de traiter les données financières caractérisées par de brusques variations multiples (et peu évidentes à formaliser),
- D'où la nécessité de disposer de méthodes statistiques plus robustes.

Test nonparamétrique : la statistique du rapport des incréments

Définition

La statistique du rapport des incréments (IR) est définie comme suit :

$$IR := \frac{1}{T - 3m} \sum_{k=0}^{T-3m-1} \frac{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) + \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) \right| + \left| \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}$$

où T est la taille d'échantillon et m une fenêtre telle que, $T/m \rightarrow \infty$.

- C'est la somme de rapports des incréments du second ordre du processus des sommes partielles $S_T := \sum_{t=1}^T X_t$

Test nonparamétrique : la statistique du rapport des incréments

Définition

La statistique du rapport des incréments (IR) est définie comme suit :

$$IR := \frac{1}{T-3m} \sum_{k=0}^{T-3m-1} \frac{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) + \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) \right| + \left| \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}$$

où T est la taille d'échantillon et m une fenêtre telle que, $T/m \rightarrow \infty$.

- C'est la somme de rapports des incréments du second ordre du processus des sommes partielles $S_T := \sum_{t=1}^T X_t$
- C'est une "statistique locale" car elle est basée sur des sous-échantillons mobiles de taille $3m$

Statistique locale : la statistique du rapport des incréments

Propriétés

Statistique locale : la statistique du rapport des incréments

Propriétés

- $0 \leq IR \leq 1$

Statistique locale : la statistique du rapport des incréments

Propriétés

- $0 \leq IR \leq 1$
- *La statistique est inchangée si X_t est remplacée par toute combinaison linéaire de X_t*

Statistique locale : la statistique du rapport des incréments

Propriétés

- $0 \leq IR \leq 1$
- *La statistique est inchangée si X_t est remplacée par toute combinaison linéaire de X_t*
- *Dépend du processus des sommes partielles S_T and non de S_T^* , c-à-d la moyenne \bar{X} n'apparaît pas (insensibilité à des changements de moyenne)*

Statistique locale : la statistique du rapport des incréments

Propriétés

- $0 \leq IR \leq 1$
- *La statistique est inchangée si X_t est remplacée par toute combinaison linéaire de X_t*
- *Dépend du processus des sommes partielles S_T and non de S_T^* , c-à-d la moyenne \bar{X} n'apparaît pas (insensibilité à des changements de moyenne)*
- *De façon plus générale, cette statistique ne dépend pas de statistiques globales telles que la moyenne d'échantillon, le périodogramme, etc.*

Statistique locale : la statistique du rapport des incréments

Propriétés

- $0 \leq IR \leq 1$
- *La statistique est inchangée si X_t est remplacée par toute combinaison linéaire de X_t*
- *Dépend du processus des sommes partielles S_T and non de S_T^* , c-à-d la moyenne \bar{X} n'apparaît pas (insensibilité à des changements de moyenne)*
- *De façon plus générale, cette statistique ne dépend pas de statistiques globales telles que la moyenne d'échantillon, le périodogramme, etc.*
- *Comme pour la statistique V/S , on choisit la fenêtre $m \in [T^{1/3}, T^{1/2}]$*

Hypothèses

Hypothèse

Hypothèse (A1) Pour $-1/2 < d < 1/2$, il existe une constante $G(d) \neq 0$ et une normalisation $\bar{G}_m = G_m(d) \rightarrow \infty, A_m = A_m(d)$ telles que

$$G_m^{-1} \left(\sum_{t_1=1+[mT_1]}^{[m(T_1+\tau_1)]} (X_{t_1} - A_m), \sum_{t_2=1+[mT_2]}^{[m(T_2+\tau_2)]} (X_{t_2} - A_m) \right) \\ \rightarrow_{\text{FDD}} G(d) \left(B_{d+1/2}^1(\tau_1), B_{d+1/2}^2(\tau_2) \right)$$

quand $m, T_1, T_2 - T_1 \rightarrow \infty$, où $B_{d+1/2}^1, B_{d+1/2}^2$ sont des copies indépendantes du fBm $B_{d+1/2}$ avec coefficient d'Hurst $H = d + 1/2 \in (0, 1)$.

Hypothèses

Hypothèse

Hypothèse (A2) Pour $.5 < d < 1.5$, il existe une constante $G(d) \neq 0$ et une normalisation $\overline{G}_m = G_m(d) \rightarrow \infty$ telles que

$$G_m^{-1} (X_{[m(T_1+\tau_1)]} - X_{[mT_1]}, X_{[m(T_2+\tau_2)]} - X_{[mT_2]}) \\ \xrightarrow{\text{FDD}} G(d)(B_{d-1/2}^1(\tau_1), B_{d-1/2}^2(\tau_2))$$

quand $m, T_1, T_2 - T_1 \rightarrow \infty$, où $B_{d-1/2}^1, B_{d-1/2}^2$ sont des copies indépendantes du fBm $B_{d-1/2}$ avec coefficient d'Hurst $H = d - 1/2 \in (0, 1)$.

Convergence vers $\Lambda(d)$

Proposition

- ① *Supposons que l'hypothèse (A1) soit vérifiée, $-.5 < d < .5$. Alors, quand $T \rightarrow \infty$, $m \rightarrow \infty$, $m/T \rightarrow 0$*

$$EIR \rightarrow \Lambda(d), \quad E(IR - \Lambda(d))^2 \rightarrow 0.$$

Remarque

La Gaussianité n'est pas requise.

Convergence vers $\Lambda(d)$

Proposition

- ① *Supposons que l'hypothèse (A1) soit vérifiée, $-.5 < d < .5$. Alors, quand $T \rightarrow \infty$, $m \rightarrow \infty$, $m/T \rightarrow 0$*

$$EIR \rightarrow \Lambda(d), \quad E(IR - \Lambda(d))^2 \rightarrow 0.$$

- ② *Supposons que l'hypothèse (A2) soit vérifiée, $.5 < d < 1.5$. Alors, quand $T \rightarrow \infty$, $m \rightarrow \infty$, $m/T \rightarrow 0$,*

$$EIR \rightarrow \Lambda(d), \quad E(IR - \Lambda(d))^2 \rightarrow 0.$$

Remarque

La Gaussianité n'est pas requise.

La fonction $\Lambda(d)$

$$\Lambda(d) := E \left[\frac{|Z_d(0) + Z_d(1)|}{|Z_d(0)| + |Z_d(1)|} \right]$$

où Z_d est un processus Gaussien stationnaire en temps continu appelé l'incrément second d'un mouvement Brownien fractionnaire

$$Z_d(\tau) := \frac{1}{\sqrt{|4 - 4^{d+.5}|}} \begin{cases} \Delta^2 B_{d+.5}(\tau), & -.5 < d < .5, \\ \sqrt{2d(2d+1)} \int_0^1 \Delta B_{d-.5}(\tau+s) ds, & .5 < d < 1.5, \end{cases}$$

$Z_d(0)$ et $Z_d(1)$ sont Gaussiennes, de moyennes zéro, variance 1, et covariance

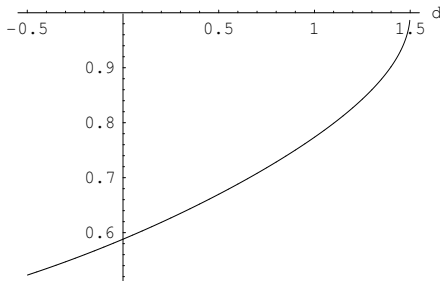
$$\rho(d) := \frac{-9^{d+1/2} + 4^{d+3/2} - 7}{2(4 - 4^{d+1/2})}$$

Graphes de la fonction $\Lambda(d)$

$$\Lambda(d) = \Lambda_0(\rho(d)),$$

où

$$\Lambda_0(r) := \frac{2}{\pi} \arctan \left(\sqrt{\frac{1+r}{1-r}} \right) + \frac{1}{\pi} \sqrt{\frac{1+r}{1-r}} \log \left(\frac{2}{1+r} \right).$$



Théorème Central Limite : cas Gaussien stationnaire

Théorème

Cas Gaussien Stationnaire Soit X_t un processus stationnaire Gaussien dont la densité spectrale $f(x)$ vérifie $f(x) = |x|^{-2d} (c_0 + O(|x|^\beta))$ ($x \rightarrow 0$), pour $-0.5 < d < 0.5$, $c_0 > 0$, $\beta > 0$. De plus, supposons que $f(x)$ est différentiable sur $(0, \pi)$ et

$$|f'(x)| \leq C|x|^{-2d-1},$$

où $C > 0$ est une constante. Alors, quand $T, m, T/m \rightarrow \infty$,

$$(T/m) \text{Var}(IR) \rightarrow \sigma^2(d),$$

et

$$(T/m)^{1/2}(IR - EIR) \rightarrow_D N(0, \sigma^2(d)),$$

Théorème Central Limite : cas intégré

Théorème

Cas intégré Soit $X_t - X_{t-1} = U_t$ un processus Gaussien stationnaire dont la densité spectrale $f(x)$ vérifie $f(x) = |x|^{2-2d} (c_0 + O(|x|^\beta))$ ($x \rightarrow 0$) pour $.5 < d < 1.25$, $c_0 > 0$, $\beta > 0$. De plus, supposons que $f(x)$ est différentiable sur $(0, \pi)$ et

$$|f'(x)| \leq C|x|^{1-2d},$$

où $C > 0$ est une constante. Alors

$$(T/m) \text{Var}(IR) \longrightarrow \sigma^2(d),$$

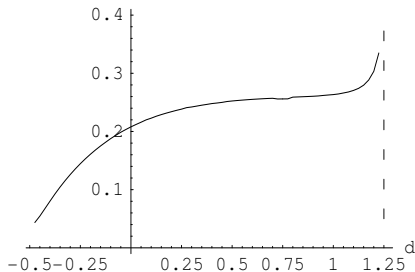
et

$$(T/m)^{1/2}(IR - EIR) \longrightarrow_D N(0, \sigma^2(d)),$$

La fonction $\sigma(d)$

$$\sigma^2(d) := 2 \int_0^\infty \text{cov} \left(\frac{|Z_d(0) + Z_d(1)|}{|Z_d(0)| + |Z_d(1)|}, \frac{|Z_d(\tau) + Z_d(\tau + 1)|}{|Z_d(\tau)| + |Z_d(\tau + 1)|} \right) d\tau,$$

où $Z_d(\tau)$ est le second incrément du mouvement Brownien fractionnaire.



Applications

La même statistique peut-être utilisée pour tester plusieurs hypothèses :

- Test de racine unité, région critique :

$$|R - \Lambda(1)| < Z_{\alpha} \sigma(1) \sqrt{\frac{m}{T - 3m}}$$

(Z_{α} est le quantile de la loi Normale $N(0, 1)$ pour un test de taille α).

Applications

La même statistique peut-être utilisée pour tester plusieurs hypothèses :

- Test de racine unité, région critique :

$$IR - \Lambda(1) < Z_\alpha \sigma(1) \sqrt{\frac{m}{T - 3m}}$$

- Test de stationnarité contre la non stationnarité, région critique :

$$IR - \Lambda(0.5) > Z_\alpha \sigma(0.5) \sqrt{\frac{m}{T - 3m}}$$

(Z_α est le quantile de la loi Normale $N(0, 1)$ pour un test de taille α).

Applications

La même statistique peut-être utilisée pour tester plusieurs hypothèses :

- Test de racine unité, région critique :

$$IR - \Lambda(1) < Z_\alpha \sigma(1) \sqrt{\frac{m}{T - 3m}}$$

- Test de stationnarité contre la non stationnarité, région critique :

$$IR - \Lambda(0.5) > Z_\alpha \sigma(0.5) \sqrt{\frac{m}{T - 3m}}$$

- Test de courte mémoire contre mémoire de longue-portée, région critique :

$$IR - \Lambda(0) > Z_\alpha \sigma(0) \sqrt{\frac{m}{T - 3m}}$$

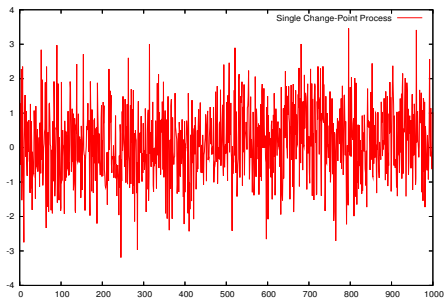
(Z_α est le quantile de la loi Normale $N(0, 1)$ pour un test de taille α).

Robustesse au cas de rupture unique en moyenne

Considérons les deux processus suivants :

- DGP A: $\mu_t = 0$, $t = 1, \dots, T$, (processus avec moyenne constante)
- DGP B: $\mu_t = 0$ pour $t = 1, \dots, [T/2]$, $\mu_t = 1/4$ pour $t = [T/2] + 1, \dots, T$.

Le processus (B) standard pour l'étude des détections de rupture est caractérisé par un faible changement de la moyenne qui ne peut être distingué à l'oeil nu (Nous reviendrons dans la section suivante sur ce processus)



Cas de rupture unique en moyenne : comparaison de la taille des tests

- DGP A: $\mu_t = 0, t = 1, \dots, T$, (processus avec moyenne constante)
- DGP B: $\mu_t = 0$ pour $t = 1, \dots, [T/2]$, $\mu_t = 1/4$ pour $t = [T/2] + 1, \dots, T$.

Table: Fréquence de rejet de l'hypothèse nulle de dépendance de courte portée pour des séries iid $N(0, 1)$. Taille du test 5%. $T = 1000$ (basé sur 10000 simulations)

	V/S		IR		$\hat{r}, \xi_t \sim \text{AR}(k)$	
DGP	$q = 10$	$q = 30$	$m = 10$	$m = 30$	$k = 1$	$k = 3$
DGP A	0.0432	0.0358	0.0514	0.0489	0.0796	0.0705
DGP B	0.8780	0.8254	0.0562	0.0585	0.7253	0.7830

Le cas de tendance avec rupture

- Considérons le processus avec tendance et rupture (dans la tendance) :

$$X_t = X_0 + c_0 t + c_1 \mathbb{I}_{\{t > [\delta T]\}}(t - [\delta T]) + \varepsilon_t, \quad \delta \in (0, 1), \quad \varepsilon_t \sim \text{iid } N(0, 1).$$

- Nous fixons $\delta = 0.5$, c-à-d la rupture a lieu au milieu de l'échantillon,

Table: Fréquence de rejet de l'hypothèse nulle de mémoire courte pour des séries iid avec une tendance déterministe linéaire et une possible rupture. Taille du test 5%. $T = 1000$ (basé sur 10000 simulations)

		V/S		IR		$\hat{r}, \xi_t \sim \text{AR}(k)$	
c_0	c_1	$q = 10$	$q = 30$	$m = 10$	$m = 30$	$k = 1$	$k = 3$
0.001	0.0	0.0444	0.0363	0.0548	0.0654	1.0000	1.0000
0.001	0.002	1.0000	1.0000	0.0581	0.2043	1.0000	1.0000

Innovations α -stables

Table: Fréquence de rejet de l'hypothèse nulle de dépendance de faible portée pour des processus FARIMA(0, d , 0) avec innovations Gaussiennes ($\alpha = 2$) et α -stables symétriques . Taille du test 5%. $T = 1000$ (basé sur 10000 simulations)

		V/S		IR		$\hat{r}, \xi_t \sim \text{AR}(k)$	
α	d	$q = 10$	$q = 30$	$m = 10$	$m = 30$	$k = 1$	$k = 3$
2.0	0.30	0.7182	0.4486	0.6752	0.3733	0.9999	0.9238
2.0	0.20	0.4816	0.2809	0.4170	0.2327	0.9956	0.8361
2.0	0.10	0.2209	0.1300	0.1864	0.1199	0.8208	0.4942
2.0	0.00	0.0432	0.0358	0.0514	0.0489	0.0814	0.0698
1.25	0.30	0.7920	0.5153	0.9656	0.6851	0.9929	0.9450
1.25	0.20	0.5660	0.2861	0.8093	0.4534	0.9910	0.9011
1.25	0.10	0.1984	0.1016	0.3966	0.2031	0.8971	0.5542
1.25	0.00	0.0224	0.0177	0.0762	0.0544	0.0387	0.0391

Modèles de volatilité GARCH (1, 1) non-homogènes

C-à-d des processus définis par :

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \theta X_{t-1}^2, \quad \varepsilon_t \sim N(0, 1) \text{ or } \varepsilon_t \sim t(7).$$

- On considère plusieurs types de processus :

Modèles de volatilité GARCH (1, 1) non-homogènes

C-à-d des processus définis par :

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \theta X_{t-1}^2, \quad \varepsilon_t \sim N(0, 1) \text{ or } \varepsilon_t \sim t(7).$$

- On considère plusieurs types de processus :
- Pour l'un de ces processus, les paramètres (ω, β, θ) sont constants, donc la variance inconditionnelle du processus $\sigma^2 = \omega / (1 - \theta - \beta)$ est constante

Modèles de volatilité GARCH (1, 1) non-homogènes

C-à-d des processus définis par :

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \theta X_{t-1}^2, \quad \varepsilon_t \sim N(0, 1) \text{ or } \varepsilon_t \sim t(7).$$

- On considère plusieurs types de processus :
- Pour l'un de ces processus, les paramètres (ω, β, θ) sont constants, donc la variance inconditionnelle du processus $\sigma^2 = \omega / (1 - \theta - \beta)$ est constante
- Pour les autres processus, les paramètres (ω, β, θ) changent à la date $t = T/2$ de telle façon que les variances inconditionnelles changent aussi avec des amplitudes différentes.

Modèles de volatilité GARCH (1, 1) non-homogènes

- DGP 0: GARCH(1,1):

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3.$$

Modèles de volatilité GARCH (1, 1) non-homogènes

- DGP 0: GARCH(1,1):

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3.$$

- DGP 1: GARCH (1,1) processus avec rupture abrupte et importante des paramètres et de la variance inconditionnelle

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3 \text{ pour } t = 1, \dots, [T/2] \quad (\sigma^2 = 0.25),$$

$$\omega = 0.15, \quad \beta = 0.65, \quad \theta = 0.25 \text{ pour } t = [T/2] + 1, \dots, T \quad (\sigma^2 = 1.5).$$

Modèles de volatilité GARCH (1, 1) non-homogènes

- DGP 0: GARCH(1,1):

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3.$$

- DGP 1: GARCH (1,1) processus avec rupture abrupte et importante des paramètres et de la variance inconditionnelle

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3 \text{ pour } t = 1, \dots, [T/2] \quad (\sigma^2 = 0.25),$$

$$\omega = 0.15, \quad \beta = 0.65, \quad \theta = 0.25 \text{ pour } t = [T/2] + 1, \dots, T \quad (\sigma^2 = 1.5).$$

- DGP 5: Deux ruptures aux dates $[T/3]$ et $[2T/3]$, c-à-d

$$\omega = 0.1, \quad \beta = 0.3, \quad \theta = 0.3 \text{ pour } t = 1, \dots, [T/3] \text{ et } t = [2T/3] + 1, \dots, T$$

$$(\sigma^2 = 0.25),$$

$$\omega = 0.125, \quad \beta = 0.6, \quad \theta = 0.1 \text{ pour } t = [T/3] + 1, \dots, [2T/3]$$

$$(\sigma^2 = 0.4667).$$

Modèles de volatilité GARCH (1, 1) non-homogènes

Table: Fréquence de rejet de l'hypothèse nulle de dépendance de courte portée pour les séries de carrés X_t^2 du processus GARCH(1,1) avec des innovations $\mathcal{N}(0, 1)$. Taille du test 5%. $T = 1000$ (basé sur 10000 simulations)

DGP	V/S		IR		$\hat{r}, \xi_t \sim \text{AR}(k)$	
	$q = 10$	$q = 30$	$m = 10$	$m = 30$	$k = 1$	$k = 3$
DGP 0	0.0648	0.0379	0.2394	0.0910	0.5621	0.0871
DGP 1	0.9958	0.9468	0.6153	0.2548	0.9950	0.8714
DGP 5	0.7844	0.6899	0.2458	0.1390	0.9566	0.7360

Cas extrême : changements aléatoires multiples

Cas du modèle “moyenne plus bruit” défini par :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad \mu_t = \mu_{t-1} + v_t$$

$$v_t = \begin{cases} 0 & \text{avec probabilité } 1 - \rho \\ w_t & \text{avec probabilité } \rho \end{cases} \quad w_t \sim N(0, 1).$$

Table: Fréquence de rejet de l'hypothèse nulle de dépendance de courte portée pour les séries du modèle “moyenne plus bruit”. Taille du test size 5%. $T = 1000$ (basé sur 10000 simulations)

p	V/S			IR		
	$q = 10$	$q = 20$	$q = 30$	$m = 10$	$m = 20$	$m = 30$
0.1	0.9988	0.9863	0.9574	1.0000	0.9999	0.9959
0.05	0.9977	0.9858	0.9544	0.9935	0.9988	0.9974
0.01	0.9941	0.9780	0.9451	0.3884	0.6592	0.7788
0.005	0.9628	0.9426	0.9139	0.1902	0.3476	0.4664
0.001	0.5238	0.5060	0.4859	0.0737	0.0905	0.1089

Conclusion de ces expériences

- A la différence de tests paramétriques, la statistique IR est peu affectée par la présence de ruptures et nonlinéarités
- Lorsque le processus étudié est bien spécifié, les tests paramétriques sont plus puissants,
- Mais une légère erreur de spécification affecte fortement les tests paramétriques, alors que la statistique IR est bien plus robuste,
- La statistique IR est d'un grand intérêt pratique pour le praticien, certaines séries étant engendrées par un processus assez complexe (voir l'exemple suivant)

Application à des séries chronologiques financières

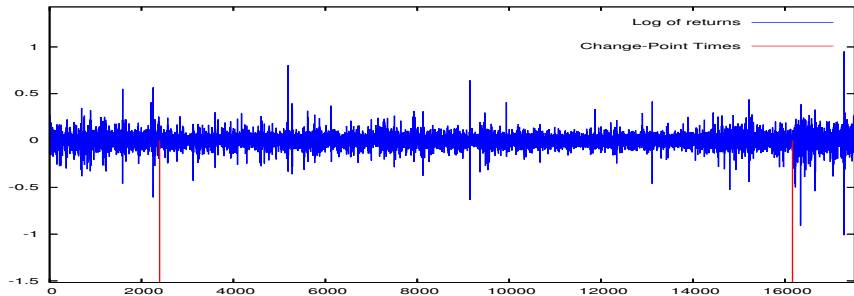


Figure: La série de rendements sur le dollar US/livre anglaise avec les dates estimées de ruptures de la variance (en utilisant la méthode Gaussienne adaptative, voir leçon 3).

La série est caractérisée par de l' "intermittence", avec deux changements de variance significatifs.

On teste la présence de dépendance de longue portée dans les séries de carrés de rendements avec les statistiques IR et V/S

Table: Les statistiques V/S et IR pour la série des carrés de rendements sur les taux de change intra-journaliers espacés de 30 minutes GBP-USD

V/S			IR		
q	V/S	P -values	m	IR	P -values
$[T^{1/3}]$	1.2072	8.9425e-11	$[T^{1/3}]$	0.6100	0.0030
40	1.0868	9.6421e-10	40	0.5918	0.3540
60	0.9798	7.9703e-09	60	0.6127	0.0222
80	0.8898	4.7048e-08	80	0.6179	0.0175
100	0.8148	2.0709e-07	100	0.6106	0.0778
$[T^{1/2}]$	0.7232	1.2618e-06	$[T^{1/2}]$	0.5992	0.2721

- La dépendance forte est présente (mais l'intensité est plus faible que ce qui est prétendu)
- Pour plusieurs séries, la statistique IR ne détecte pas de dépendance forte
- (Voir plus loin la section sur les ondelettes)

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,
 - DGP B: $\mu_t = 0$ pour $t = 1, \dots, k^*$, $\mu_t = \Delta$ pour $t = k^* + 1, \dots, T$.

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,
 - DGP B: $\mu_t = 0$ pour $t = 1, \dots, k^*$, $\mu_t = \Delta$ pour $t = k^* + 1, \dots, T$.
 - avec $k^* = \lfloor T/2 \rfloor$ (C'est le processus avec rupture vu plus haut)

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,
 - DGP B: $\mu_t = 0$ pour $t = 1, \dots, k^*$, $\mu_t = \Delta$ pour $t = k^* + 1, \dots, T$.
 - avec $k^* = \lfloor T/2 \rfloor$ (C'est le processus avec rupture vu plus haut)
- Nous considérons un test CUSUM pour discriminer entre

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,
 - DGP B: $\mu_t = 0$ pour $t = 1, \dots, k^*$, $\mu_t = \Delta$ pour $t = k^* + 1, \dots, T$.
 - avec $k^* = \lfloor T/2 \rfloor$ (C'est le processus avec rupture vu plus haut)
- Nous considérons un test CUSUM pour discriminer entre
 - H_0 : processus iid avec rupture(s) (multiples) de moyenne,

Tester l'hypothèse nulle de changement de régime contre l'alternative de dépendance de longue portée

- Considérons le processus iid :

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

- Nous considérons deux cas pour μ_t
 - DGP A: $\mu_t = 0$ pour $t = 1, \dots, T$,
 - DGP B: $\mu_t = 0$ pour $t = 1, \dots, k^*$, $\mu_t = \Delta$ pour $t = k^* + 1, \dots, T$.
 - avec $k^* = [T/2]$ (C'est le processus avec rupture vu plus haut)
- Nous considérons un test CUSUM pour discriminer entre
 - H_0 : processus iid avec rupture(s) (multiples) de moyenne,
 - H_1 : processus dépendant de longue portée,

Test CUSUM [Berkes *et al.* (2006)]

- La statistique M_T est basée sur l'estimateur de date de rupture :

$$\hat{k} = \min \left\{ k : \max_{1 \leq i \leq T} \left| \sum_{1 \leq j \leq i} X_j - \frac{i}{T} \sum_{1 \leq j \leq T} X_j \right| = \left| \sum_{1 \leq j \leq k} X_j - \frac{k}{T} \sum_{1 \leq j \leq T} X_j \right| \right\},$$

Test CUSUM [Berkes *et al.* (2006)]

- La statistique M_T est basée sur l'estimateur de date de rupture :

$$\hat{k} = \min \left\{ k : \max_{1 \leq i \leq T} \left| \sum_{1 \leq j \leq i} X_j - \frac{i}{T} \sum_{1 \leq j \leq T} X_j \right| = \left| \sum_{1 \leq j \leq k} X_j - \frac{k}{T} \sum_{1 \leq j \leq T} X_j \right| \right\},$$

- et la statistique évaluée sur les deux sous échantillons avant et après \hat{k}

$$T_{T,1} = \frac{1}{S_1(q)} \hat{k}^{-1/2} \max_{1 \leq k \leq \hat{k}} \left| \sum_{1 \leq i \leq k} X_i - \frac{k}{\hat{k}} \sum_{1 \leq i \leq \hat{k}} X_i \right|,$$

$$T_{T,2} = \frac{1}{S_2(q)} (T - \hat{k})^{-1/2} \max_{\hat{k} \leq k \leq T} \left| \sum_{\hat{k} \leq i \leq k} X_i - \frac{k - \hat{k}}{T - \hat{k}} \sum_{\hat{k} \leq i \leq T} X_i \right|,$$

Test CUSUM [Berkes *et al.* (2006)]

- La statistique M_T est basée sur l'estimateur de date de rupture :

$$\hat{k} = \min \left\{ k : \max_{1 \leq i \leq T} \left| \sum_{1 \leq j \leq i} X_j - \frac{i}{T} \sum_{1 \leq j \leq T} X_j \right| = \left| \sum_{1 \leq j \leq k} X_j - \frac{k}{T} \sum_{1 \leq j \leq T} X_j \right| \right\},$$

- et la statistique évaluée sur les deux sous échantillons avant et après \hat{k}

$$T_{T,1} = \frac{1}{S_1(q)} \hat{k}^{-1/2} \max_{1 \leq k \leq \hat{k}} \left| \sum_{1 \leq i \leq k} X_i - \frac{k}{\hat{k}} \sum_{1 \leq i \leq \hat{k}} X_i \right|,$$

$$T_{T,2} = \frac{1}{S_2(q)} (T - \hat{k})^{-1/2} \max_{\hat{k} \leq k \leq T} \left| \sum_{\hat{k} \leq i \leq k} X_i - \frac{k - \hat{k}}{T - \hat{k}} \sum_{\hat{k} \leq i \leq T} X_i \right|,$$

- $S_1(q)$ et $S_2(q)$ sont deux estimateurs spectraux de $s^2 = \sum_{j \in \mathbb{Z}} \text{cov}(X_0, X_j)$, évalués sur les sous échantillons $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_T$ respectivement.

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*
- *et la séquence des ε_t est faiblement stationnaire*

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*
- *et la séquence des ε_t est faiblement stationnaire*
- *et des hypothèses additionnelles sur la dépendance de Δ et k^* sur la taille de l'échantillon T , c-à-d, $k^* = [T\delta]$ pour quelque $0 < \delta < 1$, $T\Delta^2 \rightarrow \infty$, $\Delta^2|\hat{k} - k^*| = O_P(1)$,*

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*
- *et la séquence des ε_t est faiblement stationnaire*
- *et des hypothèses additionnelles sur la dépendance de Δ et k^* sur la taille de l'échantillon T , c-à-d, $k^* = [T\delta]$ pour quelque $0 < \delta < 1$, $T\Delta^2 \rightarrow \infty$, $\Delta^2|\hat{k} - k^*| = O_P(1)$,*
- *la fenêtre q de l'estimateur spectral de s^2 vérifie $q\Delta^2 = O(1)$, nous avons*

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*
- *et la séquence des ε_t est faiblement stationnaire*
- *et des hypothèses additionnelles sur la dépendance de Δ et k^* sur la taille de l'échantillon T , c-à-d, $k^* = [T\delta]$ pour quelque $0 < \delta < 1$, $T\Delta^2 \rightarrow \infty$, $\Delta^2|\hat{k} - k^*| = O_P(1)$,*
- *la fenêtre q de l'estimateur spectral de s^2 vérifie $q\Delta^2 = O(1)$, nous avons*
- *alors $M_T \rightarrow_D \max\left\{\sup_{0 \leq t \leq 1} |B^{(1)}(t)|, \sup_{0 \leq t \leq 1} |B^{(2)}(t)|\right\}$,*

La statistique M_T est définie par :

Définition

$$M_T = \max\{T_{T,1}, T_{T,2}\}.$$

Théorème

- *Sous H_0 les observations suivent le modèle de rupture DGP B*
- *et la séquence des ε_t est faiblement stationnaire*
- *et des hypothèses additionnelles sur la dépendance de Δ et k^* sur la taille de l'échantillon T , c-à-d, $k^* = [T\delta]$ pour quelque $0 < \delta < 1$, $T\Delta^2 \rightarrow \infty$, $\Delta^2|\hat{k} - k^*| = O_P(1)$,*
- *la fenêtre q de l'estimateur spectral de s^2 vérifie $q\Delta^2 = O(1)$, nous avons*
- *alors $M_T \rightarrow_D \max\left\{\sup_{0 \leq t \leq 1} |B^{(1)}(t)|, \sup_{0 \leq t \leq 1} |B^{(2)}(t)|\right\}$,*
- *où $B^{(1)}$ et $B^{(2)}$ sont des ponts Browniens.*

Remarques générales sur ce test CUSUM

- On peut étendre cette procédure au cas de ruptures multiples en utilisant l'algorithme de segmentation binaire.
- Il suffit de fixer le nombre maximum de ruptures des données,
- A chaque étape de cet algorithme, si H_0 est rejetée, et si le nombre maximum de ruptures n'est pas atteint, on rajoute une date de rupture, et on revient à la procédure de détection,

$$M_T := \max\{T_{T,1}, T_{T,2}, T_{T,3}\},$$

$$\rightarrow_D \max\left\{ \sup_{0 \leq t \leq 1} |B^{(1)}(t)|, \sup_{0 \leq t \leq 1} |B^{(2)}(t)|, \sup_{0 \leq t \leq 1} |B^{(3)}(t)| \right\},$$

- Le séquence de valeurs critiques pour chaque étape est calculée à partir de Kiefer (1959)
- Si le nombre de ruptures maximales est supérieur au nombre de ruptures réelles, l'hypothèse nulle n'est jamais rejetée,

Définition d'une ondelette

Principe

Utiliser quelques propriétés des coefficients des ondelettes pour estimer le paramètre de longue portée d (ou $\alpha = 2d$) en présence de tendances et de non-stationnarités.

Définition

Une ondelette est une fonction $\psi(\cdot)$ définie sur \mathbf{R} telle que

$$\int_{\mathbf{R}} \psi(t) dt = 0,$$

c-à-d vérifie la condition d'admissibilité.

Définition d'une ondelette

- On suppose que $\psi(t)$ vérifie la condition d'intégrabilité $\psi(t) \in L^1(\mathbf{R}) \cap L^2(\mathbf{R})$.
- La fonction d'ondelette est une fonction band-pass, c-à-d une petite onde dont le support est (presque) limité dans les domaines du temps et des fréquences
- Pour l'analyse par ondelettes, on choisit une ondelette de référence, appelée ondelette mère notée par ψ_0
- Base d'ondelettes : à partir de ψ_0 , une famille entière d'ondelettes est définie en utilisant deux opérateurs :
 - ① Un opérateur de translation dans le temps : $(\mathcal{T}_\tau \psi_0)(t) = \psi_0(t - \tau)$,
 - ② Un opérateur de changement d'échelle : $(\mathcal{D}_a \psi_0)(t) = \frac{1}{\sqrt{a}} \psi_0\left(\frac{t}{a}\right)$,
- On utilise ici les ondelettes de Daubechies

Estimateur de d basé sur les ondelettes

- L'ondelette mère ψ_0 a \mathcal{N} moments d'ordre zéro, avec $\mathcal{N} \geq 1$, c-à-d:

$$\int t^k \psi_0(t) dt \equiv 0, \quad k = 0, \dots, \mathcal{N} - 1.$$

- $\psi_{j,k}$ est une famille de bases de fonctions $\{\psi_{j,k} = 2^{-j/2} \psi_0(2^{-j}t - k)\}$, c-à-d une collection de changements d'échelle et de translations de l'ondelette mère ψ_0
- $j = 1, \dots, J$ sont les octaves, $k \in \mathcal{Z}$
- Les coefficients d'ondelettes sont définis par : $d_X(j, k) = \int_{\mathbf{R}} X(t) \psi_{j,k}(t) dt$
- Estimateur de Veitch et Abry (1999) : estimation de $\alpha = 2d$, basée sur la transformée discrete des ondelettes qui utilise les propriétés d'indépendance des coefficients $d_X(j, k)$ pour les processus Gaussiens fractionnaires.

Moments d'ordre zéro d'une ondelette

- L'entier \mathcal{N} est une variable clé pour l'analyse basée sur les ondelettes qui est choisi par l'utilisateur,

Moments d'ordre zéro d'une ondelette

- L'entier \mathcal{N} est une variable clé pour l'analyse basée sur les ondelettes qui est choisi par l'utilisateur,
- Par définition de \mathcal{N} , les coefficients de tout polynome d'ordre $P < \mathcal{N}$ sont strictement nuls $d_P(j, k) \equiv 0$.

Moments d'ordre zéro d'une ondelette

- L'entier \mathcal{N} est une variable clé pour l'analyse basée sur les ondelettes qui est choisi par l'utilisateur,
- Par définition de \mathcal{N} , les coefficients de tout polynôme d'ordre $P < \mathcal{N}$ sont strictement nuls $d_P(j, k) \equiv 0$.
- Pour un processus fortement dépendant (et d'espérance nulle) X_t , la superposition d'un polynôme non stationnaire n'affectera pas l'estimation du paramètre de longue portée tant que le degré du polynôme est inférieur au nombre de moments nuls \mathcal{N} .

Moments d'ordre zéro d'une ondelette

- L'entier \mathcal{N} est une variable clé pour l'analyse basée sur les ondelettes qui est choisi par l'utilisateur,
- Par définition de \mathcal{N} , les coefficients de tout polynôme d'ordre $P < \mathcal{N}$ sont strictement nuls $d_P(j, k) \equiv 0$.
- Pour un processus fortement dépendant (et d'espérance nulle) X_t , la superposition d'un polynôme non stationnaire n'affectera pas l'estimation du paramètre de longue portée tant que le degré du polynôme est inférieur au nombre de moments nuls \mathcal{N} .
- L'utilisateur choisit \mathcal{N} pour obtenir une estimation du paramètre de longue portée (coefficient d'échelle α) qui ne soit pas affectée par des tendances et des non-stationarités.

Exemple de série fortement dépendante avec tendance

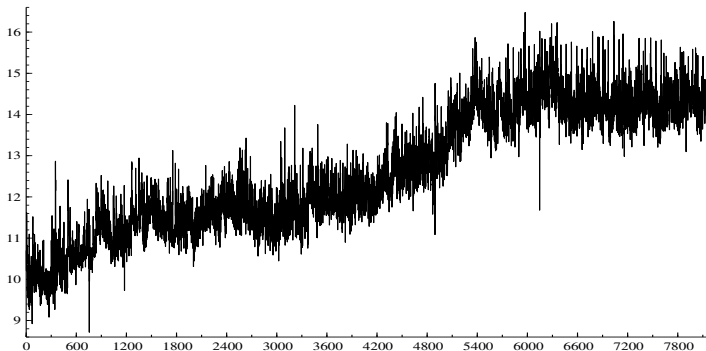


Figure: Logarithme du volume des transactions sur l'action AT&T

Cette série sera étudiée plus loin.

Principe d'estimation de d

- Le processus $d_X(j, k)$ est stationnaire (si $\mathcal{N} \geq (\alpha - 1)/2$)
- Pour de grandes échelles, sa variance vérifie la loi puissance suivante :

$$Ed_X(j, \cdot)^2 = 2^{j\alpha} c_f C, \quad \text{quand } 2^j \rightarrow \infty, \quad C = \int |\lambda|^{-\alpha} |\Psi_0(\lambda)|^2 d\lambda,$$

($\Psi_0(\lambda)$ est la transformée de Fourier de l'ondelette mère ψ_0 .)

- Le paramètre de longue portée α est estimé à partir du coefficient de la droite de régression linéaire :

$$\log_2 (Ed_X(j, \cdot)^2) = j\alpha + \log_2(c_f C) \quad \text{appellée "logscale diagram"}$$

Distribution “asymptotique” de l'estimateur

- Soit $S_X(j) = 1/n_j \sum_{k=1}^{n_j} d_X(j, k)^2$, (n_j : nombre de coefficients d'ondelettes $d_X(j, k)$ disponibles à l'octave j , n_j est de l'ordre de $2^{-j} T$)

$$\hat{\alpha}_W = \sum_{j=j_1}^{j_2} w_j (\log_2 S_X(j) - (\psi(n_j/2)/\log 2 - \log_2(n_j/2))),$$

$[j_1, j_2]$: intervalle des octaves.

- Cet estimateur a “approximativement” la distribution asymptotique suivante :

$$(\hat{\alpha} - \alpha) \sim N\left(0, \frac{1}{T \ln^2(2) 2^{1-j_1}}\right),$$

- j_1 étant l'octave la plus basse, le comportement de longue portée est capturé par les octaves plus grandes que j_1 .

Choix de l'octave basse j_1

- si j_1 est trop "petit" : fort biais car l'intervalle des octaves utilisés pour la régression inclura des octaves ne satisfaisant pas la loi de puissance à cause de dépendances de court terme (et de non-linéarités),
- si j_1 est trop "grand" : le biais est réduit, mais la variance devient très grande
- Sélection de j_1 en liaison avec le problème de fenêtre "optimale" pour l'estimateur local de Whittle (voir leçon 2)
- L'octave associée à la fenêtre optimale m_{LW}^{opt} de l'estimateur local de Whittle est donc égale à m_{LW}^{opt}/T , et correspond à l'octave 2^{-j_1} .
- En utilisant m_{LW}^{opt} on peut définir l'octave basse optimale comme suit :

$$j_1^{opt} = \left\lceil \frac{\log T - \log m_{LW}^{opt}}{\log 2} \right\rceil,$$

- Cette règle fonctionne de façon satisfaisante (en l'absence d'une autre méthode)

Estimateur de d basé sur les ondelettes : exemple du choix de j_1 pour des processus à longue portée et non linéaires

Considérons le processus de volatilité stochastique avec longue portée défini par :

$$X_t = \sigma_t \zeta_t, \quad \zeta_t \sim N(0, 1), \quad \sigma_t = \sigma \exp(Y_t/2), \quad Y_t \sim \text{FARIMA}(p, d, q)$$

Remarque

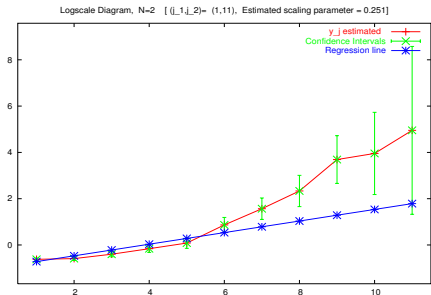
- *On peut aussi considérer le cas non-homogène :*

$$E Y_t = 0, \quad t \leq k, \quad E Y_t = \mu, \quad t > k,$$

le processus $\{X_t\}$ est à la fois fortement dépendant et contient une rupture à la date k . $T = 10000$, $k = \lfloor T/2 \rfloor$, et $\mu = 2.0$.

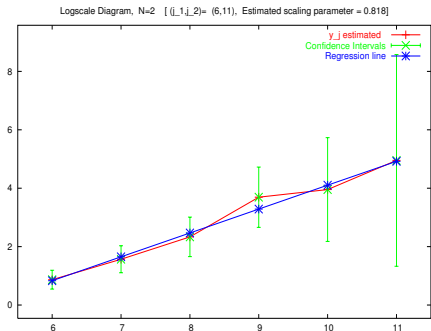
- *Dans ce cas, l'estimateur de d basé sur les ondelettes n'est pas affecté par la présence de cette rupture (à la différence des estimateurs dans le domaine temporel et spectral)*

Choix de j_1 : processus à longue portée et non linéaires



- “Logscale diagram” pour un processus LMSV, avec $\alpha = 0.90$
- On choisit ici $j_1 = 1$, $j_2 = 11$, $\mathcal{N} = 2$. $\hat{\alpha} = 0.251$, c-à-d une régression sur l'ensemble total des octaves
- Il en résulte un biais dans l'estimation du coefficient de longue portée

Choix de j_1 : processus à longue portée et non linéaires



- “Logscale diagram” pour un processus LMSV, avec $\alpha = 0.90$
- On choisit ici $j_1 = 6$, $j_2 = 11$, $\mathcal{N} = 2$. $\hat{\alpha} = 0.818$
- Le biais d'estimation du coefficient d'échelle est ici négligeable

Choix de j_1 : processus faiblement dépendant et non-homogène

Considérons un processus GARCH(1,1) avec rupture à la date (inconnue) k :

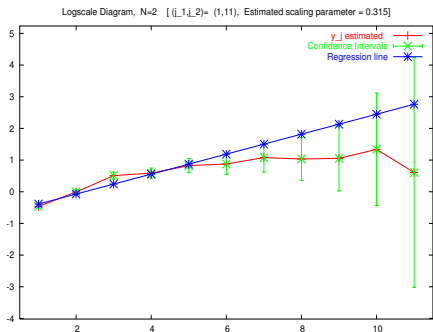
$$\begin{aligned} r_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \\ \sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma r_{t-1}^2 \quad t \leq k, \\ \sigma_t^2 &= \omega^* + \beta^* \sigma_{t-1}^2 + \gamma^* r_{t-1}^2, \quad t > k. \end{aligned}$$

avec $\omega \neq \omega^*$ ou $\beta \neq \beta^*$ ou $\gamma \neq \gamma^*$.

Remarque

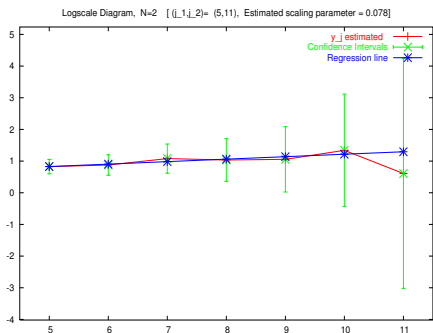
Si $\omega/(1 - \beta - \gamma) \neq \omega^/(1 - \beta^* - \gamma^*)$, les séries $|r_t|$ et r_t^2 auront un périodogramme et une ACF similaires à ceux d'un processus avec longue portée.*

Estimateur de d basé sur les ondelettes : le processus GARCH non-stationnaire



- “Logscale diagram” pour la réalisation de la valeur absolue d'un processus
- Nous choisissons $j_1 = 1$, $j_2 = 11$, $\mathcal{N} = 2$; $\hat{\alpha} = 0.315$
- Utiliser l'intervalle de toutes les octaves induit un biais d'estimation du à la présence de nonlinéarités.

Estimateur de d basé sur les ondelettes : le processus GARCH non-stationnaire

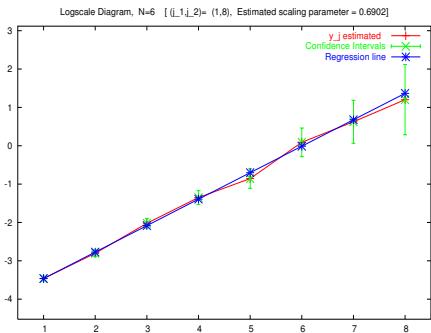


- “Logscale diagram” pour la réalisation de la valeur absolue d'un processus GARCH non stationnaire,
- Nous choisissons $j_1 = 5$, $j_2 = 11$, $\mathcal{N} = 2$; $\hat{\alpha} = 2d = 0.078$
- Choisir $j_1 \gg 1$ permet d'obtenir une estimation du paramètre de longue portée d proche de zéro, et donc non affectée par les nonlinéarités du processus.

Applications à des séries financières

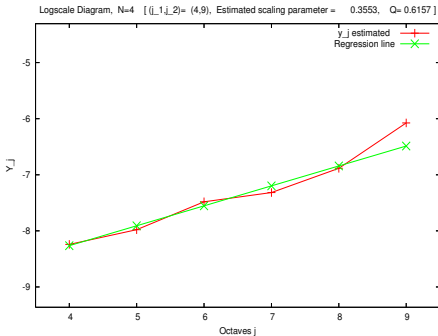
- Nous considérons les séries suivantes :
 - Taux de change intra-journaliers
 - Séries de volume de transactions (et de volatilités, c-à-d les carrés et la valeur absolue des rendements)
- Nous voulons vérifier si les séries de “volatilité” sont fortement dépendantes, c-à-d, adjuer entre forte dépendance réelle et artefact statistique
- Ceci permettrait de vérifier ou d'infirmer l'hypothèse que les séries de volatilité et de volume ont les mêmes propriétés de longue mémoire; voir Teysnière et Abry (2005).

Séries de volume de transactions : Un exemple du choix de \mathcal{N} , $\mathcal{N} = 6$;



- “Logscale diagram” pour le log du volume des transactions sur les actions AT&T, $j_1 = 1$, $\mathcal{N} = 6$; $\hat{\alpha} = 0.6902$
- Nous choisissons $\mathcal{N} = 6$ à cause de la tendance (le résultat d'estimation est stable pour $\mathcal{N} \geq 6$)
- Nous pouvons choisir toutes les octaves de car la loi d'échelle apparait dès la première octave (ce qui est assez intéressant d'un point de vue empirique)

Applications : estimation du paramètre de longue portée de la volatilité du taux de change dollar/yen observé toutes les 30 minutes



- “Logscale diagram” pour la série de volatilité (rendements absolus) du taux de change
- Pour $\mathcal{N} = 4$ et $j_1 = 4$ on obtient $\hat{\alpha} = 0.3553 \ll \hat{\alpha}_{LW} = 0.5597$
- L'estimateur basé sur les ondelettes donne une estimation de d moins élevée que celle de l'estimateur local de Whittle (LW),
- Ceci est consistant avec les résultats fournis par la statistique IR (voir au début de cette leçon)

Remarques finales sur les ondelettes

- La distribution asymptotique de l'estimateur basé sur les ondelettes (présentée dans cette leçon) est inexacte, car les coefficients d'ondelette, même distants, ne sont pas indépendants,
- La distribution exacte de cet estimateur dépend du paramètre de longue portée, n'est donc pas utilisable en pratique,
- La distribution donnée ici peut être utilisée en pratique, tout en connaissant son caractère approximatif
- Néanmoins, cet estimateur (qui requiert de grands échantillons $T \geq 5000$), permet de mieux caractériser les phénomènes de dépendance forte en finance (et de contredire une bonne partie de la littérature empirique sur le sujet dont les résultats sont sujets à caution).
- En particulier, l'hypothèse selon laquelle les processus de volatilité et de volume ont les mêmes propriétés de dépendance forte est contredite par les résultats d'estimation. (Voir Teysnière et Abry (2005) pour une étude exhaustive.)

Références



Abry, P., Veitch, D. (1998). Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory* **44**, 2–15.



Abry, P., Flandrin, P., Taqqu, M.S., Veitch, D. (2000). Wavelets for the analysis, estimation, and synthesis of scaling data, dans *Self Similar Network Traffic and Performance Evaluation*, K. Park et W. Willinger éditeurs, 39–88, Wiley.



Berkes, I., Horváth, L., Kokoszka, P.S., Shao, Q-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* **34**, 1140–1165.



Giraitis, L., Kokoszka, P., Leipus, R., Teysnière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* **112**, 265–294.



Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**, 1420–1437.



Surgailis, D., Teysnière, G., Vaičiulis, M. (2008). The increment ratio statistic. *Journal of Multivariate Analysis* **99**, 510–541.



Teyssnière, G., Abry, P. (2007). Wavelet analysis of nonlinear long-range dependent processes. Applications to financial time series, dans *Long Memory in Economics*, G. Teysnière et al. éditeurs, Springer Verlag, Berlin, pp. 173–238.