

Généralisation des tests LMC et KPSS

Apparition d'une singularité

Frédéric Proïa

Université de Bordeaux

(Ex-)Équipe ALEA (INRIA Bordeaux Sud-Ouest)
Équipe Proba-Stat (Institut de Mathématiques de Bordeaux)

Journées MAS
Toulouse

27-29 août 2014

Sommaire

- 1 Cadre de travail
 - Le processus autorégressif
 - Deux approches complémentaires
 - Quelques processus stochastiques utiles
 - Principaux résultats
- 2 Focus sur un cas particulier
- 3 Conclusion

Le processus autorégressif

Définition

On dit que la série chronologique (X_t) définie sur \mathbb{Z} est un « *processus autorégressif d'ordre p* » si elle est définie, pour tout $t \in \mathbb{Z}$, par

$$X_t = \mu + \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + \varepsilon_t$$

où (ε_t) est un bruit blanc de variance σ^2 et $\theta_p \neq 0$.

Pour tout $t \in \mathbb{Z}$, on écrit

$$\mathcal{A}(L)X_t = \mu + \varepsilon_t$$

avec, pour tout $z \in \mathbb{C}$,

$$\mathcal{A}(z) = 1 - \theta_1 z - \dots - \theta_p z^p.$$

Proposition

S'il existe $z_0 \in \mathbb{C}$ tel que $|z_0| = 1$ et $\mathcal{A}(z_0) = 0$, alors le processus (X_t) n'est pas stationnaire. Il est *intégr* et possède une *racine unitaire*.

Deux approches complémentaires

\mathcal{H}_0 : "(X_t) est intégré"

Pour tout $t \in \mathbb{Z}$,

$$\underbrace{(1-L)\mathcal{A}'(L)X_t = T_t + \varepsilon_t}_{\text{sous } \mathcal{H}_0} \quad \text{vs} \quad \underbrace{\mathcal{A}(L)X_t = T_t + \varepsilon_t}_{\text{sous } \mathcal{H}_1}.$$

Principe des tests de racine unitaire : estimer la racine supposée unitaire et tester la significativité de sa proximité avec 1.

→ Test DF, test ADF, test PP.

\mathcal{H}_0 : "(X_t) est stationnaire"

Pour tout $t \in \mathbb{Z}$,

$$\underbrace{\mathcal{A}(L)X_t = T_t + \varepsilon_t}_{\text{sous } \mathcal{H}_0} \quad \text{vs} \quad \underbrace{\mathcal{A}(L)X_t = T_t + S_t^\eta}_{\text{sous } \mathcal{H}_1}$$

où (S_t^η) est une marche aléatoire engendrée par (η_t).

Principe des tests de stationnarité : chercher la présence d'intégration dans les résidus.

→ Test KPSS, test LMC.

Quelques processus stochastiques utiles

Définition

Le processus défini, pour $t \in [0, 1]$ et $d \in \mathbb{N}$, par

$$W^{(d)}(t) = \int_0^t \int_0^{s_1} \dots \int_0^{s_{d-1}} W(s_d) ds_d \dots ds_1$$

est appelé un « *processus de Wiener intégré d'ordre d* ». Par convention, $W^{(0)}(t) \equiv W(t)$.

Par exemple

$$W^{(1)}(t) = \int_0^t W(s) ds,$$

$$W^{(2)}(t) = \int_0^t \int_0^s W(u) du ds.$$

Quelques processus stochastiques utiles

Définition

Le processus défini, pour $t \in [0, 1]$ et $r \in \mathbb{N}$, par

$$B_r(t) = h_r(W)(t)$$

est appelé un « *pont brownien généralisé d'ordre r* ». La fonction h_r de $C([0, 1])$ dans $C([0, 1])$ est donnée explicitement par MacNeill en 1978.

Par exemple

$$B_0(t) = h_0(W)(t) = W(t) - tW(1),$$

$$B_1(t) = h_1(W)(t) = W(t) + t(2 - 3t)W(1) - 6t(1 - t) \int_0^1 W(s) ds.$$

Quelques processus stochastiques utiles

Définition

Le processus défini, pour $t \in [0, 1]$ et $r, d \in \mathbb{N}$, par

$$C_{r,d}(t) = h_r(W^{(d)})(t)$$

est appelé un « *pont brownien intégré d'ordre $r \times d$* ». Par convention, $C_{r,0}(t) \equiv B_r(t)$.

Par exemple

$$C_{0,1}(t) = h_0(W^{(1)})(t) = \int_0^t W(s) ds - t \int_0^1 W(s) ds,$$

$$C_{1,1}(t) = h_1(W^{(1)})(t) = \int_0^t W(s) ds + t(3t-4) \int_0^1 W(s) ds + 6t(1-t) \int_0^1 s W(s) ds.$$

Quelques processus stochastiques utiles

Définition

Le processus défini, pour $t \in [0, 1]$ et $r, d \in \mathbb{N}$, par

$$W_{r,d}(t) = \frac{dC_{r,d+1}(t)}{dt}$$

est appelé un « *processus de Wiener recentré d'ordre $r \times d$* ».

Par exemple

$$W_{0,0}(t) = \frac{dC_{0,1}(t)}{dt} = W(t) - \int_0^1 W(s) ds,$$

$$W_{1,0}(t) = \frac{dC_{1,1}(t)}{dt} = W(t) + (6t - 4) \int_0^1 W(s) ds + (6 - 12t) \int_0^1 s W(s) ds.$$

Cadre général

Modélisation

Pour tout $0 \leq t \leq T$, soit

$$\mathcal{A}(L)X_t = (\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + S_t^\eta + \varepsilon_t \quad \text{avec} \quad S_t^\eta = \rho S_{t-1}^\eta + \eta_t$$

où κ est un indicateur, $|\rho| = 1$ et (ε_t) et (η_t) sont des bruits blancs non corrélés.

Estimation

- θ est estimé par maximum de vraisemblance (sous $\mathcal{H}_0 : \sigma_\eta^2 = 0$).
- α est estimé par moindres carrés sur les résidus issus de la régression précédente.
- On définit les **processus des sommes partielles** et la statistique de test

$$S_t = \sum_{k=1}^t \widehat{\varepsilon}_k, \quad Q_t = \sum_{k=1}^t \widehat{\varepsilon}_k^2 \quad \text{et} \quad \widehat{K}_T = \frac{1}{TQ_T} \sum_{t=1}^T S_t^2.$$

Principaux résultats

Proposition

Sous $\mathcal{H}_0 : \sigma_\eta^2 = 0$,

$$\widehat{K}_T \xrightarrow{\mathcal{L}} \int_0^1 B_r^2(s) ds.$$

De plus, sous $\mathcal{H}_1^+ : \sigma_\eta^2 > 0$ et $\rho = 1$,

$$\widehat{K}_T \xrightarrow{\mathcal{P}} +\infty \quad \text{et} \quad \frac{\widehat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 C_{r,1}^2(s) ds}{\int_0^1 W_{r,0}^2(s) ds}.$$

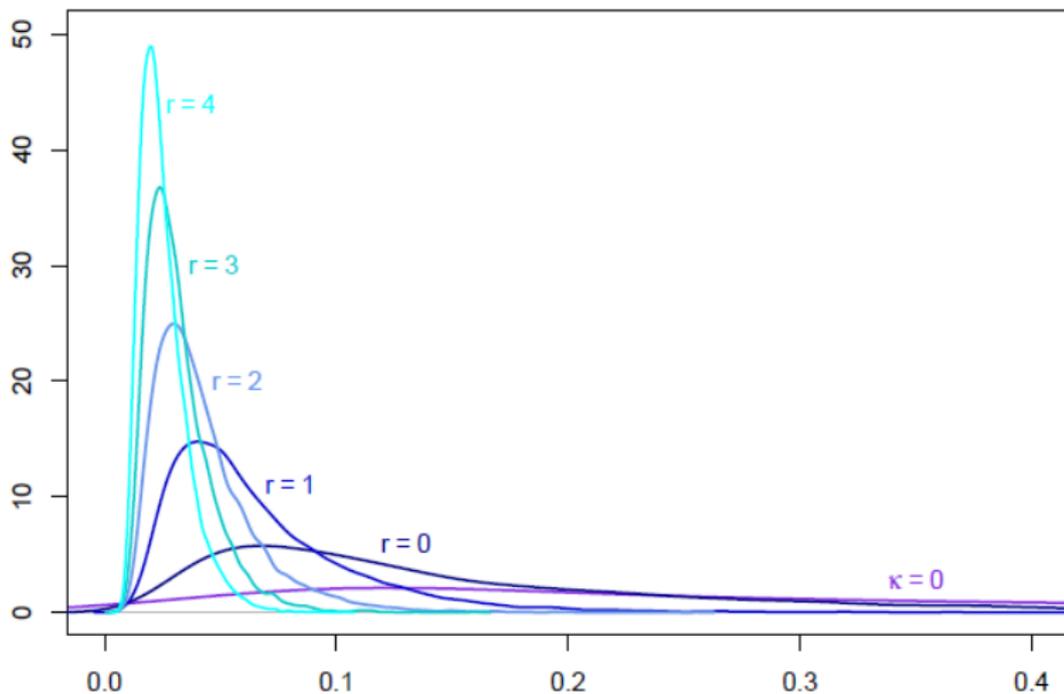
Enfin, sous $\mathcal{H}_1^- : \sigma_\eta^2 > 0$ et $\rho = -1$,

$$\widehat{K}_T \xrightarrow{\mathcal{P}} 0 \quad \text{et, lorsque } \kappa = 0, \quad T\widehat{K}_T \xrightarrow{\mathcal{L}} \frac{2\sigma_\varepsilon^2 \int_0^1 W_\varepsilon^2(s) ds + \sigma_\eta^2 \int_0^1 W_\eta^2(s) ds}{2\sigma_\eta^2 \int_0^1 W_\eta^2(s) ds}.$$

Les processus apparaissant dans les distributions limites sont identifiables et construits sur le processus de Wiener (intégrés, généralisés, ponts).

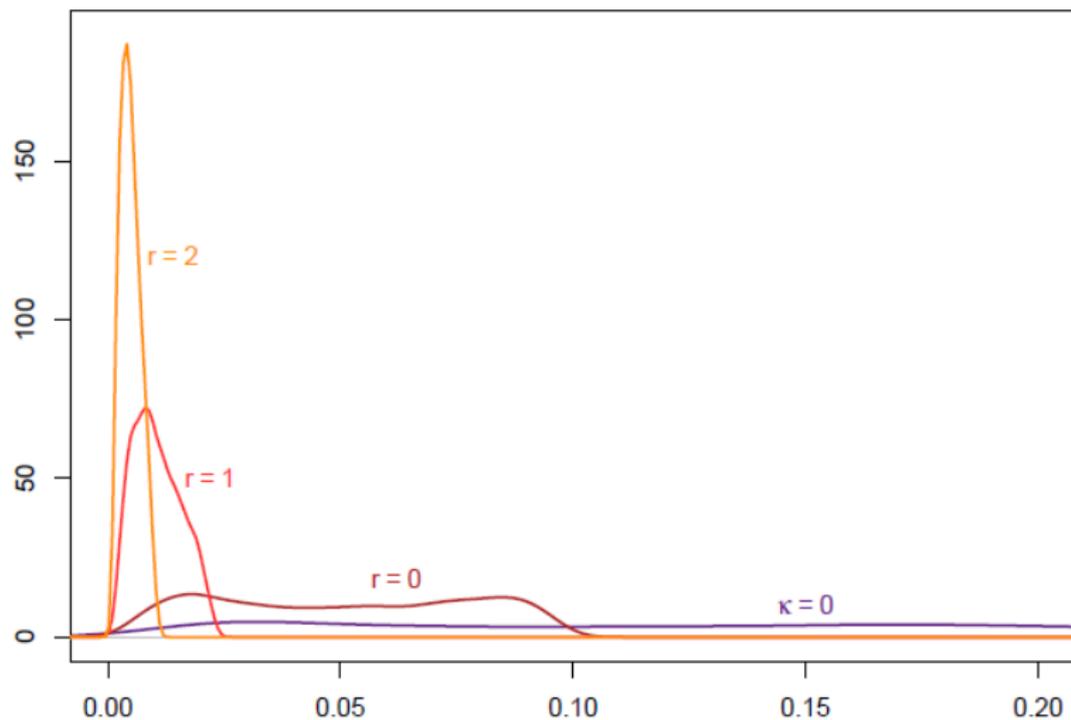
Distribution asymptotique

Sous $\mathcal{H}_0 : \sigma_{\eta}^2 = 0$



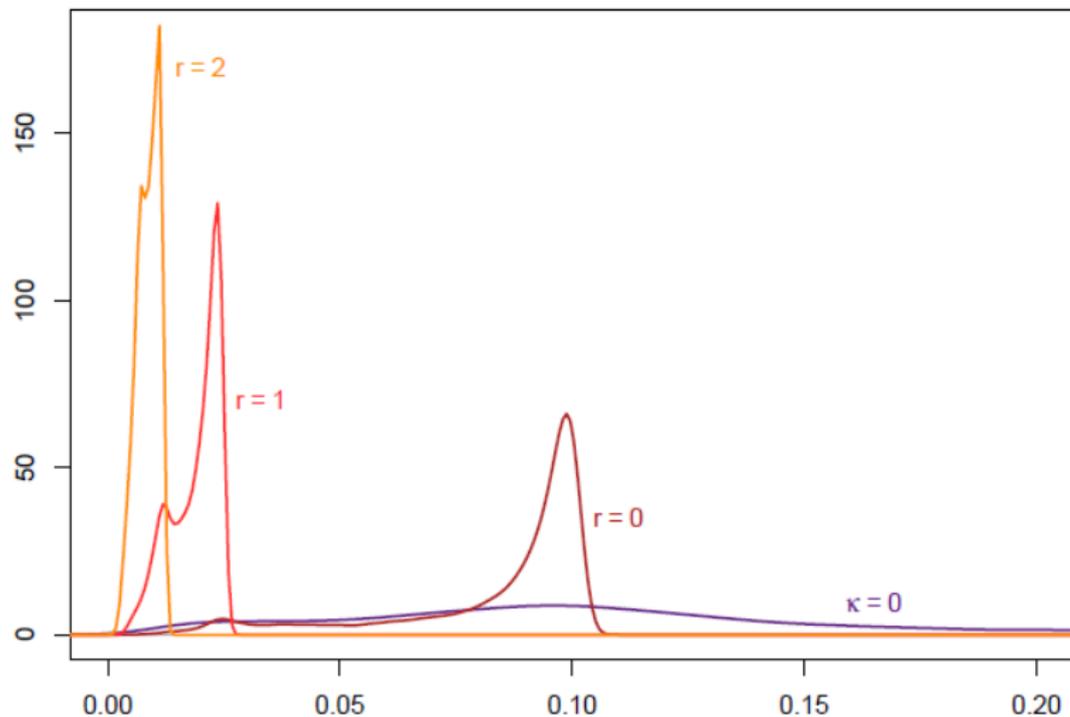
Distribution asymptotique

Sous \mathcal{H}_1^+ : " $\sigma_\eta^2 > 0, \rho = 1$ et $d = 1$ "



Distribution asymptotique

Sous \mathcal{H}_1^+ : " $\sigma_\eta^2 > 0$, $\rho_1 = \rho_2 = 1$ et $d = 2$ "



Sommaire

- 1 Cadre de travail
- 2 Focus sur un cas particulier
 - Bruit blanc ou marche aléatoire ?
 - Principes d'invariance
 - Distribution asymptotique de \widehat{K}_T
- 3 Conclusion

Bruit blanc ou marche aléatoire ?

Modélisation

Pour tout $0 \leq t \leq T$, soit

$$\mathcal{A}(L)X_t = (\alpha_0 + \alpha_1 t_T + \dots + \alpha_r t_T^r) \mathbb{I}_{\{\kappa \neq 0\}} + S_t^\eta + \varepsilon_t \quad \text{avec} \quad S_t^\eta = \rho S_{t-1}^\eta + \eta_t$$

où κ est un indicateur, $|\rho| = 1$ et (ε_t) et (η_t) sont des bruits blancs non corrélés.

Cas particulier

Considérons le cas $\rho = 0$ et $\kappa = 0$. Pour tout $0 \leq t \leq T$, soit

$$X_t = S_t^\eta + \varepsilon_t \quad \text{avec} \quad S_t^\eta = \rho S_{t-1}^\eta + \eta_t.$$

Estimation résiduelle

On a tout simplement $\hat{\varepsilon}_t = X_t$ et donc, pour tout $0 \leq t \leq T$,

$$S_t = \sum_{k=0}^t X_k, \quad Q_t = \sum_{k=0}^t X_k^2 \quad \text{et} \quad \hat{K}_T = \frac{1}{TQ_T} \sum_{t=0}^T S_t^2.$$

Sous \mathcal{H}_0 : " $\sigma_\eta^2 = 0$ "

Pour tout $0 \leq t \leq T$,

$$X_t = \varepsilon_t.$$

Le processus est **stationnaire**.

Sous \mathcal{H}_1^+ : " $\sigma_\eta^2 > 0$ et $\rho = 1$ "

Pour tout $0 \leq t \leq T$,

$$X_t = \sum_{k=0}^t \eta_k + \varepsilon_t.$$

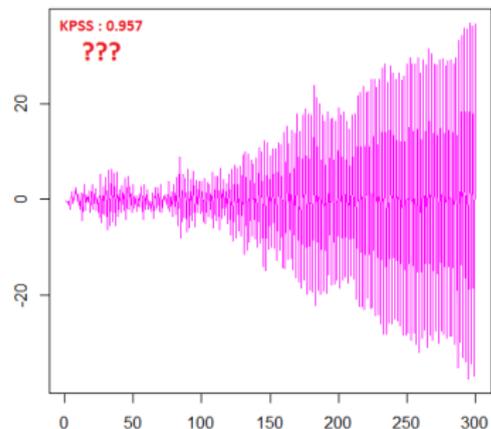
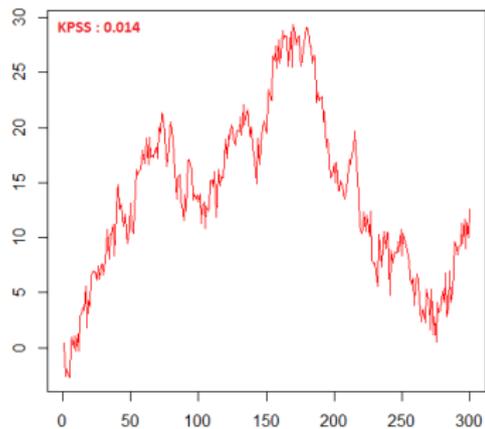
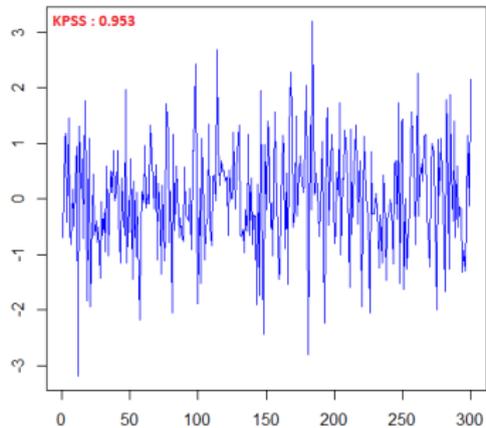
Le processus **n'**est **pas stationnaire**.

Sous \mathcal{H}_1^- : " $\sigma_\eta^2 > 0$ et $\rho = -1$ "

Pour tout $0 \leq t \leq T$,

$$X_t = \sum_{k=0}^t (-1)^{t-k} \eta_k + \varepsilon_t.$$

Le processus **n'**est **pas stationnaire**.



Principes d'invariance associés à la marche aléatoire

Théorème (Donsker)

Pour tout $1 \leq t \leq T$, soit la marche aléatoire

$$S_t^\eta = \sum_{k=0}^t \eta_k$$

où (η_t) est un bruit blanc de variance $0 < \sigma_\eta^2 < \infty$. Alors, pour $0 \leq \tau \leq 1$,

$$\frac{S_{\lfloor T\tau \rfloor}^\eta}{\sigma_\eta \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau).$$

Continuité

$$\frac{1}{\sigma_\eta T^{3/2}} \sum_{t=0}^T S_t^\eta = \sum_{t=0}^T \int_{\frac{t}{T}}^{\frac{t+1}{T}} \frac{S_{\lfloor Ts \rfloor}^\eta}{\sigma_\eta \sqrt{T}} ds \xrightarrow{\mathcal{L}} \int_0^1 W(s) ds \equiv W^{(1)}(1).$$

Principes d'invariance associés à la marche aléatoire alternée

(Dedecker-Rio)

Pour tout $1 \leq t \leq T$, soit la marche aléatoire

$$S_t^\eta = \sum_{k=0}^t (-1)^{t-k} \eta_k.$$

Alors, pour $0 \leq \tau \leq 1$,

$$\frac{S_{[T\tau]}^\eta}{\sigma_\eta \sqrt{T}} \xrightarrow{\mathcal{L}} W(\tau)$$

car le processus $((-1)^t \eta_t)$ est stationnaire de variance σ_η^2 .

Compensation

Supposons que T est pair. Alors,

$$\frac{1}{\sigma_\eta \sqrt{T}} \sum_{t=0}^T S_t^\eta = \frac{1}{\sigma_\eta \sqrt{T}} \sum_{t=0}^{T/2} \eta_{2t} \xrightarrow{\mathcal{L}} W\left(\frac{1}{2}\right) \stackrel{\mathcal{L}}{=} \frac{W(1)}{\sqrt{2}}.$$

Distribution asymptotique de \hat{K}_T

Sous \mathcal{H}_0 : " $\sigma_\eta^2 = 0$ "

Pour tout $0 \leq t \leq T$,

$$S_t = \sum_{k=0}^t \varepsilon_k \quad \text{et} \quad Q_t = \sum_{k=0}^t \varepsilon_k^2.$$

Ainsi,

$$\frac{1}{\sigma_\varepsilon^2 T^2} \sum_{t=0}^T S_t^2 = \sum_{t=0}^T \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left(\frac{S_{[Ts]}}{\sigma_\varepsilon \sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^1 W^2(s) ds,$$

$$\frac{Q_T}{T} \rightarrow \sigma_\varepsilon^2 \quad \text{p.s.}$$

Finalement,

$$\hat{K}_T \xrightarrow{\mathcal{L}} \int_0^1 W^2(s) ds.$$

Distribution asymptotique de \hat{K}_T

Sous \mathcal{H}_1^+ : " $\sigma_\eta^2 > 0$ et $\rho = 1$ "

Pour tout $0 \leq t \leq T$,

$$S_t = \sum_{k=0}^t (S_k^\eta + \varepsilon_k) \quad \text{et} \quad Q_t = \sum_{k=0}^t (S_k^\eta + \varepsilon_k)^2.$$

Ainsi,

$$\frac{1}{\sigma_\eta^2 T^4} \sum_{t=0}^T S_t^2 = \sum_{t=0}^T \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left(\frac{S_{[Ts]}}{\sigma_\eta T^{3/2}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^1 \left(\int_0^s W(u) du \right)^2 ds,$$

$$\frac{Q_T}{\sigma_\eta^2 T^2} \sim \frac{1}{\sigma_\eta^2 T^2} \sum_{t=0}^T (S_t^\eta)^2 = \sum_{t=0}^T \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left(\frac{S_{[Ts]}^\eta}{\sigma_\eta \sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \int_0^1 W^2(s) ds.$$

Finalement,

$$\hat{K}_T \xrightarrow{\mathcal{P}} +\infty \quad \text{et} \quad \frac{\hat{K}_T}{T} \xrightarrow{\mathcal{L}} \frac{\int_0^1 W^{(1)2}(s) ds}{\int_0^1 W^2(s) ds}.$$

Distribution asymptotique de \widehat{K}_T

Sous \mathcal{H}_1^- : " $\sigma_\eta^2 > 0$ et $\rho = -1$ "

Pour tout $0 \leq t \leq T$,

$$S_t = \sum_{k=0}^t (S_k^\eta + \varepsilon_k) \quad \text{et} \quad Q_t = \sum_{k=0}^t (S_k^\eta + \varepsilon_k)^2.$$

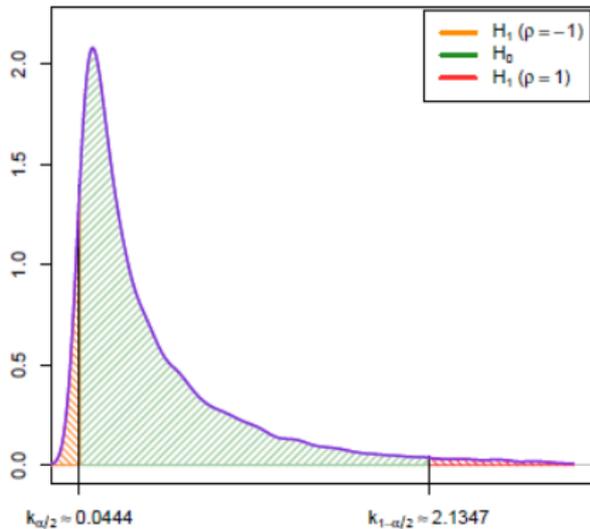
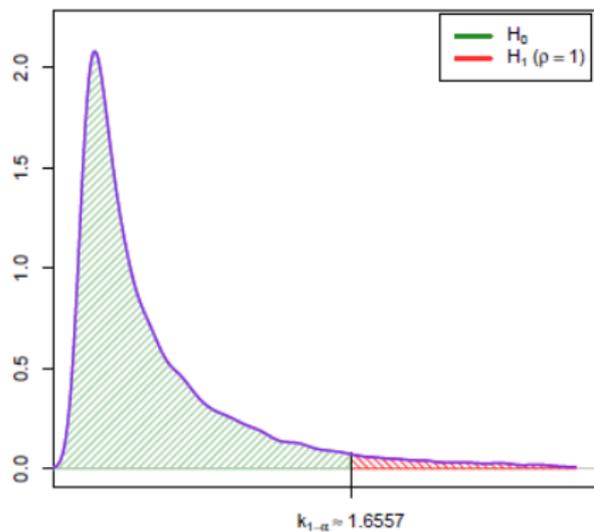
Ainsi,

$$\frac{1}{T^2} \sum_{t=0}^T S_t^2 = \sum_{t=0}^T \int_{\frac{t}{T}}^{\frac{t+1}{T}} \left(\frac{S_{[Ts]}}{\sqrt{T}} \right)^2 ds \xrightarrow{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 W_\varepsilon^2(s) ds + \frac{\sigma_\eta^2}{2} \int_0^1 W_\eta^2(s) ds,$$

$$\frac{Q_T}{\sigma_\eta^2 T^2} \xrightarrow{\mathcal{L}} \int_0^1 W_\eta^2(s) ds.$$

Finalement,

$$\widehat{K}_T \xrightarrow{\mathcal{P}} 0 \quad \text{et} \quad T\widehat{K}_T \xrightarrow{\mathcal{L}} \frac{2\sigma_\varepsilon^2 \int_0^1 W_\varepsilon^2(s) ds + \sigma_\eta^2 \int_0^1 W_\eta^2(s) ds}{2\sigma_\eta^2 \int_0^1 W_\eta^2(s) ds}.$$



Simulations en présence de racine unitaire négative

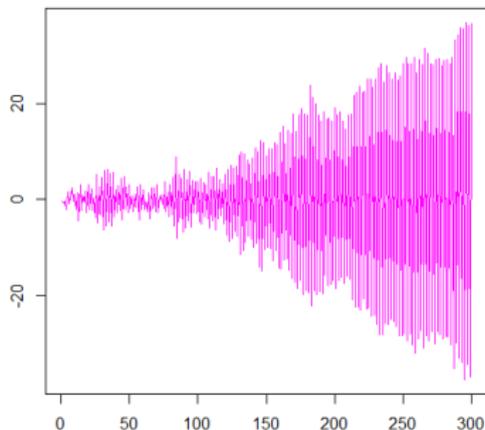
- KPSS : 0.957
- Correction apportée : 0.058

Sommaire

- 1 Cadre de travail
- 2 Focus sur un cas particulier
- 3 Conclusion

Conclusion

- Généralisation apportée aux cas $p > 0$, $\kappa \neq 0$ et $r > 0$.
- Gestion de la multi-intégration (processus ARIMA avec $d > 0$).
- Modélisation des signaux à haute fréquence par une racine unitaire négative ?
- Estimation jointe de α et de θ pour une amélioration conséquente du test LMC.



Merci de votre attention !