

Racines unitaires

Chapitre 3

Économétrie des séries temporelles

Structure du cours

Marche aléatoire et marche aléatoire avec dérive

Processus stationnaires de tendance versus processus stationnaires de différence

Régression fallacieuse

Test de racine unitaire

Tests de Dickey-Fuller

Tests de Dickey-Fuller augmentés

Exemple empirique

Conclusion

Marche aléatoire

De nombreuses séries temporelles macroéconomiques apparaissent non stationnaires lorsqu'elles sont analysées par des modèles AR :

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

Marche aléatoire ($\alpha = 1$) :

$$y_t = y_0 + \sum_{j=1}^t \varepsilon_j$$

Persistante des chocs aléatoires – **mémoire infinie**

La marche aléatoire est courante dans les marchés financiers et des changes bilatéraux
– taux de change USD/GBP.

Marche aléatoire – ne peut pas exploiter l'historique passé de (y_t, ε_t) pour systématiquement réaliser des profits en spéculant sur le taux de change futur.

Si $\alpha > 1$, y_t est explosif.

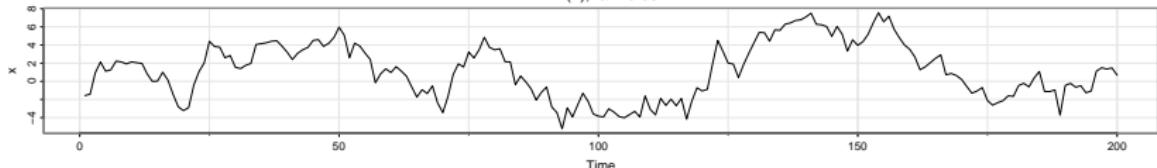
Taux de change USD/GBP

```
GBP <- getSymbols(Symbols = 'GBP=X', src = 'yahoo',
                   from='2004-12-01', to='2024-12-01', auto.assign = FALSE)
GBP$Date = as.Date.character(GBP$Date, format = "%Y/%m/%d")
GBP.date <- GBP$Date
GBP.p <- GBP$GBP.X.Close
plot(GBP.p, main="GBP=X")
```

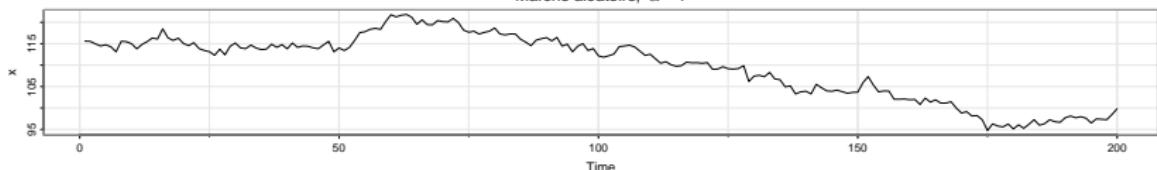


Marche aléatoire et processus explosif

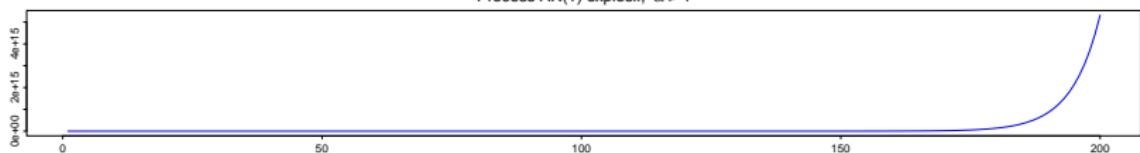
AR(1), $\alpha = 0.95$



Marche aléatoire, $\alpha = 1$



Process AR(1) explosif, $\alpha > 1$



Marche aléatoire avec dérive

$$y_t = \mu + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

μ est le terme de dérive.

Par substitution récursive :

$$y_t = y_0 + \mu t + \sum_{j=1}^t \varepsilon_j$$

La tendance déterministe est représentée par μt .

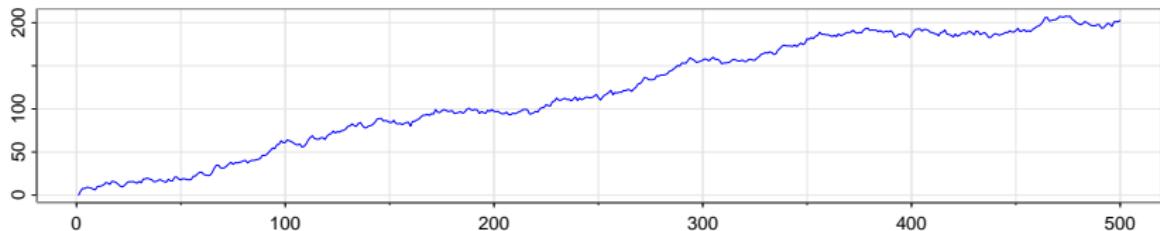
$$\mathbb{E}(y_t) = y_0 + \mu t$$

$$\mathbb{V}(y_t) = \gamma_t(0) = t\sigma^2$$

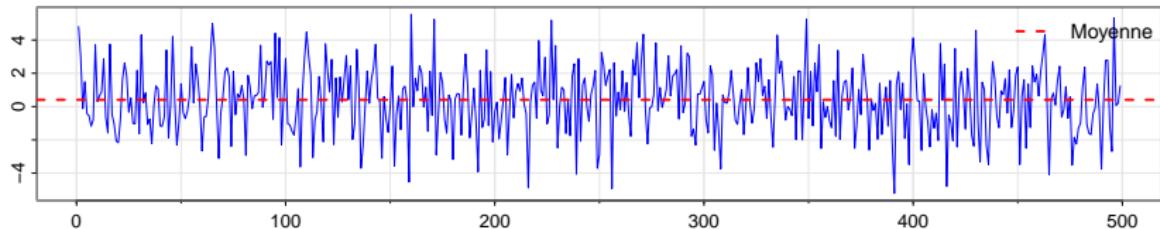
$$\text{cov}(y_t, y_{t-h}) = \gamma_t(h) = (t - h)\sigma^2$$

Marche aléatoire avec dérive - graph

Process simulé



Process différencié simulé



Processus stationnaires de différence (DS) et de tendance (TS)

Les séries temporelles non stationnaires peuvent être classées en deux grandes catégories :

- ▶ **Processus stationnaires de différence (DS - Difference Stationary)**
- ▶ **Processus stationnaires de tendance (TS - Trend Stationary)**

Processus stationnaires de différence (DS)

Un processus DS est non stationnaire, mais peut devenir stationnaire après avoir pris une ou plusieurs différences (d).

Exemple : Marche aléatoire avec ou sans dérive (cf. slides précédents)

$$y_t = \mu + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

1. Si $\mu = 0$, c'est une **marche aléatoire pure**.
2. Si $\mu \neq 0$, c'est une **marche aléatoire avec dérive**.

Après différenciation ($\Delta y_t = y_t - y_{t-1}$) :

$$\Delta y_t = \mu + \varepsilon_t.$$

- ▶ Δy_t est **stationnaire** (variance et moyenne constantes dans le temps).

Processus stationnaires de tendance (TS)

Un processus TS est stationnaire autour d'une tendance déterministe. Contrairement aux processus DS, la non-stationnarité est éliminée en supprimant la **tendance déterministe**.

Exemple : Modèle TS

$$y_t = \mu + \beta t + u_t, \quad u_t \sim \text{NID}(0, \sigma^2)$$

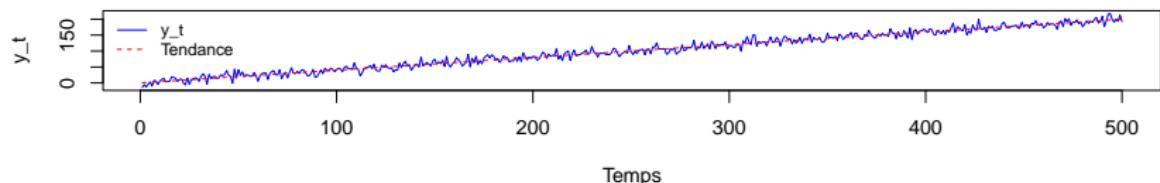
1. μ est une constante (intercept),
2. βt représente une **tendance déterministe**,
3. u_t est un processus **stationnaire** autour de la tendance.

La série devient stationnaire après dé-trendation ($y_t - E[y_t]$) :

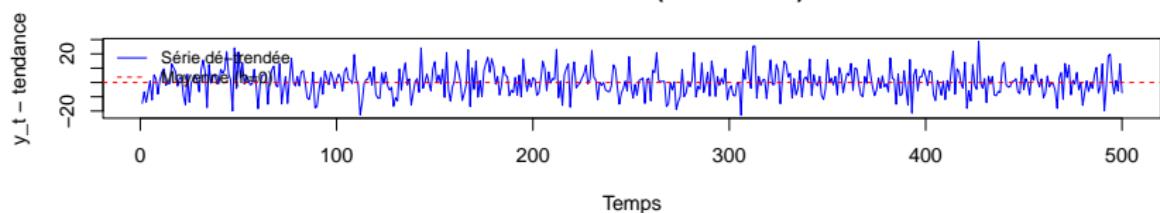
$$y_t - (\mu + \beta t) = u_t.$$

Stationnaire en tendance

Processus stationnaire en tendance (TS)



Série dé-trendée (stationnaire)



Différences principales entre DS et TS

Caractéristique	Processus DS	Processus TS
Source de non-stationnarité	Tendance stochastique	Tendance déterministe
Transformation pour stationnarité	Différenciation	Dé-trendation
Exemple	Marche aléatoire, marche aléatoire avec dérive	Série avec tendance linéaire
Stationnarité	Stationnaire après d différences	Stationnaire après suppression de la tendance

Types de modèles univariés

$$y_t = \mu + \alpha y_{t-1} + \beta t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

Paramètre	Description	Propriétés
$\mu \neq 0; \alpha < 1; \beta \neq 0$	Tendance déterministe avec composante AR(1) stationnaire	I(0)
$\mu \neq 0; \alpha = 1; \beta \neq 0$	Marche aléatoire avec dérive et tendance quadratique	I(1)
$\mu \neq 0; \alpha = 1; \beta = 0$	Marche aléatoire avec dérive	I(1)
$\mu \neq 0; \alpha = 0; \beta \neq 0$	Tendance déterministe	I(0)
$\mu = 0; \alpha = 1; \beta = 0$	Marche aléatoire pure	I(1)

Pour le cas marche aléatoire avec dérive et tendance quadratique

$$y_t = \mu + \alpha y_{t-1} + \beta t + \varepsilon_t$$

Donc :

$$\Delta y_t = \mu + \beta t + \varepsilon_t \sim I(0)$$

Si $\Delta y_t \sim I(0) \Rightarrow \sum_{i=1}^t [\Delta y_i] \sim I(1)$

$$y_t = \mu t + \beta t^2 + \sum_{i=1}^t \varepsilon_i$$

Donc, une tendance quadratique dans le niveau.

Peu probable de se produire en pratique, sauf localement.

Processus à racine unitaire

Le processus à racine unitaire est donné par :

$$\begin{aligned}y_t &= \mu + y_{t-1} + \varepsilon_t \\&= \mu t + y_0 + \sum_{j=1}^t \varepsilon_j\end{aligned}$$

où $\sum_{j=1}^t \varepsilon_j$ est une marche aléatoire ;

y_0 est la condition initiale ;

et μt est le terme de dérive.

Problème : La théorie de la distribution pour les estimateurs et les statistiques de test diffère du cas transversal/stationnaire – en raison de la dépendance temporelle.

2. Régression fallacieuse

Deux séries non apparentées $I(1)$.

Régression de y sur $z \Rightarrow$ relation apparemment significative.

Transformations pour induire la stationnarité :

- ▶ Différenciation
- ▶ Suppression de la tendance déterministe
- ▶ Cointégration

Rappel

Régression OLS standard pour des données stationnaires :

$$y_t = \beta x_t + u_t, \quad u_t \sim \text{NID}(0, \sigma^2) \quad (1)$$

$$\hat{\beta} = \left[\sum_{t=1}^T x_t^2 \right]^{-1} \sum_{t=1}^T x_t y_t = (X' X)^{-1} (X' y)$$

$$\mathbb{E} (\hat{\beta} | x_1, \dots, x_T) = \beta$$

$$\mathbb{V} (\hat{\beta} | x_1, \dots, x_T) = \sigma^2 (X' X)^{-1}$$

Résultats asymptotiques - convergence

$$\widehat{\beta} - \beta = \frac{\frac{1}{T} \sum_{t=1}^T x_t u_t}{\frac{1}{T} \sum_{t=1}^T x_t^2} \quad (2)$$

En utilisant la loi faible des grands nombres :

$$\frac{1}{T} \sum_{t=1}^T x_t u_t \xrightarrow{P} E(x_t u_t) = 0$$

$$\frac{1}{T} \sum_{t=1}^T x_t^2 \xrightarrow{P} E(x_t^2) = Q \neq 0$$

Donc :

$$\widehat{\beta} - \beta \xrightarrow{P} Q^{-1} \times 0 = 0 \quad (3)$$

ce qui établit la convergence.

Résultats asymptotiques - normalité

$$\sqrt{T} (\hat{\beta} - \beta) = \sqrt{T} \frac{\frac{1}{T} \sum_{t=1}^T x_t u_t}{\frac{1}{T} \sum_{t=1}^T x_t^2} \quad (4)$$

Comme $\frac{1}{T} \sum_{t=1}^T x_t^2 \xrightarrow{P} \mathbf{Q}$ (une constante) et $\frac{1}{\sqrt{T}} \sum_{t=1}^T x_t u_t$ a une distribution normale limite, alors $\sqrt{T} (\hat{\beta} - \beta)$ est asymptotiquement normalement distribué.

La preuve de la convergence nécessite $\text{plim} (1/T) (X' X) = \mathbf{Q}$ – avec l'augmentation de l'information de l'échantillon, les moments de l'échantillon convergent vers les moments de la population (fixes).

Par conséquent, les moments de la population doivent être stationnaires.

Que se passe-t-il si les données ne sont pas stationnaires ?

Données non stationnaires – régression fallacieuse

DGP :

$$y_t = y_{t-1} + u_t$$

$$x_t = x_{t-1} + v_t$$

$$\mathbb{E}(u_t v_s) = 0, \forall t, s; \quad \mathbb{E}(u_t u_k) = \mathbb{E}(v_t v_k) = 0, \forall k \neq 0$$

$$y_0 = x_0 = 0$$

c'est-à-dire, x_t et y_t sont des marches aléatoires non corrélées.

Hypothèse économique :

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{5}$$

Nous devrions trouver $\beta_1 = \frac{\partial y_t}{\partial x_t} \xrightarrow{P} 0$.

Yule (1926) – les corrélations fallacieuses peuvent persister dans de grands échantillons.

Sous $H_0 : \beta_1 = 0$, nous nous attendrions à ce que $E(\widehat{\beta}_1) = 0$.

Pour des régresseurs non stationnaires, cela ne tient pas.

Les statistiques t conventionnelles sur $\widehat{\beta}_0$ et $\widehat{\beta}_1$ n'ont pas de distributions t .

Elles divergent en distribution à mesure que T augmente. Par conséquent, pour $H_0 : \beta_1 = 0$, nous rejeterions plus fréquemment à mesure que T augmente.

Problème d'équilibre : $y_t \sim I(1)$ et ε_t supposé $\sim I(0)$.

Alors $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ pourrait être bien défini pour β_1 non nul comme $x_t \sim I(1)$.

Lorsque $\beta_1 = 0$, $\{\varepsilon_t\}$ doit être $I(1)$ – ce qui viole les hypothèses.

Pour les données non stationnaires, $\widehat{\beta}_1$ ne converge pas vers 0, mais converge plutôt vers une variable aléatoire.

$$\widehat{\beta}_1 = \left[T^{-1} \sum_{t=1}^T (x_t - \bar{x})^2 \right]^{-1} T^{-2} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})$$

Le numérateur et le dénominateur convergent faiblement vers des fonctionnelles du mouvement brownien.

Comme $x \sim I(1)$, la variance de l'échantillon = $f(T)$.

Régression déséquilibrée, $\varepsilon \sim I(1)$, la variance de l'échantillon = $f(T)$.

$\widehat{\beta}_0$ et $\widehat{\beta}_1$ minimisent $\widehat{\sigma}_\varepsilon^2$.

On observe que $\widehat{\sigma}_{\beta_1} = \widehat{\sigma}_\varepsilon \left(\sum_{t=1}^T (x_t - \bar{x})^2 \right)^{-\frac{1}{2}}$ est une fonction de T .

Par conséquent, $t = \frac{\widehat{\beta}_1}{\widehat{\sigma}_{\beta_1}}$ diverge lorsque $T \rightarrow \infty$.

3. Tests de racine unitaire

Dickey-Fuller test

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, \sigma^2)$$

Définissons $\gamma = \alpha - 1$:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

Si racine unitaire, $\alpha = 1 \Rightarrow \gamma = 0$.

Par conséquent, nous souhaitons tester :

$$H_0 : \gamma = 0$$

Cela suggère OLS et le rejet de l'hypothèse nulle si $\hat{\gamma}$ est significativement négatif.

Problème : il est nécessaire de connaître la distribution de la statistique de test sous l'hypothèse nulle.

Sous l'hypothèse nulle, la série est non stationnaire. Cela implique que la statistique de test n'est pas distribuée asymptotiquement selon une loi normale.

Dickey-Fuller test (suite)

Considérons l'estimateur de $\hat{\gamma}$ lorsque $\gamma = 0$:

$$\hat{\gamma} = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \varepsilon_t}{\frac{1}{T} \sum_{t=1}^T y_{t-1}^2}$$

En utilisant $y_{t-1} = \sum_{j=1}^{t-1} \varepsilon_j$:

$$T\hat{\gamma} = \frac{\frac{1}{T} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \varepsilon_j \right) \varepsilon_t}{\frac{1}{T^2} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \varepsilon_j \right)^2}$$

$\sum_{j=1}^{t-1} \varepsilon_j$ n'est pas ergodique, ni IID, donc la loi des grands nombres (LLN) ne s'applique pas.

$\sum_{j=1}^{t-1} \varepsilon_j \varepsilon_t$ n'est pas une séquence de différences martingales satisfaisant les hypothèses pour le théorème central limite (CLT).

Dickey-Fuller test (suite)

Les distributions limites de $T\hat{\gamma}$ ont été simulées par Dickey et Fuller (JASA, 1979) et MacKinnon (1991).

Ces distributions sont appelées distributions de Dickey-Fuller.

Les logiciels économétriques produisent des valeurs critiques à partir de surfaces de réponse.

Le test de racine unitaire permet de distinguer entre les données I(0) et I(1).

Mais différentes hypothèses nulles sont possibles en fonction des termes déterministes.

Par conséquent, différentes statistiques de test en fonction des hypothèses nulles.

Trois cas à considérer :

Test	Modèle	Hypothèse
$\hat{\tau}$	$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$H_0 : \gamma = 0$
$\hat{\tau}_\mu$	$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$	$H_0 : \mu = 0; \gamma = 0$
$\hat{\tau}_\beta$	$\Delta y_t = \mu + \gamma y_{t-1} + \beta t + \varepsilon_t$	$H_0 : \beta = 0; \gamma = 0$

Les valeurs critiques dépendent de la spécification des hypothèses nulles et alternatives.

Un test unilatéral est généralement utilisé pour maximiser la puissance :

$H_0 : \gamma = 0$ versus $H_1 : \gamma < 0$; car $\gamma > 0 \Rightarrow$ processus explosif.

Distribution	2.5%	5%	10%	50%	90%	95%	97.5%
$N(0, 1)$	-1.96	-1.64	-1.28	0	1.28	1.64	1.96
DF_μ	-3.12	-2.86	-2.57		-0.44	-0.07	0.23
DF_β	-3.66	-3.41	-3.12		-1.25	-0.94	-0.66

Procédure de test

1. Évaluer les données graphiquement.
2. Si une tendance est probable, tester l'hypothèse conjointe $H_0 : \gamma = 0; \beta = 0$. La non-rejection \Rightarrow processus DS sans tendance déterministe.
3. H_0 pourrait être rejeté en raison de :
 - ▶ $\gamma \neq 0; \beta = 0$: Aucun mécanisme générant une tendance.
 - ▶ $\gamma = 0; \beta \neq 0$: Correspond à une tendance quadratique dans le DGP.
 - ▶ $\gamma \neq 0; \beta \neq 0$: Tendance stationnaire ($\gamma < 0$).
4. Si les données ne présentent pas de tendance, tester l'hypothèse conjointe $H_0 : \gamma = 0; \mu = 0$.
5. H_0 pourrait être rejeté en raison de :
 - ▶ $\gamma \neq 0; \mu = 0$: Stationnarité.
 - ▶ $\gamma = 0; \mu \neq 0$: Donnera une dérive notable.
 - ▶ $\gamma \neq 0; \mu \neq 0$: Distribution normale standard asymptotique.
6. Si la moyenne est zéro, sans tendance, tester l'hypothèse $H_0 : \gamma = 0$.

Augmented Dickey-Fuller Test

Le test est considéré sur la base d'un processus AR(1). Si le DGP est un processus AR(p) :

$$y_t = \mu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t$$

mais un modèle AR(1) est ajusté :

$$y_t = \mu + \alpha_1 y_{t-1} + \nu_t$$

alors :

$$\nu_t = \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t$$

ce qui génère une autocorrélation sérielle.

Ajuster un modèle AR(p) en utilisant une stratégie de recherche générale à spécifique.

Commencer avec un grand p et tester vers le bas pour s'assurer qu'il n'y a pas d'autocorrélation sérielle.

Considérons un processus AR(2) :

$$\begin{aligned}y_t &= \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t \\&= \mu + (\alpha_1 + \alpha_2) y_{t-1} - \alpha_2 (y_{t-1} - y_{t-2}) + \varepsilon_t \\\Delta y_t &= \mu + \gamma y_{t-1} + \phi_1 \Delta y_{t-1} + \varepsilon_t\end{aligned}$$

où $\gamma = \alpha_1 + \alpha_2 - 1$ et $\phi_1 = -\alpha_2$.

Le modèle de Dickey-Fuller est augmenté par Δy_{t-1} .

Tester $H_0 : \gamma = 0$ pour $\widehat{\tau}_\mu$ comme avant.

Modèle ADF(p) avec intercept et tendance :

$$\Delta y_t = \mu + \gamma y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \beta t + \varepsilon_t$$

où

$$\begin{aligned}\gamma &= \alpha_1 + \alpha_2 + \cdots + \alpha_p - 1, \\ \phi_1 &= -(\alpha_2 + \cdots + \alpha_p) \\ &\vdots \\ \phi_{p-1} &= -\alpha_p\end{aligned}$$

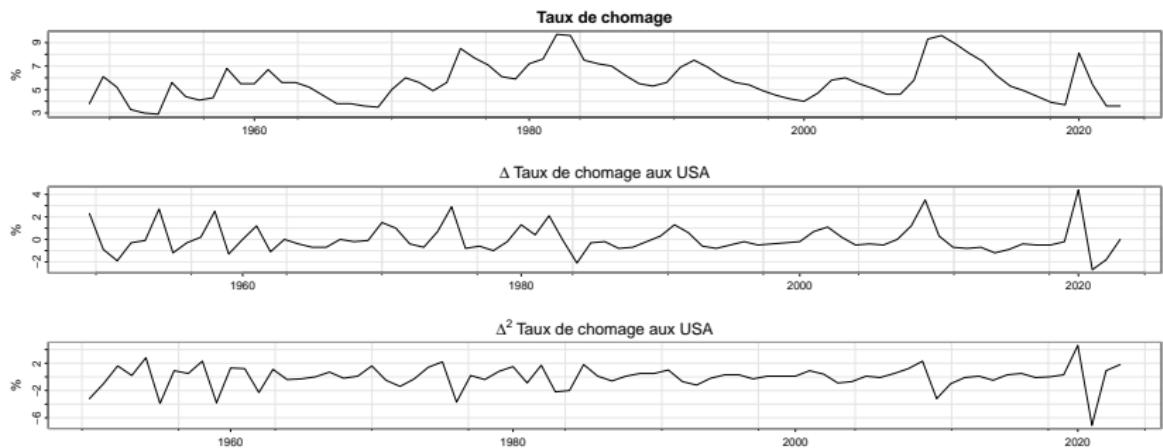
4. Illustration empirique

Données

```
USunrate = fredr(series_id = "UNRATE",
                   observation_start = as.Date("1948-01-01"),
                   observation_end = as.Date("2023-10-01"),
                   frequency = "a")
x = USunrate$value
xDate = USunrate$date
```

$USunrate$, $\Delta USunrate$, $\Delta^2 USunrate$

```
par(mfrow=c(3,1))
tsplot(xDate,x,
main="Taux de chomage", ylab = "%", xlab=' ')
# Si on différencie la série 1 fois :
tsplot(xDate[-1],diff(x),main= expression(Delta ~ "Taux de chomage aux USA"),
       ylab = "%", xlab=' ')
# Si on différencie la série 2 fois :
tsplot(xDate[-1][-1],diff(diff(x)),main=
       expression(Delta^2 ~ "Taux de chomage aux USA"),ylab = "%",xlab=' ')
```



Tester la stationnarité

ADF {tseries}

```
adf.test(x, alternative = "stationary")
```

```
##  
##  Augmented Dickey-Fuller Test  
##  
## data:  x  
## Dickey-Fuller = -2.8082, Lag order = 4, p-value = 0.2456  
## alternative hypothesis: stationary
```

```
adf.test(diff(x), alternative = "stationary")
```

```
##  
##  Augmented Dickey-Fuller Test  
##  
## data:  diff(x)  
## Dickey-Fuller = -4.8096, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(diff(diff(x)))
```

```
##  
##  Augmented Dickey-Fuller Test  
##  
## data:  diff(diff(x))  
## Dickey-Fuller = -6.0254, Lag order = 4, p-value = 0.0001
```

Tester la stationnarité

ADF {urca}

```
w0=ur.df(x, type = "drift", selectlags = "AIC")
summary(w0)$teststat ; summary(w0)$cval
```

```
##           tau2      phi1
## statistic -3.631277 6.626363

##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

```
w1=ur.df(diff(x), type = "none", selectlags = "AIC")
summary(w1)$teststat ; summary(w1)$cval
```

```
##           tau1
## statistic -7.420607

##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

```
w2=ur.df(diff(diff(x)), type = "none", selectlags = "AIC")
summary(w2)$teststat ; summary(w2)$cval
```

```
##           tau1
## statistic -10.36705

##      1pct  5pct 10pct
```

Changement structurel

```
za_test = ur.za(x)
za_summary <- summary(za_test)
za_summary@teststat
```

```
## [1] -4.084208
```

```
za_summary@cval
```

```
## [1] -5.34 -4.80 -4.58
```

```
za_summary@bpoint
```

```
## [1] 26
```

Attention - ruptures structurelles

Tester la présence de racines unitaires en présence de ruptures structurelles, par exemple, Perron (1989).

Un processus $I(0)$ avec une rupture est difficile à distinguer d'un processus $I(1)$. Un grand coefficient sur la variable dépendante retardée (LDV) est nécessaire pour décaler vers une nouvelle moyenne.

Exemple :

Générer des données données par :

$$\begin{aligned}y_t &= 0.5y_{t-1} + \varepsilon_t && \text{pour } t = 1, \dots, 50 \\&= 4 + 0.5y_{t-1} + \varepsilon_t && \text{pour } t = 51, \dots, 100\end{aligned}$$

où $\varepsilon_t \sim NID(0, 1)$

Ajuster un modèle AR(1) aux données :

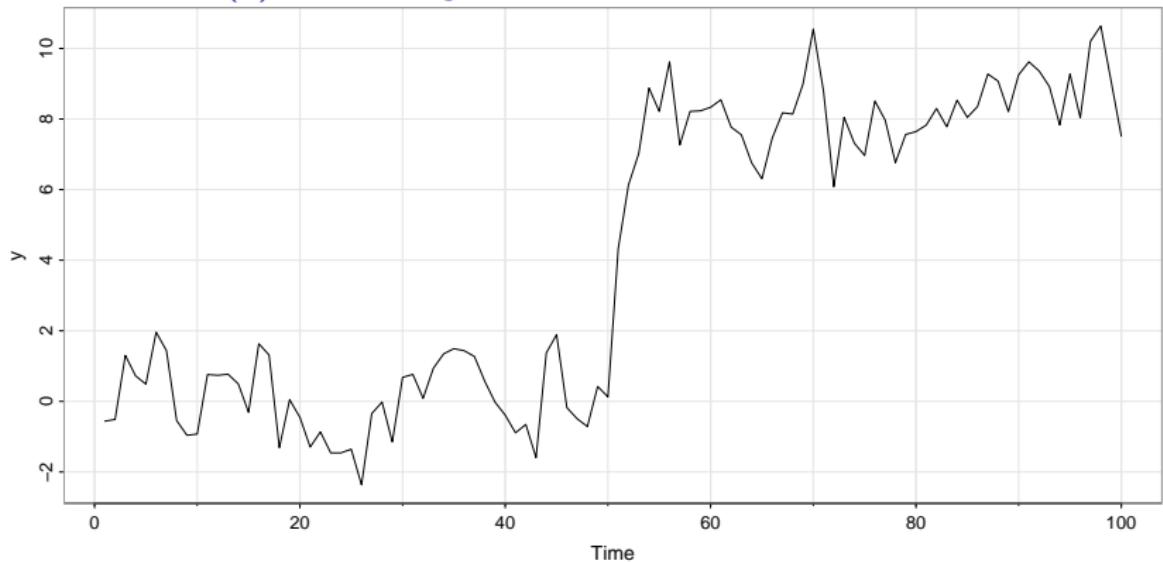
$$y_t = 0.9606y_{t-1} + 3.8998$$

(0.0000) (0.1060)

(pvalues)

Un coefficient proche de l'unité sur la variable dépendante retardée (LDV).

Processus AR(1) avec changement structurel



Ajustment AR(1)

```
fit <- arima(y, order = c(1, 0, 0))
print(fit)

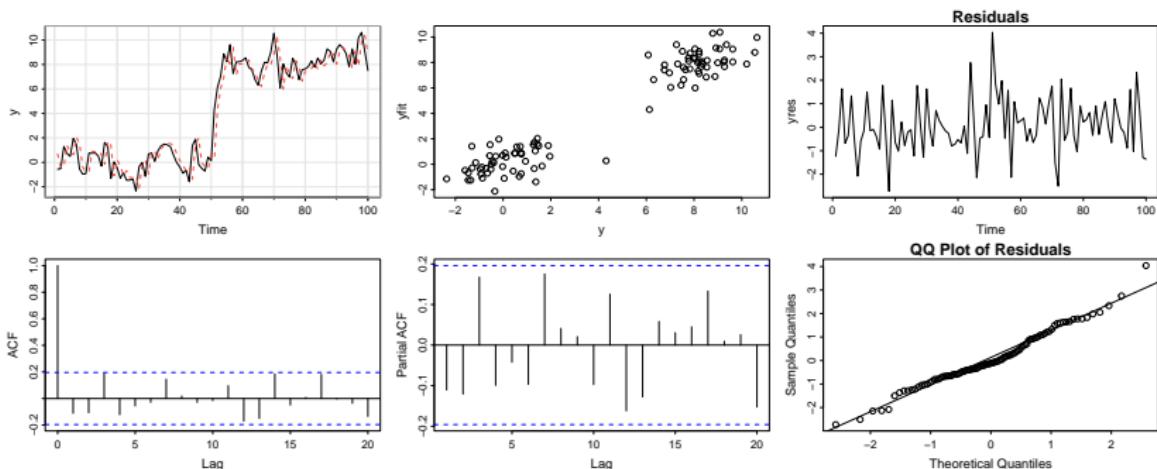
##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##         ar1  intercept
##       0.9606     3.8998
##  s.e.  0.0249     2.4127
##
## sigma^2 estimated as 1.341:  log likelihood = -157.85,  aic = 321.69
##         ar1 intercept
##  0.9606238 3.8997618

## [1] "p_values"

##         ar1 intercept
##  0.0000000 0.1060181
```

Graph des y_t , \hat{y}_t et ε_t

```
par(mfrow = c(2, 3))
tsplot(y)
points(yfit, type = "l", col = 2, lty = 2)
plot(y, yfit)
plot(yres, main = "Residuals")
acf(yres, main = "ACF of Residuals")
pacf(yres, main = "PACF of Residuals")
qqnorm(yres, main = "QQ Plot of Residuals")
qqline(yres)
```



5. Conclusion

Les tests de racine unitaire - les estimateurs sont standard - OLS

MAIS les distributions limites de LR, W, t et F diffèrent des distributions standard χ^2 , $N(0, 1)$

Fonctionnelles des mouvements brownien - Dickey-Fuller

Les distributions sont tabulées et simples à simuler

Problème - les tests de racine unitaire ont une faible puissance lorsque $\alpha \neq 1$ mais proche de 1

Difficile d'inférer les propriétés à long terme à partir d'un échantillon fini

Pour aller plus loin, étendre au cadre multivarié - cointégration