

# 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

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** ( $\alpha = 1$ ) :

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

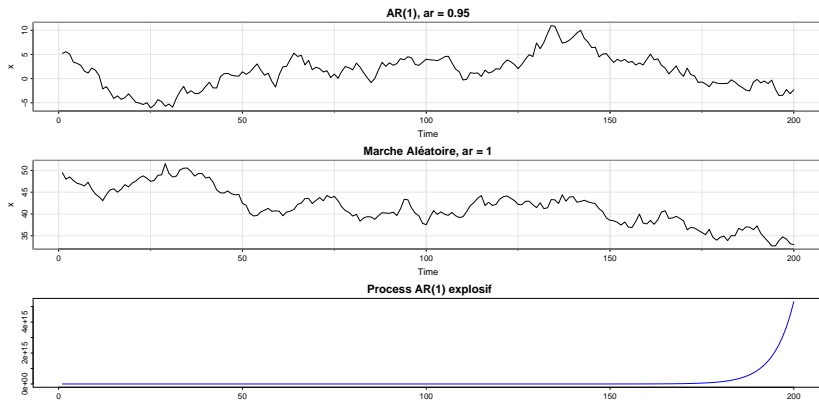
Persistence des chocs aléatoires – **mémoire infinie**

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

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

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

## Marche aléatoire et processus explosif



## Marche aléatoire avec dérive

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

$\mu$  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  $\mu t$ .

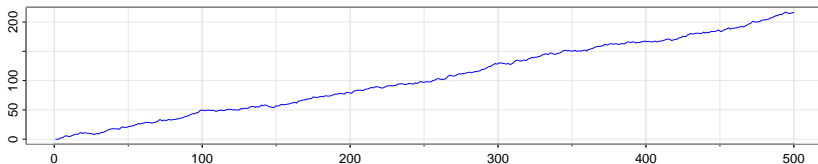
$$E(y_t) = y_0 + \mu t$$

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

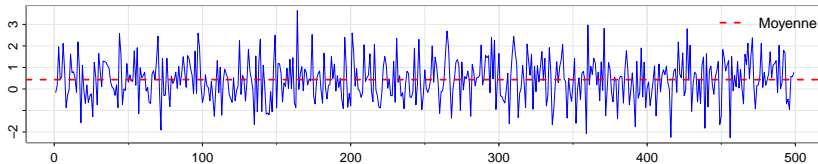
$$\text{cov}(y_t, y_{t-h}) = \gamma_t(h) = (t-h)\sigma^2$$

## Marche aléatoire avec dérive - graph

Process simulé



Process simulé différencié





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

Modèle stationnaire de tendance :

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

Tendance déterministe – rendue stationnaire en soustrayant la tendance.

Tendance exponentielle exhibant une croissance proportionnelle constante :

$$Y_t = e^{\delta t}$$

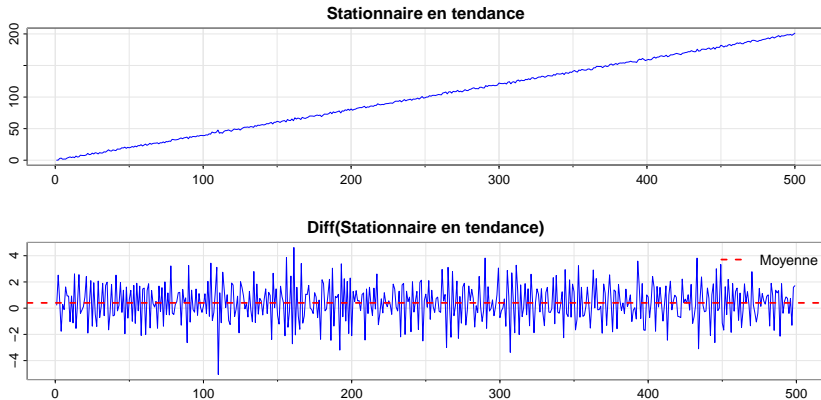
$$\frac{dY}{dt} = \delta Y_t$$

Les données sont généralement en logarithmes,  $y = \log Y$ , – donc tendance linéaire.

Il est souvent difficile de distinguer entre les modèles TS et DS.

Mais la persistance des chocs est différente entre TS et DS – chocs permanents versus chocs transitoires.

## Stationnaire en 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 :

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

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

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

Donc, une tendance quadratique dans le niveau.

Peu probable de se produire en pratique, sauf localement.

## 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 :

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

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

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

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

## Résultats asymptotiques

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

En utilisant la loi faible des grands nombres :

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T x_t u_t &\xrightarrow{p} E(x_t u_t) = 0 \\ \frac{1}{T} \sum_{t=1}^T x_t^2 &\xrightarrow{p} E(x_t^2) = \mathbf{Q} \neq 0 \end{aligned}$$

Donc :

$$\hat{\beta} - \beta \xrightarrow{p} \mathbf{Q}^{-1} \times 0 = 0 \quad (3)$$

ce qui établit la convergence.

## Résultats asymptotiques (suite)

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

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 :**

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

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

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

$$y_0 = x_0 = 0$$

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  $E(\hat{\beta}_1) = 0$ .

Pour des régresseurs non stationnaires, cela ne tient pas.

Les statistiques  $t$  conventionnelles sur  $\hat{\beta}_0$  et  $\hat{\beta}_1$  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  $\varepsilon_t$  supposé  $\sim I(0)$ .

Alors  $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$  pourrait être bien défini pour  $\beta_1$  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,  $\hat{\beta}_1$  ne converge pas vers 0, mais converge plutôt vers une variable aléatoire.

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

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  $= f(T)$ .

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

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

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

## 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
$DF_\mu$	-3.12	-2.86	-2.57		-0.44	-0.07	0.23
$DF_\beta$	-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  $\Delta y_{t-1}$ .

Tester  $H_0 : \gamma = 0$  pour  $\hat{\tau}_\mu$  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

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

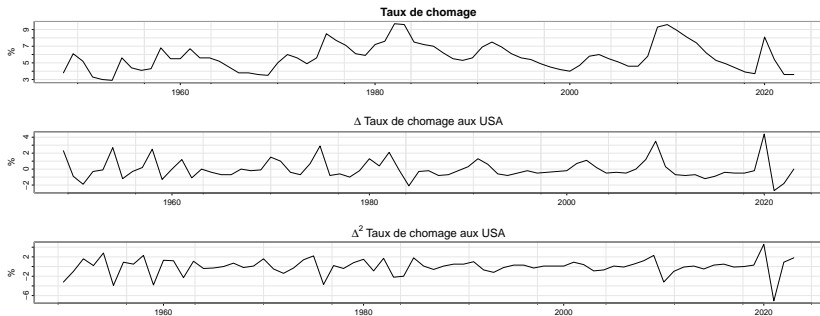
# Illustration empirique

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



# Tester la stationarité

## ADF {tseries}

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

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

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

```
##  
## 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 = 0.01  
## alternative hypothesis: stationary
```

## Tester la stationarité

### ADF {urca}

```
w0=ur.df(x, type = "none", selectlags = "AIC")
summary(w0)@teststat ; summary(w0)@cval
```

```
##                tau1
## statistic -1.078626
##          1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

```
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
## tau1 -2.6 -1.95 -1.61
```

## 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  $\varepsilon_t \sim \text{NID}(0, 1)$

Ajuster un modèle  $\text{AR}(1)$  aux données :

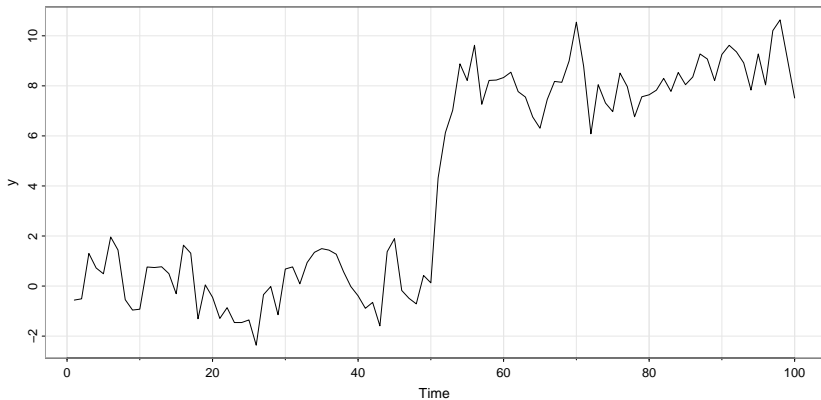
$$y_t = \underset{(0.0000)}{0.9606}y_{t-1} + \underset{(0.1060)}{3.8998}$$

(pvalues)

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



## Processus AR(1) avec changement structurel



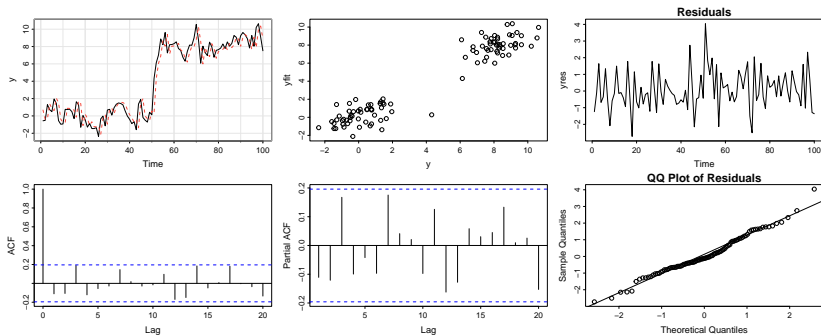
## Ajustment AR(1)

```
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 $\varepsilon_t$

```
par(mfrow = c(2, 3))
tsplot(y)
points(yfit, type = "l", col = "red", 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)
```



# 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  $\chi^2$ ,  $N(0, 1)$

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