

# Séries temporelles univariées

M1 ECAP – TD4 – Année 2024/2025

## TD4 : Analyse de la non-stationarité et racines unitaires

*Responsable d'enseignement : Benoît SÉVI*  
[benoit.sevi@univ-nantes.fr](mailto:benoit.sevi@univ-nantes.fr)

**HOUSSAIS Rémi, QUINTIN DE KERCADIO Pierre**

March 27, 2025

# Table of contents

<b>1</b>	<b>Chocs permanents vs. choc transitoires: une analyse exploratoire</b>	<b>2</b>
<b>2</b>	<b>TS vs DS : simulation de processus</b>	<b>4</b>
2.1	Loi $N(0,1/4)$ . . . . .	4
2.1.1	Seed(123) . . . . .	4
2.1.2	Seed(287) . . . . .	5
2.1.3	Seed(986) . . . . .	7
2.1.4	Conclusion . . . . .	8
2.2	Loi $N(0,1/2)$ . . . . .	8
2.2.1	Seed(123) . . . . .	8
2.2.2	Seed(287) . . . . .	10
2.2.3	Seed(986) . . . . .	12
2.2.4	Conclusion . . . . .	13
2.3	Loi $N(0,1)$ . . . . .	13
2.3.1	Seed(123) . . . . .	13
2.3.2	Seed(287) . . . . .	15
2.3.3	Seed(986) . . . . .	16
2.3.4	Conclusion . . . . .	18
2.4	Conlusion générale . . . . .	18
<b>3</b>	<b>Régressions fallacieuses</b>	<b>19</b>
3.1	$n = 200$ et $N = 5000$ . . . . .	19
3.2	Variations du nombre d'observations, $n$ , et du nombre de séries générée, $N$ .	19
3.2.1	Augmentation et diminution de $n$ . . . . .	19
3.2.2	Augmentation et diminution de $N$ . . . . .	21
3.3	Conclusion générale . . . . .	22

# 1 Chocs permanents vs. choc transitoires: une analyse exploratoire

```
library(stats)
library(ggplot2)

simulate_ar1 <- function(phi, choc_value, choc_time = 100, n = 200) {
  Y <- numeric(n)
  eps <- rnorm(n, mean = 0, sd = 1)

  for (t in 2:n) {
    Y[t] <- phi * Y[t - 1] + eps[t]
    if (t == choc_time) {
      Y[t] <- Y[t] + choc_value
    }
  }
  return(Y)
}

set.seed(123)

phi_vals <- c(0.5, 0.9, 1.0)
chocs <- c(20, 40, -20, -40)

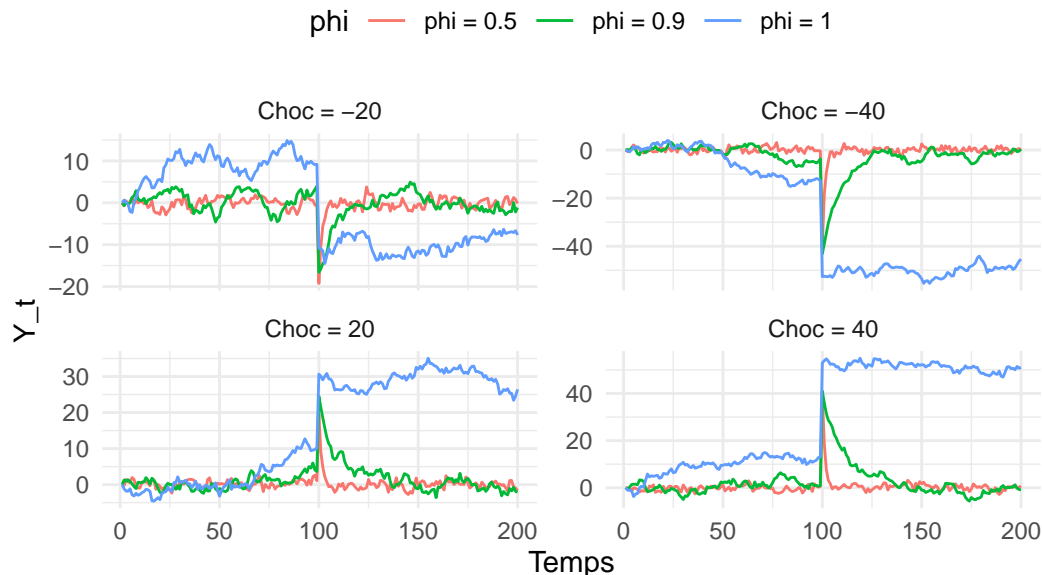
# Stocker les résultats dans une data.frame
results <- data.frame()

for (choc in chocs) {
  for (phi in phi_vals) {
    serie <- simulate_ar1(phi, choc)
    df <- data.frame(
      t = 1:200,
      Y = serie,
      phi = paste0("phi = ", phi),
      choc = paste0("Choc = ", choc)
    )
    results <- rbind(results, df)
  }
}

ggplot(results, aes(x = t, y = Y, color = phi)) +
  geom_line() +
  facet_wrap(~choc, scales = "free_y") +
  labs(title = "Impact de chocs à t=100 pour différentes valeurs de ",
       x = "Temps", y = "Y_t") +
```

```
theme_minimal() +
theme(legend.position = "top")
```

Impact de chocs à  $t=100$  pour différentes valeurs de ..



```
library(ggplot2)
simulate_ar1 <- function(phi, choc_value, choc_time = 100, n = 200) {
  Y <- numeric(n)
  eps <- rnorm(n, mean = 0, sd = 1)

  for (t in 2:n) {
    Y[t] <- phi * Y[t - 1] + eps[t]
    if (t == choc_time) {
      Y[t] <- Y[t] + choc_value
    }
  }
  return(Y)
}
```

## Interprétation

**Lorsque  $\phi=0.5$  et  $\phi=0.5$  :**

Un choc de +20 ou -20 entraîne une perturbation temporaire, mais la série retrouve rapidement son niveau d'avant-choc, en environ 10 unités de temps. Cela s'explique par le fait qu'un faible  $\phi$  signifie une mémoire courte : les valeurs passées influencent peu les valeurs futures, et l'effet du choc s'estompe rapidement.

**Lorsque  $\phi=0.9$  et  $\phi=0.9$  :**

Pour un même choc, le retour à l'état initial est bien plus lent, prenant environ 25 unités de temps. Avec une valeur plus élevée de  $\phi$ , la série possède une mémoire plus longue, ce qui signifie que les effets du choc persistent plus longtemps avant de s'atténuer.

Lorsque  $\rho = 1.0$  (marche aléatoire) :

La série ne revient pas à une moyenne stable après le choc. Une fois l'impact survenu, la trajectoire du processus est modifiée de façon durable. Cela illustre bien le fait qu'un AR(1) avec  $\rho = 1$  est non stationnaire : il ne possède pas de force de rappel vers une moyenne fixe, et l'effet du choc se propage indéfiniment avec des fluctuations autour du nouveau niveau atteint.

Enfin, lorsque l'intensité du choc passe de 20 à 40 (ou de -20 à -40), les écarts par rapport à l'état initial sont plus marqués, et le temps nécessaire pour s'en rapprocher à nouveau augmente. Cela confirme que plus le choc est fort, plus le processus aura du mal à retrouver son équilibre, surtout pour des valeurs élevées de  $\rho$ .

## 2 TS vs DS : simulation de processus

### 2.1 Loi $N(0,1/4)$

#### 2.1.1 Seed(123)

```
library(ggplot2)
set.seed(123)

n <- 200
sigma2 <- 1/4

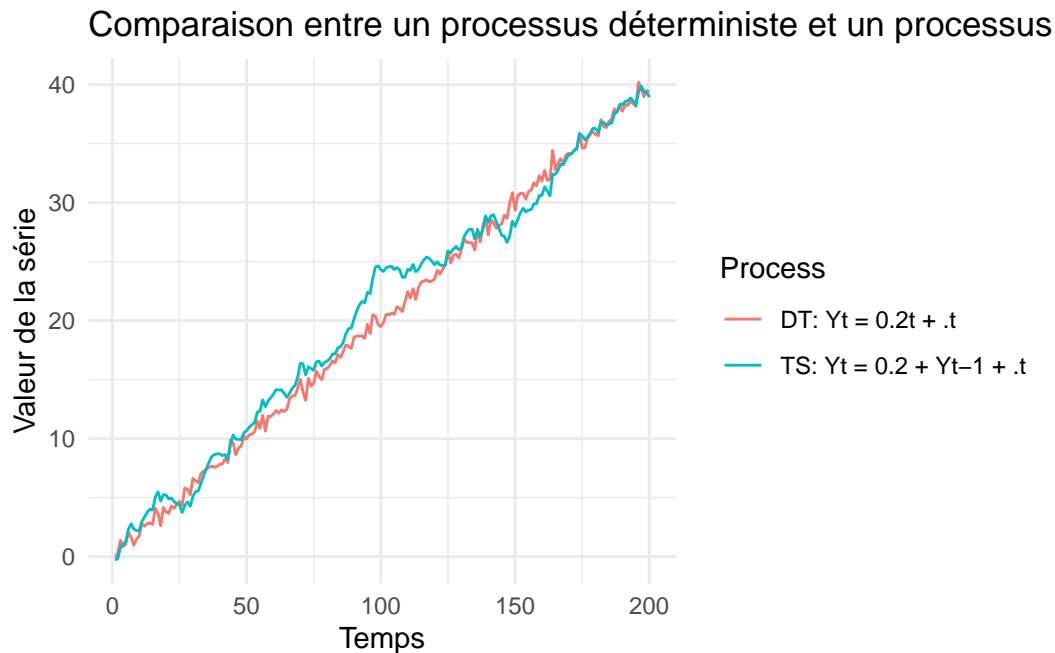
epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

Y_DT <- numeric(n)
Y_TS <- numeric(n)

for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)
```

```
# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
        x = "Temps",
        y = "Valeur de la série") +
  theme_minimal()
```



### Interprétation

Le graphique compare deux séries soumises à un même bruit  $N(0,1/4)$ : un processus déterministe ( $Y_t = 0.2t + \epsilon_t$ ), et un processus stochastique ( $Y_t = 0.2 + Y_{t-1} + \epsilon_t$ ).

La série déterministe suit une tendance linéaire régulière, perturbée temporairement par de faibles chocs : elle est stationnaire autour d'une tendance. En revanche, la série stochastique accumule les chocs dans le temps, ce qui entraîne des écarts de plus en plus marqués : elle est non stationnaire avec des effets de chocs permanents.

#### 2.1.2 Seed(287)

```
set.seed(287)

n <- 200
sigma2 <- 1/4

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc
```

```

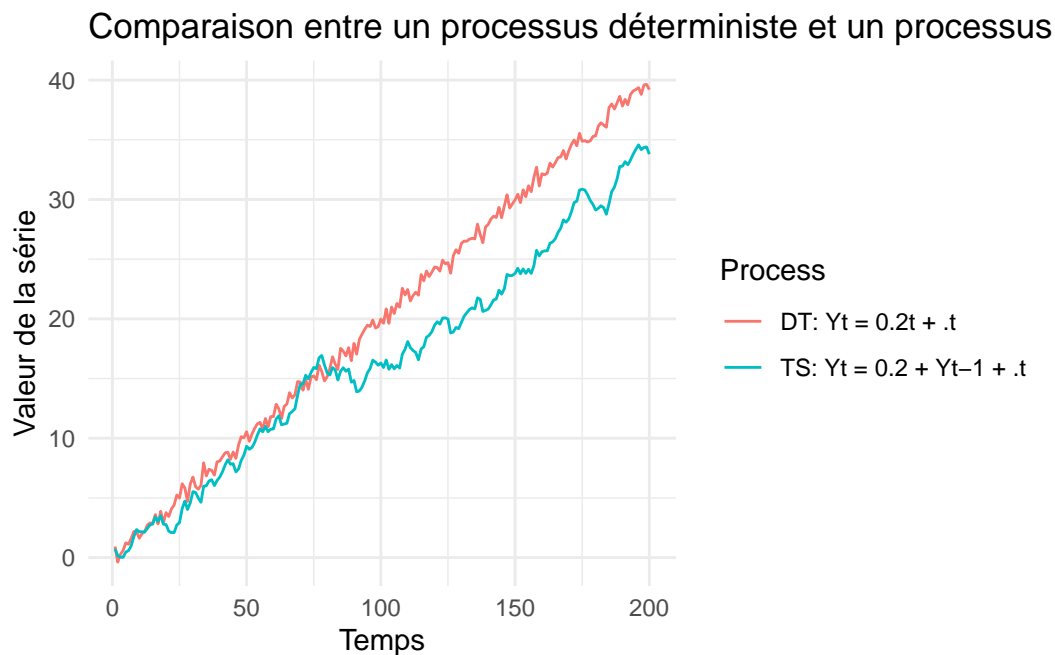
Y_DT <- numeric(n)
Y_TS <- numeric(n)

for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
       x = "Temps",
       y = "Valeur de la série") +
  theme_minimal()

```



### Interprétation

Avec ce nouveau tirage aléatoire, on retrouve la même dynamique générale : le processus déterministe suit une tendance régulière, alors que le processus stochastique évolue de

manière plus irrégulière. Cette fois, le DT reste au-dessus du TS sur presque toute la période, ce qui montre que la trajectoire du processus stochastique dépend fortement des chocs initiaux. Le fond reste inchangé : le DT absorbe les chocs, le TS les cumule dans le temps.

### 2.1.3 Seed(986)

```
set.seed(986)

n <- 200
sigma2 <- 1/4

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

Y_DT <- numeric(n)
Y_TS <- numeric(n)

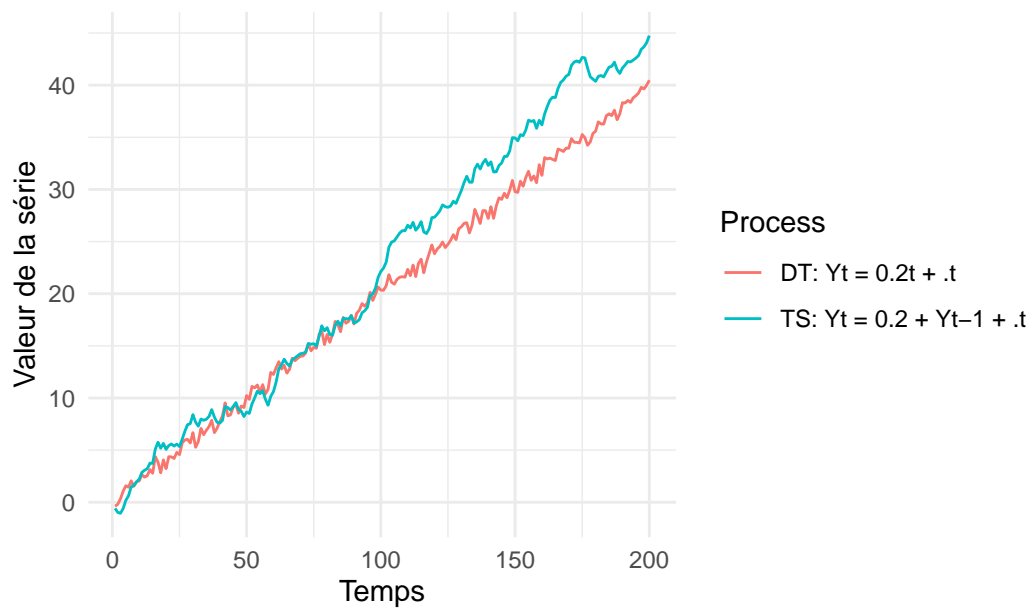
for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
    x = "Temps",
    y = "Valeur de la série") +
  theme_minimal()
```



## Comparaison entre un processus déterministe et un processus



### Interprétation

Avec ce nouveau tirage aléatoire, on retrouve la même dynamique générale: le processus déterministe (rouge) suit une tendance régulière, alors que le processus stochastique (bleu) évolue de manière plus irrégulière. Cette fois, le TS dépasse nettement le DT à partir du milieu de la période, ce qui montre une nouvelle fois que la trajectoire du processus stochastique dépend fortement des chocs accumulés. Le fond reste inchangé : le DT absorbe les chocs, le TS les cumule dans le temps.

### 2.1.4 Conclusion

La dynamique globale reste constante : le processus déterministe suit toujours une tendance régulière perturbée temporairement par le bruit, tandis que le processus stochastique affiche une trajectoire plus irrégulière, qui s'écarte progressivement en fonction des chocs accumulés. Dans certaines simulations, le TS reste en dessous du DT, dans d'autres il le dépasse, ce qui illustre parfaitement la forte sensibilité du processus stochastique à la réalisation des chocs initiaux. Ces trois graphiques confirment que le DT est stable et prévisible, alors que le TS est instable et imprévisible à long terme, malgré des conditions de départ identiques.

## 2.2 Loi $N(0,1/2)$

### 2.2.1 Seed(123)

```
set.seed(123)

# Paramètres
n <- 200 # Nombre d'observations
```

```

sigma2 <- 1/2 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

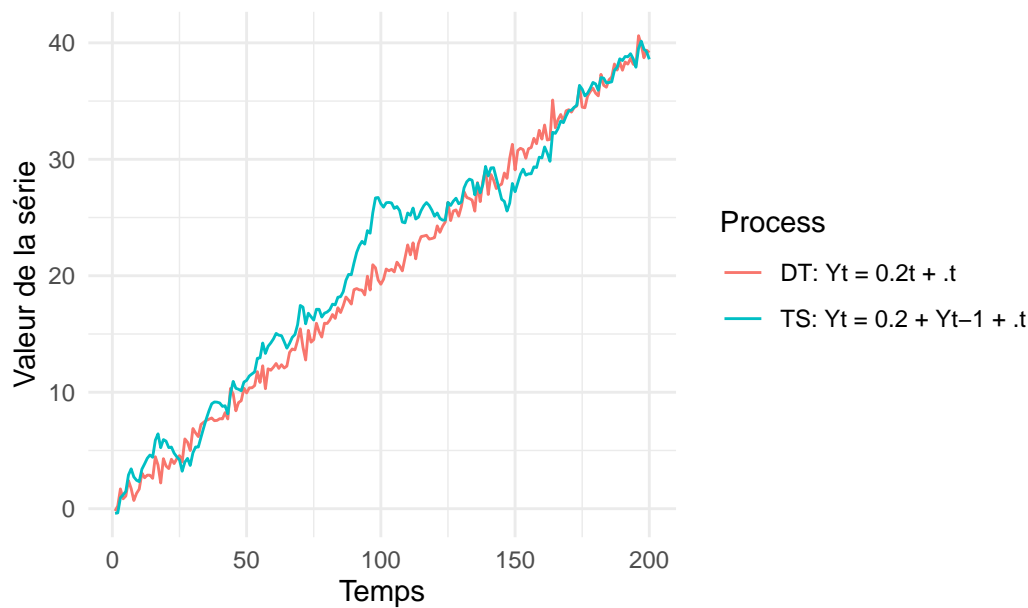
for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
       x = "Temps",
       y = "Valeur de la série") +
  theme_minimal()

```

## Comparaison entre un processus déterministe et un processus



### Interprétation

Avec ce nouveau tirage aléatoire et une variance du bruit plus élevée, on observe une dynamique toujours similaire : le processus déterministe progresse régulièrement selon sa tendance linéaire, tandis que le processus stochastique présente une trajectoire plus irrégulière. Ici, les deux séries restent proches sur l'ensemble de l'échantillon, mais on note que le TS dépasse légèrement le DT en fin de période. Cela montre que l'augmentation de la variance amplifie la variabilité des deux séries, mais affecte davantage le TS qui accumule les chocs. Le DT, lui, reste globalement stable autour de sa pente.

### 2.2.2 Seed(287)

```
set.seed(287)

# Paramètres
n <- 200 # Nombre d'observations
sigma2 <- 1/2 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {

```

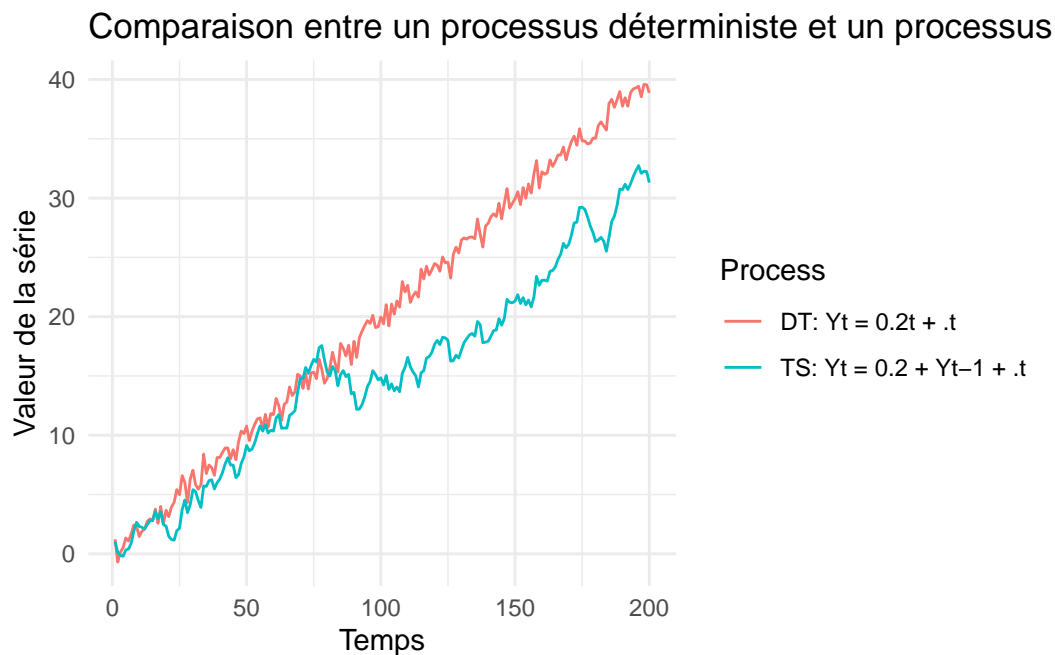
```

    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
    x = "Temps",
    y = "Valeur de la série") +
  theme_minimal()

```



### Interprétation

Avec ce nouveau tirage, on observe une configuration différente : le processus déterministe dépasse clairement le processus stochastique sur presque toute la période. Cela illustre à nouveau que, bien que les deux séries partagent la même pente moyenne (0.2), le comportement du TS est fortement influencé par les chocs initiaux. Ici, ces chocs ont freiné la progression de la série stochastique, tandis que la série déterministe, insensible à la mémoire, poursuit une trajectoire plus régulière. Le contraste est net en fin de période, où le DT atteint des niveaux beaucoup plus élevés que le TS.

### 2.2.3 Seed(986)

```
set.seed(986)

# Paramètres
n <- 200 # Nombre d'observations
sigma2 <- 1/2 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

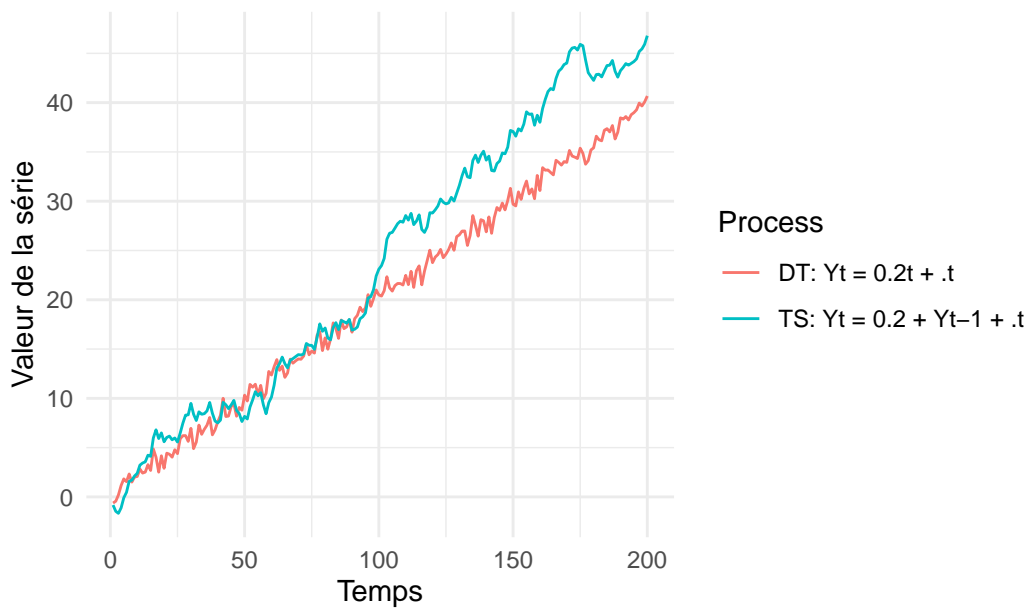
# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT:  $Y_t = 0.2t + \epsilon_t$ ", "TS:  $Y_t = 0.2 + Y_{t-1} + \epsilon_t$ "), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
       x = "Temps",
       y = "Valeur de la série") +
  theme_minimal()
```

## Comparaison entre un processus déterministe et un processus



### Interprétation

Avec ce troisième tirage, le processus stochastique est nettement au dessus sur le déterministe dès la moitié de la période. On observe ici une forte montée du TS, liée à une succession de chocs positifs qui, une fois cumulés, entraînent un écart important par rapport au DT. À l'inverse, le processus déterministe reste fidèle à sa tendance linéaire, avec des fluctuations modérées. Cette configuration montre à quel point le processus stochastique peut diverger rapidement selon les chocs, même si les deux séries partagent la même pente moyenne.

### 2.2.4 Conclusion

Pour une loi  $N(0,1/2)$  nous pouvons remarquer que la dynamique reste constante comme sur la loi  $N(0,1/4)$ . En effet, le processus déterministe suit toujours une tendance régulière, seulement perturbée de manière temporaire par un bruit modéré. En revanche, le processus stochastique présente une trajectoire plus irrégulière, qui s'écarte progressivement à mesure que les chocs s'accumulent dans le temps. Dans certaines simulations, le TS reste en dessous du DT, dans d'autres il le dépasse nettement, ce qui illustre parfaitement la forte sensibilité du processus stochastique à la réalisation des chocs aléatoires. Ces trois graphiques confirment que le DT est stable, tandis que le TS reste instable.

## 2.3 Loi $N(0,1)$

### 2.3.1 Seed(123)

```
set.seed(123)
```

```
# Paramètres
```

```

n <- 200 # Nombre d'observations
sigma2 <- 1 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

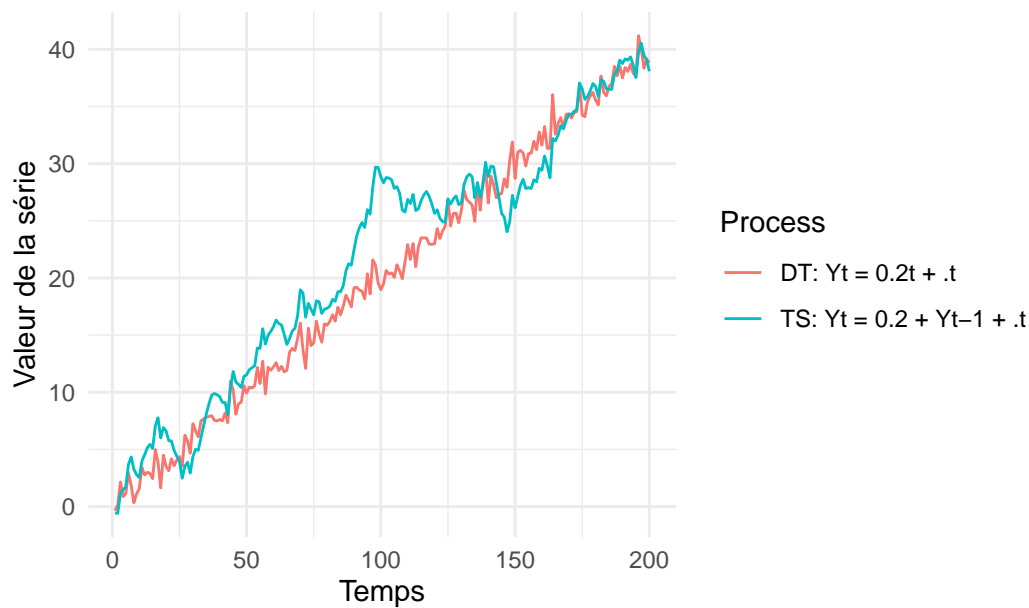
for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT:  $Y_t = 0.2t + \epsilon_t$ ", "TS:  $Y_t = 0.2 + Y_{t-1} + \epsilon_t$ "), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
    x = "Temps",
    y = "Valeur de la série") +
  theme_minimal()

```

## Comparaison entre un processus déterministe et un processus



### Interprétation

Pour cette loi  $N(0,1)$ , la dynamique reste la même observée que les deux autres lois. En effet, le processus déterministe suit une trajectoire croissante régulière, alors que le processus stochastique fluctue davantage et s'en écarte par moments. Ici, le TS dépasse nettement le DT autour de  $t=100$ . avant de redescendre et de se rapprocher en fin de période.

### 2.3.2 Seed(287)

```
set.seed(287)

# Paramètres
n <- 200 # Nombre d'observations
sigma2 <- 1 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}
```



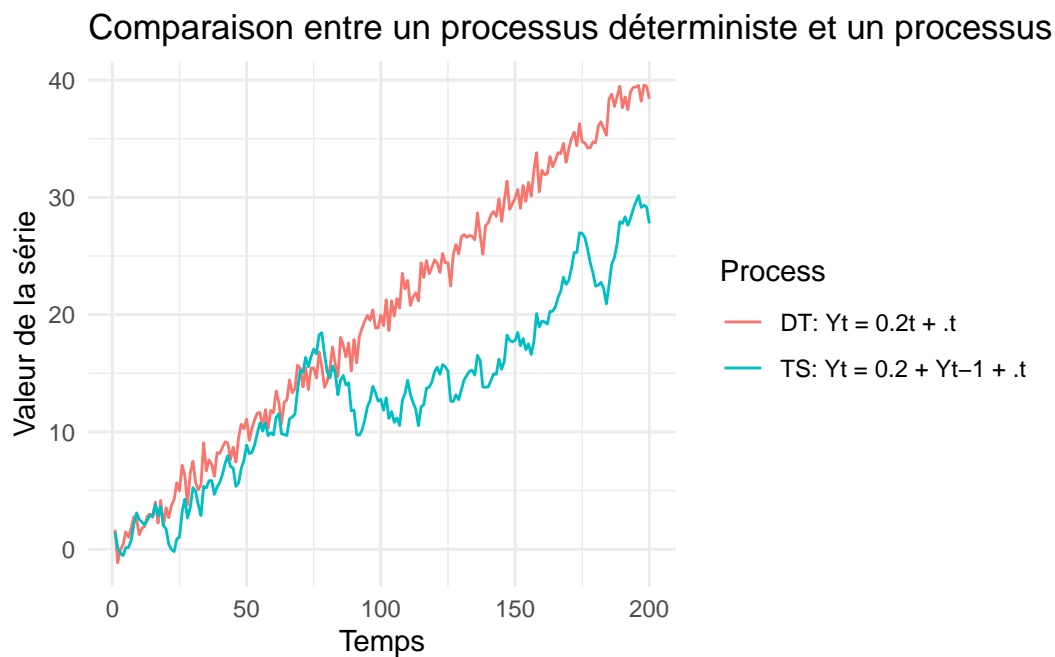
```

}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT: Yt = 0.2t + t", "TS: Yt = 0.2 + Yt-1 + t"), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
       x = "Temps",
       y = "Valeur de la série") +
  theme_minimal()

```



### Interprétation

Avec ce nouveau tirage aléatoire, nous pouvons remarquer le processus déterministe est toujours stable dans le temps et est au dessus du stochastique sur l'ensemble de la période. Nous pouvons remarquer, que à partir de  $T=90$ , TS chute brutalement et reste en dessous du processus déterministe.

### 2.3.3 Seed(986)

```

set.seed(986)

# Paramètres

```

```

n <- 200 # Nombre d'observations
sigma2 <- 1 # Variance du bruit

epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma2)) # Génération du bruit blanc

# Simulation des processus
Y_DT <- numeric(n)
Y_TS <- numeric(n)

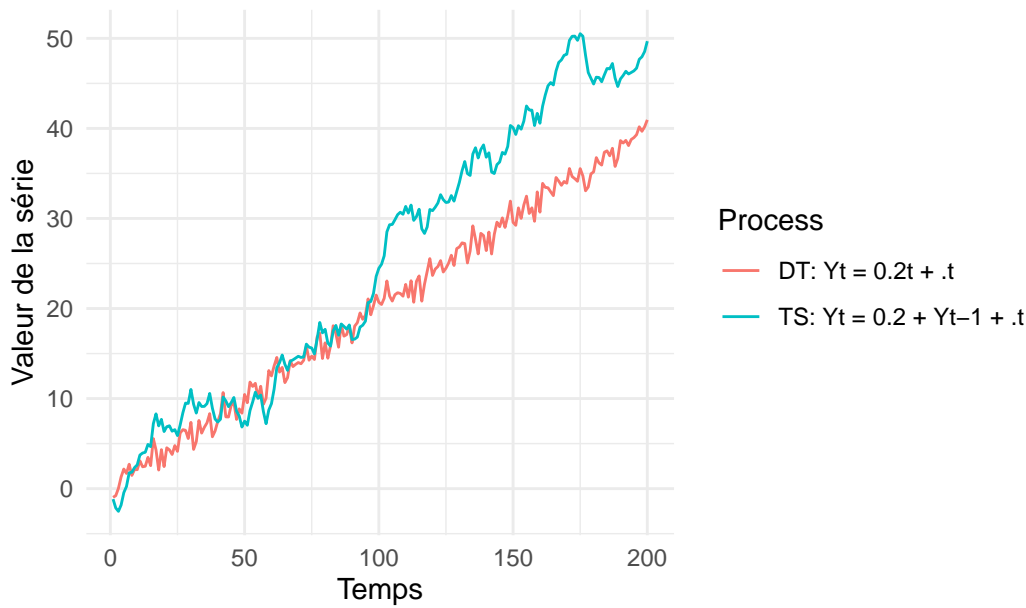
for (t in 1:n) {
  Y_DT[t] <- 0.2 * t + epsilon[t]
  if (t == 1) {
    Y_TS[t] <- epsilon[t]
  } else {
    Y_TS[t] <- 0.2 + Y_TS[t - 1] + epsilon[t]
  }
}

# Création du DataFrame
data <- data.frame(
  Time = rep(1:n, 2),
  Value = c(Y_DT, Y_TS),
  Process = rep(c("DT:  $Y_t = 0.2t + \epsilon_t$ ", "TS:  $Y_t = 0.2 + Y_{t-1} + \epsilon_t$ "), each = n)
)

# Graphique
ggplot(data, aes(x = Time, y = Value, color = Process)) +
  geom_line() +
  labs(title = "Comparaison entre un processus déterministe et un processus stochastique",
    x = "Temps",
    y = "Valeur de la série") +
  theme_minimal()

```

## Comparaison entre un processus déterministe et un processus



### Interprétation

Avec ce nouveau tirage aléatoire, nous pouvons remarquer le processus déterministe est toujours stable dans le temps et est en dessous du stochastique sur l'ensemble de la période. Nous pouvons remarquer, que à partir de  $T=90$ , TS monte brutalement et reste au dessus du processus déterministe.

### 2.3.4 Conclusion

Pour une loi  $N(0,1)$ , la dynamique globale reste constante : le processus déterministe suit toujours une tendance régulière perturbée temporairement par le bruit, tandis que le processus stochastique affiche une trajectoire plus irrégulière, qui s'écarte progressivement en fonction des chocs accumulés. Avec une variance plus élevée, ces écarts deviennent encore plus marqués. Le TS peut rester en retrait comme dans le deuxième graphique, coller au DT comme dans le premier, ou bien le dépasser largement comme dans le troisième. Cette variabilité illustre parfaitement la forte sensibilité du processus stochastique à la réalisation des chocs, qui s'amplifie avec la taille des perturbations.

## 2.4 Conclusion générale

Pour conclure sur cette partie, nous pouvons remarquer que pour chaque série c'était plus ou moins la même tendance. En effet, le processus déterministe a toujours suivi la tendance quelque soit la loi  $N$  ou le seed, et le processus stochastique a toujours été plus sensible aux fluctuations.

## 3 Régressions fallacieuses

### 3.1 $n = 200$ et $N = 5000$

```
set.seed(123)

n <- 200
N <- 5000
rejets <- numeric(N)

for (i in 1:N) {

  X <- cumsum(rnorm(n))
  Y <- cumsum(rnorm(n))

  model <- lm(Y ~ X)
  p_value <- summary(model)$coefficients[2, 4]

  rejets[i] <- ifelse(p_value < 0.05, 1, 0)
}

pourcentage_rejets <- mean(rejets) * 100
cat("Pourcentage de rejets de H0 (1 = 0) au seuil de 5% :", round(pourcentage_rejets,
```

Pourcentage de rejets de H0 (1 = 0) au seuil de 5% : 83.52 %

**Interprétation** Avec ce premier test avec  $n=200$  et  $N=5000$ , nous obtenons un taux de rejet de H0 de 83.52%, par conséquent même si les séries sont indépendantes, nous rejetons H0 très souvent. Cela s'explique par le fait que les marches aléatoires sont non stationnaires, en raison de la présence d'une racine unitaire dans leur dynamique.

Pour pouvoir observer les différences nous allons varier  $n$  et  $N$

### 3.2 Variations du nombre d'observations, $n$ , et du nombre de séries générée, $N$ .

#### 3.2.1 Augmentation et diminution de $n$

##### 3.2.1.1 $n = 50$ et $N = 5000$

```
set.seed(238)
```

```

n <- 50
N <- 5000
rejets <- numeric(N)

for (i in 1:N) {

  X <- cumsum(rnorm(n))
  Y <- cumsum(rnorm(n))

  model <- lm(Y ~ X)
  p_value <- summary(model)$coefficients[2, 4]

  rejets[i] <- ifelse(p_value < 0.05, 1, 0)
}

pourcentage_rejets <- mean(rejets) * 100
cat("Pourcentage de rejets de H0 (1 = 0) au seuil de 5% :", round(pourcentage_rejets,

```

Pourcentage de rejets de H0 (1 = 0) au seuil de 5% : 66.14 %

## Interprétation

Lorsqu'on diminue le nombre d'observation  $n$  à 50, nous pouvons remarquer que nous obtenons un taux de rejet de H0 de 66.14%. Ce qui est inférieur à celui précédent. Par conséquent, quand le nombre d'observation diminue alors on rejete moins à tort.

### 3.2.1.2 $n = 500$ et $N = 5000$

```

set.seed(238)

n <- 500
N <- 5000
rejets <- numeric(N)

for (i in 1:N) {

  X <- cumsum(rnorm(n))
  Y <- cumsum(rnorm(n))

  model <- lm(Y ~ X)
  p_value <- summary(model)$coefficients[2, 4]

```

```

    rejets[i] <- ifelse(p_value < 0.05, 1, 0)
  }

pourcentage_rejets <- mean(rejets) * 100
cat("Pourcentage de rejets de H0 (1 = 0) au seuil de 5% :", round(pourcentage_rejets,

```

Pourcentage de rejets de H0 (1 = 0) au seuil de 5% : 89.2 %

### Interprétation

Lorsqu'on augmente le nombre d'observation  $n$  à 500, nous pouvons remarquer que nous obtenons un taux de rejet de H0 de 89.2%. Ce qui est supérieur aux deux précédents. Par conséquent, quand le nombre d'observation augmente alors on rejete plus à tort.

## 3.2.2 Augmentation et diminution de $N$

### 3.2.2.1 $n = 200$ et $N = 2000$

```

set.seed(238)

n <- 200
N <- 2000
rejets <- numeric(N)

for (i in 1:N) {

  X <- cumsum(rnorm(n))
  Y <- cumsum(rnorm(n))

  model <- lm(Y ~ X)
  p_value <- summary(model)$coefficients[2, 4]

  rejets[i] <- ifelse(p_value < 0.05, 1, 0)
}

pourcentage_rejets <- mean(rejets) * 100
cat("Pourcentage de rejets de H0 (1 = 0) au seuil de 5% :", round(pourcentage_rejets,

```

Pourcentage de rejets de H0 (1 = 0) au seuil de 5% : 83.6 %

### Interprétation

Lorsqu'on diminue le nombre de séries générées  $N$  à 2000, nous pouvons remarquer que nous obtenons un taux de rejet de H0 de 83.6%. Ce qui est légèrement au dessus du modèle initial.

### 3.2.2.2 $n = 200$ et $N = 10000$

```
set.seed(238)

n <- 200
N <- 10000
rejets <- numeric(N)

for (i in 1:N) {

  X <- cumsum(rnorm(n))
  Y <- cumsum(rnorm(n))

  model <- lm(Y ~ X)
  p_value <- summary(model)$coefficients[2, 4]

  rejets[i] <- ifelse(p_value < 0.05, 1, 0)
}

pourcentage_rejets <- mean(rejets) * 100
cat("Pourcentage de rejets de H0 (1 = 0) au seuil de 5% :", round(pourcentage_rejets,
```

Pourcentage de rejets de H0 (1 = 0) au seuil de 5% : 83.49 %

### Interprétation

Lorsqu'on augmente le nombre de séries générées  $N$  à 10000, nous pouvons remarquer que nous obtenons un taux de rejet de  $H_0$  de 83.49%. Ce qui est légèrement en dessous du modèle initial.

## 3.3 Conclusion générale

Par conséquent, le fait de changer le nombre de séries générées ne varie pas énormément le taux de rejet de  $H_0$ . Alors que, le fait de changer à la baisse ou à la hausse le nombre d'observations fait varier énormément le taux de rejet de  $H_0$