

**Mémoire de Master 1 Mathématiques Appliquées
de l'Université de Paris Dauphine**

Titre : Test de linéarité dans un modèle à volatilité stochastique

Par : Ilan Palacci, Yohan Cohen, Avigail Zerrad

Directeur de Mémoire : Marc Hoffmann

Numéro de groupe : 3

Confidentialité : ☒ Non ☐ Oui (Durée : ☒ 1 an ☐ 2 ans)

Remerciement

Nous tenons à remercier notre directeur de mémoire, Marc Hoffmann, qui nous a guidé dans la réalisation de ce travail, et qui s'est montré disponible malgré toutes les situations que nous avons pu traverser cette année.

Table des matières

Introduction	3
1 Approche théorique	6
1.1 Définition du modèle	6
1.2 Cas ω inconnu	9
2 Mise en application numérique	10
2.1 Simulation de données selon le modèle	10
2.2 Tests sur données réelles	13
3 Puissance du test	20
3.1 L'hypothèse alternative	20
3.2 Implémentation numérique	21
4 Résultats théoriques	24
4.1 Niveau du test (contrôle de l'erreur de 1ère espèce)	24
4.2 Puissance du test (contrôle de l'erreur de 2nde espèce)	25
Conclusion	27
Bibliographie	27

Introduction

Cette étude est basée sur l'article de recherche *Testing Linearity in an AR Errors-in-variables Model with Application to Stochastic Volatility* (2003), écrit par D. Feldmann, W. Härdle, C. Hafner, M. Hoffmann, O. Lepski et A. Tsybakov.

La volatilité stochastique est utilisée dans le cadre de la finance quantitative, pour évaluer des produits dérivés, tels que des options. Le nom provient du fait que le modèle traite la volatilité du sous-jacent comme un processus aléatoire. Une bonne connaissance de ces modèles est particulièrement importante pour l'analyse des données haute fréquence en finance. Cela nous permet en effet de faire des prévisions sur la volatilité, d'évaluer le risque...

Parmi les modèles financiers à volatilité constante, non-stochastique, on peut citer le modèle de Cox-Ross-Rubinstein (modèle en temps discret) ou le modèle de Black-Scholes (qui correspond à la généralisation du modèle CRR en temps continu), sur lesquels nous avons eu l'occasion de travailler lors du cours de "Mouvement brownien & évaluation des actifs contingents" que nous avons suivi cette année.

Les modèles de volatilité stochastique présentent l'une des approches pour résoudre le problème posé par le modèle de Black-Scholes, qui ne prend pas en compte le fait que la volatilité sous-jacente peut ne pas être constante pendant le temps de vie du produit dérivé, et que celui-ci est affecté par le changement de valeur du sous-jacent. En supposant que la volatilité du prix du sous-jacent est un processus stochastique, plutôt qu'une constante, il devient possible de modéliser les produits dérivés avec plus de précision.

Il existe plusieurs modèles qui ont été établis au cours du temps comme les modèles auto-régressifs d'hétéroscédasticité conditionnelle (ARCH) (Engle, 1982; Gouriéroux, 1997) ou encore les modèles à volatilité stochastique (SV) (Taylor, 1986) utilisés pour la tarification des options sous une volatilité variable dans le temps.

Le modèle SV à temps discret peut être représenté comme un modèle auto-régressif (AR) à erreurs dans les variables. Il est alors intéressant de proposer un test statistique qui permet de distinguer l'hypothèse AR paramétrique linéaire, de l'ensemble des alternatives AR non paramétriques.

Dans un premier temps, nous allons définir le modèle au sens mathématique, puis présenter les tests statistiques construits. L'enjeu sera alors de tester la cohérence des estimateurs et tests qui nous sont proposés en les appliquant sur des simulations basées sur le modèle théorique. Par la suite, s'il s'avère que le modèle est cohérent, nous pourrons l'appliquer sur des jeux de données réels afin de tester l'hypothèse nulle de linéarité, pour nous intéresser ensuite à la puissance du test.

Chapitre 1

Approche théorique

1.1 Définition du modèle

Nous commençons par poser les différentes variables du modèle.

t désigne l'indice de temps, $t = 1, \dots, n$. Le prix au temps t de l'actif sous-jacent est donné par S_t . Les rendements sont notés h_t , et la log volatilité Y_t .

La modélisation mathématique habituelle des rendements est donnée par :

$$h_t = \log \left(\frac{s_t}{s_{t-1}} \right), t = 1, \dots, n$$

Par ailleurs, le modèle standard de volatilité stochastique de Taylor (1986) permet d'écrire les rendements comme :

$$h_t = \exp(Y_t/2) \xi_t^* \Rightarrow h_t^2 = \exp(Y_t) \xi_t^{*2}$$

où les ξ_t^* sont des variables centrées, indépendantes et identiquement distribuées.

Ce qui donne :

$$\log h_t^2 = Y_t + 2 \log(|\xi_t^*|)$$

On note $Z_t := \log h_t^2$. On peut donc écrire :

$$Z_t = Y_t + 2 \log(|\xi_t^*|) - 2\mathbf{E}[\log(|\xi_t^*|)] + 2\mathbf{E}[\log(|\xi_t^*|)].$$

En notant $\omega := 2\mathbf{E}[\log(|\xi_t^*|)]$ et $\xi_t := 2 \log(|\xi_t^*|) - \omega$, on obtient :

$$Z_t = \omega + Y_t + \xi_t \tag{1.1}$$

Par ailleurs, le modèle standard de Taylor propose la relation de récurrence suivante :

$$Y_t = \vartheta Y_{t-1} + \varepsilon_t \tag{1.2}$$

où les ε_t sont des variables centrées, indépendantes et identiquement distribuées.

On suppose également que ε_t et ξ_t sont mutuellement indépendantes.

On peut écrire (1.2) comme : $Y_t = m(Y_{t-1}) + \varepsilon_t$, où $m()$ est une fonction.

La question suivante se pose alors : est-il légitime de supposer que $m()$ est linéaire ?

Notre objectif sera de construire un test statistique afin de valider ou non la linéarité de $m()$ dans le cadre du modèle de volatilité stochastique.

Cela revient à tester : $\mathbf{H}_0 : m(x) = \vartheta x$, $\vartheta \in [a, b]$, avec $0 \leq a < b < 1$.

L'alternative H_1 est l'ensemble très général des fonctions non linéaires - que l'on affinera par la suite.

Construction de la statistique de test

On cherche à construire une statistique de test pivotale, i.e dont on connaît la loi sous H_0 .

Par définition de l'hypothèse nulle, on a : $m(x) = \vartheta x$.

On commence par supposer que ω est connu, donc on pose $\omega = 0$.

Dans ce cas, on a :

$$\begin{aligned} Z_t &= Y_t + \xi_t \\ \Rightarrow Z_{t-1} &= Y_{t-1} + \xi_{t-1} \\ \Rightarrow \vartheta Z_{t-1} &= \vartheta Y_{t-1} + \vartheta \xi_{t-1} \end{aligned}$$

On obtient :

$$Z_t - \vartheta Z_{t-1} = Y_t + \xi_t - \vartheta Y_{t-1} - \vartheta \xi_{t-1}$$

Par ailleurs, sous H_0 , on a $Y_t = m(Y_{t-1}) + \varepsilon_t = \vartheta Y_{t-1} + \varepsilon_t$.

Donc

$$Z_t - \vartheta Z_{t-1} = \vartheta Y_{t-1} + \varepsilon_t + \xi_t - \vartheta Y_{t-1} - \vartheta \xi_{t-1}$$

$$Z_t - \vartheta Z_{t-1} = \varepsilon_t + \xi_t - \vartheta \xi_{t-1} \tag{1.3}$$

On a alors :

$$\mathbf{E} \left[(Z_t - \vartheta Z_{t-1})^2 \right] = \mathbf{E} \left[(\varepsilon_t + \xi_t - \vartheta \xi_{t-1})^2 \right]$$

Comme les ξ_t et les ε_t sont centrées et indépendantes,

$$\mathbf{E} \left[(\varepsilon_t + \xi_t - \vartheta \xi_{t-1})^2 \right] = \text{var}(\varepsilon_t + \xi_t - \vartheta \xi_{t-1}) = \text{Var}(\varepsilon_t) + \text{var}(\xi_t) + \vartheta^2 \text{var}(\xi_{t-1})$$

On pose pour alléger les notations : $\mathbf{E}[\xi_1^2] := \eta^2$ et $\mathbf{E}[\varepsilon_1^2] := \sigma^2$.

Donc :

$$\mathbf{E} \left[(\varepsilon_t + \xi_t - \vartheta \xi_{t-1})^2 \right] = \sigma^2 + \eta^2 (1 + \vartheta^2) := M(\vartheta)$$

On pose de même : $\mathbf{E} [\xi_1^4] := \mu$ et $\mathbf{E} [\epsilon_1^4] := \nu$.

Pour la variance, on obtient :

$$\text{var} \left[(\varepsilon_t + \xi_t - \vartheta \xi_{t-1})^2 \right] = (\nu - \sigma^4) + \mu (1 + \vartheta^2)^2 - \eta^4 (1 - \vartheta^2)^2 + 4\sigma^2 \eta^2 (1 + \vartheta^2) := B(\vartheta)$$

La valeur de ϑ étant inconnue, on cherche à l'estimer. On pose $N = [n/2]$, avec $N \geq 4$. Un estimateur de ϑ sous H_0 est :

$$\bar{\vartheta}_n = \frac{\sum_{t=3}^{N-1} Z_t Z_{t-2}}{\sum_{t=3}^{N-1} Z_{t-1} Z_{t-2}} \quad (1.4)$$

En le projetant sur $[a, b]$, on obtient :

$$\hat{\vartheta}_n = \begin{cases} \bar{\vartheta}_n & \text{if } a \leq \bar{\vartheta}_n \leq b \\ a & \text{if } \bar{\vartheta}_n < a \\ b & \text{if } \bar{\vartheta}_n > b \end{cases}$$

Les variables $((Z_t - \vartheta Z_{t-1})^2)_{t=1, \dots, n}$ sont supposées intégrables, indépendantes et identiquement distribuées ; nous connaissons leur espérance et leur variance. De plus, nous avons un estimateur de ϑ . En appliquant une variante du TCL et en remplaçant ϑ par son estimateur, nous obtenons alors que T_n converge en loi vers une Normale $N(0, 1)$, où :

$$T_n := \frac{1}{\sqrt{NB_n}} \sum_{t=N+1}^n \left\{ (Z_t - \hat{\vartheta}_n Z_{t-1})^2 - M_n \right\} \quad (1.5)$$

avec $M_n = M(\hat{\vartheta}_n)$ et $B_n = B(\hat{\vartheta}_n)$.

On obtient un test asymptotique de niveau α pour tester la linéarité de $m(\cdot)$. On accepte H_0 si $T_n \leq q_{1-\alpha}$, où $q_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi $N(0, 1)$.

1.2 Cas ω inconnu

En pratique, on ne peut pas supposer que ω est connu. On peut réécrire (1.1) et (1.2) comme :

$$\begin{aligned} Z_t &= \omega(1 - \vartheta) + \vartheta Z_{t-1} + \varepsilon_t + \xi_t - \vartheta \xi_{t-1} \\ &= \gamma + \vartheta Z_{t-1} + \nu_t(\vartheta) \end{aligned}$$

où $\gamma := \omega(1 - \vartheta)$ et $\nu_t(\vartheta) := \varepsilon_t + \xi_t - \vartheta \xi_{t-1}$ avec $\mathbf{E}[\nu_t(\vartheta)] = 0$.

On remarque que $\hat{\omega} = \frac{1}{n} \sum_{t=1}^n Z_t$ est un estimateur \sqrt{n} -consistant de ω .

L'idée est d'obtenir un estimateur de ϑ dans le cas ω inconnu, et il est plus pratique de travailler avec l'expression $Z_t = \gamma + \vartheta Z_{t-1} + \nu_t(\vartheta)$. En effet, la réécriture de Z_t en fonction de γ permet une sorte de linéarisation. Si on omet le bruit $\nu_t(\vartheta)$, on peut facilement estimer ϑ à l'aide d'une procédure itérative.

Voici la procédure itérative proposée :

Initialisation

On considère les observations centrées $Z_t^* = Z_t - \hat{\omega}$. On pose :

$$\vartheta_n^{(1)} = \frac{\sum_{t=3}^n Z_t^* Z_{t-2}^*}{\sum_{t=3}^n Z_{t-1}^* Z_{t-2}^*}, \quad \gamma^{(1)} = \frac{1}{n} \sum_{t=2}^n \left(Z_t - \vartheta_n^{(1)} Z_{t-1} \right)$$

Itérations

$$\tilde{z}_t^{(i)} = Z_t - \gamma^{(i-1)}, \quad \vartheta_n^{(i)} = \frac{\sum_{t=3}^n \tilde{z}_t^{(i)} Z_{t-2}}{\sum_{t=3}^n \tilde{z}_t^{(i)} Z_{t-1}}, \quad \gamma^{(i)} = \hat{\omega}(1 - \vartheta_n^{(i)}).$$

Lorsque i tend vers l'infini, et à n fixé, $\vartheta_n^{(i)}$ converge vers une certaine limite $\bar{\vartheta}_{0n}$, que l'on projette sur $[a, b]$. On obtient l'estimateur $\hat{\vartheta}_{0n}$.

On obtient alors la statistique de test :

$$\tilde{T}_n = \frac{1}{\sqrt{nB_n}} \sum_{t=2}^n \left\{ \left(\tilde{z}_t - \hat{\vartheta}_{0n} Z_{t-1} \right)^2 - M_n \right\} \quad (1.6)$$

Chapitre 2

Mise en application numérique

2.1 Simulation de données selon le modèle

Dans un premier temps, nous utilisons le modèle décrit dans la première partie afin de simuler des données puis de vérifier la cohérence des estimateurs et statistiques de tests choisis, dans le cas ω connu.

On choisit les paramètres suivants :

```
a = 0.4
b = 0.6
n = 5000
p = 10000
mu1 = 0
nu1 = 1
mu2 = 0
nu2 = 0.01
theta = 0.5
```

FIGURE 2.1: Paramètres utilisés pour simuler les données selon le modèle.

'mu1', 'nu1', 'mu2', 'nu2' sont les paramètres des lois normales que suivent ε_t et ξ_t .

Le code ci-dessous permet de simuler $p = 10000$ jeux de $n = 5000$ données.

```
StockSampling <- function(n,p,mu1,nu1,mu2,nu2,theta){
  DATAM = matrix(0, n, p)
  Z = numeric(n)
  Y = numeric(n)
  for (b in 1:p) {
    Y[1] = 0.5
    for ( i in 1:(n-1) ) {
      Y[i+1] = theta*Y[i]+epsilon[i,b]
      Z[i] = Y[i]+xi[i,b] # w supposé nul pour l'instant
    }
    Z[n] = theta*Y[n]+xi[n,b]
    DATAM[,b] = Z
  }
  return(DATAM)
}

DATAM = StockSampling(n,p,mu1,nu1,mu2,nu2,theta)
```

FIGURE 2.2: Algorithme de simulation des jeux de données Z_1, \dots, Z_n

Une exécution donne par exemple :

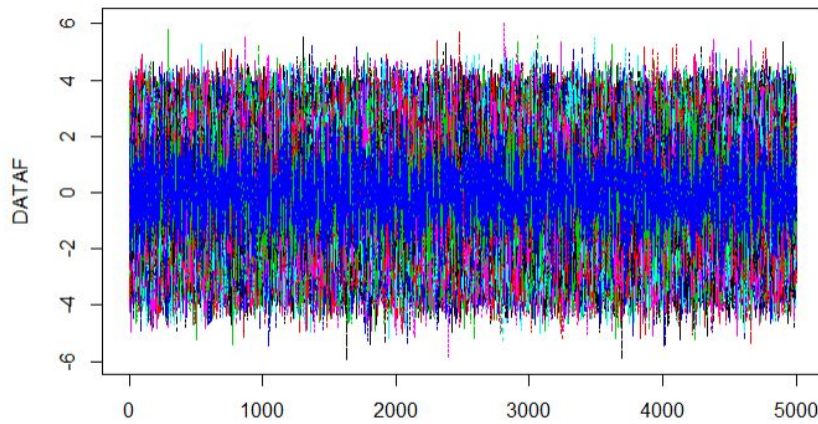


FIGURE 2.3: Graphique des données simulées en partant de $Y_0 = 0.5$

En ayant choisi $\vartheta = 0.5$ pour simuler les données, on obtient la matrice DATAM. Pour savoir si l'estimateur de ϑ proposé dans la première partie est cohérent, nous le calculons pour ces données. Si l'estimateur est bien choisi, nous devrions avoir une majorité de $\hat{\vartheta}_n$ qui sont proches de 0.5.

Voici l'histogramme obtenu :

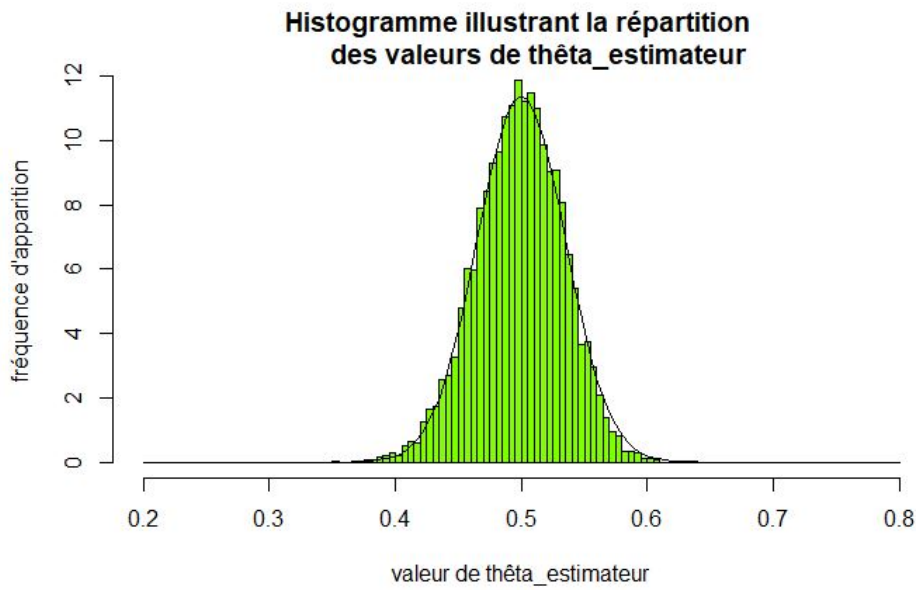


FIGURE 2.4: Histogramme représentant la répartition des valeurs des estimateurs de ϑ , ainsi que la densité d'une $N(0.5, 0.035)$

On remarque que les valeurs sont assez resserrées autour de 0.5, ce qui confirme que l'estimateur de ϑ proposé est cohérent. Par ailleurs, la forme de l'histogramme ci-dessus suggère que l'estimateur proposé est asymptotiquement normal.

Nous allons à présent calculer les statistiques de tests pour plusieurs échantillons. Puisque nous avons généré des données en supposant que $m()$ est linéaire, si la statistique de test est bien choisie, nous devrions avoir que H_0 est acceptée dans la majorité des cas.

En reprenant le test proposé dans la partie théorique, nous avons :

```
#Calcul de la somme (ds la stat)

N=(n/2)
somme<-function(a,b,n,p){
  N=(n/2)
  res = vector('numeric',p)
  for (j in 1:p){
    sum=0
    for (i in (N+1):n){
      sum=sum+(( DATAM[i,j] - THETACHAP[j]*DATAM[i-1,j])^2)-M_n[j]
    }
    res[j]=sum
  }
  return(res)
}

#statistique de test

S_n=somme(a,b,n,p)
STAT0=(1/sqrt(N*B_n))*S_n
STAT1=(1/sqrt(N*B_n))*S_n
STAT0[STAT0<=qnorm(0.95,0,1)]=0

STAT1[STAT1>qnorm(0.95,0,1)]=1
HYP0=sum(STAT0==0)
HYP1=sum(STAT1==1)
cat('On accepte H0 dans', (HYP0/p)*100, '% des cas')
cat('  On rejette H0 dans', (HYP1/p)*100, '% des cas')
```

FIGURE 2.5: Construction de la statistique de test pour chaque jeu de données & test de H_0

On obtient alors (pour $n = 5000$ et $p = 10000$) :

```
On accepte H0 dans 98.58 % des cas  On rejette H0 dans 1.42 % des cas
```

2.2 Tests sur données réelles

En pratique, la statistique de test proposée dans la première partie n'est pas utilisée, puisqu'on ne connaît pas les valeurs de B_n et de M_n . On utilise les estimateurs suivants :

$$\widehat{M}_n = \frac{1}{n-1} \sum_{t=2}^n (\tilde{z}_t - \bar{\vartheta}_{0n} Z_{t-1})^2, \quad \widehat{B}_n = \frac{1}{n-1} \sum_{t=2}^n (\tilde{z}_t - \bar{\vartheta}_{0n} Z_{t-1})^4$$

De plus, on divise les données en k groupes de tailles égales et on obtient ainsi k test statistiques définis comme suit :

$$T_n^{(j)} = \sqrt{\frac{k}{n\widehat{B}_n}} \sum_{t=n_{j-1}+1}^{n_j} \left\{ (\tilde{z}_t - \bar{\vartheta}_{0n} Z_{t-1})^2 - \widehat{M}_n \right\} \quad (2.1)$$

Pour k grand, cela revient en quelque sorte à tester localement la linéarité (i.e sur de plus petites périodes de temps).

Nous nous intéressons à présent aux tests sur données réelles (i.e pour différents indices boursiers). Sur chacune des figures représentées ci-dessous, le premier graphique correspond au prix en fonction du temps, et le deuxième représente le log-log-rendement. Les données sur lesquelles nous avons appliqué les tests sont prises depuis la création de chaque indice.

Dow Jones

Nous observons tout d'abord que le fait de diviser les données est cohérente. En effet, pour $k = 2$ ou $k = 10$, l'hypothèse de linéarité est toujours acceptée. En revanche, pour des valeurs de k plus élevées, on rejette l'hypothèse sur certaines périodes. Ces périodes devraient logiquement correspondre à des moments de crise.

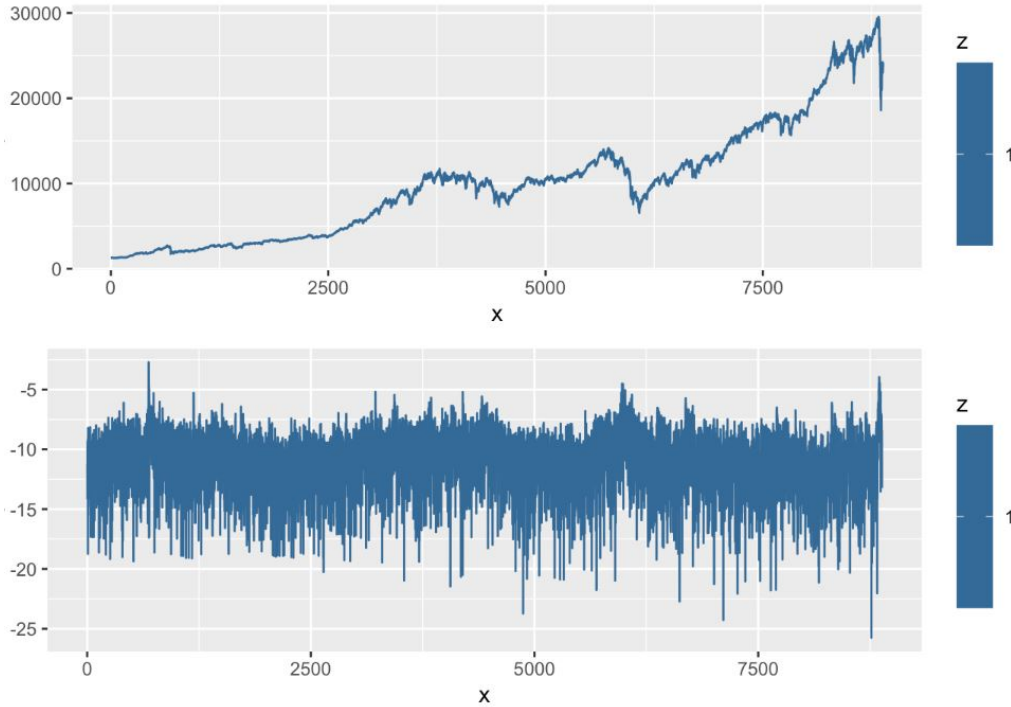


FIGURE 2.6: Dow Jones, $k = 2$

Pour $k = 50$, on accepte H_0 dans 96% des cas. Les périodes de rejet, au nombre de 2, débutent en 2003 et en 2007, et durent n/k jours, où n représente le nombre de jours depuis la création de l'indice.

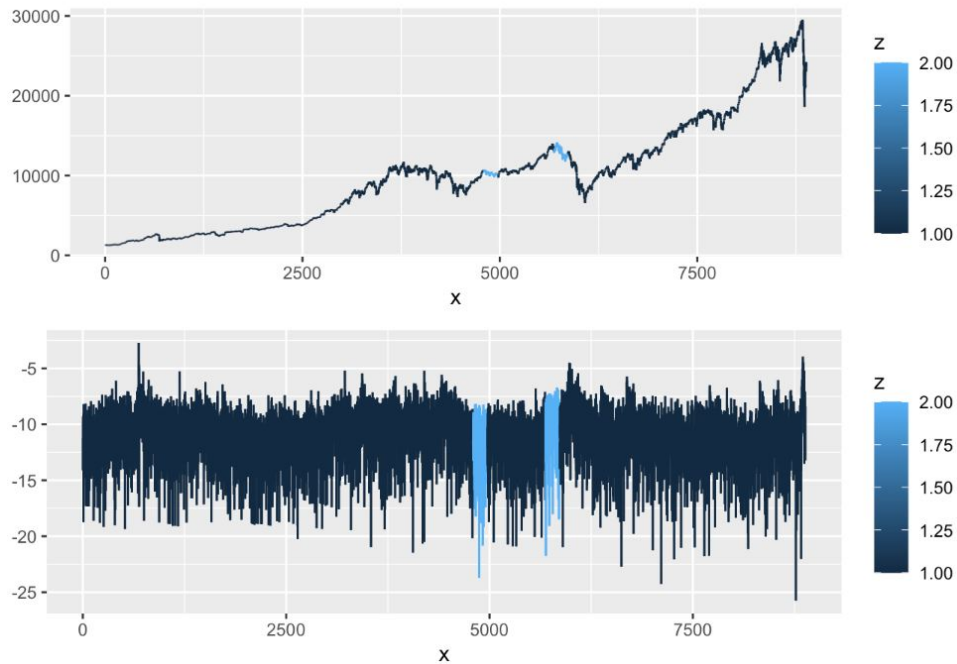


FIGURE 2.7: Dow Jones, $k = 50$

Voici ce qu'on obtient pour $k = 100$ (rejet à 10%).

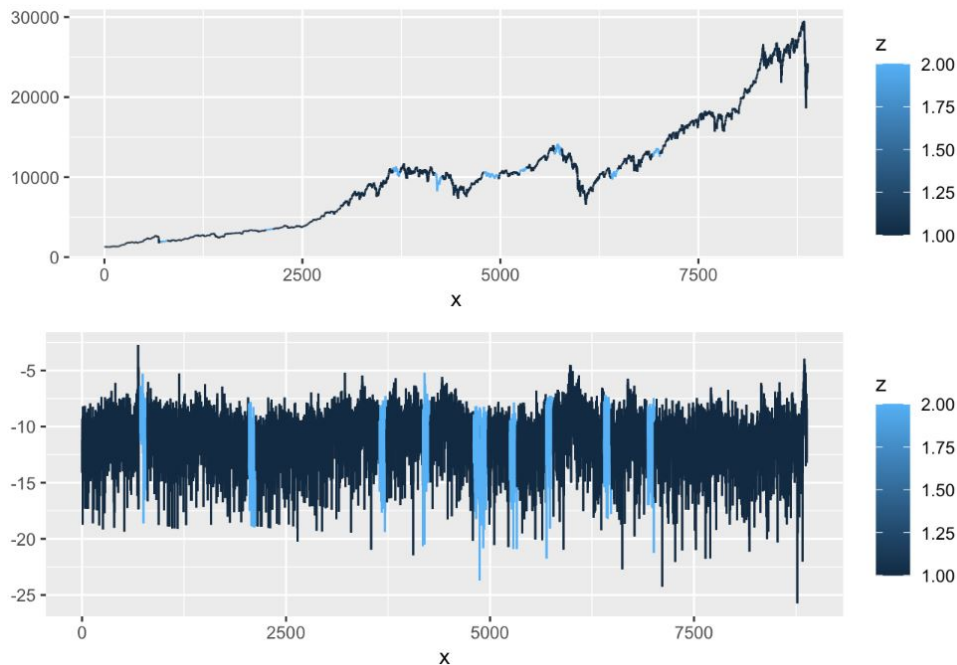


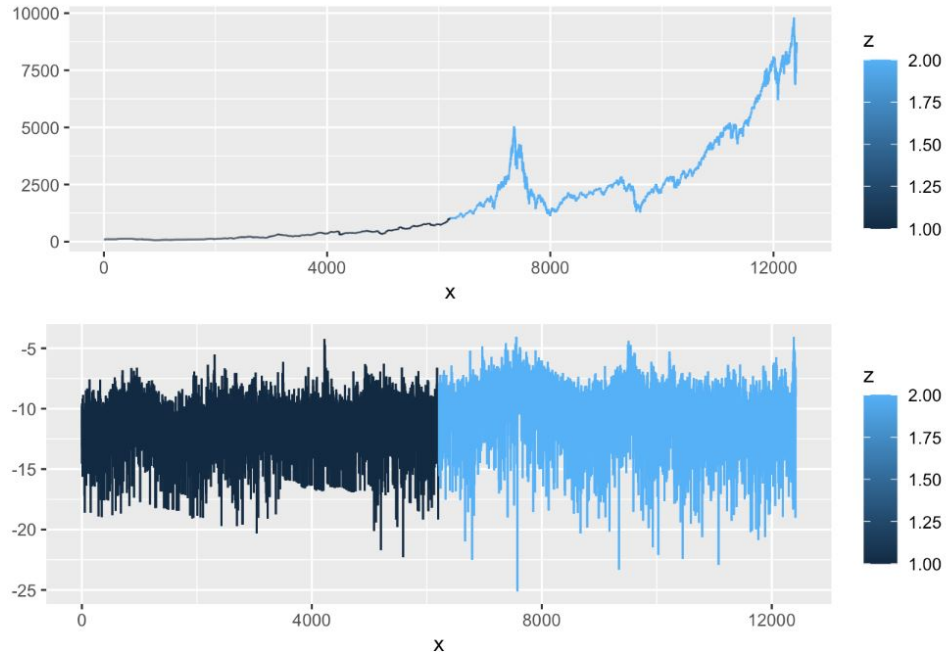
FIGURE 2.8: Dow Jones, $k = 100$

Note : pour les graphiques bicolores, la valeur 1 (bleu marine) correspond à l'acceptation de H_0 et la valeur 2 (bleu ciel) correspond au rejet.

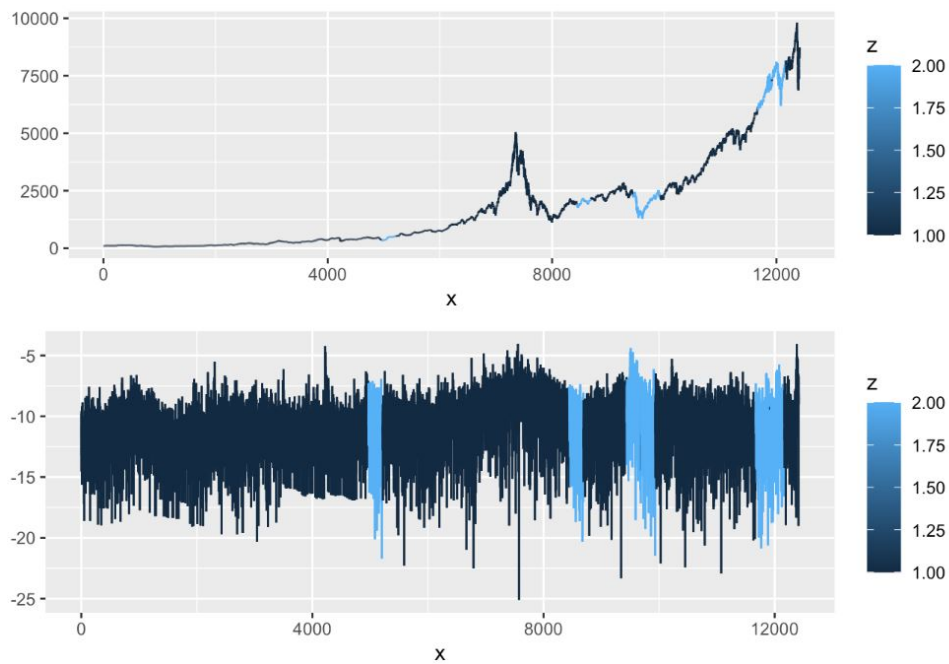
NASDAQ

Si la linéarité semble confirmée pour le Dow Jones, il semblerait que le NASDAQ ne se comporte pas aussi "bien".

Pour $k = 2$, on accepte H_0 seulement dans 50% des cas.

FIGURE 2.9: NASDAQ, $k = 2$

Pour $k = 50$, on accepte H_0 dans 88% des cas. Les périodes de rejet débutent en 1990, 2004, 2008, 2009, 2017 et 2018, et durent à chaque fois n/k jours.

FIGURE 2.10: NASDAQ, $k = 50$

S&P 500 Pour $k = 2$, le S&P 500 se comporte comme le NASDAQ, en terme de pourcentage de rejet.

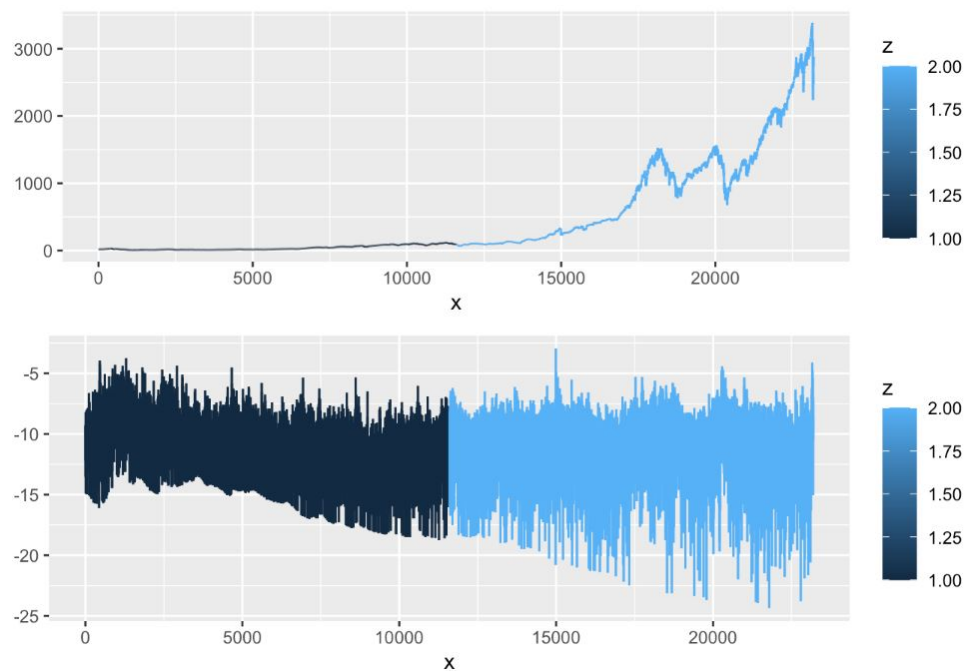


FIGURE 2.11: S&P 500, $k = 2$

Pour $k = 50$, on remarque de nombreuses périodes de rejet, avec un pourcentage d'acceptation de H_0 de 74%.

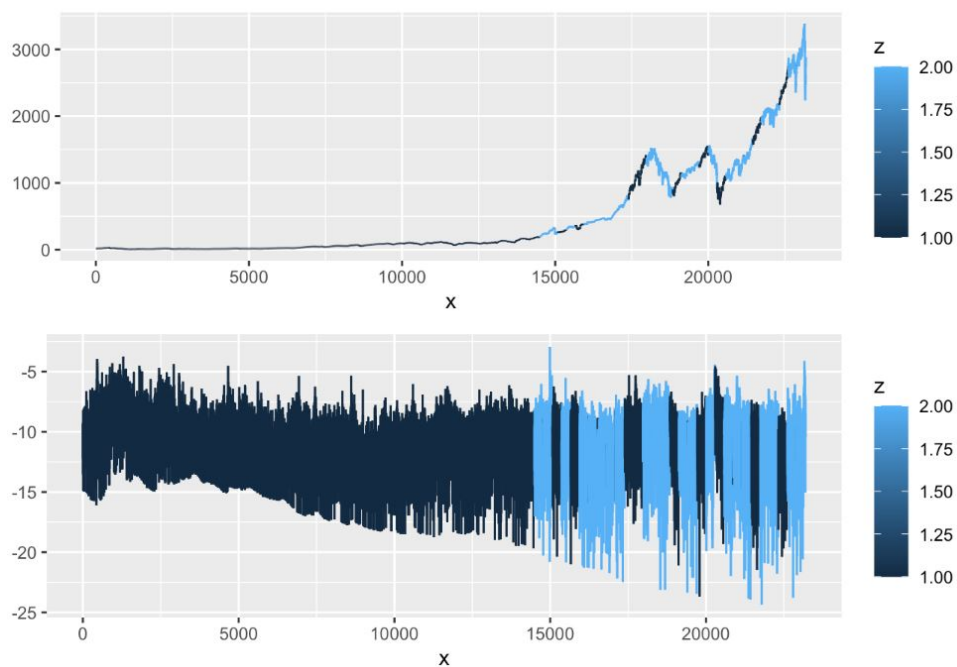


FIGURE 2.12: S&P 500, $k = 50$

Facebook Pour l'action FB, la linéarité est confirmée jusqu'à $k = 30$.

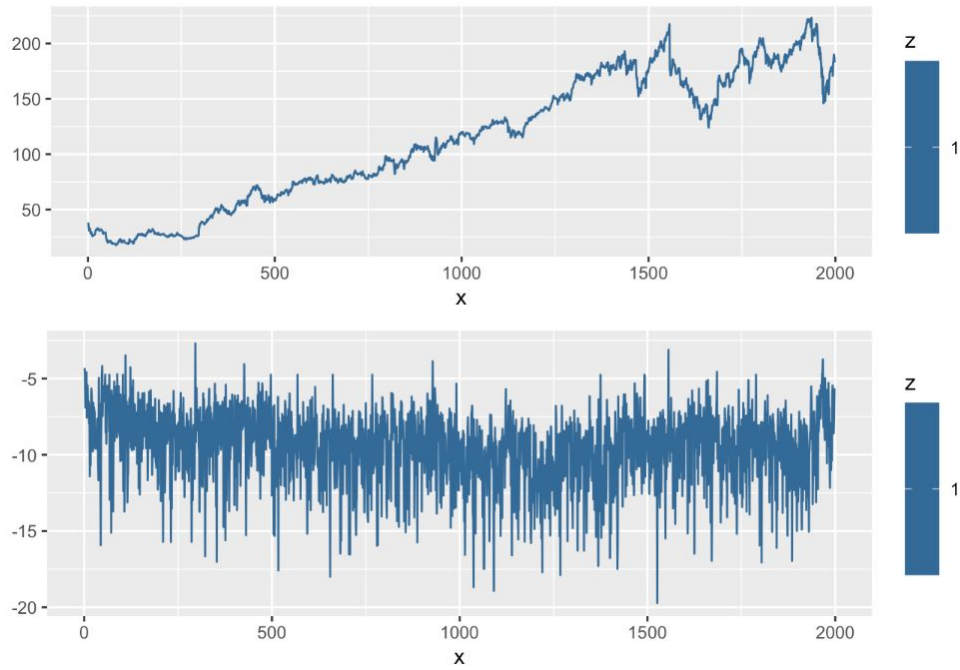


FIGURE 2.13: Facebook, $k = 30$

Pour $k = 50$, on accepte H_0 dans 90% des cas.

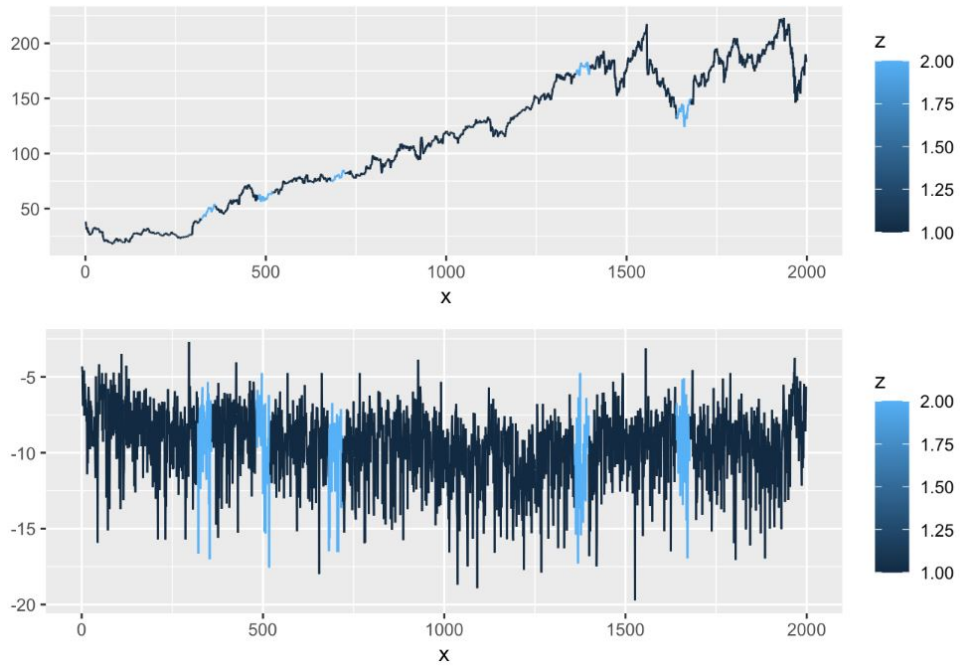


FIGURE 2.14: Facebook, $k = 50$

Amazon On rejette H_0 1 fois sur 2 pour $k = 2$.

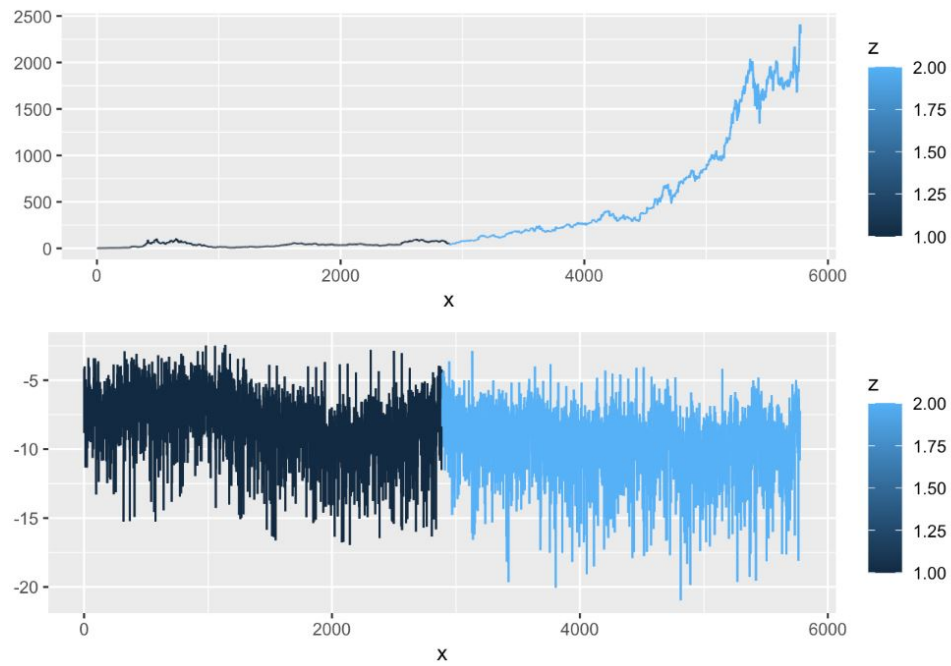


FIGURE 2.15: AMZN, $k = 2$

Pour $k = 50$, on accepte H_0 dans 96% des cas.

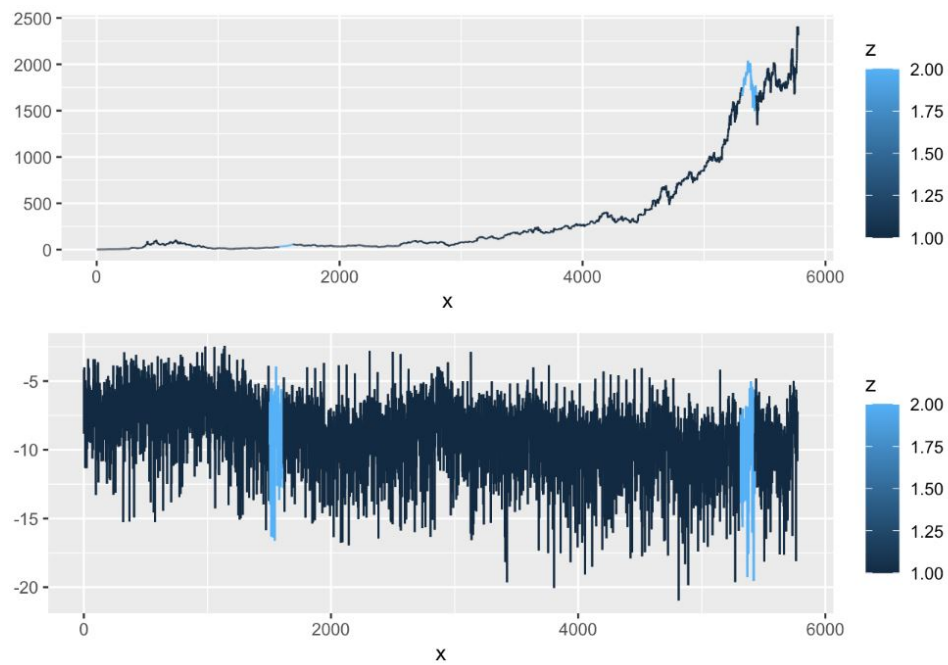


FIGURE 2.16: AMZN, $k = 50$

DEM/USD On rejette H_0 1 fois sur 2 pour $k = 2$.

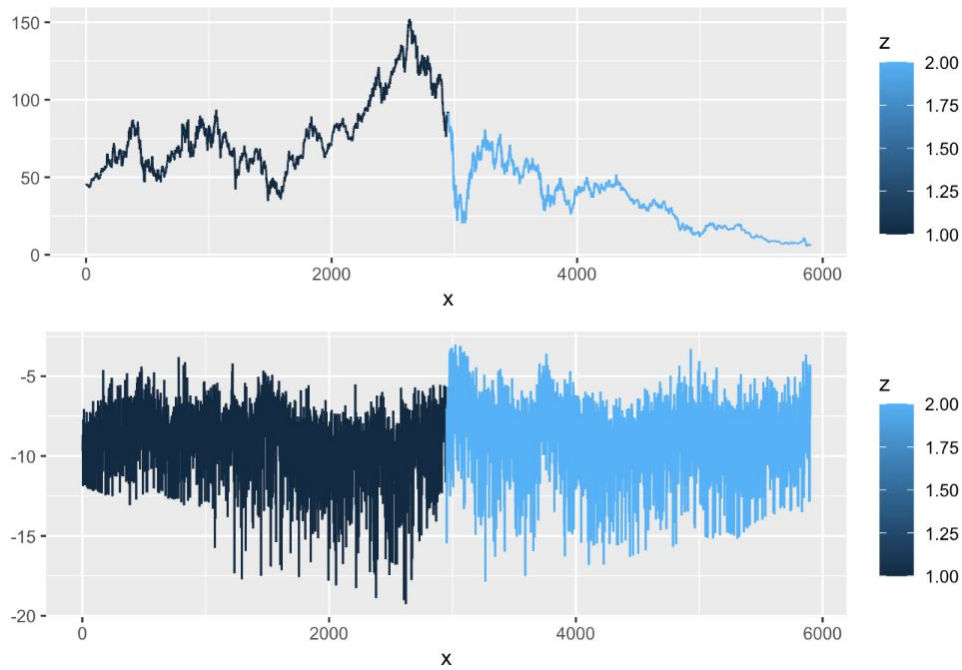


FIGURE 2.17: DEM/USD, $k = 2$

Pour $k = 50$, on accepte H_0 dans 88% des cas.

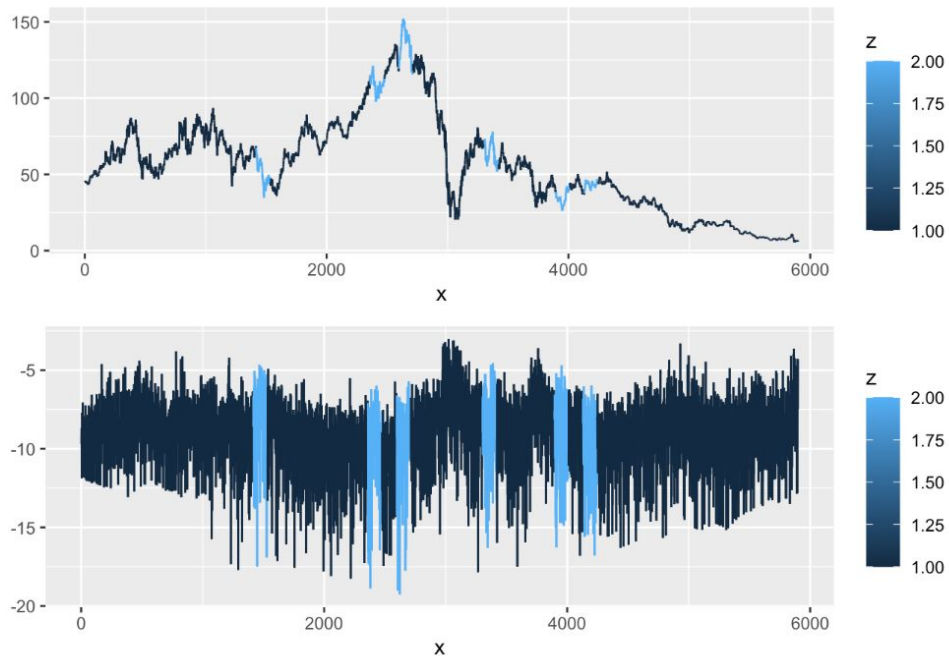


FIGURE 2.18: DEM/USD, $k = 50$

Pour ne pas surcharger notre mémoire, nous n'avons pas affiché les graphiques pour tous les k testés. L'observation globale que l'on peut faire est la suivante : en divisant en k groupes, on remarque que certaines dates ressortent, comme 2008, qui correspond effectivement à un krach boursier.

Chapitre 3

Puissance du test

3.1 L'hypothèse alternative

On s'intéresse à présent à l'hypothèse alternative.

La puissance d'un test est un élément important dans la construction de celui-ci. Elle représente la probabilité de rejeter quand il le faut, c'est à dire quand on est sous H_1 . Ainsi, plus la puissance est proche de 1, plus le test est adapté et bien construit.

Dans le cas du modèle présenté, l'alternative correspond à l'ensemble des fonctions non linéaires. C'est un ensemble assez large et non paramétrique, contrairement à H_0 .

Tout d'abord, on suppose que les fonctions de l'alternative sont telles que Y_t ne tende pas vers l'infini lorsque $t \rightarrow \infty$. On impose donc la condition suivante :

$$m(\cdot) \in \mathcal{M},$$

où $\mathcal{M} = \mathcal{M}(c, d)$ est l'ensemble des fonctions $m(\cdot)$ satisfaisant $|m(x)| \leq c|x| + d$, pour un certain $c \in (0, 1), d > 0$.

L'idée est d'établir une certaine "distance" par rapport à l'hypothèse nulle. Il serait alors pertinent de "perturber" une fonction linéaire et de tester la puissance du test avec la fonction obtenue.

On considère par exemple la fonction : $m(x) = \vartheta x + \lambda \sin(x)$, avec $\lambda > 0$ le paramètre qui détermine l'écart par rapport à la linéarité. On a bien dans ce cas que l'hypothèse ci-dessus est vérifiée : $|m(x)| \leq |x||\vartheta| + \lambda$, donc $m \in \mathcal{M}(c, d)$ pour un certain $c \in (0, 1), d > 0$, et avec $|\vartheta| < 1$.

3.2 Implémentation numérique

Nous allons générer les variables Z_t suivant le modèle suivant (on considère ici pour simplifier que ω est connu) :

$$\begin{aligned} Z_t &= Y_t + \xi_t \\ Y_t &= \vartheta Y_{t-1} + \lambda \sin(Y_{t-1}) + \varepsilon_t \end{aligned} \quad (3.1)$$

Par ailleurs, on utilisera les estimateur et test décrits dans la partie 1, mais sans diviser l'échantillon.

Nous simulons numériquement selon le modèle décrit ci-dessus, et en prenant ε_t et ξ_t indépendantes et *i.i.d* $N(0, 1)$.

On simule $p = 10000$ jeux de $n = 5000$ données. Voici les paramètres utilisés pour l'implémentation numérique.

```
a = 0.7
b = 0.99
n = 5000
p = 10000
mu1 = 0
nu1 = 1
mu2 = 0
nu2 = 1
theta = 0.95
```

FIGURE 3.1: Paramètres utilisés pour simuler des données sous H_1

On prend λ à valeurs dans $\{0, 0.2, 0.4, 0.5, 1\}$.

Nous avons représenté sur les figures suivantes les distributions de T_n pour différents λ .

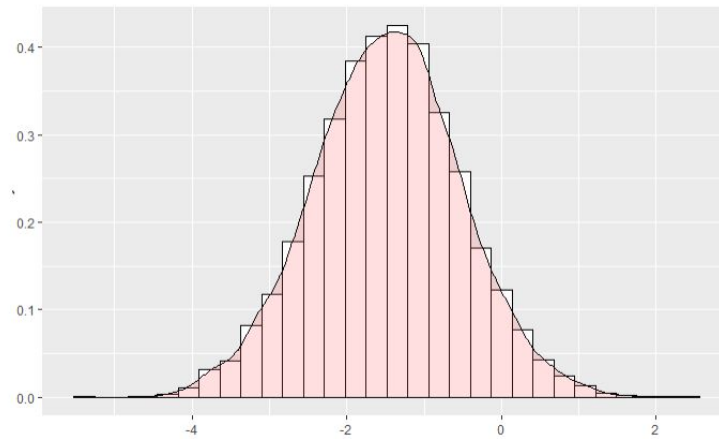
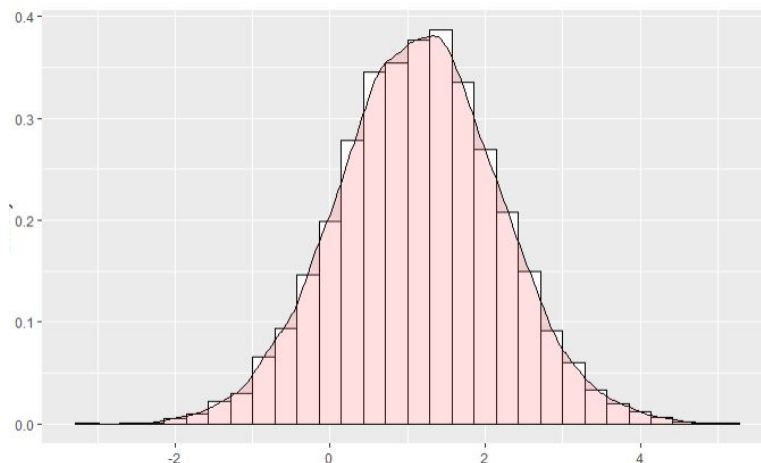
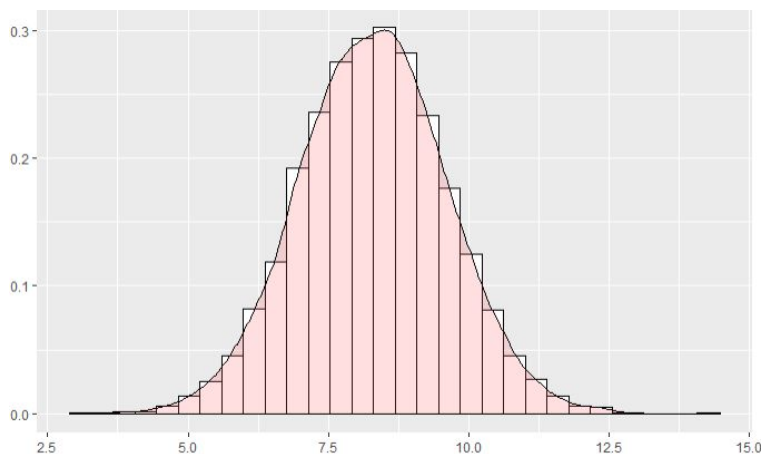


FIGURE 3.2: Distribution de T_n pour $\lambda = 0$

FIGURE 3.3: Distribution de T_n pour $\lambda = 0.5$ FIGURE 3.4: Distribution de T_n pour $\lambda = 1$

On remarque que les distributions se déplacent vers la droite (cf. voir les valeurs sur l'axe des abscisses de chaque graphique) lorsque λ augmente, ce qui traduit la consistance du test.

On effectue les tests avec un niveau de confiance de 95%. Voici par exemple le pourcentage de rejet lorsqu'on a pris $\lambda = 0.2$:

```
Data_0.2 = StockSampling3(n,p,mu1,nu1,mu2,nu2,theta,0.2)
Theta_0.2 = Theta(n,p>Data_0.2)
Thetachap_0.2 = Thetachap(a,b,n,p,Theta_0.2)
somme_0.2 = somme(a,b,n,p>Data_0.2,Thetachap_0.2)

H1=final_test2(n,p,somme_0.2)
y=H1/p
cat(' Pour lambda = 0.2, on rejette H0 dans',(H1/p)*100,'% des cas')
~~~

Pour lambda = 0.2, on rejette H0 dans 14.07 % des cas
```

FIGURE 3.5: Pourcentage de rejet pour $\lambda = 0.2$

Rappelons qu'il s'agit du pourcentage de rejet lorsque l'on s'est placé sous l'hypothèse alternative H_1 . On remarque que lorsque λ augmente, le pourcentage de rejet augmente aussi, ce qui est cohérent. En effet, lorsque λ est nul, on est sous H_0 . Ainsi, plus on "s'éloigne" de l'ensemble des fonctions linéaires, plus le pourcentage de rejet augmente.

En traçant le pourcentage de rejet en fonction de λ , on obtient :

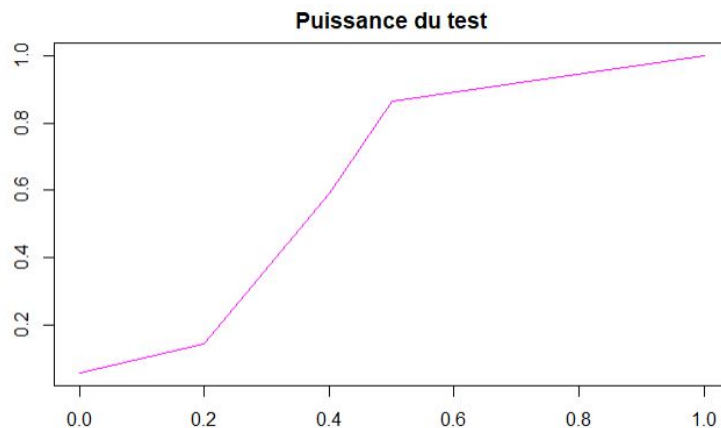


FIGURE 3.6: Puissance du test en fonction de λ

On remarque sur la figure 3.6 que la puissance tend vers 1 lorsque λ augmente. On peut donc fixer un certain λ qui permettra de définir une alternative moins large tout en assurant la puissance du test proposé.

Chapitre 4

Résultats théoriques

L'approche théorique définie au Chapitre 1 repose sur des résultats mathématiques. Cette section présente quelques uns de ces résultats qui ont permis de construire des tests et estimateurs cohérents avec le modèle.

Pour l'approche théorique, nous avons supposé les hypothèses suivantes :

A1. $\{\xi_t\}$ et $\{\varepsilon_t\}$ sont des variables aléatoires i.i.d. , et les deux séquences sont mutuellement indépendantes. La valeur $Y_0 = y_0$ est fixée.

A2. $\mathbf{E}\varepsilon_1 = \mathbf{E}\xi_1 = 0$
 $\mathbf{E}\xi_1^2 = \eta^2, \mathbf{E}\xi_1^3 = 0$ et $\mathbf{E}\xi_1^4 = \mu$
 $\mathbf{E}\varepsilon_1^2 = \sigma^2, \mathbf{E}\varepsilon_1^3 = 0$ et $\mathbf{E}\varepsilon_1^4 = \nu$
 $\mathbf{E}|\xi_1|^{4+\delta} < \infty, \mathbf{E}|\varepsilon_1|^{4+\delta} < \infty$ avec un certain $\delta > 0$.

A3. La variable aléatoire ε_1 a une densité $p(\cdot)$ par rapport à la mesure de Lebesgue, satisfaisant $p(x) > 0$ pour tout $x \in \mathbb{R}$.

4.1 Niveau du test (contrôle de l'erreur de 1ère espèce)

On rappelle le test construit dans le cas ω connu.

La statistique de test est donnée par :

$$T_n := \frac{1}{\sqrt{NB_n}} \sum_{t=N+1}^n \left\{ \left(Z_t - \hat{\vartheta}_n Z_{t-1} \right)^2 - M_n \right\}$$

On note Δ_n la règle de décision donnée par :

$$\Delta_n = \begin{cases} 0 & \text{if } T_n \leq t_\alpha \\ 1 & \text{if } T_n > t_\alpha \end{cases}$$

Le théorème suivant permet de justifier que le test construit est bien de niveau α .

THEOREME 1. Supposons **A1** et **A2**. Alors : $\limsup_{n \rightarrow \infty} \sup_{\vartheta \in [a,b]} \mathbf{P}_\vartheta \{ \Delta_n = 1 \} \leq \alpha$

4.2 Puissance du test (contrôle de l'erreur de 2nde espèce)

Comme expliqué au Chapitre 3, on peut définir l'alternative comme l'ensemble des fonctions linéaires auxquelles on a rajouté une "perturbation". En pratique, nous avons décidé de prendre comme perturbation $\lambda \sin(x)$ pour notre implémentation numérique. Nous avons ainsi pu voir que le test était puissant pour un certain λ (voir Figure 3.6 : Puissance du test en fonction de λ)

La construction théorique de l'alternative est la suivante. On pose :

$$d_n(m) = d_n(m(\cdot)) = \inf_{\vartheta \in [a,b]} \frac{1}{N} \sum_{t=N+1}^n \mathbf{E}_m (m(Y_{t-1}) - \vartheta Y_{t-1})^2$$

et :

$$\mathcal{M}_n(\lambda) = \{m \in \mathcal{M} : d_n(m) \geq \lambda/\sqrt{n}\}, \text{ avec } \lambda > 0.$$

On considère l'ensemble des alternatives :

$$\mathbf{H}_n : \quad m \in \mathcal{M}_n(\lambda)$$

Le théorème suivant permet de montrer que pour un $\lambda > 0$ assez grand, le test proposé atteint la puissance asymptotique qui est arbitrairement proche de 1 uniformément sur l'ensemble des alternatives non paramétriques $\mathcal{M}_n(\lambda)$.

THEOREME 2. On suppose A1 – A3. Alors pour tout $0 < \alpha, \beta < 1$, il existe une constante $\lambda(\alpha, \beta)$ telle que : $\limsup_{n \rightarrow \infty} \sup_{m \in \mathcal{M}_n(\lambda(\alpha, \beta))} \mathbf{P}_m \{\Delta_n = 0\} \leq \beta$.

Ce théorème montre qu'on a donc aussi un "contrôle" sur l'erreur de seconde espèce, ce qui permet d'avoir un test puissant.

Conclusion

Notre mémoire, basé sur l'article de recherche *Testing Linearity in an AR Errors-in-variables Model with Application to Stochastic Volatility*, présente un volet théorique et une mise en application numérique.

La cohérence des estimateurs & tests statistiques construits pour ω connu a été confirmée grâce à des simulations sur R. Pour le cas ω inconnu, la procédure de test proposée dans la partie théorique a été appliquée pour différents indices boursiers.

Tous les résultats que nous obtenons se basent sur des tests statistiques dont la puissance a été démontrée dans la troisième partie de ce mémoire. Les résultats théoriques sur lesquels reposent la construction des tests ont quant à eux été résumés à la partie 4.

Les dates de "non-volatilité" que l'on a obtenu pour chaque indice sont à titre indicatif, puisqu'elles dépendent de la valeur prise par k . Néanmoins, il semblerait qu'il y ait une certaine linéarité jusqu'aux années 2000. On pourrait émettre l'hypothèse qu'à partir de la crise financière de 2008, chaque indice a "réagi" différemment. On peut donc affirmer que les crises économiques ont un impact significatif sur la volatilité.

En conclusion, les résultats obtenus semblent soutenir l'hypothèse d'une structure de volatilité paramétrique pour certains des ensembles de données analysés.

Bibliographie

- [1] D. Feldmann, W. Härdle, C. Hafner, M. Hoffmann, O. Lepski, A. Tsybakov (2003), *Testing Linearity in an AR Errors-in-variables Model with Application to Stochastic Volatility*.
- [2] R. Engle (1982), *Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation*, *Econometrica* 50, 987-1008.
- [3] C. Gouriéroux (1997), *ARCH Models and Financial Applications*, Springer, New York.
- [4] S. J. Taylor (1986), *Modeling Financial Time Series*, Wiley, Chichester.
- [5] Imen Ben Tahar, *Mouvement brownien & évaluation des actifs contingents*, Cours de M1 Mathématiques Appliquées (2020), Polycopié de l'Université Paris-Dauphine.
- [6] Christian Robert, *Estimation Statistique*, Cours de L3 Mathématiques (2018-2019), Polycopié de l'Université Paris-Dauphine.
- [7] Marc Hoffmann, *Tests statistiques*, Cours de L3 Mathématiques (2019), Polycopié de l'Université Paris-Dauphine.
- [8] Julien Stoeck, *Méthodes de Monte Carlo*, Cours de M1 Mathématiques Appliquées (2019-2020), Polycopié de l'Université Paris-Dauphine.