Modèle de point de rupture pour le modèle AR(1) sur le CAC 40

Groupe 1
AMJAHID Mohamed Amin, CHEAM Richard
EL GORGI Skander, LY Sovanleng
NOUV Ratanakmuny, THUO Menghor
ZAHRAOUI Imane, ZHANG Xichen

Université Paris-Saclay
ENSIIE

Ce rapport est soumis pour
Modélisation Statistique - Séries Temporelles

April 2024

Abstract

Ce rapport examine la stabilité structurelle de l'indice boursier CAC40 en utilisant un modèle autorégressif avec un point de rupture (modèle AR(1)). L'objectif principal est de détecter d'éventuelles ruptures structurelles dans les rendements logarithmiques de l'indice CAC40, ce qui pourrait indiquer des changements significatifs dans la dynamique du marché. L'analyse commence par la collecte des prix de clôture ajustés quotidiens à partir du 1er janvier 2021. À l'aide de ces prix, les rendements logarithmiques sont calculés pour des retards d'un à quatre jours. L'étude utilise un modèle AR(1) pour ajuster ces rendements, puis teste les points de rupture à l'aide du test de Kolmogorov-Smirnov. Ce test non paramétrique évalue s'il existe des différences significatives dans la distribution des rendements avant et après chaque point de rupture potentiel. La méthodologie implique une analyse extensive des données, un ajustement du modèle et des tests d'hypothèses pour valider l'existence des points de rupture.

Les résultats pourraient fournir de nouvelles perspectives sur le timing et la nature des changements de régime au sein de l'indice CAC40, offrant des informations précieuses pour les investisseurs, les décideurs et les théoriciens économiques concernant la stabilité sous-jacente et la prévisibilité du marché.

Table of contents

Chapter 1

Introduction

Le CAC40, représentant un segment important du marché boursier français, constitue un indicateur majeur des tendances économiques et financières en Europe. Analyser sa stabilité et identifier les changements structurels potentiels au sein de sa dynamique de prix sont essentiels pour comprendre le comportement du marché et améliorer les processus de prise de décision financière. Cette étude se concentre sur la détection de ces ruptures structurelles en utilisant des modèles autorégressifs spécifiquement conçus pour prendre en compte les éventuels changements dans le processus de génération des données sous-jacentes.

Les séries temporelles financières, telles que les prix des actions, sont souvent modélisées en utilisant les rendements logarithmiques en raison de leurs propriétés de stationnarité et de comparabilité dans le temps. Le rendement logarithmique au temps t pour un retard m est défini comme suit :

$$Y_t^{(m)} = \log\left(\frac{X_t}{X_{t-m}}\right)$$

où X_t représente le prix de clôture ajusté de l'action au temps t. Cette transformation normalise les rendements, permettant une application plus aisée des modèles statistiques linéaires.

Dans ce contexte, un modèle autorégressif d'ordre 1 (AR(1)) est utilisé pour modéliser ces rendements :

$$Y_t^{(m)} = \beta Y_{t-1}^{(m)} + \varepsilon_t \tag{1.1}$$

où β est le coefficient à estimer et ε_t est le terme d'erreur, généralement supposé être indépendant et identiquement distribué avec une moyenne de zéro et une variance σ^2.

Cependant, les modèles AR(1) traditionnels supposent une constance des paramètres dans le temps, une hypothèse qui peut ne pas être vérifiée dans des marchés volatils. Pour aborder cela, notre étude introduit un modèle de rupture, permettant aux paramètres de changer à un point inconnu k dans l'échantillon :

$$Y_t^{(m)} = \begin{cases} a_1 Y_{t-1}^{(m)} + \varepsilon_t & \text{if } t \leq k \\ a_2 Y_{t-1}^{(m)} + \varepsilon_t & \text{if } t > k \end{cases}$$

Ici, a_1 et a_2 sont les paramètres après le point de rupture, reflétant un potentiel changement dans la dynamique du marché.

———————————-

Pour déterminer la présence et l'emplacement de ces points de rupture, le test de Kolmogorov-Smirnov est utilisé pour comparer la distribution de $Y_t^{(m)}$ avant et après chaque candidat k. L'hypothèse nulle H_0 suppose l'absence de point de rupture, ce qui implique que les rendements suivent la même distribution tout au long de la période :

$$H_0 : \text{ Aucune rupture n'existe et } Y_1, \ldots, Y_k \text{ ainsi que } Y_{k+1}, \ldots, Y_n \text{ suivent la même loi}$$

L'hypothèse alternative H_1 suggère une rupture structurelle, indiquant des lois différentes régissant les périodes avant et après k:

$$H_1 : \text{ Une rupture existe, indiquant des processus générateurs différents}$$

———————————-

Notre rapport est structuré pour analyser en détail l'indice CAC40 à l'aide de techniques statistiques approfondies. Le chapitre 2 traite des aspects théoriques du modèle autorégressif, des méthodes d'estimation telles que le Maximum de Vraisemblance et les Moindres Carrés Ordinaires, ainsi que des tests de Kolmogorov-Smirnov. Le chapitre 3 applique ces méthodologies au CAC40, en présentant les résultats, les visualisations et les présentations structurées des données. La conclusion du chapitre 4 passe en revue les conclusions et propose des orientations pour de futures recherches. Les appendices A fournissent le code d'analyse et approfondissent l'exploration du modèle de détection des points de rupture, assurant ainsi un traitement complet des applications statistiques en finance.

Chapter 2

Statistiques

2.1 Modèle autorégressif

Un modèle autorégressif (AR) est un type de modèle statistique largement utilisé en analyse de séries temporelles pour décrire des processus dépendants du temps. L'idée principale du modèle est que la valeur actuelle d'une série temporelle peut être expliquée comme une fonction de ses valeurs précédentes, plus un terme d'erreur stochastique. Il suppose une *relation linéaire* entre les valeurs passées.

Cette section offre une exploration complète du modèle AR(1), détaillant sa formulation mathématique(Sec.2.1.1), ainsi que les méthodes d'estimation(Sec.2.1.2).

2.1.1 Modèle AR(1)

Le modèle AR(1) général est différent de notre AR(1) (2.1.1). Dans ce modèle, la valeur de x au temps t est une fonction linéaire de la valeur de x au temps $t-1$. L'expression algébrique du modèle est la suivante [4]:

$$x_t = \delta + \phi_1 x_{t-1} + w_t$$

Hypothèses

- $w_t \sim$ i.i.d $N(0, \sigma_w^2)$, ce qui signifie que les erreurs sont indépendamment et identiquement distribuées selon une loi normale avec une moyenne de 0 et une variance constante.
- Les propriétés des erreurs w_t are sont indépendantes de x_t.

- La série $x_1, x_2, \ldots$ est (faiblement) stationnaire. Une condition préalable pour qu'un modèle AR(1) soit stationnaire est que $|\phi_1| < 1$.

Propriétés du modèle AR(1)

Les formules pour la moyenne, la variance et l'autocorrélation pour un processus de séries temporelles avec un modèle AR(1) sont les suivantes[3].

- La moyenne (théorique) de x_t est

$$E(x_t) = E(\delta + \phi_1 x_{t-1} + w_t) = E(\delta) + E(\phi_1 x_{t-1}) + E(w_t) = \delta + \phi_1 E(x_{t-1}) + 0 \quad (2.1)$$

Avec l'hypothèse de stationnarité, $E(x_t) = E(x_{t-1})$. Soit μ la moyenne commune. Ainsi $\mu = \delta + \phi_1 \mu$. Résolvons pour μ

$$\mu = \frac{\delta}{1 - \phi_1} \quad (2.2)$$

- La variance de x_t est

$$\mathrm{Var}(x_t) = \mathrm{Var}(\delta) + \mathrm{Var}(\phi_1 x_{t-1}) + \mathrm{Var}(w_t) = \phi_1^2 \mathrm{Var}(x_{t-1}) + \sigma_w^2 \quad (2.3)$$

En vertu de l'hypothèse de stationnarité, $\mathrm{Var}(x_t) = \mathrm{Var}(x_{t-1})$. Remplaçons $\mathrm{Var}(x_t)$ Par $\mathrm{Var}(x_{t-1})$ et résolvons pour $\mathrm{Var}(x_t)$. Comme $\mathrm{Var}(x_t) > 0$, il s'ensuit que $(1 - \phi_1^2) > 0$ et donc $|\phi_1| < 1$.

- La corrélation entre les observations séparées de h périodes est

$$\rho_h = \phi_1^h$$

Cela définit la fonction d'autocorrélation théorique pour une variable de série temporelle avec un modèle AR(1).

2.1.2 Méthode du Maximum de Vraisemblance

Pour $m = 1, 2, 3, 4$, on s'intéresse au comportement de :

$$Y_t^{(m)} = log(\frac{X_t}{X_{t-m}})$$

et nous considérons le modèles suivant :

$$Y_t^{(m)} = \begin{cases} a_1 Y_{t-1}^{(m)} + \varepsilon_t & si\ t \leq k \\ a_2 Y_{t-1}^{(m)} + \varepsilon_t & si\ t > k \end{cases} \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

où a_1, k, a_2, σ^2 sont inconnus que l'on souhaite estimer.

Pour chaque observation dans le Régime 1 (avant le point de rupture) :

$$L(Y_t^{(m)} | Y_{t-1}^{(m)}, a_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_t^{(m)} - a_1 Y_{t-1}^{(m)})^2}{2\sigma^2} \right) \quad \text{for } t \leq k \tag{2.4}$$

Pour chaque observation dans le Régime 2 (après le point de rupture) :

$$L(Y_t^{(m)} | Y_{t-1}^{(m)}, a_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_t^{(m)} - a_2 Y_{t-1}^{(m)})^2}{2\sigma^2} \right) \quad \text{for } t > k \tag{2.5}$$

En supposant l'indépendance entre les observations,
la vraisemblance globale $\mathcal{L}$ pour la série est :

$$\mathcal{L}(a_1, a_2, k, \sigma^2 | \text{data}) = \left(\prod_{t=1}^{k} L(Y_t^{(m)} | Y_{t-1}^{(m)}, a_1, \sigma^2) \right) \cdot \left(\prod_{t=k+1}^{n} L(Y_t^{(m)} | Y_{t-1}^{(m)}, a_2, \sigma^2) \right) \tag{2.6}$$

Cela peut être exprimé sous forme logarithmique (log-vraisemblance) pour simplifier les calculs :

$$\begin{aligned} \log \mathcal{L} &= \sum_{t=1}^{k} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(Y_t^{(m)} - a_1 Y_{t-1}^{(m)})^2}{2\sigma^2} \right) + \sum_{t=k+1}^{n} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(Y_t^{(m)} - a_2 Y_{t-1}^{(m)})^2}{2\sigma^2} \right) \\ &= -n \log(\sqrt{2\pi}\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{t=1}^{k} \left(Y_t^{(m)} - a_1 Y_{t-1}^{(m)} \right)^2 + \sum_{t=k+1}^{n} \left(Y_t^{(m)} - a_2 Y_{t-1}^{(m)} \right)^2 \right) \end{aligned}$$

$$\text{Dérivées partielles :} \begin{cases} \dfrac{\partial}{\partial a_1} \log \mathcal{L} = \dfrac{1}{\sigma^2} \displaystyle\sum_{t=1}^{k} Y_{t-1}^{(m)} (Y_t^{(m)} - a_1 Y_{t-1}^{(m)}) = 0 \\ \dfrac{\partial}{\partial a_2} \log \mathcal{L} = \dfrac{1}{\sigma^2} \displaystyle\sum_{t=k+1}^{n} Y_{t-1}^{(m)} (Y_t^{(m)} - a_2 Y_{t-1}^{(m)}) = 0 \\ \dfrac{\partial}{\partial \sigma^2} \log \mathcal{L} = -n + \dfrac{1}{2\sigma^2} \left(\displaystyle\sum_{t=1}^{k} \left(Y_t^{(m)} - a_1 Y_{t-1}^{(m)} \right)^2 + \sum_{t=k+1}^{n} \left(Y_t^{(m)} - a_2 Y_{t-1}^{(m)} \right)^2 \right) = 0 \end{cases}$$

imply that:

$$\begin{cases} \hat{a}_1(k) = \dfrac{\sum\limits_{t=1}^{k} Y_t^{(m)} Y_{t-1}^m}{\sum\limits_{t=1}^{k} Y_{t-1}^m} & (1) \\ \hat{a}_2(k) = \dfrac{\sum\limits_{t=k+1}^{n} Y_t^{(m)} Y_{t-1}^m}{\sum\limits_{t=k+1}^{n} Y_{t-1}^m} & (2) \\ \hat{\sigma^2}(k) = -n + \dfrac{1}{2\sigma^2}\left(\sum\limits_{t=1}^{k}\left(Y_t^{(m)} - a_1 Y_{t-1}^{(m)}\right)^2 + \sum\limits_{t=k+1}^{n}\left(Y_t^{(m)} - a_2 Y_{t-1}^{(m)}\right)^2\right) & (3) \end{cases}$$

Pour trouver l'estimateur de k on injecte ces valeur dans $\log \mathscr{L}$. Ce qui donne:

$$\Lambda_n(k) = -n \log(\hat{\sigma^2}(k)) + C^{te}$$

$\hat{k}$ réalise le max (indice) de $\Lambda_n(k)$

2.2 Tests d'homogénéité de Kolmogorov-Smirnov

2.2.1 Présentation théorique du test

Le test de Kolmogorov-Smirnov (K-S) est une méthode statistique couramment utilisée pour déterminer si un échantillon est conforme à une distribution de probabilité donnée ou à un autre ensemble de données. Cependant, pour ce projet, nous souhaitons rechercher l'existence d'un point de rupture; par conséquent, le test K-S à deux échantillons sera utilisé afin de vérifier si deux ensembles de données (sous-ensemble avant et après le point de rupture potentiel) suivent tous deux la même loi[2].

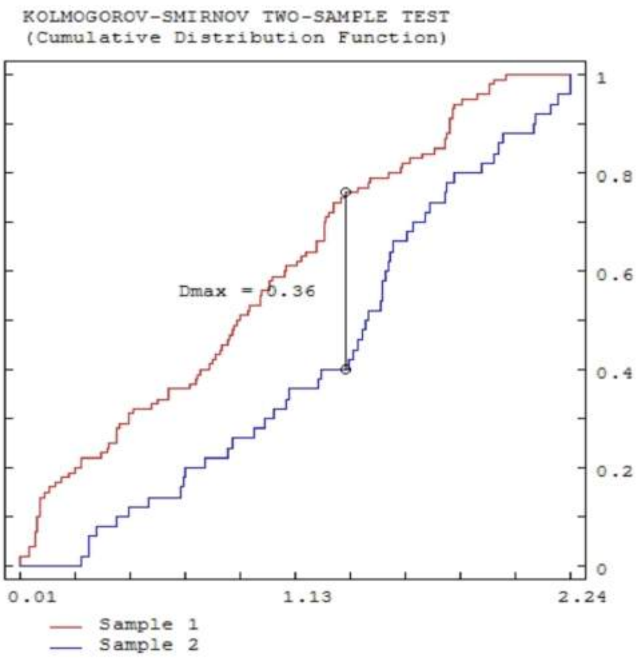

Fig. 2.1 Un exemple de représentation visuelle du test K-S à deux échantillons

Étant donné deux ensembles de données $X_1,...,X_{N_1}$ et $X_1,...,X_{N_2}$, la statistique K-S est la différence maximale entre la fonction de distribution cumulative de ces deux ensembles de données, notée par

$$D_{N_1,N_2} = \max_{-\infty < X < +\infty} |S_{N_1}(X) - S_{N_2}(X)|$$

où $S_{N_1}(X)$ and $S_{N_2}(X)$ sont les fonctions de répartition empiriques (la fraction de N valeurs de données qui sont inférieures ou égales à t (à gauche de x), également connue sous le nom de "fonction staircase" qui augmente par $\frac{1}{N}$) du premier et du second échantillon respectivement. Précisément, la fonction de répartition est une fonction en escalier, croissante, continue à droite et admettant une limite à gauche et donc la fonction de distribution cumulative empirique (eCDF) s'écrit aussi couramment comme suit[1]:

$$F_N(t) = \frac{1}{N}\sum_{i=1}^{N} 1_{X_i \leq t}$$

où 1 est la fonction indicatrice:

$$1_{X_i \leq t} = \begin{cases} 1 & \text{si } X_i \leq t \\ 0 & \text{sinon} \end{cases}$$

Le test K-S est basé sur deux hypothèses (nulle et alternative) :

$$\begin{cases} H_0 : X_1,...,X_{N_1} \text{ et } X_1,...,X_{N_2} \text{ suivent la même loi} \\ H_1 : X_1,...,X_{N_1} \text{ et } X_1,...,X_{N_2} \text{ ne suivent pas la même loi} \end{cases}$$

L'hypothèse nulle (H_0) est rejetée sous une condition si p-value $> \alpha$, où α est le niveau de signification choisi et vice versa. La valeur p du test K-S est calculée en comparant la valeur observée de la statistique K-S (D_{N_1,N_2}) à la valeur critique de la statistique K-S sous l'hypothèse nulle que les deux échantillons proviennent de la même distribution. De même, à l'aide du théorème de Glivenko-Cantelli (une généralisation de la loi forte des grands nombres au cas non-paramétrique), avec quelques manipulations et en posant une variable aléatoire $\sqrt{\frac{N_1N_2}{N_1+N_2}}D_{N_1,N_2}$, avec $\lambda > 0$, on a le résultat comme suit:

$$P(\sqrt{\frac{N_1N_2}{N_1+N_2}}D_{N_1,N_2} \leq \lambda) \to +2\sum_{k=1}^{+\infty}(-1)^k e^{-(2k^2\lambda^2)}$$

Autrement dit, l'hypothèse nulle est rejetée au niveau α si:

$$D_{N_1,N_2} > \sqrt{-\frac{1}{2}\log(\frac{\alpha}{2})}\sqrt{\frac{N_1+N_2}{N_1N_2}}$$

En outre, une statistique K-S élevée implique une différence significative entre les distributions des deux échantillons, tandis qu'une statistique K-S faible suggère une divergence moindre entre les distributions. Cependant, l'interprétation de la statistique K-S doit être considérée en conjonction avec p-value.

2.2.2 Étude de cas pratique du test

On observe deux échantillons $(X_1,...,X_k)$ et $(X_{k+1},...,X_n)$, séparées par la rupture à X_k. On note F_1 et F_2 la fonction de répartition de chacune des variables $(X_1,...,X_k)$ et $(X_{k+1},...,X_n)$ respectivement. Donc, le but est de tester si les deux échantillons sont issus d'une même loi (hypothèse H_0) et dans le cas contraire (hypothèse H_1). En considerant que les données sont ordonnées et les valeurs $t_{i=1,..,n} > 0$, donc on a:

$$D_{k,n-k}(t_i) = |F_1(t_i) - F_2(t_i)| = |\frac{1}{k}\sum_{i=1}^{k} 1_{X_i \leq t_i} - \frac{1}{n-k}\sum_{i=k+1}^{n} 1_{X_i \leq t_i}|$$

D'après le test de K-S, on prend la valeur maximale de la distance D, donc:

$$D_{k,n-k} = \max_{t \in R} |F_1(t) - F_2(t)|$$

De plus, en posant $\lambda = \sqrt{\frac{k(n-k)}{n}} D_{k,n-k}$:

$$P(\sqrt{\frac{k(n-k)}{n}} D_{k,n-k} \leq \lambda) \to 1 + 2 \sum_{i=1}^{+\infty} (-1)^i e^{-(2i^2\lambda^2)}$$

$$\text{p-value} = P(\sqrt{\frac{k(n-k)}{n}} D_{k,n-k} \geq \lambda) \to -2 \sum_{i=1}^{+\infty} (-1)^i e^{-(2i^2\lambda^2)}$$

Pour ce projet, le niveau de signification choisi est $\alpha = 0,05$. Alors, si la valeur p (p-value) obtenue est plus petite que α, nous n'avons pas assez de preuves pour conclure que les deux échantillons suivent la même loi et vice versa. En termes de codage, afin de faciliter la tâche, nous utiliserons l'aide de la bibliothèque Python `stats.kstwo.sf` (Survival Function) pour calculer p-value et après le comparer avec le test K-S à deux échantillons de `scipy.stats.ks_2samp` de Python.

Chapter 3

Application sur les données CAC 40

Dans cette section, on applique les deux méthodes présentées précédemment sur les données du CAC 40 que nous avions extraites à partir du 1er janvier 2021.

3.1 Résultats

Table 3.1 Stock Parameters (m = 1).

	AC.PA	ACA.PA	AI.PA	AIR.PA	ATO.PA	BN.PA	BNP.PA	CA.PA
a_1	-0.017536	0.218320	-0.073651	0.168762	0.478666	-0.321184	0.089181	0.166129
a_2	0.045741	0.049660	0.001992	-0.061500	-0.032631	-0.014962	0.027805	-0.008566
σ	0.000196	0.000133	0.000086	0.000159	0.001275	0.000064	0.000187	0.000115
k*	39	35	316	27	4	39	298	121

	CAP.PA	DG.PA	EN.PA	ENGI.PA	GLE.PA	HO.PA	KER.PA	LR.PA
a_1	0.005826	0.135892	0.152823	0.208768	-0.024362	0.044230	-0.302689	-0.007952
a_2	0.019079	-0.012461	-0.005329	0.007617	0.066977	-0.027123	-0.014094	-0.014099
σ	0.000172	0.000086	0.000076	0.000124	0.000259	0.000142	0.000184	0.000109
k*	47	79	225	113	48	40	6	75

	MC.PA	ML.PA	OR.PA	ORA.PA	RI.PA	SAN.PA	SGO.PA	SU.PA
a_1	-0.065865	-0.002148	-0.012072	0.011828	-0.177112	-0.016464	0.162074	-0.014067
a_2	0.020317	0.012535	-0.101010	0.248575	0.002173	-0.068646	0.034979	-0.024356
σ	0.000179	0.000134	0.000126	0.000040	0.000107	0.000129	0.000192	0.000160
k*	413	324	8	382	336	160	14	292

	SW.PA	TTE.PA	VIE.PA	VIV.PA	WLN.PA
a_1	0.104444	0.018533	0.002074	0.088898	0.049930
a_2	0.002277	-0.059648	0.018417	-0.004376	-0.003546
σ	0.000117	0.000147	0.000162	0.000090	0.001072
k*	307	19	428	65	558

Table 3.1 Stock Parameters (m = 2).

	AC.PA	ACA.PA	AI.PA	AIR.PA	ATO.PA	BN.PA	BNP.PA	CA.PA
a_1	0.016724	0.540381	0.324576	0.401198	-0.296982	0.003648	-0.109541	0.493789
a_2	0.514425	0.007936	0.031993	0.025652	0.003289	0.371933	0.551452	-0.071123
σ	0.000308	0.000205	0.000134	0.000244	0.003862	0.000094	0.000311	0.000192
k*	18	547	463	484	557	9	38	488

	CAP.PA	DG.PA	EN.PA	ENGI.PA	GLE.PA	HO.PA	KER.PA	LR.PA
a_1	-0.039043	0.348058	0.547267	0.560981	-0.106159	0.065593	-0.182234	-0.023123
a_2	0.576442	0.016623	0.061827	0.071725	0.448814	0.311802	0.396485	0.494013
σ	0.000264	0.000138	0.000123	0.000202	0.000464	0.000250	0.000439	0.000163
k*	4	463	391	276	44	73	464	18

	MC.PA	ML.PA	OR.PA	ORA.PA	RI.PA	SAN.PA	SGO.PA	SU.PA
a_1	0.540548	-0.060779	-0.045082	0.024358	0.721597	0.636184	0.609195	0.465902
a_2	-0.033580	0.522330	0.466583	0.564093	-0.057666	-0.054152	-0.020393	0.011796
σ	0.000262	0.000218	0.000188	0.000065	0.000189	0.000190	0.000299	0.000245
k*	552	36	3	84	335	467	553	544

	SW.PA	TTE.PA	VIE.PA	VIV.PA	WLN.PA
a_1	0.537393	0.065868	-0.079388	-0.056762	-0.267450
a_2	0.071163	0.250966	0.250966	0.491342	0.851537
σ	0.000194	0.000242	0.000268	0.000155	0.002530
k*	311	109	42	126	466

Table 3.1 Stock Parameters (m = 3).

	AC.PA	ACA.PA	AI.PA	AIR.PA	ATO.PA	BN.PA	BNP.PA	CA.PA
a_1	0.016176	0.712705	0.453535	0.596704	-0.420860	0.005475	-0.150022	0.723580
a_2	0.695049	0.026907	0.046107	0.031852	-0.017474	0.485863	0.664945	-0.092760
σ	0.000333	0.000217	0.000166	0.000257	0.007885	0.000113	0.000378	0.000228
k*	12	557	463	483	557	9	38	488

	CAP.PA	DG.PA	EN.PA	ENGI.PA	GLE.PA	HO.PA	KER.PA	LR.PA
a_1	-0.065674	0.454375	0.646521	0.775459	-0.155936	0.091165	-0.237005	-0.033067
a_2	0.631268	0.024623	0.085804	0.101325	0.652556	0.379860	0.596745	0.669560
σ	0.000319	0.000163	0.000151	0.000223	0.000604	0.000333	0.000725	0.000200
k*	15	462	392	273	44	73	462	17

	MC.PA	ML.PA	OR.PA	ORA.PA	RI.PA	SAN.PA	SGO.PA	SU.PA
a_1	-0.041227	-0.069706	-0.055962	0.034482	1.129260	0.609337	0.843617	0.657446
a_2	0.626485	0.610453	0.561339	0.672323	-0.080272	-0.055898	-0.026384	0.017326
σ	0.000312	0.000255	0.000219	0.000083	0.000287	0.000229	0.000352	0.000268
k*	3	36	3	83	335	526	552	543

	SW.PA	TTE.PA	VIE.PA	VIV.PA	WLN.PA
a_1	0.702852	0.312302	-0.117844	-0.075749	-0.005661
a_2	0.097459	0.072664	0.588175	0.646429	-0.394841
σ	0.000241	0.000272	0.000336	0.000185	0.004547
k*	311	387	42	125	2

Table 3.1 Stock Parameters (m = 4).

	AC.PA	ACA.PA	AI.PA	AIR.PA	ATO.PA	BN.PA	BNP.PA	CA.PA
a_1	0.022208	0.840067	0.477244	0.669405	-0.173456	0.010114	-0.184403	1.014459
a_2	0.756623	0.879299	0.056968	0.038967	-0.492877	0.532418	0.871234	-0.106753
σ	0.000362	0.000224	0.000196	0.000292	0.012994	0.000125	0.000426	0.000254
k*	11	497	463	482	1	8	37	489

	CAP.PA	DG.PA	EN.PA	ENGI.PA	GLE.PA	HO.PA	KER.PA	LR.PA
a_1	-0.069703	0.502193	0.589185	0.718765	-0.208194	0.107387	-0.266729	-0.041580
a_2	0.726584	0.030335	0.077532	0.123596	0.827937	0.371674	0.756933	0.800955
σ	0.000335	0.000177	0.000165	0.000279	0.000845	0.000420	0.001023	0.000214
k*	15	461	463	309	43	72	462	16

	MC.PA	ML.PA	OR.PA	ORA.PA	RI.PA	SAN.PA	SGO.PA	SU.PA
a_1	0.815080	-0.081567	-0.080894	0.039289	1.085267	0.721537	0.993114	0.732152
a_2	-0.012319	0.723923	0.735609	0.833105	-0.091646	-0.072683	-0.022148	0.022337
σ	0.000316	0.000269	0.000221	0.000089	0.000330	0.000237	0.000367	0.000278
k*	550	35	2	83	335	531	551	542

	SW.PA	TTE.PA	VIE.PA	VIV.PA	WLN.PA
a_1	0.514740	0.325579	-0.139567	-0.089033	-0.041815
a_2	0.067859	0.076585	0.808907	0.745238	-0.434672
σ	0.000267	0.000325	0.000377	0.000215	0.006389
k*	449	387	41	124	1

3.2 Graphiques

3.2.1 AC.PA

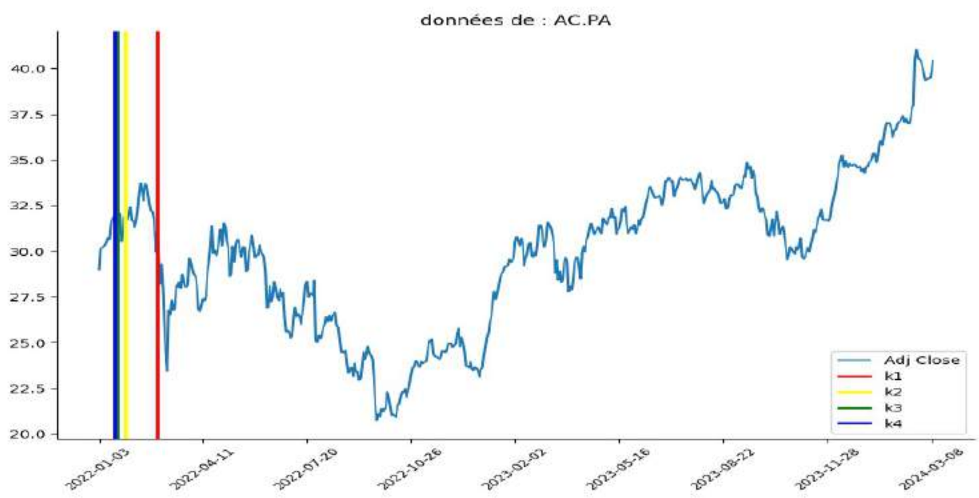

Fig. 3.1 Cours de l'action de AC.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le ’2022-01-18’, ’2022-01-19’, ’2022-01-27’ et le ’2022-02-25’.

3.2.2 ACA.PA

Fig. 3.2 Cours de l’action de ACA.PA du 01/01/2022 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le ’2022-02-21’, ’2023-12-27’, ’2024-02-20’ et le ’2024-03-05’.

3.2.3 AI.PA

Fig. 3.3 Cours de l’action de AI.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le ’2023-03-24’ et le ’2023-10-20’.

3.2.4 AIR.PA

Fig. 3.4 Cours de l'action de AIR.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-02-09', '2023-11-16', '2023-11-17' et le '2023-11-20'.

3.2.5 ATO.PA

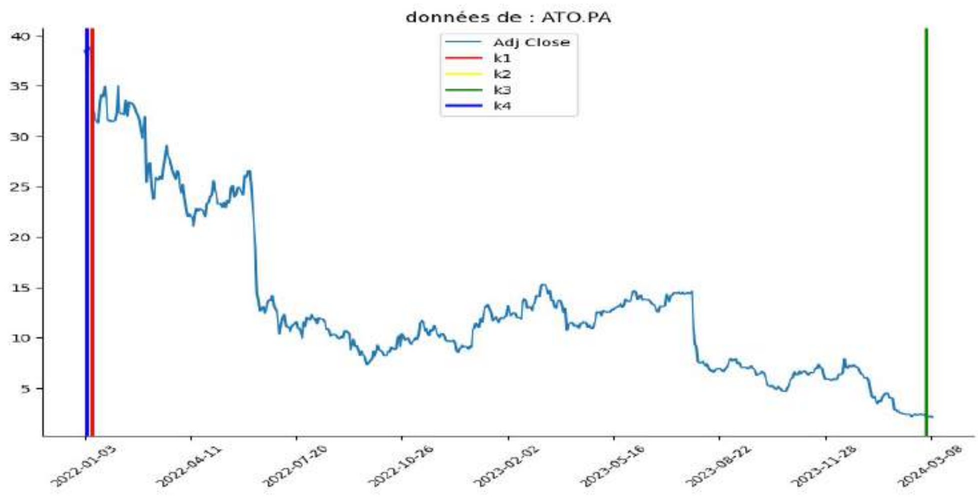

Fig. 3.5 Cours de l'action de ATO.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-04', '2022-01-07'et le '2024-03-05'.

3.2.6 BN.PA

Fig. 3.6 Cours de l'action de BN.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-13', '2022-01-14' et le '2022-02-25'.

3.2.7 BNP.PA

Fig. 3.7 Cours de l'action de BNP.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-02-23', '2022-02-24' et le '2023-02-28'.

3.2.8 CA.PA

Fig. 3.8 Cours de l'action de CA.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-06-23', '2023-11-27' et le '2023-11-24'.

3.2.9 CAP.PA

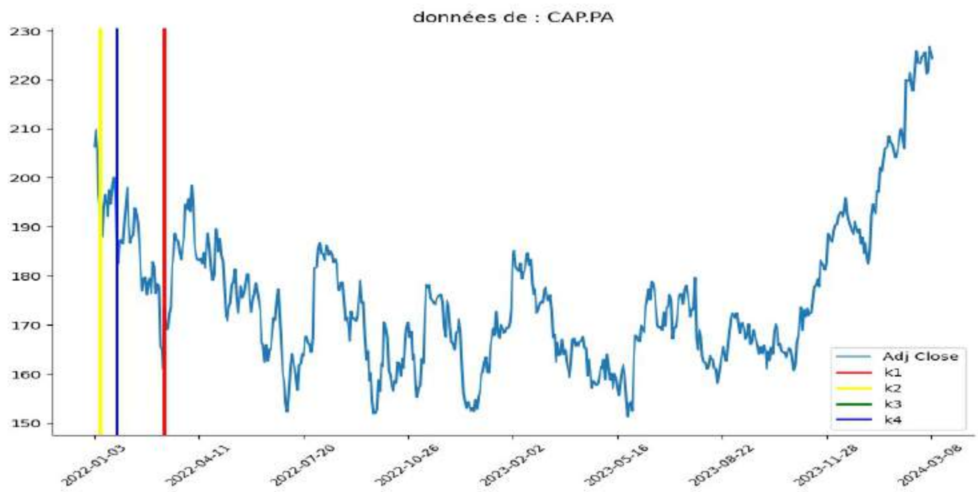

Fig. 3.9 Cours de l'action de CAP.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-07', '2022-01-24' et le '2022-03-09'.

3.2.10 DG.PA

Fig. 3.10 Cours de l'action de DG.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-04-26', '2023-10-18','2023-10-19' et le '2023-10-20'.

3.2.11 EN.PA

Fig. 3.11 Cours de l'action de EN.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-11-16', '2023-07-12','2023-07-13' et le '2023-10-20'.

3.2.12 ENGI.PA

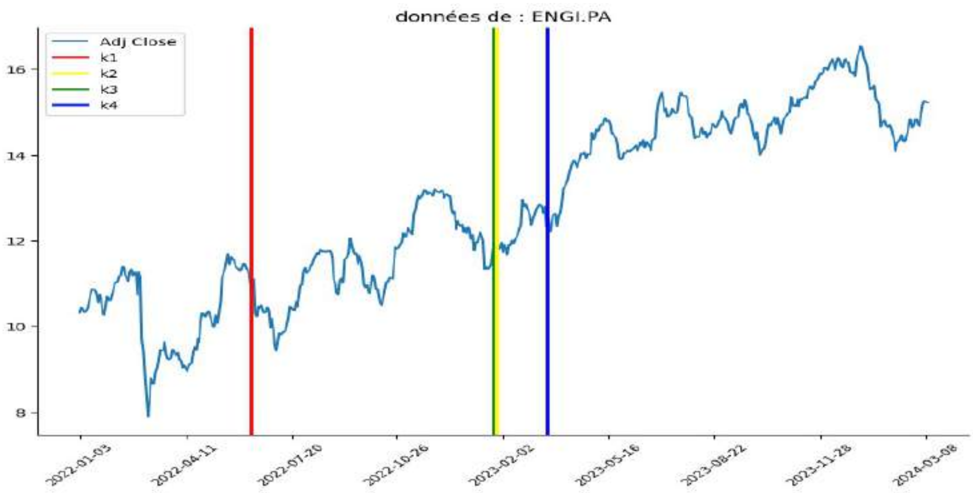

Fig. 3.12 Cours de l'action de ENGI.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-06-13', '2023-01-27','2023-01-34' et le '2023-03-15'.

3.2.13 GLE.PA

Fig. 3.13 Cours de l'action de GLE.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-03-10', '2023-03-04' et le '2022-03-03'.

3.2.14 HO.PA

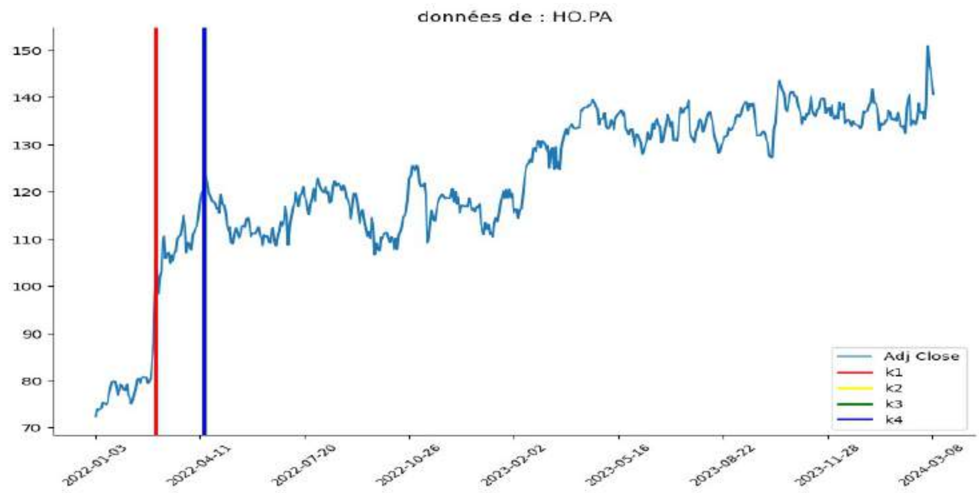

Fig. 3.14 Cours de l'action de HO.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-02-28', '2022-04-14' et le '2022-04-13'.

3.2.15 KER.PA

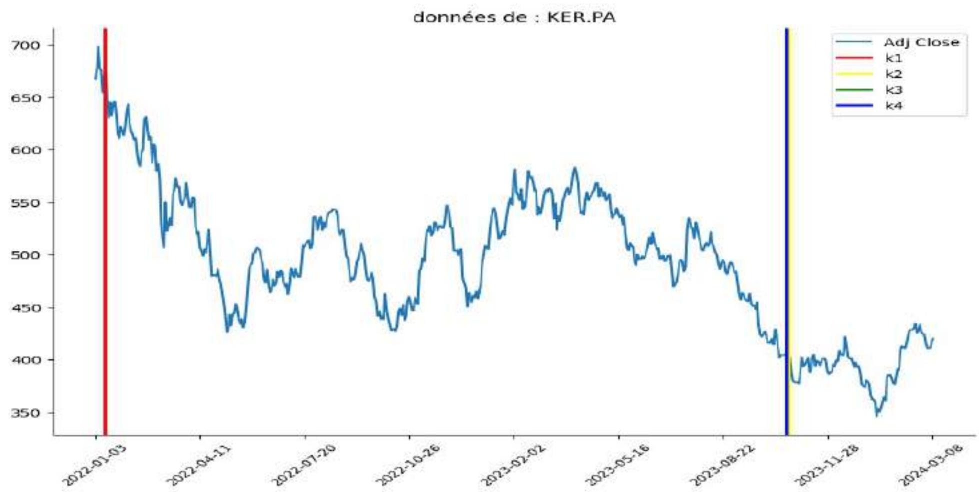

Fig. 3.15 Cours de l'action de KER.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-11', '2023-10-23' et le '2023-10-19'.

3.2.16 LR.PA

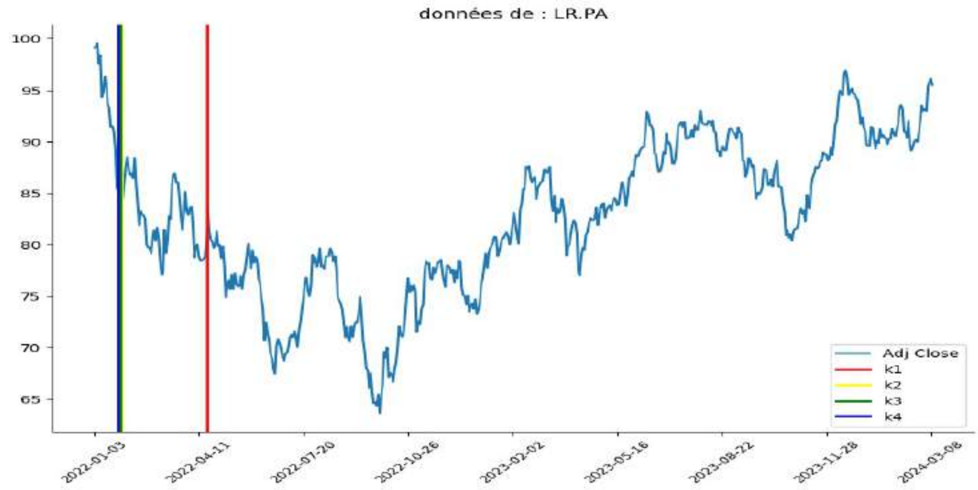

Fig. 3.16 Cours de l'action de LR.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-04-20', '2022-01-07', '2022-01-26' et le '2022-01-25'.

3.2.17 MC.PA

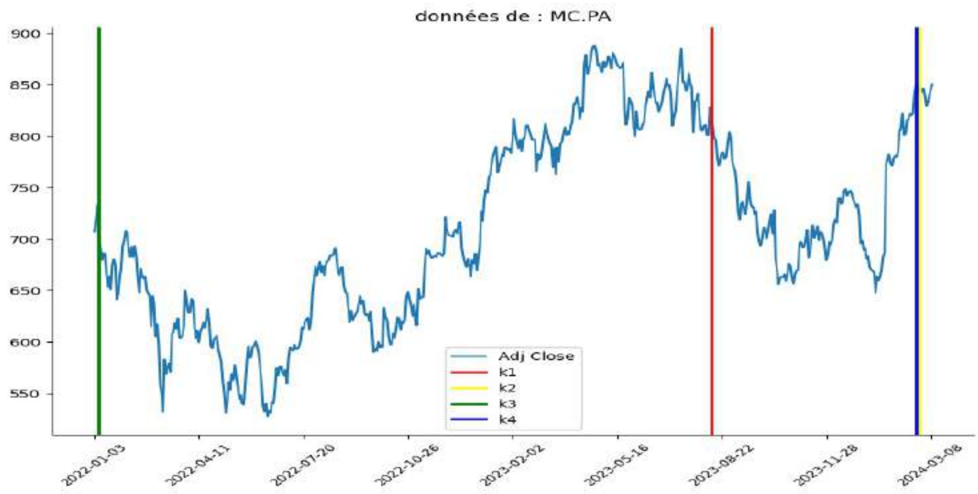

Fig. 3.17 Cours de l'action de MC.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-08-11', '2024-02-27', '2022-01-06' et le '2024-02-23'.

3.2.18 ML.PA

Fig. 3.18 Cours de l'action de ML.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-04-05', '2022-02-21'et le '2022-02-22'.

3.2.19 OR.PA

Fig. 3.19 Cours de l'action de OR.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-13', '2022-01-06'et le '2022-01-05'.

3.2.20 ORA.PA

Fig. 3.20 Cours de l'action de ORA.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-06-29', '2022-05-03'et le '2022-05-02'.

3.2.21 RI.PA

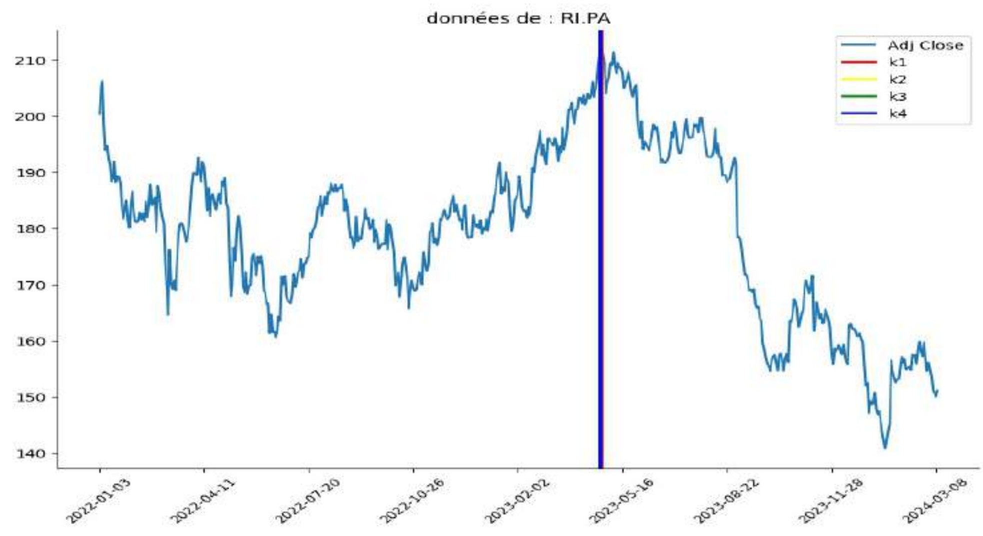

Fig. 3.21 Cours de l'action de RI.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-04-25' et le '2023-04-24'.

3.2.22 SAN.PA

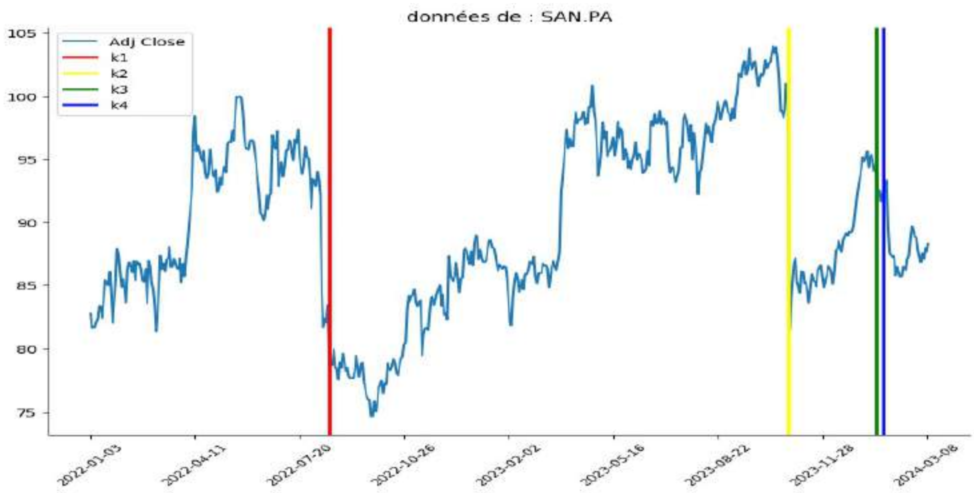

Fig. 3.22 Cours de l'action de SAN.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-08-17', '2023-10-26', '2024-01-22' et le '2024-01-29'.

3.2.23 SGO.PA

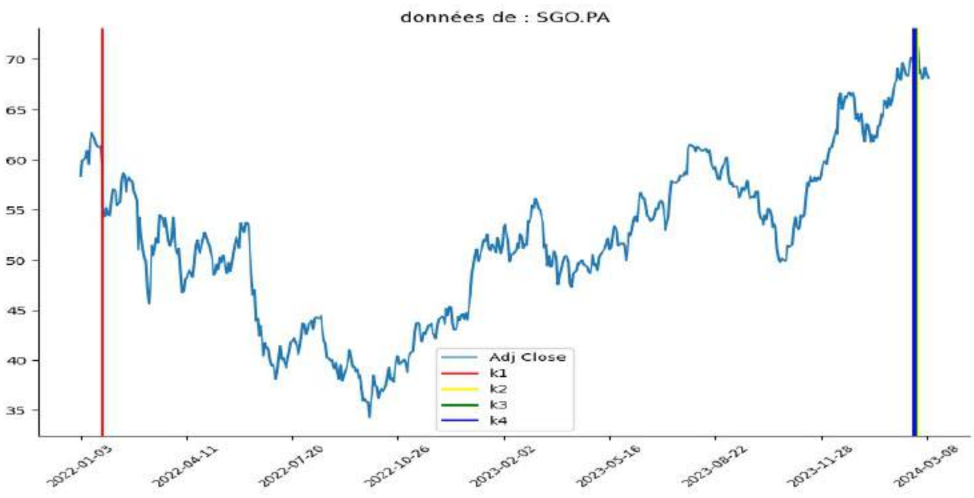

Fig. 3.23 Cours de l'action de SGO.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-21', '2024-02-28', '2024-02-27' et le '2024-02-26'.

3.2.24 SU.PA

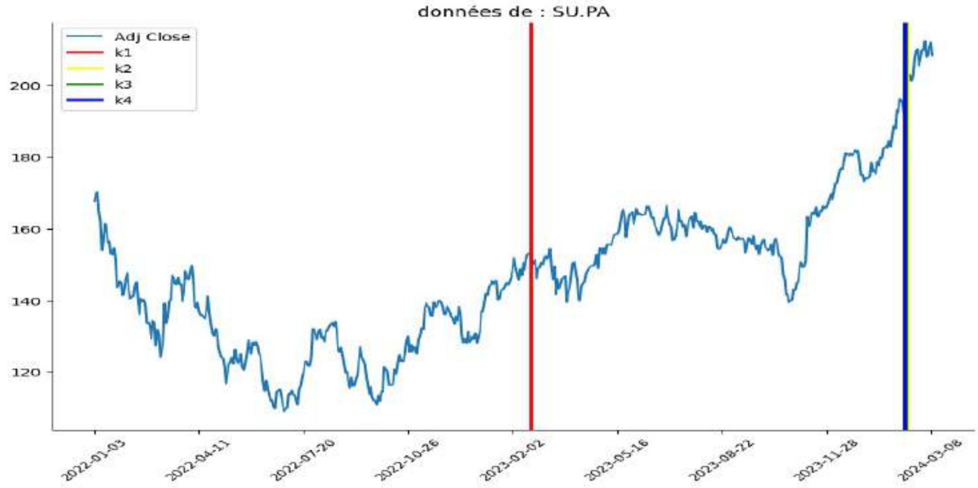

Fig. 3.24 Cours de l'action de SU.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-02-20', '2024-02-15', '2024-02-14' et le '2024-02-13'.

3.2.25 SW.PA

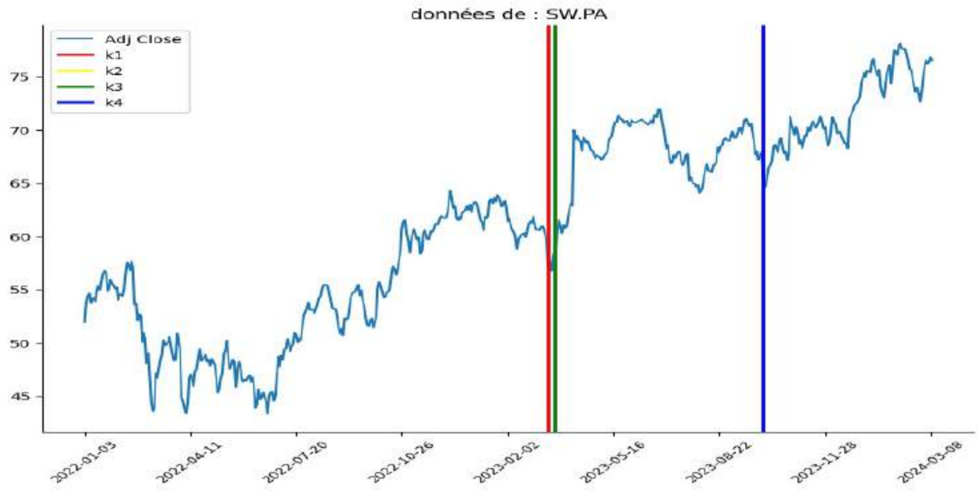

Fig. 3.25 Cours de l'action de SW.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-02-13', '2023-03-17' et le '2023-10-02'.

3.2.26 TTE.PA

Fig. 3.26 Cours de l'action de TTE.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-01-28', '2022-06-07' et le '2023-07-06'.

3.2.27 VIE.PA

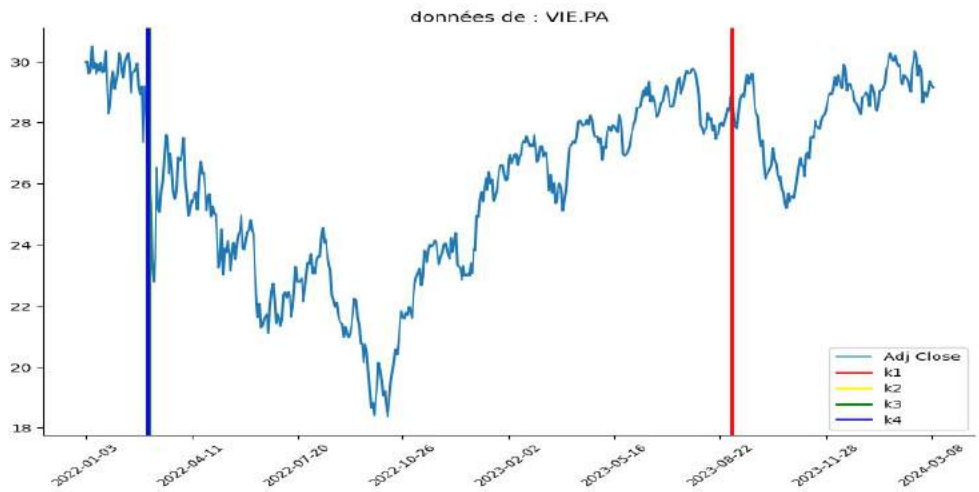

Fig. 3.27 Cours de l'action de VIE.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2023-09-01', '2022-03-01' et le '2022-03-02'.

3.2.28 VIV.PA

Fig. 3.28 Cours de l'action de VIV.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2022-04-04', '2022-06-30', '2022-06-29' et le '2022-06-08'.

3.2.29 WLN.PA

Fig. 3.29 Cours de l'action de WLN.PA du 01/01/2021 au 11/03/2024

Les dates des points de rupture possibles trouvés sont le '2024-03-06', '2023-10-25', '2022-01-05' et le '2022-01-04'.

3.2.30 Résumé du test de Kolmogorov-Smirnov

- pour m = 1, tous les fichiers données contiennent un point de rupture.
- pour m = 2, tous les fichiers données contiennent un point de rupture sauf BN.PA.csv.
- pour m = 3, tous les fichiers données contiennent un point de rupture sauf BN.PA.csv et MC.PA.csv.
- pour m = 4, tous les fichiers données contiennent un point de rupture sauf OR.PA.csv, AC.PA.csv, et BN.PA.csv.

Chapter 4

Conclusion and Perspectives

L'objectif principal était d'identifier les changements structurels significatifs dans les séries temporelles des actions, afin de fournir des informations pertinentes pour les investisseurs et les analystes financiers.

Notre étude a déployé un modèle de détection de ruptures basé sur l'estimation AR(1) pour identifier les points de rupture où des changements importants dans le comportement des actions ont été observés. Ce modèle nous a permis d'estimer des paramètres clés tels que a1, a2, sigma2, et k, ce dernier représentant le moment précis du changement. La méthode mise en œuvre nous a permis de procéder à une analyse itérative pour identifier tous les points de rupture significatifs sur nos données. En cas de multiples ruptures, nous avons appliqué une méthode de division répétée des échantillons, cherchant à isoler chaque segment où les changements se manifestent.

Pour valider nos résultats, nous avons également réalisé des tests de Kolmogorov-Smirnov pour confirmer l'homogénéité des distributions avant et après les points de rupture détectés. Toutes les ruptures identifiées ont passé avec succès ce test d'homogénéité, renforçant ainsi la signification de nos résultats. Les graphiques générés illustrent de manière visuelle les séries temporelles des actions du CAC40, mettant en évidence les points de rupture détectés grâce à notre approche AR(1).

En conclusion, notre étude a utilisé une approche analytique basée sur le modèle autoregressif AR(1) pour détecter les moments critiques de changement dans les actions du CAC40. Les résultats obtenus ont fourni des informations précieuses sur les points de rupture. pour orienter des décisions d'investissement plus éclairées et pour approfondir la compréhension des dynamiques du marché boursier.

References

[1] (2016). Statistique : Tests d'hypothèses. https://www.math.u-bordeaux.fr/~mchabano/Agreg/ProbaAgreg1213-COURS3-Stat2.pdf. Accessed: 2024-04-16.

[2] (2024). Kolmogorov–smirnov test. https://en.wikipedia.org/wiki/Kolmogorov\T1\textendashSmirnov_test#Twosample_Kolmogorov\T1\textendashSmirnov_test. Accessed: 2024-04-9.

[3] Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.

[4] Knuth, D. E. (1984). *The TEXbook*. Addison-Wesley.

Appendix A

Code

A.1 Modele de detection de ruptures

```python
import os
import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pylab as plt

### Importation des donnes du CAC40 ###

Liste_donnees=os.listdir('.\.\DonCAC40\\')
m_values = [1, 2, 3, 4]
result_dict = {}

for m in m_values:
    Frame_Dict = {}
    Frame_Dict['actions'] = [t[:-4] for t in Liste_donnees]

    estimator = ['a1', 'a2', 'sigma2', 'k']
    for x in estimator:
        Frame_Dict[x] = []

    for i in range(len(Liste_donnees)):
        F = '.\.\DonCAC40\\' + Liste_donnees[i]
        data = pd.read_csv(F, delimiter=',')
        # Remove missing data
        data = data[data[data.columns[5]].notnull()]
        X = data[data.columns[5]]
        Y = simulate_Y_m(X, m)
```

```

        D = AR1_Estimation(Y)
        for cle, valeur in D.items():
            Frame_Dict[cle].append(valeur)

    Resultats_S = pd.DataFrame.from_dict(Frame_Dict)
    Resultats_S = Resultats_S.set_index('actions')

    result_dict[m] = Resultats_S

### Code Methode analytique ###
def simulate_Y_m(x, m):
    n = len(x)
    Y = []
    for t in range(m, n):
        Y_m = np.log(x[t] / x[t-m])
        Y.append(Y_m)
    return Y

def AR1_Estimation(data):
    import numpy as np

    n = len(data)

    # Define the objective function
    def obj_function(k):
        a1 = sum([data[i-1]*data[i] for i in range(1, int(k)+1)])/sum
    ([data[i-1] for i in range(1, int(k)+1)])
        a2 = sum([data[i-1]*data[i] for i in range(int(k)+1, n)])/sum
    ([data[i-1] for i in range(int(k)+1, n)])
        sigma2 = (1/(2*n)) * (sum([(data[i]- a1*data[i-1])**2 for i
    in range(1, int(k)+1)])+ sum([(data[i]- a2*data[i-1])**2 for i in
    range(int(k)+1, n)]))
        return -n * np.log(sigma2)

    # Calculate the likelihood for each k
    mle = [obj_function(i) for i in range(1, n-1)]

    # Find the index of the maximum likelihood
    k = np.argmax(mle)  +1

    # Calculate a1 and a2 using optimized k
    a1 = sum([data[i-1]*data[i] for i in range(1, k+1)]) / sum([data[
    i-1] for i in range(1, k+1)])
```

```
    a2 = sum([data[i-1]*data[i] for i in range(k+1, n)]) / sum([data[i-1] for i in range(k+1, n)])

    # Calculate sigma2 using optimized k
    sigma2 = (1/(2*n)) * (sum([(data[i]- a1*data[i-1])**2 for i in range(1, k+1)]) + sum([(data[i]- a2*data[i-1])**2 for i in range(k+1, n)]))

    # Store results in a dictionary
    Resultats = {
        'a1': a1,
        'a2': a2,
        'sigma2': sigma2,
        'k': k
    }

### Test de Kolmogorov-Smirnov ###
def eCDF(sample, t, sort = False):
    # sorts the sample, if unsorted
    if sort:
        sample.sort()
    # counts how many observations are below x
    res = sum(sample <= t)
    # divides by the total number of observations
    res = res / len(sample)
    return res
def decision(p_value, alpha):
        if (p_value <= alpha):
            return "reject null hypothesis" #X1...Xk and Xk...XN do not follow the same distribution (breakpoint exists)
        else:
            return "accept null hypothesis" #otherwise
def ks_2sample_test(sample1, sample2):
    # gets all observations
    observations = np.concatenate((sample1, sample2))
    observations.sort()
    # sorts the samples
    sample1.sort()
    sample2.sort()
    # find K-S statistic
    list_D = []
    for x in observations:
        cdf_sample1 = eCDF(sample = sample1, t = x)
        cdf_sample2 = eCDF(sample = sample2, t = x)
```

```
        list_D.append(abs(cdf_sample1 - cdf_sample2))
    ks_stat = max(list_D)
    # calculates the p-value based on the two-sided test
    # the p-Value comes from the KS Distribution Survival Function (
   SF = 1-CDF)
    n1, n2 = float(len(sample1)), float(len(sample2))
    en = n1 * n2 / (n1 + n2)
    p_value = stats.kstwo.sf(ks_stat, np.round(en))
    dec = decision(p_value, 0.05) # alpha = 0.05

    return {"ks_stat": ks_stat, "p_value" : p_value,"decision": dec}

result_test = {}
for m in [1,2,3,4]:
    Frame_Dict = {}
    Frame_Dict['actions'] = [t[:-4] for t in Liste_donnees]
    result = ['ks_stat', 'p_value','decision']
    for x in result:
        Frame_Dict[x] = []

    for i in range(len(Liste_donnees)):
        F = '.\.\DonCAC40\\' + Liste_donnees[i]
        data = pd.read_csv(F, delimiter=',')
        # Remove missing data
        data = data[data[data.columns[5]].notnull()]
        X = data[data.columns[5]]
        k = result_dict[m]['k'][i]
        data1 = np.asarray(X[:k])
        data2 = np.asarray(X[k:])
        D = ks_2sample_test(data1, data2)
        for cle, valeur in D.items():
            Frame_Dict[cle].append(valeur)

    Resultats_test = pd.DataFrame.from_dict(Frame_Dict)
    Resultats_test = Resultats_test.set_index('actions')

    result_test[m] = Resultats_test

###  pour tracer les graphes des donnees ###
for i in range(len(Liste_donnees)):
    F='.\.\DonCAC40\\'+Liste_donnees[i]
    data = pd.read_csv(F, delimiter=',')
    data=data[data[data.columns[5]].notnull()]  # supprimer les
   donnÃľes manquantes
```

```

    ### colonnes 6  : AdjClose
    ### print(data.columns[5])
    k1 = result_dict[1]['k'][i]
    k2 = result_dict[2]['k'][i]
    k3 = result_dict[3]['k'][i]
    k4 = result_dict[4]['k'][i]

    fig = plt.figure(figsize=(10, 7))
    ax = fig.add_subplot(111)
    # cacher axes  'haut' , 'droit' et 'gauche':
    for side in ['right','top']:
        ax.spines[side].set_visible(False)

    data[data.columns[5]].plot()

    ### on affiche les dates modulo 60
    indices=[i for i in range(len(data[data.columns[0]])) if i%70==0]
    XX_ticks=[data[data.columns[0]][i] for i in indices]
    plt.xticks(indices,XX_ticks, rotation=45, fontsize=9)
    #####
    plt.axvline(k1,color='red',label='k1',lw=2)
    plt.axvline(k2,color='yellow',label='k2',lw=2)
    plt.axvline(k3,color='green',label='k3',lw=2)
    plt.axvline(k4,color='blue',label='k4',lw=2)

    plt.title("donnÃľes de : %s"%Liste_donnees[i][:-4])

    plt.legend()
    plt.show()
```