

Méthode hybride pour la détection des ruptures structurelles dans les séries financières : Cas de l'indice MASI

Hybrid method for detecting structural breaks in financial series: The case of the MASI index

OULDKHAL Asma

Enseignant chercheur
Faculté d'Economie et de Gestion
Université Hassan 1er - MAROC
Laboratoire de Modélisation Mathématique et de Calcul Economique

ASSOUDOU Souad

Enseignant chercheur
Faculté d'Economie et de Gestion
Université Hassan 1er - MAROC
Laboratoire de Modélisation Mathématique et de Calcul Economique

Date de soumission : 07/08/2026

Date d'acceptation : 28/09/2026

Pour citer cet article :

OULDKHAL. A. & ASSOUDOU. S. (2026) « Méthode hybride pour la détection des ruptures structurelles dans les séries financières : Cas de l'indice MASI », Revue Française d'Économie et de Gestion « Volume 7 : Numéro 10 » pp : 1131- 1157.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons
Attribution License 4.0 International License



Résumé

Cet article propose une approche hybride pour la détection des ruptures structurelles dans les séries financières. La procédure combine une méthode heuristique d'exploration des ruptures, avec une validation économétrique par régression ARIMAX, à variables exogènes. La stratégie d'exploration repose sur trois piliers : (i) un lissage par médiane glissante, (ii) un calcul de dérivée de niveau par différence de médianes sur fenêtres adjacentes, et (iii) un filtrage des extrema par regroupement spatial et clustering K-means visant à éliminer les faux positifs.

L'application de notre méthode à l'indice MASI a permis de détecter avec précision les dates ruptures de la période Janvier 2019- Juin 2026, coïncidant avec la crise du COVID-19, le début de la guerre en Ukraine, le conflit avec l'Iran et d'autres événements ayant pu impacter la bourse de Casablanca. L'intégration de ces ruptures dans un modèle ARIMA a permis de réduire la persistance du modèle ARIMA et d'améliorer les performances de prévision hors échantillon. En effet les mesures (RMSE, MAE et MAPE) ont été réduits de plus de 14%.

Mots-clés : Rupture structurelle, indice MASI, lissage par médiane glissante, clustering K-means, ARIMAX.

Abstract

This article proposes a hybrid approach for detecting structural breaks in financial time series. The procedure combines a heuristic method for break detection with econometric validation using ARIMAX regression (incorporating exogenous variables). The detection strategy rests on three pillars: (i) smoothing via a rolling median, (ii) calculating the derivative of the level by differencing medians across adjacent windows, and (iii) filtering extrema through spatial grouping and K-means clustering to eliminate false positives.

Applying our method to the MASI index enabled the accurate detection of break dates between January 2019 and June 2026, coinciding with the COVID-19 crisis, the onset of the war in Ukraine, the conflict with Iran, and other events that may have impacted the Casablanca Stock Exchange. Incorporating these breaks into an ARIMA model reduced model persistence and improved out-of-sample forecasting performance; specifically, error metrics (RMSE, MAE, and MAPE) were reduced by more than 14%.

Keywords: Structural breaks, MASI Index, Moving median smoothing, clustering K-means, ARIMAX.

Introduction

L'analyse des séries temporelles est confrontée à un défi récurrent : la non-stationnarité. Depuis les travaux de Nelson et Plosser (1982) ([24]) et Perron (1989) ([26]), la distinction entre non-stationnarité stochastique (racine unitaire) et déterministe (ruptures structurelles) est fondamentale. Les ruptures de niveau modifient la moyenne de la série de façon permanente et peuvent biaiser les tests économétriques et les prévisions si elles ne sont pas prises en considération. Les ruptures de variance compromettent considérablement la fiabilité des prévisions des risques.

Dans un modèle ARIMA, une rupture de niveau non modélisée entraîne une interprétation de l'estimateur, en termes de série à mémoire très longue, une racine autorégressive proche de l'unité et des prévisions qui donne une moyenne, qui ne correspond plus au niveau actuel.

L'objectif de notre travail est de détecter les ruptures de niveau dans la série brute des prix de l'indice MASI de la bourse de Casablanca, durant la période [Janvier 2019- Juin 2026].

En nous tournant vers les méthodes de détection des ruptures structurelles qu'elles soient paramétriques (Chow, Bai-Perron, CUSUM) ou non paramétriques (basées sur le noyau ou sur le rang), nous constatons qu'elles présentent des limites pour l'analyse des séries brutes non stationnaires, avec racine unitaire, à l'exemple de notre série du MASI. Les premières méthodes supposent la stationnarité des erreurs et les secondes sont sensibles au choix du noyau, supposent l'indépendance des observations ou ne détectent pas la nature du changement. De plus, des restrictions sévères sont imposées au nombre maximum de ruptures à tester, pour des contraintes statistiques et computationnelles. Ces méthodes s'appliquent généralement aux Log-rendements à cause de leur stationnarité. Cependant, les ruptures sur les prix brutes et sur les rendements ne coïncident pas nécessairement. Une rupture de niveau sur les prix peut ne pas entraîner de rupture dans la moyenne ou la volatilité des rendements.

Partant de ce constat, nous proposons dans cet article une nouvelle approche qui repose dans un premier temps sur une technique exploratoire des ruptures structurelles, utilisée en ingénierie et en traitement de signal et de données, qui combine lissage par médiane glissante, détection d'extrema, regroupement et clustering K-means ([15],[18],[19],[30],[31],[33]). Les dates de ruptures proposées par cette technique ne sont retenues qu'après validation par un modèle ARIMAX à step, préalablement spécifié, qui fournit un cadre statistique rigoureux de validation : tests de significativité, critères d'information, blancheur des résidus et validation hors échantillon.

La méthode a permis d'identifier avec précision les dates de ruptures, associées aux principales crises de la période d'analyse, à savoir le choc lié à la pandémie du COVID-19, la guerre d'Ukraine en 2021, le conflit avec l'Iran et d'autres événements ayant pu impacter la bourse de Casablanca et l'indice MASI durant cette période, sans que cela ne permette d'établir un lien de causalité direct.

Enfin, la comparaison des performances prédictives du modèle ARIMA, selon qu'il intègre ou non les points de ruptures démontre l'importance cruciale de leur détection. En effet, la prise en compte de ces ruptures engendre une réduction très significative des métriques d'erreur RMSE, MAE et MAPE.

L'article est organisé comme suit :

1. La section 1 présente une revue de la littérature
2. La section 2 détaille la méthodologie
3. La section 3 présente les données et les résultats
4. La section 4 discute des résultats et des implications constatées.

Tous les calculs et graphiques sont réalisés sur le logiciel R.

1. Revue de La littérature

L'histoire de l'étude des ruptures structurelles a commencé par l'étude de la stabilité des paramètres dans les modèles de séries temporelles, en particulier dans une régression linéaire. Cette stabilité garantit la bonne prévision du futur.

1.1. Du test de Chow à Bai-Perron

Le test de Gregory Chow en 1960 ([8]) vérifie si les coefficients d'une régression linéaire restent constants sur deux sous-périodes. Une seule date de rupture est autorisée et elle est connue. Le test de Quandt du rapport de vraisemblance, la même année, autorise une date de rupture inconnue. L'hypothèse sous-jacente était que les séries sont stationnaires.

Avec l'essor de la théorie des racines unitaires dans les années 1980, P. Perron dans son article fondateur de 1989 ([26]) démontre que la présence d'une rupture structurelle dans une série temporelle biaise les tests de racines unitaires et cette série est prise à tort pour une marche aléatoire.

Plus tard, l'article de Bai-Perron de 1998 ([3]) offre un cadre unifié pour estimer des ruptures multiples, à dates inconnues et pouvant être partielles (la rupture ne concerne pas tous les paramètres) dans les modèles de régression linéaire. Il se base sur le test de Chow mais calcule une version généralisée du test pour chaque combinaison possible de dates de ruptures puis choisit celle qui minimise la somme des carrés des résidus. Le nombre maximum de

ruptures devrait être fixé à l'avance. Des contraintes statistiques et computationnelles n'autorisent pas un nombre de ruptures m trop grand (en général m ne dépasse pas 10).

Le cadre de Bai-Perron repose sur l'hypothèse que la série étudiée est stationnaire en dehors des ruptures testées et sur l'hypothèse que le terme d'erreur du modèle suit un processus de bruit blanc, ce qui signifie que les erreurs du modèle doivent être au moins non corrélées et de variance finie. Bai-Perron (1998) convient à des modèles à erreurs stationnaires, y compris AR.

La méthode de Bai-Perron ne tolère ni les séries à queues épaisses ni le processus Garch et ne tolère pas non plus les séries à racine unitaire ou à tendance stochastique. Les tests de ruptures présentent alors un risque élevé de sur-détection. La marche aléatoire dérive de façon stochastique et chaque dérive est interprété par l'algorithme comme une succession de saut de moyenne. Ce qui génère un excès de ruptures fallacieuses ([27]).

Une marche aléatoire peut très bien subir, elle aussi, des ruptures structurelles, de niveau, de tendance ou de variance que l'algorithme Bai-Perron peine à détecter. Bai-Perron dans son article de l'année 2003 ([4]) souligne que la puissance du test est compromise en cas de forte persistance et recommande des « trimmings » (seuil minimal de taille de segments) plus élevés, au prix de ne plus pouvoir détecter des ruptures rapprochées. L'article de P. Prodon en 2008 ([27]) montre ces difficultés de sur-détection, sous-détection ou sensibilité au « trimming » qui peuvent fausser l'inférence et conduire à des conclusions erronées.

Dans le cas des indices financiers, Bai-Perron a été principalement appliqué aux séries des Log-rendements qui sont en général stationnaires. Or les ruptures sur les prix et sur les rendements ne coïncident pas nécessairement. Une rupture de niveau sur le prix peut ne pas affecter la moyenne des rendements et une rupture de variance des rendements peut être invisible sur les prix.

1.2. Le test CUSUM

Le test CUSUM (Cumulative Sum) est un test de stabilité structurelle basée sur une approche séquentielle d'analyse des résidus récursifs d'une régression estimée par moindres carrés récursifs. Sous l'hypothèse nulle de stabilité des paramètres, les résidus récursifs sont non corrélés, de moyenne nulle et de variance constante. Le cumul de ces résidus doit fluctuer autour de zéro dans un intervalle de confiance. Un écart aux bornes indique un changement des paramètres du modèle de régression.

Dans son article de 2013, Aue, A ([2]) propose une adaptation du test en cas de corrélation sérielle des résidus du modèle et traite de la distinction entre rupture structurelle et racine unitaire.

Le test CUSUM, dans sa version originale ([22]), ne fournit pas d'estimateur fiable de la date de rupture et il est conçu pour une rupture unique, donc malgré sa robustesse distributionnelle et sa simplicité, il est utilisé, dans la pratique, comme un outil exploratoire précédant le test de Bai-Perron.

Une autre faiblesse du test CUSUM standard est qu'il est valide uniquement pour les séries stationnaires à variance constante, mais des extensions très récentes l'adaptent aux séries avec racine unitaire. Dans son article de 2026 ([35]), Zhong, C. traite des séries temporelles non stationnaires, générées par un modèle ARMA à innovations dépendantes et hétéroscédastiques, en proposant deux statistiques CUSUM et « moving-sum » et en développant une théorie asymptotique pour l'estimation de la date de rupture, sous différentes amplitudes de saut. La convergence vers la vraie date de rupture est assurée mais la précision dépend de l'amplitude du saut et de la taille de l'échantillon. Le cadre de l'étude de Zhong sont les séries dont la non stationnarité est due à une tendance déterministe et non à la marche aléatoire.

1.3. Les méthodes à noyau

Les méthodes à noyau reposent sur l'idée d'estimer localement les paramètres en pondérant les observations par une fonction de noyau ([1]). La méthode permet une estimation de paramètres variant dans le temps dans une régression linéaire (Kernel Regression), apporte une correction à l'inférence quand les erreurs sont autocorrélés ou hétéroscédastiques (Kernel-based HAC Covariance Matrix Estimation), capture des dynamiques complexes comme des ruptures brutales, des variations lisses et des changements dans un ou plusieurs paramètres du modèle (Kernel-Based time-varying parameter estimation).

L'algorithme KCP (Kernel Change Point) ([16]) permet de dater les ruptures en retenant celles qui maximisent l'hétérogénéité entre segments. Le critère à minimiser est estimé en se basant sur la technique du noyau.

Si les méthodes à noyau sont souvent privilégiées pour leur flexibilité et leur caractère non paramétrique, elles n'en demeurent pas moins sujettes à plusieurs critiques, notamment leur dépendance au choix de la bande passante et l'impossibilité d'identifier quel paramètre de la distribution a été affecté par la rupture.

Les méthodes à noyau reposent sur l'hypothèse d'indépendance (ou m-dépendance) des observations pour garantir la consistance de l'estimateur. Or une marche aléatoire viole cette hypothèse et rend l'application directe à une série brute des prix financiers non justifiée.

1.4. La méthode LSTM

L'apprentissage profond par réseaux de neurones LSTM (Long Short - Term Memory) ([28], [36]) exploite une mémoire à long terme de la série pour détecter la signature statistique qui précède un changement de régime ou une crise. Le LSTM est d'abord un programme de prévision de ruptures ou de détection en ligne. En temps réel, il compare la prédiction du modèle avec la valeur observée. Cette comparaison est effectuée à l'aide d'un test statistique (comme le CUSUM) qui décide si l'écart est significatif.

Classée parmi les méthodes les plus modernes, cela n'empêche que des hypothèses telles que la stationnarité par morceaux, la dépendance temporelle apprenable à partir des données rend problématique son application à la marche aléatoire pure. Elle s'applique en général aux rendements.

2. Méthodologie

2.1. Principe de la méthode heuristique d'exploration des ruptures

2.1.1. Définition d'une rupture structurelle

Une rupture structurelle est un changement brutal et durable qui modifie une ou plusieurs propriétés fondamentales d'une série temporelle, à savoir son niveau (moyenne), sa pente (tendance) ou sa variance (volatilité).

2.1.2. Hypothèses

Les séries financières présentent deux caractéristiques empiriques. D'une part, les changements de régime (Krachs, reprise) représentent des événements "rares". Sur plusieurs années, seules quelques dates correspondent à des ruptures significatives. D'autre part, ces ruptures, lorsqu'elles surviennent, se traduisent par des mouvements brutaux et durables du niveau ou de la pente, générant des écarts importants entre la période avant et la période après. Ces deux constats permettent de formuler les hypothèses :

Hypothèse d'homogénéité des ruptures :

Les véritables ruptures produisent des valeurs extrêmes du même ordre de grandeur. En effet, un changement durable de niveau génère un écart systématique entre les médianes de la période avant et la période après rupture. Il se forme alors un cluster naturel de valeurs extrêmes des écarts entre les médianes, distinctes des fluctuations aléatoires de faible amplitude.

Hypothèse de rareté des ruptures : Les ruptures sont des événements rares, tandis que le bruit produit des minima locaux nombreux et de faible amplitude. Le cluster des ruptures est donc minoritaire en effectif par rapport au cluster du bruit.

Ces deux hypothèses sont essentielles dans le clustering par K-means. Les deux populations (ruptures, bruits) sont séparées automatiquement sans avoir à fixer de seuil qui pourrait être propre à chaque série. Le Clustering K-means avec $k=2$ devrait être confirmé empiriquement.

2.2. Principe de la sélection des dates candidates

L'approche se déploie en quatre étapes successives :

Étape 1 : Lissage du signal par médiane glissante : la médiane est calculée sur chaque fenêtre de longueur $w_s=3$ ou 5. La médiane est préférée à la moyenne pour sa robustesse face aux valeurs extrêmes et sa moindre sensibilité à l'asymétrie des distributions. En calcul de dérivée (différence de médiane), on compare deux fenêtres consécutives. Une valeur aberrante ou outlier est ignorée par la médiane mais en calcul de moyenne, elle réduirait l'écart avec la fenêtre voisine et ferait manquer la vraie rupture.

Étape 2 : Calcul de la dérivée par comparaison des médianes de fenêtres consécutifs de taille w_c . Une valeur fortement négative indique une chute de niveau et une valeur fortement positive indique une hausse de niveau.

Étape 3 : Regroupement par blocs de 20 jours, pour chaque couple de valeurs (w_s , w_c) : dans chaque bloc, on ne garde que l'extremum de forte amplitude. Cela évite de détecter plusieurs fois la même rupture à cause des fluctuations locales du signal.

Étape 4: Clustering par K-means:

Le filtrage des extrema locaux par clustering K-means ($K=2$) sépare les vraies ruptures avec une forte amplitude, du bruit qui a une faible amplitude. Cette séparation est effectuée en maximisant la distance entre le centre de gravité (moyenne arithmétique) du premier groupe et celui du deuxième. Ce clustering est effectué pour chaque couple de valeurs (w_s , w_c).

Les notations utilisées sont :

- w_s (smoothing_width) : taille de la fenêtre de lissage : 3 ou 5.
- w_c (cost_width) : taille des fenêtres pour le calcul de la dérivée (3,5,7,10,15,20).
- w_{se} (search_width) : taille de la fenêtre de regroupement des extrema locaux (fixée à 20 jours).
- K : nombre de clusters pour K-means ($k = 2$).

2.3. Algorithme

Soit $Y = \{y_1, \dots, y_T\}$ la série observée, où T est le nombre total d'observations.

Pour chaque couple de valeurs (w_s, w_c) : effectuer les étapes suivantes :

Etape 1 : Lissage par médiane glissante :

Pour chaque indice i allant de 1 à $T - w_s + 1$, on calcule :

$$L_i = \text{médiane}(y_i, y_{i+1}, \dots, y_{i+w_s-1}).$$

Etape 2 : Calcul de la fonction dérivée de niveau :

On compare les médianes de deux blocs consécutifs de taille w_c . Pour chaque position k allant de w_c à longueur(L) $- w_c$, on calcule l'écart entre les médianes de deux fenêtres adjacentes en k , de taille w_c :

$$D_{\text{niveau}}(k) = \text{mediane}(L_{k+1}, \dots, L_{k+w_c}) - \text{mediane}(L_{k-w_c+1}, \dots, L_k).$$

Etape 3 : Extraction des extrema locaux et regroupement :

On identifie les indices k où $D(k)$ est un minimum (resp. Maximum) local :

$$D(k) < D(k-1) \quad \text{et} \quad D(k) < D(k+1). \quad (\text{resp. } D(k) > D(k-1) \quad \text{et} \quad D(k) > D(k+1)).$$

On découpe ensuite les valeurs de dérivée en blocs de taille w_{se} . Pour chaque bloc, on ne conserve que le minimum (resp. Maximum) le plus bas (resp. Haut). On obtient ainsi une liste de N candidats $\mathcal{C} = \{c_1, \dots, c_N\}$.

Etape 4 : Clustering K-means pour isoler les vraies ruptures :

1. On extrait les valeurs de la dérivée aux positions des candidats : $\mathcal{X} = \{D(c_1), D(c_2), \dots, D(c_N)\}$.

2. On initialise deux centres de clusters, μ_1 et μ_2 .

3. Pour chaque valeur $x \in \mathcal{X}$, on l'affecte au cluster dont le centre est le plus proche : $\text{cluster}(x) = \text{argmin}_{j \in \{1,2\}} (x - \mu_j)^2$.

4. On recalcule chaque centre comme la moyenne des valeurs qui lui sont affectées : $\mu_j^{(\text{nouveau})} = \frac{1}{|\text{cluster}_j|} \sum_{x \in \text{cluster}_j} x$.

5. On répète les étapes 3 et 4 jusqu'à convergence.

Le cluster minoritaire est retenu comme contenant les véritables ruptures. $\text{argmin}_{j \in \{1,2\}}$: désigne la valeur 1 ou 2 selon que $(x - \mu_j)^2$ est la plus faible.

* $|\text{cluster}_j|$: désigne l'effectif du cluster.

2.4. Validation des dates retenues

Les dates retenues par l'algorithme précédent devraient faire l'objet d'un processus de validation.

Dans notre cas, nous avons effectué une modélisation ARIMA de la série, après étude de la stationnarité et détermination du degré de différenciation du modèle ARIMA.

Une date est validée comme une rupture de niveau si le coefficient de la variable « step » associée à cette date est significatif au seuil de 5%, si les critères AIC et BIC diminuent par rapport au modèle ARIMA sans « step » et si les résidus du modèle ARIMAX sont stationnaires et non autocorrélés.

Une validation hors échantillon, avec un modèle ARIMAX intégrant l'ensemble des ruptures retenues devrait confirmer l'amélioration de la performance prédictive par rapport au modèle initial ARIMA.

3. Données et modélisation ARIMA de la série

3.1. Etude de la série de l'indice MASI

Nous appliquons la méthode à l'indice MASI (Moroccan All Shares Index) de la bourse de Casablanca. La période retenue pour l'étude s'étale du 1^{er} Janvier 2019 au 26 Juin 2026. Les données sont journalières :

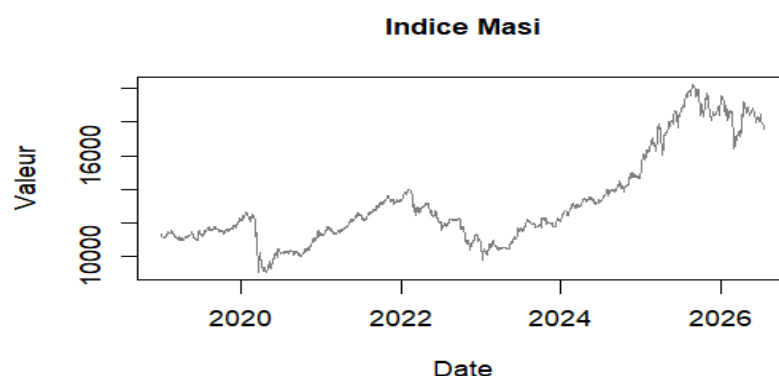
3.2. Statistiques descriptives

Tableau N°1 : Statistiques descriptives

Statistique	Valeur
Nombre d'observations	1823
Moyenne (μ)	12976
Médiane	12147
Maximum	20233
Minimum	8988
Écart-type (σ)	2646
Jarque-Bera (JB)	
P-value	0,0000.. Rejet de la normalité à 1%

Sources : Auteurs -Logiciel R

Figure N°1 : Courbe de l'indice MASI



Source : Auteurs -logiciel R

3.3. Tests de stationnarité

Le test ADF (Augmented Dickey-Fuller) est un test statistique qui permet de détecter la racine unitaire dans une série temporelle. Le rejet de l'hypothèse nulle de présence de racine unitaire conclut à la stationnarité de la série et le non rejet de l'hypothèse nulle stipule l'existence d'une racine unitaire et la nécessité de différencier la série. La statistique ADF pour tous les tests du tableau 2 (avec constante, avec constante et tendance, sans les deux) est supérieure à la valeur critique. L'hypothèse de présence de racine unitaire n'est pas rejetée. La série nécessite une première différenciation. Les tests de stationnarité sont de nouveau appliqués à la série différenciée, y compris le test KPSS, qui a pour hypothèse nulle la stationnarité de la série, contrairement au test ADF.

Les résultats du tableau 3 montre que la série différenciée de l'indice MASI est stationnaire. La série du MASI est intégrée d'ordre 1.

Tableau N°2 : Test ADF sur la série des prix du MASI

Modèle de test ADF	Statistique ADF	Valeur critique (5%)	Conclusion
Avec constante	-0.4018	-2,86	Racine unitaire
Avec constante et tendance	-1,6514	-3,41	Racine unitaire
Aucun	1.0846	-1,95	Racine unitaire

Source : Auteurs – Calcul sur logiciel R

Tableau N°3: Test de stationnarité sur la série différenciée du MASI

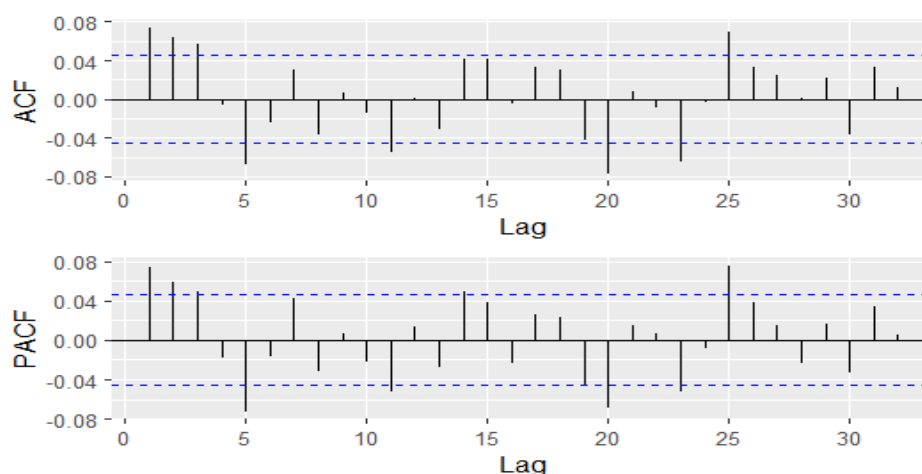
Type de test	Statistique	Valeur critique (5%)	Conclusion
ADF avec constante	-13,84	-2,86	Stationnaire
ADF avec constante et tendance	-13,87	-3,41	Stationnaire
ADF (sans constante, ni tendance)	-13.78	-1.95	Stationnaire
Kpss (constante, tendance, constante et tendance)	—	—	$p - value = 0,1$ Stationnaire

Source : Auteurs – Calcul sur logiciel R

3.4. Modélisation ARIMA

3.4.1. ACF et PACF de la série différenciée - Masi

Figure N°2 : Autocorrélations et autocorrélations partielles - Série différenciée - MASI



Source : Auteurs – logiciel R

3.4.2. Modèle ARIMA sélectionné

La modélisation ARIMA (p,q) (Autoregressive Integrated Moving Average) ([6]) est la méthode statistique de référence pour l'analyse et la prévision d'une série temporelle. La procédure de Box-Jenkins nécessite de stationnariser la série puis d'utiliser les graphiques ACF et PACF pour identifier l'ordre p et q.

Dans le cas de l'indice MASI sur la période étudiée [2019 - 2026], la série est intégrée d'ordre 1 (tableau N°3). Les graphiques ACF et PACF de ΔY_t de notre série (Figure N°2) montrent des corrélations faibles mais significatives aux trois premiers retards pour le ACF et aux deux premiers retards pour le Pacf. Le choix du modèle ARIMA doit résulter d'un compromis entre qualité d'ajustement et parcimonie, évalué à l'aide des deux critères AIC et BIC.

Tableau N° 4 : Modèles ARIMA candidats

P	d	q	AIC	BIC	Ljung-Box-p-value
2	1	2	22576.34	22603.88	0.08132
0	1	3	22579.04	22601.10	0.07369
2	1	0	22582.32	22598.84	0.01594

Source : Auteurs – Calcul sur logiciel R

Note : Les p-values du test de Ljung-Box supérieures à 0.05 indiquent qu'il n'y a pas d'autocorrélation significative des résidus.

Le modèle ARIMA retenu est un processus (2,1,2) (tableau N°4), sélectionné aussi par la procédure automatisée de Box-Jenkins (auto.arima sur R). Ce modèle minimise l'autocorrélation résiduelle, ce qui en fait un modèle solide pour la validation des ruptures. Ce modèle est défini par :

$$\Delta Y_t = AR_1 \Delta Y_{t-1} + AR_2 \Delta Y_{t-2} + \varepsilon_t + MA_1 \varepsilon_{t-1} + MA_2 \varepsilon_{t-2}$$

Où ε_t est un bruit blanc. Les coefficients AR_1 , AR_2 , MA_1 , MA_2 sont donnés dans le tableau N°5 :

Tableau N°5 : Coefficients du modèle ARIMA (2,1,2)

Coefficients	AR(1)	AR(2)	MA(1)	MA(2)
Valeurs	0.8795	-0.6358	-0.8264	0.6751
s.e.	0.1344	0.1378	0.1304	0.1292
p-value	5.894e-11	3.969e-06	2.374e-10	1.719e-07

Source : Auteurs – Calcul sur logiciel R

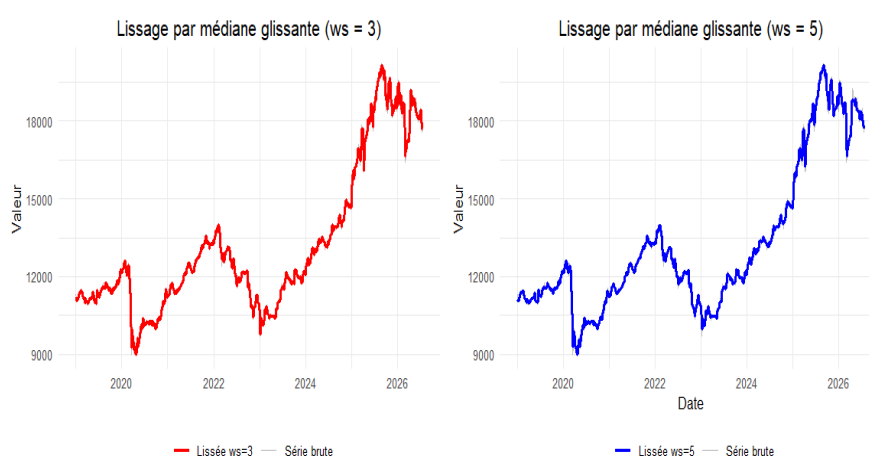
Note : Les p-values inférieures à 0.05 indiquent des coefficients significatifs.

4. Résultats

4.1. Lissage de la série, par médiane glissante

Nous choisissons d'effectuer deux types de lissages : Une fenêtre de 5 jours ($w_s = 5$) est choisi car elle correspond à une semaine de transactions. Elle filtre efficacement le bruit haute fréquence. La fenêtre à trois jours $w_s = 3$ est aussi testée pour capturer les chocs les plus courts. La figure N° 3 donne les deux types de lissages.

Figure N°3 : Lissage $w_s = 3$ et $w_s = 5$



Source : Auteurs – Calcul sur logiciel R

4.2. Validation empirique K-means – K = 2 - par la méthode du coude et le score de silhouette

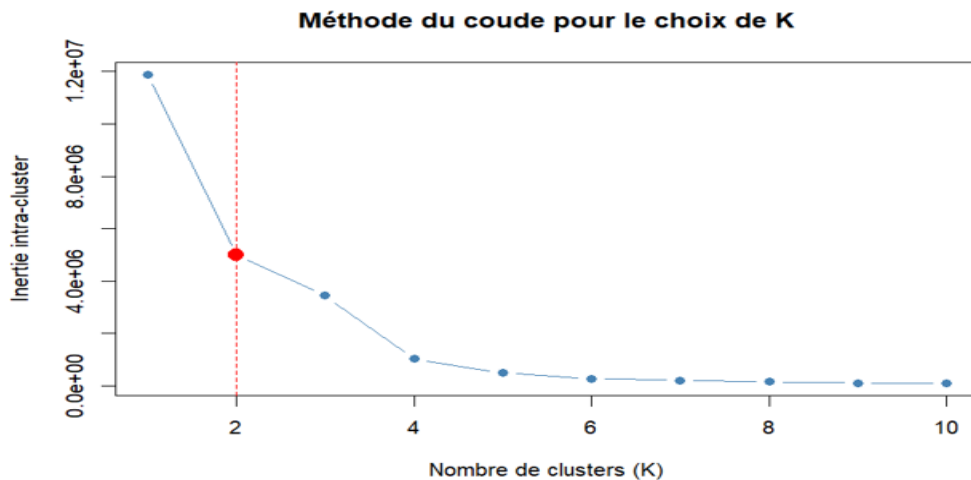
Le choix K-means de $K=2$, formulé dans les hypothèses est confirmé empiriquement par la méthode du coude et le score de silhouette, ainsi que le choix du cluster minoritaire comme cluster des ruptures.

Pour toutes les configurations testées ($w_s=3,5$ et $w_c=3,5,7,10,15,20$), le score de silhouette est maximal pour $K=2$. Les résultats du score de silhouette sont donnés en Annexe 2 (tableau N°10).

La figure N°4 illustre la méthode du coude dans le cas particulier $w_s=5$ et $w_c=10$. Une inflexion de l'inertie intra-cluster à $K = 2$, appelée coude, prouve que l'ajout de clusters supplémentaires n'améliore plus significativement la compacité des clusters.

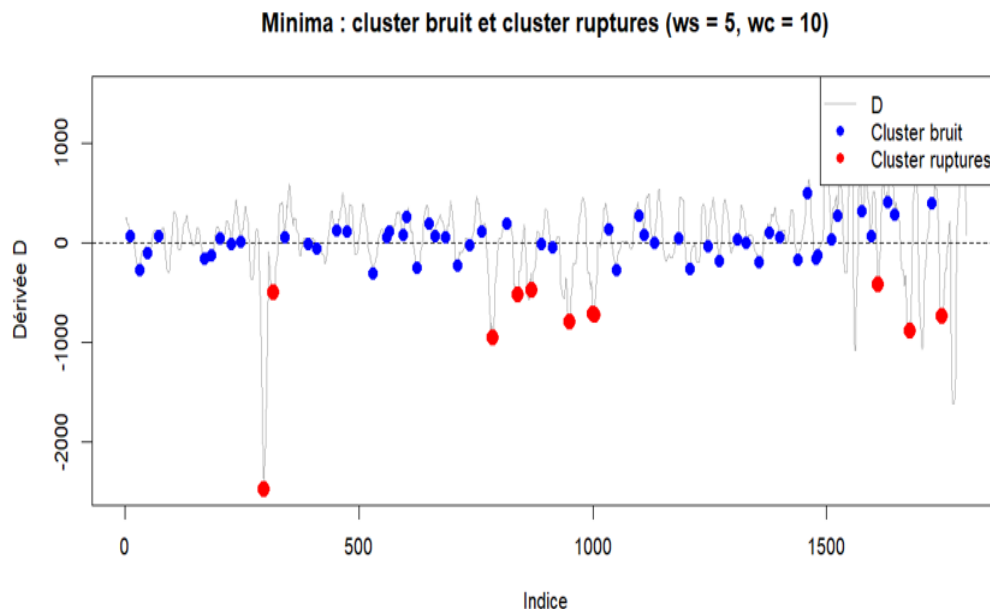
La figure N°5 illustre graphiquement, pour ce même cas, la séparation entre le cluster minoritaire qui est celui des ruptures du cluster majoritaire qui est celui du bruit.

Figure N° 4 : Méthode du coude (cas $w_s=5$, $w_c=10$)



Source : Auteurs – Calcul sur logiciel R

Figure N° 5 : Séparation des deux clusters (cas $w_s=5$, $w_c=10$)

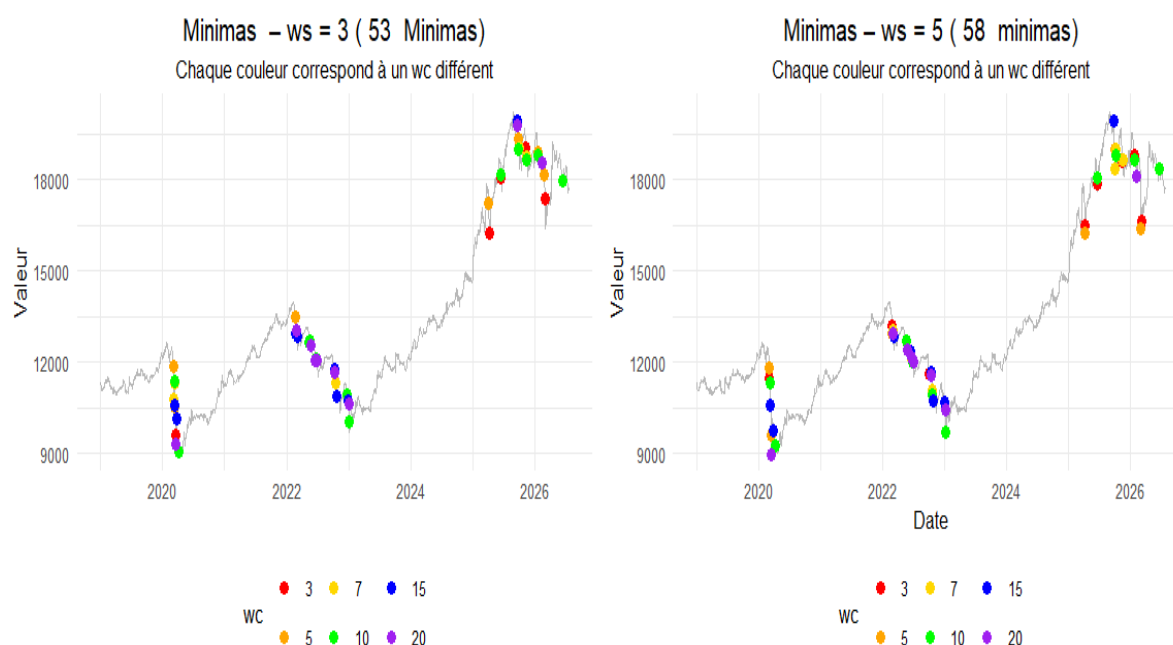


Source : Auteurs – Calcul sur logiciel R

4.3. Extrema locaux

La figure N°6 présente les minima locaux détectés par l'algorithme, à partir des deux types de lissages retenus $w_s = 5$ et $w_s = 3$ et avec les différents paramètres de coût w_c testés.

Figure N°6 : Lissage $w_s = 3$ et $w_s = 5$



Source : Auteurs – calcul sur logiciel R

4.4. Tableau des ruptures de niveau validées

Les extrema locaux détectés par l'algorithme (Etape 1 à 4) sont soumis à validation individuelle par un modèle ARIMA, augmenté d'une variable indicatrice de type step, relative à chaque date retenue. Une date est validée comme une rupture de niveau, lorsque la p-value du coefficient de la variable indicatrice est inférieur à 0.05, le critère d'Akaike AIC diminue par rapport au modèle ARIMA initial sans ruptures et les résidus restent un bruit blanc. Le tableau N° 6 liste les dates ainsi retenues, avec leur sens (hausse ou baisse de niveau), le coefficient δ de la variable indicatrice, la p-value associée et le ΔAIC , différence entre le AIC du modèle avec et sans la rupture.

Tableau N° 6 : Ruptures de niveau retenues (22)

Date	Sens	Coefficient δ	P-value	ΔAIC	w_s
2020-03-06	Baisse	-399 ;998	0.0007	9.30	5
2020-03-10	Hausse	711.465	0.0000	31.81	3,5
2020-03-12	Baisse	-728.535	0.0000	34 ;74	3, 5
2020-03-16	Baisse	-882.621	0.0000	42.84	3, 5
2020-03-18	Baisse	-272.590	0.0206	3.38	5
2022-02-24	Baisse	-557.162	0.0000	20.51	3, 5
2022-02-25	Hausse	245.941	0.0379	2.22	5

Tableau N°6 (suite)

Date	Sens	Coefficient δ	P-value	ΔAIC	w_s
2022-05-10	Baisse	-297.923	0.0111	4.42	5
2022-10-20	Baisse	-233.563	0.0467	2.01	5
2023-01-05	Baisse	-376.488	0.0014	8.19	3
2023-01-06	Baisse	-306.505	0.0091	4.85	5
2025-04-07	Baisse	-984.751	0.0000	58.43	3, 5
2025-04-08	Hausse	344.728	0.0040	6.32	5
2025-09-18	Hausse	250.669	0.0352	2.43	3
2025-09-29	Baisse	-297.934	0.0116	4.35	3, 5
2025-10-03	Hausse	471.314	0.0001	13.73	5
2025-11-11	Baisse	-455.405	0.0001	12.77	5
2026-02-05	Baisse	-308.968	0.0094	4.61	5
2026-02-27	Baisse	-292.582	0.0157	3.90	3
2026-03-02	Baisse	-762.412	0.0000	37.99	3
2026-03-03	Baisse	-926.956	0.0000	60.51	5
2026-03-04	Hausse	409.151	0.0007	9.43	5

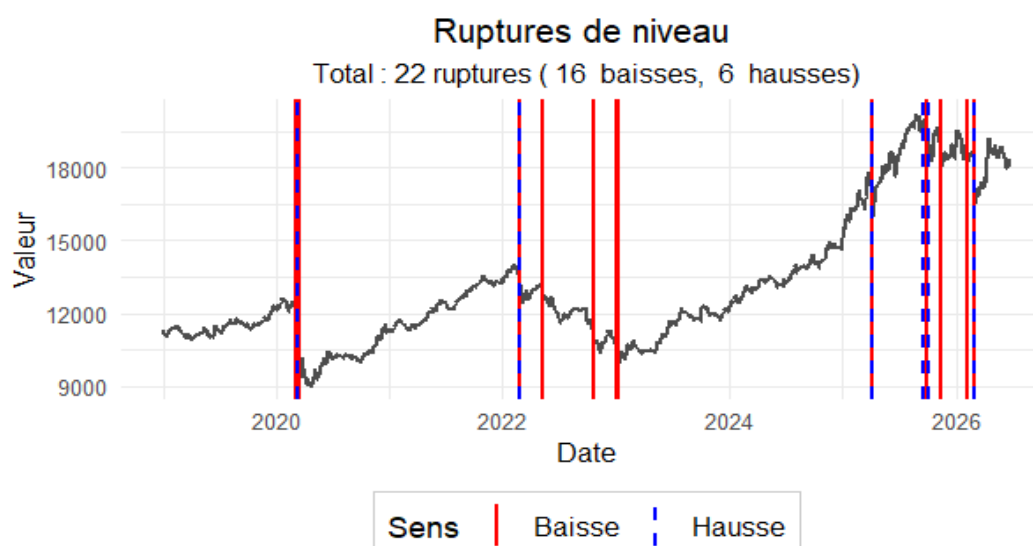
Source : Auteurs – Calcul sur logiciel R

Note : $\Delta AIC = AIC_{avec} - AIC_{sans}$. Une valeur positive indique une amélioration du modèle après introduction de la rupture.

4.5. Tracé des ruptures de niveau validées

La figure N°7 représente les ruptures (baisses et hausses de niveau) détectées et validées par l'algorithme.

Figure N°7: Ruptures de niveau validées



Source : Auteurs – Calcul sur logiciel R

4.6. Concordance temporelle des ruptures avec les épisodes de crises

Le tableau N°7 met en perspective les ruptures détectées avec les crises économiques et géopolitiques documentées, de la période d'analyse (Sources: Boursenews.ma, Le360.ma, média24.com, SNRTnews.com).

Tableau N°7 : Concordance temporelles des ruptures détectées avec les crises économiques de la période d'analyse.

Dates	Événements économiques ou géopolitiques
2020-03-06	- Début officiel du Krach boursier lié à la crise du (Covid-19) . - Décision de l'Autorité marocaine du Marché des Capitaux de réduire drastiquement les seuils de variations maximales des cours.
2020-03-10	Rebond technique spectaculaire et éphémère.
2020-03-12	Second Krach boursier, faisant suite à l'annonce de l'OMS du 11 Mars, qualifiant le covid de pandémie.
2020-03-16	Baisse du cours malgré la réduction drastique des seuils.
2020-03-18	Point bas historique du Krach du Covid-19, le plus bas enregistré depuis l'année 2016. Il marque la fin de la phase de capitulation.
2022-02-24	Jour de déclenchement de l'invasion de l'Ukraine par la Russie : → Fuite vers la liquidité et peur de l'inflation énergétique.
2022-02-25	Rebond technique correctif suite au choc psychologique de la veille.
2022-05-10	Baisse des cours : Publication ce jour-là de l'indice de confiance des investisseurs en Bourse qui passait sous la barre des 50 points : pessimisme et anticipation de mauvais résultats financiers à cause de l'inflation.
2022-10-20	Plongeon brutal suite à la présentation ce jeudi 20 octobre de la loi de finance, annonçant une hausse drastique de l'Impôt sur les Sociétés. Le même jour, le HCP annonce que l'inflation venait d'atteindre 8,3 % en Septembre

Tableau N°7 : (suite)

2023-01-05	Semaine noire marquant le début de l'année 2025 : - Ajustement du taux directeur - Entrée en vigueur des nouvelles mesures fiscales d'alignement progressif des taux de l'Impôt sur les Sociétés.
2023-01-06	→ Fuite des capitaux et arbitrages en défaveur des actions. Ce vendredi 6 Janvier : Fin de la phase de capitulation pure, malgré la baisse Marginale enregistrée ce jour-là.
2025-04-07	Krash boursier brutal provoqué par un choc géopolitique mondial, surnommé "Effet Trump" : Annonce de nouveaux droits de douanes massifs et punitifs ravivant les craintes d'une explosion de l'inflation à l'international Et un coup d'arrêt à la croissance économique
2025-04-08	Puissant rebond technique au lendemain du Krach historique de l'Effet Trump, prouvant la résilience et la maturité des institutionnels marocains.
2025-09-18	Forte impulsion boursière : Contexte macroéconomique favorable, marqué par de grandes introductions en bourse et l'arrivée de 20 000 nouveaux petits porteurs particuliers.
2025-09-29	Forte baisse due à plusieurs facteurs : 1. Correction technique, faisant entrer l'indice dans une phase de consolidation. 2. Publications financières du 1er semestre 2025 (TGCC). 3. Climat psychologique en raison des mouvements de protestation dans plusieurs villes marocaines (GenZ 212).
2025-10-03	Rebond technique puissant, marquant une performance annuelle exceptionnelle. Lancement officiel, ce jour-là d'un programme d'accompagnement de 31 entreprises, dédié à l'accès aux marchés des capitaux.
2025-11-11	Forte correction, s'inscrivant dans la volatilité accrue du dernier trimestre.
2026-02-05	Baisse s'inscrivant dans le mouvement de consolidation.

Tableau N°7 : (suite)

2026-02-27	Prolongement de la phase de consolidation, accompagné d'un comportement d'attentisme et de prudence des opérateurs de marché.
2026-03-02	Krach boursier d'origine géopolitique, suite aux frappes menées par les Etats-Unis et Israël contre l'Iran pendant le week-end du 28 Février-2 Mars.
2026-03-03	Krach consécutif à celui de la veille, sous l'effet d'une escalade militaire internationale
2026-03-04	Rebond technique, signalant un apaisement progressif de la psychologie des investisseurs d'attentisme et de prudence des opérateurs de marché..

Source : Informations relayées par les médias marocains

Conclusion : Les dates de ruptures de niveau détectées, coïncident avec précision avec les dates des crises économiques et géopolitiques majeures et d'autres événements économiques et financiers importants, relayés par les médias marocains et qui ont affecté la bourse de Casablanca. Néanmoins, il convient de souligner que les concordances temporelles observées ne permettent pas d'établir une relation de causalité entre ces événements et les dates détectées mais ces coïncidences viennent renforcer la crédibilité de nos résultats.

4.7. Prévisions hors échantillon

4.7.1. Méthode

Pour évaluer la capacité prédictive du modèle ARIMA augmenté de variables indicatrices relatives aux dates des ruptures structurelles identifiées, nous avons procédé à une prévision hors échantillon, utilisant l'intégralité de la série historique (T =1823 observations) décrivant la période du début janvier 2019 à la fin du mois de juin 2026. Les données utilisées pour la prévision ne font pas partie de l'échantillon d'analyse et n'ont servi ni à la détection des ruptures, ni à l'estimation du modèle ARIMA sélectionné. Ce sont de nouvelles observations recueillies à partir du début du mois de Juillet 2026 (3 Juillet 2026) jusqu'au 14 Septembre 2026 (36 observations). Les deux modèles comparés sont :

1. Un modèle ARIMA (2,1,2) sans rupture établi au paragraphe (3.4.2), servant de référence.
2. Un modèle ARIMAX, intégrant les ruptures de niveau détectées.

La sélection automatique de Box-Jenkins, implémenté via la fonction `auto.arima` du package `Forecast` pour R a permis de sélectionner un modèle ARIMAX(1,1,1) avec Drift, dont les caractéristiques (coefficients, p-values, critères AIC et BIC) sont données dans le tableau N° 8. Le modèle avec ruptures s'écrit :

$$\Delta Y_t = Drift + AR(1)\Delta Y_{t-1} + MA(1)\varepsilon_{t-1} + \sum_{i=1}^m \delta_i \cdot D_{i,t} + \varepsilon_t$$

où, $D_{i,t}$ sont les variables indicatrices associées à chaque date de rupture et δ_i sont leurs coefficients, listés à la deuxième colonne du tableau N° 8 . Les coefficients « Drift » et AR(1) et MA(1) sont également donnés dans le tableau N° 8.

4.7.2. Caractéristiques des deux modèles ARIMA : Avec et Sans ruptures

Le tableau N°8 présente les coefficients estimés des deux modèles, avec leurs écarts-types, p-values, critères AIC et BIC ainsi que le résultat du test de Ljung-Box sur la blancheur de leurs résidus.

Tableau N°8 : Caractéristiques des deux modèles ARIMA « Avec » et « Sans » ruptures

Modèle	Coef.	Écart-type	P-value
Avec ruptures			
AR(1)	0.52669	0.19348	0.0064853**
MA(1)	-0.47101	0.20019	0.0186320 *
Drift	7.1620	2.72579	0.0086016 **
2020-03-06	-321.38	104.89	0.0022 **
2020-03-10	626.18	104.82	2.31e-10 ***
2020-03-12	-747.59	103.89	6.22e-13 ***
2020-03-16	-918.48	103.74	< 2.2e-16 ***
2020-03-18	-327.47	103.49	0.0022467**
2022-02-24	-551.60	103.70	1.045e-07 ***
2022-02-25	225.52	103.70	0.0296483 *
2022-05-10	-316.00	103.42	0.0015555 *
2022-10-20	-237.85	103.52	0.0215819 *

Tableau N°8 (Suite)

2023-01-05	-397.40	103.61	0.0001254 ***
2023-01-06	-334.61	103.53	0.0012299 **
2025-04-07	-958.15	103.85	< 2.2e-16 ***
2025-04-08	234.19	104.02	0.024369 *
2025-09-18	207.48	103.57	0.045161 *
2025-09-29	-342.96	104.12	0.00098 ***
2025-10-03	470.15	103.93	6.07e-06 ***
2025-11-11	-429.86	103.61	3.34e-05 ***
2026-02-05	-340.78	103.44	0.000986***
2026-02-27	-389.08	103.60	0.000172***
2026-03-02	-771.72	103.73	1.01e-13 ***
2026-03-03	-991.11	103.76	< 2.2e-16 ***
2026-03-04	235.08	103.79	0.0235222 *
Diagnostic	AIC= 22131.19	BIC=22274.39	Ljung-Box (résidus): $p - value = 0.05312$
Sans rupture			
AR(1)	0.8795	0.1344	7.345e-15
AR(2)	-0.6358	0.1378	4.580e-09
MA(1)	-0.8264	0.1304	1.163e-14
MA(2)	0.6751	0.1292	4.043e-11
Diagnostic	AIC= 22576.34	BIC= 22603.88	Ljung-Box (résidus): $p - value = 0.08132$

Source : Auteurs – Calcul sur logiciel R

Pour le modèle avec ruptures, on observe un drift significatif ($\mu = 7.1620$, $p < 0.05$), indiquant une tendance de long terme positive, ainsi que des coefficients correspondant aux dates de ruptures de niveau, qui sont tous très significatifs.

Le modèle avec rupture réalise de bien meilleures performances en terme d'AIC et de BIC, par rapport au modèle initial sans rupture.

4.7.3. Comparaison des deux modèles : AIC, BIC et erreurs de prévision

Le tableau N°9 compare la qualité d'ajustement (AIC, BIC) entre les deux modèles (avec et sans ruptures) détaillés au tableau N°8 et les erreurs de prévision (RMSE, MAE, MAPE) à partir des 36 nouvelles observations [3 Juillet - 14 Septembre] 2026.

Tableau N° 9 : Comparaison des modèles : Avec et sans ruptures

Critère	Sans rupture	Avec ruptures	Amélioration (%)
AIC	22576.34	22131.19	2.0
BIC	22603.88	22274.39	1.5
RMSE	594.25	486.20	18.2
MAE	511.78	437.17	14.6
MAPE (%)	2.75	2.36	14.2

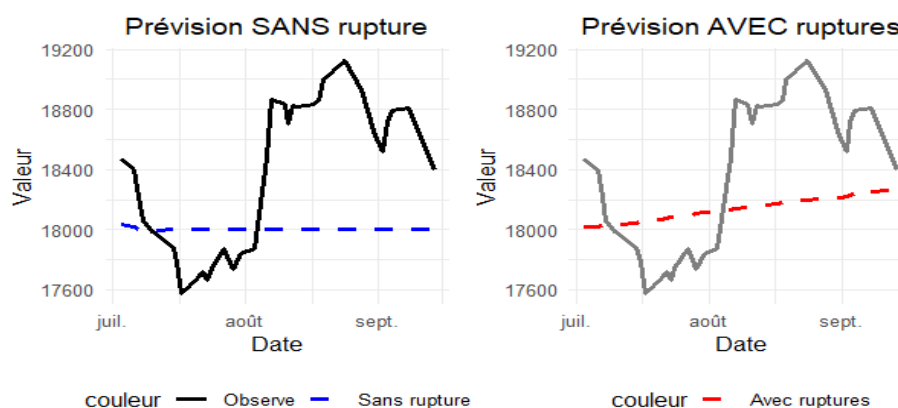
Source : Auteurs – Calcul sur logiciel R

Les résultats montrent que le modèle avec ruptures présente un AIC et un BIC plus faibles, indiquant un meilleur compromis entre ajustement et parcimonie sur la période d'estimation. Les erreurs de prévision hors échantillon sont plus faibles pour le modèle avec ruptures : le RMSE diminue de 18.2 %, le MAE de 14.6% et le MAPE de 14.2 % par rapport au modèle sans rupture.

4.7.4. Graphique des prévisions

La figure N°8 illustre les prévisions des deux modèles superposés aux valeurs observées sur la période hors échantillon.

Figure N° 8 : Prévisions hors échantillon « Avec » et « Sans » Ruptures



Source : Auteurs – calcul sur logiciel R

4.8. Conclusion

L'intégration des ruptures de niveau dans le modèle ARIMA a d'abord permis de réduire la persistance du modèle, faisant passer d'un ARIMA(2,1,2) à un ARIMAX(1,1,1). Ensuite, L'introduction des ruptures a rendu possible la sélection d'un modèle ARIMAX avec drift, permettant ainsi de capturer la tendance de long terme. Un modèle avec drift dans le modèle sans ruptures n'a pas pu être retenu, faute de critères améliorés par rapport à l'ARIMA(2,1,2). Enfin, ce modèle ARIMAX avec ruptures a accru la précision des prévisions hors échantillon, mesurée à l'aide des métriques RMSE, MAE et MAPE de 18,2%, 14,6% et 14,2%. Le modèle ARIMAX(1,1,1) avec ruptures et drift apparaît donc comme un outil de prévision plus fiable que l'ARIMA(2,1,2) sans ruptures.

Ces résultats soulignent l'utilité d'une détection préalable des ruptures pour la modélisation et la prévision des séries financières.

Conclusions

Cet article a présenté une méthodologie hybride pour la détection des ruptures de niveau dans les séries financières, appliquée au cas de l'indice MASI de la bourse de Casablanca sur la période [2019 - 2026].

L'approche combine une méthode heuristique de détection des ruptures, largement utilisée en traitement de signal et de données (lissage médian, dérivée de niveau, détection d'extrema, regroupement et clustering K-means) avec une validation économétrique par modèle ARIMAX.

L'originalité de notre approche repose sur l'adaptation de la méthode heuristique de détection des ruptures au cas des séries financières et son couplage avec une validation ARIMA. Les dates ruptures identifiées par notre algorithme coïncident très précisément avec les épisodes de crises économiques et géopolitiques documentées, sans que cela ne permette de prétendre à un lien de causalité direct.

Par ailleurs, le modèle final ARIMAX(1,1,1) avec variables indicatrices associées aux ruptures structurelles identifiées, offre un meilleur compromis entre parcimonie et qualité d'ajustement, en comparaison avec le modèle ARIMA initial. De plus, la prise en compte des ruptures de niveau dans le modèle de prévision permet d'améliorer très significativement la précision des prévisions hors échantillon du prix de l'indice MASI . Ces résultats ouvrent des perspectives pour la détection de rupture structurelles dans les séries brutes non stationnaires, à racine unitaire.

Une première perspective est de considérer d'autres séries financières qui seraient intégrés à un ordre supérieur comme l'ordre 2.

Sur le plan méthodologique et dans un objectif de complémentarité, une comparaison avec les méthodes existantes (Bai-Perron, KCP,...) appliquées à la série des rendements apparaît nécessaire. Cette comparaison permettrait de comprendre les liens entre les changements de niveau dans la série brute et le changement de dynamique dans la série des rendements à travers les paramètres autorégressifs AR et moyennes mobiles MA.

Annexe 1 : Algorithme de détection des ruptures

Pour chaque valeur de w_s et w_c , effectuer les étapes 1 à 6 :

Etape 1 : Lissage par médiane glissante

Pour $i \leftarrow 1 : T - w_s + 1$, $L[i] \leftarrow \text{médiane}(y_i, y_{i+1}, \dots, y_{i+w_s-1})$

Etape 2 : Calcul de la dérivée. Pour $n \leftarrow (w_c : \text{longueur}(L) - w_c)$

pour $k \leftarrow 1 : n$

$G \leftarrow L[k - w_c + 1], \dots, L[k]$

$D \leftarrow L[k + 1], \dots, L[k + w_c]$

$\Delta[k] \leftarrow \text{médiane}(D) - \text{médiane}(G)$

Etape 3 : Extraction des extrema locaux. Pour $i \leftarrow 2 : \text{longueur}(\Delta) - 1$

Si $\Delta[i] < \Delta[i - 1]$ et $\Delta[i] < \Delta[i + 1]$ minima $\leftarrow \text{minima} \cup \{i\}$

Si $\Delta[i] > \Delta[i - 1]$ et $\Delta[i] > \Delta[i + 1]$ maxima $\leftarrow \text{maxima} \cup \{i\}$

Etape 4 : Regroupement par blocs. Pour $a \leftarrow 1 : \text{longueur}(\Delta)$ par pas de w_{se}

$b \leftarrow a + w_{se} - 1$

bloc $\leftarrow \{i \in \text{minima} \mid a \leq i \leq b\}$; $c \leftarrow \text{argmin}_{i \in \text{bloc}} \Delta[i]$; candidats $\leftarrow \text{candidats} \cup \{c\}$

Etape 5 : Clustering K-means

$X \leftarrow \{\Delta[c] \mid c \in \text{candidats}\}$

Exécuter K-means($X, K = 2$) avec une initialisation aléatoire et répéter plusieurs fois jusqu'à convergence. Obtenir deux clusters $\mathcal{C}_1$ et $\mathcal{C}_2$ de tailles n_1 et n_2

cluster_rupture $\leftarrow \text{argmin}(n_1, n_2)$

retenus $\leftarrow \{c \in \text{candidats} \mid \text{cluster}(c) = \text{cluster_rupture}\}$

Etape 6 : Validation individuelle par ARIMA(2,1,2)

Pour $indice \in \text{retenus}$

Créer $D_t = 0$ si $t < indice$, sinon 1

Estimer ARIMAX(2,1,2) avec $xreg = D_t$

Extraire le coefficient δ de $xreg$ et sa p-value

Si $p < 0.05$ et AIC diminue: validées $\leftarrow \text{validées} \cup \{indice\}$

Annexe 2 : Score de silhouette et K optimal pour les différentes configurations (W_s, W_c):

Tableau N°10 : K optimal par configuration et score de silhouette correspondant

W_s	W_c	K	Score de Silhouette
3	3	2	0.7948
3	5	2	0.7567
3	7	2	0.7796
3	10	2	0.6804
3	15	2	0.7430
3	20	2	0.7019
5	3	2	0.7392
5	5	2	0.7236
5	7	2	0.7604
5	10	2	0.6936
5	15	2	0.7477
5	20	2	0.7179

Bibliographie

- [1] Arlot, S., Celisse, A., & Harchaoui, Z. (2019). A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162), 1-56.
- [2] Aue, A. et al. (2013). Structural Breaks in Times Series. *Journal of Time Series Analysis*, 34, 1-16
- [3] Bai, J., & Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural changes, *Econometrica*, 66(1), 47-78.
- [4] Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1-22.
- [5] Barry, D., & Hartigan, J. A. ([1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421), 309-319.
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Mai 2015. 720 pages.
- [7] Brown, R. L., Durbin, J., & Evans, J.M. (1975). Techniques for testing the constancy of Regression relationships Over times. *Journal of Royal Statistical Society: series B (Methodological)*, 37(2), 149-163.
- [8] Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3), 591-605.
- [9] Chu, L., & Chen, H. (2019). Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *The Annals of Statistics*, 47(1), 382-414.
- [10] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1), 33-43.
- [11] Erdman., C., & John W. E., (2007), bcp: An R Package for Performing a Bayesian Analysis of Change Point Problems. *Journal of Statistical Software*, 23(3), 1-13.
- [12] Flynn, T., & Yoo, S. (2019). Change detection with the kernel cumulative sum algorithm. *IEEE 58th Conference on Decision and Control*, 6092-6099
- [13] Fried, R. (2004). Robust filtering of time series with trends. *Journal of nonparametric statistics*, 16(3-4), 313-328.
- [14] Fryzlewicz, P. 2014). Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42(6), 2243-2281.
- [15] Garreau, D., & Arlot, S. (2018). Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2), 4448-4486.

- [16] Harchaoui, Z., & Cappé, O. (2007). Retrospective Multiple Change-point Estimation with Kernels ; 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, 768-772.
- [17] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.
- [18] Karrouche, I. (2020), Détection des ruptures et d'anomalies dans les séries temporelles, Rapport de stage - MASTER 2, Université - Paris Descartes - ENGIE digital.
- [19] Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590-1598.
- [20] Kulasekera, K. B., & Saxena, K. M. L. (1991). Estimation of change point in failure rate models. *Journal of Statistical Planning and Inference*, 29(1-2), 111-124.
- [21] Li, X., & Zhang, X. (2026). fastcpd: Fast change point detection in R. *Journal of Statistical Software*, 116(6), 1-53.
- [22] Lucas, J. M. (1985). Cumulative sum (CUSUM) control schemes. *Communications in Statistics - Theory and Methods*, 14(11), 2689-2704.
- [23] MacQueen, J. (1966). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.
- [24] C. R., Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-162.
- [25] Otto, S., & Breitung, J. (2023). Backward CUSUM for Testing and Monitoring Structural Change with an Application to COVID-19 Pandemic Data. *Econometric Theory*, 39(4), 659-692.
- [26] Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361-1401.
- [27] Prodon, R. (2008). Potential pitfalls in determining multiple structural changes with an Application to Purchasing Power Parity. *Journal of Business & Economic Statistics*, 26(1), 50-65
- [28] Ran, M., Tang, Z., Chen, Y., & Wang, Z. (2024). Early warning of systemic risk in stock market based on EEMD-LSTM. *PLoS ONE*, 19(5), e0300741.
- [29] Reynolds, M. R., & Stoumbos, Z. G. (2000). A general approach to modeling CUSUM charts for a proportion. *IIE Transactions*, 32(6), 515-535.

- [30] Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3), 507-512
- [31] Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267-276.
- [32] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- [33] Wibawa. A., -Median Smoothing Backpropagation Neural Network to Forecast Unique Visitors Time Series of Electronic Journal, *Journal of applied data sciences*, vol 4(3), 163-174.
- [34] Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and Dating of Structural Changes in Practice.. *Computational Statistics & Data Analysis*, 44(1-2), 109-123.
- [35] Zhong, C. (2026) Statistical inference and structural break detection in nonstationary ARMA model with dependent innovation. *Journal of Time Series Analysis*. 47(3). Advanced on line publication: doi.org/10.1111/jtsa.700080
- [36] Zhao, L., Li, Z., Zhang, J., & Teng, B. (2023). An integrated CEEMDAN-LSTM joint model for significant wave height forecasting. *Journal of Marine Science and Engineering*, 11(2).
- [37] Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3), 251-270.