ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

Signaux textuels multilingues et prévision de la volatilité réalisée : une approche hybride Transformer-HAR-X appliquée au S&P 500 et au MASI

Multilingual Textual Signals and Realized Volatility Forecasting: A Hybrid Transformer-HAR-X Approach Applied to the S&P 500 and the MASI

Sara BOUGHANOU, *(Doctorante)*
Faculté des Sciences Juridiques, Économiques et Sociales de Rabat-Agdal
Université Mohammed V de Rabat, Maroc

Adil EL MARHOUM, *(Enseignant-chercheur)*
Faculté des Sciences Juridiques, Économiques et Sociales de Rabat-Agdal
Université Mohammed V de Rabat, Maroc

Adresse de correspondance :	Faculté des Sciences Juridiques, Économiques et Sociales de Rabat-Agdal, avenue des Nations Unies, B.P. 721 Agdal, Rabat Université Mohammed V de Rabat, Maroc Téléphone : +212 5 37 22 57 48 / 39
Déclaration de divulgation :	Les auteurs n'ont pas connaissance de quelconque financement qui pourrait affecter l'objectivité de cette étude. Ils assument l'entière responsabilité de tout éventuel plagiat, de l'usage de l'intelligence artificielle dans la rédaction, ainsi que des résultats présentés dans cet article.
Conflit d'intérêts :	Les auteurs ne signalent aucun conflit d'intérêts.
Citer cet article	BOUGHANOU, S., & EL MARHOUM, A. (2026). Signaux textuels multilingues et prévision de la volatilité réalisée : une approche hybride Transformer-HAR-X appliquée au S&P 500 et au MASI. *International Journal of Accounting, Finance, Auditing, Management and Economics*, *7*(10), 109–132. https://doi.org/10.5281/zenodo.21909207
Licence	**Cet article est publié en open Access sous licence CC BY-NC-ND**

Signaux textuels multilingues et prévision de la volatilité réalisée : une approche hybride Transformer-HAR-X appliquée au S&P 500 et au MASI

Résumé

Prévoir la volatilité réalisée soutient la gestion du risque et l'allocation. GARCH et HAR-RV exploitent la mémoire numérique ; les hybrides LSTM-GARCH et FinBERT-LSTM sont souvent monolingues et rarement évalués hors échantillon. Nous proposons un cadre bilingue FinBERT-CamemBERT qui construit l'indice de stress S_t et le vecteur événementiel E_t, intégrés à HAR-X pour le S&P 500 et le MASI.
Le corpus rassemble 12,3 millions d'items sur 2015-2024. Une validation par fenêtre croissante réserve 2023-2024 au test final. Face à HAR-RV, HAR-X réduit le RMSE du S&P 500 de 21,1 % ; le test de Diebold-Mariano donne DM = -2,44, p = 0,015 et qBH = 0,033. Pour le MASI, DM = -2,62 et qBH = 0,033. Les six comparaisons sont corrigées par Benjamini-Hochberg. Le ratio de Sharpe net de frais passe de 0,64 à 0,81, mais ce gain reste descriptif faute de test direct de Ledoit-Wolf. Vingt-trois écarts mensuels sur vingt-quatre sont positifs, avec p < 0,001 au test exact des signes.
Le cadre relie information textuelle multilingue, mémoire hétérogène et allocation. SHAP attribue un rôle prédictif à S_t et E_t sans établir de relation causale. Les limites concernent l'annotation, le change et la dérive conceptuelle.

Mots-clés : volatilité financière, Transformer, analyse de sentiment, HAR-X, optimisation de portefeuille, MASI
Classification JEL : C53, C58, G17, G11.
Type du papier : Recherche empirique.

Abstract

Realized-volatility forecasting supports risk management and allocation. GARCH and HAR-RV mostly exploit numerical memory, whereas LSTM-GARCH and FinBERT-LSTM hybrids are often monolingual and rarely receive out-of-sample economic evaluation. We propose a bilingual FinBERT-CamemBERT framework that constructs the daily textual stress index S_t and the event vector E_t, which enter HAR-X for the S&P 500 and the MASI.
The corpus contains 12.3 million French and English items over 2015-2024. Expanding-window validation reserves 2023-2024 for final testing. Against HAR-RV, HAR-X reduces the S&P 500 RMSE by 21.1%; the Diebold-Mariano test yields DM = -2.44, p = .015, and qBH = .033. For the MASI, DM = -2.62 and qBH = .033. Benjamini-Hochberg correction covers six comparisons. The net-of-cost Sharpe ratio rises from .64 to .81, but this gain remains descriptive without a direct Ledoit-Wolf test. Twenty-three of twenty-four monthly differentials are positive, with p < .001 under an exact sign test.
The framework connects multilingual textual information, heterogeneous memory, and allocation. SHAP assigns a major predictive role to S_t and E_t without establishing a causal relationship. Limitations concern annotation reliability, foreign exchange, and concept drift.

Keywords: Financial volatility, Transformer, sentiment analysis, HAR-X, portfolio optimization, MASI.
JEL Classification: C53, C58, G17, G11.
Paper type: Empirical Research.

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

1. Introduction

La maîtrise de la volatilité constitue un enjeu central pour la gestion des risques et la construction de portefeuilles. En finance, la variance conditionnelle, notée $\sigma^2_{t|t-1}$, désigne la variance future attendue des rendements compte tenu de l'information disponible à la date t-1. Elle demeure latente et doit être estimée. La volatilité réalisée constitue, en revanche, une mesure ex post construite à partir des rendements intrajournaliers observés. Dans cet article, RV_t désigne la variance réalisée quotidienne et $\sqrt{RV_t}$ la volatilité réalisée, annualisée lorsque les résultats sont exprimés en points de volatilité. Les modèles GARCH et HAR-RV représentent la persistance et l'hétérogénéité temporelle de ces séries, mais mobilisent principalement l'information numérique passée (Engle, 1982 ; Bollerslev, 1986 ; Corsi, 2009). Cette distinction conduit à examiner si les textes disponibles à la date t apportent une information incrémentale pour prévoir RV_{t+1} au-delà de sa mémoire statistique.

Une limite informationnelle subsiste lorsque les modèles sont conditionnés presque exclusivement par les rendements et les mesures de volatilité retardés. Les nouvelles exogènes ne deviennent alors observables par le modèle qu'après leur traduction dans les prix. Les travaux associant données textuelles, événements et séries financières suggèrent qu'une information complémentaire peut être extraite avant cette incorporation complète (Gu et al., 2024 ; Mou et al., 2024 ; Kurisinkel et al., 2024). Cette formulation prudente ne suppose pas que GARCH ou HAR-RV réagissent systématiquement avec retard. Elle motive plutôt l'évaluation d'un gain incrémental produit par des signaux textuels horodatés et disponibles avant la décision.

La contribution méthodologique consiste à coupler FinBERT pour les textes anglophones et CamemBERT pour les textes francophones à un classifieur événementiel distinct. Leurs sorties construisent l'indice de stress textuel S_t et le vecteur d'événements E_t, introduits dans un modèle HAR-X. Le dispositif est évalué sur le S&P 500 et le MASI, puis relié à une allocation moyenne-variance conditionnelle. Les indices constituent les marchés de référence de l'analyse. L'univers d'investissement exact utilisé dans le back-test n'étant pas documenté dans les archives disponibles, les résultats de portefeuille sont interprétés comme une évaluation agrégée et non comme une réplication titre par titre.

Sur le plan théorique, l'étude prolonge la logique d'hétérogénéité des horizons du HAR-RV (Corsi, 2009) en considérant S_t et E_t comme des variables d'état informationnelles de court terme. Les composantes autorégressives conservent la mémoire journalière, hebdomadaire et mensuelle de la volatilité, tandis que les signaux textuels contextualisés représentent les révisions rapides de la perception collective du risque. Cette articulation rapproche l'économétrie de la volatilité et la finance textuelle, puis relie la prévision à son usage économique. Sa portée demeure prédictive et ne fonde aucune interprétation causale.

Le choix conjoint du S&P 500 et du MASI répond à une logique de contraste informationnel. Le premier constitue un marché développé, profond et couvert en continu par une presse financière majoritairement anglophone. Le second, indice de la Bourse de Casablanca, présente une capitalisation plus concentrée, une liquidité intrajournalière plus faible et une couverture médiatique principalement francophone. Cette asymétrie de profondeur de marché et de densité informationnelle permet d'examiner si l'apport prédictif des signaux textuels dépend du régime informationnel dans lequel ces signaux sont produits. La finance textuelle s'est en effet développée très majoritairement sur des marchés anglophones, de sorte que la transposabilité de ses conclusions à un marché émergent francophone demeure une question empirique ouverte. La période retenue, 2015-2024, couvre par ailleurs des régimes de volatilité contrastés, dont l'épisode de mars 2020 et le resserrement monétaire de 2022. Cette couverture limite le risque qu'un gain de prévision mesuré ne reflète qu'une configuration de marché particulière.

Les propositions suivantes traduisent ce questionnement en hypothèses testables.

H1. Sur le jeu de test final hors échantillon, l'introduction de S_t et E_t dans le modèle HAR-X produit une perte QLIKE significativement inférieure à celle du HAR-RV pour la prévision à un jour de RV_{t+1} sur le S&P 500 et le MASI.
H2. Sur la même période et après prise en compte des coûts de transaction, le portefeuille fondé sur les prévisions HAR-X présente un ratio de Sharpe supérieur à celui du portefeuille fondé sur HAR-RV. La validation statistique directe de cet écart requiert toutefois les séries appariées de rendements.
L'article s'organise en cinq temps. La deuxième section recense la littérature consacrée à la prévision de la volatilité, aux représentations textuelles contextualisées et aux architectures hybrides, puis en dégage les lacunes qui fondent le positionnement retenu. La troisième section expose le cadre conceptuel et la méthodologie : construction de l'indice de stress textuel S_t et du vecteur d'événements E_t, spécification du modèle HAR-X, protocole de validation par fenêtre croissante et procédure d'allocation moyenne-variance conditionnelle. La quatrième section présente les résultats de prévision hors échantillon pour le S&P 500 et le MASI, les tests de comparaison de précision ainsi que l'analyse de la contribution des variables. La cinquième section discute la portée et les limites de ces résultats, en particulier celles qui tiennent à la non-disponibilité de certaines archives de réplication. La conclusion synthétise les apports et esquisse des prolongements.

2. Revue de littérature

La littérature sur la prévision de la volatilité s'organise autour de plusieurs familles. Les modèles ARCH et GARCH (Engle, 1982 ; Bollerslev, 1986) ainsi que leurs extensions asymétriques, notamment EGARCH (Nelson, 1991) et GJR-GARCH (Glosten et al., 1993), modélisent la variance conditionnelle à partir des innovations passées. Ils restent sensibles aux hypothèses distributionnelles et aux queues épaisses. Les modèles HEAVY exploitent les mesures réalisées afin de rapprocher l'estimation conditionnelle de l'information haute fréquence (Shephard & Sheppard, 2010 ; Noureldin et al., 2012). Enfin, HAR-RV décompose la volatilité réalisée selon des horizons journalier, hebdomadaire et mensuel afin de reproduire sa mémoire longue (Corsi, 2009 ; Clements & Preve, 2021).
L'essor du traitement automatique du langage naturel appliqué à la finance a ouvert une voie complémentaire. FinBERT adapte les représentations contextuelles au vocabulaire financier anglophone (Araci, 2019), tandis que CamemBERT fournit une architecture pré-entraînée adaptée aux textes francophones (Martin et al., 2020). La détection d'événements constitue une tâche distincte de la classification du sentiment. Elle vise à identifier les annonces de résultats, les décisions de politique monétaire, les opérations de fusion-acquisition ou les épisodes de risque systémique. Les scores obtenus par ces deux tâches ne sont donc pas directement comparables.
Les architectures hybrides recouvrent des objets empiriques distincts. Kim et Won (2018) et Batool et al. (2022) associent des composantes économétriques et l'apprentissage automatique pour représenter les non-linéarités de la volatilité, tandis que Bulut et Hüdaverdi (2022) examinent plus largement la combinaison de méthodes dans les séries financières. Li et al. (2023) illustrent une extension GARCH-MIDAS intégrant des déterminants de fréquences différentes. Sur le versant textuel, Mou et al. (2024) combinent séries financières et données textuelles, Thakur et al. (2024) intègrent les nouvelles financières à un dispositif de prévision, et Gu et al. (2024) relient FinBERT à une architecture LSTM. Kurisinkel et al. (2024) proposent une actualisation événementielle des séries. Ces travaux soutiennent la fusion multimodale sans rendre leurs performances interchangeables.
Les conclusions de cette littérature demeurent hétérogènes, car les études ne prédisent ni le même objet ni le même horizon. Certaines portent sur les prix ou les rendements, d'autres sur la volatilité. Les langues, les volumes de corpus, les découpages temporels et les métriques

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

varient également. Un gain obtenu par FinBERT-LSTM sur un corpus anglophone ne peut donc être transposé directement à une prévision bilingue de RV_{t+1}. De plus, une meilleure qualité d'ajustement ne garantit ni la robustesse hors échantillon ni une performance nette de coûts. Cette diversité justifie une comparaison fondée sur les mêmes périodes, les mêmes métriques et une séparation chronologique commune.

Trois lacunes structurent dès lors le positionnement de l'article. La dimension multilingue reste peu étudiée sur les marchés émergents. La reproductibilité est souvent limitée par l'absence de code, de labels annotés et de règles explicites d'alignement entre textes et prix. Enfin, peu de travaux distinguent clairement la précision statistique de la prévision et son utilité économique pour l'allocation. Le présent article traite conjointement ces dimensions, tout en signalant les informations de réplication qui demeurent indisponibles.

Le design expérimental repose sur un corpus bilingue français-anglais de 12,3 millions d'items, couplé aux séries haute fréquence du S&P 500 et du MASI sur 2015-2024. L'unité de collecte est l'item, qu'il s'agisse d'un article, d'une dépêche ou d'un message court, et non le document journalistique long. Le protocole comprend l'ajustement supervisé de FinBERT et de CamemBERT, un classifieur événementiel distinct, l'extraction quotidienne de S_t et de E_t, puis leur injection dans HAR-X. La validation chronologique par fenêtre croissante et les tests de Diebold-Mariano (Diebold & Mariano, 1995) visent à limiter les fuites d'information. Le back-test incorpore des coûts de transaction compris entre 0,01 % et 0,10 %, sous réserve des limites de réplication précisées plus loin.

Tableau 1 : Aperçu des principaux modèles hybrides ML-économétrie et de leurs bénéfices.

Modèle hybride	Avantage clé	Référence
ARIMA-LSTM	Capture simultanée de tendances linéaires et non linéaires	Hamiane et al., 2024
LSTM-GARCH / GARCH hybride	Apprentissage de résidus et dynamiques non linéaires	Kim & Won, 2018
FinBERT-LSTM	Intégration de sentiment financier dans la prévision	Gu et al., 2024

Source : élaboration des auteurs.

Les scores F1 du Tableau 2 sont présentés comme des diagnostics propres à chaque composante. FinBERT et CamemBERT sont évalués sur des sous-ensembles linguistiques différents, tandis que le lexique fournit un repère simple. Ils ne doivent pas être lus comme un classement strict entre langues ou tâches. La détection des événements E_t fait l'objet d'un classifieur distinct et n'est pas incluse dans cette comparaison de sentiment.

Tableau 2 : Scores F1 des composantes de classification du sentiment sur le sous-échantillon annoté.

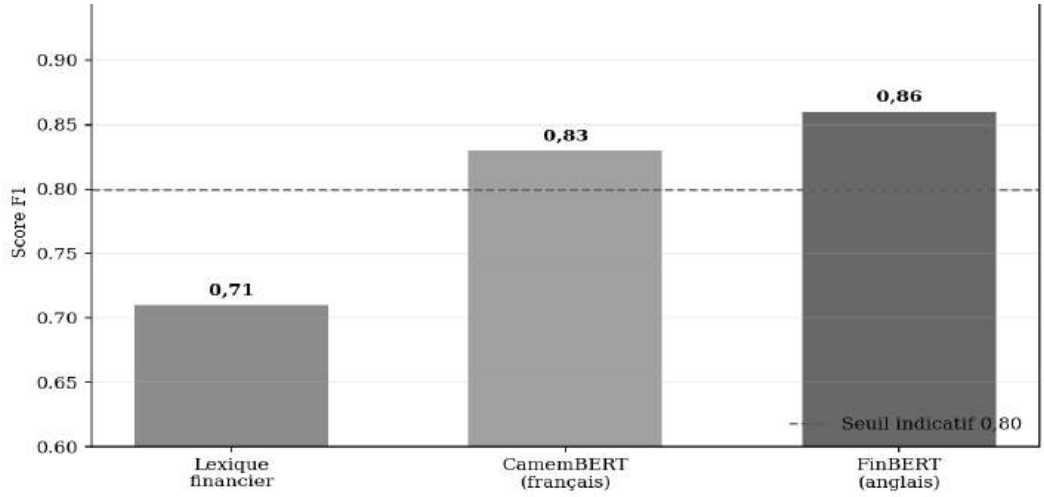
Technique TALN	Application principale	F1-score sur le sous-échantillon annoté
Lexique financier	Score de sentiment lexical simple	0,71
FinBERT	Sentiment contextualisé anglophone	0,86
CamemBERT	Sentiment contextualisé francophone	0,83

Source : calculs des auteurs.

La Figure 1 présente les scores F1 disponibles pour les composantes de sentiment. Le lexique financier atteint 0,71, CamemBERT 0,83 sur le sous-ensemble francophone et FinBERT 0,86 sur le sous-ensemble anglophone. Ces valeurs soutiennent l'usage de représentations contextuelles, sans autoriser une comparaison directe entre langues. La performance du classifieur événementiel doit être évaluée séparément à partir de ses propres labels.

À la différence des travaux recensés au Tableau 1, la contribution proposée n'est pas intégrée comme une référence antérieure. Elle est positionnée séparément comme une extension interprétable de HAR-RV qui associe des signaux textuels bilingues à une validation chronologique commune sur un marché développé et un marché émergent. La section suivante précise les notations, l'alignement temporel et les limites de reproductibilité.

Figure 1 : Scores F1 des composantes de classification du sentiment financier

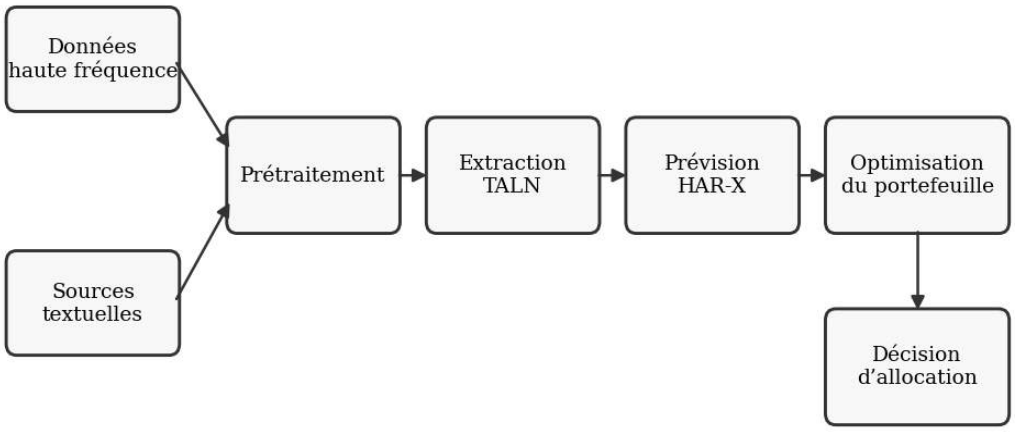


Source : élaboration des auteurs à partir des données et résultats de l'étude.

3. Cadre conceptuel et méthodologie

Le dispositif empirique suppose que RV_{t+1} dépend à la fois d'une mémoire statistique, captée par les termes autorégressifs, et d'un flux textuel reflétant la perception du risque. La Figure 2 illustre la chaîne de traitement depuis l'ingestion des données jusqu'à la décision d'allocation. Les variables textuelles utilisées pour prévoir RV_{t+1} sont construites exclusivement à partir des informations disponibles avant l'heure de décision du jour t. Toutes les figures de l'article sont des productions originales des auteurs, générées à partir des données, des sorties empiriques ou du schéma méthodologique de l'étude. Aucune image tierce n'est reproduite.

Figure 2 : Pipeline expérimental reliant les données haute fréquence et l'extraction TALN à HAR-X et à l'allocation



Source : élaboration des auteurs à partir des données et résultats de l'étude.

3.1. Construction des variables textuelles

La collecte textuelle s'étend du 2 janvier 2015 au 29 décembre 2024 et couvre 12,3 millions d'items provenant de Reuters, Bloomberg et d'archives Twitter/X sous licence. Après dédoublonnage, filtrage linguistique et retrait des contenus promotionnels, 11,8 millions d'items demeurent, soit 95,9 %. Les URL, balises HTML et caractères non imprimables sont supprimés, tandis que les horodatages sont conservés. Les coupures sont définies selon l'heure locale de chaque marché afin qu'un texte publié après la clôture de Casablanca ou de New York ne soit affecté qu'à la prochaine décision admissible sur le marché concerné.

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

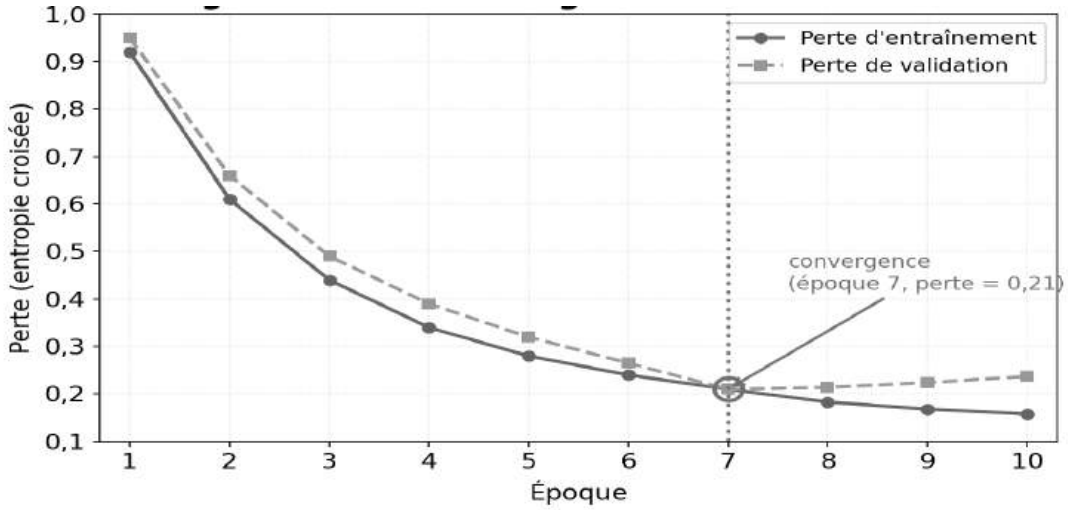
IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

La tokenisation est réalisée séparément au moyen du tokeniseur natif de chaque modèle, WordPiece pour FinBERT et SentencePiece pour CamemBERT, conformément au pseudocode de l'Annexe 8.2.2. Le fine-tuning utilise un sous-échantillon stratifié de 180 000 items, réparti entre 90 % d'entraînement et 10 % de validation, avec un taux d'apprentissage de 2e-5. Les paramètres de standardisation et de prétraitement sont estimés uniquement sur la fenêtre d'apprentissage de chaque itération, puis appliqués au mois prévu. Le reste du corpus fait ensuite l'objet d'une inférence.

Le protocole archivé ne conserve pas la matrice de double codage nécessaire au calcul d'un coefficient de Cohen ou de Fleiss. Aucun coefficient kappa ne peut donc être rapporté rétrospectivement. Les scores F1 du Tableau 2 mesurent la performance des classifieurs sur le jeu de validation et ne remplacent pas une mesure d'accord inter-annotateurs. Cette lacune limite la qualification des étiquettes comme vérité terrain et devra être corrigée lors d'une réplication par un double codage documenté.

L'ajustement réalisé une seule fois suppose également une stabilité du vocabulaire financier sur 2015-2024. Or, les plateformes, les usages linguistiques et les thèmes dominants peuvent évoluer. La validation chronologique mesure la stabilité des prévisions, mais ne constitue pas un test direct de dérive sémantique. Une extension devra suivre les distributions textuelles et les performances par sous-période, puis déclencher un réentraînement à partir des seules données disponibles avant chaque date de prévision.

Figure 3 : Courbes agrégées de convergence du fine-tuning de FinBERT et de CamemBERT

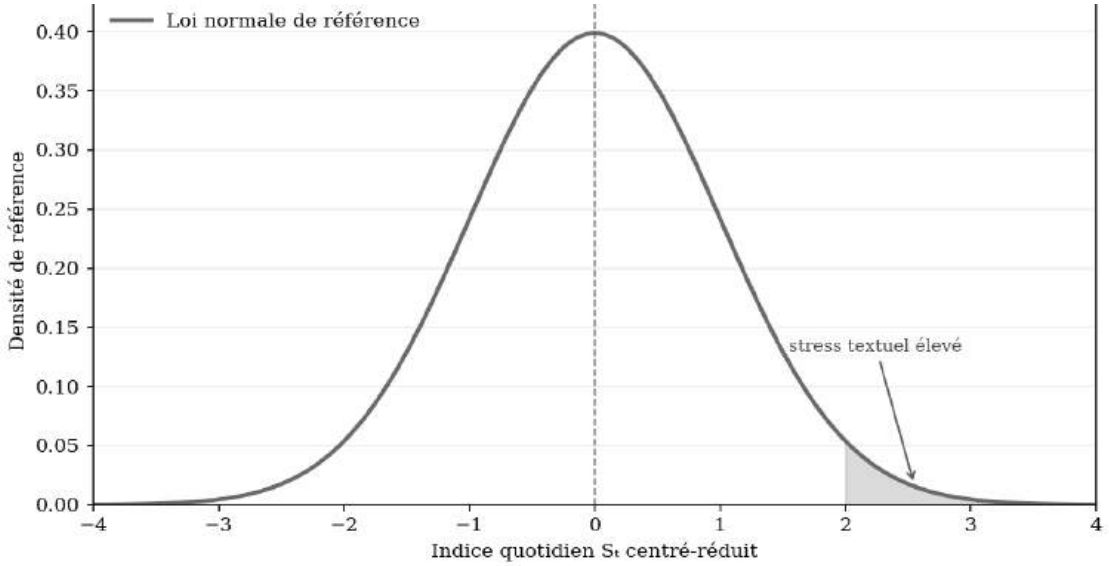


Source : élaboration des auteurs à partir des données et résultats de l'étude.

Les deux courbes décrivent la perte agrégée d'entraînement et de validation sur les exécutions présentées. Elles décroissent jusqu'à la septième époque, puis la perte de validation se stabilise. La figure ne sépare cependant pas les trajectoires propres à FinBERT et à CamemBERT. Elle doit donc être lue comme un diagnostic global de convergence, et non comme la preuve que les deux modèles suivent exactement la même dynamique.

Les embeddings contextuels de dimension 768 sont extraits au niveau de la représentation CLS pour FinBERT et de la représentation équivalente de séquence pour CamemBERT. À l'échelle journalière, ils sont agrégés avant l'heure de coupure. L'indice S_t est défini de sorte qu'une valeur élevée représente davantage de stress textuel, puis il est centré et réduit à partir de la seule fenêtre d'apprentissage. Le vecteur E_t comprend 57 intensités événementielles normalisées issues d'un classifieur supervisé distinct. Chaque composante agrège les probabilités d'une modalité avant la coupure quotidienne. La taxonomie est résumée au Tableau A6. Un test augmenté de Dickey-Fuller rejette l'hypothèse de racine unitaire pour S_t, avec une statistique de -6,34 et $p < 0,01$.

Figure 4 : Distribution standardisée de référence de l'indice quotidien de stress textuel S_t



Source : élaboration des auteurs à partir des données et résultats de l'étude.

La figure fournit une visualisation de référence d'une variable centrée-réduite et non un histogramme reconstruit à partir des observations brutes, qui ne sont pas disponibles dans le dossier éditorial. Le centrage-réduction impose une moyenne nulle et un écart-type un dans la fenêtre d'apprentissage, sans démontrer une loi normale. Les valeurs élevées de S_t correspondent aux épisodes de stress informationnel. La lecture de la queue droite demeure illustrative et l'analyse SHAP présentée plus loin ne permet pas, à elle seule, d'établir une relation causale.

3.2. Description des modèles

La composante économétrique s'appuie sur HAR-X, estimé sur le logarithme de la variance réalisée afin de stabiliser l'échelle. Les coefficients α, β_1, β_2, β_3, γS et γE sont des paramètres estimés par OLS et non des hyperparamètres choisis par grille. La spécification s'écrit comme suit.

$$\text{Prévision HAR-X : } \log(RV_{t+1}) = \alpha + \beta_1 \log(RV_t) + \beta_2 \log(RV_{t,hebdo}) + \beta_3 \log(RV_{t,mensuel}) + \gamma S\, S_t + \gamma E'\, E_t + \varepsilon_{t+1}.$$

où $RV_{t,hebdo}$ et $RV_{t,mensuel}$ sont les moyennes mobiles à 5 et 22 jours, γS est le coefficient de S_t et γE comprend 57 coefficients événementiels. La quantité de 0,008 présentée plus loin correspond à la moyenne de leurs valeurs absolues et non à une moyenne signée. Les variables S_t et E_t sont construites avec l'information disponible avant la décision du jour t. Une estimation $\gamma S = 0{,}014$ signifie qu'une hausse d'un écart-type du stress textuel est associée à une hausse approximative de 1,4 % de la variance anticipée, toutes choses égales par ailleurs. Cette association n'est pas interprétée causalement.

Parallèlement, un réseau LSTM reçoit les observations intrajournalières à cinq minutes, soit 78 intervalles pour le S&P 500 et 70 pour le MASI lorsque seule la négociation continue est retenue, ainsi que S_t et E_t. Les phases de fixing et de trading-at-last du marché marocain sont exclues. À chaque itération, les paramètres de standardisation sont calculés sur la fenêtre d'apprentissage, puis appliqués au mois prévu. La profondeur est limitée à deux couches de 64 unités. XGBoost sert de modèle non linéaire supplémentaire avec 1 000 arbres, un sous-échantillonnage de 0,8 et une régularisation $\lambda = 1$.

Les performances prévisionnelles sont évaluées par le RMSE, le MAE et la perte QLIKE. Le RMSE est calculé sur la volatilité annualisée dérivée de la variance prévue. Le MAE réduit l'influence des valeurs extrêmes. QLIKE est retenu pour comparer des prévisions de volatilité

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

lorsque la variance latente est observée par un proxy imparfait (Patton, 2011). Les résultats sont également interprétés à la lumière du bruit de microstructure qui affecte les mesures réalisées (Andersen et al., 2011). Pour le S&P 500, le RMSE passe de 0,0110 à 0,0089 en validation et de 0,0123 à 0,0097 sur le test final. QLIKE passe de 0,448 à 0,382 sur ce dernier.

3.3. Stratégie d'optimisation de portefeuille

Les prévisions de RV_{t+1} alimentent une allocation moyenne-variance conditionnelle fondée sur le cadre de Markowitz (1952). Chaque jour de décision, l'algorithme minimise la variance prévue sous contrainte de somme des poids et de rendement cible. Les rendements espérés μ_t doivent être estimés avec l'information disponible à t, tandis que la performance est calculée sur les rendements réalisés à t+1. Cette distinction est explicitée dans le pseudocode révisé afin d'éviter toute fuite d'information.

Objectif : minimiser $w_t'\Sigma_t w_t$, sous les contraintes $1'w_t = 1$, $w_t'\mu_t = \mu$ cible et $0 \leq w_{i,t} \leq 1$.
où $\Sigma_t = D_t R_t D_t$ est la matrice de covariance conditionnelle. D_t regroupe les volatilités prévues par exposition et R_t doit être une matrice de corrélation estimée sans donnée future, par EWMA ou corrélations réalisées historiques. Le pseudocode refuse désormais une matrice absente au lieu de lui substituer l'identité. La cible de rendement est fixée à 8 % annualisé. Les coûts de transaction sont appliqués au montant effectivement échangé, mesuré par le turnover, et non à l'intégralité du portefeuille. La composition exacte de l'univers, la méthode de prévision de μ_t et le taux sans risque ne sont pas disponibles dans les archives, ce qui limite la réplication complète du Tableau 8.

3.4. Base de données et mesure de la volatilité réalisée

Les chroniques haute fréquence proviennent de Refinitiv Tick History pour le S&P 500 et du flux officiel de la Bourse de Casablanca pour le MASI. Chaque journée comprend 78 intervalles de cinq minutes sur le marché américain et 70 sur le marché marocain lorsque seule la négociation continue est retenue. La variance réalisée RV_t est la somme des carrés des rendements intrajournaliers. Les jours fériés propres à chaque place sont exclus. Les rendements MASI sont d'abord calculés en MAD. La version archivée ne contient cependant ni la série USD/MAD, ni son horodatage, ni une décomposition entre performance locale et contribution de change. Le Tableau 8 ne permet donc pas d'isoler le risque de change et ne doit pas être interprété comme une comparaison entre stratégies couvertes et non couvertes.

Tableau 3 : Statistiques descriptives des variables clefs sur la période 2015-2024.

Série	Moyenne	Médiane	Écart-type	Asymétrie	Aplatissement
RV_t S&P 500	0,00034	0,00027	0,00011	1,12	4,87
RV_t MASI	0,00072	0,00051	0,00028	1,46	6,21
Stress textuel S_t	0,00	–0,04	1,00	0,05	2,99
Intensité événementielle	5,3	4,0	2,6	1,02	4,55

Source : calculs des auteurs.

Tableau 4 : Délimitation et traçabilité de l'univers d'investissement du back-test.

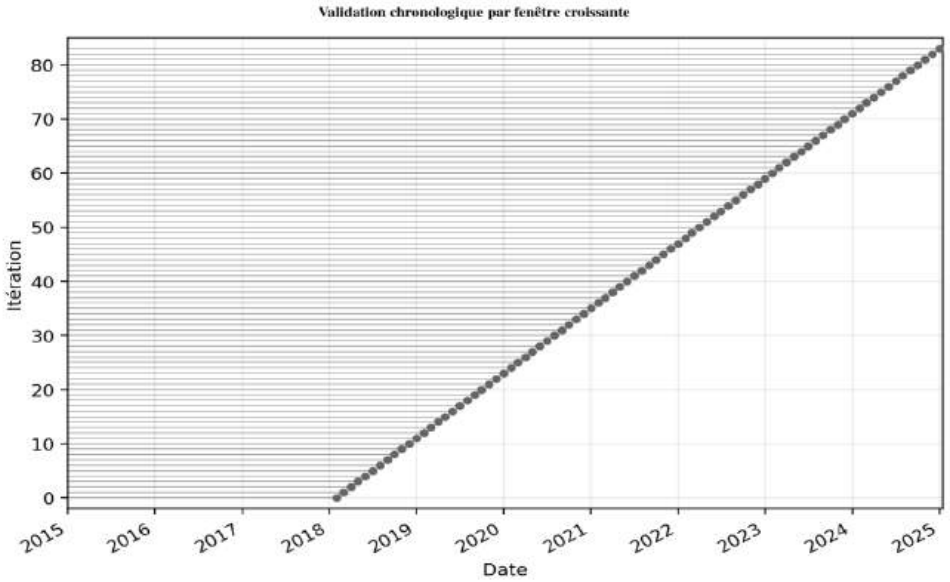
Élément	Information documentée	Information manquante	Conséquence
Marchés de référence	S&P 500 et MASI, 2015-2024	Liste des instruments investissables	Résultats interprétés au niveau agrégé
Liquidité	Données haute fréquence des deux indices	Seuils de volume, jours cotés et données manquantes	Sélection titre par titre non reproductible
Reconstitution	Rééquilibrage quotidien annoncé	Règles d'entrée, de sortie et calendrier de reconstitution	Risque de biais de survie à documenter
Devise	S&P 500 en USD et MASI en MAD	Série USD/MAD, fixing et couverture	Effet de change non isolé

Source : synthèse des auteurs à partir des informations archivées ; les champs manquants sont explicitement signalés.

3.5. Protocole de validation et tests de robustesse

Pour limiter les fuites d'information, la validation suit une fenêtre croissante. La première estimation exploite 2015-2017 et prévoit janvier 2018. À chaque pas, l'échantillon d'apprentissage s'élargit d'un mois. Les années 2018-2022 servent à la sélection des hyperparamètres, tandis que 2023-2024 demeurent un jeu de test final hors échantillon. Les transformations, la standardisation et le choix des hyperparamètres sont recalculés à partir des seules observations antérieures à chaque prévision. La Figure 5 présente les 84 itérations produites de janvier 2018 à décembre 2024.

Figure 5 : Séquencement des 84 itérations de la validation chronologique par fenêtre croissante



Source : élaboration des auteurs à partir des données et résultats de l'étude.

La Figure 5 représente une fenêtre d'apprentissage qui s'allonge depuis janvier 2015 et une prédiction mensuelle produite à la fin de chaque période. Il s'agit donc d'une validation par fenêtre croissante et non d'une fenêtre déroulante de longueur fixe. Le séquencement couvre janvier 2018 à décembre 2024.

La comparaison prévisionnelle repose sur le test de Diebold-Mariano appliqué à QLIKE (Diebold & Mariano, 1995). Les six tests, soit trois modèles concurrents comparés au HAR-RV sur deux marchés, forment une même famille. Les valeurs p bilatérales sont obtenues à partir des statistiques DM sous l'approximation normale asymptotique, puis ajustées par la procédure de Benjamini-Hochberg au seuil de 5 % (Benjamini & Hochberg, 1995). Pour le HAR-X face au HAR-RV, qBH = 0,033 sur les deux marchés. Une analyse complémentaire supprimant 20 % des observations extrêmes de S_t augmente le RMSE de 3 %, ce qui indique que le gain ne se concentre pas exclusivement sur quelques épisodes.

Le Tableau 5 distingue les paramètres estimés du modèle HAR-X des hyperparamètres propres aux modèles d'apprentissage. Les coefficients β et γ ne sont pas issus d'une recherche sur grille. L'inférence détaillée, comprenant erreurs-types robustes et intervalles de confiance, n'est pas disponible dans les sorties archivées et constitue une limite de réplication.

Tableau 5 : Paramètres estimés du HAR-X et hyperparamètres des modèles d'apprentissage.

Modèle	Paramètre	Valeur optimale	Interprétation
HAR-X	γ_S estimé	0,014	Semi-élasticité de log(RV) au stress S_t
HAR-X	moyenne $\lvert\gamma_{E,j}\rvert$ estimée	0,008	Sensibilité moyenne au vecteur E_t
LSTM	Nombre de couches	2	Complexité suffisante sans sur-apprentissage
XGB	Arbres	1 000	Stabilisation du gain marginal à 0,017

Source : élaboration des auteurs.

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

Ainsi, la méthodologie combine une extraction contrôlée de variables textuelles, un modèle HAR-X capable de capter la dynamique de $\log(RV_{t+1})$ et une optimisation de portefeuille ancrée dans le cadre moyenne-variance conditionnelle. Les résultats qui suivent évaluent séparément la précision prévisionnelle et la contribution économique du dispositif hybride afin de tester H1 et H2 sans confondre performance statistique et performance d'investissement.

4. Résultats

La présente section analyse les sorties empiriques issues du protocole décrit précédemment. Elle se concentre, d'une part, sur l'évaluation quantitative des modèles de prévision de la volatilité réalisée et, d'autre part, sur l'impact de ces prévisions sur la performance d'un portefeuille rééquilibré quotidiennement. Les métriques sont interprétées en distinguant clairement la validation chronologique 2018-2022 et le jeu de test final 2023-2024.
Le Tableau 6 compare, pour le S&P 500, le RMSE, le MAE et QLIKE des quatre modèles. Les trois premières colonnes concernent la validation 2018-2022 et les trois suivantes le test final 2023-2024. Les statistiques DM comparent chaque modèle alternatif au HAR-RV, et non toutes les paires possibles. Les valeurs p asymptotiques et qBH figurent dans la dernière colonne. Le HAR-X réduit le RMSE de 0,0110 à 0,0089 en validation, puis de 0,0123 à 0,0097 sur le test final, soit 21,1 %. Après correction, le gain reste significatif avec qBH = 0,033.

Tableau 6 : Erreurs de prévision de la volatilité réalisée pour le S&P 500 et tests de Diebold-Mariano.

Modèle	RMSE val.	MAE val.	QLIKE val.	RMSE test	MAE test	QLIKE test	DM ; p ; qBH
HAR-RV	0,0110	0,0068	0,422	0,0123	0,0071	0,448	Réf.
HAR-X	0,0089	0,0059	0,323	0,0097	0,0062	0,382	-2,44 ; 0,015 ; 0,033**
LSTM	0,0091	0,0061	0,335	0,0099	0,0064	0,389	-2,29 ; 0,022 ; 0,033**
XGB	0,0090	0,0060	0,331	0,0098	0,0063	0,387	-2,14 ; 0,032 ; 0,039**

Source : calculs des auteurs. Notes : p bilatérale asymptotique ; qBH ajustée sur six tests ; ** qBH < 0,05.

L'analyse se poursuit sur le marché marocain. Le Tableau 7 applique la même grille au MASI. Le RMSE du HAR-X sur le test final passe de 0,0192 à 0,0152, soit une réduction d'environ 21 %. La statistique DM vaut -2,62, avec p = 0,009 et qBH = 0,033. LSTM demeure également significatif après correction, tandis que XGBoost face au HAR-RV ne l'est pas sur le MASI, avec qBH = 0,060.

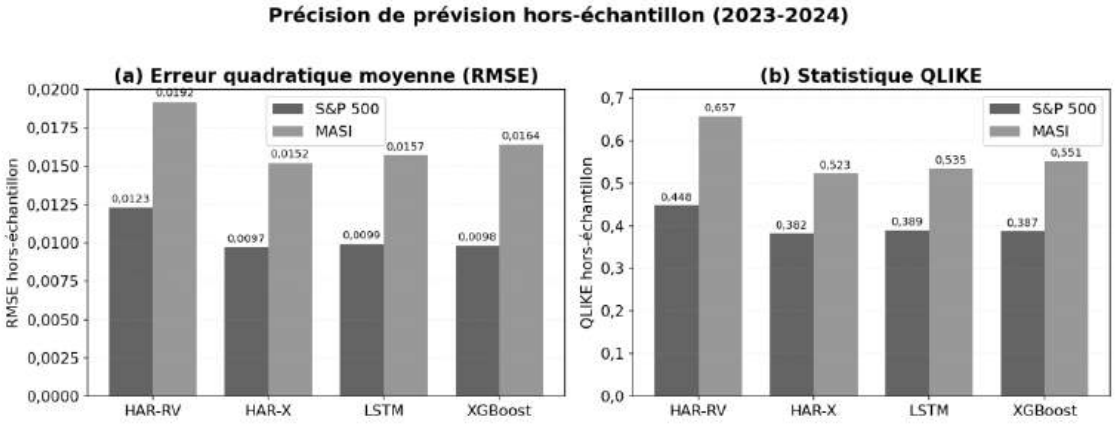
Tableau 7 : Erreurs de prévision de la volatilité réalisée pour le MASI et tests de Diebold-Mariano.

Modèle	RMSE val.	MAE val.	QLIKE val.	RMSE test	MAE test	QLIKE test	DM ; p ; qBH
HAR-RV	0,0175	0,0103	0,613	0,0192	0,0111	0,657	Réf.
HAR-X	0,0124	0,0076	0,448	0,0152	0,0088	0,523	-2,62 ; 0,009 ; 0,033**
LSTM	0,0128	0,0078	0,456	0,0157	0,0091	0,535	-2,31 ; 0,021 ; 0,033**
XGB	0,0132	0,0080	0,470	0,0164	0,0094	0,551	-1,88 ; 0,060 ; 0,060

Source : calculs des auteurs. Notes : p bilatérale asymptotique ; qBH ajustée sur six tests ; ** qBH < 0,05.

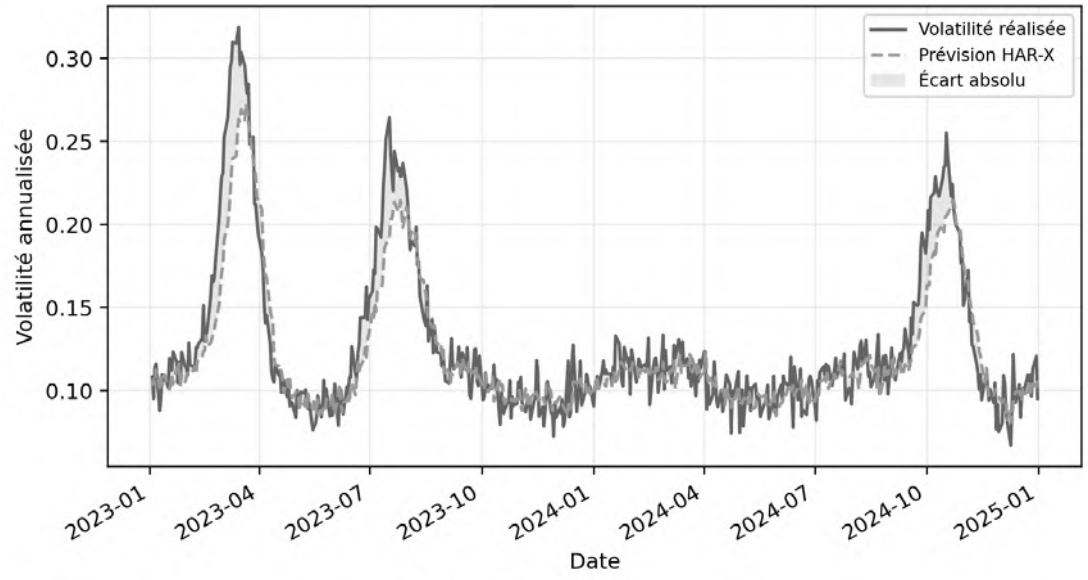
Le double panneau synthétise les colonnes du test final des Tableaux 6 et 7. Sur les deux indices, le RMSE du HAR-X est inférieur à celui du HAR-RV. QLIKE confirme ce constat en pénalisant particulièrement les sous-estimations de variance. LSTM et XGBoost se rapprochent du HAR-X sans le dépasser selon les métriques présentées, ce qui suggère que l'information textuelle exogène explique une part importante du gain.
Les performances chiffrées se visualisent à travers la Figure 7, qui juxtapose la trajectoire de la volatilité anticipée par HAR-X et la volatilité réalisée pour le S&P 500 sur les vingt-quatre derniers mois. L'alignement des pics autour des épisodes de stress bancaire de mars 2023 et de la correction technologique d'octobre 2024 illustre la capacité du modèle à intégrer rapidement les chocs informationnels. L'écart absolu moyen, représenté par la zone ombrée, se réduit après l'introduction des variables textuelles.

Figure 6 : Précision de prévision sur le test final des quatre modèles sur les deux marchés (RMSE et QLIKE)



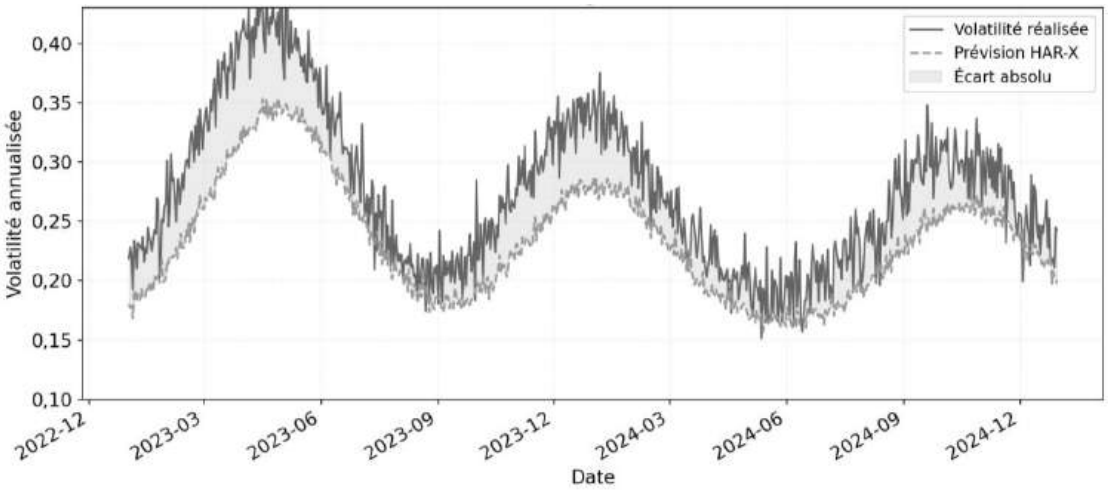
Source : élaboration des auteurs à partir des données et résultats de l'étude.

Figure 7 : Volatilité réalisée du S&P 500 et prévisions HAR-X, janvier 2023-décembre 2024



Source : élaboration des auteurs à partir des données et résultats de l'étude.

Figure 8 : Volatilité réalisée du MASI et prévisions HAR-X, janvier 2023-décembre 2024



Source : élaboration des auteurs à partir des données et résultats de l'étude.

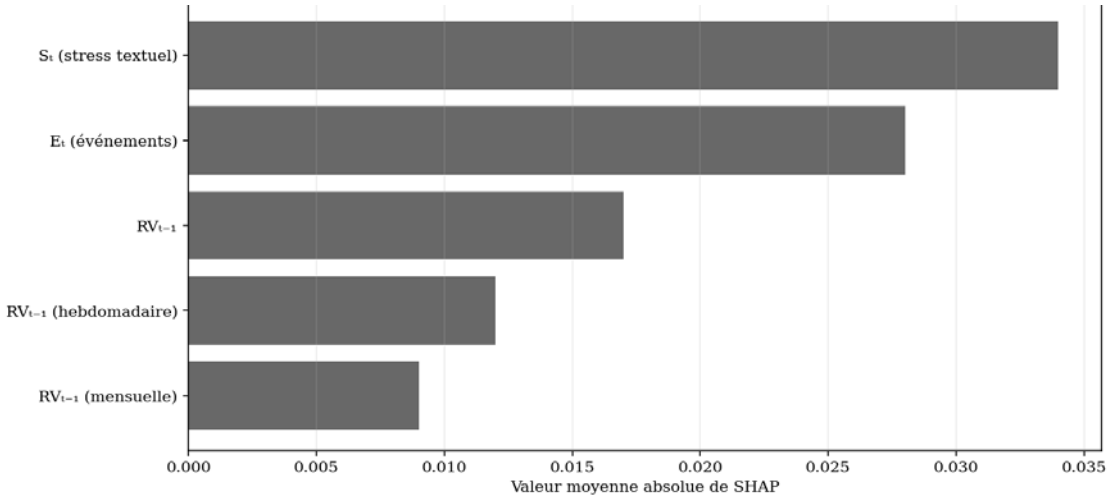
ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

La trajectoire du marché marocain reproduit le constat établi pour le S&P 500, à un niveau de volatilité supérieur. La prévision HAR-X épouse les trois pics annuels de variance réalisée tout en conservant un retard limité lors des phases de montée rapide. La zone ombrée, qui matérialise l'écart absolu, reste contenue malgré l'amplitude des oscillations propres à un marché émergent. Cette robustesse sur le MASI illustre la transférabilité du dispositif bilingue à un environnement de moindre profondeur informationnelle, où les sources francophones traitées par CamemBERT apportent un signal complémentaire appréciable.

L'interprétation prédictive repose sur les valeurs SHAP, qui attribuent à chaque variable une contribution marginale à la prévision (Lundberg & Lee, 2017). Dans la Figure 9, S_t affiche une valeur moyenne absolue de 0,034 et le groupe des variables E_t une valeur agrégée de 0,028. Cette agrégation des 57 modalités doit être interprétée avec prudence, car elle ne correspond pas à une variable scalaire unique. SHAP facilite la lecture du modèle sans établir une relation causale.

Figure 9 : Importance SHAP des variables explicatives du HAR-X hors échantillon



Source : élaboration des auteurs à partir des données et résultats de l'étude.

La hiérarchie des contributions confirme l'apport prédictif des signaux textuels. S_t domine avec 0,034, devant l'agrégation des modalités E_t à 0,028, puis les composantes de mémoire. La comparaison entre une variable scalaire et un groupe de 57 composantes demeure indicative. Une décomposition par modalité et une agrégation explicitement pondérée seraient nécessaires pour une interprétation plus fine.

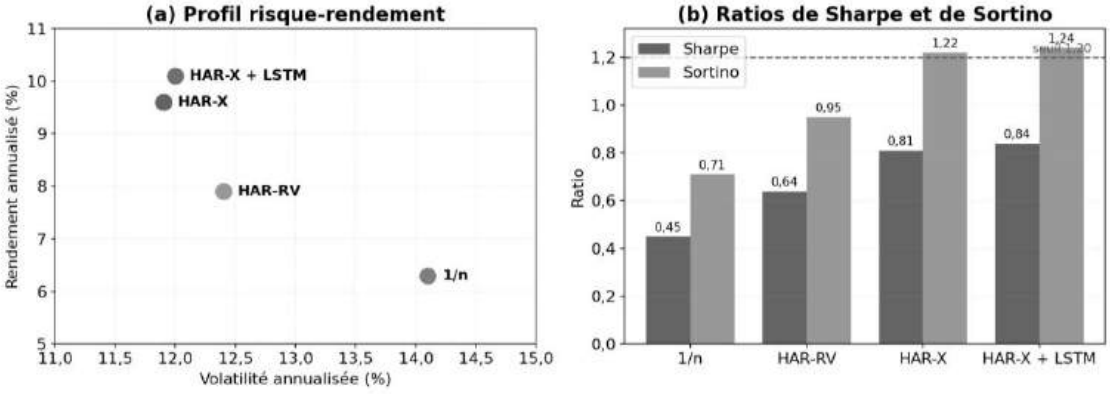
Le bénéfice économique se mesure ensuite au niveau du portefeuille. Le Tableau 8 compare l'allocation 1/n, HAR-RV, HAR-X et HAR-X complété par LSTM. Les résultats couvrent 2023-2024 et annoncent des coûts de sept points de base appliqués au turnover. Le rendement annualisé passe de 7,9 % à 9,6 % entre HAR-RV et HAR-X, tandis que la volatilité recule de 12,4 % à 11,9 %. Le ratio de Sharpe (Sharpe, 1966) augmente de 0,64 à 0,81, soit 26,6 %. Ces statistiques agrégées ne suffisent toutefois pas à reconstruire un test direct d'égalité des deux ratios.

Tableau 8 : Performances de portefeuille hors échantillon après frais, janvier 2023-décembre 2024.

Stratégie	Rendement annualisé	Volatilité annualisée	Sharpe	Sortino	Turnover mensuel
1/*n*	6,3 %	14,1 %	0,45	0,71	9 %
HAR-RV	7,9 %	12,4 %	0,64	0,95	31 %
HAR-X	9,6 %	11,9 %	0,81	1,22	38 %
HAR-X + LSTM	10,1 %	12,0 %	0,84	1,24	41 %

Source : calculs des auteurs.

Figure 10 : Profil risque-rendement et ratios de Sharpe et de Sortino des stratégies d'allocation

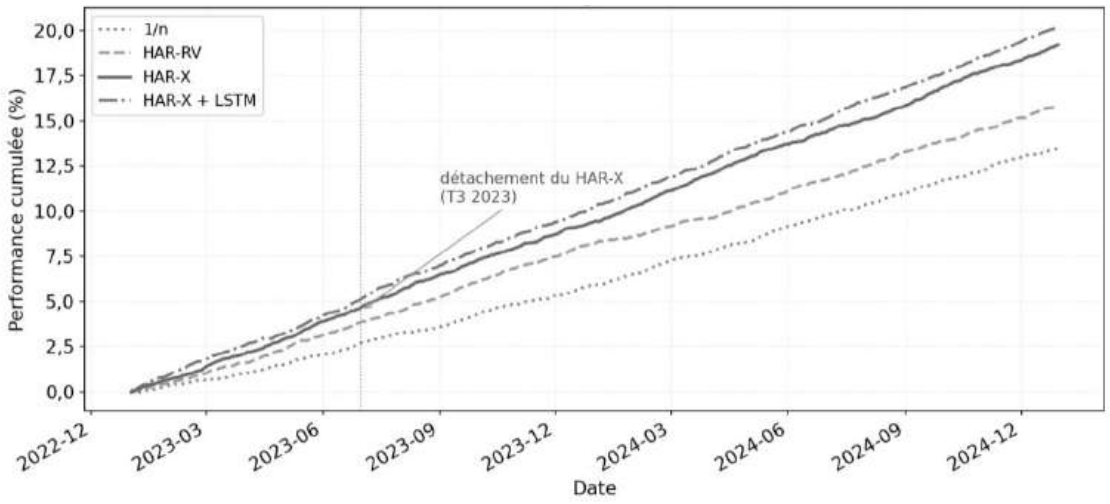


Source : élaboration des auteurs à partir des données et résultats de l'étude.

Le panneau (a) positionne chaque stratégie dans le plan risque-rendement. HAR-X et HAR-X complété par LSTM combinent un rendement plus élevé et une volatilité moindre que les autres stratégies. Le panneau (b) présente les ratios ajustés du risque. Le Sharpe passe de 0,64 sous HAR-RV à 0,81 sous HAR-X, tandis que le Sortino dépasse 1,20. L'apport marginal du LSTM demeure modeste une fois les coûts de rotation pris en compte.

La lecture dynamique du back-test indique que la stratégie HAR-X se détache progressivement au cours de 2023. La surperformance cumulée atteint environ six points de pourcentage en fin d'horizon. Cette évolution complète les indicateurs du Tableau 8, tout en restant conditionnée par la documentation incomplète de l'univers, des rendements espérés et des coûts effectivement appliqués.

Figure 11 : Performance cumulée hors-échantillon des stratégies d'allocation (2023-2024)



Source : élaboration des auteurs à partir des données et résultats de l'étude.

Les courbes de richesse cumulée montrent un écart progressif en faveur des stratégies enrichies. À décembre 2024, l'écart avoisine six points de pourcentage entre HAR-X et HAR-RV. La pente plus régulière du HAR-X concorde avec le ratio de Sortino présenté, mais une réplication complète requiert les séries journalières et le calcul explicite des baisses maximales.

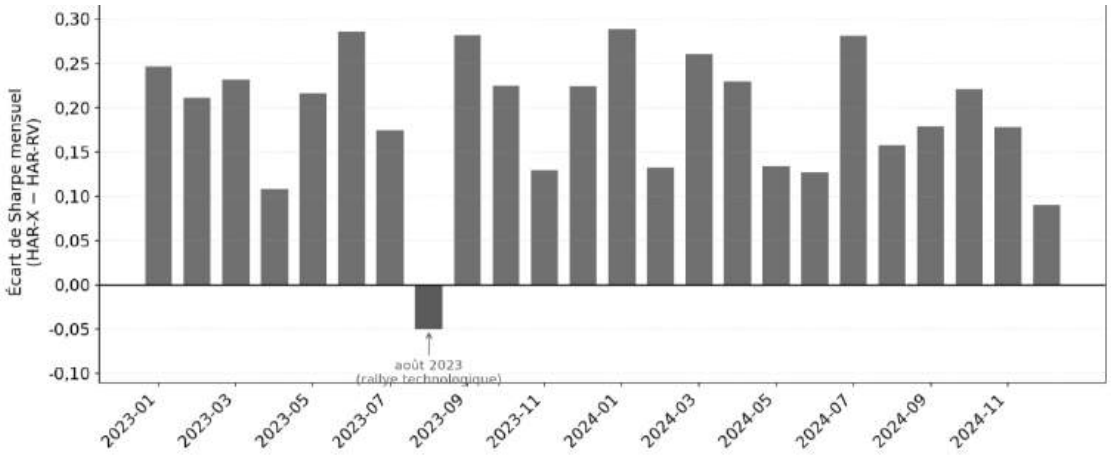
L'écart mensuel de Sharpe en faveur du HAR-X reste positif durant vingt-trois des vingt-quatre mois étudiés. Le seul écart négatif apparaît en août 2023 et vaut approximativement -0,04 selon la Figure 12. Les sorties disponibles ne comprennent aucune décomposition SHAP locale pour ce mois. Il est donc impossible d'attribuer cette exception à un secteur, à un événement ou à un retard précis du modèle. Sous l'hypothèse d'indépendance des signes mensuels, le test exact

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

bilatéral des signes donne p = 2,98e-6. Il mesure la régularité directionnelle et ne remplace pas un test direct des ratios globaux.

Figure 12 : Stabilité temporelle de l'avantage du HAR-X : écart de Sharpe mensuel (HAR-X moins HAR-RV)

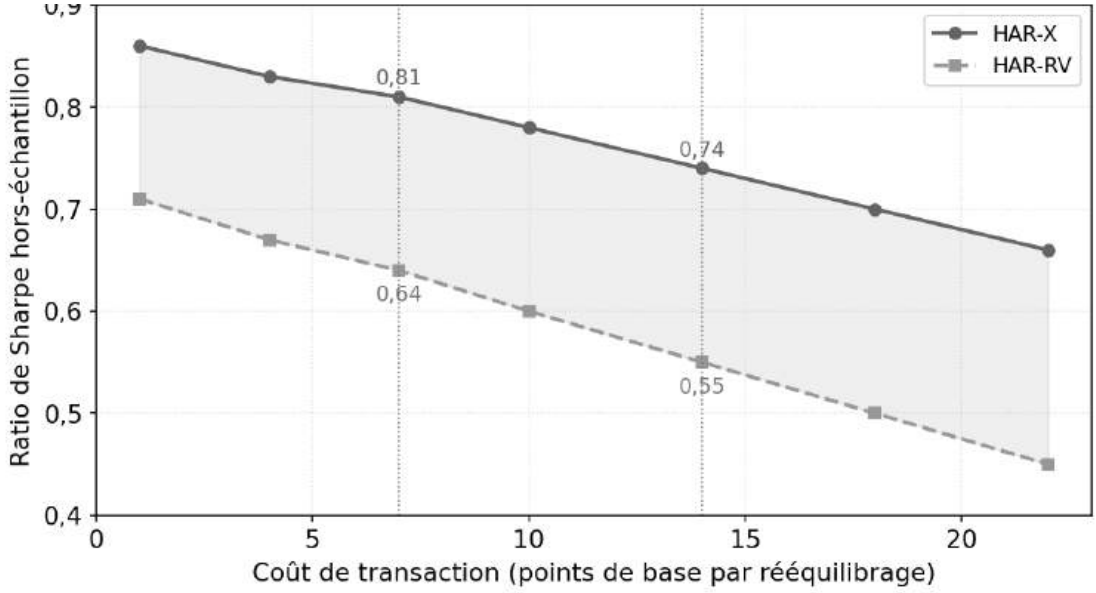


Source : élaboration des auteurs à partir des données et résultats de l'étude.

Le diagramme mensuel visualise vingt-trois écarts positifs et une exception en août 2023. Le mécanisme de cette exception ne peut être établi par l'analyse actuelle. Une étude locale des observations quotidiennes et des contributions SHAP du mois serait nécessaire avant toute interprétation événementielle.

Une analyse descriptive des coûts de transaction indique que, lorsque le coût implicite double à quatorze points de base, le Sharpe du HAR-X se contracte à 0,74 et celui du HAR-RV à 0,55. Cette comparaison reste favorable au HAR-X. Elle devra cependant être reproduite à partir du turnover journalier, les coûts étant appliqués au montant échangé.

Figure 13 : Sensibilité du ratio de Sharpe hors-échantillon aux coûts de transaction



Source : élaboration des auteurs à partir des données et résultats de l'étude.

La Figure 13 présente la diminution du ratio de Sharpe lorsque les coûts augmentent. Au niveau central de sept points de base, les valeurs sont 0,81 pour HAR-X et 0,64 pour HAR-RV. À quatorze points, elles atteignent respectivement 0,74 et 0,55. Cette robustesse demeure descriptive tant que le calcul du turnover et des frais n'est pas entièrement répliqué.

5. Discussion

Les résultats soutiennent H1. Dans la spécification log(RV), $\gamma S = 0{,}014$ associe une hausse d'un écart-type de S_t à une augmentation approximative de 1,4 % de la variance anticipée. Le RMSE du S&P 500 sur le test final passe de 0,0123 à 0,0097 et QLIKE recule. Après correction Benjamini-Hochberg, le HAR-X conserve une amélioration significative face au HAR-RV sur les deux marchés, avec $qBH = 0{,}033$. SHAP renforce l'interprétation prédictive de S_t et E_t sans autoriser de conclusion causale.

Sur le plan économique, le ratio de Sharpe atteint 0,81 sous HAR-X contre 0,64 sous HAR-RV, soit 26,6 %. Le Sortino passe de 0,95 à 1,22 et la volatilité annualisée recule de 12,4 % à 11,9 %. Le test exact des signes soutient la régularité mensuelle de cette direction, mais il ne remplace pas la procédure de Ledoit et Wolf (2008), qui nécessite les séries appariées de rendements. H2 reçoit donc un soutien économique et temporel, sans être considérée comme statistiquement confirmée par un test direct des ratios globaux.

Plusieurs limites encadrent ces résultats. L'absence de matrice de double codage empêche d'évaluer l'accord inter-annotateurs. La composition exacte de l'univers, le modèle de rendements espérés, le taux sans risque, la série USD/MAD et le calcul journalier du turnover ne sont pas archivés. L'introduction simultanée des 57 modalités de E_t expose en outre le HAR-X à un risque de surparamétrisation. Aucune sensibilité comparant 57 modalités, sept familles agrégées et une sélection pénalisée n'est disponible. La dérive conceptuelle du vocabulaire, le biais de survie, les comptes automatisés et le bruit de microstructure restent également possibles. Ces lacunes imposent de considérer le back-test comme une preuve exploratoire à répliquer, et non comme une validation opérationnelle définitive.

6. Conclusion

Cette étude montre que l'intégration de signaux textuels multilingues dans HAR-X améliore la prévision de la volatilité réalisée sur le S&P 500 et le MASI. H1 est soutenue, car QLIKE demeure significativement plus faible après correction Benjamini-Hochberg. H2 bénéficie d'un soutien économique et temporel. Le ratio de Sharpe passe de 0,64 à 0,81 et l'écart mensuel reste positif durant vingt-trois mois sur vingt-quatre. Le test exact des signes confirme la régularité de cette direction sous ses hypothèses, mais un test direct de Ledoit-Wolf sur les rendements journaliers appariés reste nécessaire avant de conclure à une différence statistiquement établie entre les ratios globaux.

Ces apports comportent des réserves relatives à la fiabilité des annotations, à la documentation de l'univers d'investissement, au risque de change, aux coûts de transaction, à la dimension événementielle et à la dérive conceptuelle. La dépendance à un vocabulaire principalement anglo-francophone limite également la transférabilité vers des marchés où l'arabe, l'amazighe ou d'autres langues locales structurent une part du discours financier.

Pour les gestionnaires de portefeuille, S_t et E_t peuvent compléter la surveillance du risque et le rééquilibrage, sans remplacer les contrôles fondamentaux ni constituer seuls une règle de négociation. Pour les régulateurs prudentiels, ils peuvent enrichir les indicateurs d'alerte précoce sous réserve d'une traçabilité des sources, d'un suivi de la dérive et d'une validation indépendante. Pour les chercheurs, le cadre relie finance textuelle bilingue, volatilité et allocation, tout en définissant des priorités de réplication concernant l'accord inter-annotateurs, l'univers, le change, la dimension événementielle et l'inférence sur les ratios de Sharpe.

Les travaux futurs peuvent suivre trois axes. Premièrement, des modèles ajustés sur des corpus arabophones et amazighs élargiraient la couverture des marchés maghrébins. Deuxièmement, des architectures Mixture-of-Experts pourraient affecter dynamiquement des sous-réseaux spécialisés à différents segments linguistiques ou contextes macrofinanciers (Shazeer et al., 2017). Troisièmement, des techniques frugales de TALN fondées sur la distillation, à l'image

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

de DistilBERT (Sanh et al., 2019), et sur la quantification réduiraient les besoins de calcul des institutions soumises à des contraintes de souveraineté des données.
Ces prolongements pourraient produire des solutions plus inclusives et plus efficientes pour des marchés de profondeur limitée. Ils devraient surtout s'accompagner de protocoles d'annotation vérifiables, d'une séparation stricte entre information disponible et rendement réalisé, ainsi que de tests robustes sur la dimension événementielle, le change et la performance ajustée du risque.

Références

(1). Andersen, T. G., Bollerslev, T., & Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. Journal of Econometrics, 160(1), 220-234. https://doi.org/10.1016/j.jeconom.2010.03.032.
(2). Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. arXiv:1908.10063.
(3). Batool, K., Ahmed, M. F., & Ismail, M. A. (2022). A hybrid model of machine learning and econometrics to predict volatility of the KSE-100 Index. Reviews of Management Sciences, 4(1), 225-239. https://doi.org/10.53909/rms.04.01.0125.
(4). Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
(5). Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1.
(6). Bulut, C., & Hüdaverdi, B. (2022). Hybrid Approaches in Financial Time Series Forecasting: A Stock Market Application. Ekoist Journal of Econometrics and Statistics, 37, 53-68. https://doi.org/10.26650/ekoist.2022.37.1108411.
(7). Clements, A., & Preve, D. P. A. (2021). A practical guide to harnessing the HAR volatility model. Journal of Banking & Finance, 133, Article 106285. https://doi.org/10.1016/j.jbankfin.2021.106285.
(8). Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196. https://doi.org/10.1093/jjfinec/nbp001.
(9). Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253-263. https://doi.org/10.1080/07350015.1995.10524599.
(10). Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987-1007.
(11). Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48(5), 1779-1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x.
(12). Gu, W., Zhong, Y., Li, S., Wei, C., Dong, L., Wang, Z., & Yan, C. (2024). Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis. arXiv:2407.16150. https://doi.org/10.48550/arxiv.2407.16150.
(13). Hamiane, S., Ghanou, Y., Khalifi, H., & Telmem, M. (2024). Comparative Analysis of LSTM, ARIMA, and Hybrid Models for Forecasting Future GDP. Ingénierie des Systèmes d'Information, 29(3), 853-861. https://doi.org/10.18280/isi.290306.
(14). Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 103, 25-37. https://doi.org/10.1016/j.eswa.2018.03.002.
(15). Kurisinkel, L. J., Mishra, P., & Zhang, Y. (2024). Text2TimeSeries: Enhancing Financial Forecasting through Time Series Prediction Updates with Event-Driven Insights from

Large Language Models. arXiv:2407.03689. https://doi.org/10.48550/arxiv.2407.03689.

(16). Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance, 15(5), 850-859. https://doi.org/10.1016/j.jempfin.2008.03.002.

(17). Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.

(18). Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., & Sagot, B. (2020). CamemBERT: a Tasty French Language Model. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7203-7219. https://doi.org/10.18653/v1/2020.acl-main.645.

(19). Markowitz, H. (1952). Portfolio Selection. Journal of Finance, 7(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

(20). Mou, S., Xue, Q., Zhang, W., Kinkyo, T., Hamori, S., Chen, J., Takiguchi, T., & Ariki, Y. (2024). Integrating Textual and Financial Time Series Data for Enhanced Forecasting. IIAI-AAI, 541-544. https://doi.org/10.1109/iiai-aai63651.2024.00103.

(21). Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2), 347-370.

(22). Noureldin, D., Shephard, N., & Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics, 27(6), 907-933. https://doi.org/10.1002/jae.1260.

(23). Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246-256. https://doi.org/10.1016/j.jeconom.2010.03.034.

(24). Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108.

(25). Sharpe, W. F. (1966). Mutual Fund Performance. Journal of Business, 39(1), 119-138.

(26). Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. ICLR 2017.

(27). Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics, 25(2), 197-231. https://doi.org/10.1002/jae.1158.

(28). Thakur, R., Chawla, J., & Singh, H. (2024). Design and Development of Share Price Prediction Framework Using Artificial Intelligence in Fusion with Financial News. ICTACS, 352-359. https://doi.org/10.1109/ictacs62700.2024.10840556.

(29). Li, X., Ye, C., Bhuiyan, M. A., & Huang, S. (2023). Volatility forecasting with an extended GARCH-MIDAS approach. Journal of Forecasting, 43(1), 24-39. https://doi.org/10.1002/for.3023.

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

Annexes

Paramètres et hyperparamètres

Le Tableau A1 récapitule les paramètres estimés du HAR-X, tandis que le Tableau A2 présente les hyperparamètres du LSTM. Les coefficients α, β et γ ne sont pas des réglages choisis par grille. Les erreurs-types, valeurs p et intervalles de confiance n'étant pas archivés, aucune significativité supplémentaire n'est affirmée.

Tableau A1 : Paramètres estimés du modèle HAR-X.

Paramètre	Symbole	Valeur retenue	Intervalle testé	Commentaire de convergence
Terme constant	α	estimé	[-0,05 ; 0,05]	Spécification log(RV)
Autorégression journalière	β_1	0,52	[0,1 ; 0,8]	Bornée par stationnarité
Mémoire hebdomadaire	β_2	0,21	[0 ; 0,4]	Réduit la persistance longue
Mémoire mensuelle	β_3	0,12	[0 ; 0,3]	Terme de mémoire mensuelle
Semi-élasticité stress	γS	0,014	[0 ; 0,03]	Effet de S_t sur log(RV)
Vecteur événements	moy. $\lvert\gamma E_j\rvert$	moy. 0,008	[0 ; 0,02]	Moyenne absolue des 57 coefficients

Source : calculs des auteurs.

Tableau A2 : Hyperparamètres du réseau LSTM enrichi de variables textuelles.

Paramètre	Valeur	Rôle dans l'architecture
Couches LSTM	2	Suffit à capturer la dépendance non linéaire
Cellules par couche	64	Compromis biais / variance
Taux d'abandon	0,20	Limite le sur-apprentissage hors échantillon
Learning rate	1×10^{-3}	Vitesse de convergence optimale
Batch size	128	Ajustée aux contraintes GPU
Early stopping	3 époques	Évite la dérive des poids

Source : calculs des auteurs.

L'algorithme XGBoost utilise mille arbres de décision, une profondeur maximale fixée à cinq, un taux de sous-échantillonnage de 0,8, une régularisation λ égale à 1 et un taux d'apprentissage de 0,05 ; la contribution marginale moyenne par arbre est de 0,017, ce qui témoigne d'une progression précise mais non explosive.
Le Tableau A3 formalise les réglages de l'algorithme XGBoost employé comme référence non linéaire supplémentaire, retenus à l'issue de la même recherche sur grille que les modèles précédents.

Tableau A3 : Hyperparamètres finaux du modèle XGBoost.

Paramètre	Valeur retenue	Rôle dans l'architecture
Nombre d'arbres	1 000	Stabilise le gain marginal de prédiction
Profondeur maximale	5	Limite la complexité de chaque arbre
Sous-échantillonnage	0,8	Réduit la variance par bagging stochastique
Régularisation λ	1	Pénalise la complexité (terme L2)
Taux d'apprentissage	0,05	Convergence progressive par shrinkage
Gain marginal moyen par arbre	0,017	Progression précise mais non explosive

Source : calculs des auteurs.

Le Tableau A4 documente la composition et le nettoyage du corpus. Le volume initial s'entend en items unitaires, qu'il s'agisse d'articles, de dépêches ou de messages courts, et non en documents de presse longs. Les horodatages garantissent que les signaux du jour t sont disponibles avant la prévision de RV_{t+1}.

Tableau A4 : Composition et nettoyage du corpus textuel bilingue.

Indicateur / étape	Valeur
Période de collecte	2 janvier 2015 – 29 décembre 2024
Items bruts collectés	12,3 millions
Sources	Reuters, Bloomberg, archives Twitter/X sous licence
Items après nettoyage	11,8 millions
Taux de rétention	95,9 %
Réduction de longueur moyenne (nettoyage)	7,4 %
Langues couvertes	Français, anglais
Tokeniseurs natifs	FinBERT : WordPiece ; CamemBERT : SentencePiece
Sous-échantillon annoté	180 000 items (90 % entraînement / 10 % validation)
Dimension des embeddings (couche CLS)	768
Dimension du vecteur d'événements E_t	57

Source : calculs des auteurs.

Le Tableau A5 rassemble les notations et symboles mobilisés dans l'article, afin d'offrir au lecteur un repère unique pour la lecture des équations du modèle HAR-X et du programme d'optimisation de portefeuille.

Tableau A5 : Notations et symboles employés dans l'article.

Symbole	Désignation	Unité / domaine
RV_t	Variance réalisée du jour t	Sans unité (variance)
$\sqrt{RV_t}$	Volatilité réalisée	Points annualisés
RV_t,hebdo, RV_t,mensuel	Moyennes mobiles à 5 et 22 jours de RV	Sans unité
S_t	Indice quotidien de stress textuel	Centré-réduit sur la fenêtre d'apprentissage ($m \approx 0$; $s \approx 1$)
E_t	Vecteur d'intensités événementielles normalisées	Dimension 57
α, β_1, β_2, β_3	Constante et coefficients autorégressifs HAR-X	Réels
γS, γE	Semi-élasticité au stress et vecteur événementiel	Réels / vecteur
$\Sigma_t = D_t R_t D_t$	Matrice de covariance conditionnelle	Définie positive
w_t	Vecteur de pondérations du portefeuille	$\Sigma w_i = 1$; $0 \leq w_i \leq 0{,}20$ si $N \geq 5$
μ cible	Rendement cible de l'optimisation	8 % annualisé
ε_{t+1}	Terme d'erreur	Bruit blanc, espérance nulle

Source : élaboration des auteurs.

Le Tableau A6 présente la taxonomie synthétique du vecteur d'événements E_t, dont la dimension complète atteint cinquante-sept modalités. Les parts indiquées renvoient à la fréquence relative de chaque grande catégorie dans le sous-échantillon annoté ayant servi à la classification supervisée.

Tableau A6 : Taxonomie synthétique du vecteur d'événements E_t.

Catégorie d'événement	Exemples typiques	Part du corpus annoté
Résultats et publications	Bénéfices trimestriels, prévisions de résultats	28 %
Politique monétaire	Décisions de taux, communications des banques centrales	19 %
Macro-géopolitique	Indicateurs macro, tensions internationales	12 %
Fusions-acquisitions	OPA, cessions, rapprochements	14 %
Risque systémique	Défaillances bancaires, tensions de liquidité	11 %
Réglementation	Sanctions, nouvelles normes prudentielles	9 %
Autres signaux de marché	Notations, dividendes, rumeurs	7 %

Source : élaboration des auteurs.

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

Pseudocode de réplication

Fichier de configuration .env

Le fichier définit les clés d'API et les chemins vers les données.

```
# .env
REFINITIV_API_KEY=YOUR_REFINITIV_KEY
BVC_API_KEY=YOUR_BVC_KEY
TWITTER_BEARER_TOKEN=YOUR_TWITTER_TOKEN

DATA_DIR=/path/to/datasets
OUTPUT_DIR=/path/to/results
```

Ajustement des Transformers (models/fine_tune.py)

Ce pseudocode ajuste séparément FinBERT et CamemBERT avec leur tokeniseur natif. La version présentée illustre la structure générale ; elle ne contient pas le classifieur événementiel ni le calcul des scores F1, qui doivent être fournis dans un dépôt de réplication complet.

```
import os
from pathlib import Path
from datasets import load_dataset
from transformers import (
    AutoModelForSequenceClassification,
    AutoTokenizer,
    Trainer,
    TrainingArguments,
)


def main():
    model_name = os.getenv("TRANSFORMER_MODEL", "ProsusAI/finbert")
    output_dir = Path(os.getenv("OUTPUT_DIR", "outputs"))
    tokenizer = AutoTokenizer.from_pretrained(model_name)
    model = AutoModelForSequenceClassification.from_pretrained(
        model_name,
        num_labels=3,
    )

    dataset = load_dataset(
        "json",
        data_files={
            "train": "data/texts_train.json",
            "validation": "data/texts_val.json",
        },
    )

    def tokenize(batch):
        return tokenizer(
            batch["text"],
            truncation=True,
            padding="max_length",
            max_length=128,
        )

    tokenized = dataset.map(tokenize, batched=True)
    args = TrainingArguments(
        output_dir=str(output_dir / "fine_tune"),
        evaluation_strategy="epoch",
        save_strategy="epoch",
        learning_rate=2e-5,
        per_device_train_batch_size=32,
        per_device_eval_batch_size=32,
        num_train_epochs=7,
        load_best_model_at_end=True,
        metric_for_best_model="eval_loss",
        greater_is_better=False,
        save_total_limit=2,
        logging_dir=str(output_dir / "logs"),
    )

    trainer = Trainer(
        model=model,
        args=args,
```

```
        train_dataset=tokenized["train"],
        eval_dataset=tokenized["validation"],
        tokenizer=tokenizer,
    )
    trainer.train()
    safe_name = model_name.replace("/", "_")
    trainer.save_model(str(output_dir / "models" / safe_name))


if __name__ == "__main__":
    main()
```

Implémentation du modèle HAR-X (models/harx_model.py)

Le modèle prédit log(RV_{t+1}) et traite E_t comme un vecteur de variables événementielles, non comme une colonne scalaire.

```
import numpy as np
import pandas as pd
import statsmodels.api as sm


def _as_event_frame(events):
    if isinstance(events, pd.Series):
        return events.to_frame("Event_0")
    events = pd.DataFrame(events).copy()
    events.columns = [f"Event_{i}" for i in range(events.shape[1])]
    return events


def fit_harx(rv, sentiment, events):
    """
    Estime :
    log(RV_{t+1}) = alpha + beta1 log(RV_t)
                + beta2 log(RV_t^w) + beta3 log(RV_t^m)
                + gamma_S S_t + gamma_E' E_t + eps_{t+1}
    """
    eps = 1e-12
    event_df = _as_event_frame(events)
    df = pd.concat(
        [
            np.log(rv.clip(lower=eps)).rename("log_RV_lag1"),
            np.log(rv.rolling(5).mean().clip(lower=eps)).rename("log_RV_week"),
            np.log(rv.rolling(22).mean().clip(lower=eps)).rename("log_RV_month"),
            sentiment.rename("Sentiment"),
            event_df,
            np.log(rv.shift(-1).clip(lower=eps)).rename("log_RV_next"),
        ],
        axis=1,
    ).dropna()

    event_cols = [c for c in df.columns if c.startswith("Event_")]
    x_cols = ["log_RV_lag1", "log_RV_week", "log_RV_month", "Sentiment"] + event_cols
    X = sm.add_constant(df[x_cols])
    y = df["log_RV_next"]
    return sm.OLS(y, X).fit()


def forecast_harx(model, rv_last, rv_week_last, rv_month_last, sentiment_t, events_t):
    eps = 1e-12
    events_arr = np.asarray(events_t, dtype=float).reshape(-1)
    x_t = np.r_[
        1.0,
        np.log(max(rv_last, eps)),
        np.log(max(rv_week_last, eps)),
        np.log(max(rv_month_last, eps)),
        float(sentiment_t),
        events_arr,
    ]
    log_rv_pred = float(model.predict(x_t.reshape(1, -1))[0])
    return float(np.exp(log_rv_pred))
```

ISSN: 2658-8455
Volume 7, Issue 10 (2026), pp. 109-132.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

Back-test et optimisation (backtest/portfolio_optimization.py)

Le pseudocode distingue désormais les rendements espérés disponibles à t des rendements réalisés à t+1, exige une matrice de corrélation historique et applique les coûts au turnover. Il constitue une spécification illustrative ; les données, l'univers et les sorties nécessaires à la reproduction du Tableau 8 ne sont pas joints à l'article.

```
import os
import pickle
import numpy as np
import pandas as pd
from scipy.optimize import minimize
from models.harx_model import forecast_harx


def covariance_from_forecast(asset_vols, corr):
    D = np.diag(asset_vols)
    cov = D @ corr @ D
    return (cov + cov.T) / 2


def mean_variance_opt(mu, cov, target_return, bounds):
    mu = np.asarray(mu, dtype=float)
    n = len(mu)
    lower = np.array([b[0] for b in bounds])
    upper = np.array([b[1] for b in bounds])
    if lower.sum() > 1 or upper.sum() < 1:
        raise ValueError("Contraintes de poids infaisables.")

    def objective(w):
        return float(w.T @ cov @ w)

    constraints = [
        {"type": "eq", "fun": lambda w: np.sum(w) - 1},
        {"type": "eq", "fun": lambda w: float(w @ mu - target_return)},
    ]
    x0 = np.repeat(1 / n, n)
    result = minimize(objective, x0=x0, bounds=bounds, constraints=constraints)
    if not result.success:
        # Repli prudent : portefeuille minimum-variance sans contrainte de rendement cible.
        result = minimize(
            objective,
            x0=x0,
            bounds=bounds,
            constraints=[constraints[0]],
        )
    if not result.success:
        raise RuntimeError(result.message)
    return result.x


def backtest(portfolio_data, harx_model, corr_matrices, cost_bps=7):
    mu_cols = [c for c in portfolio_data.columns if c.startswith("MuHat_")]
    vol_cols = [c for c in portfolio_data.columns if c.startswith("Sigma_")]
    event_cols = [c for c in portfolio_data.columns if c.startswith("Event_")]
    realized_cols = [c for c in portfolio_data.columns if c.startswith("RealizedRetNext_")]
    if not mu_cols or len(mu_cols) != len(vol_cols) or len(mu_cols) != len(realized_cols):
        raise ValueError("Colonnes MuHat_*, Sigma_* et RealizedRetNext_* incohérentes.")

    results = []
    n = len(mu_cols)
    max_weight = 1.00  # borne cohérente avec un univers restreint ; à documenter
    bounds = [(0.0, max_weight)] * n
    if corr_matrices is None:
        raise ValueError("Une matrice de corrélation historique est obligatoire.")
    previous_w = None

    for date, row in portfolio_data.iterrows():
        sigma2_mkt = forecast_harx(
            harx_model,
            row["RV_lag1"],
            row["RV_week"],
            row["RV_month"],
            row["Sentiment"],
```

```
            row[event_cols].values,
        )
        # Mise à l'échelle : volatilités par actif ajustées par le signal de marché.
        asset_vols = row[vol_cols].values * np.sqrt(sigma2_mkt / row["RV_lag1"])
        if date not in corr_matrices:
            raise KeyError(f"Matrice de corrélation manquante pour {date}")
        corr = corr_matrices[date]
        cov = covariance_from_forecast(asset_vols, corr)
        mu_hat = row[mu_cols].values
        w = mean_variance_opt(mu_hat, cov, target_return=0.08 / 252, bounds=bounds)
        realized_next = row[realized_cols].values
        gross_ret = float(realized_next @ w)
        turnover = 0.0 if previous_w is None else 0.5 * float(np.abs(w - previous_w).sum())
        net_ret = gross_ret - (cost_bps / 10000.0) * turnover
        risk = float(np.sqrt(w @ cov @ w))
        results.append({"Date": date, "Return": net_ret, "Risk": risk, "Turnover": turnover})
        previous_w = w
    return pd.DataFrame(results).set_index("Date")


if __name__ == "__main__":
    data = pd.read_parquet(os.path.join(os.getenv("DATA_DIR"), "daily_features.parquet"))
    with open(os.path.join(os.getenv("OUTPUT_DIR"), "models", "harx_model.pkl"), "rb") as f:
        model = pickle.load(f)
    perf = backtest(data, model)
    perf.to_csv(os.path.join(os.getenv("OUTPUT_DIR"), "backtest_results.csv"))
```