

HEC MONTRÉAL

**Impact de la Fréquence des Nouvelles sur la Distribution des Rendements Boursiers
(Skewness, Kurtosis, Volatilité, Rendement Moyen Arithmétique)**

Par
Kévin-Christopher Tiacoh

Sous la direction de
Martin Boyer, Ph. D.

**Sciences de la gestion
(Option Finance)**

**Mémoire présenté en vue de l'obtention du grade de maîtrise des sciences en gestion
(M. Sc.)**

Avril 2025
© Kévin-Christopher Tiacoh, 2025

Résumé

Bien que peu nombreuses, toutes les études ayant examiné les moments d'ordre supérieur de la distribution des rendements boursiers, notamment la skewness et la kurtosis, s'accordent sur la pertinence de leur intégration dans la gestion des risques et l'amélioration de la performance des investissements. Ces moments s'avèrent ainsi utiles tant pour les investisseurs et les gestionnaires d'actifs que pour les autorités de régulation des marchés financiers. Toutefois, la littérature existante s'est jusqu'ici surtout concentrée sur l'effet de l'information sur les prix des actifs et leur volatilité, par le biais de ses différents canaux de diffusion. Son impact potentiel sur ces deux autres dimensions de la distribution des rendements a, en revanche, été largement négligé.

Ce mémoire vise à combler cette lacune en analysant principalement l'impact de la fréquence des nouvelles sur la skewness et la kurtosis des rendements boursiers, tout en examinant également ses effets sur la volatilité et le rendement moyen arithmétique.

À partir d'un modèle de régression OLS groupée, corrigé pour l'endogénéité (ainsi que pour d'autres hypothèses économétriques testées), les résultats suggèrent qu'une fréquence accrue des nouvelles est associée à une probabilité plus élevée d'observer des rendements extrêmes dans la distribution future des rendements boursiers, ce qui corrobore l'hypothèse initiale relative à la kurtosis. À l'inverse, un effet statistiquement significatif est également observé sur la volatilité, mais son signe est contraire à celui attendu : une fréquence plus élevée des nouvelles est associée à une baisse, et non à une hausse, de la volatilité future des rendements. Quant à la skewness et le rendement moyen arithmétique, aucun effet statistiquement significatif n'est mis en évidence au seuil conventionnel de 5%, et les effets observés vont dans le sens inverse des relations négatives initialement supposées.

Sur la base d'une classification en quartiles, cinq portefeuilles d'entreprises ont été construits en fonction de la fréquence des nouvelles observée lors de la période précédente, utilisée ici comme proxy du niveau de couverture médiatique. Ce mémoire examine donc également : (1) l'effet différencié de la fréquence des nouvelles selon le degré d'exposition médiatique des portefeuilles, et (2) la stabilité de cet effet à travers différents cycles économiques.

Mise en garde : Il convient toutefois de souligner que les données de fréquence des nouvelles utilisées dans cette étude proviennent de la base RavenPack et n'ont pas fait l'objet d'une validation indépendante. Par conséquent, les résultats présentés doivent être interprétés avec prudence, dans la mesure où toute faiblesse dans la qualité des données pourrait introduire un biais de type « *garbage in, garbage out* ».

Mots-clés : skewness, kurtosis, volatilité, rendement moyen arithmétique, fréquence des nouvelles, portefeuilles

Méthodes de recherche : régression OLS groupée, estimation IV-2SLS (méthode des variables instrumentales en deux étapes), classification de portefeuilles selon les quartiles

Pendant la préparation de ce travail, j'ai utilisé ChatGPT afin de résoudre certains problèmes de programmation spécifiques et d'améliorer la qualité rédactionnelle globale. Après avoir utilisé cet outil, j'ai relu et corrigé le contenu si nécessaire et j'assume l'entièvre responsabilité du contenu de cette publication.

Table des matières

<i>Liste des tableaux</i>	1
<i>Liste des figures</i>	2
1. <i>Remerciements</i>	3
2. <i>Introduction</i>	5
3. <i>Revue de Littérature</i>	8
3.1 Fréquence des nouvelles et efficacité des marchés: réactions immédiates ou différées	8
3.2 Réactions comportementales à la fréquence des nouvelles: surréaction et sentiment des investisseurs	9
3.3 Fréquence des nouvelles et dynamique de la volatilité	11
3.4 Couverture médiatique et sauts de prix boursiers: implications pour la skewness et la kurtosis	13
3.5 Nouvelles macroéconomiques et politiques: amplification des distributions de rendements	14
3.6 Inattention des investisseurs et fréquence des nouvelles: effets sur les queues de distribution des rendements	16
3.7 Annonces de politique monétaire: effets sur la volatilité et la distribution des rendements	18
3.8 Intégrer la skewness et la kurtosis dans la gestion de portefeuille: repenser le risque dans un environnement influencé par l'information	20
4. <i>Méthodologie</i>	22
4.1 Collecte de Données	22
4.2 Variables dépendantes	23
4.3 Variables indépendantes	27
4.4 Construction des portefeuilles	33
4.5 Présentation des hypothèses	37
4.6 Modèle économétrique retenu : Régression OLS groupée	40
4.7 Spécification finale du modèle	42
4.8 Vérification des hypothèses sous-jacentes du modèle	45
4.8.1 Exogénéité des variables explicatives	46
4.8.2 Corrélation entre les variables explicatives et multicolinéarité	51
4.8.3 Homoscédasticité conditionnelle des erreurs	54
4.8.4 Indépendance des erreurs	56
4.8.5 Linéarité de la spécification fonctionnelle du modèle	58
5. <i>Présentation et discussion des résultats</i>	61
6. <i>Tests de robustesse</i>	74

6.1 Hétérogénéité de l'effet de la fréquence des nouvelles selon les portefeuilles	74
6.2 Analyse par sous-périodes autour de la crise financière de 2008	84
7. Résumé et interprétation des résultats	93
7.1 Limites de l'étude	93
7.2 Conclusion	95
8. Bibliographie	97
9. Annexes	100

Liste des tableaux

TABLEAU 1 - MATRICE DE TRANSITION DES PORTEFEUILLES.....	35
TABLEAU 2 - STATISTIQUES DESCRIPTIVES DE MEANFREQP,T, VARIABLE EXPLICATIVE PRINCIPALE	36
TABLEAU 3 - RÉSULTATS DE LA RÉGRESSION OLS GROUPÉE: SPÉCIFICATION FINALE APRÈS SÉLECTION RÉTROGRADE (SEUIL DE SIGNIFICATIVITÉ: VALEUR P ≤ 10 %)	43
TABLEAU 4 - RÉSULTATS DU TEST DE DURBIN-HAUSMAN CONCERNANT L'ENDOGÉNÉITÉ DE MEANFREQP,T (MÉTHODE DE RÉGRESSION AUXILIAIRE) DANS L'ANALYSE DE LA SKEWNESS SEMESTRIELLE.....	47
TABLEAU 5 - RÉSULTATS DE LA RÉGRESSION IV-2SLS CORRIGEANT L'ENDOGÉNÉITÉ DE MEANFREQP,T.....	49
TABLEAU 6 - RÉGRESSION DE PREMIÈRE ÉTAPE: PERTINENCE INSTRUMENTALE DE MEANFREQP,T_LAG POUR EXPLIQUER MEANFREQP,T49	
TABLEAU 7 - VÉRIFICATION DE LA MULTICOLINÉARITÉ BASÉE SUR LE VIF: SPÉCIFICATION FINALE	53
TABLEAU 8 - RÉSULTATS DES TESTS DE BREUSCH-PAGAN ET DE WHITE SUR LES RÉSIDUS DE LA RÉGRESSION IV DANS L'ANALYSE DE LA SKEWNESS SEMESTRIELLE	55
TABLEAU 9 - RÉSULTATS DES TESTS DE WOOLDRIDGE ET DE BREUSCH-GODFREY SUR LES RÉSIDUS DE LA RÉGRESSION IV DANS L'ANALYSE DE LA SKEWNESS SEMESTRIELLE	57
TABLEAU 10 - EFFET DE LA FRÉQUENCE DES NOUVELLES SUR LA SKEWNESS SEMESTRIELLE: ESTIMATION IV-2SLS APRÈS CORRECTION DE L'ENDOGÉNÉITÉ	62
TABLEAU 11 - EFFET DE LA FRÉQUENCE DES NOUVELLES SUR LA KURTOSIS SEMESTRIELLE: ESTIMATION IV-2SLS APRÈS CORRECTION DE L'ENDOGÉNÉITÉ	65
TABLEAU 12 - EFFET DE LA FRÉQUENCE DES NOUVELLES SUR LA VOLATILITÉ SEMESTRIELLE: ESTIMATION IV-2SLS APRÈS CORRECTION DE L'ENDOGÉNÉITÉ	68
TABLEAU 13 - EFFET DE LA FRÉQUENCE DES NOUVELLES SUR LE RENDEMENT MOYEN ARITHMÉTIQUE SEMESTRIEL: ESTIMATION IV-2SLS APRÈS CORRECTION DE L'ENDOGÉNÉITÉ	72
TABLEAU 14 - EFFET HÉTÉROGÈNE DE LA FRÉQUENCE DES NOUVELLES SUR LA SKEWNESS SEMESTRIELLE: ESTIMATION IV-2SLS AVEC INTERACTIONS DE PORTEFEUILLE.....	75
TABLEAU 15 - EFFET HÉTÉROGÈNE DE LA FRÉQUENCE DES NOUVELLES SUR LA KURTOSIS SEMESTRIELLE: ESTIMATION IV-2SLS AVEC INTERACTIONS DE PORTEFEUILLE.....	77
TABLEAU 16 - EFFET HÉTÉROGÈNE DE LA FRÉQUENCE DES NOUVELLES SUR LA VOLATILITÉ SEMESTRIELLE: ESTIMATION IV-2SLS AVEC INTERACTIONS DE PORTEFEUILLE.....	79
TABLEAU 17 - EFFET HÉTÉROGÈNE DE LA FRÉQUENCE DES NOUVELLES SUR LE RENDEMENT MOYEN ARITHMÉTIQUE SEMESTRIEL: ESTIMATION IV-2SLS AVEC INTERACTIONS DE PORTEFEUILLE	81
TABLEAU 18 - IMPACT DE LA FRÉQUENCE DES NOUVELLES SUR LA SKEWNESS SEMESTRIELLE AVANT, PENDANT, ET APRÈS LA CRISE DE 2008: ESTIMATIONS IV-2SLS	85
TABLEAU 19 - IMPACT DE LA FRÉQUENCE DES NOUVELLES SUR LA KURTOSIS SEMESTRIELLE AVANT, PENDANT, ET APRÈS LA CRISE DE 2008: ESTIMATIONS IV-2SLS	87
TABLEAU 20 - IMPACT DE LA FRÉQUENCE DES NOUVELLES SUR LA VOLATILITÉ SEMESTRIELLE AVANT, PENDANT, ET APRÈS LA CRISE DE 2008: ESTIMATIONS IV-2SLS	89
TABLEAU 21 - IMPACT DE LA FRÉQUENCE DES NOUVELLES SUR LE RENDEMENT MOYEN ARITHMÉTIQUE SEMESTRIEL AVANT, PENDANT, ET APRÈS LA CRISE DE 2008: ESTIMATIONS IV-2SLS.....	91
TABLEAU 22 - RÉSULTATS DU TEST DE DURBIN-HAUSMAN CONCERNANT L'ENDOGÉNÉITÉ DE LIQUIDITYP,T (MÉTHODE DE RÉGRESSION AUXILIAIRE) DANS L'ANALYSE DE LA SKEWNESS SEMESTRIELLE	100
TABLEAU 23 - RÉSULTATS DU TEST DE DURBIN-HAUSMAN CONCERNANT L'ENDOGÉNÉITÉ DE SMB SEMI-ANNUAL RETURN (MÉTHODE DE RÉGRESSION AUXILIAIRE) DANS L'ANALYSE DE LA SKEWNESS SEMESTRIELLE	101

Liste des figures

FIGURE 1 – TAUX DE ROTATION DES PORTEFEUILLES PAR SEMESTRE	34
FIGURE 2 – MATRICE DE CORRÉLATION DES VARIABLES EXPLICATIVES POUR L’ANALYSE DE LA SKEWNESS SEMESTRIELLE.....	52
FIGURE 3 – GRAPHIQUES DE DISPERSION AVEC COURBE LOWESS ET RÉSULTATS DU TEST RESET DE RAMSEY (INDIVIDUEL) POUR L’ANALYSE DE LA SKEWNESS SEMESTRIELLE	59
FIGURE 4 - ÉVOLUTION DE LA SKEWNESS SEMESTRIELLE SELON LES PORTEFEUILLES CONSTRUITS EN FONCTION DE LA FRÉQUENCE DES NOUVELLES	61
FIGURE 5 - ÉVOLUTION DE LA KURTOSIS SEMESTRIELLE SELON LES PORTEFEUILLES CONSTRUITS EN FONCTION DE LA FRÉQUENCE DES NOUVELLES	64
FIGURE 6 - ÉVOLUTION DE LA VOLATILITÉ SEMESTRIELLE SELON LES PORTEFEUILLES CONSTRUITS EN FONCTION DE LA FRÉQUENCE DES NOUVELLES	67
FIGURE 7 - ÉVOLUTION DU RENDEMENT MOYEN ARITHMÉTIQUE SEMESTRIEL SELON LES PORTEFEUILLES CONSTRUITS EN FONCTION DE LA FRÉQUENCE DES NOUVELLES	71

1. Remerciements

Tout d'abord, je tiens à remercier le corps professoral de HEC Montréal, en commençant par mon directeur de mémoire, Monsieur Martin Boyer. Lorsque mon premier choix de sujet, centré sur l'immobilier et qui me tenait particulièrement à cœur, s'est révélé irréalisable faute de données, c'est lui qui m'a proposé ce sujet en alternative. Bien qu'il ne s'agisse pas de mon choix initial, celui-ci a progressivement suscité mon intérêt et s'est avéré tout aussi passionnant.

Je suis également profondément reconnaissant envers Monsieur Grégoire Vincent, sans qui ce projet de recherche n'aurait pas été possible. Grâce à lui, j'ai pu obtenir les données relatives à la couverture médiatique des entreprises étudiées, après avoir constaté qu'il n'était plus possible de les extraire automatiquement via Factiva.

Je ne saurais oublier de remercier Monsieur Jaber Mohamed, administrateur du laboratoire, qui s'est toujours rendu disponible et dont le soutien technique, tant pour le traitement que pour l'extraction des données, a été d'une valeur inestimable.

À vous, messieurs, j'aimerai adresser ma plus profonde gratitude pour votre disponibilité, malgré vos lourdes responsabilités et vos emplois du temps chargés. Merci pour votre patience, pour m'avoir plongé dans vos univers respectifs et pour avoir partagé vos connaissances avec tant de générosité. Tout cela, conjugué à vos précieux conseils, m'a permis de mener à terme ce projet, qui me semblait au départ presque hors de portée. Je veillerai à mettre en pratique ces enseignements enrichissants tant dans mon développement personnel que dans ma future carrière professionnelle.

Je souhaite également remercier mon père, Jean-François, pour son soutien financier indéfectible depuis le début de mon parcours à HEC Montréal en 2017. Merci Papa de m'avoir permis de poursuivre des études supérieures et de m'avoir soutenu financièrement pendant toutes ces années. Tu n'imagines pas le poids que tu as retiré de mes épaules. Merci aussi pour ta patience, malgré ta santé fragile, face à l'allongement de ce projet sur plusieurs sessions.

Je suis tout aussi reconnaissant envers ma mère, Zahara, ainsi que mes sœurs Mélie, Andréa et Inès, pour leur soutien affectif et moral constant. Merci d'avoir toujours cru en moi et en mon potentiel, même lorsque je peinais moi-même à voir la lumière au bout du tunnel.

Enfin, je tiens à me remercier de ne pas avoir abandonné. Ce projet de recherche a sans doute été l'épreuve la plus difficile et exigeante de tout mon parcours académique. Je me félicite d'avoir résisté à l'envie de tout laisser tomber, mais d'avoir plutôt fait preuve de discipline, de résilience, de patience, de curiosité, et surtout, d'avoir appris à relativiser. Je suis fier de ne pas avoir laissé mon anxiété de performance et mes problèmes de dépression prendre le dessus. Malgré tous les obstacles, je suis convaincu que ce projet m'a permis de grandir et de mieux me préparer à affronter le monde professionnel.

2. Introduction

Depuis de nombreuses années, les acteurs des marchés financiers s'interrogent sur les facteurs qui influencent réellement le comportement des marchés. Parmi cette multitude de facteurs, l'information a toujours été considérée comme un élément central dans la formation des prix. Toutefois, c'est l'étude fondatrice de Fama, Fisher, Jensen et Roll (1969), *The Adjustment of Stock Prices to New Information*, qui a formalisé cette idée en introduisant l'hypothèse d'efficience des marchés. Celle-ci postule que les prix des actifs intègrent de manière instantanée et impartiale toute nouvelle information, rendant ainsi impossible la réalisation systématique de rendements excédentaires à partir de l'information publique. Depuis lors, cette hypothèse est devenue une pierre angulaire de la recherche académique visant à explorer la complexité de la relation entre information et dynamiques des marchés financiers.

Cependant, des études ultérieures ont montré que cette relation est plus nuancée que ce qui avait été initialement supposé. Des phénomènes tels que la sous-réaction et la surréaction aux nouvelles ont été documentés (De Bondt & Thaler, 1985 ; Veronesi, 1999), suggérant ainsi que les prix des actifs peuvent subir des distorsions temporaires en raison de biais comportementaux ou de contraintes informationnelles. Ces constats ont ouvert la voie à de nouvelles approches visant à mieux comprendre l'impact de l'information non seulement sur le niveau des prix, mais également sur la distribution des rendements.

En approfondissant cette relation, la littérature s'est principalement concentrée sur les réactions des prix et la volatilité, suite à la publication d'informations. Engle et Ng (1991) ont mis en lumière un effet asymétrique des chocs informationnels sur la volatilité conditionnelle des rendements : les mauvaises nouvelles, assimilées à des chocs négatifs, génèrent une augmentation plus forte de la volatilité que les bonnes, un phénomène souvent associé à l'effet de levier. Plus récemment, Jeon, McCurdy et Zhao (2021) ont quant à eux, mis en évidence que les informations relayées par les médias peuvent déclencher des sauts brusques dans les rendements boursiers, soulignant ainsi l'importance non seulement du volume d'informations, mais aussi de la manière dont celle-ci est perçue et diffusée par les médias.

Au-delà de leur influence immédiate sur les prix des actifs et la volatilité, certaines études se sont penchées sur le pouvoir prédictif des nouvelles et du contenu médiatique sur les rendements futurs.

Tetlock (2007) a montré dans son étude que le ton des articles financiers publiés dans la presse influence les rendements boursiers à court terme : un ton négatif est souvent suivi d'une baisse des prix, tandis qu'un ton positif peut entraîner une hausse des rendements. De manière similaire, Schumaker, Zhang, Huang et Chen (2012) ont constaté que l'analyse du sentiment des articles de presse financière peut contribuer à anticiper les mouvements de prix avec un certain degré de précision, suggérant que les informations médiatiques contiennent des signaux utiles et exploitables pour les investisseurs cherchant à prévoir les fluctuations des cours boursiers. Ainsi, l'information ne se limite pas à déclencher des réactions immédiates sur les marchés, mais peut également servir d'outil prédictif des tendances futures.

Bien que de nombreuses études aient analysé l'impact de l'information sur les marchés financiers, elles se concentrent généralement sur une catégorie spécifique de nouvelles, telles que les annonces macroéconomiques, politiques ou monétaires, et évaluent leurs effets sur les rendements ou la volatilité. Même lorsque ces recherches adoptent une approche plus large, sans distinction entre les types d'informations, elles demeurent principalement axées sur deux dimensions : la réaction des prix et la volatilité des rendements, ou le pouvoir prédictif des nouvelles sur ces caractéristiques de la distribution des rendements. Ainsi, malgré la richesse des travaux existants, l'impact des nouvelles sur des mesures plus fines de la distribution des rendements, comme la skewness et la kurtosis, demeure largement inexplorée.

Outre les caractéristiques traditionnelles telles que la moyenne et la variance, la skewness et la kurtosis sont des mesures statistiques essentielles pour analyser la forme de la distribution des rendements boursiers. La skewness capture l'asymétrie de la distribution par rapport à une distribution normale symétrique. Une asymétrie marquée peut traduire une propension des marchés à générer des rendements extrêmes, positifs ou négatifs, influençant ainsi la perception du risque par les investisseurs. La kurtosis, quant à elle, mesure le degré de concentration ou d'aplatissement de la distribution et reflète la fréquence des événements extrêmes. Une kurtosis élevée suggère une plus grande prévalence de chocs importants, ce qui a des implications majeures en matière de gestion des risques. « La modélisation de l'asymétrie des rendements devient de plus en plus importante en économie et en finance. [...] En connaissant la forme de la distribution des rendements d'un portefeuille, les investisseurs seront en mesure de porter de meilleurs jugements en fonction de leurs préférences en matière de risque » (Premaratne & Bera, 2000). Ainsi, une

meilleure compréhension de la distribution des rendements permet non seulement aux investisseurs d'ajuster leurs stratégies d'investissement selon leur tolérance au risque, mais elle offre également aux autorités de régulation des marchés financiers un levier pour anticiper d'éventuels déséquilibres.

S'appuyant sur cette lacune dans la littérature, la présente étude vise à analyser l'influence de la fréquence des nouvelles sur la distribution des rendements boursiers, en mettant particulièrement l'accent sur la skewness et la kurtosis. À cette fin, des portefeuilles d'entreprises sont construits en fonction de la fréquence des articles publiés à leur sujet au cours de la période précédente. Dans le cadre de cette étude, cette fréquence est considérée comme un indicateur de l'intensité de la couverture médiatique. À l'aide d'une régression OLS groupée, l'étude examine comment la fréquence des nouvelles affecte différemment la skewness et la kurtosis des rendements boursiers, apportant ainsi une contribution originale à la littérature financière. Par ailleurs, l'impact de la fréquence des nouvelles sur la volatilité des rendements et le rendement moyen arithmétique est également analysé, non pas en tant qu'objet principal d'analyse, mais afin d'inscrire cette recherche dans une perspective plus large de la dynamique des rendements boursiers.

Mise en garde : Il convient toutefois de souligner que les données de fréquence des nouvelles utilisées dans cette étude proviennent de la base RavenPack et n'ont pas fait l'objet d'une validation indépendante. Par conséquent, les résultats présentés doivent être interprétés avec prudence, dans la mesure où toute faiblesse dans la qualité des données pourrait introduire un biais de type « garbage in, garbage out ».

Cette étude est structurée comme suit : la section suivante présente une revue de littérature, en mettant l'accent sur l'influence de la fréquence et du type d'information sur la distribution des rendements, avec une attention particulière portée à la skewness, à la kurtosis, à la volatilité et aux réactions comportementales. Elle explore également le rôle du sentiment des investisseurs, de l'attention médiatique et du contenu des nouvelles dans l'amplification des risques extrêmes et des asymétries de rendements. Les sections suivantes détaillent le processus de collecte des données et la méthodologie employée, avant de procéder à l'analyse des résultats obtenus. L'étude se conclut par une discussion des limites méthodologiques et des pistes de recherche futures.

3. Revue de Littérature

3.1 Fréquence des nouvelles et efficacité des marchés: réactions immédiates ou différées

La fréquence de diffusion des nouvelles financières influence directement la rapidité et la précision avec lesquelles les prix des actifs intègrent l'information. Ce processus affecte non seulement les rendements et la volatilité, mais également les moments statistiques d'ordre supérieur, tels que la skewness et la kurtosis, qui caractérisent respectivement l'asymétrie des distributions et les risques extrêmes. L'étude fondatrice de Fama, Fisher, Jensen et Roll (1969), analysant 940 fractionnements d'actions sur le NYSE entre 1927 et 1959, conclut à une réaction quasi instantanée des prix, en accord avec l'hypothèse d'efficience des marchés (EMH). Ce cadre théorique, fondé sur des anticipations parfaitement rationnelles, s'avère toutefois de plus en plus restrictif dans un environnement dominé par le trading algorithmique et la complexité croissante des flux informationnels.

À rebours de cette vision classique, plusieurs travaux récents mettent en évidence des inefficiences dans le traitement de l'information. À partir d'un corpus de plus de 250 000 articles du WSJ.com examinés à l'aide d'un classificateur bayésien, Antweiler et Frank (2006) observent que les périodes d'intense activité médiatique sont suivies de rendements anormaux corrigés ultérieurement, suggérant ainsi des surréactions initiales. De manière complémentaire, Chan (2001) montre que des actions associées à des nouvelles négatives subissent des baisses prolongées avant ajustement, ce qui révèle des réactions différées potentiellement liées à des biais cognitifs ou à une attention sélective des investisseurs. Ces études, bien qu'utilisant des sources et méthodes distinctes, convergent vers l'idée que l'information n'est pas systématiquement intégrée de manière immédiate ni rationnelle, et que cette inefficience pourrait contribuer à l'apparition d'asymétries, notamment une skewness négative dans les distributions de rendements. Cependant, aucune d'entre elles ne mesure directement les moments d'ordre supérieur, limitant ainsi la portée de leurs conclusions quant à l'effet précis de la fréquence des nouvelles sur la forme des distributions.

La littérature empirique a également exploré les liens entre fréquence de l'information, activité boursière et volatilité. Berry et Howe (1994), en étudiant les dépêches Reuters entre 1990 et 1991, identifient une corrélation positive entre fréquence des nouvelles et volume de transactions, sans effet significatif sur la volatilité. Une limite importante de leur étude réside dans l'homogénéisation des contenus, sans distinction selon leur importance économique. En

contrepoint, Peress (2011) exploite des grèves de journaux comme chocs exogènes affectant la diffusion de l'information. Il montre qu'en période de grève, le volume des transactions chute de 14 % et la volatilité intrajournalière diminue de 9 %, indiquant que la rareté de l'information ralentit à la fois l'activité boursière et la dynamique des prix. Bien que cette approche renforce l'idée d'un rôle structurant de la fréquence informationnelle, sa focalisation sur la presse écrite limite sa portée dans un contexte marqué par l'essor des canaux numériques. Si Berry et Howe (1994) et Peress (2011) s'accordent sur le lien entre fréquence des nouvelles et activité des investisseurs, leurs résultats divergent concernant la volatilité. Cette différence, liée à leurs méthodologies respectives et aux périodes étudiées, suggère qu'un flux intense d'information pourrait accentuer la concentration des rendements extrêmes, traduite par une kurtosis plus élevée. Toutefois, l'absence de mesures explicites de skewness et de kurtosis dans ces travaux empêche de confirmer cette hypothèse.

Ainsi, la fréquence de diffusion de l'information apparaît à la fois comme un vecteur d'activité et un facteur potentiel d'inefficience. En période de forte activité informationnelle, les marchés peuvent connaître des ajustements différés, des déséquilibres transitoires ou des comportements mimétiques, surtout lorsque la capacité à filtrer le bruit informationnel est limitée. Ces distorsions peuvent altérer la symétrie des distributions de rendements et amplifier les risques extrêmes. Le manque récurrent de mesures directes des moments d'ordre supérieur constitue une lacune importante dans l'évaluation de ces dynamiques. Le présent mémoire se propose précisément de combler cette lacune en analysant empiriquement le lien entre fréquence des nouvelles et caractéristiques distributionnelles des rendements.

3.2 Réactions comportementales à la fréquence des nouvelles: surréaction et sentiment des investisseurs

La diffusion fréquente d'informations joue un rôle déterminant dans la formation des comportements d'investissement, en activant des réactions cognitives et émotionnelles susceptibles d'influencer les prix des actifs, la volatilité et la structure des rendements. La littérature en finance comportementale souligne que les décisions des investisseurs sont souvent façonnées par des biais psychologiques et des dynamiques émotionnelles. Dans cette perspective, De Bondt et Thaler (1985) mènent une étude emblématique sur les actions du NYSE entre 1926 et 1982, en formant des portefeuilles contraires regroupant les titres ayant connu les plus fortes hausses (“gagnants”) et

baisses (“perdants”) passées. Ils observent une surperformance notable des titres “perdants” sur un horizon de trois ans, particulièrement marquée en janvier, possiblement liée à des ventes fiscales. Cette dynamique, interprétée comme une correction de sous-évaluation, suggère que les mauvaises nouvelles fréquentes peuvent induire des surréactions et générer des opportunités de retour à la moyenne, ce qui favoriserait une asymétrie négative. Toutefois, l'étude repose sur les rendements anormaux cumulés, sans ajustement pour des facteurs de risque tels que la taille ou le momentum, limitant ainsi la portée explicative.

Dans un cadre théorique, Veronesi (1999) propose un modèle d'apprentissage en incertitude, selon lequel les effets des nouvelles dépendent du contexte macroéconomique. Le modèle prévoit que les mauvaises nouvelles suscitent des surréactions plus fortes en période d'expansion, tandis que les bonnes nouvelles génèrent des réponses plus modérées en récession. Ce résultat souligne le caractère contextuel du sentiment des investisseurs, dont la réactivité dépend du niveau d'attentes *ex ante*. Néanmoins, l'absence de validation empirique ou d'application concrète réduit la portée du modèle, notamment quant à ses implications sur la forme des distributions de rendements.

Adoptant une approche empirique, Tetlock (2007) analyse l'impact du ton médiatique sur les prix boursiers, à partir de la chronique “Abreast of the Market” publiée dans le Wall Street Journal entre 1984 et 1999. Il constate qu'un ton pessimiste est suivi de baisses de cours, souvent compensées les jours suivants, ainsi que d'une hausse des volumes de transactions. Ces réactions sont interprétées comme des manifestations de surréactions induites par le sentiment négatif. Cependant, l'analyse étant centrée sur une seule chronique, la représentativité des résultats peut s'en trouver limitée. García (2012) étend cette approche à partir d'un siècle d'archives du New York Times et montre que, durant les récessions, les nouvelles négatives ont un effet plus marqué sur les rendements quotidiens du Dow Jones qu'en période d'expansion. Ce résultat renforce l'idée d'une sensibilité accrue des investisseurs au pessimisme médiatique dans les périodes d'incertitude. Toutefois, la méthode lexicale mobilisée pour quantifier le sentiment reste relativement rudimentaire, ce qui peut affecter la précision des inférences.

L'étude de Huberman et Regev (1999) illustre de manière frappante l'impact de la rediffusion d'une information déjà connue. Les auteurs documentent une hausse spectaculaire de 330 % du cours d'EntreMed à la suite d'un article du New York Times relayant des résultats scientifiques publiés antérieurement. Cette réaction souligne que la visibilité ou la répétition d'une

information peut, indépendamment de sa nouveauté, déclencher des mouvements de prix significatifs. Bien que révélatrice, cette étude repose sur une approche anecdotique, ce qui en limite la généralisation empirique.

Dans l'ensemble, ces travaux indiquent que la fréquence de l'information peut amplifier certains biais comportementaux, en particulier les surréactions, et contribuer à des distorsions de prix. Une asymétrie négative peut émerger lorsque les investisseurs réagissent de manière excessive aux nouvelles défavorables, tandis qu'un afflux d'informations en période d'incertitude peut engendrer des mouvements de prix abrupts, renforçant la kurtosis. Toutefois, peu d'études mobilisent explicitement des mesures statistiques de skewness ou de kurtosis, ce qui constitue une limite importante. Dans une analyse centrée sur les moments d'ordre supérieur, intégrer ces mesures permettrait d'appréhender plus finement les effets de la fréquence informationnelle sur la structure des distributions de rendements financiers.

3.3 Fréquence des nouvelles et dynamique de la volatilité

La fréquence des nouvelles financières entretient un lien étroit avec la volatilité des marchés, en influençant directement les anticipations de risque et, de manière indirecte, les moments d'ordre supérieur tels que la kurtosis. En tant qu'indicateur clé de la perception du risque, la volatilité occupe une place centrale dans la valorisation des actifs et la gestion des portefeuilles. Engle et Ng (1991) figurent parmi les premiers à modéliser l'asymétrie de la réponse de la volatilité aux nouvelles à l'aide de spécifications GARCH non linéaires (EGARCH et GJR-GARCH), appliquées aux données du marché Japonais. Ils montrent que les mauvaises nouvelles, assimilées à des chocs négatifs, ont un effet plus prononcé sur la volatilité conditionnelle que les nouvelles positives de même intensité, ce qui se traduit par des mécanismes tels que l'aversion au risque ou l'effet de levier. En introduisant la notion de « courbe d'impact des nouvelles », leur cadre formalise la dissymétrie des réactions du marché face aux chocs informationnels. Toutefois, leur modèle ne tient pas compte de l'hétérogénéité des nouvelles en termes de contenu, de source ou de contexte, ce qui peut nuancer la portée de leurs conclusions.

Adoptant une perspective complémentaire, Chiang (2016) examine la relation entre pessimisme médiatique et anticipation de la volatilité, à travers l'analyse d'articles du Wall Street Journal entre 2002 et 2009 et leur lien avec le VIX. Il identifie une corrélation positive entre la

fréquence des nouvelles à tonalité négative et les pics de volatilité implicite, suggérant que le ton médiatique influence les anticipations de risque. Contrairement à Engle et Ng, l'approche privilégie la dimension anticipative plutôt que les effets *ex post* sur les rendements. Toutefois, la méthode lexicale mobilisée, fondée sur des dictionnaires statiques, ne capture pas toujours la richesse sémantique du langage économique, ce qui en limite la précision du signal.

Les travaux d'Ederington et Lee (1996) apportent un éclairage distinct en différenciant les effets des annonces prévues (telles que les chiffres de l'emploi) et des nouvelles imprévues sur la volatilité implicite, dans les marchés obligataires et des changes. Leurs résultats indiquent que la prévisibilité réduit l'intensité des réactions, tandis que les nouvelles inattendues, surtout lorsqu'elles sont négatives, provoquent des hausses marquées de volatilité. Cette idée est prolongée par Bomfim (2000), qui montre que les ajustements monétaires inattendus du FOMC entre 1989 et 1998, entraînent une réaction de volatilité plus marquée, notamment lorsque les taux sont relevés. Les deux études mettent en évidence un effet asymétrique selon la direction des nouvelles. Toutefois, la portée empirique de leurs résultats demeure partielle : Bomfim, en s'appuyant sur des données journalières, ne permet pas de capturer les ajustements intrajournaliers souvent plus vifs, alors qu'Ederington et Lee, bien qu'utilisant des données à haute fréquence, se concentrent exclusivement sur les marchés obligataires et des changes, limitant ainsi la généralisation aux marchés actions.

Pris ensemble, ces travaux soulignent que la fréquence, la tonalité et le caractère inattendu des nouvelles façonnent la dynamique de la volatilité. Les mauvaises nouvelles, notamment lorsqu'elles sont fréquentes ou surprenantes, tendent à accentuer les fluctuations de prix, accroissant ainsi la probabilité d'événements extrêmes, une caractéristique associée à une kurtosis élevée. Pourtant, ces études se concentrent sur la volatilité en tant que second moment, sans intégrer explicitement les moments d'ordre supérieur. La relation entre volatilité, fréquence informationnelle et asymétrie (skewness) demeure en grande partie inexplorée. Dans le cadre d'un travail centré sur la distribution des rendements, cette lacune méthodologique est notable. Une analyse conjointe des moments supérieurs et de la volatilité permettrait de mieux comprendre les effets différenciés des nouvelles sur la forme des distributions, en tenant compte à la fois de la dispersion, de l'asymétrie et de la concentration des queues. Ainsi, si les recherches existantes éclairent le rôle des nouvelles dans la transmission du risque, elles appellent à des investigations plus complètes intégrant la skewness et la kurtosis dans les dynamiques de volatilité.

3.4 Couverture médiatique et sauts de prix boursiers: implications pour la skewness et la kurtosis

La fréquence de la couverture médiatique influence la forme des distributions de rendements, en agissant à la fois sur l'asymétrie (skewness) et l'épaisseur des queues (kurtosis). Jeon et al. (2021) proposent une analyse empirique fondée sur plus de 21 millions d'articles publiés entre 1980 et 2020. Ils montrent que les sauts de prix sont fortement corrélés à l'intensité médiatique : les hausses sont plus fréquentes en cas de forte exposition, tandis que les baisses s'amplifient lorsque le ton est négatif. En combinant fréquence et tonalité, l'étude offre une lecture fine de l'influence des médias sur les rendements. Toutefois, la méthode employée pour détecter les sauts peut confondre signaux informationnels et bruit de marché, en particulier dans un contexte de trading algorithmique, ce qui complique l'interprétation causale. Malgré cette limite, les auteurs contribuent de manière notable à la compréhension des mécanismes reliant les caractéristiques médiatiques aux discontinuités de prix, et suggèrent l'émergence d'une skewness négative et d'une kurtosis élevée en période d'intensité informationnelle.

En contrepoint, Fang et Peress (2008) adoptent une perspective centrée sur l'absence de couverture médiatique. En étudiant les marchés boursiers américains entre 1993 et 2002, ils constatent que les entreprises non couvertes surperforment celles exposées médiatiquement, avec un écart annuel moyen de 3 %, particulièrement marqué pour les petites capitalisations. Cette surperformance est attribuée à une asymétrie d'information, favorisant des corrections haussières différées lorsqu'une entreprise initialement ignorée attire soudainement l'attention des investisseurs. Une faible fréquence de couverture est ainsi associée à une volatilité moindre et à une atténuation des réactions émotionnelles, susceptibles de limiter les distorsions liées au sentiment de marché. Néanmoins, l'absence de prise en compte de la tonalité des nouvelles constitue une limite, car le cadrage affectif de l'information peut influencer significativement les comportements d'investissement. Malgré cela, l'étude illustre comment un déficit d'attention médiatique peut engendrer des asymétries positives dans la distribution des rendements.

Ces deux approches mettent en évidence des effets contrastés de la fréquence médiatique sur les extrêmes de marché. Jeon et al. soulignent l'amplification des risques baissiers en cas d'exposition négative et soutenue, tandis que Fang et Peress illustrent le potentiel haussier d'un

environnement médiatiquement silencieux. Ainsi, la skewness apparaît sensible à la fois à l'intensité et à la visibilité médiatique.

L'impact de formats alternatifs est également exploré par Antweiler et Frank (2002), qui analysent plus de 1,5 million de messages postés sur les forums Yahoo! Finance entre 1999 et 2001. Ils observent qu'un volume élevé de discussions est associé à une volatilité accrue et à des rendements futurs plus faibles, en ligne avec l'hypothèse de divergence d'opinion. Toutefois, leur analyse ne distingue ni la tonalité des messages ni leur contenu informationnel, ce qui limite la compréhension du rôle exact joué par le sentiment exprimé dans ces échanges numériques. En complément, Strycharz et al. (2018) examinent la couverture de trois entreprises Néerlandaises entre 2014 et 2015, et observent qu'une fréquence élevée de couverture accroît la volatilité, avec des effets modulés selon le secteur. Cette hétérogénéité sectorielle suggère que la sensibilité des investisseurs à l'information dépend du contexte industriel. Néanmoins, la portée externe de l'étude reste limitée en raison de la taille restreinte de l'échantillon et de sa concentration géographique.

Parmi ces travaux, seule l'étude de Jeon et al. établit un lien explicite entre fréquence des nouvelles et sauts de prix, ce qui met en lumière une lacune persistante dans la littérature : l'intégration systématique de la skewness et de la kurtosis demeure rare. Si Antweiler et Frank ou Strycharz et al. confirment l'impact médiatique sur la volatilité, ils n'en examinent pas les conséquences sur la structure complète des distributions, notamment les risques de queue. Pris ensemble, ces résultats confirment que la couverture médiatique, qu'elle soit excessive ou insuffisante, façonne les extrêmes de rendements et les asymétries observées. L'absence de modélisation explicite des moments d'ordre supérieur en limite toutefois l'exploitation dans un cadre centré sur les risques extrêmes. Cela souligne la nécessité d'études intégrant pleinement la ces dimensions pour mieux appréhender l'effet structurel de l'environnement informationnel sur les marchés financiers.

3.5 Nouvelles macroéconomiques et politiques: amplification des distributions de rendements

Les annonces macroéconomiques et politiques exercent une influence déterminante sur la distribution des rendements financiers, en particulier dans des contextes d'incertitude accrue. Ces chocs informationnels affectent à la fois le niveau moyen des rendements et l'intensité de leur dispersion, ce qui contribuent à l'épaisissement des queues de distribution et à des déformations

asymétriques. Flannery et Protopapadakis (2002) proposent l'une des analyses empiriques les plus complètes sur ce sujet, en évaluant l'effet de 17 indicateurs macroéconomiques sur les actions américaines entre 1980 et 1996. Parmi ces facteurs, l'indice des prix à la consommation, la croissance monétaire et les données sur l'emploi se distinguent par leur impact significatif sur la volatilité. L'influence de la croissance monétaire s'étend d'ailleurs aux rendements eux-mêmes, avec des effets captés via la mesure MAD. Toutefois, leur recours à des données agrégées et à faible fréquence limite la détection des ajustements rapides, et restreint l'analyse des dynamiques extrêmes associées à la kurtosis.

Dans une perspective complémentaire, Balduzzi et al. (2001) utilisent des données haute fréquence pour analyser les effets des annonces économiques planifiées sur les obligations du Trésor américain entre 1991 et 1995. Ils montrent que les publications majeures, telles que les chiffres de l'emploi ou de l'inflation, génèrent des variations immédiates des rendements ainsi qu'une concentration de la volatilité durant l'heure suivant leur diffusion. Ce schéma post-annonce, marqué par des ajustements de prix souvent excessifs suivis de corrections, est susceptible d'engendrer des distributions à queues épaisses. Toutefois, l'analyse se limite aux transactions inter-courtiers, ce qui restreint la portée des résultats à une partie seulement des acteurs de marché.

Ces deux contributions confirment le rôle structurant des annonces macroéconomiques sur la dynamique des rendements et de la volatilité, sans pour autant modéliser explicitement les moments d'ordre supérieur. Elles suggèrent plutôt, de manière indirecte, que la fréquence de diffusion de ces annonces, notamment lorsqu'elles sont anticipées ou surviennent en période de turbulence, accentue les risques extrêmes. Cette divergence méthodologique, entre approche multifactorielle de Flannery et Protopapadakis et analyse intrajournalière de Balduzzi et al., reflète des granularités différentes, sans permettre une évaluation précise de l'impact sur la skewness ou la kurtosis.

Dans des environnements politiquement ou économiquement instables, l'influence des nouvelles fréquentes semble s'intensifier. Al-Maadid et al. (2018), en analysant les marchés boursiers du Conseil de coopération du Golfe entre 2010 et 2018, constatent que les nouvelles économiques fréquentes affectent significativement les rendements en période de forte volatilité, avec un effet encore plus prononcé pour les annonces politiques, probablement en raison du rôle central des autorités politiques dans ces marchés. Néanmoins, leur méthodologie repose sur un simple comptage d'articles, sans prise en compte de la tonalité ni du contexte, ce qui limite la

précision des résultats. De surcroît, l'absence de mesures des réactions intrajournalières renforce l'idée d'une absorption partielle ou retardée de l'information, un phénomène particulièrement pertinent dans l'analyse des mouvements extrêmes liés à la kurtosis.

Dans un cadre plus localisé, Essaddam (2002) étudie les effets des nouvelles politiques durant la période du référendum Québécois entre 1990 et 1996, et observe une hausse significative de la volatilité, exacerbée pour les nouvelles politiques perçues comme défavorables. Cette asymétrie de réaction suggère une skewness négative plus marquée en période de tension politique. Toutefois, le recours exclusif au Wall Street Journal comme source d'information constitue une limite, notamment en ce qu'il néglige les perceptions locales, pourtant essentielles pour interpréter les comportements d'investisseurs régionaux.

Pris ensemble, les travaux d'Al-Maadid et d'Essaddam illustrent comment des annonces fréquentes, dans un contexte instable, exacerbent la volatilité des rendements. Néanmoins, aucun ne mesure l'effet sur la skewness ou la kurtosis. Bien que les effets sur la volatilité soient établis, l'impact sur la forme complète des distributions reste encore peu exploré. Or, pour évaluer rigoureusement les risques extrêmes, il est essentiel de comprendre comment ces chocs modifient les queues de distribution et introduisent des asymétries structurelles. En définitive, ces études soulignent l'influence majeure des annonces macroéconomiques et politiques sur la volatilité, en particulier lorsqu'elles sont fréquentes ou inattendues. Bien qu'elles ne testent pas explicitement les moments d'ordre supérieur, elles laissent entrevoir une amplification du risque de queue dans les phases d'instabilité. Elles soulignent ainsi la nécessité de recherches futures intégrant la skewness et la kurtosis dans l'analyse des réactions de marché aux nouvelles, afin de concevoir des modèles de gestion du risque plus robuste face aux chocs informationnels.

3.6 Inattention des investisseurs et fréquence des nouvelles: effets sur les queues de distribution des rendements

La fréquence des nouvelles, conjuguée à l'inattention des investisseurs, constitue un mécanisme clé dans la formation des asymétries de rendement. La cadence de diffusion de l'information influence le degré de vigilance du marché, affectant ainsi le moment et l'intensité des ajustements de prix, et, par conséquent, la forme de la distribution des rendements. Une attention insuffisante peut entraîner des déséquilibres persistants ou des ajustements retardés, tandis qu'un excès d'information peut induire des surréactions. DellaVigna et Pollet (2005) analysent plus de

143 000 annonces de résultats publiées entre 1984 et 2006, et montrent que celles diffusées le vendredi, moment associé à une moindre attention, suscitent des réactions initiales atténées, suivies de corrections tardives. Cette lenteur d'ajustement se traduit par des rendements anormaux mensuels de 3,84 %, révélant que la synchronisation de l'information joue un rôle crucial. Toutefois, leur indicateur d'attention, fondé sur le jour de publication, ne tient pas compte d'autres facteurs tels que la visibilité de l'entreprise ou la concurrence informationnelle, ce qui limite la portée causale des résultats.

Dans une perspective complémentaire, Tetlock (2010) s'intéresse à la redondance médiatique à l'aide d'un indice de similarité (Jaccard) appliqué à des nouvelles financières. Il observe que la répétition excessive d'informations peut engendrer des surréactions, notamment chez les investisseurs individuels susceptibles de mal interpréter ou de surévaluer des contenus déjà connus. Ces excès sont souvent suivis de corrections, générant une volatilité transitoire et accentuant potentiellement la queue gauche de la distribution des rendements. Bien que novatrice, sa méthode peut confondre contenu redondant et nouveauté informationnelle, ce qui risque de nuire à la précision du signal capté.

Ces deux études montrent que l'inattention médiatise la relation entre fréquence de diffusion et asymétrie des rendements. Une information rare ou diffusée à des moments de faible vigilance favorise des ajustements tardifs, pouvant s'accompagner de rendements anormalement positifs et donc d'une asymétrie positive. À l'inverse, une couverture intensive, surtout lorsqu'elle est redondante, tend à induire des réactions excessives, susceptibles de favoriser une skewness négative. Cette divergence tient également à la nature des nouvelles étudiées : DellaVigna et Pollet se concentrent sur des annonces de résultats à l'échelle microéconomique, tandis que Tetlock analyse des cycles médiatiques globaux et leur interprétation comportementale.

Ferguson et al. (2015) prolongent cette réflexion en montrant que la fréquence des nouvelles améliore la prévisibilité des rendements, notamment pour les petites capitalisations ou les entreprises peu visibles. Leur analyse, qui couvre la période 1981–2010, révèle qu'une couverture médiatique intensive amplifie les réactions du marché, surtout lorsque l'attention initiale des investisseurs est limitée. Cela appuie l'idée que la fréquence des nouvelles réduit l'asymétrie d'information, en diffusant des signaux pertinents aux acteurs insuffisamment informés. Néanmoins, l'usage du dictionnaire de sentiment de Loughran–McDonald peut entraîner une

classification imprécise de certaines nuances linguistiques, affectant ainsi la robustesse des résultats.

Dans une optique algorithmique, Schumaker et al. (2012) mobilisent des techniques d'apprentissage automatique pour prédire les mouvements de prix à partir de nouvelles négatives. Leur modèle, appliqué à l'année 2005, atteint une précision de 59 % et génère un rendement de 3,04 %, ce qui suggère que des flux informationnels négatifs et fréquents peuvent anticiper des variations extrêmes. Bien que la période d'analyse soit courte, leur approche à haute fréquence renforce l'idée d'un lien entre le contenu médiatique et le risque de queue, potentiellement associé à une hausse de la kurtosis.

Parmi ces études, seule celle de Tetlock aborde explicitement les retournements de rendements, élément central de l'analyse des asymétries. Ensemble, ces travaux confirment que la fréquence des nouvelles module l'attention des investisseurs qui influence par la suite, le timing et l'amplitude des ajustements de prix. Une diffusion limitée, mal synchronisée, tend à générer des réactions différées, souvent associées à une asymétrie positive. À l'inverse, une surabondance d'information, surtout émotionnellement chargée ou redondante, accroît la probabilité de surréactions et de corrections ultérieures, ce qui pourrait se traduire par une skewness négative et une concentration accrue des rendements extrêmes. Ces constats soulignent la nécessité d'une modélisation plus fine du traitement de l'information par les marchés financiers, intégrant à la fois la fréquence, le contenu et le moment de diffusion des nouvelles. Une telle approche permettrait de mieux saisir l'incidence de l'environnement informationnel sur les moments d'ordre supérieur des rendements, et ainsi d'évaluer plus rigoureusement les risques extrêmes dans les marchés financiers.

3.7 Annonces de politique monétaire: effets sur la volatilité et la distribution des rendements

Les annonces de politique monétaire, en particulier lorsqu'elles sont fréquentes et inattendues, figurent parmi les sources d'information les plus influentes sur les marchés financiers. En modifiant les anticipations relatives aux taux d'intérêt et à l'inflation, elles génèrent souvent des pics de volatilité qui affectent la forme des distributions de rendements sur différentes classes d'actifs. Bomfim (2000) propose l'une des premières contributions empiriques solides sur ce sujet en examinant la réaction de l'indice S&P 500 entre 1989 et 1998. Il observe une volatilité modérée

dans les jours précédant les réunions du Federal Open Market Committee (FOMC), suivie d'une hausse marquée immédiatement après les annonces, surtout lorsque les décisions monétaires diffèrent des attentes. Les hausses de taux inattendues suscitent des réactions plus fortes que les baisses, ce qui suggère une asymétrie comportementale cohérente avec une aversion aux pertes. À l'aide de modèles GARCH, l'auteur capture les dynamiques de volatilité avec une précision temporelle satisfaisante, bien que l'usage de données journalières limite l'analyse des ajustements intrajournaliers, particulièrement pertinents en cas de chocs de grande ampleur.

Rigobon et Sack (2001) adoptent une approche méthodologiquement distincte en recourant à une identification fondée sur l'hétéroscléasticité pour isoler l'impact causal des annonces du FOMC. Leur étude révèle qu'un relèvement inattendu de 25 points de base entraîne une baisse de 1,9 % de l'indice S&P 500, avec des effets encore plus marqués sur le Nasdaq. Leur stratégie repose sur l'hypothèse que la variance des chocs monétaires peut être distinguée de celle d'autres événements de marché. Toutefois, cette hypothèse suppose que les autres facteurs macroéconomiques restent constants autour de l'annonce, une condition difficile à vérifier empiriquement, ce qui introduit un risque d'endogénéité. Malgré cette réserve, leur approche permet une estimation plus fine des effets directionnels et instantanés des décisions monétaires.

Les deux études s'accordent à montrer qu'une fréquence élevée d'annonces de politique monétaire, en particulier lorsqu'elles sont inattendues, engendre une hausse significative de la volatilité. Cette observation suggère un lien potentiel avec une kurtosis plus élevée, résultant d'un épaisseissement des queues de distribution. En outre, la sensibilité accrue des marchés aux hausses de taux, plus marquée que celle observée lors des baisses, souligne une asymétrie dans les réactions. Ce constat est compatible avec la théorie des perspectives, selon laquelle les pertes provoquent une détresse émotionnelle plus intense que les gains de même amplitude. Sur le plan méthodologique, les deux travaux illustrent des choix contrastés. Bomfim privilégie une modélisation dynamique de la volatilité à l'aide des modèles GARCH, tout en restant contraint par la fréquence journalière des données. Rigobon et Sack, quant à eux, mobilisent une approche haute fréquence plus adaptée à la capture des chocs intrajournaliers, mais reposent sur des hypothèses restrictives qui en limitent la validité empirique. Cette tension méthodologique illustre les compromis auxquels sont confrontés les chercheurs, partagés entre la recherche de précision temporelle et la robustesse des hypothèses, dans l'analyse de l'effet des annonces monétaires sur les rendements.

Néanmoins, aucune de ces études n'évalue directement l'impact des annonces sur les moments d'ordre supérieur que sont la skewness ou la kurtosis. Si elles suggèrent une intensification du risque de queue et une possible asymétrie des rendements, elles n'en fournissent pas de mesure explicite. Cette absence limite leur applicabilité directe dans une recherche portant précisément sur les déformations des distributions de rendements induites par la fréquence informationnelle. Malgré leurs approches divergentes, les deux contributions convergent sur le rôle central des annonces de politique monétaire, en particulier en période de resserrement. Les résultats indiquent que les décisions soudaines et inattendues, notamment les hausses de taux, ont tendance à accroître la volatilité, à favoriser le risque de queue et à accentuer les asymétries dans la distribution des rendements. Ces effets soulignent l'importance des chocs monétaires comme vecteurs de transmission des incertitudes macroéconomiques et leur capacité à modifier durablement les propriétés statistiques des actifs financiers.

3.8 Intégrer la skewness et la kurtosis dans la gestion de portefeuille: repenser le risque dans un environnement influencé par l'information

L'inclusion des moments d'ordre supérieur, tels que la skewness et la kurtosis, dans les modèles de valorisation des actifs et les stratégies de gestion des risques apparaît de plus en plus nécessaire, notamment dans les environnements marqués par des rendements extrêmes. Plusieurs travaux empiriques récents mettent en évidence leur valeur ajoutée en matière de prévision et d'allocation. Amaya et al. (2015), à partir de données intrajournalières issues de la base TAQ couvrant la période de 1993 à 2008, montrent que les actifs affichant une skewness réalisée élevée tendent à générer des rendements futurs plus faibles, tandis qu'une kurtosis réalisée élevée est associée à des rendements futurs plus élevés. Ces relations demeurent significatives après contrôle par les facteurs de Fama et French, ce qui renforce la pertinence de ces moments comme variables explicatives des performances futures et justifie leur intégration dans les décisions d'allocation d'actifs.

Dans une optique complémentaire, Jondeau et Rockinger (2000) utilisent un modèle GARCH avec résidus suivant une loi de Student pour estimer la dynamique temporelle de la skewness et de la kurtosis. Leurs résultats révèlent que ces moments ne sont ni constants ni indépendants. Ils varient dans le temps et présentent des co-mouvements entre classes d'actifs, ce qui souligne la possibilité de propagation de chocs extrêmes à travers les marchés. Ces observations

plaident en faveur d'une gestion des risques dynamique, capable de dépasser les indicateurs standards tels que la Value-at-Risk, en intégrant les propriétés temporelles des distributions de rendements.

Bien que fondés sur des méthodologies différentes, à savoir l'utilisation de données haute fréquence pour Amaya et al. et de séries temporelles pour Jondeau et Rockinger, ces deux travaux convergent sur l'utilité opérationnelle des moments d'ordre supérieur. Cette convergence est renforcée par les travaux de Lahbous et Zammar (2021), qui testent l'effet de la skewness et de la kurtosis dans une approche d'optimisation multi-objectifs appliquée à la Bourse de Casablanca sur entre 2010 à 2019. Leur étude montre que maximiser la skewness favorise l'occurrence de rendements positifs, tandis que minimiser la kurtosis réduit l'exposition aux risques extrêmes. Toutefois, la spécificité du marché étudié, en l'occurrence un marché émergent, limite la généralisabilité de leurs conclusions aux places financières plus développées.

Dans un cadre plus généralisable, Lezmi et al. (2019) proposent une stratégie d'allocation fondée sur la skewness, à l'aide de modèles de type mélange gaussien et jump-diffusion. Ils mettent en évidence que les approches traditionnelles, telles que le risk parity, sous-estiment les risques extrêmes. À l'inverse, l'intégration conjointe de la skewness et de mesures de risque comme l'Expected Shortfall permet des allocations plus robustes, mieux adaptées aux périodes de forte instabilité. Leur approche offre ainsi une alternative prometteuse pour les décideurs cherchant à renforcer la résilience des portefeuilles face aux chocs de marché.

Dans l'ensemble, ces études s'accordent sur l'intérêt d'inclure la skewness et la kurtosis dans l'évaluation et la gestion des risques financiers, bien qu'elles divergent dans leur champ d'application. Lahbous et Zammar s'intéressent aux marchés émergents, tandis que Lezmi et al. proposent une méthodologie à portée plus universelle. Il convient toutefois de souligner que seuls Amaya et al. établissent un lien empirique direct entre moments d'ordre supérieur et rendements futurs, ce qui révèle une lacune dans la validation empirique des approches d'allocation fondées sur la skewness et la kurtosis. Malgré cette limite, ces travaux constituent un socle conceptuel robuste pour étendre les modèles classiques d'évaluation du risque en y intégrant les caractéristiques distributionnelles des rendements.

4. Méthodologie

4.1 Collecte de Données

Les données utilisées dans cette étude ont été collectées via deux plateformes principales : Wharton Research Data Services (WRDS) et RavenPack. L'étude couvre la période allant du 3 janvier 2000 au 26 août 2015, et repose sur un échantillon de 545 entreprises cotées sur le marché boursier Américain.

Les données financières ont été extraites à partir des bases bien établies de WRDS, notamment CRSP et Compustat. Plus précisément, les variables suivantes ont été collectées : les rendements boursiers journaliers (RET), le prix acheteur le plus bas de la journée (BIDLO), le prix vendeur le plus élevé de la journée (ASKHI), le nombre d'actions en circulation (SHROUT), ainsi que le prix de clôture ajusté quotidien (PRCCD).

En complément, les rendements quotidiens du S&P 500, ainsi que les valeurs journalières des cinq facteurs de Fama-French et du facteur Momentum, ont été intégrés à l'analyse.

Les données relatives aux nouvelles, initialement prévues pour être obtenues via Factiva, ont finalement été collectées via RavenPack en raison des restrictions récentes de Factiva, qui ont rendu l'extraction automatisée des données impossible.

Les informations extraites incluent le nombre total de nouvelles publiées quotidiennement (RP_NEWS_COUNT) pour l'ensemble des 545 entreprises, sur la période 2000-2015, soit 153 582 nouvelles au total. Aucune distinction n'a été faite quant à la tonalité des nouvelles (positive, négative, neutre) ni quant à leur catégorie (macroéconomique, politique, monétaire, etc.).

Enfin, une correspondance entre les identifiants CUSIP des entreprises sur WRDS et RavenPack a été effectuée afin d'agréger et d'harmoniser l'ensemble des données collectées.

La période d'analyse a été choisie stratégiquement car elle couvre plusieurs cycles économiques et événements majeurs, tels que l'éclatement de la bulle Internet (2000-2002), la crise financière mondiale de 2008, ainsi que la période de reprise post-crise.

4.2 Variables dépendantes

La caractérisation statistique des rendements de portefeuille repose sur le calcul des principaux moments de leur distribution: la moyenne arithmétique, l'écart type, la skewness et la kurtosis. Les formulations standards non pondérés de ces moments sont présentées ci-dessous, conformément à Press et al. (1992):

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad (1)$$

$$\sigma(x_1 \dots x_N) = \sqrt{\text{Var}(x_1 \dots x_N)} \quad (2)$$

$$\text{Skew}(x_1 \dots x_N) = \frac{1}{N} \sum_{j=1}^N \left(\frac{x_j - \bar{x}}{\sigma} \right)^3 \quad (3)$$

$$\text{Kurt}(x_1 \dots x_N) = \left\{ \frac{1}{N} \sum_{j=1}^N \left[\frac{x_j - \bar{x}}{\sigma} \right]^4 \right\} - 3 \quad (4)$$

Bien que ces formules aient été introduites à une époque antérieure de la statistique computationnelle, elles demeurent largement utilisées dans l'économétrie financière contemporaine et constituent la base conceptuelle des calculs effectués au niveau des portefeuilles dans la présente étude.

Dans les calculs préliminaires, une pondération par la capitalisation boursière a été appliquée afin d'assurer une représentation fidèle de l'importance relative de chacune des entreprises au sein des portefeuilles. Cette approche est conforme aux méthodologies utilisées dans la construction des indices boursiers de référence, telles que décrites par l'Autorité des marchés financiers (AMF) dans *Le marché des indices financiers: opportunités et risques* (Grillet-Aubert, 2020).

La capitalisation boursière $M_{i,j}$ d'une entreprise i à une date j est définie comme le produit entre le nombre d'actions en circulation et le prix de clôture ajusté :

$$M_{i,j} = \text{SHROUT}_{i,j} \times \text{PRCCD}_{i,j} \quad (5)$$

où:

- $SHROUT_{i,j}$ représente le nombre d'actions en circulation de l'entreprise i au jour j .
- $PRCCD_{i,j}$ correspond au prix de clôture ajusté de l'action de l'entreprise i au jour j .

Le cours de clôture ajusté (PRCCD) est utilisé à la place du cours de clôture brut afin de tenir compte des opérations sur titres telles que les fractionnements d'actions et le versement de dividendes. Cet ajustement permet d'assurer une représentation plus précise et cohérente de la capitalisation boursière dans le temps.

À partir de la capitalisation boursière, un poids relatif $W_{i,j}$ est attribué à chaque entreprise au sein de son portefeuille en divisant sa capitalisation boursière par la somme des capitalisations boursières de toutes les entreprises appartenant au même portefeuille:

$$W_{i,j} = \frac{M_{i,j}}{\sum_{i \in p} M_{i,j}} \quad (6)$$

Le rendement journalier pondéré $R_{p,j}$ d'un portefeuille p au jour j est ensuite obtenu en effectuant une moyenne pondérée des rendements journaliers des entreprises du portefeuille :

$$R_{p,j} = \frac{\sum_{i \in p} W_{i,j} \cdot R_{i,j}}{\sum_{i \in p} W_{i,j}} \quad (7)$$

où:

- $R_{i,j}$ représente le rendement boursier journalier de l'entreprise i au jour j .

Afin de réduire le bruit de haute fréquence tout en conservant les tendances pertinentes dans le comportement des rendements, une fréquence semestrielle a été retenue. Cet horizon intermédiaire permet de trouver un équilibre entre la volatilité à court terme des données quotidiennes ou trimestrielles et le lissage excessif lié à une agrégation annuelle. Une fréquence annuelle aurait en

effet pu masquer des évolutions importantes survenant au cours de l'année, telles que les ajustements rapides des rendements induits par des évènements macroéconomiques, financiers ou informationnels identifiés dans la revue de littérature.

Ainsi, le **rendement moyen arithmétique semestriel** $\bar{R}_{p,t}$ du portefeuille p pour le semestre t est calculé comme suit :

$$\bar{R}_{p,t} = \frac{1}{N_t} \sum_{j \in t} R_{p,j} \quad (8)$$

où:

- N_t est le nombre total de jours dans le semestre t.

La **volatilité semestrielle** $\sigma_{p,t}$ du portefeuille p est ensuite obtenu à partir de la racine carrée de la variance des rendements journaliers :

$$\sigma_{p,t} = \sqrt{\frac{1}{N_t} \sum_{j \in t} (R_{p,j} - \bar{R}_{p,t})^2} \quad (9)$$

Afin de capturer les caractéristiques d'ordre supérieur de la distribution des rendements, nous calculons les troisième et quatrième moments centrés réduits, à savoir la skewness et la kurtosis. Ces mesures sont obtenues à partir des formules standards, pondérées par la capitalisation boursière, et tiennent également compte de l'agrégation des rendements selon la fréquence semestrielle choisie.

La **skewness semestrielle** $S_{p,t}$ du portefeuille p au semestre t est calculée à partir de la formule classique du troisième moment centré :

$$S_{p,t} = \frac{\frac{1}{N_t} \sum_{j \in t} (R_{p,j} - \bar{R}_{p,t})^3}{\sigma_{p,t}^3} \quad (10)$$

La **kurtosis semestrielle** $K_{p,t}$ du portefeuille p au semestre t, corrigée de la valeur de référence de 3 associée à la distribution normale, est calculée comme suit :

$$K_{p,t} = \frac{\frac{1}{N_t} \sum_{j \in t} (R_{p,j} - \bar{R}_{p,t})^4}{\sigma_{p,t}^4} - 3 \quad (11)$$

4.3 Variables indépendantes

Les variables indépendantes présentées dans cette section ont été sélectionnées sur la base des résultats empiriques des études précédemment abordées dans la revue de littérature. Nous détaillons ici la pertinence de chaque variable ainsi que la méthodologie utilisée pour leur calcul.

Inspirés par les travaux de Jeon, McCurdy et Zhao (2021) dans leur étude *News as Sources of Jumps in Stocks*, nous adoptons leur approche consistant à mesurer la fréquence des nouvelles à partir du nombre total d'articles publiés quotidiennement sur chaque entreprise. Toutefois, dans le cadre de cette étude, nous proposons une adaptation méthodologique afin d'analyser l'impact de l'information sur un horizon temporel plus long. Nous définissons ainsi une fréquence moyenne semestrielle $F_{i,t}$, calculée comme suit :

$$F_{i,t} = \frac{\text{total number of news articles published for firm } i \text{ during semester } t}{\text{number of days in semester } t} \quad (12)$$

Dans cette étude, la principale variable explicative, désignée par **MeanFreq_{p, t}**, représente la moyenne pondérée des fréquences semestrielles des nouvelles pour les entreprises composant le portefeuille p au cours du semestre t. Cette variable est construite de manière à refléter la composition effective du portefeuille en tenant compte des poids relatifs des entreprises au cours du semestre considéré, tout en s'appuyant sur la fréquence des nouvelles observée au semestre précédent $F_{i,t-1}$, qui sert de critère de sélection pour l'inclusion dans le portefeuille. Elle est calculée selon la formule suivante :

$$\text{MeanFreq}_{p,t} = \frac{\sum_{i \in p} W_{i,t} \cdot F_{i,t-1}}{\sum_{i \in p} W_{i,t}} \quad (13)$$

où:

- $F_{i,t-1}$ correspond à la fréquence moyenne des nouvelles publiées au sujet de l'entreprise i au cours du semestre précédent $t - 1$, utilisée comme critère d'inclusion dans le portefeuille p au semestre t.
- $W_{i,t}$ représente le poids relatif de l'entreprise i dans le portefeuille p au cours du semestre t, déterminé à partir de sa capitalisation boursière actuelle et calculé comme suit :

$$W_{i,t} = \frac{\bar{M}_{i,t}}{\sum_{i \in p} \bar{M}_{i,t}} \quad (14)$$

avec:

- $\bar{M}_{i,t}$ correspondant à la capitalisation boursière moyenne de l'entreprise i au cours du semestre t et définie comme suit :

$$\bar{M}_{i,t} = \frac{\sum_{j \in t} M_{i,j}}{N_t} \quad (15)$$

où:

- $M_{i,j}$ représente la capitalisation boursière de l'entreprise i au jour j du semestre t .
- N_t représente le nombre total de jours composant le semestre t .

Bien que la variable $\text{MeanFreq}_{p,t}$ permette de capturer l'intensité de la couverture médiatique des entreprises dans cette étude, il est important de noter qu'elle présente plusieurs limites méthodologiques. D'une part, elle ne tient pas compte de la pertinence informationnelle, du contenu ni de la tonalité des nouvelles publiées. Elle se limite à quantifier la fréquence brute des mentions médiatiques des entreprises de l'échantillon, sans distinguer les articles véritablement informatifs de ceux qui sont redondants ou anecdotiques. D'autre part, un même article peut être associé à plusieurs entreprises lorsqu'elles sont mentionnées conjointement, ce qui peut entraîner une surestimation de la fréquence agrégée. Par ailleurs, le choix d'un horizon semestriel, retenu pour garantir une cohérence temporelle avec les variables dépendantes analysées (skewness, kurtosis, volatilité et rendement moyen), repose sur la volonté de réduire le bruit de haute fréquence tout en conservant les tendances structurelles. Ce compromis permet de trouver un équilibre entre la volatilité des données de haute fréquence et le lissage excessif que produirait une agrégation annuelle. Cependant, cet horizon relativement long peut atténuer certains chocs informationnels de court terme et ne permet pas de distinguer la nouveauté des nouvelles qui pourrait être un aspect potentiellement essentiel lorsque l'on souhaite évaluer l'effet marginal du flux informationnel.

Berry et Howe (1994) montrent que les grandes entreprises, ou celles bénéficiant d'une forte visibilité médiatique et d'une couverture analytique importante, ont tendance à réagir plus fortement aux nouvelles. Chan (2001) confirme ces résultats en soulignant que les petites

entreprises réagissent généralement plus lentement aux nouvelles que les grandes, en particulier lorsque les nouvelles sont négatives.

Afin de contrôler les effets potentiels de la taille des entreprises sur la distribution des rendements, il est pertinent d'inclure la capitalisation boursière comme variable de contrôle. Ainsi, nous définissons la variable $M_{p,t}$, qui représente la moyenne pondérée des capitalisations boursières moyennes semestrielles des entreprises appartenant à chaque portefeuille p . Cette variable est calculée selon la formule suivante :

$$M_{p,t} = \frac{\sum_{i \in p} W_{i,t} \cdot \bar{M}_{i,t}}{\sum_{i \in p} W_{i,t}} \quad (16)$$

Afin d'assurer une meilleure comparabilité entre les portefeuilles et d'atténuer l'influence des très grandes entreprises, nous utilisons le logarithme de $M_{p,t}$:

$$\log(M_{p,t})$$

Fang et Peress (2008) montrent que l'effet médiatique est particulièrement marqué pour les actions de petites capitalisations et celles présentant un écart bid-ask élevé. Leurs résultats suggèrent que les titres caractérisés par un écart important, et donc une liquidité plus faible, sont les plus sensibles à la couverture médiatique, ce qui accroît leur réactivité aux chocs informationnels. Parmi les quatre mesures de liquidité analysées dans leur étude, l'écart bid-ask et le volume d'échange sont identifiés comme les principaux indicateurs de l'effet médiatique. Inspirés par leur méthodologie, nous incluons une variable de contrôle pour la liquidité, notée $\text{Liquidity}_{p,t}$, afin d'isoler son effet spécifique sur la distribution des rendements. Cette variable correspond à la moyenne pondérée des écarts bid-ask moyens des entreprises composant le portefeuille p durant le semestre t , et est calculée comme suit :

$$\text{Liquidity}_{p,t} = \frac{\sum_{i \in p} W_{i,t} \cdot \overline{\text{Spread}}_{i,t}}{\sum_{i \in p} W_{i,t}} \quad (17)$$

où:

- $\overline{\text{Spread}}_{i,t}$ correspond à l'écart bid-ask moyen quotidien de l'entreprise i sur l'ensemble du semestre t , calculé comme suit :

$$\overline{\text{Spread}_{i,t}} = \frac{\sum_{j \in t} \text{Spread}_{i,j}}{N_t} \quad (18)$$

avec:

- $\text{Spread}_{i,j}$ représentant l'écart bid-ask quotidien de l'entreprise i au jour j du semestre t , défini comme suit :

$$\text{Spread}_{i,j} = \frac{\text{Ask}_{i,j} - \text{Bid}_{i,j}}{\frac{\text{Ask}_{i,j} + \text{Bid}_{i,j}}{2}} \quad (19)$$

où:

- $\text{Ask}_{i,j}$ correspond au prix vendeur de l'action i observé au jour j , c'est-à-dire le prix le plus bas auquel un investisseur est prêt à vendre l'action.
- $\text{Bid}_{i,j}$ correspond au prix acheteur de l'action i observé au jour j , c'est-à-dire le prix le plus élevé auquel un investisseur est prêt à acheter l'action.

Amaya, Christoffersen, Jacobs et Vasquez (2008) ont montré que les actifs ayant un bêta élevé tendent à présenter des variations plus marquées de la skewness et de la kurtosis réalisées. Cela suggère une influence structurelle de ce facteur sur la distribution des rendements boursiers. L'inclusion du bêta dans notre analyse se justifie donc par son rôle fondamental dans la mesure de la sensibilité des actifs aux chocs de marché. Dans cette étude, nous calculons un bêta de portefeuille, noté $\beta_{p,t}$, qui correspond à la moyenne pondérée des bêtas individuels des entreprises composant le portefeuille p au cours du semestre t . Ce bêta est défini par la relation suivante :

$$\beta_{p,t} = \frac{\sum_{i \in p} W_{i,t} \cdot \beta_i}{\sum_{i \in p} W_{i,t}} \quad (20)$$

où:

- β_i représente le bêta de l'entreprise i sur l'ensemble de la période d'analyse. Il est estimé en régressant les rendements journaliers de l'entreprise sur ceux du S&P 500, conformément à la relation linéaire suivante :

$$R_{i,j} = \alpha + \beta_i R_{SP500,j} + \epsilon_{i,j} \quad (21)$$

avec:

- $R_{i,j}$: le rendement quotidien de l'entreprise i au jour j .
- $R_{SP500,j}$: le rendement quotidien du S&P 500 au jour j .

- α : l'ordonnée à l'origine de la régression.
- β_i : la sensibilité des rendements de l'entreprise aux variations du S&P 500.
- $\varepsilon_{i,j}$: le terme d'erreur capturant les variations spécifiques à l'entreprise non expliquées par le marché.

Bien que l'estimation théorique du bêta dans le modèle CAPM repose sur les rendements excédentaires, nous avons opté pour les rendements bruts en raison de la disponibilité partielle du taux sans risque journalier sur la période étudiée (2000–2015). Par ailleurs, cette approximation est fréquente dans la littérature empirique, notamment lorsque l'horizon d'analyse est long ou que les taux sans risque sont faibles et relativement stables à court terme.

Le bêta de chaque entreprise est ainsi estimé à l'aide de la formule standard suivante :

$$\beta_i = \frac{\text{Cov}(R_i, R_{\text{SP500}})}{\text{Var}(R_{\text{SP500}})} \quad (22)$$

où:

- $\text{Cov}(R_i, R_{\text{SP500}})$ désigne la covariance entre les rendements de l'entreprise i et ceux du S&P 500.
- $\text{Var}(R_{\text{SP500}})$ représente la variance des rendements du S&P 500.

Estimer le bêta de chaque entreprise sur l'ensemble de la période d'analyse (15 ans) revient à supposer une stabilité temporelle du risque systématique, ce qui est peu réaliste. Afin d'atténuer ce biais, la variable $\beta_{p,t}$ a été standardisée dès le début de l'analyse, devenant ainsi $\beta_{p,t_standardized}$, de sorte à recentrer sa distribution et rendre ses variations relatives plus comparables. Cette approche permet de corriger partiellement l'instabilité du bêta dans le temps et d'améliorer la robustesse des résultats des régressions ultérieures.

Fama et French (2015) ont démontré que l'ajout de facteurs supplémentaires au modèle traditionnel d'évaluation des actifs financiers (CAPM) améliore significativement le pouvoir explicatif des rendements boursiers. Par ailleurs, Carhart (1997) a souligné l'importance du facteur de momentum dans la dynamique à court terme des rendements. Dans cette étude, l'inclusion des cinq facteurs de Fama-French (Excess Return on the Market, Small Minus Big, High Minus Low, Robust Minus

Weak, Conservative Minus Aggressive) ainsi que du facteur Momentum vise à capter les effets structurels du marché tout en contrôlant les caractéristiques propres aux entreprises, telles que la taille, la rentabilité, la politique d'investissement et le momentum. Ces facteurs influencent la sensibilité des actifs aux chocs de marché et, par conséquent, leur réaction face aux nouvelles. En conséquence, chacun de ces six facteurs est intégré au modèle, et nous calculons leurs moyennes semestrielles respectives selon la formule suivante :

$$\text{Factor}_t = \frac{\sum_{j \in t} \text{Factor}_j}{N_t} \quad (23)$$

où:

- Factor_j correspond à la valeur quotidienne du facteur au jour j.

4.4 Construction des portefeuilles

Dans le cadre de cette étude, les portefeuilles ont été construits en fonction de la fréquence moyenne semestrielle des articles de presse publiés pour chaque entreprise au cours du semestre précédent, notée $F_{i,t-1}$. L'objectif est d'examiner dans quelle mesure la couverture médiatique passée influence la distribution actuelle des rendements boursiers.

Cependant, une faible dispersion des fréquences calculées a été observée, de nombreuses entreprises recevant peu voire aucune couverture médiatique. Afin d'assurer une segmentation pertinente et d'éviter une concentration excessive dans certaines catégories, nous nous sommes inspirés de la méthodologie d'Antweiler et Frank (2002) dans leur étude *Internet Stock Message Boards and Stock Returns*. Contrairement à leur classification en quintiles, la distribution effective de nos données nous a conduits à adopter une répartition en quartiles, basée sur $F_{i,t-1}$.

Ainsi, au début de chaque semestre, les entreprises sont triées selon leur fréquence moyenne de nouvelles du semestre précédent, $F_{i,t-1}$, puis réparties dans cinq portefeuilles distincts, définis comme suit :

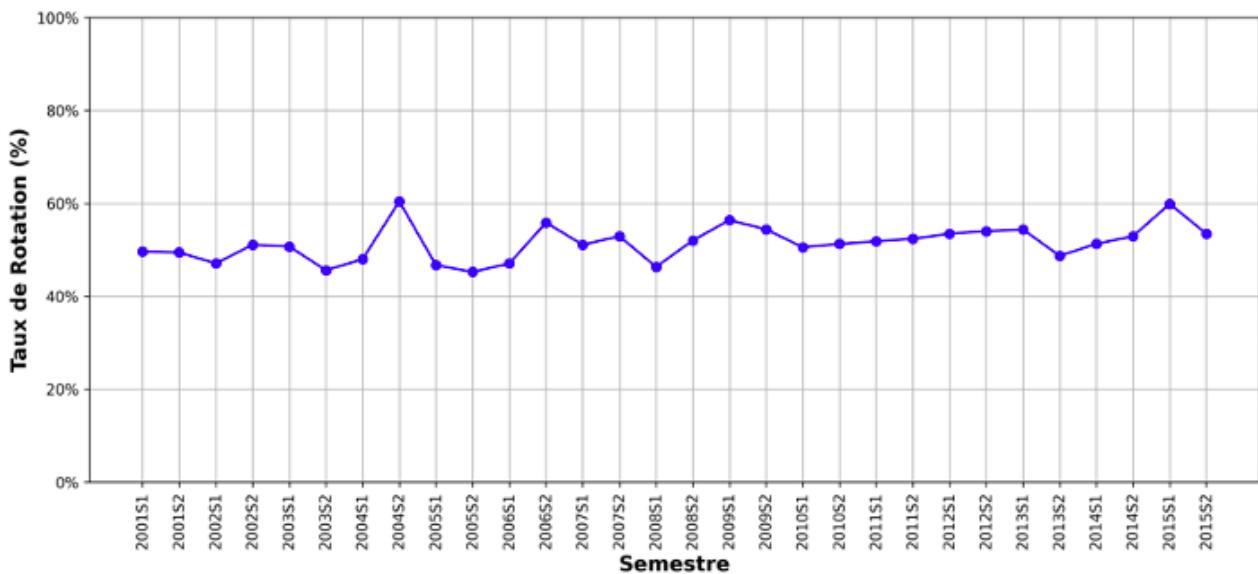
Le portefeuille Zéro qui regroupe les entreprises pour lesquelles aucun article de presse n'a été publié durant le semestre précédent ($F_{i,t-1} = 0$). Ce portefeuille servira ultérieurement de groupe de référence dans l'analyse, afin d'évaluer l'impact d'une absence totale de couverture médiatique.

Les quatre autres portefeuilles sont définis en fonction des quartiles de la distribution de la fréquence moyenne des nouvelles ($F_{i,t-1}$). Le portefeuille Q1 regroupe les entreprises appartenant au premier quartile, caractérisé par une couverture médiatique faible mais non nulle. Le portefeuille Q2 contient celles du deuxième quartile, correspondant à une couverture médiatique modérée. Le portefeuille Q3 inclut les entreprises du troisième quartile, ayant bénéficié d'une fréquence relativement élevée de nouvelles. Enfin, le portefeuille Q4 comprend les entreprises du quartile supérieur, c'est-à-dire celles ayant reçu le niveau de couverture médiatique le plus élevé durant le semestre précédent.

Cette classification, mise à jour à chaque semestre, assure un rééquilibrage dynamique des portefeuilles. En procédant de cette manière, la formation des portefeuilles repose uniquement sur des informations passées, ce qui renforce la dimension prédictive du modèle tout en évitant tout biais d'anticipation.

Il est important de souligner que la construction des portefeuilles repose exclusivement sur $F_{i, t-1}$, et non sur la variable explicative principale MeanFreq_{p, t} utilisée dans l'analyse économétrique. Cette dernière est calculée après la formation des portefeuilles et correspond à la moyenne pondérée des fréquences de nouvelles des entreprises composant chaque portefeuille, comme indiqué dans l'**Équation (13)**.

Figure 1 – Taux de rotation des portefeuilles par semestre



La Figure 1 illustre l'évolution semestrielle du taux de rotation des portefeuilles sur la période d'analyse, allant du premier semestre de l'année 2001 (2001S1) au second semestre de l'année 2015 (2015S2). Le taux de rotation indique la proportion d'entreprises (parmi les 545 de l'échantillon) qui changent de portefeuille d'un semestre à l'autre, reflétant ainsi la dynamique ou la stabilité des classifications au fil du temps. Chaque point représente le taux de rotation calculé pour un semestre donné et est relié par une ligne bleue afin de faciliter l'observation des fluctuations dans le temps.

Sur l'ensemble de la période d'étude, ce taux fluctue autour de 50 %, ce qui indique qu'en moyenne, près d'une entreprise sur deux ne reste pas dans le même portefeuille d'un semestre à l'autre. Le taux de rotation le plus élevé est observé au second semestre de 2004 (2004S2), avec 60.4 % des entreprises changeant de portefeuille par rapport au premier semestre de la même année.

À l'inverse, le second semestre de 2005 (2005S2) présente le taux de rotation le plus faible, avec seulement 45.2 % des entreprises ayant changé de portefeuille par rapport à la période précédente.

Tableau 1 - Matrice de transition des portefeuilles

Ptf. actuel	Zero	Q1	Q2	Q3	Q4
Ptf. précédent					
Zero	0.710	0.184	0.070	0.028	0.009
Q1	0.144	0.445	0.249	0.121	0.039
Q2	0.044	0.250	0.346	0.258	0.102
Q3	0.014	0.133	0.259	0.351	0.242
Q4	0.003	0.034	0.095	0.250	0.618

Le tableau 1 présente les probabilités de transition des entreprises entre les portefeuilles Zero, Q1, Q2, Q3 et Q4 sur l'ensemble de la période d'analyse. Les lignes indiquent le portefeuille occupé par les entreprises au semestre précédent, tandis que les colonnes représentent leur portefeuille au semestre actuel. Chaque cellule indique la proportion d'entreprises passant d'un portefeuille à un autre. Les valeurs diagonales correspondent à la proportion d'entreprises restées dans le même portefeuille d'un semestre à l'autre. Note : « Ptf. » désigne l'abréviation de « portefeuille ».

L'une des premières observations que l'on peut faire est que les valeurs situées sur la diagonale sont les plus élevées, ce qui indique que les entreprises ont tendance à rester dans le même portefeuille d'un semestre à l'autre. Avec 71.0 % des entreprises du portefeuille Zéro qui y demeurent le semestre suivant, ce portefeuille apparaît comme le plus stable des cinq, suivi de près par le portefeuille Q4 avec 61.8 %. En revanche, les portefeuilles intermédiaires (Q1, Q2 et Q3) connaissent des transitions plus fréquentes. Par exemple, 25.9 % des entreprises classées en Q3 au semestre précédent passent dans le portefeuille Q2 au semestre suivant.

Tableau 2 - Statistiques descriptives de MeanFreqp,t, variable explicative principale

Portefeuille	Effectif	Moyenne	Écart-type	Min	25%	Médiane	75%	Max
Zero	31.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q1	31.000	0.022	0.009	0.008	0.015	0.024	0.028	0.039
Q2	31.000	0.049	0.023	0.014	0.024	0.055	0.069	0.078
Q3	31.000	0.086	0.039	0.028	0.047	0.097	0.123	0.138
Q4	31.000	0.263	0.046	0.188	0.229	0.261	0.295	0.354

Le tableau 2 présente les statistiques descriptives de MeanFreqp,t, la variable explicative principale qui mesure la fréquence des nouvelles pour chaque portefeuille à chaque semestre, sur l'ensemble de la période d'analyse.

Comme prévu, toutes les valeurs sont nulles pour le portefeuille Zéro, ce qui est cohérent avec sa construction, puisqu'il ne comprend que des entreprises n'ayant reçu aucune couverture médiatique sur l'ensemble de la période d'analyse. Pour les autres portefeuilles, on observe une tendance clairement croissante de la fréquence moyenne des nouvelles à mesure que l'on passe du portefeuille Q1 au portefeuille Q4. Les valeurs moyenne et médiane augmentent de 0.022 et 0.024 pour Q1 à 0.263 et 0.261 pour Q4, ce qui représente une multiplication par dix. L'écart-type suit également cette tendance, indiquant une dispersion plus importante à mesure que la fréquence des nouvelles augmente : il passe de 0.009 pour Q1 à 0.023 pour Q2, puis à 0.039 pour Q3, et enfin à 0.046 pour Q4. Cette évolution se reflète également dans l'étendue interquartile, qui s'élargit à mesure que les entreprises bénéficient d'une couverture médiatique plus importante. Enfin, les valeurs extrêmes confirment la classification en portefeuilles : le portefeuille Q4 regroupe les valeurs les plus élevées, avec un minimum de 0.188, tandis que le portefeuille Q1, composé des entreprises les moins couvertes, présente une valeur maximale de seulement 0.039.

4.5 Présentation des hypothèses

Étant donné que la relation entre la fréquence des nouvelles et la skewness ou la kurtosis n'a, à ce jour, jamais été explorée de manière directe et approfondie, les hypothèses initiales et les plus importantes de cette étude ont été formulées de façon intuitive, tout en s'appuyant sur les résultats empiriques présentés dans la revue de littérature.

À l'exception de quelques études, la grande majorité des résultats empiriques examinés dans la littérature semblent appuyer l'idée qu'une augmentation de la fréquence des nouvelles conduit à une diminution de la skewness et à une augmentation de la kurtosis des rendements. Par exemple, Jeon, McCurdy et Zhao (2021) ont montré que les entreprises bénéficiant d'une couverture médiatique plus importante sont davantage sujettes à des variations de prix brutales, appelées « sauts ». En d'autres termes, une forte exposition médiatique tend à accroître la probabilité de mouvements soudains, aussi bien à la hausse qu'à la baisse. Un tel phénomène peut donc affecter la distribution des rendements en accentuant l'asymétrie et en augmentant la fréquence des mouvements extrêmes. D'une part, puisque cette étude n'intègre pas l'analyse de sentiment des nouvelles, et étant donné que les investisseurs ont tendance à réagir plus fortement aux mauvaises nouvelles qu'aux bonnes (comme le souligne la revue de littérature), il est raisonnable de supposer qu'une attention médiatique accrue pourrait entraîner une prédominance des baisses de prix. Cela se traduirait par une distribution des rendements asymétrique vers la gauche, impliquant une skewness négative et donc un risque accru de pertes extrêmes. D'autre part, en augmentant la probabilité de sauts, une couverture médiatique plus intense pourrait également amplifier les écarts par rapport aux rendements moyens, ce qui se traduirait par une distribution à queues épaisses.

En conséquence, les deux premières hypothèses fondamentales de cette étude sont formulées comme suit :

- $H_{0;1}$: La fréquence des nouvelles n'a pas d'effet ou a un effet positif sur la skewness des rendements boursiers.
- **$H_{1;1}$: La fréquence des nouvelles a un effet négatif sur la skewness des rendements boursiers.**

- $H_{0;2}$: La fréquence des nouvelles n'a pas d'effet ou a un effet négatif sur la kurtosis des rendements boursiers.
- **$H_{1;2}$: La fréquence des nouvelles a un effet positif sur la kurtosis des rendements boursiers.**

Contrairement aux deux hypothèses précédemment formulées, qui reposaient principalement sur l'intuition en raison de l'absence de preuves empiriques directes, l'hypothèse suivante s'appuie solidement sur la littérature empirique. En effet, la majorité des études examinées ont démontré qu'une augmentation de la fréquence des nouvelles est associée à une volatilité plus élevée. À titre d'illustration, Antweiler et Frank (2002) montrent que les actions les plus commentées présentent des rendements futurs anormalement faibles et une volatilité nettement plus élevée, ce qui suggère que l'intensité de la couverture médiatique pourrait agir comme un facteur amplificateur des fluctuations de prix. En nous appuyant sur ces résultats, ainsi que sur ceux de Jeon, McCurdy et Zhao (2021), et sur d'autres éléments issus de la littérature, nous formulons l'hypothèse suivante :

- $H_{0;3}$: La fréquence des nouvelles n'a pas d'effet ou a un effet négatif sur la volatilité des rendements boursiers.
- **$H_{1;3}$: La fréquence des nouvelles a un effet positif sur la volatilité des rendements boursiers.**

Les résultats empiriques issus de la littérature sont un peu plus nuancés en ce qui concerne cette relation, certains suggérant une baisse des rendements, d'autres une hausse à mesure que la fréquence des nouvelles augmente. Cependant, en mettant de côté les cas spécifiques au contexte et en tenant compte de la réaction asymétrique du marché face aux mauvaises nouvelles, cette étude adopte l'hypothèse selon laquelle une fréquence plus élevée des nouvelles est associée à des rendements boursiers plus faibles. À titre d'exemple, Fang et Peress (2008) montrent que les actions ne bénéficiant d'aucune couverture médiatique génèrent des rendements excédentaires significativement plus élevés que celles très couvertes, avec un écart moyen de performance de 3 % par an après ajustement pour les facteurs de risque. En nous appuyant sur ces résultats, ceux d'Antweiler et Frank (2002), ainsi que sur un ensemble plus large de preuves empiriques, nous formulons l'hypothèse suivante :

- $H_{0;4}$: La fréquence des nouvelles n'a pas d'effet ou a un effet positif sur le rendement moyen arithmétique.
- **$H_{1;4}$: La fréquence des nouvelles a un effet négatif sur le rendement moyen arithmétique.**

4.6 Modèle économétrique retenu : Régression OLS groupée

Pour tester les hypothèses de l'étude, la régression OLS groupée a été retenue comme le modèle économétrique le plus approprié. Bien que le jeu de données présente une structure en panel à deux dimensions, temporelle (31 semestres) et transversale (5 portefeuilles), la nature des données empêche l'utilisation des modèles standards à effets fixes ou aléatoires.

La régression OLS groupée repose sur l'hypothèse que les coefficients des variables explicatives sont constants à travers les portefeuilles et les semestres. Elle suppose également que les observations sont indépendantes, ce qui est justifiable dans ce contexte en raison du taux élevé de rotation des entreprises au sein des portefeuilles : environ 50 % des entreprises changent de portefeuille d'un semestre à l'autre. Ce taux de rotation empêche donc d'identifier des caractéristiques stables propres à chaque portefeuille sur l'ensemble de la période d'étude.

Les modèles à effets fixes, qui visent à capturer des effets non observables propres aux entités ou aux périodes, ne sont pas appropriés dans notre contexte. Bien que la structure des données soit de type panel, chaque portefeuille n'est observé qu'une seule fois par semestre, soit un total de 155 observations (31 semestres \times 5 portefeuilles). Cette absence d'observations répétées au sein de chaque portefeuille au fil du temps empêche l'estimation des effets fixes de portefeuille, en raison du manque de variation intra-groupe nécessaire à l'identification des coefficients. Dans ce cas, les effets fixes absorberaient toute la variabilité, ce qui entraînerait une colinéarité parfaite avec les variables explicatives. Le même problème se pose pour les effets fixes temporels lorsque la variation entre les semestres est trop faible ou déjà captée par les variables explicatives.

Quant aux modèles à effets aléatoires, ils reposent sur l'hypothèse clé selon laquelle les effets spécifiques aux portefeuilles sont non corrélés avec les variables explicatives. Or, compte tenu du fort taux de rotation des entreprises et de la composition changeante des portefeuilles, cette hypothèse est difficilement défendable dans notre contexte. De plus, l'estimation de la variance des effets aléatoires nécessite également plusieurs observations par portefeuille et par période, ce qui n'est pas le cas ici.

Enfin, bien que des tests économétriques tels que les tests de Hausman, de Breusch-Pagan ou de Fisher soient généralement utilisés pour choisir entre une régression OLS groupée et des modèles

à effets fixes ou aléatoires, leur application serait ici vaine, car l'absence de variabilité intraportefeuille rend ces modèles non estimables. Ainsi, la régression OLS groupée s'impose comme le choix économétrique le plus approprié compte tenu de la structure des données. Elle permet d'estimer de manière directe et cohérente l'effet des variables explicatives sur chacune des variables dépendantes, sans introduire de biais lié à des effets spécifiques non observés.

4.7 Spécification finale du modèle

Afin d'identifier la spécification finale la plus appropriée du modèle, nous avons utilisé une procédure de sélection par élimination rétrograde fondée sur un seuil de significativité de 5 %. Compte tenu du caractère exploratoire de la sélection de variables, qui vise à identifier les relations statistiquement significatives quel que soit le sens de l'effet, une approche bilatérale a été retenue.

Bien que des techniques de régularisation comme LASSO ou Ridge soient efficaces dans des contextes de forte dimensionnalité, elles n'ont pas été utilisées ici, car elles tendent à éliminer un grand nombre de variables explicatives, voire la quasi-totalité dans certains cas, ce qui aurait limité la portée interprétative du modèle. Étant donné le nombre modéré de variables testées et la finalité explicative de l'étude, centrée sur l'évaluation de l'impact de la fréquence des nouvelles (MeanFreq_{p,t}), la méthode de sélection par élimination rétrograde a été privilégiée pour sa transparence et sa meilleure compatibilité avec une lecture économétrique fondée sur les coefficients. Bien que parfois considérée comme exploratoire, cette méthode a été appliquée en complément d'une sélection initiale fondée sur la littérature théorique, les variables du modèle de départ étant toutes justifiées par des travaux antérieurs sur les déterminants de la skewness, de la kurtosis, de la volatilité ou des rendements.

Partant du modèle initial intégrant l'ensemble des variables explicatives, cette méthode consiste à retirer, de manière itérative, la variable présentant la plus forte valeur p à chaque étape, jusqu'à ce que toutes les variables restantes soient statistiquement significatives au seuil retenu.

Dans un premier temps, cette méthode permet de réduire la complexité du modèle tout en conservant les variables apportant une information statistiquement pertinente.

Dans un second temps, elle contribue à limiter les risques de sur-ajustement, renforçant ainsi la robustesse et l'interprétabilité du modèle, ce qui est un enjeu particulièrement important compte tenu du nombre relativement modéré d'observations (155 au total).

Cependant, afin de ne pas exclure notre variable principale d'intérêt, qui constitue le cœur des hypothèses de recherche, la variable MeanFreq_{p,t} a été systématiquement conservée tout au long de la procédure, indépendamment de sa significativité statistique.

Ainsi, le processus d'élimination s'arrête automatiquement dès lors que toutes les variables explicatives restantes, à l'exception de MeanFreq_{p,t}, présentent des valeurs p inférieures au seuil de 5 %.

Cela dit, une exception a été faite pour le modèle expliquant la skewness semestrielle, pour lequel un seuil de 10 % a été retenu. En effet, le maintien du seuil plus strict de 5 % aboutissait à un modèle ne conservant que la variable principale d'intérêt. En relâchant légèrement le critère de sélection pour cette variable dépendante spécifique, l'objectif était de préserver un modèle économiquement interprétable tout en conservant un niveau raisonnable de rigueur statistique.

Tableau 3 - Résultats de la régression OLS groupée: spécification finale après sélection rétrograde (seuil de significativité: valeur p ≤ 10 %)

Variable dépendante:		Semi-Annual Skewness				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	-0.330***	0.093	-3.555	0.001	-0.513	-0.147
MeanFreq _{p,t}	0.015	0.308	0.049	0.961	-0.594	0.625
Liquidity _{p,t}	11.194***	3.082	3.632	0.000	5.105	17.284
SMB Semi-Annual Return	109.700*	57.674	1.902	0.059	-4.257	223.650
R² global	0.117					
Statistique F	6.655					
Valeur p du test F	0.000					

Le Tableau 3 présente les résultats de la régression OLS groupée après application d'une procédure de sélection par élimination rétrograde fondée sur un seuil de significativité de 10 %. La variable dépendante est la skewness semestrielle, et les variables explicatives retenues dans le modèle final sont présentées avec leurs coefficients estimés, erreurs standards, statistiques t, valeurs p et intervalles de confiance. La principale variable d'intérêt, MeanFreq_{p,t}, a été systématiquement conservée tout au long du processus d'élimination, indépendamment de sa significativité statistique. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10 %, 5 % et 1 %.

En complément de la procédure de sélection par élimination rétrograde, l'identification de la spécification finale du modèle s'est également appuyée sur l'évaluation de critères d'information, à savoir l'AIC, le BIC et la log-vraisemblance. L'utilisation conjointe de ces deux approches visait

à renforcer la validité du choix retenu, en s'assurant que la spécification finale optimise à la fois la parcimonie et la qualité d'ajustement aux données. Plusieurs spécifications alternatives, intégrant diverses combinaisons de variables explicatives, ont également été testées. Comme attenu, la log-vraisemblance avait tendance à augmenter, traduisant un meilleur ajustement brut, mais cette amélioration s'accompagnait souvent d'une hausse de l'AIC, signalant que le gain en précision ne compensait pas la complexité additionnelle induite. Ces éléments ont conforté le choix de la spécification finale obtenue par sélection rétrograde comme meilleur compromis empirique.

Le modèle présenté dans le tableau 3 ci-dessus a été utilisé exclusivement pour identifier les variables explicatives les plus informatives. Une fois cette spécification établie, l'ensemble des tests économétriques standards détaillés dans la section suivante ont été réalisés afin de valider les hypothèses fondamentales du modèle de régression OLS groupée. La détection d'une endogénéité a conduit à l'estimation du modèle final à l'aide de la méthode des variables instrumentales en deux étapes (IV-2SLS, voir **Tableau 10**), à partir duquel les inférences et conclusions relatives aux hypothèses de recherche ont été établies. Dans le cadre de ce modèle IV-2SLS, des tests spécifiques portant sur la structure des erreurs ont été menés préalablement, en complément des tests économétriques standards déjà présentés. Aucune anomalie n'ayant été détectée dans le cas de la skewness semestrielle (voir **Tableau 8** et **Tableau 9**), les erreurs standards classiques ont été conservées. En revanche, des erreurs robustes ont été systématiquement appliquées dans les autres modèles portant sur la kurtosis, la volatilité et le rendement moyen arithmétique, dès lors que les diagnostics économétriques en justifiaient l'usage.

4.8 Vérification des hypothèses sous-jacentes du modèle

Nous avons vérifié les principales hypothèses du modèle de régression OLS groupée pour chacune de nos hypothèses de recherche. Afin d'éviter les répétitions, nous nous concentrerons ici uniquement sur la première hypothèse, qui concerne la skewness semestrielle.

D'après Wooldridge (2010), les hypothèses fondamentales des modèles de régression OLS groupée sont les suivantes :

- Exogénéité conditionnelle
- Absence de colinéarité parfaite
- Homoscédasticité conditionnelle
- Absence d'autocorrélation conditionnelle
- Linéarité de la relation modélisée

Les sous-sections suivantes évaluent la pertinence de ces hypothèses dans notre contexte, présentent les tests réalisés pour les vérifier, ainsi que les ajustements éventuellement apportés.

4.8.1 Exogénéité des variables explicatives

L'exogénéité des variables explicatives constitue une hypothèse fondamentale dans le cadre d'un modèle de régression OLS groupée (Wooldridge, 2010). Une variable est dite exogène lorsqu'elle est indépendante du terme d'erreur du modèle, ce qui implique qu'elle n'est pas influencée par des facteurs non observés susceptibles de biaiser l'estimation des coefficients.

À l'inverse, la présence d'une variable endogène conduit à des estimateurs des moindres carrés ordinaires (MCO) biaisés et inconsistants. Cela compromet non seulement la validité des inférences statistiques, mais complique également l'identification de relations causales robustes.

L'endogénéité peut provenir de plusieurs sources, notamment l'omission d'une variable pertinente, une causalité inverse entre la variable explicative et la variable dépendante, ou encore des erreurs de mesure.

Pour détecter une éventuelle endogénéité des variables explicatives dans un modèle de régression OLS groupée, le test de Durbin-Hausman est couramment utilisé. Ce test repose sur la comparaison de deux estimateurs :

- L'estimateur MCO, qui est efficace lorsque les variables sont exogènes, mais biaisé en cas d'endogénéité ;
- L'estimateur des moindres carrés en deux étapes (2SLS), qui demeure consistant même en présence d'endogénéité, bien qu'il soit moins efficace si les variables sont en réalité exogènes.

L'objectif du test est donc d'évaluer si les coefficients estimés par MCO et 2SLS diffèrent significativement. Si les résultats sont similaires, la variable est considérée comme exogène. En revanche, une différence statistiquement significative entre les deux estimations suggère une endogénéité, ce qui implique que les résultats obtenus par MCO sont biaisés.

Sur la base de cette approche, le test de Durbin-Hausman a été appliqué à notre modèle en utilisant sa version fondée sur une régression auxiliaire. La procédure se déroule en plusieurs étapes :

Dans un premier temps, le modèle principal est estimé par moindres carrés ordinaires (MCO) et les résidus sont extraits. Ensuite, pour chaque variable explicative suspectée d'être endogène, une régression auxiliaire est réalisée, dans laquelle cette variable est régressée sur l'ensemble des autres variables explicatives du modèle. L'objectif est d'isoler la composante inexplicée de la variable, c'est-à-dire la partie susceptible d'être corrélée au terme d'erreur du modèle principal. Les résidus issus de cette régression auxiliaire sont alors inclus dans une nouvelle estimation de la skewness semestrielle, en conservant l'ensemble des variables explicatives initiales.

Si le coefficient associé à ces résidus est statistiquement significatif (valeur $p < 5\%$), cela indique l'existence d'une corrélation entre la variable concernée et le terme d'erreur du modèle, confirmant ainsi la présence d'endogénéité. Dans le cas contraire (valeur $p \geq 5\%$), la variable peut être considérée comme exogène.

Tableau 4 - Résultats du test de Durbin-Hausman concernant l'endogénéité de MeanFreqp,t (méthode de régression auxiliaire) dans l'analyse de la skewness semestrielle

Variable dépendante:		Semi-Annual Skewness					
Variables explicatives	Coefficient	Erreur standard	Statistique t	Valeur p	IC inférieur	IC supérieur	
const	2.354	1.187	1.984	0.049	0.009	4.698	
MeanFreqp,t	-18.459	8.174	-2.258	0.025	-34.609	-2.308	
Liquidityp,t	-32.930	19.910	-1.654	0.100	-72.269	6.408	
SMB Semi-Annual Return	84.286	46.486	1.813	0.072	-7.561	176.134	
residuals_MeanFreqp,t	18.474**	8.179	2.259	0.025	2.314	34.633	
R² global	0.117						
R² ajusté	0.099						
Statistique F	6.654						
Valeur p du test F	0.000						

Le Tableau 4 présente les résultats du test de Durbin-Hausman appliqué au modèle de régression OLS groupée. Ce test évalue l'exogénéité de la variable explicative MeanFreqp, t, en comparant les estimateurs OLS et 2SLS. Une différence statistiquement significative entre les deux séries d'estimateurs indique la présence d'endogénéité, tandis que son absence soutient l'hypothèse d'exogénéité de la variable. La méthode utilisée repose sur une régression auxiliaire, dans laquelle la significativité du coefficient associé aux résidus permet de conclure à la présence ou non d'endogénéité. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Les résultats du test de Durbin-Hausman appliqué à la variable MeanFreq_{p, t} indiquent une valeur p de 0.025 pour le coefficient associé aux résidus issus de la régression auxiliaire. Étant donné que cette valeur est inférieure au seuil de significativité de 5 %, l'hypothèse nulle H₀, selon laquelle MeanFreq_{p, t} est exogène, est rejetée. Cela implique que la composante inexpliquée de cette variable est statistiquement corrélée au terme d'erreur du modèle principal. Par conséquent, l'estimateur des moindres carrés ordinaires (MCO) est biaisé pour cette variable explicative, et les inférences tirées d'un modèle standard peuvent être invalides.

En revanche, les valeurs p obtenues pour les variables Liquidity_{p, t} (0.071) et SMB Semi-Annual Return (0.059) sont toutes deux supérieures au seuil conventionnel retenu de 5 %. Toutefois, ce seuil étant arbitraire, il convient de noter que la valeur de 0.059 reste très proche de celui-ci. Dans le cadre de cette étude, et en raison du seuil choisi, ces résultats ne permettent pas de rejeter l'hypothèse nulle H₀ d'exogénéité pour ces deux variables, qui sont par conséquent considérées comme exogènes dans notre premier modèle (voir **Tableau 22** et **Tableau 23** en annexe).

Bien que MeanFreq_{p,t} ne soit pas significative dans le modèle IV-2SLS final appliqué à la skewness (voir **Tableau 10**) la correction de son endogénéité s'inscrit dans une logique de cohérence méthodologique. Cette approche a été adoptée pour l'ensemble des hypothèses de recherche, et la variable s'est révélée à la fois endogène et significative dans certaines analyses, notamment celles portant sur la kurtosis. Comme précisé plus haut dans cette sous-section, les tests économétriques sont détaillés pour la skewness à titre représentatif. L'objectif est de garantir que l'absence de significativité observée ne découle pas d'un biais d'endogénéité, mais repose sur une estimation robuste et économétriquement valide.

Afin de corriger le problème d'endogénéité identifié pour la variable MeanFreq_{p, t}, nous avons eu recours à la méthode des moindres carrés en deux étapes (2SLS), en instrumentant cette variable à l'aide de sa propre valeur retardée d'une période, MeanFreq_{p, t_lag}. Cette approche repose sur l'hypothèse selon laquelle la variable retardée est corrélée avec la variable endogène actuelle, tout en demeurant exogène vis-à-vis du terme d'erreur du modèle. Une régression de première étape a donc été menée afin d'évaluer la pertinence de l'instrument retenu, en régressant MeanFreq_{p, t} sur MeanFreq_{p, t_lag} ainsi que sur les autres variables explicatives du modèle.

Tableau 5 - Résultats de la régression IV-2SLS corrigéant l'endogénéité de MeanFreqp,t

Variable dépendante:		Semi-Annual Skewness				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	-0.361	0.105	-3.431	0.001	-0.567	-0.155
Liquidityp,t	11.908	2.869	4.150	0.000	6.284	17.531
SMB Semi-Annual Return	107.600	55.596	1.935	0.053	-1.370	216.560
MeanFreqp,t	0.184	0.318	0.578	0.563	-0.440	0.808
R²	0.117					
R² ajusté	0.099					
Statistique F	26.504					
Valeur p du test F	0.000					

Le Tableau 5 présente les résultats de la régression IV-2SLS corrigéant l'endogénéité de la variable explicative MeanFreqp,t. La valeur retardée de MeanFreqp,t (MeanFreqp,t_lag) est utilisée comme instrument. Le tableau affiche les estimations des coefficients, les erreurs standards, les statistiques t, les valeurs p ainsi que les intervalles de confiance obtenus à l'aide de la méthode des moindres carrés en deux étapes (2SLS).

Tableau 6 - Régression de première étape: pertinence instrumentale de MeanFreqp,t_lag pour expliquer MeanFreqp,t

Variable dépendante:		MeanFreqp,t				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.002	0.008	0.210	0.834	-0.014	0.017
MeanFreqp,t_lag	0.965***	0.025	38.120	0.000	0.915	1.015
Liquidityp,t	0.110	0.256	0.432	0.667	-0.395	0.615
SMB Semi-Annual Return	1.095	4.684	0.234	0.815	-8.160	10.351
R²	0.912					
R² ajusté	0.910					
Statistique F	515.399					
Valeur p du test F	0.000					

Le Tableau 6 présente les résultats de la régression de première étape utilisée pour évaluer la pertinence de l'instrument MeanFreqp,t_lag. La variable dépendante est MeanFreqp,t, et l'instrument est inclus aux côtés des autres variables explicatives. Une relation forte et statistiquement significative entre l'instrument et la variable endogène confirme la validité de l'instrument. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Le choix d'utiliser MeanFreq_{p,t_lag} comme instrument repose sur deux critères fondamentaux : sa pertinence empirique et sa validité économique.

D'une part, la pertinence est confirmée par les résultats de la régression de première étape, dans laquelle le coefficient associé à l'instrument est élevé (0.965), avec une valeur p de 0.000 et une statistique F de 515.399, largement supérieure au seuil critique de 10 recommandé par Staiger et Stock (1997) pour écarter le risque d'un instrument faible.

D'autre part, la validité économique de l'instrument découle directement de la construction de la variable instrumentée. MeanFreq_{p,t} est définie comme la moyenne des fréquences de nouvelles observées au semestre précédent ($F_{i,t-1}$), pondérée selon la capitalisation boursière des entreprises constituant le portefeuille p au semestre t ($W_{i,t}$). La variable dépend ainsi uniquement d'une information passée, ce qui la rend pré-déterminée par rapport à la skewness observée au semestre t.

Il est donc raisonnable de supposer que MeanFreq_{p,t_lag}, en tant qu'instrument, est orthogonale aux perturbations non observées susceptibles d'affecter la skewness au semestre actuel, ce qui est une condition indispensable à son exogénéité.

Cette démarche s'inscrit par ailleurs dans les recommandations méthodologiques de Wooldridge (2010), qui préconise l'utilisation de variables retardées comme instruments valides, à condition que l'endogénéité ne soit pas liée à un phénomène de persistance extrême ou de racine unitaire.

En l'occurrence, même si l'instrument MeanFreq_{p,t_lag} explique fortement MeanFreq_{p,t} (coefficient de 0.965), cette relation pourrait s'expliquer par le fait que les portefeuilles sont reconstitués chaque semestre en fonction de la fréquence médiatique observée au semestre précédent. La composition de la variable évoluant d'une période à l'autre, un processus de racine unitaire stricte apparaît donc peu plausible.

4.8.2 Corrélation entre les variables explicatives et multicolinéarité

Pour garantir l'identifiabilité des coefficients dans un modèle de régression, notamment dans le cadre d'une régression OLS groupée, Wooldridge (2010) souligne l'importance de l'absence de colinéarité parfaite entre les variables explicatives. Pour que l'estimateur des moindres carrés ordinaires (MCO) soit bien défini et calculable, la matrice des variables explicatives doit être de rang complet. Autrement dit, les variables explicatives ne doivent pas être parfaitement corrélées, car cela empêcherait l'estimation des coefficients. Lorsqu'une variable explicative est une combinaison linéaire exacte d'une ou plusieurs autres variables, elle devient redondante, compromettant ainsi la spécification du modèle.

Même en l'absence de colinéarité parfaite, une corrélation suffisamment élevée entre certaines variables explicatives peut nuire à la robustesse des estimations en gonflant la variance des estimateurs. Cela réduit non seulement la précision des tests statistiques, mais aussi la puissance des tests d'hypothèses. Par ailleurs, lorsque deux variables sont fortement corrélées, il devient difficile d'isoler leurs effets individuels sur la variable dépendante, ce qui complique l'interprétation des résultats.

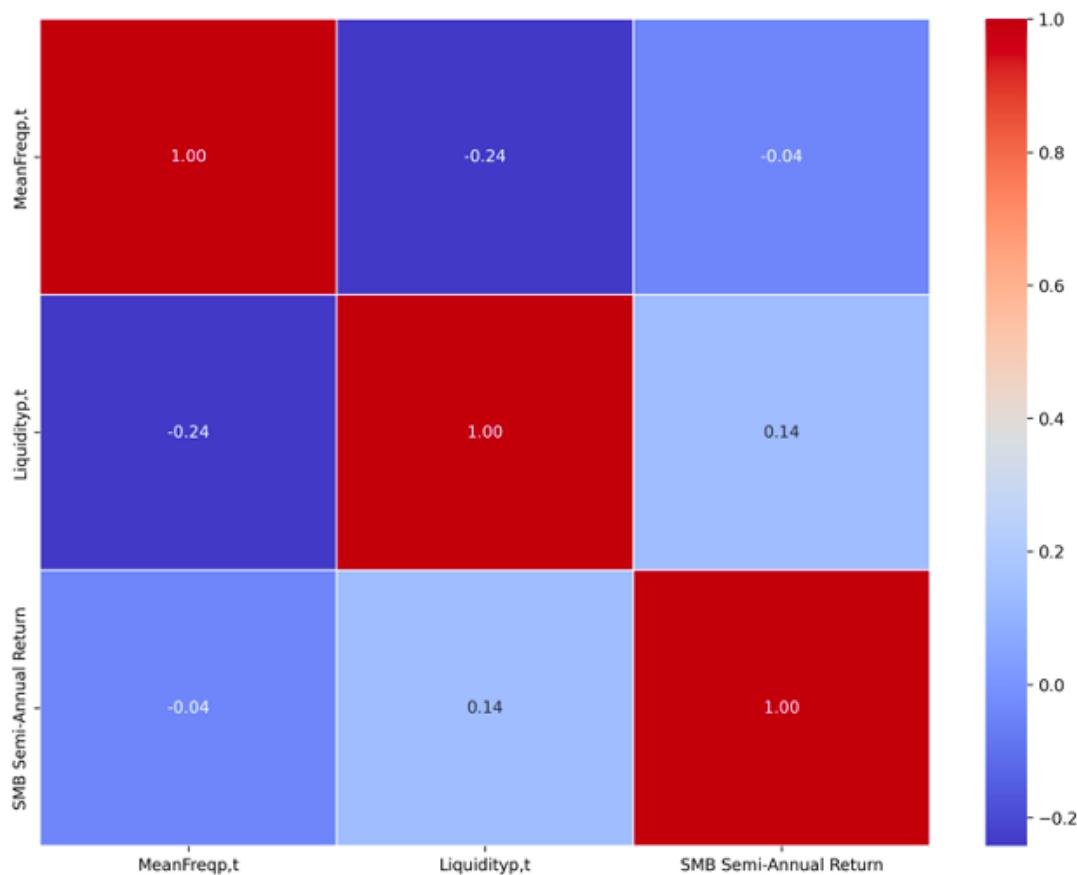
Pour détecter la présence de multicolinéarité, deux outils complémentaires sont couramment utilisés : la matrice de corrélation et le facteur d'inflation de la variance (VIF).

La matrice de corrélation permet d'évaluer la relation linéaire entre les variables explicatives du modèle et d'identifier les paires de variables présentant une forte corrélation. En pratique, un coefficient supérieur à 0,8 est généralement considéré comme un seuil critique, au-delà duquel la multicolinéarité peut apparaître et potentiellement fausser les estimations des coefficients. Toutefois, la matrice de corrélation ne mesure que les corrélations simples (bivariées) et ne prend pas en compte les effets combinés des autres variables.

Le facteur d'inflation de la variance (VIF) pallie cette limitation en quantifiant dans quelle mesure la variance d'un coefficient de régression est gonflée en raison de la colinéarité entre une variable donnée et l'ensemble des autres variables explicatives du modèle. Un VIF supérieur à 10 est généralement interprété comme révélateur d'une multicolinéarité problématique, suggérant que la variable concernée est fortement redondante avec d'autres.

Étant donné que ces outils sont exclusivement conçus pour explorer les interdépendances entre les variables explicatives, indépendamment de leur relation avec le terme d'erreur, l'analyse de la matrice de corrélation et du facteur d'inflation de la variance (VIF) a été réalisée sans corriger l'endogénéité de la variable MeanFreq_{p,t}. En effet, la détection de la multicolinéarité ne fait pas intervenir la variable dépendante. Cette étape de l'analyse est donc indépendante du caractère endogène ou exogène des variables explicatives.

Figure 2 – Matrice de corrélation des variables explicatives pour l'analyse de la skewness semestrielle



La Figure 2 illustre la force et la direction des relations linéaires entre les variables. Un coefficient proche de 1 (ou de -1) indique une corrélation linéaire forte et positive (ou négative), tandis que des valeurs proches de 0 suggèrent une relation linéaire faible ou inexistante. Cette analyse permet de détecter d'éventuels problèmes de multicolinéarité, avec une attention particulière portée aux corrélations dépassant le seuil conventionnel de 0.8. Étant donné que la multicolinéarité ne concerne que les variables explicatives, ce diagnostic est effectué sans correction de l'endogénéité de MeanFreq_{p,t}.

Comme l'illustre la figure ci-dessus, aucune paire de variables explicatives ne présente une corrélation suffisamment élevée pour indiquer un risque de multicolinéarité. En effet, le coefficient de corrélation absolu le plus élevé est observé entre $Liquidity_{p,t}$ et $MeanFreq_{p,t}$, avec une valeur de 0.24, soit bien en dessous du seuil critique de 0.8. Ainsi, les relations linéaires entre les variables explicatives ne semblent poser aucun problème majeur.

Tableau 7 - Vérification de la multicolinéarité basée sur le VIF: spécification finale

Variable	VIF
$Liquidity_{p,t}$	1.507
$MeanFreq_{p,t}$	1.410
SMB Semi-Annual Return	1.088

Le Tableau 7 présente les valeurs du facteur d'inflation de variance (VIF) associées aux variables explicatives incluses dans l'analyse de la skewness semestrielle. Le VIF quantifie dans quelle mesure la variance d'un coefficient estimé est amplifiée en raison de la multicolinéarité. Des valeurs nettement supérieures à 10 sont généralement considérées comme révélatrices d'une multicolinéarité problématique.

L'analyse du facteur d'inflation de la variance (VIF) présentée ci-dessus montre que toutes les variables explicatives retenues dans la spécification finale du modèle affichent des valeurs de VIF largement inférieures au seuil critique de 10. Plus précisément, les valeurs observées sont de 1.507 pour $Liquidity_{p,t}$, 1.410 pour $MeanFreq_{p,t}$, et 1.088 pour SMB Semi-Annual Return. Ces résultats suggèrent que les variables explicatives ne sont pas redondantes et que leurs effets peuvent être identifiés de manière distincte au sein du modèle. Dans l'ensemble, cela confirme l'absence de multicolinéarité problématique.

4.8.3 Homoscédasticité conditionnelle des erreurs

La troisième hypothèse fondamentale à vérifier dans le cadre d'un tel modèle de régression concerne l'homoscédasticité des erreurs. On parle d'erreurs homoscédastiques lorsque leur variance reste constante, quelles que soient les valeurs prises par les variables explicatives. À l'inverse, l'hétéroscléasticité désigne une situation dans laquelle la variance des erreurs varie d'une observation à l'autre. Bien que les estimateurs des moindres carrés ordinaires (MCO) issus d'un modèle OLS groupée restent non biaisés en présence d'hétéroscléasticité, celle-ci compromet néanmoins la validité de l'inférence statistique en faussant l'estimation des erreurs standards.

Parmi les tests couramment utilisés pour évaluer cette hypothèse figurent les tests de Breusch-Pagan et de White.

Le test de Breusch-Pagan examine si la variance conditionnelle des erreurs dépend linéairement des variables explicatives, en s'appuyant sur une régression auxiliaire des résidus au carré.

Le test de White est plus général, car il ne suppose aucune forme fonctionnelle spécifique d'hétéroscléasticité. Il se distingue par sa capacité à détecter des formes plus complexes de dépendance de la variance des erreurs, en intégrant dans la régression auxiliaire des termes au carré et des produits croisés des variables explicatives. Les deux tests reposent sur la même hypothèse nulle (H_0) : la variance des erreurs est constante conditionnellement aux variables explicatives (c'est-à-dire que les erreurs sont homoscédastiques).

Si cette hypothèse est rejetée, il est couramment recommandé d'utiliser des erreurs standards robustes (comme celles de White ou de Huber-White), qui ajustent l'estimation de la variance des coefficients afin de garantir la validité de l'inférence statistique, même en cas de violation de l'hypothèse d'homoscédasticité.

Tableau 8 - Résultats des tests de Breusch-Pagan et de White sur les résidus de la régression IV dans l'analyse de la skewness semestrielle

Test	Statistique	Valeur p
Breusch-Pagan	6.703	0.082
White	13.372	0.147

Le Tableau 8 présente les résultats des tests de Breusch-Pagan et de White visant à évaluer la présence d'hétérosécédasticité dans les résidus du modèle de régression IV-2SLS. Ces deux tests examinent l'hypothèse nulle d'homoscédasticité, selon laquelle la variance des erreurs est constante entre les observations. Une valeur p statistiquement significative indiquerait une violation de cette hypothèse, suggérant le recours à des erreurs standards robustes.

Les tests de Breusch-Pagan et de White ont été conjointement appliqués aux résidus issus de l'estimation du modèle à variables instrumentales (IV), lequel corrige l'endogénéité de la variable MeanFreq_{p, t}. Les valeurs p obtenues sont de 0.082 pour le test de Breusch-Pagan et de 0.147 pour le test de White. Ces résultats ne permettent pas de rejeter l'hypothèse nulle H₀ d'homoscédasticité au seuil de significativité de 5 %. Autrement dit, il n'existe pas de preuve statistique suffisante, au seuil retenu, qui suggère que la variance des erreurs dépend linéairement (test de Breusch-Pagan) ou de manière plus complexe (test de White) des variables explicatives du modèle. Par conséquent, ces deux tests ne remettent pas en cause la validité des inférences statistiques qui seront tirées de l'estimation IV.

4.8.4 Indépendance des erreurs

Au-delà de l'hypothèse d'homoscédasticité conditionnelle, il est également important de vérifier l'hypothèse d'indépendance des erreurs entre les observations, une autre condition fondamentale de l'estimation par les moindres carrés ordinaires (MCO) (Wooldridge, 2010).

L'indépendance des erreurs implique l'absence de corrélation entre les termes d'erreur du modèle. À l'inverse, l'autocorrélation survient lorsque les erreurs d'une observation (dans notre cas, d'un portefeuille) sont corrélées dans le temps, par exemple d'un semestre à l'autre. Tout comme l'hétéroscléasticité, l'autocorrélation compromet la validité de l'inférence statistique en faussant l'estimation des erreurs standards des estimateurs.

Un modèle de régression OLS groupée suppose par défaut que les erreurs sont indépendantes entre les observations. Toutefois, étant donné la structure en panel des données, qui inclut à la fois une dimension temporelle et une dimension en coupe transversale, il est particulièrement important de tester cette hypothèse afin d'assurer la robustesse des résultats et de bien capturer les dynamiques temporelles. Parmi les tests permettant d'évaluer cette hypothèse figurent les tests de Wooldridge et de Breusch-Godfrey.

Le test de Wooldridge est spécifiquement conçu pour détecter l'autocorrélation d'ordre 1 (AR (1)) dans les termes d'erreur lorsqu'on observe plusieurs entités sur plusieurs périodes. Il repose sur une régression des premières différences des résidus sur leurs valeurs retardées et ne nécessite pas l'hypothèse de normalité des erreurs.

Le test de Breusch-Godfrey, quant à lui, est plus général et permet de détecter une autocorrélation d'ordre supérieur (jusqu'à un ordre p spécifié par l'utilisateur), y compris en présence de variables explicatives retardées. Il évalue si les résidus estimés dépendent significativement de leurs propres valeurs passées, ce qui en fait un outil robuste pour identifier des formes plus complexes d'autocorrélation. Les deux tests reposent sur la même hypothèse nulle H_0 , selon laquelle les termes d'erreur sont indépendants entre les observations.

Si cette hypothèse est rejetée, une solution couramment recommandée consiste à utiliser des erreurs standards groupées, qui ajustent les erreurs standards tout en conservant les coefficients des MCO.

En tenant compte simultanément de l'autocorrélation intra-groupe et de l'hétéroscléasticité, cette approche permet de garantir la validité de l'inférence statistique, même lorsque les hypothèses d'indépendance des erreurs ou d'homoscédasticité sont violées.

Tableau 9 - Résultats des tests de Wooldridge et de Breusch-Godfrey sur les résidus de la régression IV dans l'analyse de la skewness semestrielle

Test	Statistique	Valeur p
Wooldridge	-1.143	0.255
Breusch-Godfrey	1.782	0.410

Le Tableau 9 présente les résultats des tests de Wooldridge et de Breusch-Godfrey réalisés afin de détecter la présence d'autocorrélation dans les résidus du modèle de régression IV-2SLS. Ces deux tests évaluent l'hypothèse nulle d'indépendance des erreurs entre les observations, autrement dit, l'absence d'autocorrélation. Une valeur p statistiquement significative indiquerait la présence d'autocorrélation, suggérant le recours à des erreurs standards robustes, regroupées au niveau du portefeuille, afin d'assurer la validité des inférences.

Pour évaluer la présence d'autocorrélation dans les termes d'erreur, le test de Breusch-Godfrey (avec $p = 2$) et le test de Wooldridge ont été appliqués aux résidus issus du modèle à variables instrumentales (IV), lequel corrige l'endogénéité de la variable MeanFreq_{p,t}. Les résultats affichent des valeurs p de 0.255 pour le test de Wooldridge et de 0.410 pour le test de Breusch-Godfrey. Étant donné que ces valeurs sont largement supérieures au seuil de significativité de 5 %, il n'existe aucune preuve statistique permettant de rejeter l'hypothèse nulle H_0 d'indépendance des erreurs. Autrement dit, les tests ne révèlent ni dépendance temporelle d'ordre 1 dans les résidus (test de Wooldridge), ni structure d'autocorrélation plus large jusqu'à l'ordre 2 (test de Breusch-Godfrey). Ainsi, les dynamiques temporelles des termes d'erreur ne semblent pas compromettre la fiabilité des inférences statistiques qui seront tirées de l'estimation IV.

4.8.5 Linéarité de la spécification fonctionnelle du modèle

Bien que l'hypothèse de linéarité ne fasse pas partie des conditions fondamentales à vérifier dans le cadre d'une régression OLS groupée, elle reste pertinente pour évaluer la bonne spécification fonctionnelle du modèle. Wooldridge (2010) souligne que, même dans un cadre MCO standard, une mauvaise spécification de la relation entre la variable dépendante et les variables explicatives peut entraîner des estimations biaisées des coefficients, des erreurs standards inexactes, des tests de significativité invalides et une capacité prédictive affaiblie en raison d'un ajustement global médiocre du modèle.

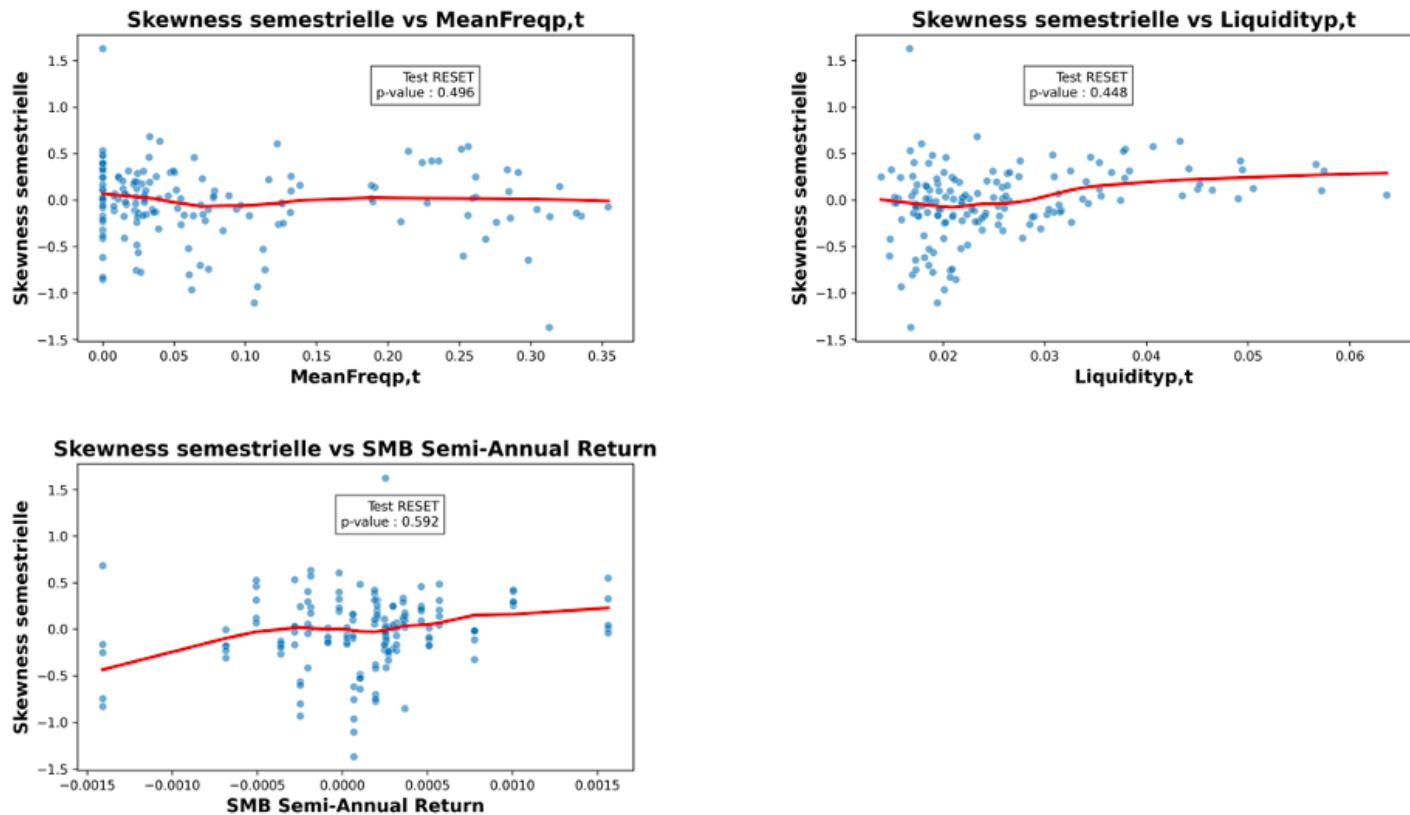
Pour tester cette hypothèse, le test RESET de Ramsey est l'outil recommandé par Wooldridge (2010). Il permet de détecter d'éventuelles non-linéarités omises en évaluant si une transformation non linéaire des valeurs ajustées de la variable dépendante possède un pouvoir explicatif significatif. Si le modèle est correctement spécifié, aucune transformation supplémentaire ne devrait améliorer l'explication de \hat{Y} .

Les hypothèses du test RESET sont les suivantes : l'hypothèse nulle H_0 , selon laquelle le modèle est correctement spécifié, et l'hypothèse alternative H_1 , qui suggère que le modèle est mal spécifié et pourrait être amélioré par l'inclusion de termes non linéaires.

Dans cette analyse, le test RESET de Ramsey a été appliqué au modèle explicatif complet de la skewness semestrielle, en utilisant les variables explicatives telles que spécifiées dans le modèle final sans correction de l'endogénéité de la variable MeanFreq_{p,t}. Cela se justifie par le fait que le test ne repose pas sur la structure des erreurs du modèle, mais évalue plutôt sa spécification fonctionnelle. Il s'agit donc d'un test basé sur la forme du modèle, indépendant du caractère exogène ou endogène des variables explicatives.

Les résultats ont donné une valeur p de 0.346, ce qui ne permet pas de rejeter l'hypothèse nulle au seuil de significativité de 5 % (voir annexe pour le **code Python** utilisé). Il n'existe donc pas de preuve statistique suffisante pour conclure que le modèle bénéficierait d'une transformation non linéaire des variables explicatives. Le modèle est donc considéré comme correctement spécifié dans le cadre de notre analyse.

Figure 3 – Graphiques de dispersion avec courbe LOWESS et résultats du test RESET de Ramsey (individuel) pour l'analyse de la skewness semestrielle



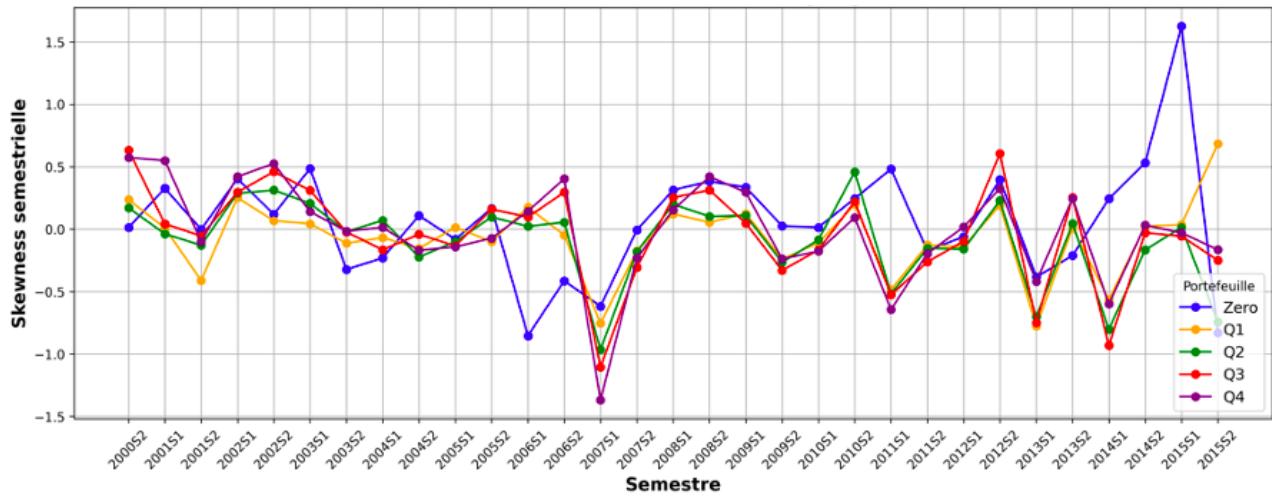
La Figure 3 présente les graphiques de dispersion illustrant les relations individuelles entre la skewness semestrielle et chacune des variables explicatives. La courbe rouge correspond à l'ajustement LOWESS, utilisée pour visualiser d'éventuelles non-linéarités dans les données. Le test RESET a été appliqué séparément pour chaque variable, et les p-values non significatives confirment l'absence de problème majeur de spécification au niveau individuel, renforçant ainsi la validité de la spécification du modèle global.

Bien que le test RESET de Ramsey appliqué au modèle global constitue la méthode la plus appropriée pour évaluer la spécification fonctionnelle d'un modèle de régression (car il examine la relation entre la variable dépendante et l'ensemble des variables explicatives de manière simultanée), nous avons choisi d'aller plus loin en examinant également chaque relation individuelle. À cette fin, les nuages de points présentés ci-dessus ont été générés afin d'illustrer visuellement la relation entre la variable dépendante et chaque variable explicative prise individuellement. Une courbe LOWESS a été ajoutée à chaque graphique afin de détecter d'éventuelles non-linéarités locales. De plus, le test RESET de Ramsey a été appliqué à chaque relation individuelle. Les résultats ont donné des valeurs p de 0.496 pour $\text{MeanFreq}_{p,t}$, 0.448 pour $\text{Liquidity}_{p,t}$, et 0.592 pour SMB Semi-Annual Return, toutes largement supérieures au seuil de

significativité de 5 %. Ainsi, aucune non-linéarité significative n'est observée à l'échelle individuelle, ce qui vient renforcer les résultats du test RESET appliqué au modèle global et confirmer la bonne spécification fonctionnelle du modèle retenu.

5. Présentation et discussion des résultats

Figure 4 - Évolution de la skewness semestrielle selon les portefeuilles construits en fonction de la fréquence des nouvelles



La Figure 4 présente l'évolution de la skewness semestrielle des différents portefeuilles au cours du temps. La ligne bleue représente le portefeuille Zero (entreprises sans couverture médiatique), la ligne orange représente Q1, la ligne verte Q2, la ligne rouge Q3, et la ligne violette Q4 (entreprises avec la couverture médiatique la plus élevée). Chaque ligne retrace les valeurs observées de la skewness semestrielle du second semestre de 2000 (2000S2) jusqu'au second semestre de 2015 (2015S2).

Le portefeuille Zéro, qui regroupe les entreprises n'ayant fait l'objet d'aucune couverture médiatique, présente des fluctuations extrêmes de sa skewness. On observe notamment un pic fortement positif au premier semestre 2015, suivi d'une chute brutale au second semestre de la même année, ce qui illustre clairement son instabilité générale en comparaison avec les autres portefeuilles tout au long de la période.

À l'inverse, les portefeuilles Q3 et Q4, composés des entreprises les plus médiatisées, affichent une skewness plus stable, généralement centrée autour de zéro, avec des variations plus limitées.

Bien que le profil reste quelque peu irrégulier, le portefeuille Q2 tend à enregistrer des valeurs de skewness de plus en plus négatives au fil du temps. Le portefeuille Q1 demeure relativement stable, tandis que le portefeuille Q4 présente une légère tendance baissière de la skewness sur l'ensemble de la période.

Durant la crise financière de 2007-2008, tous les portefeuilles ont connu une baisse de leur skewness, en particulier Q4 et Q3, qui étaient les plus exposés à l'information. Cela suggère que, dans un contexte systémique, une fréquence élevée de nouvelles ne suffit pas à atténuer les mouvements extrêmes des rendements.

Néanmoins, de manière générale, il semble que plus la fréquence des nouvelles publiées est élevée, plus la skewness est contenue et moins sujette à des variations extrêmes.

Tableau 10 - Effet de la fréquence des nouvelles sur la skewness semestrielle: estimation IV-2SLS après correction de l'endogénéité

Variable dépendante:		Semi-Annual Skewness				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	-0.328***	0.096	-3.412	0.001	-0.517	-0.140
Liquidity _t	10.352***	3.275	3.161	0.002	3.934	16.770
SMB Semi-Annual Return	121.567**	57.999	2.096	0.036	7.891	235.244
MeanFreq _t	0.149	0.325	0.458	0.647	-0.488	0.785
R² global	0.102					
R² ajusté	0.084					
Statistique F	17.685					
Valeur p du test F	0.001					

Le Tableau 10 présente les résultats de la régression IV-2SLS, qui estime l'effet de la fréquence des nouvelles sur la skewness semestrielle après correction de l'endogénéité. La variable dépendante est la skewness semestrielle. Les variables explicatives incluent MeanFreq_t, Liquidity_t et SMB Semi-Annual Return. L'estimation utilise MeanFreq_{t_lag} comme instrument de MeanFreq_t. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Le modèle estimé pour expliquer la skewness semestrielle est globalement significatif, avec une statistique F de 17.685 et une valeur p de 0.001.

La variable MeanFreq_t présente un coefficient positif de 0.149, ce qui suggère que toutes choses égales par ailleurs, une augmentation de la fréquence des nouvelles observée au semestre précédent est, en moyenne, associée à une hausse de la skewness au semestre suivant. De plus, la valeur p associée de 0.647, nettement supérieure au seuil conventionnel de 5 %, indique que cette relation

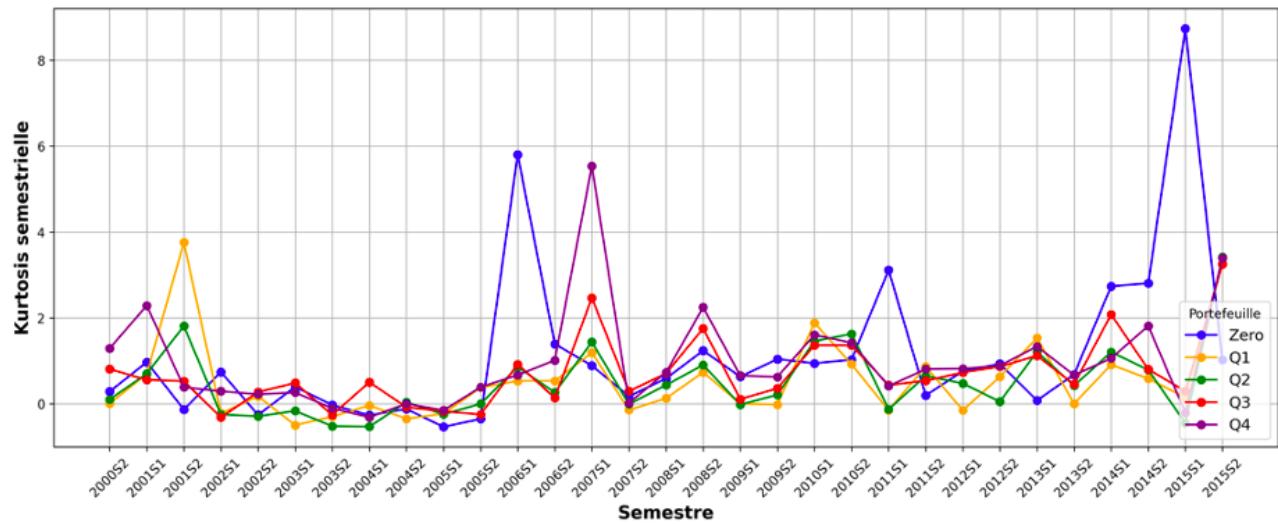
n'est pas statistiquement significative. En considérant à la fois le signe positif du coefficient, contraire à celui attendu, et l'absence de significativité statistique, nous ne pouvons pas rejeter l'hypothèse nulle $H_{0;1}$.

En ce qui concerne les autres variables explicatives du modèle, $\text{Liquidity}_{p,t}$ et SMB Semi-Annual Return présentent toutes deux des effets positifs et significatifs.

Plus précisément, cela indique que, pour un semestre donné, une hausse de $\text{Liquidity}_{p,t}$, qui mesure le bid-ask spread moyen pondéré (et reflète donc une baisse de la liquidité au sein d'un portefeuille), est liée à une augmentation significative de 10.352 points de la skewness semestrielle (valeur $p = 0.002$). Intuitivement, lorsque les portefeuilles deviennent moins liquides, les ajustements de prix tendent à se produire moins fréquemment, mais lorsqu'ils surviennent, ils peuvent être plus brusques et orientés à la hausse. Cela entraîne des rendements ponctuellement très positifs, qui contribuent à une skewness plus élevée.

Par ailleurs, une augmentation d'une unité du SMB Semi-Annual Return (indiquant la surperformance des actions à petite capitalisation par rapport aux grandes capitalisations sur le marché) s'accompagne d'une hausse significative de 121.567 points de la skewness semestrielle (valeur $p = 0.036$). En d'autres termes, pour un semestre donné, la surperformance des petites capitalisations sur l'ensemble du marché se traduit par une distribution des rendements plus asymétrique à droite pour le portefeuille.

Figure 5 - Évolution de la kurtosis semestrielle selon les portefeuilles construits en fonction de la fréquence des nouvelles



La Figure 5 présente l'évolution de la kurtosis semestrielle des différents portefeuilles au cours du temps. La ligne bleue représente le portefeuille Zero (entreprises sans couverture médiatique), la ligne orange représente Q1, la ligne verte Q2, la ligne rouge Q3, et la ligne violette Q4 (entreprises avec la couverture médiatique la plus élevée). Chaque ligne retrace les valeurs observées de la kurtosis semestrielle du second semestre 2000 (2000S2) jusqu'au second semestre 2015 (2015S2).

Le portefeuille Zéro, composé d'entreprises n'ayant fait l'objet d'aucune couverture médiatique, se distingue légèrement des autres portefeuilles par une kurtosis plus élevée et une plus grande instabilité sur l'ensemble de la période d'analyse. Il présente des pics particulièrement marqués au première semestre de 2006 (environ 6) et un sommet extrême au premier semestre de 2015, où sa kurtosis dépasse 8, soit la valeur observée la plus élevée parmi tous les portefeuilles. Ces pics reflètent une probabilité accrue de rendements extrêmes, ce qui suggère que ces entreprises sont particulièrement vulnérables aux chocs de marché.

À l'inverse, les portefeuilles Q1 à Q4, qui regroupent des entreprises bénéficiant d'un niveau croissant de couverture médiatique, affichent des niveaux de kurtosis beaucoup plus modérés et stables, généralement compris entre 0 et 3.

La réaction à la crise financière de 2007–2008 est particulièrement marquée pour le portefeuille Q4, probablement en raison de la forte exposition médiatique des entreprises qui le composent, ce qui les rend plus sensibles aux chocs macroéconomiques. À l'inverse, les autres portefeuilles semblent avoir été relativement moins affectés.

Dans l'ensemble, les données montrent que plus la fréquence des nouvelles augmente, plus la distribution des rendements tend à se normaliser, avec moins de valeurs extrêmes et une plus grande stabilité au fil du temps. Cette tendance devient particulièrement visible après l'année 2010, à partir de laquelle les portefeuilles Q1 à Q4 présentent une kurtosis plus stable, contrairement au portefeuille Zéro, qui continue d'afficher un comportement irrégulier.

Tableau 11 - Effet de la fréquence des nouvelles sur la kurtosis semestrielle: estimation IV-2SLS après correction de l'endogénéité

Variable dépendante:		Semi-Annual Kurtosis				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.355*	0.192	1.854	0.064	-0.020	0.731
MKT-RF Semi-Annual Return	-346.961***	122.306	-2.837	0.005	-586.676	-107.246
HML Semi-Annual Return	-387.088*	205.468	-1.884	0.060	-789.798	15.622
MeanFreqp,t	6.586***	2.049	3.214	0.001	2.570	10.603
Betap,t_standardized	0.985***	0.273	3.603	0.000	0.449	1.521
R² global	-0.082					
R² ajusté	-0.111					
Statistique F	19.518					
Valeur p du test F	0.001					

Le Tableau 11 présente les résultats de la régression IV-2SLS, qui estime l'effet de la fréquence des nouvelles sur la kurtosis semestrielle après correction de l'endogénéité. La variable dépendante est la kurtosis semestrielle. Les variables explicatives incluent MeanFreqp,t, Betap,t_standardized, MKT-RF Semi-Annual Return et HML Semi-Annual Return. L'estimation utilise respectivement MeanFreqp,t_lag et Betap,t_standardized_lag comme instruments pour MeanFreqp,t et Betap,t_standardized. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 19.518 et une valeur p de 0.001, le modèle estimé pour expliquer la kurtosis semestrielle est globalement significatif.

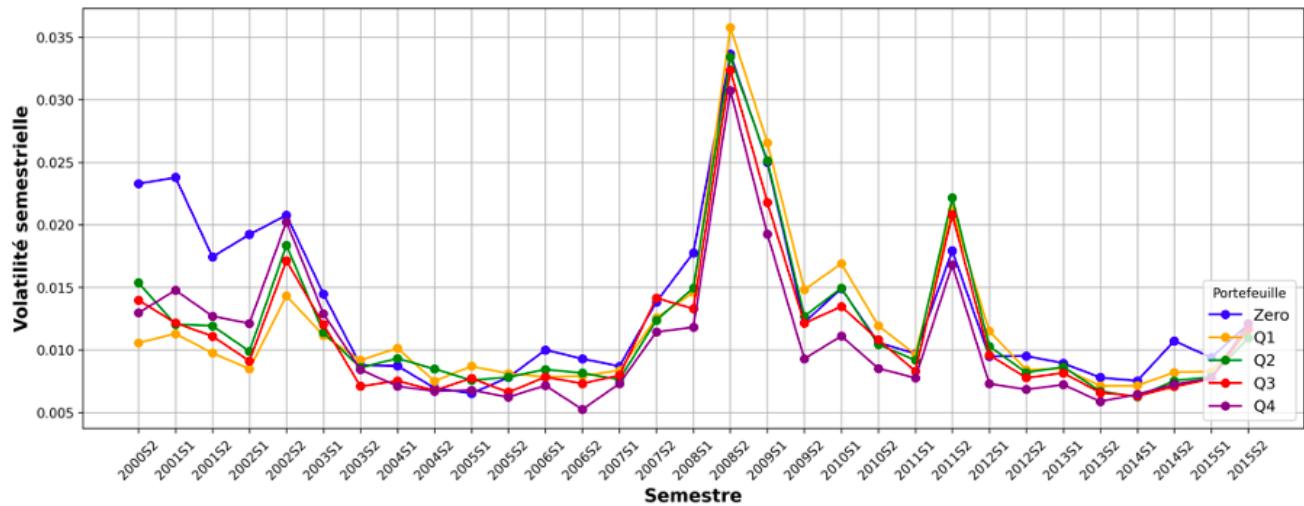
La variable MeanFreq_{p,t}, qui mesure la fréquence des nouvelles, présente un coefficient positif de 6.586, statistiquement significatif au seuil de 5 % (valeur p = 0.001). Cela signifie que toutes choses égales par ailleurs, une augmentation de la fréquence des nouvelles observée au semestre précédent est, en moyenne, associée à une hausse significative de 6.586 points de la kurtosis au semestre suivant. Autrement dit, une fréquence plus élevée de couverture médiatique tend à accroître la probabilité d'observer des rendements extrêmes dans la distribution des rendements au cours de la période suivante. Dans le cadre du test unilatéral mené, ce résultat est conforme à la relation attendue et statistiquement significatif. On peut donc rejeter l'hypothèse nulle H_{0;2} au seuil de 5 %, et valider l'hypothèse alternative H_{1;2}, qui postule un effet positif.

Concernant la variable MKT-RF Semi-Annual Return, celle-ci présente un coefficient négatif et significatif de -346.961 (valeur p = 0.005). Cela suggère que, pour un semestre donné, une augmentation du rendement excédentaire du marché (par rapport au taux sans risque) est liée à une diminution significative de 346.961 points de la kurtosis semestrielle des rendements de portefeuille. Ce résultat peut s'expliquer par le fait que lorsque le marché affiche de bonnes performances durant un semestre, les rendements des portefeuilles ont tendance à se concentrer davantage autour de leur moyenne, ce qui réduit la probabilité d'observer des rendements extrêmes.

Par ailleurs, la variable Beta_{p,t}_standardized présente un coefficient positif et significatif de 0.985, avec une valeur p de 0.000. Cela indique que, pour un semestre donné, une exposition plus élevée des portefeuilles au risque systématique s'accompagne d'une hausse significative de 0.985 points de la kurtosis semestrielle. Ce résultat est cohérent avec l'idée selon laquelle les portefeuilles à bêta élevé sont plus sensibles aux chocs de marché et donc plus susceptibles de générer des rendements extrêmes.

Enfin, bien que l'effet estimé de -387.088 (pour la variable HML Semi-Annual Return) sur la kurtosis semestrielle soit proche du seuil de significativité, la valeur p de 0.060 reste légèrement supérieure au seuil conventionnel choisi de 5 %. Cet effet n'est donc pas considéré comme statistiquement significatif dans le cadre de cette analyse.

Figure 6 - Évolution de la volatilité semestrielle selon les portefeuilles construits en fonction de la fréquence des nouvelles



La Figure 6 présente l'évolution de la volatilité semestrielle des différents portefeuilles au cours du temps. La ligne bleue représente le portefeuille Zero (entreprises sans couverture médiatique), la ligne orange représente Q1, la ligne verte Q2, la ligne rouge Q3, et la ligne violette Q4 (entreprises avec la couverture médiatique la plus élevée). Chaque ligne retrace les valeurs observées de la volatilité semestrielle du second semestre 2000 (2000S2) jusqu'au second semestre 2015 (2015S2).

Une première observation frappante concerne la flambée de la volatilité durant la crise financière mondiale de 2007-2008: tous les portefeuilles, quel que soit leur niveau de couverture médiatique, ont connu une hausse importante. À ce moment-là, chacun a atteint son niveau de volatilité maximal. Le portefeuille Q1 enregistre la valeur la plus élevée de toute la période, suivi respectivement des portefeuilles Zéro, Q2, Q3 et Q4. Ainsi, les portefeuilles les moins médiatisés occupent le haut du classement, tandis que ceux bénéficiant d'une plus grande couverture apparaissent en bas.

Sur l'ensemble de la période, le portefeuille Zéro affiche également une volatilité systématiquement plus élevée que les autres portefeuilles, en particulier au début et à la fin de la période d'analyse. Le manque de couverture médiatique des entreprises qui le composent pourrait donc être associé à un risque accru pour les investisseurs.

À la suite du pic observé au second semestre 2011, qui correspond très certainement à la crise de la dette souveraine en Europe, la volatilité tend à se stabiliser. Du premier semestre 2012 jusqu'à

la fin de la période étudiée, les écarts de volatilité entre les portefeuilles se réduisent, ce qui traduit un environnement de marché plus calme.

Enfin, sur l'ensemble de la période, les portefeuilles Q3 et Q4, composés des entreprises bénéficiant des plus hauts niveaux de couverture médiatique, se distinguent par une volatilité plus faible et plus stable que les autres. La plus grande disponibilité d'information pour ces entreprises semble avoir réduit l'incertitude perçue par les investisseurs, ce qui se reflète dans une volatilité moindre de leurs rendements.

Tableau 12 - Effet de la fréquence des nouvelles sur la volatilité semestrielle: estimation IV-2SLS après correction de l'endogénéité

Variable dépendante:		Semi-Annual Volatility				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.013***	0.000	61.470	0.000	0.012	0.013
Log(Mp,t)_standardized	0.001***	0.000	5.097	0.000	0.001	0.001
SMB Semi-Annual Return	-1.284***	0.318	-4.036	0.000	-1.908	-0.661
HML Semi-Annual Return	-0.593*	0.304	-1.950	0.051	-1.190	0.003
RMW Semi-Annual Return	-0.153	0.329	-0.465	0.642	-0.797	0.491
<i>MeanFreqp,t</i>	-0.004**	0.002	-2.187	0.029	-0.008	-0.000
Liquidityp,t_standardized	0.006***	0.000	28.011	0.000	0.005	0.006
R² global	0.908					
R² ajusté	0.905					
Statistique F	1171.230					
Valeur p du test F	0.000					

Le Tableau 12 présente les résultats de la régression IV-2SLS, qui estime l'effet de la fréquence des nouvelles sur la volatilité semestrielle après correction de l'endogénéité. La variable dépendante est la volatilité semestrielle. Les variables explicatives incluent MeanFreqp,t, Liquidityp,t_standardized, Log(Mp,t)_standardized, SMB Semi-Annual Return, HML Semi-Annual Return et RMW Semi-Annual Return. L'estimation utilise respectivement MeanFreqp,t_lag et Liquidityp,t_standardized_lag comme instruments de MeanFreqp,t et Liquidityp,t_standardized. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Le modèle estimé pour expliquer la volatilité semestrielle est globalement significatif, avec une statistique F de 1171.230 et une valeur p de 0.000.

La variable MeanFreq_{p, t}, présente un coefficient négatif de -0.004, statistiquement significatif au seuil de 5 % (valeur p = 0.029). Cela signifie que toutes choses égales par ailleurs, une augmentation de la fréquence des nouvelles observée au semestre précédent est, en moyenne, associée à une diminution significative de 0.004 points de la volatilité semestrielle au semestre suivant. Bien que statistiquement significatif, le signe négatif du coefficient va à rebours de l'hypothèse de recherche H_{1;3}, qui postule un effet positif de la fréquence des nouvelles sur la volatilité. Ainsi, dans le cadre du test unilatéral mené, ce résultat ne permet pas de rejeter l'hypothèse nulle H_{0;3}.

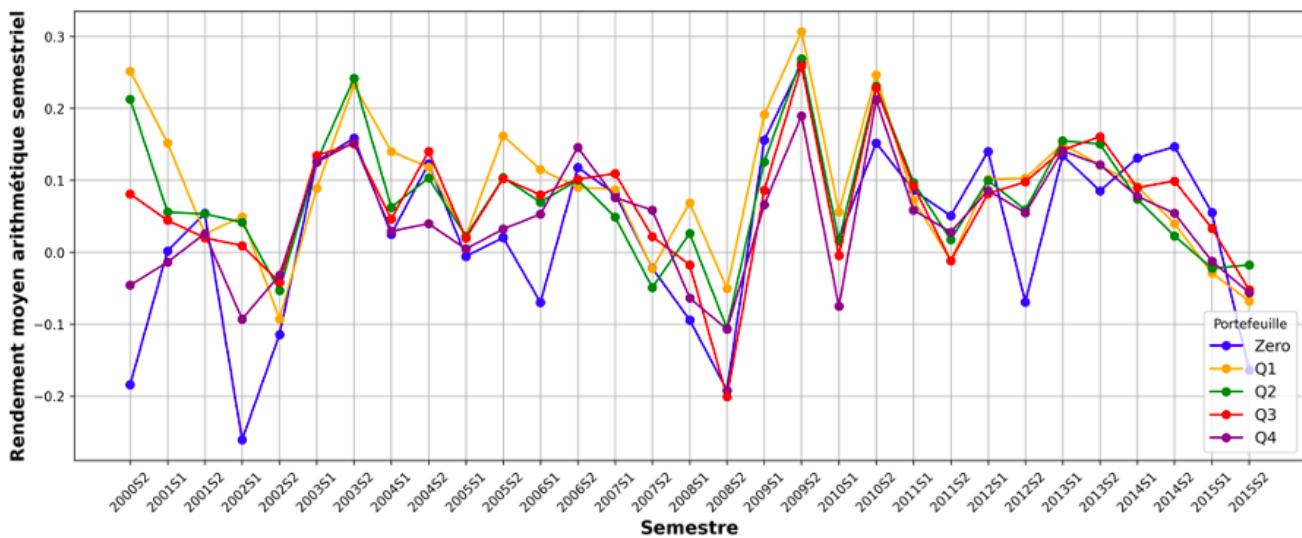
En ce qui concerne les facteurs de risque de Fama-French, seule la variable SMB Semi-Annual Return présente un effet négatif et significatif (coefficient de -1.284; valeur p = 0.000), ce qui indique que, pour un semestre donné, la surperformance des actions à petite capitalisation par rapport aux grandes capitalisations sur le marché est liée à une baisse significative de 1.284 points de la volatilité semestrielle. Ce type de surperformance survient généralement dans des conditions de marché plus favorables, au cours desquelles les investisseurs ont tendance à réallouer leur capital vers des actifs plus risqués. Ainsi, ce résultat peut s'expliquer par le fait qu'une redistribution des flux vers un plus large éventail de titres contribue à réduire la pression exercée sur certains segments spécifiques du marché, ce qui atténue les fluctuations globales et permet ainsi de faire baisser la volatilité agrégée.

Par ailleurs, les variables Liquidity_{p, t}_standardized et Log(M_{p, t})_standardized présentent toutes deux des effets positifs et significatifs sur la volatilité semestrielle, avec des coefficients respectifs de 0.006 et 0.001, et des valeurs p égales à 0.000. D'une part, l'interprétation de ces résultats suggère que, pour un semestre donné, les portefeuilles moins liquides se caractérisent par une plus grande instabilité de leurs rendements boursiers. Cela peut s'expliquer par le fait que, dans un environnement moins liquide, les coûts de transaction sont généralement plus élevés et les ajustements de prix moins fluides. En conséquence, les investisseurs sont plus susceptibles de réagir de manière abrupte à l'arrivée de nouvelles informations, ce qui se traduit par des mouvements de prix plus volatils. En d'autres termes, une faible liquidité amplifie la transmission des chocs informationnels, ce qui accentue la volatilité. D'autre part, le résultat selon lequel, au

cours d'un semestre donné, une augmentation de la capitalisation boursière moyenne des portefeuilles s'accompagne d'une hausse significative de leur volatilité peut sembler contre-intuitif à première vue. Toutefois, cela peut s'expliquer par le fait que les valeurs à grande capitalisation incluses dans ces portefeuilles sont généralement plus suivies par les analystes financiers et bénéficient d'une exposition médiatique accrue. Cette forte visibilité peut accroître la sensibilité de leurs prix aux flux d'information, ce qui rend les réactions du marché plus rapides et parfois plus marquées. Par conséquent, bien que ces entreprises soient structurellement perçues comme plus stables, leur forte réactivité à l'information publique peut, paradoxalement, se traduire par une volatilité importante.

Enfin, avec des valeurs p respectives de 0.051 (légèrement supérieure au seuil de significativité retenu de 5%) et de 0.642 (nettement supérieure à ce seuil), les effets négatifs des variables HML Semi-Annual Return et RMW Semi-Annual Return sur la volatilité semestrielle ne sont pas considérées comme statistiquement significatifs dans le cadre de cette analyse.

Figure 7 - Évolution du rendement moyen arithmétique semestriel selon les portefeuilles construits en fonction de la fréquence des nouvelles



La Figure 7 présente l'évolution du rendement moyen arithmétique semestriel des différents portefeuilles au cours du temps. La ligne bleue représente le portefeuille Zero (entreprises sans couverture médiatique), la ligne orange représente Q1, la ligne verte Q2, la ligne rouge Q3, et la ligne violette Q4 (entreprises avec la couverture médiatique la plus élevée). Chaque ligne retrace les valeurs observées du rendement moyen arithmétique semestriel du second semestre 2000 (2000S2) jusqu'au second semestre 2015 (2015S2).

La crise financière a eu un impact majeur sur l'ensemble des portefeuilles en entraînant une forte baisse des rendements moyens arithmétiques. Certains portefeuilles ont même atteint leurs plus faibles niveaux de rendement durant cette période, en particulier Q2, Q3 et Q4.

Bien que la différence reste modérée, le portefeuille Zéro affiche des rendements moyens arithmétiques légèrement inférieurs par rapport aux autres portefeuilles qui bénéficient d'un certain niveau de couverture médiatique, quelle qu'en soit l'intensité. Ce portefeuille a notamment enregistré ses rendements les plus faibles aux semestres 2002S1, 2006S1 et 2012S2.

Cependant, à partir du premier semestre de 2013, on observe une tendance générale à la baisse des rendements dans l'ensemble des portefeuilles, accompagnée de fluctuations légèrement atténées, bien qu'encore présentes.

Dans l'ensemble, bien que le portefeuille Zéro présente des rendements légèrement plus faibles à certaines périodes, cette tendance n'est ni constante ni monotone sur l'ensemble de la période

d'analyse. Il n'est donc pas possible d'établir un lien clair entre l'évolution des rendements moyens arithmétiques et le niveau de couverture médiatique à ce stade.

Tableau 13 - Effet de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel: estimation IV-2SLS après correction de l'endogénéité

Variable dépendante:		Semi-Annual Arithmetic Average Return				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.715***	0.211	3.394	0.001	0.302	1.129
MKT-RF Semi-Annual Return	86.201***	4.829	17.849	0.000	76.736	95.666
CMA Semi-Annual Return	15.136	13.667	1.107	0.268	-11.652	41.923
<i>MeanFreqp,t</i>	0.297*	0.162	1.837	0.066	-0.020	0.613
log(Mp,t)	-0.029***	0.009	-3.128	0.002	-0.047	-0.011
Betap,t_standardized	0.020	0.014	1.401	0.161	-0.008	0.048
R² global	0.738					
R² ajusté	0.729					
Statistique F	393.782					
Valeur p du test F	0.000					

Le Tableau 13 présente les résultats de la régression IV-2SLS, qui estime l'effet de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel après correction de l'endogénéité. La variable dépendante est le rendement moyen arithmétique semestriel. Les variables explicatives incluent *MeanFreqp,t*, log(Mp,t), Betap,t_standardized, MKT-RF Semi-Annual Return et CMA Semi-Annual Return. L'estimation utilise respectivement *MeanFreqp,t_lag*, log(Mp,t)_lag et Betap,t_standardized_lag comme instruments de *MeanFreqp,t*, log(Mp,t) et Betap,t_standardized. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 393.782 et une valeur p de 0.000, le modèle estimé pour expliquer le rendement moyen arithmétique semestriel est globalement significatif.

La variable *MeanFreqp,t*, qui mesure la fréquence des nouvelles, présente un coefficient positif de 0.297. Cela signifie que toutes choses égales par ailleurs, une augmentation de la fréquence des

nouvelles observée au semestre précédent est, en moyenne, associée à une hausse de 0.297 points du rendement moyen arithmétique au semestre suivant. Par ailleurs, avec une valeur p de 0.066, légèrement supérieure au seuil conventionnel choisi de 5 %, cette relation n'est pas considérée comme statistiquement significative dans le cadre de cette analyse. En tenant compte à la fois du signe positif du coefficient, contraire à celui attendu, et de l'absence de significativité statistique, l'hypothèse nulle $H_{0;4}$, ne peut pas être rejetée.

Concernant le facteur de risque MKT-RF Semi-Annual Return, celui-ci exerce un effet fortement positif et statistiquement significatif (coefficient de 86.201, valeur p = 0.000). Ce résultat indique que, pour un semestre donné, une augmentation du rendement excédentaire du marché (par rapport au taux sans risque) est liée à une hausse significative de 86.201 points du rendement moyen arithmétique semestriel des portefeuilles. Cette relation est cohérente avec les prédictions du CAPM, selon lesquelles le rendement d'un actif est positivement lié à la prime de risque de marché.

À l'inverse, la variable Log($M_{p,t}$) présente un effet négatif et significatif de -0.029, avec une valeur p de 0.002. Cette relation suggère que, pour un semestre donné, une augmentation de la capitalisation boursière moyenne au sein d'un portefeuille se traduit par une baisse du rendement moyen arithmétique. Ce résultat peut s'expliquer par le fait que les grandes entreprises, souvent plus matures et moins dynamiques, offrent un potentiel de croissance plus limité, ce qui se traduit généralement par des rendements plus faibles.

Pour conclure, avec des valeurs p respectives de 0.268 et 0.161, toutes deux nettement supérieures au seuil de significativité de 5 %, les effets positifs des variables CMA Semi-Annual Return et Beta_{p,t}_standardized sur le rendement moyen arithmétique semestriel ne sont pas statistiquement significatifs.

6. Tests de robustesse

6.1 Hétérogénéité de l'effet de la fréquence des nouvelles selon les portefeuilles

Bien que la construction des portefeuilles en fonction de la fréquence des nouvelles constitue un élément central de la méthodologie, l'inclusion des termes d'interaction entre la variable explicative principale, $\text{MeanFreq}_{p,t}$, et les variables muettes de portefeuille, a volontairement été réservée aux tests de robustesse. Cette décision a été prise afin d'isoler clairement, dans le modèle principal, l'effet moyen global de la fréquence des nouvelles sur la distribution des rendements, sans introduire dès le départ une hétérogénéité structurelle entre les portefeuilles. L'introduction de ces termes d'interaction dans les tests de robustesse vise ainsi à vérifier la stabilité et la cohérence des résultats, tout en testant l'hypothèse selon laquelle l'impact de la fréquence des nouvelles pourrait varier selon le niveau d'exposition médiatique des entreprises (allant du portefeuille Zéro, sans couverture, au portefeuille Q4, le plus médiatisé). Plus précisément, ce premier test de robustesse analyse l'effet différencié de la fréquence des nouvelles sur chacune des variables dépendantes. Pour ce faire, des variables muettes ont été créées pour représenter chacun des quatre portefeuilles exposés médiatiquement (Q1 à Q4), le portefeuille Zéro servant de catégorie de référence. Des termes d'interaction entre $\text{MeanFreq}_{p,t}$ et chacune de ces variables muettes ont ensuite été construits afin d'évaluer si l'effet de la fréquence des nouvelles sur les variables dépendantes varie significativement d'un portefeuille à l'autre, et donc selon l'intensité de la couverture médiatique reçue par les entreprises au cours du semestre précédent.

Tableau 14 - Effet hétérogène de la fréquence des nouvelles sur la skewness semestrielle: estimation IV-2SLS avec interactions de portefeuille

Variable dépendante:		Semi-Annual Skewness					
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur	
const	-0.230**	0.114	-2.012	0.044	-0.455	-0.006	
Liquidityp,t	9.509***	3.288	2.892	0.004	3.065	15.953	
SMB Semi-Annual Return	106.723*	57.906	1.843	0.065	-6.771	220.218	
MeanFreqp,t_Portfolio_Q1	-4.452	3.903	-1.141	0.254	-12.102	3.197	
MeanFreqp,t_Portfolio_Q2	-2.422	1.760	-1.376	0.169	-5.870	1.027	
MeanFreqp,t_Portfolio_Q3	-0.849	1.013	-0.838	0.402	-2.834	1.136	
MeanFreqp,t_Portfolio_Q4	-0.060	0.354	-0.168	0.867	-0.753	0.634	
R² global	0.130						
R² ajusté	0.094						
Statistique F	20.943						
Valeur p du test F	0.002						

Le Tableau 14 présente les résultats de la régression IV-2SLS qui estime l'effet hétérogène de la fréquence des nouvelles sur la skewness semestrielle selon les portefeuilles. La variable dépendante est la skewness semestrielle. Les variables explicatives incluent Liquidityp,t, SMB Semi-Annual Return, ainsi que les interactions entre MeanFreqp,t et les dummys de portefeuille (Q1, Q2, Q3 et Q4). Le portefeuille Zero sert de catégorie de référence. L'estimation utilise les valeurs retardées de MeanFreqp,t, en interaction avec les dummys de portefeuille, comme instruments. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 20.943 et une valeur p de 0.002, le modèle est globalement significatif.

La constante représente l'estimation de la skewness semestrielle pour le portefeuille Zéro, composé d'entreprises sans couverture médiatique. Dans un contexte où les autres variables explicatives sont à leur niveau de référence, cette skewness est estimée à -0.230. Ce résultat, statistiquement

significatif (avec une valeur p de 0.044), indique une asymétrie négative des rendements de ces entreprises et donc une exposition accrue aux chocs négatifs.

Les termes d'interaction entre la variable $\text{MeanFreq}_{p,t}$ et les variables indicatrices de portefeuille présentent tous des coefficients négatifs, conformément à l'hypothèse alternative $H_{1;1}$, bien qu'aucun ne soit statistiquement significatif (avec des valeurs p comprises entre 0.169 pour Q2 et 0.867 pour Q4). L'effet négatif estimé est plus marqué pour les portefeuilles faiblement exposés médiatiquement (coefficient de -4.452 pour Q1) et s'atténue à mesure que la couverture médiatique augmente, jusqu'à devenir presque nul pour le portefeuille le plus couvert (coefficient de -0.060 pour Q4). Bien que cette relation ne soit pas confirmée sur le plan statistique, la dynamique observée suggère qu'une plus grande visibilité médiatique est associée à une moindre exposition aux chocs négatifs, en cohérence avec l'analyse graphique de la **Figure 4**.

Comparativement au modèle principal, dans lequel la fréquence des nouvelles exerçait un effet globalement positif mais non significatif sur la skewness, l'introduction des termes d'interactions révèle des dynamiques opposées selon le portefeuille. Tandis que les entreprises bénéficiant d'une couverture médiatique, quelle que soit son intensité, ne présentent pas de relation significative avec la fréquence des nouvelles, celles sans couverture médiatique affichent une skewness structurellement plus négative.

Tableau 15 - Effet hétérogène de la fréquence des nouvelles sur la kurtosis semestrielle: estimation IV-2SLS avec interactions de portefeuille

Variable dépendante:		Semi-Annual Kurtosis				
Variables explicatives	Coefficient	Erreur standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.083	0.382	0.218	0.828	-0.666	0.832
MKT-RF Semi-Annual Return	-385.800***	133.751	-2.885	0.004	-647.946	-123.654
HML Semi-Annual Return	-380.290*	210.645	-1.805	0.071	-793.146	32.566
MeanFreqp,t_Portfolio_Q1	16.903	17.709	0.955	0.340	-17.806	51.611
MeanFreqp,t_Portfolio_Q2	12.126	8.987	1.349	0.177	-5.489	29.740
MeanFreqp,t_Portfolio_Q3	13.110**	6.197	2.116	0.034	0.964	25.256
MeanFreqp,t_Portfolio_Q4	7.919***	2.861	2.768	0.006	2.312	13.526
Betap,t_standardized	1.156***	0.359	3.218	0.001	0.452	1.859
R² global	-0.136					
R² ajusté	-0.192					
Statistique F	19.238					
Valeur p du test F	0.008					

Le Tableau 15 présente les résultats de la régression IV-2SLS qui estime l'effet hétérogène de la fréquence des nouvelles sur la kurtosis semestrielle selon les portefeuilles. La variable dépendante est la kurtosis semestrielle. Les variables explicatives incluent MKT-RF Semi-Annual Return, HML Semi-Annual Return, Betap,t_standardized, ainsi que les interactions entre MeanFreqp,t et les dummys de portefeuille (Q1, Q2, Q3 et Q4). Le portefeuille Zero sert de catégorie de référence. L'estimation utilise les valeurs retardées de MeanFreqp,t, en interaction avec les dummys de portefeuille, et Betap,t_standardized_lag, comme instruments des variables endogènes. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 19.238 et une valeur p de 0.008, le modèle est globalement significatif.

La constante représente l'estimation de la kurtosis semestrielle pour le portefeuille Zéro, composé d'entreprises sans couverture médiatique. Lorsque les autres variables explicatives sont à leur niveau de référence, la kurtosis semestrielle est estimée à 0.083. Bien que cette valeur suggère une légère propension aux rendements extrêmes pour ces entreprises, elle n'est pas statistiquement significative (valeur p = 0.828).

Les termes d'interaction entre MeanFreq_{p,t} et les variables indicatrices de portefeuille présentent tous des coefficients positifs, en accord avec l'hypothèse alternative H_{1;2}. Toutefois, seuls les effets associés aux portefeuilles Q3 et Q4, qui correspondent aux entreprises les plus médiatisées, sont statistiquement significatifs (avec des valeurs p respectives de 0.034 et 0.006). Ces résultats suggèrent qu'une augmentation de la fréquence des nouvelles au semestre précédent accroît significativement la probabilité d'observer des rendements extrêmes, mais uniquement pour ces entreprises fortement exposées médiatiquement.

L'analyse des coefficients montre que cet effet tend à diminuer à mesure que la couverture médiatique augmente, passant d'un coefficient de 16.903 pour Q1 à 7.919 pour Q4 malgré une légère irrégularité observée entre le portefeuille Q3 dont le coefficient (13.110) dépasse celui de Q2 (12.126).

Ainsi, bien que l'effet estimé soit plus fort pour les entreprises peu couvertes médiatiquement, il reste non significatif (avec des valeurs p de 0.340 pour Q1 et 0.177 pour Q2) tandis qu'il devient modéré mais significatif pour les entreprises les plus médiatisées. Ce schéma, partiellement confirmé, reste cohérent avec l'analyse graphique de la **Figure 5**.

Comparativement au modèle sans interactions, dans lequel la fréquence des nouvelles exerçait un effet globalement positif et significatif sur la kurtosis, l'inclusion des termes d'interaction met en évidence une hétérogénéité selon le niveau d'exposition médiatique. L'effet moyen global masque en réalité des dynamiques contrastées : plus marquées mais non significatives pour les entreprises peu visibles, et plus modérées mais statistiquement significatives pour celles plus suivies. Cette différence pourrait s'expliquer par une intégration de l'information plus fluide, régulière et davantage anticipée par les marchés pour les entreprises médiatisées.

Tableau 16 - Effet hétérogène de la fréquence des nouvelles sur la volatilité semestrielle: estimation IV-2SLS avec interactions de portefeuille

Variable dépendante:		Semi-Annual Volatility				
Variables explicatives	Coefficient	Erreur standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	0.013***	0.000	42.333	0.000	0.012	0.013
SMB Semi-Annual Return	-1.252***	0.330	-3.790	0.000	-1.899	-0.604
HML Semi-Annual Return	-0.603**	0.305	-1.974	0.048	-1.201	-0.004
RMW Semi-Annual Return	-0.153	0.324	-0.471	0.638	-0.789	0.483
Log(Mp,t)_standardized	0.001***	0.000	3.834	0.000	0.001	0.002
MeanFreqp,t_Portfolio_Q1	-0.002	0.023	-0.078	0.938	-0.046	0.043
MeanFreqp,t_Portfolio_Q2	0.002	0.010	0.181	0.856	-0.017	0.021
MeanFreqp,t_Portfolio_Q3	-0.003	0.005	-0.515	0.607	-0.012	0.007
MeanFreqp,t_Portfolio_Q4	-0.004**	0.002	-2.210	0.027	-0.008	-0.001
Liquidityp,t_standardized	0.006***	0.000	27.298	0.000	0.005	0.006
R² global	0.910					
R² ajusté	0.904					
Statistique F	1217.635					
Valeur p du test F	0.000					

Le Tableau 16 présente les résultats de la régression IV-2SLS qui estime l'effet hétérogène de la fréquence des nouvelles sur la volatilité semestrielle selon les portefeuilles. La variable dépendante est la volatilité semestrielle. Les variables explicatives incluent SMB Semi-Annual Return, HML Semi-Annual Return, RMW Semi-Annual Return, Log(Mp,t)_standardized, Liquidityp,t_standardized, ainsi que les interactions entre MeanFreqp,t et les dummys de portefeuille (Q1, Q2, Q3 et Q4). Le portefeuille Zero sert de catégorie de référence. L'estimation utilise les valeurs retardées de MeanFreqp,t, en interaction avec les dummys de portefeuille, et Liquidityp,t_standardized_lag comme instruments des variables endogènes. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 1217.635 et une valeur p de 0.000, le modèle est globalement significatif.

La constante du modèle représente l'estimation de la volatilité semestrielle pour le portefeuille Zéro, composé d'entreprises sans aucune couverture médiatique. Lorsque toutes les autres variables explicatives sont à leur niveau de référence, cette volatilité est estimée à 0.013.

Statistiquement significative (avec une valeur p de 0.000), cette estimation reflète un niveau de volatilité structurelle qui pourrait être lié à un manque d'information publique ou à une attention réduite des investisseurs, deux facteurs susceptibles d'accroître l'incertitude et donc, les fluctuations des rendements.

Les termes d'interaction entre la fréquence des nouvelles et les variables indicatrices de portefeuille présentent des coefficients majoritairement négatifs et non significatifs. Seule l'interaction avec le portefeuille Q4, qui regroupe les entreprises les plus médiatisées, est significative (coefficient = -0.004 ; valeur $p = 0.027$), mais son signe est contraire à celui de l'hypothèse alternative $H_{1;3}$, qui anticipait une hausse de la volatilité. L'effet attendu n'apparaît que pour le portefeuille Q2 (coefficient = 0.002), sans atteindre la significativité (valeur $p = 0.856$).

Dans le cas de notre test unilatéral, aucun effet significatif n'est observé dans le sens attendu. Toutefois, certains résultats restent cohérents avec les observations visuelles de la **Figure 6**. La volatilité structurellement plus élevée des entreprises sans couverture médiatique transparaît dans la constante du modèle, tandis que, pour les portefeuilles exposés médiatiquement, les coefficients d'interaction suivent globalement une tendance décroissante à mesure que la couverture médiatique augmente (à l'exception de Q2). Ce constat contredit à la fois l'hypothèse de recherche, mais aussi les conclusions de Jeon, McCurdy et Zhao (2021), ainsi que d'Antweiler et Frank (2002), qui soutenaient que les entreprises bénéficiant d'une plus grande couverture médiatique tendent à être associées à une volatilité plus élevée ou à des mouvements de prix plus brusques.

Comparativement au modèle sans interactions, dans lequel la fréquence des nouvelles exerçait un effet globalement négatif et significatif, l'inclusion des termes d'interaction met en évidence une forte hétérogénéité entre les portefeuilles. L'effet positif anticipé n'est observé que pour le portefeuille Q2, sans être significatif. Enfin, l'estimation obtenue pour le portefeuille Zéro confirme un niveau de volatilité structurellement élevé en l'absence totale de couverture médiatique.

Tableau 17 - Effet hétérogène de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel: estimation IV-2SLS avec interactions de portefeuille

Variable dépendante:		Semi-Annual Arithmetic Average Return				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	1.034***	0.341	3.031	0.002	0.365	1.703
MKT-RF Semi-Annual Return	85.927***	5.380	15.971	0.000	75.382	96.472
CMA Semi-Annual Return	10.159	14.187	0.716	0.474	-17.646	37.964
log(Mp,t)	-0.042***	0.015	-2.903	0.004	-0.071	-0.014
Betap,t_standardized	0.026	0.018	1.484	0.138	-0.008	0.060
MeanFreqp,t_Portfolio_Q1	-1.024	1.073	-0.955	0.340	-3.127	1.078
MeanFreqp,t_Portfolio_Q2	-0.034	0.343	-0.099	0.921	-0.707	0.639
MeanFreqp,t_Portfolio_Q3	0.410	0.265	1.546	0.122	-0.110	0.931
MeanFreqp,t_Portfolio_Q4	0.388**	0.196	1.976	0.048	0.003	0.772
R² global	0.718					
R² ajusté	0.702					
Statistique F	349.356					
Valeur p du test F	0.000					

Le Tableau 17 présente les résultats de la régression IV-2SLS qui estime l'effet hétérogène de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel selon les portefeuilles. La variable dépendante est le rendement moyen arithmétique semestriel. Les variables explicatives incluent MKT-RF Semi-Annual Return, CMA Semi-Annual Return, log(Mp,t), Betap,t_standardized, ainsi que les interactions entre MeanFreqp,t et les dummys de portefeuille (Q1, Q2, Q3 et Q4). Le portefeuille Zero sert de catégorie de référence. L'estimation utilise les valeurs retardées de MeanFreqp,t, en interaction avec les dummys de portefeuille, ainsi que log(Mp,t)_lag et Betap,t_standardized_lag, comme instruments des variables endogènes. Le tableau affiche les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p et les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour chaque variable explicative. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Avec une statistique F de 349.356 et une valeur p de 0.000, le modèle est globalement significatif.

La constante du modèle représente la moyenne arithmétique semestrielle estimée des rendements pour le portefeuille Zéro, composé d'entreprises n'ayant fait l'objet d'aucune couverture

médiatique et servant de catégorie de référence dans les variables indicatrices de portefeuille. Lorsque toutes les autres variables explicatives sont à leur niveau de référence, ce rendement est estimé à 1.034, avec une valeur p de 0.002. Ce résultat ne porte pas directement sur l'effet marginal de la fréquence des nouvelles, mais suggère que l'absence totale de couverture médiatique n'est pas nécessairement associée à des rendements plus faibles. En effet, les entreprises du portefeuille Zéro présentent, en moyenne, des rendements positifs et significatifs. Cette observation rejoint partiellement les travaux de Fang et Peress (2008), qui mettent en évidence une prime liée à l'absence de couverture médiatique : selon eux, les entreprises non médiatisées génèrent des rendements excédentaires plus élevés que celles fortement couvertes. Cette dynamique est ici particulièrement marquée lorsqu'on compare le portefeuille Zéro aux portefeuilles Q3 et Q4.

Les termes d'interaction entre la fréquence des nouvelles MeanFreq_{p, t}, et les variables indicatrices de portefeuille révèlent des dynamiques hétérogènes selon le niveau de couverture médiatique. Seuls les portefeuilles Q1 et Q2, correspondant aux entreprises les moins médiatisées, présentent des coefficients négatifs, conformes au signe attendu selon l'hypothèse alternative H_{1;4}. Toutefois, ces effets ne sont pas statistiquement significatifs : -1.024 avec une valeur p de 0.340 pour Q1, et -0.034 avec une valeur p de 0.921 pour Q2.

En revanche, les portefeuilles Q3 et Q4, regroupant respectivement les entreprises bénéficiant d'une couverture médiatique élevé et très élevé, présentent des coefficients positifs (0.410 pour Q3 et 0.388 pour Q4), à l'inverse du signe anticipé. Seul l'effet associé au portefeuille Q4 est statistiquement significatif (valeur p = 0.048), tandis que celui de Q3 ne l'est pas (valeur p = 0.122). Ces résultats contredisent une fois de plus ceux d'Antweiler et Frank (2002), qui soutenaient que les entreprises bénéficiant d'une forte couverture médiatique avaient tendance à afficher des rendements futurs anormalement faibles.

Dans le cadre de notre test unilatéral, aucun effet significatif n'est observé dans le sens attendu. Comparativement au modèle sans interactions, dans lequel la fréquence des nouvelles exerçait un effet globalement positif mais non significatif, l'introduction des termes d'interactions met en évidence une hétérogénéité marquée selon le niveau de couverture médiatique. L'effet négatif

attendu ne se manifeste que dans les portefeuilles les moins médiatisées (Q1 et Q2), sans atteindre toutefois le seuil de significativité.

6.2 Analyse par sous-périodes autour de la crise financière de 2008

Ce second test de robustesse vise à déterminer si l'effet de la fréquence des nouvelles ($\text{MeanFreq}_{p,t}$) sur les caractéristiques de la distribution des rendements varie selon les différentes phases du cycle économique. Pour cela, une analyse par sous-périodes a été menée autour de la crise financière de 2008.

L'ensemble des données a ainsi été divisé en trois sous-périodes distinctes:

- Période pré-crise: du second semestre 2000 au premier semestre 2007;
- Période de crise: du second semestre 2007 au premier semestre 2009;
- Période post-crise: du second semestre 2009 au second semestre 2015.

Tableau 18 - Impact de la fréquence des nouvelles sur la skewness semestrielle avant, pendant, et après la crise de 2008: estimations IV-2SLS

Variable dépendante:		Semi-Annual Skewness	
Variable explicative:		MeanFreqp,t	
Paramètres estimés	Pré-crise	Crise Financière 2008	Post-crise
Coefficient	0.917**	-0.030	-0.948*
Erreur standard	0.450	0.349	0.574
Statistique t	2.039	-0.087	-1.653
Valeur p	0.042	0.931	0.098
IC inférieur	0.036	-0.714	-2.072
IC supérieur	1.799	0.654	0.176
R² global	0.152	0.647	0.058
R² ajusté	0.110	0.581	0.011
Statistique F	17.566	36.766	5.133
Valeur p du test F	0.001	0.000	0.162

Le Tableau 18 présente les résultats de l'estimation IV-2SLS analysant l'impact de la fréquence des nouvelles sur la skewness semestrielle avant, pendant et après la crise financière de 2008. L'échantillon est divisé en trois sous-périodes distinctes : la période pré-crise (2000S2–2007S1), la crise (2007S2–2009S1) et la période post-crise (2009S2–2015S2). La variable dépendante est la skewness semestrielle. Le tableau présente les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p ainsi que les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour la variable explicative d'intérêt, MeanFreqp,t. L'estimation utilise MeanFreqp,t_lag comme instrument de MeanFreqp,t. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10 %, 5 % et 1 %.

Avec des statistiques F et des valeurs p respectives de 17.566 et 0.001 (période pré-crise), 36.766 et 0.000 (période de crise), et 5.133 et 0.162 (période post-crise), les modèles sont globalement significatifs pour les périodes pré-crise et de crise, mais pas pour la période post-crise.

L'effet de la fréquence des nouvelles (MeanFreqp,t) sur la skewness semestrielle varie sensiblement selon la phase du cycle économique. Avant la crise, cet effet est positif et statistiquement significatif (coefficients de 0.917 ; valeur p = 0.042), ce qui indique qu'une couverture médiatique

accrue est associée à une asymétrie des rendements davantage orientée vers les gains. Ce résultat n'est pas surprenant, dans la mesure où il s'inscrit dans un contexte économique globalement stable, malgré les turbulences liées à l'éclatement de la bulle technologique au début des années 2000. Durant cette période, une intensité médiatique plus élevée aurait pu amplifier les comportements optimistes des investisseurs sur les marchés.

En revanche, pendant la crise financière, l'effet estimé devient légèrement négatif et non significatif (coefficients de -0.030 ; valeur $p = 0.931$), ce qui suggère qu'aucun impact significatif de la fréquence des nouvelles sur l'asymétrie des rendements n'est détectable dans un contexte de forte instabilité, de panique, et de surcharge informationnelle.

Après la crise, l'effet de la fréquence des nouvelles reste négatif et s'intensifie (coefficients de -0.948), tout en demeurant non significatif (valeur $p = 0.098$). Toutefois, compte tenu de la non-significativité globale du modèle pour cette période, il n'est pas possible de conclure de manière fiable à l'existence d'un effet systématique de la fréquence des nouvelles sur la skewness des rendements dans un contexte de reprise post-crise.

Dans le cadre de notre test unilatéral, seuls les effets estimés pendant la crise et la période post-crise vont dans le sens attendu par l'hypothèse de recherche, bien que ces relations ne soient pas significatives. À l'inverse, le seul effet statistiquement significatif, observé pendant la période pré-crise, contredit cette hypothèse en affichant un coefficient positif.

Tableau 19 - Impact de la fréquence des nouvelles sur la kurtosis semestrielle avant, pendant, et après la crise de 2008: estimations IV-2SLS

Variable dépendante:		Semi-Annual Kurtosis	
Variable explicative:		MeanFreqp,t	
Paramètres estimés	Pré-crise	Crise Financière 2008	Post-crise
Coefficient	9.329**	3.744***	6.657*
Erreurs standard	3.968	1.266	3.490
Statistique t	2.351	2.958	1.907
Valeur p	0.019	0.003	0.057
IC inférieur	1.552	1.263	-0.183
IC supérieur	17.106	6.225	13.497
R² global	-0.110	0.752	-0.271
R² ajusté	-0.184	0.686	-0.356
Statistique F	7.182	61.584	13.977
Valeur p du test F	0.127	0.000	0.007

Le Tableau 19 présente les résultats de l'estimation IV-2SLS analysant l'impact de la fréquence des nouvelles sur la kurtosis semestrielle avant, pendant et après la crise financière de 2008. L'échantillon est divisé en trois sous-périodes distinctes : la période pré-crise (2000S2–2007S1), la crise (2007S2–2009S1) et la période post-crise (2009S2–2015S2). La variable dépendante est la kurtosis semestrielle. Le tableau présente les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p ainsi que les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour la variable explicative d'intérêt, MeanFreqp,t. L'estimation utilise MeanFreqp,t_lag comme instrument de MeanFreqp,t. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10 %, 5 % et 1 %.

Avec des statistiques F et des valeurs p respectives de 7.182 et 0.127 (période pré-crise), 61.584 et 0.000 (période de crise), et 13.977 et 0.007 (période post-crise), les modèles sont globalement significatifs pour les périodes de crise et post-crise, mais pas pour la période pré-crise.

L'effet de la fréquence des nouvelles sur la kurtosis reste positif quelle que soit la phase du cycle économique, bien que son niveau de significativité varie selon les périodes. Avant la crise financière de 2008, cet effet est positif et statistiquement significatif (coefficient de 9.329 ; valeur p = 0.019), ce qui suggère qu'une visibilité médiatique accrue est associée à une probabilité plus

élevée d'observer des rendements extrêmes dans un environnement économique relativement stable. Toutefois, la non-significativité globale du modèle invite à interpréter ce résultat avec prudence.

Pendant la crise, l'effet demeure positif, bien que plus modéré (coefficient de 3.744), et reste hautement significatif (valeur $p = 0.003$), ce qui indique qu'une fréquence accrue de nouvelles s'accompagne d'une intensification des rendements extrêmes dans un contexte de forte instabilité des marchés.

Après la crise, l'effet reste également positif (coefficient de 6.657), mais devient statistiquement non significatif avec une valeur p de 0.057, légèrement supérieure au seuil conventionnel de 5% retenu dans cette étude. Ainsi, selon le cadre méthodologique adopté ici, l'impact de la fréquence des nouvelles sur les rendements extrêmes apparaît moins marqué dans un contexte de reprise post-crise. Cependant, la proximité de cette valeur avec le seuil invite à nuancer cette interprétation.

Dans le cadre de notre test unilatéral, les effets estimés au cours des trois périodes du cycle économique restent globalement cohérents avec l'hypothèse de recherche, qui postule une relation positive entre fréquence des nouvelles et kurtosis. Toutefois, seuls les effets observés en périodes pré-crise et de crise sont statistiquement significatifs au seuil de 5%, tandis que l'effet post-crise demeure proche de ce seuil sans l'atteindre.

Tableau 20 - Impact de la fréquence des nouvelles sur la volatilité semestrielle avant, pendant, et après la crise de 2008: estimations IV-2SLS

Variable dépendante:		Semi-Annual Volatility	
Variable explicative:		MeanFreqp,t	
Paramètres estimés	Pré-crise	Crise Financière 2008	Post-crise
Coefficient	-0.006***	-0.006	-0.005
Erreurs standard	0.002	0.005	0.003
Statistique t	-2.667	-1.364	-1.532
Valeur p	0.008	0.173	0.126
IC inférieur	-0.010	-0.015	-0.011
IC supérieur	-0.002	0.003	0.001
R² global	0.941	0.992	0.685
R² ajusté	0.935	0.988	0.653
Statistique F	733.130	5354.000	110.170
Valeur p du test F	0.000	0.000	0.000

Le Tableau 20 présente les résultats de l'estimation IV-2SLS analysant l'impact de la fréquence des nouvelles sur la volatilité semestrielle avant, pendant et après la crise financière de 2008. L'échantillon est divisé en trois sous-périodes distinctes : la période pré-crise (2000S2–2007S1), la crise (2007S2–2009S1) et la période post-crise (2009S2–2015S2). La variable dépendante est la volatilité semestrielle. Le tableau présente les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p ainsi que les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour la variable explicative d'intérêt, MeanFreqp,t. L'estimation utilise MeanFreqp,t_lag comme instrument de MeanFreqp,t. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10 %, 5 % et 1 %.

Avec des statistiques F respectives de 733.130 (période pré-crise), 5354.000 (période de crise) et 110.170 (période post-crise), toutes associées à une valeur p de 0.000, les modèles sont globalement significatifs pour les trois phases du cycle économique.

Bien que la significativité de l'effet varie selon les périodes, la fréquence des nouvelles conserve un effet négatif sur la volatilité, quelle que soit la phase du cycle économique. Avant la crise financière, cet effet est à la fois négatif et statistiquement significatif (coefficient de -0.006 ; valeur

$p = 0.008$), ce qui indique qu'une couverture médiatique plus intense s'accompagne d'une volatilité plus faible dans un environnement de marché relativement stable.

En revanche, cet effet devient non significatif pendant la crise (coefficients de -0.006 ; valeur $p = 0.173$) ainsi que durant la période post-crise (coefficients de -0.005 ; valeur $p = 0.126$), ce qui suggère que, dans des contextes de forte incertitude ou de reprise économique, l'impact de la fréquence des nouvelles sur la volatilité des rendements n'est pas statistiquement établi.

Dans le cadre de notre test unilatéral, les coefficients estimés sont systématiquement négatifs, ce qui va à l'encontre de l'hypothèse de recherche selon laquelle une fréquence accrue de nouvelles amplifierait la volatilité des rendements. Ces résultats suggèrent plutôt un effet modérateur. Toutefois, seul l'effet observé en période pré-crise est statistiquement significatif, ce qui limite la portée des conclusions quant à la robustesse et la constance de cet effet à travers le cycle économique.

Tableau 21 - Impact de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel avant, pendant, et après la crise de 2008: estimations IV-2SLS

Variable dépendante:	Semi-Annual Arithmetic Average Return		
Variable explicative:	MeanFreqp,t		
Paramètres estimés	Pré-crise	Crise Financière 2008	Post-crise
Coefficient	0.537**	0.259	-0.129
Erreurs standard	0.247	1.011	0.288
Statistique t	2.177	0.256	-0.446
Valeur p	0.030	0.798	0.656
IC inférieur	0.054	-1.723	-0.694
IC supérieur	1.021	2.241	0.437
R² global	0.493	0.705	0.777
R² ajusté	0.450	0.599	0.758
Statistique F	152.050	130.100	161.390
Valeur p du test F	0.000	0.000	0.000

Le Tableau 21 présente les résultats de l'estimation IV-2SLS analysant l'impact de la fréquence des nouvelles sur le rendement moyen arithmétique semestriel avant, pendant et après la crise financière de 2008. L'échantillon est divisé en trois sous-périodes distinctes : la période pré-crise (2000S2–2007S1), la crise (2007S2–2009S1) et la période post-crise (2009S2–2015S2). La variable dépendante est le rendement moyen arithmétique semestriel. Le tableau présente les coefficients estimés, les erreurs standards, les statistiques t, les valeurs p ainsi que les intervalles de confiance à 95 % (IC inférieur et IC supérieur) pour la variable explicative d'intérêt, MeanFreqp,t. L'estimation utilise MeanFreqp,t_lag comme instrument de MeanFreqp,t. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10 %, 5 % et 1 %.

Avec des statistiques F de 152.050 (période pré-crise), 130.100 (période de crise) et 161.390 (période post-crise), toutes associées à une valeur p de 0.000, les modèles sont globalement significatifs pour les trois phases du cycle économique.

L'effet de la fréquence des nouvelles (MeanFreqp,t) sur le rendement moyen arithmétique fluctue selon les phases du cycle économique, tant en termes de signe que de significativité statistique. Avant la crise financière, cet effet est positif et significatif (coefficient = 0.537 ; valeur p = 0.030),

ce qui indique que, dans un contexte de stabilité économique, une intensité informationnelle accrue est associée à des rendements moyens plus élevés.

Pendant la crise, bien que l'effet reste positif (coefficients = 0.259), la relation devient largement non significative (valeur $p = 0.798$), ce qui suggère que la fréquence des nouvelles ne permet pas d'expliquer la performance boursière dans un contexte d'incertitude extrême.

Dans la période post-crise, la relation s'inverse et devient légèrement négative (coefficients = -0.129), tout en restant non significative (valeur $p = 0.656$). Ce résultat, bien que non significatif, pourrait refléter un changement dans la manière dont les marchés interprètent les nouvelles, en particulier dans les phases de reprise économique, notamment en ce qui concerne les rendements moyens.

Dans le cadre de notre test unilatéral, les effets observés en périodes pré-crise et de crise s'écartent de l'hypothèse de recherche, en suggérant qu'une fréquence plus élevée des nouvelles serait associée à des rendements moyens plus élevés. De plus, seul l'effet observé avant la crise est significatif. En revanche, le coefficient estimé est conforme au signe anticipé par cette hypothèse en période post-crise, bien qu'il demeure non significatif. Ces résultats nuancent donc la robustesse et la stabilité temporelle de l'effet observé.

7. Résumé et interprétation des résultats

7.1 Limites de l'étude

L'une des forces de cette étude réside dans le fait qu'elle compte parmi les premières à explorer la relation entre l'information et les moments d'ordre supérieur de la distribution des rendements, à savoir la skewness et la kurtosis, en utilisant la fréquence des nouvelles comme proxy de l'information. De plus, la robustesse de cette relation a été évaluée sur une période étendue de 15 ans, permettant d'englober plusieurs cycles économiques majeurs, notamment la crise financière de 2008.

Néanmoins, les résultats doivent être interprétés avec prudence. En raison de contraintes de ressources, l'accès aux données sur les articles de presse publiés pour chaque entreprise de l'échantillon s'est révélé particulièrement difficile. Même après l'obtention de ces données, la faible dispersion observée (notamment une forte concentration autour de faibles niveaux de fréquence) a compliqué la construction des portefeuilles, en dépit de l'approche par quartiles retenue.

Par ailleurs, comme évoqué dans la revue de littérature, la relation entre les nouvelles et la dynamique des marchés financiers, qu'il s'agisse des réactions de prix ou de la volatilité des actifs, est plus complexe qu'il n'y paraît. À ce titre, la fréquence brute des nouvelles ne permet pas, à elle seule, de saisir toute cette complexité. Des recherches futures gagneraient ainsi à intégrer des analyses textuelles avancées (par exemple, via des outils de traitement automatique du langage naturel) afin de mesurer et d'incorporer des indicateurs tels que le sentiment (positif, négatif, neutre), l'incertitude, la redondance, l'objectivité ou la subjectivité du contenu, ainsi que de distinguer les types de nouvelles (macroéconomiques, politiques, monétaires).

D'un point de vue méthodologique, le choix d'une granularité semestrielle constituait un compromis : s'il a permis de faire émerger des tendances structurelles, il a probablement entraîné une perte de précision, en lissant certains effets de court terme dans la moyenne sur six mois. De même, le recours à la sélection par élimination rétrograde, combinée aux critères AIC, BIC et à la log-vraisemblance pour déterminer la spécification finale du modèle, a très certainement conduit à un optimum local. Autrement dit, la combinaison finale de variables retenue dans chaque

modèle ne constitue pas nécessairement la meilleure combinaison globale possible, dans la mesure où l'approche adoptée a pu reléguer au second plan la cohérence économique du modèle.

Par ailleurs, dans les analyses portant sur la kurtosis et le rendement moyen arithmétique, les variables explicatives représentant les facteurs de Fama-French ont été traitées comme exogènes, malgré la détection occasionnelle d'endogénéité par le test de Durbin-Hausman. Conformément à Wooldridge (2010), la théorie économique justifie le traitement d'une variable comme exogène lorsque la mise en évidence empirique de son endogénéité semble économiquement peu plausible. En ce sens, les facteurs de Fama-French, construits pour refléter les risques systématiques du marché, ont été considérés comme exogènes dans ces cas particuliers.

Au cours du processus d'estimation, des problèmes d'hétéroscédasticité ont été détectés dans les analyses de la volatilité et du rendement moyen arithmétique, tandis qu'une autocorrélation a également été observée dans l'analyse de la volatilité. Pour corriger ces problèmes, des erreurs standards robustes ont été utilisées. Bien que la clusterisation des erreurs standards par entité soit techniquement possible, le très faible nombre d'unités (cinq portefeuilles) soulevait, dans la présente étude, de sérieuses préoccupations quant à la fiabilité des inférences obtenues. Par conséquent, des corrections robustes classiques ont été privilégiées afin d'assurer une estimation plus cohérente, y compris dans l'analyse de la volatilité, malgré la présence d'autocorrélation.

Dans une perspective future, les études portant sur un sujet similaire gagneraient à s'appuyer sur des bases de données plus complètes et diversifiées pour collecter les informations médiatiques au niveau des entreprises, afin de dépasser les limitations liées à une faible dispersion. Par ailleurs, l'intégration d'une analyse de sentiment constituerait une amélioration méthodologique précieuse. Cela permettrait de mieux comprendre les mécanismes complexes par lesquels l'information influence la distribution des rendements boursiers, tout en offrant une interprétation plus fine et nuancée.

7.2 Conclusion

Au regard des résultats obtenus, nos hypothèses postulant qu'une augmentation de la fréquence des nouvelles entraîne respectivement une diminution significative de la skewness des rendements boursiers et du rendement moyen arithmétique ne sont pas confirmées empiriquement. Les effets observés vont même à l'encontre des relations négatives anticipées. Concernant la volatilité, une couverture médiatique plus intense est liée à une baisse significative de celle-ci. Ce constat, bien que significatif, va à l'encontre de l'effet amplificateur attendu et infirme ainsi l'hypothèse de recherche initiale.

En revanche, les résultats confirment l'hypothèse formulée pour la kurtosis : une augmentation de la couverture médiatique s'accompagne d'une hausse significative de la kurtosis des rendements boursiers futurs. Cela implique une probabilité accrue d'observer des rendements extrêmes, qu'ils soient positifs ou négatifs, et révèle ainsi une distribution à queues épaisses.

Pour affiner cette analyse, des termes d'interaction entre la fréquence des nouvelles et des variables indicatrices de portefeuille ont été intégrés au modèle. Cela a permis de mettre en évidence un effet hétérogène de la couverture médiatique sur les différentes caractéristiques de la distribution des rendements étudiées.

En ce qui concerne la skewness, une intensité informationnelle plus forte est associée à une asymétrie négative plus marquée pour les portefeuilles faiblement médiatisés ; asymétrie qui s'atténue à mesure que la couverture augmente. Toutefois, ces effets ne sont pas statistiquement significatifs. Le portefeuille de référence (Zero) présente quant à lui une skewness négative significative, ce qui traduit une asymétrie structurelle des rendements orientée davantage vers les pertes en l'absence totale d'information publique.

S'agissant de la kurtosis, une fréquence plus élevée des nouvelles coïncide avec une probabilité accrue d'observer des rendements extrêmes, en particulier dans les portefeuilles les plus exposés. Bien que cette relation soit visible sur l'ensemble des portefeuilles, seuls ceux fortement médiatisés présentent des coefficients significatifs, ce qui suggère que pour les entreprises les plus visibles, la couverture médiatique contribue à accentuer les extrêmes dans la distribution future des rendements.

Contrairement aux travaux de Jeon, McCurdy et Zhao, ainsi que d'Antweiler et Frank, qui mettaient en évidence un effet amplificateur sur la volatilité, nos résultats révèlent qu'une fréquence accrue des nouvelles est généralement liée à une diminution de la volatilité future. Bien que la plupart des portefeuilles ne présentent pas de relation significative, l'effet négatif devient plus marqué avec le niveau d'exposition médiatique, et significatif uniquement pour le portefeuille le plus couvert (Q4). À l'inverse, le portefeuille Zéro, en l'absence d'information publique, se caractérise par une volatilité structurellement élevée et statistiquement significative.

Enfin, l'analyse du rendement moyen arithmétique révèle des dynamiques contrastées. Une fréquence accrue des nouvelles se manifeste par des rendements futurs plus faibles dans les portefeuilles peu médiatisés et plus élevés dans les portefeuilles fortement couverts. Toutefois, seul l'effet observé pour Q4 est statistiquement significatif. De son côté, le portefeuille Zéro affiche un rendement moyen significativement positif, suggérant ainsi que l'absence d'information publique ne nuit pas nécessairement à la performance boursière.

Au-delà de ces résultats, l'analyse par phase du cycle économique indique que la fréquence des nouvelles conserve un effet positif sur la kurtosis et négatif sur la volatilité des rendements, même si la significativité de ces relations varie selon les périodes. En revanche, l'impact sur la skewness et le rendement moyen arithmétique demeure plus incertain, les coefficients fluctuant tant dans leur direction que dans leur significativité au fil des phases du cycle.

Mise en garde : Il convient toutefois de souligner que les données de fréquence des nouvelles utilisées dans cette étude proviennent de la base RavenPack et n'ont pas fait l'objet d'une validation indépendante. Par conséquent, les résultats présentés doivent être interprétés avec prudence, dans la mesure où toute faiblesse dans la qualité des données pourrait introduire un biais de type « *garbage in, garbage out* ».

8. Bibliographie

1. **Al-Maadid, A., Caporale, G. M., Spagnolo, F., & Spagnolo, N. (2018).** The Impact of Business and Political News on the GCC Stock Markets. *CESifo Working Paper No. 7353*. Available at SSRN: <https://ssrn.com/abstract=3338680>
2. **Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015).** Does realized skewness predict the cross-section of equity returns? *SSRN Electronic Journal*. <https://ssrn.com/abstract=1898735>
3. **Antweiler, W., & Frank, M. Z. (2002).** Internet stock message boards and stock returns. *Journal of Business*, 77(2), 125–163.
4. **Antweiler, W., & Frank, M. Z. (2006).** *Do U.S. stock markets typically overreact to corporate news stories?* SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.878091>
5. **Baldazzi, P., Elton, E. J., & Green, T. C. (2001).** Economic news and bond prices: Evidence from the U.S. Treasury market. *Journal of Financial and Quantitative Analysis*, 36(4), 523–543. <https://doi.org/10.2307/2676223>
6. **Berry, T. D., & Howe, K. M. (1994).** Public information arrival. *The Journal of Finance*, 49(4), 1331–1346. <https://doi.org/10.1111/j.1540-6261.1994.tb02456.x>
7. **Bomfim, A. N. (2000).** *Pre-announcement effects, news, and volatility: Monetary policy and the stock market.* Finance and Economics Discussion Series, 2000-50. Board of Governors of the Federal Reserve System. <https://doi.org/10.17016/FEDS.2000.50>
8. **Chan, W. S. (2001).** Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*, 70(2), 223–260. [https://doi.org/10.1016/S0304-405X\(03\)00146-6](https://doi.org/10.1016/S0304-405X(03)00146-6)
9. **Chiang, S. Y. (2016).** *News which moves the market: Assessing the impact of published financial news on the stock market* [Thèse de doctorat, Singapore Management University]. SMU Institutional Knowledge.
10. **DellaVigna, S., & Pollet, J. M. (2005).** *Investor inattention and Friday earnings announcements* (Working Paper No. 14511). National Bureau of Economic Research. <http://dx.doi.org/10.2139/ssrn.843786>
11. **De Bondt, W. F. M., & Thaler, R. (1985).** Does the stock market overreact? *The Journal of Finance*, 40(3), 793–805. <https://doi.org/10.1111/j.1540-6261.1985.tb05004.x>
12. **Ederington, L. H., & Lee, J. H. (1996).** The creation and resolution of market uncertainty: The impact of information releases on implied volatility. *Journal of Financial and Quantitative Analysis*, 31(4), 513–539. <https://doi.org/10.2307/2331235>
13. **Engle, R. F., & Ng, V. K. (1991).** Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5), 1749–1778.
14. **Essaddam, M. N. (2002).** *Impact du risque politique sur les taux de rendement boursier et leur volatilité : Le cas du Québec* (Thèse de doctorat, Université Laval). Thèses Canada.
15. **Fang, L., & Peress, J. (2008).** Media coverage and the cross-section of stock returns. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.971202>
16. **Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969).** The adjustment of stock prices to new information. *International Economic Review*, 10(1), 1–21. <https://doi.org/10.2307/2525569>

17. Ferguson, N. J., Philip, D., Lam, H. Y. T., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, 19(3), 157–200.
18. Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *The Review of Financial Studies*, 15(3), 751–782.
19. García, D. (2012). Sentiment during recessions. *SSRN Electronic Journal*.
<http://dx.doi.org/10.2139/ssrn.1571101>
20. Huberman, G., & Regev, T. (1999). Speculating on a cure for cancer: A non-event that made stock prices soar. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.172448>
21. Jeon, Y., McCurdy, T. H., & Zhao, X. (2021). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *SSRN Electronic Journal*. Available at SSRN: <https://ssrn.com/abstract=3911398>
22. Jondeau, E., & Rockinger, M. (2000). *Conditional volatility, skewness, and kurtosis: Existence and persistence* (NER #77). Banque de France.
<http://dx.doi.org/10.2139/ssrn.260910>
23. Lahbous, A., & Zammar, R. (2021). *L'importance de la prise en considération des moments d'ordre supérieur dans l'optimisation du portefeuille*. African Scientific Journal, 3(8), 126–141.
24. Lezmi, E., Malongo, H., Roncalli, T., & Sobotka, R. (2019). *Portfolio allocation with skewness risk: A practical guide*. SSRN Electronic Journal.
<http://dx.doi.org/10.2139/ssrn.3201319>
25. Peress, J. (2011). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *SSRN Electronic Journal*.
<http://dx.doi.org/10.2139/ssrn.2019834>
26. Premaratne, G., & Bera, A. K. (2000). Modeling asymmetry and excess kurtosis in stock return data. *University of Illinois Working Paper, Department of Economics*, No. 00-0123. <https://doi.org/10.2139/ssrn.259009>
27. Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C: The art of scientific computing* (2nd ed.). Cambridge University Press. ISBN 0-521-43108-5.
28. Rigobon, R., & Sack, B. (2001). Measuring the reaction of monetary policy to the stock market. *National Bureau of Economic Research Working Paper Series* (No. 8794). <http://dx.doi.org/10.2139/ssrn.268288>
29. Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), 458–464.
<https://doi.org/10.1016/j.dss.2012.03.001>
30. Strycharz, J., Strauss, N., & Trilling, D. (2018). The role of media coverage in explaining stock market fluctuations: Insights for strategic financial communication. *International Journal of Strategic Communication*, 12(1), 67–85.
<https://doi.org/10.1080/1553118X.2017.1378220>
31. Tetlock, P. C. (2010). *All the news that's fit to reprint: Do investors react to stale information?* [Unpublished manuscript]. Columbia University.
<http://dx.doi.org/10.2139/ssrn.1018221>
32. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.685145>

- 33. Veronesi, P. (1999).** Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *The Review of Financial Studies*, 12(5), 975–1007.
<https://doi.org/10.1093/rfs/12.5.975>
- 34. Wooldridge, J. M. (2010).** *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press. ISBN 978-0-262-23258-6.

9. Annexes

Tableau 22 - Résultats du test de Durbin-Hausman concernant l'endogénéité de Liquidityp,t (méthode de régression auxiliaire) dans l'analyse de la skewness semestrielle

Variable dépendante:		Semi-Annual Skewness					
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur	
const	-1.185	0.474	-2.499	0.014	-2.122	-0.248	
MeanFreqp,t	0.767	0.516	1.487	0.139	-0.252	1.787	
Liquidityp,t	42.640	17.057	2.500	0.013	8.939	76.342	
SMB Semi-Annual Return	28.829	13.304	2.167	0.032	2.544	55.114	
residuals_Liquidityp,t	-31.446*	17.270	-1.821	0.071	-65.567	2.676	
R² global	0.117						
R² ajusté	0.099						
Statistique F	6.654						
Valeur p du test F	0.000						

Le Tableau 22 présente les résultats du test de Durbin-Hausman appliqué au modèle de régression OLS groupée. Ce test évalue l'exogénéité de la variable explicative Liquidityp, t, en comparant les estimateurs OLS et 2SLS. Une différence statistiquement significative entre les deux séries d'estimateurs indique la présence d'endogénéité, tandis que son absence soutient l'hypothèse d'exogénéité de la variable. La méthode utilisée repose sur une régression auxiliaire, dans laquelle la significativité du coefficient associé aux résidus permet de conclure à la présence ou non d'endogénéité. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Tableau 23 - Résultats du test de Durbin-Hausman concernant l'endogénéité de SMB Semi-Annual Return (méthode de régression auxiliaire) dans l'analyse de la skewness semestrielle

Variable dépendante:		Semi-Annual Skewness				
Variables explicatives	Coefficient	Erreurs standard	Statistique t	Valeur p	IC inférieur	IC supérieur
const	-0.333	0.093	-3.586	0.000	-0.516	-0.149
MeanFreqp,t	0.013	0.308	0.042	0.967	-0.596	0.622
Liquidityp,t	11.597	3.060	3.790	0.000	5.551	17.643
SMB Semi-Annual Return	54.890	28.836	1.904	0.059	-2.084	111.865
residuals_SMB Semi-Annual Return	54.805*	28.838	1.900	0.059	-2.173	111.783
R² global	0.117					
R² ajusté	0.099					
Statistique F	6.654					
Valeur p du test F	0.000					

Le Tableau 23 présente les résultats du test de Durbin-Hausman appliqué au modèle de régression OLS groupée. Ce test évalue l'exogénéité de la variable explicative SMB Semi-Annual Return, en comparant les estimateurs OLS et 2SLS. Une différence statistiquement significative entre les deux séries d'estimateurs indique la présence d'endogénéité, tandis que son absence soutient l'hypothèse d'exogénéité de la variable. La méthode utilisée repose sur une régression auxiliaire, dans laquelle la significativité du coefficient associé aux résidus permet de conclure à la présence ou non d'endogénéité. Les symboles *, ** et *** indiquent que les coefficients sont statistiquement significatifs aux seuils respectifs de 10%, 5% et 1%.

Code Python utilisé pour le diagnostic de la linéarité du modèle OLS via le test RESET de Ramsey

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.stats.diagnostic import linear_reset
file_path = "/Users/kevintiacoh/Desktop/NOUVELLES
DONNÉES3/FICHIER_EQUATION_REGRESSION_SKEWNESS_FINAL_POOLEDOLS.csv"
"
data = pd.read_csv(file_path, delimiter=';')
# We verify that all the variables are present in the data.
explanatory_variables = [
    "MeanFreqp,t",
    "Liquidityp,t",
    "SMB Semi-Annual Return"
]
for var in ["Semi-Annual Skewness"] + explanatory_variables:
```

```

if var not in data.columns:
    raise ValueError(f'Error : The variable '{var}' does not exist in the data.')

# We make sure that all explanatory variables are numeric.
for var in explanatory_variables:
    data[var] = pd.to_numeric(data[var], errors="coerce")
# We define the dependent variable and the matrix of explanatory variables.
y = data["Semi-Annual Skewness"]
X = data[explanatory_variables]
X = sm.add_constant(X)
# We fit the OLS regression model.
model = sm.OLS(y, X, missing='drop').fit()
# We apply the Ramsey RESET test on the overall model.
reset_test = linear_reset(model, power=2, use_f=True)
p_value_reset = reset_test.pvalue
print("Ramsey RESET test on the global model involving all variables.")
print(f"P-value : {p_value_reset:.5f}")
if p_value_reset > 0.05:
    print("The overall model appears to be well-specified and linear.")
else:
    print("Non-linearity is detected in the overall model.")
    print("It might be useful to test additional transformations.")

```

