



Les données numériques pour la prévision des rendements boursiers : Applications de l’outil Google Trends

Mémoire

Marc-André Bergeron

Maîtrise en économie - avec mémoire
Maître ès arts (M.A.)

Québec, Canada

© Marc-André Bergeron, 2020

LES DONNÉES NUMÉRIQUES POUR LA PRÉVISION DES RENDEMENTS
BOURSIERS : APPLICATIONS DE L'OUTIL GOOGLE TRENDS

Mémoire de Maitrise
Par Marc-André Bergeron

Sous la direction de
Gordon, Stephen (Directeur)
Samson, Lucie (Co-Directrice)

Département d'économie
Faculté des sciences sociales
Université Laval

Remis le 27 avril 2020

Résumé

L'accès à l'information financière est un déterminant important pour la prévision des rendements boursiers (mesurés par la variation de la capitalisation boursière) et la littérature propose d'utiliser les comportements de recherches numériques des individus comme indicateur de l'attention des marchés financiers. Ce mémoire explore la relation entre les volumes de recherches pour trois indices de marchés boursiers (Dow Jones, Nasdaq et Nyse) et sept entreprises cotées en bourse (Amazon, Google, Apple, Microsoft, Johnson& Jonhson, Berkshire Hathaway et JP Morgan& Chase) et les rendements boursiers (du 1er janvier 2006 au 31 décembre 2019) sans apporter de preuves catégoriques pour l'utilité des volumes de recherche. On trouve une relation statistiquement significative entre les volumes de recherches et les rendements boursiers absolus des indices de marché (Dow Jones, Nasdaq et Nyse) pendant la période de crise économique. On trouve également une relation statistiquement significative pour la prédiction des rendements bruts d'Apple sur toute la période à l'étude. Les relations significatives suggèrent un comportement économique cohérent avec la littérature : les chocs d'information deviennent progressivement moins importants alors qu'ils sont intégrés par le marché.

Table des matières

Résumé	ii
Table des matières	iii
Liste tableaux.....	v
Liste figures	vi
Introduction	1
Chapitre 1 - Revue de la littérature.....	3
1.1. Les supports numériques	3
1.2 Modélisation financière	5
1.3 La stabilité paramétrique.....	7
1.4 La causalité de Granger	9
1.5 Recensement de la littérature	11
Chapitre 2 – Données et méthodologie.....	12
2.1 Données de recherche et les données boursières	12
2.2 Transformation des données	14
2.3 Modèles estimés.....	15
Chapitre 3 - Résultats	18
3.1 Évaluation de la corrélation	18
3.2 Comportement de la corrélation	21
3.3 La stabilité paramétrique.....	22
3.4 La prédiction des rendements	23
Chapitre 4 - Discussion	27
4.1 Convergences et divergences avec la littérature	27
4.2 Apports originaux	28
4.3 Limites et sources d'erreur	29
Conclusion	30
Bibliographie	32
Annexe A – Démonstration du biais pour l'équation des anticipations rationnelles	34
Annexe B – Tableau résumé de la recension des études	35
Annexe C – Tableau résumé des critères d'informations (régressions sans bris structurels)	36
Annexe D – Tableau résumé des critères d'informations (régressions avec bris structurels)	37
Annexe E – Régressions sur les rendements bruts (sans bris structurels).....	38

Annexe F – Régressions sur les rendements bruts (avec bris structurels).....	39
Annexe G – régressions sur les rendements absolus (sans bris structurels).....	40
Annexe H – Régressions sur les rendements absolus (avec bris structurels)	41
Annexe I – Comparaison entre les bris sélectionnés pour les modèles boursiers et complets	42

Liste des tableaux

Tableau 1 - Périodicité des statistiques recueillies par GT.....	4
Tableau 2 – Statistiques descriptives des séries de volumes de recherche.....	15
Tableau 3 – Estimation des coefficients de la régression pour les rendements absolus avec bris structurels entre le 7 septembre 2008 et 15 mai 2011	19
Tableau 4 – Estimation des coefficients pour l’estimation des rendements bruts de la compagnie Apple.....	20
Tableau 5 - Comparaison de l'erreur quadratique moyenne pour la prédiction des rendements boursiers absolus avec et sans l’inclusion des volumes de recherche	24

Liste des figures

Figure 1 - Série d'observations hebdomadaires pour le terme « Dow Jones » entre le 1 ^{er} janvier 2014 et le 31 décembre 2019.....	5
Figure 2 - Unification des séries pour le terme de recherche « Google » entre 2006 et 2019	13
Figure 3 - Variation des recherches pour le terme "Dow Jones" entre 2006 et 2019.....	14
Figure 4 – Erreur cumulative pour la prédiction des rendements absolus du Dow Jones, Nasdaq et Nyse avec l'application d'une fenêtre mobile de 200 observations	25
Figure 5 - Erreur cumulative pour la prédiction des rendements absolus du Dow Jones, Nasdaq et Nyse avec l'application d'une fenêtre mobile de 100 observations	25

Introduction

Depuis les années 1970, la littérature économique argumente qu'il est impossible de prédire le rendement des marchés financiers. Selon l'hypothèse des marchés efficients (HME), le prix des actifs doit complètement représenter l'ensemble de l'information disponible aux investisseurs. Si les prix diffèrent de l'évaluation par les investisseurs, la demande ou l'offre pour l'actif devrait s'ajuster de sorte qu'il soit impossible de réaliser un profit. (Fama, 1970) pose trois conditions à l'établissement de marchés efficients : de faibles coûts de transaction, de l'information financière facilement accessible et un consensus sur les méthodes d'évaluation d'actifs. On peut facilement défendre la validité empirique des trois conditions. Malgré l'omniprésence de la théorie dans les écrits académiques et son attrait heuristique, sa validité demeure sujet à débat. Les détracteurs de l'HME suggèrent que plusieurs périodes économiques ont été marquées par des déviations significatives et durables de l'équilibre d'un marché efficient. Plus de la moitié des articles recensés par Sewell (2012) se prononcent contre l'HME, mais la déconstruction de l'hypothèse demeure largement anecdotique (les déviations du marché ne sont pas systématiquement expliquées) et ex post (les déviations sont identifiables seulement lorsque le marché se corrige) (Malkiel, 2005; Sewell, 2012).

La théorie de l'attention est l'une des nombreuses théories tentant de construire une critique cohérente de l'hypothèse des marchés efficients. Son postulat principal est simple : les humains ont des capacités cognitives limitées. L'information économique est complexe et son existence n'est pas une condition suffisante à l'équilibre des marchés, il faut y prêter attention et faire l'effort de la comprendre (Odean et Barber, 2008). Si une entreprise publie des états financiers désastreux mais que personne ne les analyse, on ne doit pas forcément s'attendre à une correction du marché. Pour vérifier la théorie de l'attention, il faut opérationnaliser le concept d'attention et l'intégrer à un modèle de prévision financière. Da et al. (2011) trouvent que les volumes de recherches produits par Google Trends approximent bien les outils traditionnellement utilisés pour mesurer l'attention (rendements extrêmes, volumes d'échanges, analyses des médias, etc.). Il s'en suit qu'on peut intégrer cette nouvelle source d'information aux modèles de prévision financière et étudier l'impact de l'attention sur les rendements boursiers.

Dans le premier chapitre, ce mémoire recense la littérature sur les aspects techniques de l'information numérique, les méthodes d'évaluation économétrique pour les séries financières et les publications qui ont fait usage de l'information numérique pour la prédiction des marchés boursiers. Dans le second chapitre, une méthodologie est établie afin de vérifier comment les données numériques peuvent être intégrées dans les modèles de prédiction et l'étendue de leur pertinence. Il y est défini une série de régressions pour tester les différentes propriétés de l'information numérique sur la prédiction boursière. Pour le troisième chapitre, on présente les résultats obtenus aux régressions présentées au second chapitre. Au quatrième chapitre, on analyse finalement les résultats en fonction des objectifs de la recherche.

Ce mémoire propose ainsi trois objectifs de recherches :

1. Vérifier s'il existe une relation de causalité entre les volumes de recherche au sujet des grandes entreprises américaines (en fonction de leur capitalisation boursière) et leurs rendements boursiers respectifs. La robustesse de cette relation ne nécessite pas qu'il soit possible d'intégrer les volumes de recherche dans une stratégie d'investissement.
2. Advenant qu'il existe une relation de causalité robuste, le second objectif de ce mémoire consiste à vérifier la valeur prédictive des volumes de recherches sur les rendements boursiers.
3. Étudier la relation entre les volumes de recherches, les rendements boursiers et la littérature économique.

Chapitre 1 - Revue de la littérature

Ce chapitre recense la littérature pertinente à la prédiction boursière et aux données numériques est recensée. Le chapitre est divisé en trois parties. La première partie présente *Google Trends*, l'outil de collecte des données du moteur de recherche Google. On y explique le fonctionnement de l'outil et les différentes données sur les comportements de recherche qu'il est possible de collecter. La seconde présente les méthodes économétriques en évaluation des séries temporelles, avec une attention particulière à l'analyse des séries financières. On y développe le modèle autorégressif multivarié d'Abel et Mishkin (1983) et ses hypothèses économétriques sous-jacentes. On s'attarde également à la mesure de la stabilité paramétrique des modèles. Dans la troisième partie, on recense l'ensemble de la littérature s'étant intéressée à la prédiction des marchés boursiers en faisant usage des données numériques. Les conclusions et les caractéristiques principales de ces études sont systématiquement présentées.

1.1. Les supports numériques

Google est l'un des moteurs de recherche le plus utilisés au monde. Avec des millions d'utilisateurs quotidiens, le volume des recherches effectuées sur des sujets pertinents pourrait être une source de données puissante pour les analyses financières. À l'aide de l'outil *Google Trends* (GT), gratuitement fourni par Google, il est possible d'avoir accès à cette information. Trois options de recherche sont accessibles à l'utilisateur pour procéder à l'extraction des données : le terme de recherche, la géographie et l'horizon temporel.

Le terme de recherche est le mot-clé que l'utilisateur emploie pour guider sa recherche sur le web. Ces termes sont systématiquement enregistrés par GT et il est ensuite possible d'avoir accès à ces statistiques de recherche. Il est important de noter que l'exploitation des données par GT est très littérale. Par exemple, les statistiques recueillies pour les termes de recherches « Dow Jones », « DJ » et « DowJones » ne seraient pas les mêmes, alors qu'ils réfèrent au même indice boursier.

Il est ensuite possible de filtrer les statistiques sur deux dimensions : la répartition géographique et l'horizon temporel. Les options de raffinement géographique dépendent naturellement de l'endroit du monde auquel on s'intéresse. Dans le cas des États-Unis, il est

possible de raffiner jusqu'aux zones urbaines pour tous les états américains. Il est également possible de restreindre la recherche à un horizon temporel précis. Par défaut, Google Trends offre un choix de sélection varié (voir le tableau 1 pour une liste complète). Pour les horizons temporels sélectionnés qui sont courts, la périodicité est généralement plus fine. Dans le cas des horizons personnalisés, la périodicité dépend de la longueur de l'horizon temporel sélectionné.

Tableau 1 - Périodicité des statistiques recueillies par GT

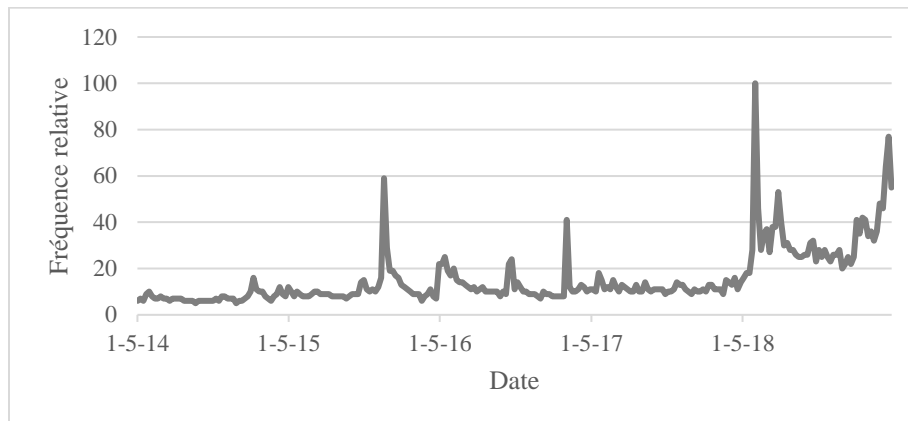
Horizon temporel	Périodicité
Dernière heure	1 observation / minute
4 dernières heures	1 observation / minute
Dernière journée	1 observation / 8 minutes
7 derniers jours	1 observation / heure
30 derniers jours	1 observation / jour
90 derniers jours	1 observation / jour
12 derniers mois	1 observation / semaine
5 dernières années	1 observation / semaine
2004 - présent	1 observation / mois
Personnalisé	Variable

La Figure 1 présente le résultat d'une exploitation de données au sujet du terme « Dow Jones » sur un horizon de 5 ans (1^{er} janvier 2014 au 1^{er} janvier 2019) aux États-Unis. Les observations ont un maximum borné à 100, suivant une formule de normalisation du genre :

$$Obs_{normalisée} = \frac{Obs_{initiale}}{Obs_{max}} \times 100$$

L'ajustement des données laisse la distribution inchangée et la valeur maximale de la série est bornée à 100. On peut donc lire la valeur d'une observation comme un pourcentage du maximum de la série.

Figure 1 - Série d'observations hebdomadaire pour le terme « Dow Jones » entre le 1^{er} janvier 2014 et le 31 décembre 2019



À titre d'exemple, la série d'observations pour la fréquence relative du terme de recherche « Dow Jones » atteint sa valeur maximale de 100 au 4 février 2018 et une valeur de 60 au 2 août 2015. On comprend donc que le volume de recherche brut dans la première semaine d'août équivaut à 60% du volume brut de recherche à la première semaine de janvier 2018.

1.2 Modélisation financière

Le concept d'anticipations économiques rationnelles est opérationnalisé dans les années 1960 par Muth (1961) et est repris par plusieurs auteurs dont Abel et Mishkin (1983). Dans le même esprit que l'hypothèse des marchés efficients, il stipule que les prix sont construits sur nos meilleures anticipations du futur. Abel et Mishkin (1983) formalisent le problème par 2 équations :

$$R_t^e = R_t - E(R_t | \theta_{t-1}) \quad (1.2.1)$$

$$E(R_t^e | \theta_{t-1}) = 0 \quad (1.2.2)$$

R_t est le rendement observé au temps t

$E(R_t | \theta_{t-1})$ est le rendement prédit pour le temps t en fonction de l'information disponible au temps $t-1$

$E(R_t^e | \theta_{t-1})$ est la déviation anticipée entre le rendement prédit et le rendement réalisé

L'équation 1.2.1 définit qu'on observe un rendement excédentaire lorsque le rendement observé est différent du rendement anticipé par le marché en fonction de l'information

disponible. La déviation du rendement anticipé R_t^e a une espérance nulle sachant que tous les acteurs ont accès à l'information disponible θ_{t-1} .

Cette formalisation d'un marché respectant l'axiome des anticipations rationnelles a l'avantage d'être facile à tester empiriquement. Il est possible de décomposer le terme d'information en variables explicatives pour une régression et vérifier que cette information n'est pas corrélée avec la déviation des rendements :

$$R_t^e = \beta X_{t-1} + \varepsilon_t \quad (1.2.3)$$

X_{t-1} est un vecteur de variables informatives.

L'équation 1.2.3 respecte la condition posée par l'équation 1.2.2 lorsque le coefficient β est égal à 0 et lorsque le terme d'erreur est nul en moyenne. La valeur estimée du coefficient est évaluée sur une distribution de Fisher et on obtient alors que l'information disponible n'apporte aucune explication pour la déviation des rendements.

(Mankiw et Shapiro, 1985) avertissent toutefois que la régression de l'équation 1.2.3 peut être biaisée lorsque la variable explicative suit un comportement fortement autorégressif. On définit alors la variable X_t comme un processus autorégressif avec un terme d'erreur v_t d'espérance nulle et de variance fixe:

$$X_t = \rho X_{t-1} + v_t \quad (1.2.4)$$

Pour l'estimation de l'équation 1.2.3, on suppose que le terme d'erreur ε_t ne peut pas être corrélé avec les régresseurs. Par ailleurs, le modèle n'exclut pas qu'il puisse exister une relation temporelle entre les termes d'erreurs ε_t et v_t . Il arrive donc, pour l'équation 1.2.2, que le terme d'erreur soit corrélé avec le régresseur pour différents niveaux de délais, violant l'hypothèse de Gauss-Markov.

(Lewellen, 2004) évalue l'importance du biais par l'équation :

$$\hat{\beta} - \beta = \gamma(\hat{\rho} - \rho) + \eta \quad (1.2.5)$$

Où $\hat{\beta} - \beta$ est la différence entre l'estimation et la valeur réelle du vecteur de coefficients pour l'équation 1.2.2, γ est le niveau de corrélation entre les termes d'erreurs pour l'équation 1.2.2 et 1.2.3, et $\hat{\rho} - \rho$ est la différence entre l'estimation du vecteur de coefficients autorégressifs

de l'équation 1.2.3. Le développement de l'équation du biais est présenté formellement à l'annexe A.

Le biais qui survient dans l'équation 1.2.5 tient donc à l'importance de la corrélation entre les termes d'erreurs des équations 1.2.3 et 1.2.4 et un biais d'estimation du coefficient autorégressif dans l'équation 1.2.4. A priori, il est impossible de déterminer l'importance du paramètre γ . Par ailleurs, le biais dans l'estimation du coefficient autorégressif s'approxime (Lewellen, 2004):

$$\hat{\rho} - \rho = \frac{-(1+3\rho)}{T} \quad (1.2.6)$$

Pour un échantillon suffisamment nombreux, le biais tend vers 0 et le biais sur l'estimateur $\hat{\beta}$ tend également vers 0 indépendamment de la relation entre les termes d'erreur.

1.3 La stabilité paramétrique

Les séries temporelles et plus particulièrement les séries financières peuvent être sujettes à des périodes d'instabilité où les relations économétriques normales ne sont plus appropriées. À titre d'exemple, la crise financière de 2008 a pris les marchés financiers par surprise et contraint les autorités à revoir leurs modèles prévisionnels. Dans un climat de panique et d'incertitude, il y a un changement fondamental dans le comportement d'investissement des acteurs financiers (Mishkin, 2013). Plus récemment, on peut également penser à la guerre commerciale entre la Chine et les États-Unis et le divorce entre l'Union européenne et le Royaume-Uni comme des périodes d'instabilité qui pourraient significativement altérer la dynamique des marchés.

Économétriquement, ce problème peut être traité de plusieurs façons. La réponse la plus simple consiste à placer une variable indicatrice pour isoler une période de la série. Toujours dans le cas de la crise financière, on pourrait indexer les données de 2008 pour ajuster l'estimation des coefficients à la crise. Cette approche comporte toutefois deux problèmes très importants. D'abord, il faut connaître l'existence des points de rupture de la série temporelle. Cette connaissance requiert une expertise pointue du domaine d'étude, sans quoi le chercheur peut isoler des périodes qui ne changent pas forcément le modèle. Ensuite, même si le chercheur évalue correctement un événement pertinent, il peut être difficile d'en borner

les effets. Toujours pour l'analyse de la crise financière de 2008, il pourrait être embêtant et subjectif de définir à quel moment la crise débute et à quel moment elle prend fin (Bai et Perron, 1998). (Andrews, 1993; Bai et Perron, 1998) proposent des méthodes d'estimation plus flexibles pour l'évaluation des bris structurels.

(Andrews, 1993) définit mathématiquement l'hypothèse nulle pour vérifier le problème de stabilité structurel dans un modèle :

$$H_0 : \beta_t = \beta_0 \text{ pour toutes les valeurs de } t \quad (1.3.1)$$

L'équation 1.3.1 stipule qu'un modèle ne contient pas de bris structurel si l'estimation des coefficients reste inchangée pour toutes les partitions possibles de la série temporelle. On peut ensuite définir l'hypothèse alternative :

$$H_1 : \beta_{1,\dots,\pi} \neq \beta_{\pi,\dots,T} \quad (1.3.2)$$

À l'équation 1.3.2, la variable π identifie un moment de la série temporelle qui contient T observations. Dans le cas d'un bris structurel, il existe une valeur de π pour laquelle l'estimation du coefficient pour un échantillon avec toutes les données précédents π est différente du coefficient estimé par l'échantillon pour toutes les données suivant π . Ce cas est toutefois limité à un scénario où il n'y aurait qu'un bris dont on connaît le moment. Ce test est plus communément connu comme étant le test de Chow.

Andrews (1993) propose donc de tester l'hypothèse alternative pour un ensemble de valeur possible de π . Ce sous-ensemble de l'échantillon est borné entre 1 et T . La statistique de Wald peut ensuite être calculée pour toutes les valeurs de π contenues dans le sous-ensemble. La valeur supérieure désigne le point où le bris structurel est le plus prononcé. L'évaluation du bris structurel par l'évaluation de la limite statistique supérieure permet d'établir le point de bris structurel avec plus de flexibilité qu'une décision arbitraire comme dans le cas de l'indicatrice. Cependant, il faut encore définir un intervalle crédible pour la position de chacun des bris à tester.

L'approche de Bai et Perron (1998) ne requiert pas une connaissance préalable du problème pour l'identification du bris structurel. Ils proposent plutôt un problème de minimisation des

MCO pour les partitions possibles de la série temporelle. On définit le modèle à l'équation 1.3.3 :

$$y_t = \beta_j X_t + \varepsilon_t \quad (1.3.3)$$

L'équation est similaire à celle obtenue pour une régression linéaire normale, à l'exception que les coefficients sont indexés pour les différentes partitions définies préalablement. Par exemple, si l'on stipule qu'il existe deux bris structurels dans la série temporelle (T_1, T_2), il devrait y avoir 3 partitions. Le problème devient donc de situer le moment des bris structurels de sorte que l'on minimise la somme des carrés des résidus pour les trois partitions qui découleront des dates de bris :

$$(T_1, \dots, T_i) = \operatorname{argmin}_{(t_1 \dots t_i)} \sum_1^j \text{SSR}|j \quad (1.3.4)$$

Contrairement à Andrews (1993), le problème de minimisation de (Bai et Perron, 1998) est complètement impartial. Un auteur pourrait donc trouver un point de bris qui minimise la somme des carrés des résidus grâce à l'identification d'un bris structurel dont il n'avait pas envisagé l'existence.

Par ailleurs, bien que le processus de segmentation de la série temporelle ne le garantisse pas, il est possible et même probable qu'il corrige certains problèmes d'hétéroscédasticité conditionnelle si la variance de l'erreur change simultanément avec le coefficient.

1.4 La causalité à la Granger

Nous avons déjà défini précédemment comment la littérature construit des séries temporelles pour les analyses financières. Le concept peut facilement être généralisé pour toutes les séries temporelles :

$$Y_t = \beta U_{t-1} + \varepsilon_t \quad (1.4.1)$$

On tente d'estimer la variable Y_t avec l'ensemble de l'information disponible U_{t-1} . On peut ensuite mesurer le succès de l'estimation en évaluant la variance de la série de résidus. Une variance élevée (faible) implique que les régresseurs n'expliquent pas très bien (expliquent bien) la variable d'intérêt. (Granger, 1969) argumente en faveur d'une définition mathématique formelle du principe de causalité pour une série temporelle :

$$\sigma^2(Y_t|U_{t-1}) < \sigma^2(Y_t|U_{t-1} - X_{t-1}) \quad (1.4.2)$$

L'équation 1.4.2 admet une relation causale de $X \rightarrow Y$ lorsque la variance des résidus pour une régression incluant toute l'information ($Y_t|U_{t-1}$) est plus faible que dans un modèle où l'on retire l'information X ($Y_t|U_{t-1} - X_{t-1}$). Lorsque la variable X est effectivement profitable à l'estimation, son inclusion dans le modèle économétrique devrait faire diminuer l'erreur de prédiction et par définition la variance.

Si l'inclusion de variables supplémentaires ne peut être que profitable à la précision pour l'échantillon de la régression, elle n'implique pas forcément une meilleure qualité prédictive en dehors de l'échantillon. En ajoutant plusieurs variables au modèle, les paramètres seront spécialement estimés pour bien caractériser les données de l'échantillon, mais perdront leur capacité à généraliser et transférer leur sens à de nouvelles données. La sélection des variables à inclure au modèle doit donc pénaliser l'inclusion excessive de variables explicatives (Greene, 2012). (Akaike, 1974) introduit un critère de sélection qui génère un arbitrage entre la vraisemblance d'un modèle et le nombre de paramètres :

$$AIC(K) = \ln\left(\frac{\epsilon'\epsilon}{N}\right) + 2\left(\frac{K}{N}\right) \quad (1.4.3)$$

Le critère d'information d'Akaike est utile à titre comparatif. En comparant deux modèles, on remarque que le modèle incluant plus de variables sera pénalisé (le critère d'information augmente) par le second terme et donc défavorisé à moins que l'augmentation de la valeur explicative engendre une diminution plus forte du critère. Le modèle sélectionné est celui qui minimise le critère. Par ailleurs, la valeur absolue du critère d'information ne transmet aucune information sur la valeur statistique du modèle.

Le critère d'information de Bayes (BIC) est une autre métrique fréquemment employée dans la littérature pour sélectionner un modèle. Semblable au critère d'Akaike, son équation s'écrit :

$$BIC(K) = \ln\left(\frac{\epsilon'\epsilon}{N}\right) + \left(\frac{K \ln(N)}{N}\right) \quad (1.4.4)$$

Le critère de Bayes pénalise davantage les modèles complexes en dépit des échantillons plus importants (Greene, 2012).

1.5 Recensement de la littérature

Dans un article paru en 2011, Da et Al. (2011) utilisent les données de recherches de Google pour expliquer les rendements des actions de l'indice Russel 3000. Leur hypothèse est qu'un acteur financier prend généralement le temps de s'informer avant de prendre une décision d'investissement et qu'il devrait donc exister une relation de causalité entre les volumes de recherche et les rendements boursiers. Ils évaluent la relation de causalité dans un modèle de type « vecteur autorégressif » (VAR) avec 5 variables : le volume de recherche, le volume d'actions transigées, un indice englobant la couverture de plusieurs agences d'informations traditionnelles et une mesure de rendement anormal. L'article de Da et al. (2011) est particulièrement significatif, car il donne le ton aux autres recherches qui reprennent essentiellement le même cadre d'analyse.

Ces études se distinguent généralement les unes des autres selon quatre dimensions : l'objectif de performance mesuré, le type de terme de recherche analysé, les dimensions temporelles (la fréquence des recherches et la période d'analyse) et les indices boursiers sélectionnés. Les résultats des principales études sont résumés dans un tableau à l'annexe B.

Les objectifs de performance mesurés sont classifiés dans deux catégories. La première catégorie définit une étude dont l'objectif explicitement décrit est d'intégrer les données de recherche dans une stratégie d'investissement. La seconde catégorie regroupe les études qui définissent une relation de causalité entre les termes de recherche et une métrique financière de rendement. L'objectif n'est pas forcément de définir une stratégie de placement rentable.

Les termes de recherche sont également classifiés dans deux catégories. La première regroupe tous les termes spécifiquement reliés aux indices boursiers analysés. Par exemple, les résultats pour le terme « Dow Jones » ou le libellé d'une entreprise seraient considérés comme étant spécifiques. La seconde catégorie inclut toutes les études qui font usage de termes financiers plus généraux. Par exemple, le terme de recherche « Inflation » serait considéré comme un terme général.

Le recensement littéraire présente un portrait nuancé de l'utilité des volumes de recherche pour la prédiction boursière. L'absence de consensus peut s'expliquer par la multiplicité des déterminants méthodologiques.

Chapitre 2 – Données et méthodologie

Dans ce chapitre, on présente les données de recherche et la méthodologie mise en place pour valider les objectifs de recherche. On y présente d'abord les données brutes collectées, leur provenance (les termes de recherches et les indices boursiers) et le traitement qui s'en suit pour les adapter à l'analyse quantitative du travail. Cette analyse est détaillée par la suite.

2.1 Données de recherche et les données boursières

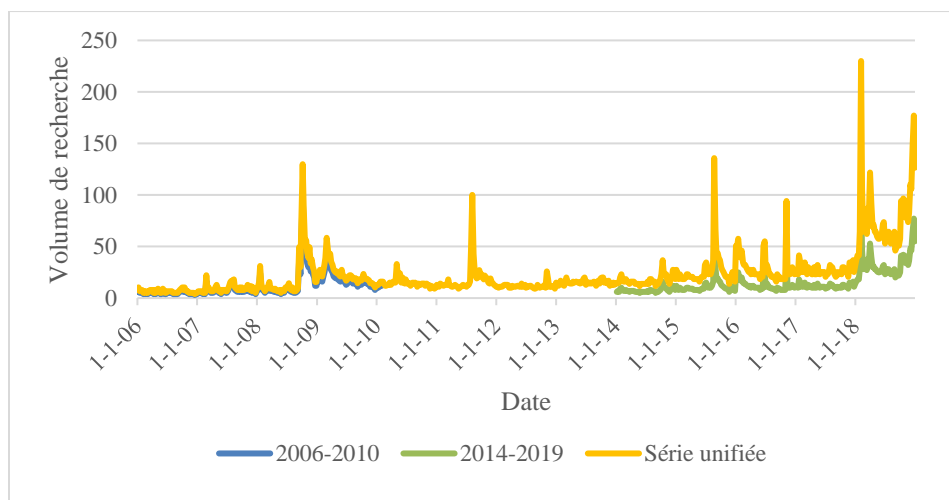
Pour les fins de ce travail, 2 types de termes de recherches ont été sélectionnés : le libellé des compagnies inscrites en bourse et les indices boursiers populaires. Pour les indices boursiers, on sélectionne le Dow Jones, le Nasdaq et le NYSE. Ces trois indices ont l'avantage de couvrir une grande diversité d'entreprises tout en jouissant d'une popularité importante auprès des acteurs financiers.

Pour le libellé des entreprises américaines, 7 compagnies ont été arbitrairement sélectionnées en fonction de leur popularité et de leur capitalisation boursière :

1. Amazon : Il s'agit premièrement d'une compagnie de commerce en ligne, mais avec des activités en croissances dans divers secteurs.
2. Alphabet : Alphabet est la compagnie mère du moteur de recherche le plus utilisé au monde. Ses activités sont regroupées en bourse sous le libellé «Goog».
3. Apple : Les activités d'Apple se concentrent principalement autour des technologies de communication.
4. Microsoft : Microsoft est une entreprise du secteur des technologies qui produit principalement des systèmes d'exploitation pour ordinateurs.
5. Johnson & Johnson : Johnson et Johnson est une compagnie spécialisée dans la vente de matériel médical et de biens pharmaceutiques grand public.
6. Berkshire Hathaway : Attachée à la notoriété de Warren Buffet, Berkshire Hathaway est un conglomérat d'investissement qui est propriétaire de plusieurs entreprises américaines.
7. JPMorgan & Chase : JPMorgan & Chase est une banque d'investissement américaine et spécialisée dans la vente de services financiers.

Le volume des recherches pour ces termes est recueilli sur un horizon temporel de 13 ans (janvier 2006 à décembre 2019) avec une périodicité hebdomadaire. Comme les données hebdomadaires ne sont disponibles que sur un horizon de 5 ans, il a fallu décomposer l'horizon de 2006 à 2019 en trois séries individuelles (1^{er} janvier 2006 au 1^{er} février 2010; 1^{er} janvier 2010 au 1^{er} février 2014; 1^{er} janvier 2014 au 31 décembre 2019). Le dernier mois des deux premières séries et le premier mois des deux dernières séries servent à ajuster la série temporelle sur une échelle commune et centrée autour des données exploitées entre 2010 et 2014. À la figure 2, on peut observer les 3 séries temporelles individuelles pour le terme « Dow Jones » ainsi que le résultat de leur unification.

Figure 2 - Unification des séries pour le terme de recherche « Google » entre 2006 et 2019



On remarque que la nouvelle série unifiée ne varie plus entre 0 et 100. Au moment de réunifier les séries temporelles, les arrondissements de Google Trends introduisent de la variance. Comme on n'ajuste que deux séries, cette variance ne devrait pas dénaturer les données brutes.

Les données boursières pour chaque terme de recherche sont recueillies depuis le moteur de recherche Yahoo!. Les données brutes ont une périodicité quotidienne et incluent les prix à l'ouverture, les prix à la fermeture et le volume d'actions transigées.

2.2 Transformation des données

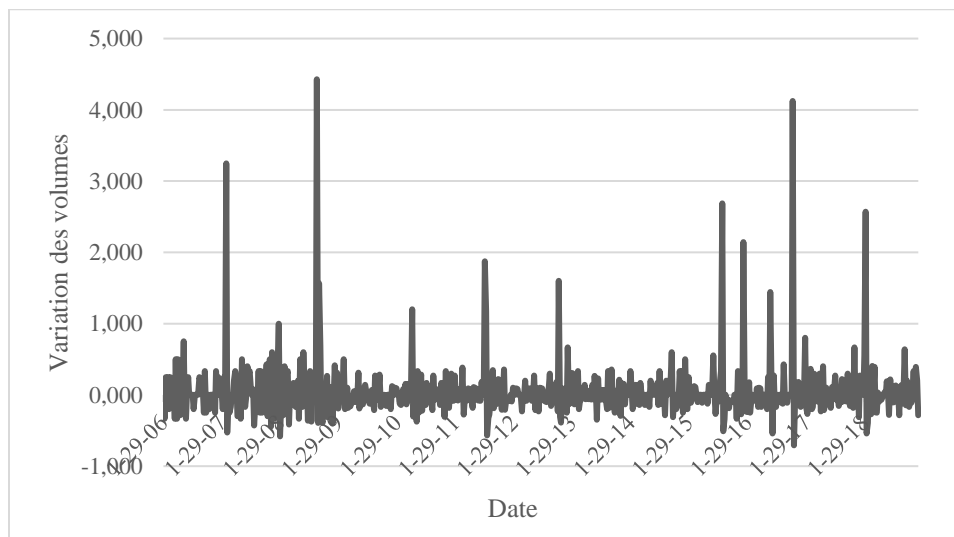
Pour procéder à l'analyse quantitative de la relation entre les termes de recherche et les rendements boursiers, il faut d'abord retravailler les données pour corriger la stationnarité des séries de Google Trends et faire ressortir certains comportements. Une fois les corrections apportées, toutes les séries sont stationnaires.

Pour stationnariser les données de recherches, nous appliquons un calcul de la variation :

$$\text{VarGt} = \frac{\text{GT}_t - \text{GT}_{t-1}}{\text{GT}_{t-1}} \quad (2.2.1)$$

La transformation permet donc de centrer la série temporelle autour de sa croissance moyenne. À titre d'exemple, la Figure 3 présente la série transformée pour la recherche du terme « Dow Jones ». Les futures mentions aux séries de recherche font référence aux séries transformées.

Figure 3 - Variation des recherches pour le terme "Dow Jones" entre 2006 et 2019



En comparant la Figure 2 et la Figure 3, on remarque que la transformation fait ressortir l'importance des changements lorsque les volumes de recherche sont faibles. Il n'est pas clair a priori si cette interaction est favorable à l'établissement de la causalité. On présente également au tableau 2 les statistiques descriptives pour toutes les séries.

Tableau 2 – Statistiques descriptives des séries de volumes de recherche

	Observations	Moyenne	Variance	80e percentile	90e percentile
Indices de marché					
Dow Jones	678	0,047	0,158	0,167	0,319
Nasdaq	678	0,015	0,035	0,103	0,189
Nyse	678	0,026	0,091	0,120	0,207
Indices					
Amazon	678	0,005	0,005	0,038	0,055
Google	678	0,003	0,004	0,032	0,060
Apple	678	0,013	0,030	0,051	0,148
Microsoft	678	0,000	0,004	0,033	0,067
J&Johnson	678	0,011	0,047	0,092	0,163
Berk	678	0,046	0,142	0,169	0,313
JP Morgan	678	0,017	0,040	0,142	0,241

Les données boursières sont également retravaillées pour faire ressortir la variation hebdomadaire des rendements (en gain ou perte de capital). Les données quotidiennes sont transformées en rendement hebdomadaire suivant la formule :

$$\text{Rendements}_{\text{hebdo}} = \frac{\text{Prix}_{\text{fermeture}} - \text{Prix}_{\text{ouverture}}}{\text{Prix}_{\text{ouverture}}} \quad (2.2.2)$$

Les rendements hebdomadaires sont donc bornés à l'ouverture de la semaine, généralement les lundis (mais peut être repoussés dans le cas des jours fériés) et à la fermeture de la semaine. Cette définition assure qu'il n'y aura pas d'enchevêtrement avec les données de recherche lorsque les délais hebdomadaires seront appliqués.

2.3 Modèles estimés

La relation prédictive entre les données de recherches et les rendements boursiers est évaluée dans un modèle d'anticipations rationnelles pour les 7 entreprises et les 3 indices en fonction de leur terme de recherche respectif et pour deux périodes de délais hebdomadaires :

$$R_t = C + \sum_{i=1}^p \rho_i R_{t-i} + \sum_{i=1}^a \alpha_i GT_{t-i} + \sum_{i=1}^b b_i GT_{t-i} * I_{gt\{0,9\}} + I_{gt\{0,9\}} + \varepsilon_t$$

$$|R_t| = C + \sum_{i=1}^p \rho_i R_{t-i} + \sum_{i=1}^a \alpha_i GT_{t-i} + \sum_{i=1}^b b_i GT_{t-i} * I_{gt\{0,9\}} + I_{gt\{0,9\}} + \varepsilon_t$$

R_t , le rendement observé d'un indice boursier

$|R_t|$, le rendement absolu observé d'un indice boursier

C , une constante au modèle

R_{t-i} , le rendement observé d'un indice boursier i périodes précédant le rendement prédit

GT_{t-i} , le volume de recherches pour l'indice boursier i périodes précédant le rendement prédit

$I_{gt\{0,9\}}$, une variable binaire qui prend la valeur 1 lorsque le volume de recherche est supérieur au 90^e centile. La variable prend la valeur 0 pour les volumes de recherche inférieurs au 90^e centile.

Le terme de gauche de l'équation, R_t , représente le rendement d'un indice boursier à tout moment. Les rendements sont prédits en fonction des valeurs retardées des rendements R_{t-i} et des données de recherches GT_{t-i} . On note également que les volumes de recherche sont indexés selon leur répartition par centile afin de mieux distinguer l'effet de la variance normale dans le comportement de recherche des démarcations plus significatives. La seule différence entre les deux régressions est la transformation en valeur absolue des rendements boursiers pour la seconde équation. La première régression suppose un effet unidirectionnel du volume de recherche, c'est-à-dire que la hausse des recherches entraîne systématiquement une hausse ou une baisse du rendement boursier. La seconde régression est moins restrictive, car elle permet un effet bidirectionnel. Avec une connaissance sur la nature de l'information, il est possible qu'un individu puisse déterminer la direction d'un choc économique. Parce que nous ne sommes pas capables de vérifier l'interprétation des nouvelles par les investisseurs, ce modèle n'assure pas qu'il existe une relation de causalité entre les variations du volume de recherches et les rendements boursiers. Néanmoins, l'hypothèse n'est pas complètement déraisonnable. Ces deux modèles seront évalués pour les 7 indices boursiers et les 3 indices de marché.

Ces deux modèles contribuent également au premier objectif de ce mémoire : évaluer la pertinence d'inclure les données de recherche dans un modèle de prédiction boursière. Si les données de recherche n'apportent pas de plus-value à ce qui serait autrement un simple modèle autorégressif, on peut alors douter de leur pertinence. Pour ce faire, nous évaluons la causalité de Granger entre les modèles présentés par les équations précédentes et un simple modèle de rendements autorégressifs avec le critère d'information d'Akaike:

$$R_t = C + \sum_{i=1}^p \rho_i R_{t-i} + \varepsilon_t$$

$$|R_t| = C + \sum_{i=1}^p \rho_i R_{t-i} + \varepsilon_t$$

Suivant la stratégie développée par Bai et Perron (1998) et présentée au premier chapitre, l'ensemble des régressions précédentes sont réévaluées pour estimer deux bris structurels par série. Ces deux bris impliquent que les paramètres estimés aux régressions initiales seront triplés pour les trois segments résultants des bris. Certaines limites sont mises en place pour assurer des résultats cohérents pour toutes les séries. D'abord, on fixe un niveau minimum pour l'étendue des segments : chaque segment doit contenir au moins 20% des données de la série (135 observations). Cette mesure assure qu'il y aura suffisamment de données pour convenablement estimer les paramètres à chaque segment. Par contre, cette mesure réduit également la flexibilité de l'algorithme à trouver les points de segmentation optimaux. Finalement, on fixe le nombre de bris structurel à 2, indépendamment du critère d'information. Il est possible que certaines séries ne nécessitent qu'un bris et que l'ajout de paramètres ne soit pas compensé par une hausse marquée de la précision.

Chapitre 3 - Résultats

Dans ce chapitre, nous présentons les résultats des régressions décrites au chapitre précédent. Ces résultats sont ensuite expliqués afin de répondre aux objectifs de recherche de ce mémoire et pour identifier les faits saillants des modèles estimés.

À titre de rappel, les trois objectifs de recherches sont :

1. Vérifier s'il existe une relation de causalité entre les volumes de recherche pour les grandes entreprises américaines (en fonction de leur capitalisation boursière) et leurs rendements boursiers respectifs.
2. Étudier le processus d'intégration de l'information numérique par les marchés financiers.
3. Admettant l'existence d'une relation de causalité, vérifier s'il est possible d'intégrer les données de recherche aux modèles de prédictions boursières.

S'en suivra finalement une discussion sur la portée des résultats. Plus particulièrement, on commente sur les possibles sources d'erreurs du travail, les apports originaux et les sujets à développer pouvant faire l'objet de recherches futures.

3.1 Évaluation de la corrélation

Le premier objectif de recherche de ce mémoire est d'évaluer la relation de causalité entre les données de recherche et les variations du marché boursier. Pour ce faire, on compare un modèle autorégressif simple (qui ne prend que les données boursières) avec un modèle complet qui intègre les données de recherche. Pour l'estimation des rendements bruts et des rendements absolus, avec ou sans bris structurels, on trouve que le modèle complet est préféré au modèle boursier autorégressif pour seulement 5 régressions. Les résultats du critère d'information d'Akaike sont présentés à l'annexe C et D. Les cinq régressions acceptées par le critère d'information sont :

1. La régression sur les rendements bruts de Apple – Sans bris structurels
2. La régression sur les rendements bruts du Nyse – Avec bris structurel

3. La régression sur les rendements absolus de Berkshire Hathaway – Sans bris structurel
4. La régression sur les rendements absolus du Dow Jones – Avec bris structurel
5. La régression sur les rendements absolus du Nyse – Avec bris structurel

À la vue de ces résultats variés, il est difficile de donner de la crédibilité à l'hypothèse que les volumes de recherches pour les libellés boursiers et les indices de marché sont systématiquement corrélés avec leurs rendements respectifs.

Même si la majorité des régressions pour les deux modèles est rejetée par le critère d'information, certaines trouvent un effet temporaire, mais robuste entre les rendements et les variations du volume de recherches. Les estimations pour les rendements absolus des indices de marchés (Dow Jones, Nasdaq et Nyse) avec un bris structurel sont particulièrement intéressantes (deux d'entre elles sont d'ailleurs acceptées par le critère d'information).

On remarque une relation robuste de plusieurs coefficients pour le second segment de l'estimation (qui couvre la période du 7 septembre 2008 au 15 mai 2011). Ces résultats sont tirés du tableau présenté à l'annexe H et condensés au Tableau 3.

Tableau 3 – Estimation des coefficients de la régression pour les rendements absolus avec bris structurels entre le 7 septembre 2008 et 15 mai 2011

Coefficients	Dow Jones	Nasdaq	Nyse
Constante (C)	0,0079 (0,0025)	0,0157 (0,0033)	0,0122 (0,0033)
R_{t-1}	0,1544 (0,0640)	0,1602 (0,0604)	0,3062 (0,0590)
GT_{t-1}	0,0165 (0,0144)	0,0075 (0,0248)	0,0224 (0,0196)
$I_{gt\{0,9\},(t-1)}$	-0,0004 (0,0088)	-0,0523 (0,0185)	-0,0918 (0,0145)
$I_{gt\{0,9\},(t-1)} * GT_{t-1}$	0,0470 (0,0184)	0,2227 (0,0589)	0,2609 (0,0415)
R_{t-2}	0,1726 (0,0566)	0,2144 (0, 0588)	0,2344 (0,0523)
GT_{t-2}	0,0228 (0,0141)	-0,0202 (0,0248)	-0,0242 (0,0197)
$I_{gt\{0,9\},(t-2)}$	0,0208 (0,0083)	-0,0275 (0,0110)	0,0374 (0,0147)
$I_{gt\{0,9\},(t-2)} * GT_{t-2}$	-0,0534 (0,0176)	0,0795 (0,0349)	-0,1078 (0,0427)

Les coefficients significatifs au seuil de 5% sont présentés en caractère gras. Les écarts-types sont présentés entre parenthèses.

Conformément à l'intuition, les volumes de recherches anormalement élevés (les volumes supérieurs au 90^e centile de l'échantillon) sont pratiquement tous statistiquement significatifs au seuil de 5% (exception pour l'indicatrice du Dow Jones à la première semaine de délai).

Une semaine suivant l'observation du rendement, les variations importantes du volume de recherche engendrent un effet marginal positif sur la variation du rendement des trois indices boursiers. Ce résultat implique que les nouvelles avec un important profil public affectent la capitalisation boursière d'une entreprise. Inversement, les variations plus faibles (inférieures au 90^e centile) ne sont significatives que pour la régression sur le Nyse deux semaines suivant l'observation initiale.

Pour le modèle d'estimation des rendements bruts, les volumes de recherche sont statistiquement significatifs uniquement pour la compagnie Apple lorsqu'on ne considère pas de bris structurels. Les résultats sont présentés à l'annexe E et condensés au tableau 4 pour la compagnie Apple

Tableau 4 – Estimation des coefficients pour l'estimation des rendements bruts de la compagnie Apple

Coefficients	Apple
Constante (C)	0,0066 (0,0020)
R_{t-1}	-0,0221 (0,0385)
GT_{t-1}	-0,0101 (0,0297)
$I_{gt-1\{0,9\}}$	-0,0177 (0,0096)
$I_{gt-1\{0,9\}} * GT_{t-1}$	0,0364 (0,0354)
R_{t-2}	0,0083 (0,0384)
GT_{t-2}	0,0664 (0,0216)
$I_{gt-2\{0,9\}}$	0,0009 (0,0099)
$I_{gt-2\{0,9\}} * GT_{t-2}$	-0,0965 (0,0312)

Les coefficients significatifs au seuil de 5% sont présentés en caractère gras. Les écarts-types sont présentés entre parenthèses.

Les rendements absolus sont positivement corrélés avec la variation des volumes de recherche à la seconde semaine de délai. Par contre, à partir du 90^e centile, l'effet marginal des recherches est renversé et devient négatif. Cette relation ne tient pas pour la régression sur les rendements absolus et témoigne d'une relation unilatérale de l'effet des volumes de recherche sur les rendements.

3.2 Comportement de la corrélation

Le second objectif de ce mémoire est d'étudier les mécanismes d'intégration de l'information par les marchés. Pour ce faire, nous interprétons l'impact des délais hebdomadaires sur l'estimation des rendements pour les régressions où les données de recherches sont statistiquement significatives; ce sont les quatre régressions qui ont fait l'objet d'une analyse au point précédent. Plus particulièrement, on s'intéresse à deux comportements :

1. Les volumes de recherches exposent un comportement autorégressif. C'est-à-dire qu'on vérifie si un choc de recherche a un impact plus important sur les rendements lorsqu'il survient une semaine avant l'observation plutôt qu'à la seconde.
2. L'effet marginal des volumes de recherche est positivement corrélé pour les délais hebdomadaires. Si les volumes de recherches ont un impact positif sur les rendements une semaine avant l'observation du rendement, on vérifie que cette relation n'est pas inversée pour la seconde semaine précédente.

Pour les régressions sur les rendements absolus, les volumes de recherches élevés sont statistiquement significatifs uniquement pour les indices Dow Jones, Nasdaq et Nyse avec l'inclusion des bris structurels. Les résultats de ces estimations sont déjà présentés au Tableau 3.

On peut vérifier le comportement régressif des variations du volume en évaluant l'apport des volumes de recherches après une et deux semaines de délais. Pour les trois indices de marché, l'effet marginal des volumes de recherche est plus faible à la seconde semaine; le coefficient estimé est systématiquement moins important à la seconde semaine de délai. Une hausse dans le volume de recherche a donc un impact plus important sur les rendements la semaine suivante qu'elle n'en a sur la seconde semaine suivante.

Par ailleurs, cette différence entre les deux semaines de délais n'est pas apparente lorsque les volumes de recherches avoisinent le 90^e centile, car les constantes du modèle camouflent l'effet marginal des volumes recherches. Ce n'est qu'avec l'augmentation de la variation du volume de recherche que l'effet régressif prend suffisamment d'ampleur pour créer une disparité entre le premier et le second délai.

Le second comportement auquel on s'intéresse est vérifié en analysant le signe de la pente pour les deux délais hebdomadaires de nos régressions. Pour les régressions du Dow Jones et du Nyse, la variation des volumes de recherches pour la semaine précédant l'observation a un effet marginal positif sur les rendements absolus (les pentes à la figure 4a sont positives). Par contre, au second délai, l'effet est plutôt négatif pour le Dow Jones et le Nyse. Ce renversement implique qu'un choc de variation des volumes de recherche induit d'abord de la volatilité dans le marché (puisque l'on observe les rendements absolus, la variation du rendement peut être négative ou positive en fonction de la nature du volume de recherche), pour ensuite inciter à la prudence. Ce renversement de causalité est également noté par (Da et al., 2011; Joseph et al., 2011), mais disputé par (Bijl et al., 2016). La régression du Nasdaq ne présente pas ce comportement et l'effet marginal à la première et seconde semaine varie positivement.

Pour les régressions sur les rendements bruts, seulement la compagnie Apple présente une estimation où le volume de recherche est une variable statistiquement significative. Comme précisé précédemment, seulement les volumes de recherches élevés (supérieurs au 90^e centile) deux semaines précédant l'observation du rendement ont un impact sur les rendements. Ce comportement est particulier, car il va à l'encontre de l'hypothèse de régression des volumes de recherches. À l'instar des indices de marché analysés précédemment, on s'attendrait à ce que le premier délai hebdomadaire apporte un effet net plus important sur la variation des rendements. Il serait raisonnable de croire que l'effet du premier délai hebdomadaire sur les rendements bruts est masqué par une relation bidirectionnelle, mais cet effet n'est pas trouvé dans la régression sur les rendements absolus de la compagnie.

3.3 La stabilité paramétrique

Ce mémoire est le premier à intégrer une méthode impartiale pour l'identification des bris structurels. Ces estimations apportent quelques résultats originaux. Les résultats seront discutés plus amplement dans la discussion du prochain chapitre, car ils apportent beaucoup de nouveaux éléments à la littérature existante.

D'abord, l'estimation des dates de bris structurels est particulièrement robuste pour le premier bris des régressions sur les rendements bruts et absolus. Dans le premier cas, à

l'exception notable de la régression pour Johnson & Johnson, le bris se situe entre octobre 2008 et février 2009. Le second bris se situe dans un intervalle relativement large et les fondements économiques du bris sont moins évidents. On observe un résultat semblable pour le modèle d'estimation des rendements absolus. Dans un cas comme dans l'autre, on note que la régression des indices de marché est très robuste à la fois au premier et au second bris. La variance plus élevée dans l'estimation du second bris implique qu'il n'était peut-être pas favorable d'inclure deux bris structurels dans la régression, particulièrement pour l'estimation des rendements d'entreprise. Ce point sera soulevé dans la discussion ultérieure.

En partitionnant les données en trois segments qui minimisent respectivement l'erreur quadratique, on peut identifier les périodes qui poussent au rejet du modèle incluant les données de recherche. Pour toutes les régressions, rendements bruts et absolus confondus, les volumes de recherche au troisième segment ne sont pas statistiquement significatifs. Il apparaît donc que les données les plus récentes ont pu nuire à l'acceptation d'un modèle incluant les volumes de recherche. À part la relation robuste identifiée pour les indices de marché, on note occasionnellement une relation entre les volumes de recherche et les rendements au premier et au second segment.

Finalement, il est possible d'attribuer de l'importance aux variables de volume de recherche dans la sélection du bris structurel optimal. Les bris sélectionnés pour les modèles boursiers et les modèles complets sont présentés à l'annexe I. On note immédiatement que la date du bris peut varier sensiblement lorsqu'on inclut les variables de recherche.

3.4 La prédiction des rendements

Considérant la relation statistiquement significative documentée ci-dessus dans l'estimation des rendements absolus pour les régressions avec bris structurels du Dow Jones, du Nasdaq et du Nyse, on vérifie maintenant s'il est possible d'appliquer le modèle dans le cadre d'une prévision financière.

Pour ce faire, les modèles de prévision des rendements sont testés sur des données différentes de celles ayant servi à l'estimation du modèle. Nous appliquons donc une fenêtre mobile de 100 observations sur notre base de données pour l'estimation des régressions. Pour une série de 100 observations, nous estimons les paramètres du modèle autorégressif spécifié au

chapitre précédent pour ensuite tester le modèle sur l'observation suivante. On répète le processus en déplaçant la fenêtre de 100 observations jusqu'à l'évaluation de l'ensemble des données. Puisque les 100 premières observations sont utilisées pour l'estimation initiale, on teste le modèle sur les données entre le 1^{er} janvier 2008 et le 30 décembre 2018. On répète finalement le processus avec une fenêtre mobile de 200 observations.

Les résultats montrent que l'inclusion des données de recherches n'améliore pas l'erreur quadratique moyenne. Tant pour une fenêtre mobile de 100 ou 200 observations, l'erreur quadratique moyenne du modèle incluant les données de recherche est plus élevée que pour le modèle où elles sont exclues (tableau 5). Ce résultat pourrait être attribué à deux hypothèses.

Tableau 5 - Comparaison de l'erreur quadratique moyenne pour la prédiction des rendements boursiers absolus avec et sans l'inclusion des volumes de recherche

Fenêtre mobile	Erreur	Dow Jones	Nasdaq	Nyse
100	Avec Google	2,93E-04	4,05E-03	4,08E-04
	Sans Google	2,69E-04	3,55E-04	3,40E-04
200	Avec Google	1,66E-04	2,58E-04	2,29E-04
	Sans Google	1,55E-04	2,56E-04	2,04E-04

Premièrement, la méthodologie avec une fenêtre mobile est plus restrictive, car on ne peut plus évaluer les coefficients de la régression en incluant des données qui ne sont pas encore réalisées et en segmentant l'échantillon pour présenter les périodes pertinentes. Pour que l'estimation par la fenêtre mobile apporte un résultat similaire, il faudrait donc que la relation soit statistiquement significative et stable sur une longue période de temps. Cependant, avec l'inclusion des bris structurels, nous trouvons une seule période significative entre septembre 2008 et mai 2011.

Deuxièmement, il est possible qu'on rencontre un problème de surajustement des paramètres du modèle. Parce que le modèle évalue plusieurs coefficients comparativement à la largeur de la fenêtre, on estime des valeurs aux coefficients qui n'ont pas de sens économique en dehors de la fenêtre où ils sont évalués. Cette hypothèse est toutefois moins crédible pour la fenêtre mobile de 200 observations.

Les figures 4 et 5 offrent un portrait plus détaillé des erreurs dans un modèle évoluant avec une fenêtre mobile de 100 ou 200 observations. On y présente la différence entre l'erreur cumulative du modèle incluant et excluant les données de recherche. Lorsque la courbe est positive, l'erreur cumulée du modèle incluant les données de recherche est plus importante. C'est-à-dire qu'il fait une moins bonne prévision des rendements cumulés jusqu'à ce moment. Inversement, lorsque la courbe est négative, l'inclusion des données est favorable au modèle.

Figure 4 – Erreur cumulative pour la prédiction des rendements absolus du Dow Jones, Nasdaq et Nyse avec l'application d'une fenêtre mobile de 200 observations

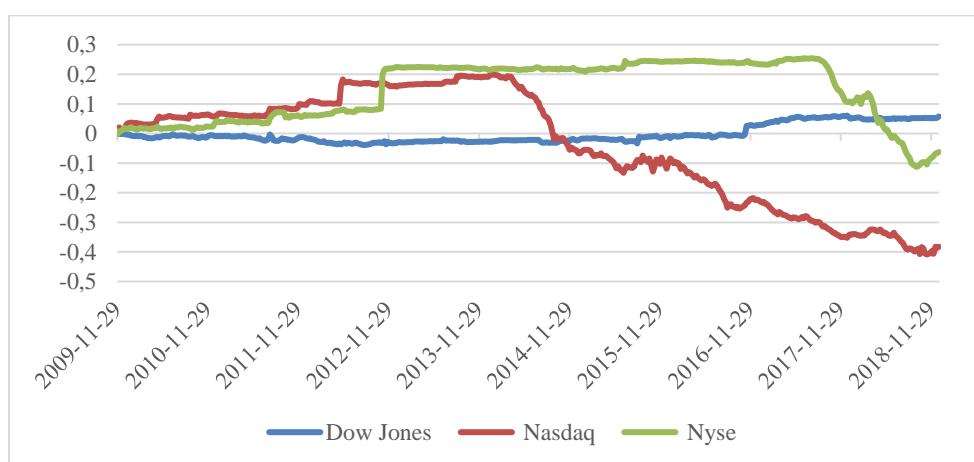
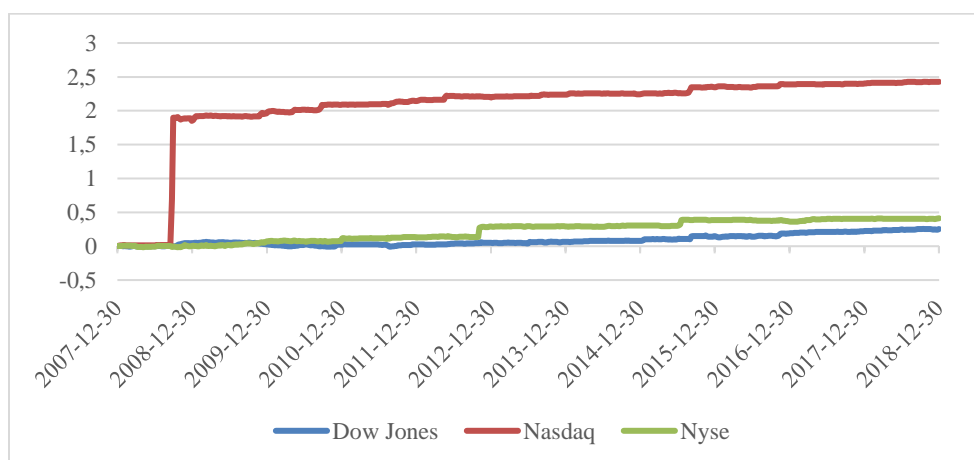


Figure 5 - Erreur cumulative pour la prédiction des rendements absolus du Dow Jones, Nasdaq et Nyse avec l'application d'une fenêtre mobile de 100 observations



On observe deux comportements intéressants. D'abord, la différence entre les erreurs cumulative peut évoluer abruptement. Une seule observation mal prédite fait significativement augmenter (ou diminuer) la mesure. C'est le cas pour l'observation au 20 mai 2012 et du 3 novembre 2012 pour la série du Nasdaq et du Nyse respectivement à la Figure 4 et pour le 28 septembre 2008 du Nasdaq à la Figure 5.

Ensuite, pour le Nasdaq à la Figure 4, on voit que la différence des erreurs diminue progressivement pour les 5 dernières années de l'échantillon (2014 à 2019), suggérant que l'inclusion des données de recherches est profitable à la prévision. Cette relation n'est toutefois pas suffisamment importante pour que l'erreur quadratique moyenne du modèle incluant les données de recherches soit plus faible.

Chapitre 4 - Discussion

Au chapitre 3, nous avons objectivement présenté les résultats des régressions estimées et certains constats surprenants seront le sujet d'une discussion plus poussée à ce quatrième chapitre. D'abord, nous identifions les points de convergence et de divergence entre la littérature et nos résultats. Ensuite, nous identifions la plus-value et les apports originaux de ce mémoire. Finalement, on conclut en commentant l'atteinte des objectifs et les limites du mémoire.

4.1 Convergences et divergences avec la littérature

Ce mémoire couvre les données de recherche pour 7 compagnies à forte capitalisation boursières et 3 indices de marché importants sur l'horizon temporel de 2006 à 2019. Cette période permet une comparaison intéressante avec la littérature qui s'est jusqu'à présent intéressée à des périodes plus courtes. Une première vague d'article évalue la relation entre les termes de recherches (sigles boursiers) pour des entreprises inscrites à la bourse et leurs rendements boursiers (Da et al., 2011; Joseph et al., 2011; Challet et Ayed, 2013; Preis et al., 2013). Ces articles trouvent généralement une relation positive et significative entre les rendements et les termes de recherche entre le milieu des années 2000 et le début des années 2010. Autrement que pour l'estimation du rendement absolu des indices de marché, nous n'avons pas trouvé de variables de recherche qui soit statistiquement significative pour toutes les régressions.

La différence peut s'expliquer par plusieurs divergences méthodologiques, les plus importantes étant sans doute le choix des termes de recherche et la panélisation des données pour un plus grand nombre d'entreprises. Les premières études analysent l'impact des recherches pour les sigles boursiers des entreprises plutôt que les libellés. Par exemple, plutôt que d'observer le volume de recherche pour le terme « apple », ils observent le volume pour le sigle « appl ». Les premières études regroupent également plus de données dans un panel. (Da et al., 2011) recueille de l'information pour 3 600 entreprises américaines. L'important volume de données permet sans doute une plus grande précision dans l'estimation de l'intervalle de confiance des coefficients. Il est également possible que le plus grand échantillon intègre des entreprises dont la nature est foncièrement différente des entreprises à forte capitalisation boursière dont le profil est très public.

La seconde vague de recherche sur la relation entre les volumes de recherche et les rendements boursiers présente des résultats semblables à ceux de ce mémoire. En effet, Perlin et al. (2016), Neri et Rovaiel (2017), Granell et Carlsson (2018) ne trouvent pas de relation de causalité entre les deux variables et contrairement aux auteurs précédents, leurs échantillons de données couvrent la période d'après la crise économique, ce qui semble être un facteur important pour l'établissement de la relation.

Pour les régressions sur les rendements absolus des indices de marché, les résultats sont cohérents avec la littérature qui trouve une relation progressivement décroissante entre les volumes de recherche et les rendements. Da et al. (2011) trouve également un effet de renversement à plus long terme. Les résultats de ce mémoire suggèrent un processus beaucoup plus rapide.

4.2 Apports originaux

Contrairement à la littérature existante, ce mémoire propose deux avancements méthodologiques : l'inclusion d'un effet pour les volumes de recherche élevés et une identification impartiale des bris structurels dans l'estimation des régressions. Ces deux avancements sont permis par la construction d'un échantillon de données couvrant une plus grande période de temps (janvier 2006 à décembre 2019).

Pour les régressions dont les coefficients de recherche sont significatifs, on démontre bien qu'une forte variation des volumes de recherche a un impact plus significatif qu'une faible variation. C'est donc de dire que les nouvelles qui attirent plus largement l'attention du public devraient avoir un effet sur les rendements boursiers.

Ensuite, l'application de la méthode d'identification des bris structurels de (Bai et Perron, 1998) a efficacement identifié l'effet de la crise économique sur l'estimation des régressions. Cet avancement est particulièrement notable, car il apporte une démarche impartiale pour définir la période de crise économique avec une précision hebdomadaire et individuelle. Les résultats montrent bien que l'effet de la crise ne se fait pas sentir au même moment pour toutes les entreprises. La démarche permet également d'identifier une limite à l'effet de la crise économique pour les indices de marché ; l'effet de la crise prend nettement fin au mois de mai 2011.

4.3 Limites et sources d'erreur

Le premier objectif de ce mémoire était d'évaluer si les volumes de recherches avaient un rapport de causalité avec les rendements boursiers. La littérature est partagée sur le sujet (voir la synthèse des articles rédigés en annexe) et ce mémoire l'est tout autant. En se basant sur le critère de sélection d'Akaike, on accepte l'inclusion des variables de recherche pour cinq régressions de rendements. Parmi celles-ci, les résultats les plus robustes sont trouvés pour l'estimation des rendements absolus des indices de marché lorsqu'on inclut des bris structurels dans l'estimation. On trouve donc une relation de causalité robuste, mais dont la portée est incertaine.

Trois raisons justifient le doute face à la portée des résultats. D'abord, les régressions les plus robustes sont celles où on prédit le rendement absolu. Comme il l'a été mentionné dans le chapitre méthodologique, cette estimation est beaucoup moins rigoureuse pour établir la causalité entre les volumes de recherche et les rendements boursiers, car elle implique implicitement que nous sommes capables d'identifier l'effet directionnel qu'un choc d'information devrait avoir sur le marché. Cette hypothèse est raisonnable, mais pas formellement vérifiée pour notre échantillon. Ensuite, même si on admet une corrélation robuste entre les variables de recherche et les rendements, la méthodologie de ce travail ne permet pas d'évaluer avec certitude la valeur prédictive des données. En segmentant les séries à leurs points optimaux, nous avons créé des sous-ensembles de données qui sont plus homogènes que l'échantillon complet. Cependant, on ne vérifie pas que le processus est effectivement homogène. Finalement, le devis méthodologique empêche la construction d'un panel de données beaucoup plus large. Avec la segmentation des échantillons, les modèles estimés dans ce mémoire incluent plusieurs paramètres pour un nombre limité d'observations et le risque de surajuster le modèle.

Conclusion

Ce mémoire de maîtrise a trois objectifs : vérifier l'existence d'une relation de causalité entre les rendements des entreprises cotées en bourse et leur volume de recherche respectif, intégrer cette information financière dans une méthodologie axée sur la prédiction des rendements boursiers et comprendre les interactions entre le volume de recherche et les rendements boursiers.

Nous remplissons le premier objectif en comparant quatre modèles d'anticipations rationnelles. Les deux premiers évaluent l'effet des volumes de recherches sur les rendements bruts et absolus respectivement. Les deux autres modèles réévaluent cette relation en incluant deux bris structurels. Parmi l'ensemble des régressions estimées sur ces modèles, nous trouvons une relation statistiquement significative entre les rendements absolus pour le Dow Jones, le Nasdaq et le Nyse pour la période de la crise économique (septembre 2008 à mai 2011).

Le second objectif est rempli en évaluant les modèles sur une fenêtre mobile de 100 observations. On teste ensuite la valeur prédictive du modèle sur l'observation suivante et on ne trouve aucune valeur prédictive pour les rendements absolus des indices de marchés en incluant les volumes de recherche.

Le troisième objectif vérifie deux comportements déjà ciblés par la littérature. D'abord une hausse du volume de recherche affecte les rendements suivant le choc et l'effet s'estompe graduellement (Da et al., 2011). Les résultats obtenus supportent cette littérature. Ensuite, l'effet marginal des volumes de recherche sur les rendements absolus s'inverse à la deuxième semaine suivant le choc de recherche. Cet effet est disputé dans la littérature et partiellement validé par les résultats de ce mémoire.

En plus d'être cohérents avec la littérature, les résultats de ce mémoire apportent de nouvelles nuances sur l'effet des volumes de recherche : la crise économique affecte significativement la relation entre les volumes de recherche et les rendements. L'inclusion des bris structurels estimés par la méthode de (Bai et Perron, 1998) permet d'identifier une période de deux ans et demi (octobre 2008 à mai 2011) attribuables à la crise économique où la relation entre les volumes et rendement change. Cette observation permet de partiellement réconcilier la

littérature qui ne trouve pas de corrélation entre les rendements et les volumes de recherche avec celle qui la trouve.

Ce mémoire échoue toutefois à trouver une relation suffisamment robuste pour utiliser les volumes de recherche dans une stratégie de prédiction des rendements boursiers. Sachant qu'une partie de la littérature prouve l'existence d'une telle relation, il apparaît important d'identifier les hypothèses méthodologiques permettant d'y arriver. Les recherches futures devraient également exploiter davantage les données de recherches fournies par Google Trends. La vaste majorité de la littérature se concentre sur des périodicités hebdomadaires alors que Google Trends permet un raffinement quotidien.

Bibliographie

- Abel, A. et F. Mishkin (1983), « On the Econometric Testing of Rationality-Market Efficiency », *The Review of Economics and Statistics*, Vol 65. No 2. P318-323
- Akaike H. (1974), « A New Look at the Statistical Model Identification », *Transactions on automatic controls*, Vol 19. No 6. P716-723
- Andrews D. (1993), « Test for Parameter Instability and Structural Change with Unknown Change Point », *Econometrica*, Vol 61. No 4. P821-856
- Bai J. et P. Perron (1998), « Estimating and Testing Linear Models with Multiple Structural Changes », *Econometrica*, Vol 66. No 1. P47-78
- Baltagi Badi (2005), *Econometric Analysis of Panel Data*. John Wiley and Sons, 291 P.
- Barber B. et Odean T. (2008), « All That Glitters : The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors », *Review of Financial Studies*, Vol.21. No 2. P785-818
- Bijl L. et al. (2016), « Google Searches and Stock Returns », *International Review of Financial Analysis*, Vol 45. P150-156
- Dimpfl T et S. Jank (2016). « Can Internet Search Queries Help to Predict Stock Market Volatility? », *European Financial Management*. Vol 22. No 2. P171-192
- Challet D. et A. Ayed (2013), *Predicting Financial Markets with Google Trends and not some Random Keywords*, SSRN. 9P.
- Da Z. et al. (2011), « In Search of Attention », *The Journal of Finance*, Vol 46. No 5. P1461-1499
- Fama E. (1970), « Efficient Capital Markets: A Review of Theory and Empirical Work », *The Journal of Finance*, Vol 25. No 2. P383-417
- Granell A. et F. Carlsson (2018), *How Google Search Trends Can Be Used as Technical Indicators for the S&P500-Index*, Royal Institute of Technology, 45P.
- Granger C. (1969), « Investigating Causal Relations by Econometric Models and Cross-spectral Methods », *Econometrica*, Vol 37. No 3. P424-438
- Greene W. (2012). *Econometric analysis – seventh edition*. Pearson. Harlow, Angleterre 1211P.
- Joseph K. et al. (2011), « Forecasting Abnormal Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search », *International Journal of Forecasting*, 21P.

- Kim N. et R. Villa (2017), *Google and Financial Markets: Can Google Trends Describe and Predict the Dynamics of Norwegian Stock Market?*, University of Stavanger. 51P.
- Lewellen J. (2004), « Predicting Returns with Financial Ratios », *Journal of Financial Economics*, Vol 74. P209-235
- Malkiel G. (2005). « Reflections on the Efficient Market Hypothesis ». *The Financial Review*. No 40. P1-9
- Mankiw G et M. Shapiro (1985). « Do we reject too often? Small Sample Properties of Tests of Rational Expectations Models », *Nber technical Paper*, No 51. 12P.
- Mishkin F. (2013), *Monnaie, Banque et Marchés Financiers – 10e édition*. Pearson. 1025P.
- Muth J. (1961), « Rational Expectations and the Theory of Price Movements », *Econometrica*, Vol 29. No 3. P315-335
- Neusser K (2016). *Time Series Econometrics*. Springer Tests in Business and Economics. 409P.
- Perlin M. et al. (2016), « Can we predict the Financial Markets Based on Google's Search Queries? », *Journal of Forecasting*. 18P.
- Preis T. et al. (2013), « Quantifying Trading Behavior in Financial Markets Using Google Trends », *Scientific Reports*. 6P.
- Sewell M. (2012), « The Efficient Market Hypothesis: Empirical Evidence », *International Journal of Statistics and Probability*, Vol 1. No 2. P164-178
- Varian H. et Choi H. (2009), « Predicting the Present with Google Trends », Google Research Blog, 23P.
- Westerlund J et P. Narayan (2015). « Testing for Predictability in Conditionnaly Heteroskedastic Stock Returns ». *Journal of Financial Econometrics*. Vol 13. No 2. P342-375

Annexe A – Démonstration du biais pour l'équation des anticipations rationnelles

Preuve 1 : $\hat{\beta} - \beta = \gamma(\hat{\rho} - \rho) + \eta$

Le modèle des anticipation rationnelles formalisé par (Mankiw et Shapiro, 1983) prend la forme d'un système de deux équations :

$$Y_t = \beta X_{t-1} + \varepsilon_t \quad (1)$$

$$X_t = \rho X_{t-1} + v_t \quad (2)$$

La première équation représente une régression normale où on cherche à évaluer l'effet de plusieurs variables explicative X_t sur une variable d'intérêt Y_t . La seconde équation représente plutôt le caractère autorégressif des variables explicatives de la première équation.

Pour arriver au résultat de (Lewellen, 2002), on doit commencer par accepter trois équations :

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon \quad (3)$$

$$\hat{\rho} = \rho + (X'X)^{-1}X'v \quad (4)$$

$$\varepsilon_t = \gamma v_t + \mu \quad (5)$$

Les deux premières équations sont la solution par MCO des coefficients pour les régressions (1) et (2). La troisième équation admet l'existence d'une relation linéaire entre les termes d'erreurs des deux régressions du modèle.

En substituant (5) dans (3) :

$$\hat{\beta} = \beta + (X'X)^{-1}X'(\gamma v_t + \mu)$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\gamma v_t + (X'X)^{-1}X'\mu$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\gamma v_t + \eta \quad (6)$$

Pour procéder avec la simplification, on peut réarranger l'équation (4) pour isoler le terme d'erreur.

$$\hat{\rho} = \rho + (X'X)^{-1}X'v$$

$$(\hat{\rho} - \rho) = (X'X)^{-1}X'v$$

$$[(X'X)^{-1}X']^{-1}(\hat{\rho} - \rho) = v$$

Et ensuite, le remplacer dans l'équation (6) :

$$\hat{\beta} = \beta + (X'X)^{-1}X'\gamma v_t + \eta$$

$$(\hat{\beta} - \beta) = (X'X)^{-1}X'\gamma[(X'X)^{-1}X']^{-1}(\hat{\rho} - \rho) + \eta$$

$$(\hat{\beta} - \beta) = \gamma(X'X)^{-1}X'[(X'X)^{-1}X']^{-1}(\hat{\rho} - \rho) + \eta$$

$$(\hat{\beta} - \beta) = \gamma I(\hat{\rho} - \rho) + \eta$$

Annexe B – Tableau résumé de la recension des études

Auteurs	Objectif de performance	Termes	Horizon temporel	Indices boursiers	Résultat
Da et al., 2011	Relation de causalité	Symbols d'actions	Données hebdomadaires de 2004 à 2008	Russel 3000	Prédiction positive des rendements
Kiassan et al., 2011	Relation de causalité	Symbols d'actions	Données hebdomadaires de 2005 à 2008	S&P 500	Prédiction positive des rendements et prédiction de la volatilité
Challet et Ayed, 2013	Stratégie d'investissement	Termes généraux	Données hebdomadaires de 2005 à 2013	SPY	Prédiction positive des rendements
Preis et al., 2013	Stratégie d'investissement	Termes généraux	Données hebdomadaires de 2004 à 2011	Dow Jones	Prédiction positive des rendements, sans coûts de transaction
Perlin et al., 2016	Stratégie d'investissement	Termes généraux	Données hebdomadaires de 2005 à 2014	S&P500, FTSE, TSX, ASX 200	Prédiction positive des rendements, particulièrement pendant la crise financière de 2008
Biil, Kringhaug et al., 2016	Stratégie d'investissement	Libellés d'entreprise	Observations hebdomadaires de 2007 à 2013	S&P 500	Prédiction positive des rendements sans coûts de transaction
Dimpfl et Jank, 2016	Relation de causalité	Libellés d'indices boursiers	Observations hebdomadaires de 2006 à 2011	Dow Jones	Prédiction de la volatilité
Neri et Rovarel, 2017	Relation de causalité	Libellés d'entreprise	Observations hebdomadaires de 2012 à 2017	OBX	Aucune prédiction systématique des rendements
Granell et Carlsson, 2018	Stratégie d'investissement	Termes généraux	Données hebdomadaires de 2013 à 2018	S&P 500	Aucune prédiction systématique des rendements

Annexe C – Tableau résumé des critères d’informations (régressions sans bris structurels)

Index	Données de rendement bruts			Données de rendement absolus		
	Modèle boursier	Modèle complet	Décision	Modèle boursier	Modèle complet	Décision
Dow Jones	-7,5700	-7,56099	Bourse	-8,3766	-8,36536	Bourse
Nasdaq	-7,2759	-7,27550	Bourse	-8,0955	-8,08260	Bourse
Nyse	-7,3591	-7,35080	Bourse	-8,1613	-8,15405	Bourse
Indices						
Amazon	-5,9439	-5,93531	Bourse	-6,6325	-6,62387	Bourse
Google	-6,5008	-6,49005	Bourse	-7,2933	-7,27770	Bourse
Apple	-6,2621	-6,26650	Complet	-7,1426	-7,13240	Bourse
Microsoft	-6,7816	-6,77306	Bourse	-7,4947	-7,48230	Bourse
Johnson & J	-7,7788	-7,76459	Bourse	-8,6096	-8,59178	Bourse
Berkshire	-7,0697	-7,05721	Bourse	-7,7032	-7,70883	Complet
JP Morgan	-5,9682	-5,96450	Bourse	-6,5666	-6,56277	Bourse

Annexe D – Tableau résumé des critères d’informations (régressions avec bris structurels)

Index	Données de rendement bruts			Données de rendement absolus		
	Modèle boursier	Modèle complet	Décision	Modèle boursier	Modèle complet	Décision
Dow Jones	-7,5728	-7,55251	Bourse	-8,4191	-8,49044	Complet
Nasdaq	-7,2955	-7,28101	Bourse	-8,1383	-8,11136	Bourse
Nyse	-7,3651	-7,41022	Complet	-8,2197	-8,29261	Complet
Indices						
Amazon	-5,9440	-5,91235	Bourse	-6,6744	-6,63561	Bourse
Google	-6,5052	-6,47449	Bourse	-7,3508	-7,31419	Bourse
Apple	-6,2673	-6,25156	Bourse	-7,1953	-7,15832	Bourse
Microsoft	-6,7838	-6,77019	Bourse	-7,5125	-7,45636	Bourse
Johnson &J	-7,7754	-7,74035	Bourse	-8,6251	-8,58029	Bourse
Berkshire	-7,0679	-7,01990	Bourse	-7,7398	-7,70768	Bourse
JP Morgan	-5,9867	-5,96221	Bourse	-6,6889	-6,56596	Bourse

Annexe E – Régressions sur les rendements bruts (sans bris structurels)

Les coefficients en italique sont significatif au seuil de 10% et ceux en gras au seuil de 5%

	Dow Jones	Nasdaq	Nyse	Amazon	Google	Apple	Microsoft	Johnson & J	Berkshire H.	JP Morgan
Constante	0,0019	0,0031	0,0003	0,0069	0,0027	0,0066	0,0023	0,0014	0,0014	0,0043
bourse_délais1	-0,1103	-0,0426	-0,0646	-0,0113	-0,0317	-0,0221	-0,0833	-0,0488	-0,1654	-0,1716
bourse_délais2	0,0206	-0,0101	0,0402	0,0215	0,0729	0,0083	0,0368	0,0294	0,0211	0,0015
google_délais1	-0,0051	-0,0115	-0,0119	0,0798	-0,0013	-0,0101	-0,0085	-0,0093	-0,0008	-0,0401
google90_délais1	0,0007	0,0001	0,0047	0,0003	0,0120	-0,0177	-0,0136	0,0028	0,0077	-0,0090
google_délais2	-0,0023	0,0124	-0,0088	0,0538	0,0015	0,0664	-0,0497	0,0022	-0,0007	0,0074
google90_délais2	-0,0016	-0,0089	0,0012	0,0104	-0,0096	0,0009	0,0101	-0,0009	0,0031	-0,0008
google_délais1*google90_délais1	0,0001	0,0061	0,0078	-0,0653	-0,0830	0,0364	0,0801	0,0055	-0,0032	0,0463
google_délais2*google90_délais2	-0,0014	-0,0173	0,0002	-0,0951	0,0900	-0,0965	-0,0036	-0,0057	-0,0037	-0,0200
R ² ajusté	0,0098	0,0081	0,0040	-0,0020	0,0020	0,0130	0,0060	-0,0040	0,0200	0,0360
Somme du carré des résidus	0,3420	0,4550	0,4220	1,7380	0,9980	1,2480	0,7520	0,2790	0,5660	1,6880

Annexe F – Régressions sur les rendements bruts (avec bris structurels)

Les coefficients en italique sont significatif au seuil de 10% et ceux en gras au seuil de 5%

	Dow Jones	Nasdaq	Nyse	Amazon	Google	Apple	Microsoft	Johnson & J	Berkshire H.	JP Morgan
Segment 1 - Constante	0,0023	0,0002	-0,0011	0,0077	-0,0010	0,0021	-0,0031	0,0002	-0,0042	0,0020
Segment 2 - Constante	0,0029	0,0051	0,0022	0,0116	<i>0,0041</i>	0,0100	0,0010	<i>0,0039</i>	0,0048	0,0026
Segment 3 - Constante	0,0017	0,0033	0,0001	<i>0,0057</i>	0,0040	0,0042	0,0062	0,0016	0,0014	0,0044
Segment 1 - Bourse_délais1	-0,1295	-0,0596	-0,0167	-0,0350	-0,0173	0,1584	-0,1278	-0,0712	-0,2059	-0,3389
Segment 2 - Bourse_délais1	-0,0772	0,1393	-0,1010	0,0499	0,0626	-0,0695	0,0265	0,1501	-0,1279	0,0088
Segment 3 - Bourse_délais1	-0,1455	-0,1614	-0,0822	-0,0475	-0,1727	-0,1315	-0,1490	-0,1244	-0,2108	-0,1239
Segment 1 - Bourse_délais2	0,0909	0,1672	0,2081	<i>0,1106</i>	0,1934	0,0399	0,1141	0,0798	-0,0112	-0,1276
Segment 2 - Bourse_délais2	-0,0433	-0,1879	-0,0572	-0,1575	0,0237	-0,0746	-0,0105	-0,0796	0,0666	<i>0,1055</i>
Segment 3 - Bourse_délais2	-0,0176	-0,0059	0,0240	0,0290	-0,0787	0,1206	0,0066	-0,0249	0,0030	0,0541
Segment 1 - Google_délais1	0,0060	-0,0202	0,0200	0,1654	-0,0833	-0,0418	0,0667	-0,0152	-0,0036	-0,0547
Segment 2 - Google_délais1	-0,0290	0,0046	-0,0391	0,1824	0,1065	-0,0196	-0,0007	-0,0054	-0,0046	-0,0365
Segment 3 - Google_délais1	-0,0049	-0,0178	-0,0116	0,0199	0,0102	0,0124	-0,0319	0,0011	0,0093	-0,0402
Segment 1 - Google90_délais1 (indicatrice)	-0,0003	0,0198	<i>0,0176</i>	-0,0055	0,0197	-0,0076	0,0677	0,0058	0,0282	<i>-0,0334</i>
Segment 2 - Google90_délais1 (indicatrice)	0,0001	-0,0605	<i>0,0572</i>	0,0070	0,0123	-0,0005	-0,0047	-0,0032	-0,0042	<i>0,0571</i>
Segment 3 - Google90_délais1 (indicatrice)	-0,0013	-0,0021	0,0017	-0,0057	0,0194	-0,0339	-0,0193	0,0153	-0,0022	0,0239
Segment 1 - Google_délais2	0,0074	0,0098	0,0148	0,3424	0,1574	0,0592	-0,0320	0,0071	-0,0081	<i>0,0538</i>
Segment 2 - Google_délais2	-0,0192	0,0211	-0,0104	-0,0775	0,0023	0,0829	<i>-0,0905</i>	-0,0104	0,0007	-0,0104
Segment 3 - Google_délais2	-0,0037	0,0089	-0,0147	0,0042	-0,0349	0,0390	-0,0111	0,0010	0,0168	-0,0200
Segment 1 - Google90_délais2 (indicatrice)	-0,0021	-0,0129	<i>0,0196</i>	-0,0114	-0,0230	0,0339	0,0095	-0,0009	0,0123	0,0097
Segment 2 - Google90_délais2 (indicatrice)	0,0069	0,0273	-0,0184	0,0165	-0,0348	-0,0180	0,0036	0,0016	0,0016	-0,0284
Segment 3 - Google90_délais2 (indicatrice)	0,0001	-0,0031	0,0045	0,0120	-0,0078	0,0094	-0,0049	0,0212	-0,0061	-0,0146
Segment 1 - Google90_délais1*Google_délais1	-0,0142	-0,0200	-0,0620	-0,1199	-0,0104	0,0045	-0,5991	0,0060	-0,0084	0,1213
Segment 2 - Google90_délais1*Google_délais1	0,0503	0,2097	-0,0960	-0,1368	-0,3008	0,0260	0,0551	0,0044	0,0050	-0,1134
Segment 3 - Google90_délais1*Google_délais1	0,0018	0,0201	0,0117	-0,0103	-0,1298	0,0388	0,0789	-0,0479	-0,0265	-0,0609
Segment 1 - Google90_délais2*Google_délais2	-0,0175	-0,0422	-0,0468	-0,3096	0,0218	-0,1464	0,0987	-0,0192	0,0026	<i>-0,0730</i>
Segment 2 - Google90_délais2*Google_délais2	-0,0280	-0,1882	0,0352	0,0238	0,5150	-0,0993	0,0414	0,0082	-0,0100	0,0417
Segment 3 - Google90_délais2*Google_délais2	0,0065	-0,0121	0,0079	-0,0626	<i>0,0834</i>	-0,0640	0,0912	-0,0702	-0,0122	0,0333
R ² ajusté	0,026	0,046	0,028	0,021	0,021	0,027	0,017	-0,002	0,024	0,070
Somme du carré des résidus	0,327	0,429	0,377	1,686	0,961	1,201	0,715	0,271	0,557	1,604
Bits 1	2008-09-28	2008-10-12	2008-09-28	2008-09-28	2008-11-23	2008-09-07	2008-10-19	2012-03-11	2009-02-15	2009-02-15
Bits 2	2011-06-12	2011-07-10	2011-06-19	2011-07-03	2014-09-21	2014-10-26	2012-10-21	2014-11-09	2013-04-21	2012-05-06

Annexe G – régressions sur les rendements absolus (sans bris structurels)

Les coefficients en italique sont significatif au seuil de 10% et ceux en gras au seuil de 5%

	Dow Jones	Nasdaq	Nyse	Amazon	Google	Apple	Microsoft	Johnson & J	Berkshire H.	JP Morgan
Constante	0,0085	0,0117	0,0084	0,0329	0,0247	0,0260	0,0180	0,0120	0,0119	<i>0,0095</i>
bourse_déla1	0,2338	0,1863	0,2560	0,0539	0,0711	0,0840	0,0672	0,1638	0,1901	0,3335
bourse_déla2	0,1605	0,2021	0,2356	0,1016	<i>0,1166</i>	0,1252	0,1710	0,0546	0,1670	0,1303
google_déla1	0,0007	-0,0059	0,0054	-0,0733	-0,0095	0,0070	-0,0346	-0,0042	0,0021	0,0371
google90_déla1	0,0006	0,0046	0,0042	-0,0157	0,0003	<i>0,0019</i>	-0,0001	-0,0011	0,0125	0,0017
google_déla2	0,0056	0,0022	-0,0091	0,0039	-0,0575	0,0059	0,0351	0,0028	-0,0053	0,0369
google90_déla2	0,0012	-0,0016	0,0009	0,0097	0,0016	-0,0085	-0,0014	0,0021	0,0011	0,0016
google_déla1*google90_déla1	0,0012	0,0044	-0,0037	0,1431	0,0195	-0,0029	0,0168	0,0020	-0,0035	-0,0290
google_déla2*google90_déla2	-0,0034	-0,0050	0,0088	-0,0695	0,0401	-0,0038	0,0039	-0,0054	0,0026	-0,0484
R ² ajusté	0,1190	0,089	0,173	0,009	0,013	0,02	0,028	0,025	0,097	0,168
Somme du carré des résidus	0,1530	0,203	0,189	0,873	0,454	0,525	0,37	0,122	0,295	0,928

Annexe H – Régressions sur les rendements absolus (avec bris structurels)

Les coefficients en italique sont significatif au seuil de 10% et ceux en gras au seuil de 5%

	Dow Jones	Nasdaq	Nyse	Amazon	Google	Apple	Microsoft	Johnson & J	Berkshire H.	JP Morgan
Segment 1 - Constante	0,0183	0,0173	0,0133	0,0512	0,0240	0,0343	0,0201	0,0136	0,0256	0,0130
Segment 2 - Constante	0,0079	0,0157	0,0122	0,0454	<i>0,0291</i>	0,0307	0,0182	<i>0,0103</i>	0,0118	0,0189
Segment 3 - Constante	0,0084	0,0110	0,0095	<i>0,0252</i>	0,0263	0,0279	0,0167	0,0170	0,0117	0,0177
Segment 1 - Bourse_délaiss1	<i>-0,0992</i>	<i>-0,0211</i>	<i>-0,0789</i>	<i>-0,0119</i>	0,0154	0,0492	<i>0,1361</i>	<i>-0,0133</i>	0,1541	0,3997
Segment 2 - Bourse_délaiss1	0,1544	0,1602	<i>0,3062</i>	<i>-0,0764</i>	<i>-0,0368</i>	<i>-0,0034</i>	0,0309	0,2788	<i>0,2582</i>	0,2553
Segment 3 - Bourse_délaiss1	0,1935	0,1537	0,1322	0,0849	0,1632	0,0429	0,1003	<i>0,0046</i>	0,1083	0,1289
Segment 1 - Bourse_délaiss2	0,0327	0,1843	0,1097	<i>0,0833</i>	0,2919	0,0647	0,0746	0,0092	0,1900	0,1301
Segment 2 - Bourse_délaiss2	0,1726	0,2144	0,2344	0,0513	<i>-0,0188</i>	0,1252	0,2511	0,0529	0,0808	0,0817
Segment 3 - Bourse_délaiss2	0,1483	0,1350	0,1989	<i>-0,0065</i>	<i>-0,1072</i>	<i>-0,0083</i>	0,0632	<i>-0,0201</i>	0,0921	-0,0011
Segment 1 - Google_délaiss1	<i>-0,0139</i>	<i>-0,0091</i>	0,0056	<i>-0,1860</i>	0,1626	0,0770	<i>-0,1658</i>	<i>-0,0231</i>	<i>-0,0150</i>	0,0533
Segment 2 - Google_délaiss1	0,0165	0,0075	0,0224	<i>-0,0704</i>	<i>-0,0615</i>	<i>-0,0232</i>	<i>-0,0167</i>	0,0082	0,0065	0,0309
Segment 3 - Google_délaiss1	0,0027	<i>-0,0056</i>	0,0004	0,0086	<i>-0,0305</i>	<i>-0,0993</i>	0,0085	<i>-0,0184</i>	0,0027	0,0018
Segment 1 - Google90_délaiss1 (indicatrice)	<i>-0,0025</i>	<i>-0,0024</i>	<i>0,0019</i>	<i>-0,0405</i>	0,0252	0,0029	0,0756	<i>-0,0258</i>	0,0174	<i>-0,0073</i>
Segment 2 - Google90_délaiss1 (indicatrice)	<i>-0,0004</i>	<i>-0,0523</i>	<i>-0,0918</i>	<i>-0,0085</i>	0,0295	0,0022	<i>-0,0068</i>	0,0009	0,0059	<i>0,0048</i>
Segment 3 - Google90_délaiss1 (indicatrice)	<i>-0,0016</i>	0,0009	0,0019	<i>-0,0092</i>	<i>-0,0107</i>	<i>-0,0224</i>	0,0051	0,0043	0,0164	<i>-0,0301</i>
Segment 1 - Google_délaiss2	<i>-0,0058</i>	<i>-0,0054</i>	<i>-0,0014</i>	-0,0265	<i>-0,0537</i>	0,0470	0,1355	0,0179	<i>-0,0417</i>	<i>0,0598</i>
Segment 2 - Google_délaiss2	0,0228	<i>-0,0202</i>	<i>-0,0242</i>	<i>-0,1733</i>	<i>-0,0578</i>	-0,0344	<i>0,0104</i>	<i>-0,0044</i>	0,0112	0,0250
Segment 3 - Google_délaiss2	0,0038	0,0131	<i>-0,0067</i>	0,0672	<i>-0,0416</i>	0,0162	0,0220	0,0060	<i>-0,0012</i>	0,0038
Segment 1 - Google90_délaiss2 (indicatrice)	<i>-0,0023</i>	0,0033	<i>-0,0018</i>	<i>-0,0161</i>	0,0103	<i>-0,0208</i>	0,0344	<i>-0,0919</i>	<i>-0,0081</i>	<i>-0,0037</i>
Segment 2 - Google90_délaiss2 (indicatrice)	0,0208	<i>-0,0275</i>	0,0374	0,0041	<i>-0,0047</i>	<i>-0,0112</i>	0,0009	<i>-0,0007</i>	0,0070	0,0401
Segment 3 - Google90_délaiss2 (indicatrice)	0,0018	0,0011	<i>-0,0004</i>	0,0212	<i>-0,0002</i>	0,0057	<i>-0,0107</i>	<i>-0,0071</i>	<i>-0,0179</i>	<i>-0,0374</i>
Segment 1 - Google90_délaiss1*Google_délaiss1	0,0156	0,0229	-0,0017	0,4552	<i>-0,2362</i>	<i>-0,0800</i>	<i>-0,4569</i>	0,1222	0,0105	-0,0462
Segment 2 - Google90_délaiss1*Google_délaiss1	0,0470	0,2227	0,2609	0,0708	<i>-0,1448</i>	0,0216	0,0168	<i>-0,0104</i>	<i>-0,0053</i>	<i>-0,0141</i>
Segment 3 - Google90_délaiss1*Google_délaiss1	<i>-0,0021</i>	0,0059	0,0009	0,0315	0,0934	0,1563	<i>-0,0289</i>	<i>-0,0096</i>	<i>-0,0242</i>	0,0988
Segment 1 - Google90_délaiss2*Google_délaiss2	0,0154	<i>-0,0253</i>	0,0279	0,0561	0,0421	<i>-0,0213</i>	<i>-0,3965</i>	0,3841	0,0586	<i>-0,0640</i>
Segment 2 - Google90_délaiss2*Google_délaiss2	<i>-0,0534</i>	0,0795	<i>-0,1078</i>	0,1145	0,2052	0,0381	0,0233	0,0020	<i>-0,0183</i>	<i>-0,1426</i>
Segment 3 - Google90_délaiss2*Google_délaiss2	<i>-0,0041</i>	<i>-0,0162</i>	0,0070	<i>-0,1556</i>	<i>0,0130</i>	<i>-0,0023</i>	0,0528	<i>0,0287</i>	<i>0,0246</i>	0,0939
R ² ajusté	0,6129	0,5828	0,62	0,522	0,576	0,606	0,514	0,568	0,502	0,486
Somme du carré des résidus	0,128	0,187	0,156	0,818	0,415	0,485	0,36	0,117	0,28	0,877
Bris 1	2008-09-07	2008-08-31	2008-09-07	2009-08-23	2009-01-11	2008-12-28	2008-08-31	2008-08-10	2009-02-22	2009-02-08
Bris 2	2011-05-08	2011-05-01	2011-05-15	2012-03-25	2015-05-31	2015-05-24	2014-11-30	2015-11-29	2011-11-06	2012-06-17

Annexe I – Comparaison entre les bris sélectionnés pour les modèles boursiers et complets

	Rendements réels				Rendements absolus			
	Modèle boursier		Modèle complet		Modèle boursier		Modèle complet	
	Premier bris	Second bris	Premier bris	Second bris	Premier bris	Second bris	Premier bris	Second bris
Dow Jones	2008-10-12	2011-07-24	2008-09-28	2011-06-12	2008-08-31	2011-11-06	2008-09-07	2011-05-08
Nadag	2008-10-12	2011-07-10	2008-10-12	2011-07-10	2008-11-02	2011-11-06	2008-08-31	2011-05-01
Nyse	2008-09-14	2015-10-25	2008-09-28	2011-06-19	2008-08-24	2011-11-06	2008-09-07	2011-05-15
Amazon	2008-09-28	2011-07-03	2008-09-28	2011-07-03	2009-01-18	2012-05-13	2009-08-23	2012-03-25
Google	2008-10-26	2014-09-21	2008-11-23	2014-09-21	2009-01-18	2011-11-06	2009-01-11	2015-05-31
Apple	2008-09-07	2014-11-30	2008-09-07	2014-10-26	2009-03-15	2013-08-18	2008-12-28	2015-05-24
Microsoft	2008-12-28	2013-04-07	2008-10-19	2012-10-21	2009-06-28	2016-04-10	2008-08-31	2014-11-30
Johnson & J	2012-03-11	2014-11-09	2012-03-11	2014-11-09	2008-08-24	2011-11-06	2008-08-10	2015-11-29
Berkshire H.	2009-01-25	2013-04-21	2009-02-15	2013-04-21	2008-08-31	2011-10-02	2009-02-22	2011-11-06
JP Morgan	2009-02-15	2012-05-06	2009-02-15	2012-05-06	2009-02-15	2012-07-01	2009-02-08	2012-06-17

