

UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

MÉMOIRE DE MAÎTRISE
PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

COMME EXIGENCE PARTIELLE
DE LA MAITRISE EN ÉCONOMIE FINANCIÈRE

PAR
BERTRAND LEMIEUX

LA RELATION ENTRE LA DIVERGENCE D'OPINIONS ET LE RENDEMENT FUTUR
DES TITRES SUR LE MARCHÉ BOURSIER

JUIN 2017

TABLE DES MATIÈRES

Sommaire.....	iv
Remerciements	v
Liste des tableaux	vi
1. Introduction.....	1
2. Revue de littérature	3
2.1. Revue générale.....	3
2.2. Les travaux de Miller	13
2.2.1. Introduction.....	13
2.2.2. Description du modèle.....	14
2.2.3. L'excès spéculatif et l'hypothèse de l'efficience des marchés.....	14
2.2.4. Risque	15
2.2.5. Changement dans le risque et nouvelles émissions de titres.....	15
2.2.6. Risque systématique	16
2.2.7. Biais dans les valeurs utilisées pour établir les prévisions	16
2.2.8. Utilisation d'une prime de risque dans l'évaluation des titres	17
2.2.9. Baisse de la rigueur des hypothèses du modèle.....	17
2.2.10. Vente à découvert	18
2.2.11. Visibilité d'un titre.....	19
2.3. Conclusion	20
3. Échantillon et méthodologie.....	21
3.1. Source des données et variables.....	21
3.2. Manipulation des données dans EXCEL et GRETL.....	23
4. Statistiques descriptives et méthodologie	26
5. Régressions linéaires et méthodologie	33
5.1. Régression simple, aucun retard	33
5.2. Régression simple, avec retards.....	38
5.3. Discussion.....	43
5.4. Régression multiple, avec retards.....	44
5.4.1. Test de la statistique F	47
5.4.2. Test de normalité des résidus.....	47
5.4.3. Test de White – Hétéroscédasticité	50
5.4.4. Test de Durbin-Watson – Autocorrélation de premier ordre.....	51
5.4.5. Test de Breusch-Godfrey – Autocorrélation.....	52
5.5. Discussion.....	54
6. Régression avec le modèle des trois facteurs de Fama-French.....	55
6.1. Présentation du modèle.....	55

6.2. Équation du modèle de Fama-French	56
6.3. Explication détaillée de la méthodologie utilisée	58
6.4. Régression avec le modèle Fama-French – Aucun retard	60
6.5. Régression avec le modèle Fama-French – Retard d'un mois.....	62
6.6. Régression avec le modèle Fama-French – Retard de deux mois.....	63
6.7. Discussion.....	65
Conclusion	66
Bibliographie	68

SOMMAIRE

Miller (1977) suggère une relation théorique négative entre la divergence d'opinions sur la valeur d'un titre et son rendement ultérieur. Puisque plus la divergence d'opinions sur un titre est grande, plus le volume de transaction sur ce titre devrait être élevé, l'auteur a analysé la relation qui existe entre le volume des transactions sur un titre et son rendement futur à court terme. En établissant les équations de régression linéaire avec aucun retard, puis d'un à quatre retards, l'analyse des coefficients de la variable indépendante (volume des transactions) obtenus démontrent clairement que plus le volume de transactions sur un titre augmente, plus le coefficient de la variable indépendante diminue. Cet effet prend naissance durant le mois en cours et se résorbe après une période de quatre mois. L'utilisation du modèle des trois facteurs de Fama-French démontre d'ailleurs qu'il y a d'autres variables explicatives que celles retenues par ce modèle. Une de ces variables manquantes pourrait être reliée à la divergence d'opinions sur la valeur des titres. Le phénomène observé pourra être utilisé dans une stratégie de portefeuille à court terme afin d'en tirer profit.

REMERCIEMENTS

Je tiens d'abord à remercier sincèrement madame Céline Gauthier, Ph.D., pour la direction de ce mémoire, ainsi que pour sa disponibilité et ses précieux commentaires ou suggestions qui ont contribué à alimenter ma réflexion.

J'aimerais aussi remercier les membres du jury d'évaluation, soit messieurs David Tessier, Ph.D. et Christian Calmès, Ph.D., pour leur lecture attentive de ce mémoire, ainsi que pour l'évaluation qu'ils en ont fait.

Des remerciements s'adressent tout particulièrement aux professeurs de la maîtrise en économie financière, qui ont su transmettre leur expérience et professionnalisme au travers de leur enseignement, ainsi que tout le personnel professionnel de l'Université du Québec en Outaouais.

Enfin, je ne peux passer sous silence les nombreuses marques d'encouragements de ma conjointe, Manon LeBlanc, Ph.D., qui a été une source d'inspiration et de motivation dans la poursuite de mon cheminement académique. De même, je voudrais exprimer ma reconnaissance à mes parents, qui ont toujours reconnu la valeur des études universitaires et qui m'ont toujours encouragé à les poursuivre.

LISTE DES TABLEAUX

Tableau I-A	Statistiques descriptives par année.....	26
Tableau I-B	Statistiques descriptives par décile	27
Tableau II-A	Moyenne des rendements et dispersion.....	29
Tableau II-B	Évolution des rendements en fonction du volume	30
Tableau II-C	Évolution des rendements en fonction de la capitalisation boursière	31
Tableau III	Statistiques de la régression et analyse de variance – Aucun retard.....	34
Tableau IV-A	Équations de la régression – Un retard.....	39
Tableau IV-B	Équations de la régression – Deux retards	39
Tableau V-A	Statistiques de la régression et analyse de variance – Un retard	39
Tableau V-B	Statistiques de la régression et analyse de variance – Deux retards.....	40
Tableau VI-A	Équations de la régression – Trois retards.....	41
Tableau VI-B	Équations de la régression – Quatre retards.....	41
Tableau VII-A	Statistiques de la régression et analyse de variance – Trois retards	41
Tableau VII-B	Statistiques de la régression et analyse de variance – Quatre retards	42
Tableau VIII-A	Coefficients de la variable « Volume »	42
Tableau VIII-B	Évolution visuelle des coefficients de la variable « Volume ».....	43
Tableau IX-A	Écran de sortie de la régression multiple du quintile D1.....	44
Tableau IX-B	Restrictions linéaires – Quintile D1	45
Tableau IX-C	Écran de sortie de la régression multiple du quintile D2	45
Tableau IX-D	Écran de sortie de la régression multiple du quintile D3	46
Tableau IX-E	Écran de sortie de la régression multiple du quintile D4.....	46
Tableau IX-F	Écran de sortie de la régression multiple du quintile D5.....	47
Tableau X	Test de spécificité de Jarque-Bera	48

Tableau XI	Graphique de la distribution de fréquence des résidus – Quintile D1	49
Tableau XII	Graphique de la distribution de fréquence des résidus – Quintile D2	50
Tableau XIII	Test de spécificité de White	51
Tableau XIV	Graphique de la série temporelle des variances de résidus – Quintiles D1 et D2	51
Tableau XV	Test de spécificité de Durbin-Watson.....	52
Tableau XVI	Test de spécificité de Breusch-Godfrey	53
Tableau XVII	Statistiques de la régression et analyse de variance – Aucun retard.....	60
Tableau XVIII	Statistiques de la régression et analyse de variance – Retard d’un mois	62
Tableau XIX	Statistiques de la régression et analyse de variance – Retard de deux mois	64

1. INTRODUCTION

Un des fondements traditionnels de la finance moderne est que les investisseurs qui ont à prendre des décisions sont rationnels. Les recherches récentes ont toutefois démontré que les investisseurs ne prennent pas toujours leurs décisions de cette façon. Plusieurs études en psychologie tendent d'ailleurs à démontrer qu'il y a énormément de biais dans un processus de prise de décision. La finance comportementale a donc émergé en prenant en compte ces diverses théories et en cherchant à identifier dans quelle mesure celles-ci influencent réellement le processus de prise de décision.

La divergence d'opinions sur la valeur d'un titre peut s'exprimer de plusieurs façons, dont notamment par les écarts entre les prévisions des analystes sur la valeur future ou les bénéfices futurs d'un titre sur un certain horizon temporel, ou, de façon plus simple, par le volume des transactions sur un titre. Pour chaque investisseur qui pense que le titre a atteint une valeur maximale et veut le vendre, il existe un investisseur qui croit, au contraire, que la valeur du titre continuera à progresser. Plus les écarts précédents sont prononcés, plus la valeur d'un titre tend à augmenter. Comme il existe des contraintes pour les ventes à découvert (coûts de transactions, interdictions, etc.) et d'autres coûts de friction, certains investisseurs pessimistes pourraient décider ne pas vendre leur titre et de le conserver, ce qui aurait comme effet qu'une partie de l'opinion pessimiste ne se reflèterait pas dans le prix du titre, biaisant donc sa valeur de façon positive. Ce faisant, le rendement subséquent ne peut qu'être inférieur au rendement réel, étant mesuré sur une valeur de départ supérieure à ce qu'elle devrait réellement être. C'est ce que Miller (1977) invoque dans son modèle suggéré.

Plusieurs phénomènes ont une influence sur la valeur d'un titre et son rendement futur et une bonne compréhension de ceux-ci pourra permettre de mieux orienter le choix des titres dans l'établissement d'un portefeuille de placement. Un des phénomènes qui a fait l'objet d'une attention de plusieurs chercheurs est sans contredit le volume des transactions (Lee et Swaminathan (2000), Piqueira (2006), Hong et Stein (2007)). D'autres auteurs (Diether, Malloy et Scherbina (2002)) verront plutôt la variabilité dans les prévisions des analystes sur un titre comme plus révélateur que le volume des transactions pour expliquer la divergence d'opinions qui existe.

Au fil du temps, plusieurs chercheurs ont utilisé diverses techniques mathématiques et statistiques pour établir certains liens entre le volume des transactions et le rendement futur, ou du moins, pour démontrer qu'il existe un lien, même de manière intuitive. Ce document reprendra certaines techniques et prémisses de base qui sous-tendent les travaux effectués par les divers auteurs, dont notamment celles de Diether, Malloy et Scherbina (2002), mais en modifiant ou utilisant conjointement certaines des techniques de mesure quantitative qu'ils ont utilisées. Ce document tentera de mettre en lumière si la divergence d'opinion, mesurée par le volume des transactions, a un effet significatif ou non sur le rendement obtenu. L'échantillon des données sera puisé sur une période postérieure à celle utilisée par Diether, Malloy et Scherbina (2002) et tentera de démontrer si les investisseurs ont tenu compte ou non des résultats de cette étude dans leur processus décisionnel. En effet, si les investisseurs savent que les titres sont surévalués parce que ce sont les optimistes qui dominent dans le processus de formation des prix, ils devraient avoir tendance à ajuster à la baisse leur évaluation du titre.

La première section présentera une revue de la littérature qui entoure le sujet discuté dans ce mémoire et qui ajoutera à la compréhension de l'évolution des diverses recherches au fil du temps. Dans la deuxième partie seront présentées la description de l'échantillon et la méthodologie utilisée. La troisième partie présentera diverses statistiques descriptives sur les données de l'échantillon. Dans la quatrième partie, les résultats des régressions linéaires simples entre le volume des transactions et le rendement futur des titres seront présentés, en incorporant des retards entre ces deux variables. La cinquième partie fera état des résultats de l'application du modèle des trois facteurs de Fama-French afin de vérifier si ce modèle offre toujours la même valeur explicative du rendement des titres à mesure que le volume des transactions augmente. La sixième et dernière partie présentera une discussion et une conclusion sur les résultats obtenus.

2. REVUE DE LITTÉRATURE

2.1. REVUE GÉNÉRALE

Deux des auteurs ayant le plus marqué le monde de la décision en contexte d'incertitude sont sans contredit Kahneman et Tversky¹. Leurs travaux ont été dans les premiers, pour ainsi dire, à expliquer et démontrer plusieurs biais et erreurs que commettent les individus en prenant leurs décisions. Un des concepts importants qu'ils ont définis a été la représentativité².

Kahneman et Tversky (1972) définissent la représentativité comme suit :

« une personne qui se fie à la représentativité évalue la probabilité d'un événement incertain, ou d'un échantillon, par : (i) : la similitude entre ses caractéristiques et celles de la population et (ii) : tient compte des faits saillants du processus de génération de cet événement ou échantillon. » (Traduction libre).

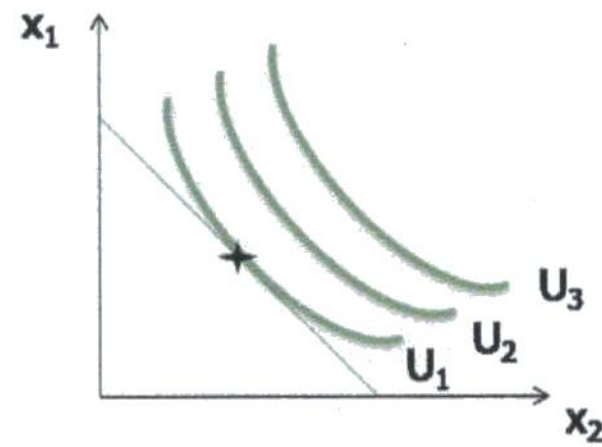
En d'autres termes, l'hypothèse de la représentativité porte sur le fait que si l'événement A est plus représentatif pour une personne que l'événement B, l'événement A se verra accorder par cette personne une plus grande probabilité d'occurrence que l'événement B.

La théorie de Kahneman et Tversky modifie, en quelque sorte, la théorie classique de l'utilité espérée. Cette dernière théorie provient de Daniel Bernoulli (1738) et a été reprise et développée par John von Neuman et Oscar Morgenstern (1944). Cette théorie classique a démontré que lorsque certains axiomes de comportements rationnels sont satisfaits, un individu qui fait face à une décision comportant un risque et des probabilités associés à chacun des scénarios possibles voudra maximiser la valeur espérée d'une certaine fonction définie pour un certain horizon temporel. Cette fonction est la « fonction d'utilité » de von Neuman et Morgenstern. Ces derniers ont démontré que si les préférences des individus satisfont quatre axiomes, il est possible de dégager une fonction d'utilité pour chacun d'eux et que celle-ci maximisera toujours l'utilité espérée. Selon cette théorie, tout individu qui ne respecte pas ces axiomes adopte un comportement irrationnel.

Le tableau suivant montre une fonction d'utilité classique par rapport à deux biens, soit x_1 et x_2 , la tangente représentant la contrainte de revenu.

¹ **Daniel Kahneman** (1934-) est un économiste américano-israélien, professeur à l'université de Princeton, lauréat du Prix Nobel d'économie en 2002 pour ses travaux fondateurs sur la théorie des perspectives, base de la finance comportementale. Il est aussi connu pour ses travaux sur l'économie du bonheur. Ses principales découvertes, sur les anomalies boursières et les biais cognitifs et émotionnels qui les causent, se sont faites en association avec Amos Tversky. Kahneman, expert en psychologie cognitive, et Tversky, expert en psychologie mathématique, développent ensemble des applications des mathématiques à la psychologie et l'économie. **Amos Tversky** (1937-1996) est un psychologue israélien. Docteur de l'université du Michigan, il enseigne à Jérusalem puis à Stanford. Il participe aux travaux sur le jugement dans l'incertitude et la théorie des perspectives pour lesquels son collègue Daniel Kahneman obtiendra, en 2002, le Prix Nobel d'économie.

² Le terme anglais est « *representativeness* ».



Les quatre axiomes sous-jacents à la théorie de von Neuman et Morgenstern sont les suivants :

Axiome 1 : Complétude

Pour deux loteries L et M , un des trois éléments suivants se vérifie nécessairement :

$$L < M, M < L, L \sim M$$

Soit M est préféré à L , soit L est préféré à M ou soit l'individu est indifférent aux deux choix possibles.

Axiome 2 : Transitivité

$$\text{Si } L < M \text{ et } M < N, \text{ alors } L < N$$

Si M est préféré à L et si N est préféré à M , alors N sera préféré à L . L'hypothèse sous-jacente à la transitivité est que la préférence est consistante pour les trois options possibles.

Axiome 3 : Continuité

Si $L < M < N$, alors il existe une probabilité $p \in [0,1]$ tel que $pL + (1 - p)N \sim M$.

L'hypothèse sous-jacente à la continuité est qu'il existe un point de renversement entre les options *mieux que* et *pire que*, en ayant un point milieu qui les sépare.

Axiome 4 : Indépendance

Si $L < M$, alors pour tout N et pour tout $p \in [0,1]$, $pL + (1 - p)N < pM + (1 - p)N$.

Le modèle développé par von Neuman et Morgenstern édicte que pour tout individu rationnel (c.-à-d. qui satisfait les quatre axiomes ci-dessus), il existe une fonction u qui attribue à chaque résultat possible A , un nombre entier $u(A)$ tel que pour n'importe quelles deux loteries :

$$L < M \text{ si et seulement si } E(u(L)) < E(u(M))$$

Où $E(u(L))$ représente la valeur espérée de l'utilité u pour L (ci-après $Eu(L)$):

$$Eu(p_1A_1 + \dots + p_nA_n) = p_1u(A_1) + \dots + p_nu(A_n)$$

Le terme u peut être déterminé en colligeant les préférences entre diverses loteries simples, soit des loteries de forme $pA + (1 - p)B$, ayant deux résultats possibles. Un individu qui maximise l'espérance d'une fonction u satisfera les quatre axiomes ci-dessus. Une telle fonction est celle communément appelée la fonction d'utilité de von Neuman et Morgenstern.

$$U(L) = \sum_{i=1}^n P_i u(A_i)$$

Pour un investisseur averse au risque, $u(A_i)$ est concave. Pour un investisseur qui est enclin à prendre des risques, $u(A_i)$ est convexe. Enfin, pour un investisseur neutre au risque, $u(A_i)$ est linéaire.

Partant de la fonction de répartition des proportions émanant de la théorie classique de l'utilité espérée de von Neuman et Morgenstern, Kahneman et Tverski (1979) ont avancé que cette fonction est modifiée par la valeur qu'accordent les individus à diverses pondérations. Ils incorporent leur théorie des perspectives à celle de l'utilité espérée en introduisant un symbole (π) représentant le facteur de changement de la pondération des probabilités qui est occasionné par le biais qu'amène la théorie des perspectives. En d'autres termes, c'est l'espérance de l'utilité de la loterie en incorporant la théorie des perspectives. S'il n'existe aucun biais dans la prise de décision des individus, le symbole (π) devrait être équivalent au chiffre un, soit qu'il n'affecte aucunement la probabilité de l'issue d'une loterie. En supposant que π de P représente la valeur que les gens accordent à une pondération³, logiquement, π de P devrait évaluer P s'il n'y avait aucun biais dans la perception que se font les individus de ce qu'est la probabilité par rapport à la probabilité réelle de l'issue d'une loterie. En prenant l'exemple des billets de loterie et de l'assurance, les individus prendront une assurance sur des biens de valeur, mais avec un risque infime qu'un incident se produise et qu'ils aient besoin de réclamer. Paradoxalement, ces mêmes individus achèteront des billets de loterie, avec une espérance infime de gagner quoi que ce soit, se soldant par une perte due au prix de leur billet. Ces comportements démontrent que les individus sont enclins au risque dans un contexte de gains espérés, mais averse au risque dans un contexte de pertes possibles.

L'étude de Kahneman et Tverski (1979) est venue, en quelque sorte, opposer ce point de vue. Ils ont posé deux questions aux mêmes 100 répondants, qui devaient répondre sur le choix qu'ils feraient dans les deux cas suivants :

1) Une probabilité de 0,001 d'obtenir un gain de 5 000 \$ ou 100 % de chances d'avoir un gain de 5 \$. 72 répondants ont choisi l'option du 5000 \$, 28 ont choisi le gain certain.

2) Une probabilité de 0,001 d'obtenir une perte de 5 000 \$ ou 100 % de chances d'avoir une perte de 5 \$. 83 répondants ont choisi l'option de la perte certaine, 17 ont choisi l'option du 5 000 \$.

Dans les deux cas, l'utilité espérée est la même, soit 5 \$ pour la première question et -5 \$ pour la deuxième question. Les résultats de cette expérience sont intéressants.

³ La lettre grecque π représente une certaine fonction qui modifiera la probabilité P .

Dans le cas de la première question, qui s'apparente à une loterie, lorsque les répondants se situent dans un contexte de gain, ils sont enclins au risque à 72 %, mais placés dans un contexte de perte, qui s'apparente à prendre de l'assurance, ils sont averses au risque à 83 %.

Si la valeur de la fonction est concave dans un contexte de gains et convexe dans un contexte de perte, les participants ont donc tendance à surévaluer les pondérations dans la zone des faibles probabilités.

Selon la fonction d'utilité classique :

Si la courbe est concave : $v(5\ 000) < 1\ 000v(5)$, ce qui implique que $\frac{v(5)}{v(5\ 000)} > 0,001$.

Selon l'étude ci-dessus sur la théorie des perspectives :

$\pi(0,001)v(5\ 000) > v(5)$, ce qui implique que $\pi(0,001) > \frac{v(5)}{v(5\ 000)} > 0,001$.

Où : π représente l'espérance de l'utilité de la loterie selon la théorie des perspectives

La même démonstration mathématique peut être effectuée avec la deuxième question, en posant l'hypothèse que la courbe d'utilité est convexe dans un contexte de perte.

Donc, cette étude démontre que $\pi(0,001) \neq 0,001$ et que les individus ont tendance à surévaluer les pondérations lorsque les probabilités sont faibles.

Dans un article subséquent, Tversky et Kahneman (1974) démontrent que pour évaluer diverses situations ou prendre une décision en contexte d'incertitude, les individus se fient à des principes heuristiques⁴, ce qui leur permet d'éviter d'avoir à évaluer un ensemble de probabilités et de simplement se fier à leur meilleur jugement de la situation. Leur article décrit trois critères heuristiques de décision qui sont employés par les individus pour évaluer les probabilités et pour prédire des valeurs. Ces trois éléments heuristiques sont la représentativité (en anglais : « *representativeness* »), la disponibilité (en anglais : « *availability* »), ainsi que l'ajustement et l'ancrage, la fixation (en anglais : « *adjustment and anchoring* »). Des différents scénarios et exemples utilisés par les auteurs pour expliquer ces éléments heuristiques, certains sont particulièrement utiles dans le contexte d'une évaluation du rendement futur ou de la valeur future des titres financiers et ils se retrouvent dans la première catégorie, soit la représentativité.

Faisant partie de la représentativité, la « conception erronée de la chance, du hasard » est une erreur fréquemment commise par les individus. Les individus s'attendent qu'une séquence d'événements aléatoires représentera les caractéristiques essentielles d'un processus, et ce, même à court terme. Dans une expérience utilisant le tir d'une pièce de monnaie, les individus considèrent qu'il est plus probable d'obtenir la séquence F(ace)-P(ile)-F-P-P-F que la séquence F-F-F-P-P-P, car cette dernière donne

⁴ Le dictionnaire Larousse définit le terme « heuristique » comme étant la « discipline qui se propose de dégager les règles de la recherche scientifique et de la découverte ». Les auteurs de l'article, quant à eux, utilisent le terme dans un contexte plus psychologique comme étant une méthode, un processus qui permet aux individus d'apprendre, de découvrir ou de comprendre les problèmes par eux-mêmes et par le truchement de leur propre expérience.

l'impression d'être moins issue du hasard. De plus, les individus considèrent que la séquence F-F-F-F-P-F pourrait être indicateur d'une pièce de monnaie truquée, plutôt que de reconnaître qu'il pourrait facilement s'agir du hasard. Une autre conséquence de ce phénomène est la tendance bien connue de l'« illusion du joueur » (en anglais « *gambler's fallacy* »). Après avoir observés que la bille s'arrête toujours sur la couleur rouge sur le jeu de la roulette dans un casino, les individus s'attendent que le prochain résultat sera noir, car la couleur noire se « doit » d'être présente dans une séquence de jeu pour que celle-ci semble plus représentative de la réalité. En se fondant sur ce jugement, les individus commettent l'erreur d'évaluer de façon erronée que la chance d'obtenir le noir est plus grande que la chance d'obtenir le rouge au prochain tir de la bille. La chance est souvent perçue comme un processus qui s'autocorrige, à savoir qu'une tendance dans un sens doit nécessairement être compensée par une tendance opposée pour rétablir l'équilibre, alors que dans les faits, les tendances dans une direction ne sont pas « corrigées » au fil des différents tirs de la bille, mais seulement diluées dans un processus plus grand.

Ce phénomène sera très important dans l'analyse des perceptions des individus sur la tendance d'un titre financier sur la bourse. En adoptant une école de pensée que l'efficience des marchés est présente, sous une forme quelconque, nul ne peut prévoir le cours du titre dans un très court laps de temps, car tous les individus ont accès à l'information en même temps⁵. Ce qui différenciera les investisseurs sera la façon dont ceux-ci interprètent cette information et comment leurs biais personnels influencent leurs décisions.

Une autre erreur fréquemment commise et faisant partie de la catégorie de la représentativité est l'« insensibilité à la prévisibilité ». Ce phénomène ressort clairement dans les études sur le processus décisionnel issu de la représentativité. Si un individu reçoit une description favorable d'un titre, cet individu aura tendance à prédire un rendement favorable pour ce titre, car la prévision qu'il en fait sera plus représentative de la description reçue. L'inverse est tout aussi vrai, une prévision de rendement médiocre d'un titre sera plus représentative d'une description médiocre de ce titre. La prévision tendra à ne pas tenir compte de la fiabilité de la description reçue ou de sa remise en question et si oui ou non cette description permet de prédire avec une certaine précision le rendement futur d'un titre. Si les individus fondent leurs prévisions en fonction de la description reçue du titre en question, ils tendront donc à être insensibles à la fiabilité des données sous-jacentes nécessaires à la formulation d'une prévision qui soit fondée et représentative de la réalité.

Cette erreur aura tendance à apparaître lorsqu'un titre en particulier devient la coqueluche de la presse financière et que toutes les analyses sont très positives et alléchantes. En vertu du phénomène précédent, les analystes et investisseurs auront tendance à se faire une image très positive d'un titre et par conséquent, leurs prévisions de bénéfices tendront à être favorablement biaisées. Il pourra donc y avoir une forme d'effet momentum pour un titre sur une certaine période, dû en grande partie par un biais favorable généralisé sur un titre, amenant une plus grande demande pour ce titre et donc un plus grand volume. Les dirigeants d'une entreprise qui auront compris ce phénomène et qui tiennent compte de la structure de rémunération des analystes

⁵ Il est fait abstraction ici du fait que certains investisseurs puissent avoir un accès privilégié à l'information. Cependant, au-delà de cette catégorie d'investisseurs, la quasi-totalité des individus aura accès à toute l'information en même temps et le prix des titres reflète à tout moment cette nouvelle information.

financiers et courtiers en valeurs mobilières pourront mousser une campagne de publicité favorable sur un titre en particulier afin de profiter du biais favorable que cela pourra créer. Cette dernière conclusion est supportée par la conclusion dégagée par Tversky et Kahneman à l'effet que la consistance d'une série de faits est déterminante pour établir le niveau de certitude qu'un individu confèrera à ses propres prévisions. Dit autrement, plus la majeure partie des analystes et la presse financière seront favorables à un titre, plus les investisseurs tendront à avoir confiance dans leurs propres prévisions, même si l'erreur de l'insensibilité à la prévisibilité décrite précédemment est commise.

Enfin, le dernier phénomène d'intérêt issu de la famille de la représentativité est la « conception erronée de la régression ». Pour expliquer ce phénomène, les auteurs ont soumis un grand nombre d'étudiants à deux versions d'un test d'aptitude de niveau de difficulté identique. Un échantillon de dix étudiants a été retenu de la tranche ayant le mieux réussi pour le premier test et l'analyse de leurs résultats a démontré qu'ils ont fait moins bonne figure au second test. L'inverse a aussi été démontré, à savoir que les dix étudiants retenus de la tranche ayant fait moins bonne figure au premier test ont atteint de meilleurs résultats au second test. Ces résultats démontrent un phénomène appelé la « régression vers la moyenne ». L'erreur qu'entraîne ce phénomène est de conclure à tort, par exemple, qu'une punition à une mauvaise performance sera bénéfique pour expliquer la meilleure performance subséquente ou que le fait de donner une rétroaction positive à une performance entraînera une surconfiance et cela sera la cause d'une moins bonne performance subséquente. Appliqué au monde de la finance et de l'investissement, ce phénomène ne doit pas biaiser les évaluations que les investisseurs se font d'un titre si celui-ci suit une trajectoire à la hausse ou à la baisse. Tant les titres à la hausse qu'à la baisse tendront à revenir vers la moyenne de rendement du marché. L'effet des phénomènes ci-dessus ne peut donc qu'être, au mieux, temporaire et les investisseurs se doivent de bien comprendre ce phénomène et ne pas conclure à tort sur les causes de la tendance haussière ou baissière d'un titre.

L'économiste David Grether (1980) a formulé certaines critiques à l'égard des travaux de recherche de Kahneman et Tversky, en invoquant que les résultats de leurs recherches pourraient difficilement s'appliquer dans le contexte d'une décision économique. Les critiques de Grether se résument comme suit :

- 1) Il est difficile de contrôler l'information donnée lorsque les descriptions sont données verbalement ou que des situations conceptualisées sont présentées à des participants. En fait, reformuler un problème mathématiquement identique peut amener des réponses différentes;
- 2) Il arrive souvent que les participants ne se fassent pas donner toute l'information véridique sur le processus aléatoire sous examen. Dès lors, les participants peuvent ne pas répondre adéquatement aux questions, ou ne pas y investir autant de sérieux dans les réponses qu'ils donneront. Il est donc important que les participants comprennent et vivent l'expérience du processus de génération des données; et
- 3) En l'absence d'un incitatif (notamment monétaire), les participants pourraient ne pas être motivés à vouloir répondre le mieux possible aux questions à l'étude.

Afin de pallier à certaines des lacunes décrites précédemment, Grether a conçu une expérience qui a servi de tremplin à plusieurs recherches subséquentes dans le domaine. Bien que son expérience ait réussi à pallier à certaines des lacunes, ou du

moins, en annuler ou à en réduire l'effet sur son expérience, elle a toutefois démontré que les conclusions générales de Kahneman et Tversky demeuraient toujours valables. L'expérience de Grether a démontré qu'effectivement, les individus tendent à partiellement ignorer l'information *a priori*. De même, les individus tendent à utiliser la représentativité dans leurs décisions, notamment en faisant des choix fondés sur les caractéristiques de l'information *a posteriori*. Ces caractéristiques tendent à être semblables aux caractéristiques des échantillons desquels provient l'information.

Selon Kahneman et Tversky (1982), avec de l'information nouvelle, les individus tendent à revoir, modifier leurs prévisions en accordant plus de poids aux événements et informations récents et en accordant moins aux événements originaux ou de départ⁶. De plus, ils ont démontré que les gens tendent à choisir une valeur prédite qui tendra à produire un résultat dont la distribution de l'ensemble de ces différents résultats reflétera l'image qu'ils se font d'une distribution fondée sur leurs impressions. Kahneman et Tversky invoquent que cette façon de faire est en contradiction avec les principes statistiques de base qui veulent que soient prises en compte les diverses considérations de prévisibilité. DeBondt et Thaler (1985) ont démontré que ces phénomènes ont une influence dans le monde de la finance et que les analystes financiers ont tendance à faire preuve de ce biais de réaction excessive.

La recherche expérimentale en psychologie suggère que les individus tendent à réagir de façon excessive à de nouvelles informations soudaines ou hors de l'ordinaire sur un phénomène quelconque, en violation du théorème de Bayes. Partant de ce principe, les auteurs DeBondt et Thaler (1985) ont effectué une recherche sur l'efficacité de marché en tentant de voir si ce phénomène a une influence sur le prix des titres financiers. Le terme « excessif » employé ici suggère une comparaison implicite avec une réaction dite « appropriée » dans les circonstances. Dans le domaine financier, une réaction appropriée en serait une qui fait appel à l'utilisation du théorème de Bayes dans le processus décisionnel faisant suite à l'obtention de nouvelles informations. DeBondt et Thaler ont construit leur recherche en se fondant sur certains principes qu'ont fait ressortir les recherches précédentes de certains auteurs. D'abord, Kleidon (1981) a démontré que le mouvement du prix des titres est fortement corrélé avec les changements dans les bénéfices de l'année à venir, ce qui confirme qu'il y a une tendance à une réaction excessive en réponse à de nouvelles informations. Ensuite Basu (1977), quant à lui, a démontré que les titres avec de faibles ratios cours/bénéfice obtiendront de plus grands rendements que les titres dont les ratios cours/bénéfice sont plus grands. Une des explications plausibles est à l'effet que les titres démontrant de faibles ratios cours/bénéfice sont temporairement sous-évalués, car les investisseurs deviennent exagérément pessimistes suite à une succession d'annonces de rendements inférieurs à ceux prévus ou suite à d'autres informations négatives. À mesure que les rendements obtenus sur ces titres seront supérieurs aux prévisions pessimistes, les prix se réajusteront. Inversement, les titres montrant de forts ratios cours/bénéfice sont vus comme surévalués et leurs prix devraient se réajuster à la baisse.

Si les prix de titres tendent à être sur ou sous-évalués, alors la tendance au renversement devrait être prévisible et ce, qu'en se fondant sur les données des rendements passés et en n'utilisant aucunement l'information comptable telle que le bénéfice par action, par exemple. DeBondt et Thaler (1985) suggèrent deux hypothèses, à savoir, en premier lieu, que les mouvements extrêmes du prix des titres feront

⁶ Il s'agit du biais de la récence des informations.

subséquentement l'objet de mouvements dans la direction opposée. En deuxième lieu, plus le mouvement initial du prix d'un titre sera important, plus le rajustement subséquent dans le prix de ce titre sera important. Ces deux hypothèses impliquent une violation de la forme faible de la théorie de l'efficacité des marchés. Les auteurs tentent de démontrer l'hypothèse que le marché réagit de façon excessive est prévisible, ce qui représente une anomalie aux postulats économiques de base que tous les investisseurs sont rationnels et prennent leurs décisions de façon rationnelle.

La plupart des tests classiques (dont Beaver et Landsman (1981)) qui tentent de mesurer la forme semi-forte de l'efficacité des marchés utilisent la méthodologie suivante. Au temps $t = 0$, généralement à un moment où un événement significatif se produit et qui a une influence sur tous les titres, par exemple une annonce des résultats trimestriels, des portefeuilles sont créés. Le bêta de chacun des portefeuilles est calculé et avec le modèle d'évaluation des actifs financiers⁷, on estime le rendement du portefeuille après une période et on le compare avec le rendement réel obtenu. Si la forme semi-forte d'efficacité des marchés fonctionne, la différence entre les deux valeurs obtenues devrait être de zéro, car le bêta devrait théoriquement incorporer la variation qui inclue toute l'information disponible aux investisseurs. Toute valeur statistiquement différente de zéro soutiendrait la thèse que les marchés ne sont pas efficaces, du moins, telle que la forme semi-forte le prescrit. Les tests de cette nature reposent sur l'hypothèse que le MÉDAF est suffisamment solide pour tout incorporer cette information et que les alphas et bêtas sont bien estimés.

DeBondt et Thaler (1985) ont modifié la méthodologie classique expliquée au paragraphe précédent. Plutôt que de former des portefeuilles au temps $t = 0$ après un événement significatif, par exemple, une annonce des bénéfices, ils bâtissent des portefeuilles au temps $t < 0$ en prenant des titres qui ont connu les plus gros gains en capital ((W) pour « *winner* ») et les plus grosses pertes en capital ((L) pour « *loser* ») pour tenter de démontrer s'il existe ou non une relation entre les rendements résiduels statistiquement différents de zéro des portefeuilles construit au temps $t < 0$ d'avec les rendements résiduels statistiquement différents de zéros à une période $t > 0$ et ce, de façon systématique. Le modèle développé par les auteurs est le suivant :

$$E(R_{jt} - E_m(R_{jt}|F_{t-1}^m)|F_{t-1}) = E(u_{jt}|F_{t-1}) = 0$$

Où :

F_{t-1} : représente l'information complète disponible au temps $t - 1$;

R_{jt} : représente le rendement du titre j au temps t ; et

$E_m(R_{jt}|F_{t-1}^m)$: représente l'espérance de R_{jt} tel que prédite par le marché, sachant l'information fournie par F_{t-1}^m .

La théorie de l'efficacité des marchés implique que $E(u_{Wt}|F_{t-1}) = E(u_{Lt}|F_{t-1}) = 0$. Les auteurs ont démontré que la relation qui est obtenue avec les portefeuilles W et L en est une dite « excessive » et les résultats obtenus pour les deux portefeuilles sont les suivants :

$$E(u_{Wt}|F_{t-1}) < 0 \text{ et } E(u_{Lt}|F_{t-1}) > 0$$

⁷ L'acronyme « MÉDAF » est le terme français du terme *CAPM*, soit « *Capital Asset Pricing Model* ».

Selon les résultats obtenus, un portefeuille de 35 titres *L* produira un rendement subséquent supérieur à celui du marché de 19,6 % en moyenne, et ce, 36 mois après la formation initiale du portefeuille. Un portefeuille de titres *W*, en revanche, produira un rendement moyen d'environ 5 % inférieur au rendement du marché. La différence entre les rendements moyens des deux portefeuilles est donc de 24,6 % avec une statistique *t* de 2,20. Les auteurs en arrivent donc aux conclusions suivantes :

- a) Il y a une asymétrie dans les résultats de cet effet de réaction excessive, à savoir qu'elle est beaucoup plus grande pour les titres *W* que pour les titres *L* ;
- b) La grande majorité du rendement excédentaire est réalisé en janvier, ce qui est en accord avec les théories des rendements saisonniers et l'effet « janvier » ; et
- c) Ce phénomène de réaction excessive devient plus prononcé durant la seconde et troisième année de la période retenue pour les analyses.

Les auteurs concluent donc que la plupart des investisseurs réagissent de façon excessive à des nouvelles financières qui sont non prévues ou qui sortent de l'ordinaire.

Selon Lee et Swaminathan (2000), il existe plusieurs écoles de pensée quant à la façon de manipuler et d'interpréter l'effet du volume des transactions. Selon ces auteurs, il est d'autant plus difficile de déterminer une relation entre le volume des transactions passées et les rendements passés, afin de pouvoir faire des prévisions sur le rendement futur des titres. Leur article tente d'établir ce lien. Dans un premier temps, ils ont démontré que l'effet du momentum⁸ des prix se renverse à long terme, soit dans un horizon de trois à cinq ans. Ce faisant, ils démontrent que le rendement des portefeuilles composés de titres en vogue, dont la valeur augmente, sera systématiquement inférieur au rendement de portefeuilles composés de titres moins populaires. Dans un deuxième temps, ils démontrent aussi que le volume des transactions passées comporte une capacité prévisionnelle sur l'ampleur et la persistance de l'effet du momentum futur. Les résultats de leur recherche démontrent que les portefeuilles composés de titres à fort volume de transactions auront un renversement plus rapide dans la tendance du momentum que les portefeuilles composés de titres avec un volume de transactions moindres.

Diether, Malloy et Scherbina (2002) ont démontré que les titres qui font l'objet de la plus grande variabilité dans les prévisions des analystes tendent à produire un rendement futur inférieur à celui que dégagent des titres similaires, mais pour lesquels les prévisions des analystes comportent une moins grande variabilité. En particulier, l'étude démontre qu'un portefeuille de titres se situant dans le plus haut quintile de dispersion des opinions produira une performance inférieure à un portefeuille se situant dans le plus bas quintile. Comme plusieurs autres auteurs, ces chercheurs postulent que le niveau de dispersion des prévisions de bénéfices que les analystes financiers estiment sur les divers titres qu'ils suivent est un bon indicateur de la divergence d'opinions entre ceux-ci. Les auteurs démontrent qu'une réduction (augmentation) de l'ampleur des prévisions des investisseurs qui détiennent le titre a comme conséquence de réduire (augmenter) les rendements futurs, tel que prédit par le modèle de Miller. Enfin, l'étude rejette l'idée que la dispersion dans les prévisions des analystes financiers puisse être

⁸ L'effet « momentum » est un phénomène qui a été démontré à l'effet que les titres ayant dégagé un rendement plus élevé durant les douze derniers mois continueront typiquement à dégager un rendement supérieur aux titres ayant connu un faible rendement dans le passé.

utilisée comme une mesure de risque. Bien que la dispersion des prévisions ait une corrélation positive avec la variabilité des rendements futurs, avec l'écart-type des rendements et le bêta du marché, il existe une relation négative très forte entre la dispersion des prévisions et les rendements futurs comme tels dans le contexte d'un modèle incorporant le risque. Ces auteurs ont produit des analyses transversales⁹ avec diverses variables en divisant certaines catégories en quintiles. Les auteurs ont testé si la constante du modèle de Fama-French (1996) était significative ou non. Pour ce faire, les auteurs estiment une équation pour chaque quantile et ils utilisent le test statistique de Gibbons, Ross et Shanken (1989), communément appelé le test « GRS », pour vérifier la nullité conjointe des constantes de chaque équation.

Un autre concept véhiculé par les chercheurs de la finance comportementale est à l'effet que plus il y aura une divergence d'opinions sur les titres, plus leur volume de transaction sera grand, ce qui se répercutera par un prix plus élevé du titre, mais cela aura une relation négative directe sur le rendement espéré du titre. Plusieurs auteurs, dont Piqueira (2006), ont remis en question l'utilisation du volume des transactions comme indicateur de la liquidité d'un titre en s'appuyant sur le fait que l'ampleur des effets du volume des transactions sur les analyses transversales du rendement des titres est trop élevée pour n'être expliquée seulement que par la variable de la liquidité. L'auteure suggère plutôt qu'un plus grand volume de transactions peut être le reflet d'une plus grande divergence d'opinions que se font les investisseurs sur la valeur actuelle et espérée d'un titre. Cette hypothèse appuie donc directement l'intuition de Miller (1977) et elle confirme l'existence d'un effet négatif fort du taux de roulement des titres sur le NASDAQ et le NYSE avec les rendements des titres.

Hong et Stein (2007) tentent de mettre l'accent sur une catégorie de modèles se fondant sur la divergence dans les opinions. Ils intègrent trois mécanismes sous-jacents à cette catégorie de modèle, soit l'obtention graduelle d'information, une attention limitée, et une information *a priori* hétérogène. Les modèles fondés sur la divergence d'opinions offrent plusieurs caractéristiques intéressantes, dont une est de permettre d'examiner directement les liens entre les mouvements de prix d'un titre et le volume des transactions. Ils démontrent que dans la plupart des marchés financiers dans le monde, le volume des transactions semble être un bon indicateur du « sentiment » des investisseurs.

Shefrin (2008), quant à lui, a développé un modèle d'évaluation des titres qui tente d'incorporer à la fois la partie fondamentale et la partie « sentiment » (divergence d'opinions). Ce modèle calcule essentiellement un arbre de décision fondé sur le théorème de Bayes, en y incorporant le biais qu'introduit la divergence d'opinions. Comme il existe de plus en plus d'évidence empirique sur l'impact de la divergence d'opinion dans l'évaluation d'un titre, le modèle de Shefrin est donc une tentative de modéliser les divers concepts introduits au fil du temps en finance comportementale.

⁹ L'expression « analyse transversale » qui sera utilisée dans ce texte est la traduction libre de l'expression anglaise « *cross-section analysis* ».

2.2. LES TRAVAUX DE MILLER

2.2.1. INTRODUCTION

Miller (1977) a été un des premiers chercheurs à introduire le concept de « divergence d'opinions », qui influencera grandement la valeur qu'accordent les investisseurs à divers titres. Il invoque que plus il y aura une divergence d'opinions quant à la valeur d'un titre, plus le titre aura un prix élevé sur le marché. Le fondement de son intuition est à l'effet que les prix au marché sont déterminés par les investisseurs les plus optimistes, et ce, même si de façon concomitante, les investisseurs pessimistes deviennent encore plus pessimistes. Une des raisons qu'il invoque pour expliquer cette assertion est l'interaction entre le niveau de divergence d'opinions et la contrainte quant aux ventes à découvert. De façon élargie, ce concept s'applique aussi à toutes les frictions¹⁰. En appliquant ces frictions aux ventes à découvert, elles empêchent donc les investisseurs plus pessimistes de pouvoir implémenter leur stratégie de vente. Ce faisant, l'évaluation que se font les investisseurs les plus optimistes se reflètera dans le prix du titre, alors que l'évaluation des plus pessimistes ne se reflètera pas dans le prix, n'ayant pu adopter une stratégie de vente à découvert sur des titres qu'ils considèrent surévalués. Ce modèle intuitif suggère que plus un titre sera surévalué, moins grand sera son rendement futur.

La finance traditionnelle considère que tous les investisseurs forment un jugement homogène de la valeur des titres. L'hypothèse fondamentale de la finance traditionnelle est à l'effet que tous les investisseurs s'attendent au même rendement espéré et à une même distribution des probabilités de ces rendements, et ce, pour tous les titres disponibles.

Miller (1977) a développé et présenté un modèle intuitif qui tente de démontrer qu'en maintenant fixe l'évaluation moyenne que font les investisseurs d'un titre, la surévaluation accordée à la valeur d'un titre augmentera avec la divergence d'opinions quant aux valeurs accordées à ce titre. Dit autrement, plus il y aura une divergence d'opinions (appelée aussi « jugement hétérogène ») quant à la valeur d'un titre, plus le titre aura un prix élevé sur le marché. L'auteur fait ressortir clairement que les prix au marché sont déterminés par les investisseurs les plus optimistes. Plus ceux-ci font preuve d'optimisme, plus les prix seront élevés, et ce, même si de façon concomitante, les investisseurs pessimistes deviennent encore plus pessimistes. Une des raisons pour expliquer cette assertion est l'interaction entre le niveau de jugement hétérogène et la contrainte quant aux ventes à découvert. Ce faisant, l'évaluation que se font les plus optimistes se reflètera dans le prix du titre, alors que l'évaluation des plus pessimistes ne se reflètera pas, n'ayant pu adopter une stratégie de vente à découvert sur des titres qu'ils considèrent surévalués.

¹⁰ Une des conditions pour qu'un marché soit dit « parfait » est qu'il n'y ait pas de coûts de transactions ou obstacles qui interfèrent avec l'offre et la demande d'un titre. Les économistes conviennent d'appeler ces coûts ou obstacles des « frictions ». Ces frictions sont considérées comme une barrière psychologique à une transaction en ayant comme conséquence que l'acheteur doit payer plus et que le vendeur doit recevoir moins. Les commissions des courtiers, les écarts bid-ask, les divers frais de transactions, les impôts sur les gains en capital, les coûts d'obtention de l'information sur un titre, les diverses restrictions sur les transactions, etc. ne sont que quelques exemples des frictions possibles.

2.2.2. DESCRIPTION DU MODÈLE

Le modèle de Miller est assez simple. Il considère d'abord qu'une obligation du gouvernement avec un terme d'un an représentant un titre sans risque. Il n'existe qu'une seule société sur le marché, donc un seul type d'action ordinaire à acheter pour un investisseur. Cette société envoie un groupe de personnes dans une expédition commerciale. À la fin d'une année, la société est liquidée et les actifs sont répartis parmi tous les actionnaires. Il s'agit d'un modèle classique de deux périodes. Après la constitution de la société, il se développe un marché pour son titre, avec comme restriction l'absence de ventes à découvert sur ce titre.

Un investisseur voulant maximiser la valeur présente de son investissement comparera le rendement espéré de la société par rapport au rendement de l'obligation. Il investira alors dans celui qui lui apparaît le plus rentable. Dans ce modèle, un investisseur est limité à l'achat d'une seule action. Étant donné l'incertitude liée à l'investissement dans le titre de la société, les investisseurs potentiels se forgent tous leur propre idée sur le rendement espéré de la société.

Le modèle¹¹ permet de tirer les conclusions suivantes :

- a) En posant l'hypothèse que tous les titres disponibles de la société sont acquis par une minorité d'investisseurs potentiels, ce qui est souvent représentatif de la réalité, le prix du titre sera plus élevé par rapport à ce qu'il aurait été, en moyenne, si l'ensemble des investisseurs représentant le marché avait eu l'occasion de participer à son évaluation en offrant un prix pour l'acheter;
- b) En maintenant l'hypothèse que tous les titres disponibles de la société sont acquis par une minorité d'investisseurs potentiels, une plus grande divergence d'opinions aura comme conséquence de faire augmenter la valeur du titre. L'inverse est aussi vrai, une réduction de la divergence d'opinions aura comme conséquence de faire diminuer la valeur du titre;
- c) Si les investisseurs potentiels fondent leurs prévisions de la distribution statistique de la moyenne prévue du rendement d'autres investissements représentatifs et investissent en se fondant sur la théorie moderne de portefeuille (modèles se fondant sur la relation risque-rendement), le prix du titre sur le marché va excéder ce qu'un investisseur moyen ayant accès à une information parfaite sur la distribution de probabilités des rendements sera prêt à payer pour ce titre; et
- d) Les décisions d'investissements prises sur les marchés financiers reflèteront l'opinion des investisseurs optimistes. Plus le nombre d'individus à décider d'investir ou non dans un titre sera grand, plus le prix reflétera la valeur que lui accorde un investisseur représentatif. À l'inverse, plus ce nombre d'individus sera réduit, plus le prix du titre tendra à s'éloigner de la valeur moyenne accordée par les investisseurs représentatifs du marché en général.

2.2.3. L'EXCÈS SPÉCULATIF ET L'HYPOTHÈSE DE L'EFFICIENCE DES MARCHÉS

Les excès spéculatifs se produisent lorsque la valeur d'un titre atteint des sommets anormaux (avec le bénéfice de la rétrospective). Étant donné le faible niveau de ventes

¹¹ Le lecteur est prié de consulter le graphique intitulé « Figure 1 » à la page 1152 de l'article de Miller. Le but du présent résumé n'est que de faire ressortir les conclusions importantes du modèle.

à découvert, un petit groupe d'investisseurs mal informés, ou se montrant très optimistes à l'égard d'un titre peuvent en venir à créer un excès spéculatif sur ce titre. Un marché comportant un grand nombre d'investisseurs bien informés risque d'avoir moins de titres sous- ou surévalués. Cependant, si ces investisseurs ne peuvent ou ne veulent faire des ventes à découvert, certains titres risqueront alors d'être surévalués.

Un petit groupe d'investisseurs très optimiste sur un titre tendra aussi à utiliser l'effet de levier pour accroître son rendement par l'utilisation des bons de souscription d'actions¹² (en anglais « *warrants* »).

2.2.4. RISQUE

Le prix d'un titre tend à être plus élevé en présence d'une grande divergence d'opinions sur le rendement espéré de celui-ci. En pratique, l'incertitude, la divergence d'opinions et le risque sont des concepts interreliés.

Le risque est lorsque le futur est inconnu, alors que la distribution de probabilités l'est. L'incertitude, quant à elle, survient lorsque la distribution de probabilités elle-même est inconnue. En pratique, peu d'importance est accordée à la distinction des deux termes et ils sont utilisés de façon interchangeable.

De façon générale, lorsque plusieurs événements peuvent avoir une conséquence sur le rendement d'un titre, tant à la hausse qu'à la baisse (risque), il y aura diverses évaluations des conséquences qu'auront les divers événements et il y aura alors présence d'une divergence d'opinions. Par conséquent, pour des investisseurs risque-neutres, la prévision du prix du titre sur le marché augmentera en fonction du risque.

Il a été démontré par Soldofsky et Miller (1969), de même que Pratt (1971) que les catégories de titres les plus risqués avaient des rendements subséquents plus bas que les catégories moins risquées de titres. Cela est contraire à la théorie voulant que les investisseurs soient indemnisés pour le risque additionnel qu'ils prennent en achetant un titre, c'est-à-dire en exigeant un rendement additionnel qui incorpore une prime pour le risque. Une explication à ce phénomène est que les titres les plus risqués sont aussi ceux qui suscitent la plus grande divergence d'opinions.

2.2.5. CHANGEMENT DANS LE RISQUE ET NOUVELLES ÉMISSIONS DE TITRES

Un changement dans la divergence d'opinions quant à un titre en particulier devrait entraîner un changement dans le prix de ce titre. Si un titre risqué devient moins risqué avec le temps, son prix devrait diminuer. Il y a raison de croire que les titres qui font l'objet de la plus grande divergence d'opinions tendent à voir cette divergence d'opinions se réduire avec le temps. Ceci est dû au fait qu'avec le passage du temps, l'incertitude se dissipe et le prix de ces titres tend à diminuer et tendre vers la valeur moyenne accordée par l'ensemble du marché. Cela est une raison additionnelle pour expliquer que le rendement des titres risqués soit inférieur au rendement des titres moins risqués.

Un autre aspect à considérer est le fait que la plupart du temps, le prix des nouvelles émissions d'actions n'est pas établi par les investisseurs sur le marché, mais plutôt par

¹² Cet article ayant été écrit en 1977, le marché des options était beaucoup moins développé. Dès lors, les bons de souscription étaient, pour ainsi dire, le précurseur ou la première génération des options.

un petit groupe d'investisseurs qui voient en cette émission une valeur suffisamment grande pour qu'ils en fassent un achat important. La divergence d'opinions est donc à son maximum au moment de l'émission du titre comme tel. S'il s'agit d'une nouvelle société en bourse, n'ayant peu ou pas d'historique de ses bénéfices, l'incertitude sera grande sur sa capacité à se tailler une place, à croître et à développer de nouveaux produits. Avec le passage du temps, elle acquiert de l'expérience et les investisseurs peuvent alors analyser le passé de cette nouvelle société et peuvent se forger une meilleure idée de la valeur à laquelle le titre devrait se transiger. Au fur et à mesure que le temps avance et que l'incertitude se dissipe, la valeur du titre déterminée par un petit groupe d'investisseurs optimistes tendra à diminuer, et ce, même si la valeur déterminée par l'investisseur moyen reste inchangée. Si le prix d'une première émission d'un titre est déterminé par le souscripteur, ce dernier risque d'établir un prix selon ses propres prévisions et risque d'être inférieur au prix que serait prêt à payer un petit groupe d'investisseurs optimistes qui eux, forment essentiellement le marché qui achètera une nouvelle émission du titre. Cela représente une explication partielle pour la sous-évaluation du prix des nouvelles émissions de titre.

2.2.6. RISQUE SYSTÉMATIQUE

Comme les échanges de titres sur le marché se font essentiellement entre investisseurs pessimistes et optimistes quant à un titre, le volume des transactions est donc une mesure de la diversité des opinions quant aux titres transigés. Cooley et Roenfeldt (1975) ont démontré qu'un volume élevé de transactions sur un titre était suivi par des rendements plus bas sur ce titre. Il existe aussi une corrélation positive entre le volume et le bêta d'un titre, suggérant qu'il y ait effectivement plus de diversité dans les opinions sur des titres présentant un bêta élevé que ceux présentant un bêta moins élevé.

Normalement, le risque systématique et l'incertitude se produisent ensemble et si les investisseurs prennent leurs décisions sur leurs meilleures prévisions du rendement espéré et du risque estimé, le prix des titres sera plus élevé et leur rendement s'en trouvera diminué. Il s'ensuit logiquement que le rendement de titres risqués (titres avec un bêta élevé) sera réduit à un montant inférieur au rendement qui serait prédit par le modèle d'évaluation des actifs financiers (en anglais : « *capital asset pricing model (CAPM)* »). Plus précisément, les titres les plus risqués produiront un rendement qui sera inférieur à la droite de marché (en anglais : « *capital market line (CML)* », qui établit la relation entre le taux d'intérêt sans risque et le portefeuille de marché. En somme, la droite de marché est une relation d'équilibre entre le risque et le rendement espéré qui ne vaut que pour les portefeuilles parfaitement diversifiés.

2.2.7. BIAIS DANS LES VALEURS UTILISÉES POUR ÉTABLIR LES PRÉVISIONS

La qualité des prévisions est cruciale pour la plupart des théories sur les marchés financiers. Les investisseurs tendent à utiliser les rendements passés des titres comme guide pour établir les prévisions des rendements futurs. Le présent article tend à démontrer que les prix ne reflètent pas les prévisions des investisseurs moyens, mais bien d'une minorité qui achète un titre en particulier. Par conséquent, les rendements subséquents ne peuvent servir de guide pour estimer les rendements futurs d'un titre.

2.2.8. UTILISATION D'UNE PRIME DE RISQUE DANS L'ÉVALUATION DES TITRES

Les investisseurs privés pourront utiliser des primes de risque parce qu'ils sont dans un marché compétitif avec d'autres investisseurs avides de dénicher les titres les plus lucratifs. Plus l'incertitude sera grande, plus grande sera la divergence d'opinions et, par conséquent, les rendements subséquents seront plus bas à la suite de l'évaluation que se fera chaque investisseur, fondée sur ses propres prévisions personnelles.

Toutefois, une différence importante entre un investissement effectué par des investisseurs publics¹³ et celui effectué par des investisseurs privés¹⁴ réside dans le fait qu'un investisseur privé peut vendre ses titres à un moment où il a besoin de liquidités, et ce, sans aucune contrainte, alors qu'un investisseur public pourrait ne pas être capable de vendre ces titres aussi rapidement, car il est régi par des objectifs de placement précis et par des contraintes pour ce faire. Par conséquent, les investissements provenant d'investisseurs privés comportent une certaine liquidité qui en réduit le risque, alors que pour les investissements provenant d'investisseurs publics ne peuvent offrir le même niveau de liquidité. Un investisseur rationnel exigera un rendement moindre pour un investissent privé, alors qu'il en exigerait un plus élevé pour un investissement public moins liquide.

2.2.9. BAISSÉ DE LA RIGUEUR DES HYPOTHÈSES DU MODÈLE

La plupart des modèles adoptent l'hypothèse que les investisseurs sont neutres face au risque. Cette hypothèse peut être modifiée sans porter atteinte à une autre hypothèse fondamentale, soit celle que les investisseurs utilisent les principes établis par Markowitz, Sharpe et Lintner et qui incorpore l'aversion au risque. Parce que l'incertitude est une source à la fois de risque et de divergence d'opinions, d'incorporer ce dernier concept dans un modèle qui limite, voire empêche toute vente à découvert, aura pour effet de réduire la prime pour le risque sur les titres les plus risqués et pourrait même rendre cette prime négative.

Si la demande est inférieure à l'offre pour un titre en particulier sur le marché, son prix diminuera. Au fur et à mesure que son prix va diminuer, le titre deviendra intéressant pour de plus en plus d'investisseurs, normalement, ceux qui font le plus preuve d'optimisme quant à son rendement espéré et à son risque. En maintenant constant le prix de tous les autres titres sur le marché, il y aura donc un plus grand nombre d'investisseurs détenant ce titre en particulier, le prix devrait diminuer jusqu'à ce que les investisseurs détiennent la quantité émise et en circulation du titre. Le prix d'équilibre sera atteint même si un petit nombre d'investisseurs détiennent de ce titre en particulier. Dit autrement, même avec un petit nombre d'investisseurs, le prix d'équilibre peut être atteint. Par conséquent et selon les principes discutés précédemment, dans cette situation en particulier, le prix du titre sera supérieur à la valeur qu'auraient octroyée les investisseurs moyens, représentatifs du marché. Le prix du titre au moment où toute l'émission sera achetée par les investisseurs augmentera plus il y aura de la divergence d'opinions quant au rendement espéré. Le titre se retrouvera sur la frontière efficiente

¹³ S'entend d'un investisseur ou d'une firme d'investissement qui agit pour le compte de clients et qui est régi par des règles et des contraintes de placement. À ce titre, une firme ne pourrait alors vendre sur-le-champ un titre si une contrainte l'en empêchait, d'où une moins grande liquidité.

¹⁴ S'entend d'un investisseur qui peut acheter et vendre un titre sur un marché, et ce, sans aucune contrainte. De plus, cet investisseur n'a que ses propres objectifs de placement à tenir compte.

des investisseurs optimistes, alors qu'à ce prix, il se retrouvera à l'intérieur de la frontière efficiente de l'investisseur moyen, représentatif du marché.

2.2.10. VENTE À DÉCOUVERT

Permettre la vente à découvert revient à accroître l'offre d'un titre sur le marché d'un montant équivalent à la somme de toutes les ventes à découvert sur le marché. Ce faisant, le prix du titre tendra à diminuer en augmentant le nombre d'investisseurs le possédant¹⁵. Cela aura pour effet d'accroître la divergence d'opinions concernant ce titre. Du fait que les investisseurs pessimistes tendront à augmenter avec l'augmentation de la divergence d'opinions sur un titre, les ventes à découvert auront comme effet de diminuer la tendance à la hausse de titres plus risqués. Il en est ainsi, car les ventes à découvert peuvent seulement faire diminuer la valeur d'un titre, car celles-ci produiront un profit seulement si la diminution du prix du titre est supérieure au montant des dividendes que celui qui a effectué la vente à découvert doit au « prêteur » du titre¹⁶.

Il est impossible de faire un profit sur la vente à découvert d'un titre qui est surévalué, car ce titre aura un rendement plus faible, si ce rendement est positif¹⁷. La raison en est que le vendeur à découvert ne reçoit pas le produit de disposition du titre, pas plus qu'il ne reçoit un intérêt sur le montant de la vente à découvert. Le produit de la vente à découvert est déposé dans un compte ne portant pas intérêt appartenant à la personne (ou entité légale) qui prête le titre, lui procurant une forme de dépôt de sécurité quant au risque relié au retour du titre prêté, ainsi que lui procurant une forme de rémunération sur le prêt en soi. Cette imperfection du marché signifie que même un très grand éventail d'investisseurs pouvant effectuer des ventes à découvert ne seront pas en mesure de s'assurer d'un marché efficient en présence d'attentes hétérogènes quant à ce marché. Cette dernière explication, y incluant la présence d'un seul investisseur mal informé, est suffisante pour annuler l'hypothèse que le marché suive une marche aléatoire.

La vente à découvert n'influence pas le modèle d'évaluation des actifs financiers (« MÉDAF », ou « CAPM » en anglais), car une des hypothèses de ce modèle est que les attentes des investisseurs sont homogènes et que chaque investisseur détient une proportion identique de tous les titres émis sur le marché. Les investisseurs averses au risque ne souhaiteront détenir qu'une quantité positive de titres. Étant donné que les investisseurs averses au risque ne voudront prendre une position à découvert, même si elle est offerte à eux, un modèle avec attentes homogènes, mais comportant une restriction pour la vente à découvert donnera exactement les mêmes résultats qu'un modèle ne comportant aucune restriction sur les ventes à découvert, en tenant pour acquis que les investisseurs sont averses ou neutres face au risque, puisque tous les

¹⁵ Le lecteur se souviendra que le modèle original proposé par l'auteur de cet article veut qu'un investisseur ne puisse posséder plus d'une action. Dès lors, c'est le nombre d'investisseurs qui augmentera, plutôt que le nombre d'actions possédées par investisseur.

¹⁶ Il est important de noter que le prêteur du titre n'a nullement connaissance si le titre qu'il possède sur le marché a été vendu à découvert. Ce même principe est suivi par les banques pour tout dépôt d'argent par un client, cet argent peut alors être prêté de nouveau par la banque sans que le client en sache quoi que ce soit. Ce qui est important, c'est que le client, ou le propriétaire légal du titre puisse obtenir son argent de la banque ou du courtier à la suite de la vente de son titre à un moment de son choix.

¹⁷ C'est-à-dire que sa valeur augmente plutôt que descendre. Il est à noter qu'un profit est réalisé sur un titre lorsque l'investisseur l'ayant vendu à découvert doit le racheter sur le marché pour un prix inférieur au prix qu'il a obtenu lors de la vente à découvert. Une personne qui vend à découvert s'attend donc à une baisse de prix du titre.

investisseurs détiendront la même proportion (positive) de chacun des titres. Si toutefois certains des investisseurs sont enclins au risque, ils pourraient se concentrer davantage sur les titres risqués et vendre à découvert certains titres moins risqués.

Toutefois, les propos du paragraphe précédent pourraient ne pas s'avérer vrais si les attentes ne sont pas homogènes. En l'absence de vente à découvert, le prix d'un titre tendra à augmenter plus il y aura divergence d'opinions. La présence possible de ventes à découvert pourrait avoir comme effet d'accroître le volume du titre en circulation et ainsi faire pression sur celui-ci pour ramener son prix à celui qui est représentatif de la majorité des investisseurs sur le marché.

Lintner (1969) a démontré qu'un modèle avec des attentes hétérogènes et aucune restriction sur la vente à découvert confèrera un résultat similaire à un modèle traditionnel utilisant le MÉDAF. En théorie, il n'y a aucune limite au nombre de titres qui peuvent être créés par la vente à découvert. Toutefois, dans un monde réel, il y a des limites aux ventes à découvert et le nombre de positions courtes représente généralement une fraction infime du nombre total d'actions en circulation. Étant donné que les titres créés par les ventes à découvert entrent directement en compétition avec les titres réellement émis par les sociétés, des restrictions sur les ventes à découvert pourraient être très utiles pour aider les sociétés à solliciter du financement par actions. Cela pourrait avoir comme contrepartie que le titre pourrait fluctuer de façon plus prononcée et les prix reflèteraient moins la réalité pour le capital de risque.

2.2.11. VISIBILITÉ D'UN TITRE

Le modèle de base développé par Miller se fonde sur l'hypothèse que chaque investisseur procède à l'analyse de tous les titres disponibles et choisit ceux qui contribueront le plus à son portefeuille. De façon réaliste, il y a plus de titres disponibles sur le marché que le nombre de titres que l'investisseur moyen peut analyser. La probabilité qu'un titre en particulier puisse susciter l'intérêt de l'investisseur moyen est la probabilité qu'il s'adonne à analyser ce titre dans la myriade de titres admissibles, combiné avec la probabilité conditionnelle qu'après cette analyse, il décide d'en faire l'acquisition.

Étant donné que tous les titres doivent être détenus par une personne, pour toute diminution du nombre d'investisseurs intéressés dans le titre d'une société, il doit y avoir une augmentation équivalente dans le nombre d'investisseurs intéressés par ce même titre. Il est évident que le prix du titre doit diminuer pour que plus d'investisseurs soient intéressés à l'acquérir dans leur portefeuille. Si le prix du titre se devait de diminuer à un niveau trop bas, il est évident que la presse financière commenterait, ce qui aurait comme conséquence d'attirer l'attention des investisseurs sur ce titre.

Il est aussi clair que la visibilité d'un titre d'une société qui procède à une nouvelle émission de titres sur le marché s'en trouvera augmentée plus cette société fait l'objet d'une publicité par la presse financière. À titre d'exemple, les sociétés qui développent de nouveaux produits ou une nouvelle technologie reçoivent généralement plus d'attention des médias financiers. Mais, ce sont les sociétés qui ont le plus de variabilité dans leurs données fondamentales qui écoperont de la plus grande part de l'attention des médias financiers.

La publication d'informations financières sur un titre peut amener les détenteurs actuels du titre à augmenter ou diminuer leur participation dans la société en question. Toutefois, la réticence à effectuer des ventes à découvert limite, en quelque sorte, les investisseurs ne détenant pas le titre à soit acheter le titre en question ou simplement de ne rien faire. Dû à la divergence d'opinions et aux coûts de transactions, les détenteurs du titre en question sont, en général, restreints à une minorité par rapport à l'ensemble des investisseurs. Par conséquent, on peut poser l'hypothèse que tout ce qui accroît la connaissance d'un titre pour un investisseur en fera augmenter le prix. L'inverse pourrait ne pas être aussi clair, à l'effet qu'une information négative pourrait simplement faire en sorte que les investisseurs ne détenant pas le titre ne fassent simplement rien et préfèrent le *statu quo*.

Les principes précédents se reflètent souvent sur les marchés financiers par le fait que pour les quelques journées entourant une nouvelle émission de titres, certains petits groupes tenteront d'accroître le volume en achetant et vendant un titre au même moment. En théorie, un volume élevé n'est pas un indicateur que le prix du titre augmentera, quoiqu'un fort volume de vente du titre puisse augmenter son prix, toutefois, le fait qu'il y ait un fort volume de transactions sur ce titre pourrait amener une visibilité accrue du titre, du fait que les investisseurs en feront une analyse plus poussée.

2.3. CONCLUSION

Miller conclut qu'en pratique, l'incertitude et le risque sont synonymes de divergence d'opinions. Dans un marché avec peu ou pas de ventes à découvert, la demande pour un titre en particulier viendra d'une minorité d'investisseurs qui ont des attentes très optimistes sur le titre. Étant donné que la divergence d'opinions est sujette à accroître avec le risque, il est fort possible que le rendement espéré soit plus bas pour des titres risqués, plutôt que plus élevé. Même des investisseurs neutres face au risque incorporeront des primes de risque dans leur modèle respectif d'évaluation des titres. La présence d'un grand nombre d'investisseurs bien informés diminuera la présence de titres sous-évalués, mais il pourra toujours y avoir la présence de titres surévalués, due à la présence d'une minorité d'investisseurs mal informés, ce qui contredit l'hypothèse d'efficience des marchés.

L'objectif de ce mémoire est de tenter de mesurer si l'ampleur de la divergence d'opinion sur un titre a un effet sur son rendement futur. En prenant un échantillon plus récent, il sera possible de constater si les divers phénomènes décrits dans la littérature ci-dessus se vérifient, ce qui démontrerait que les investisseurs ont pris en compte ces conclusions dans le choix de leurs titres.

3. ÉCHANTILLON ET MÉTHODOLOGIE

3.1. SOURCE DES DONNÉES ET VARIABLES

Les données sous-jacentes aux diverses variables ont été colligées à partir du site de *Standard & Poor's Capital IQ – Compustat* (ci-après « S&P Capital IQ »). Ce moteur de recherche a l'avantage de puiser des informations de différentes bases de données et de les regrouper en un seul fichier¹⁸. Afin d'effectuer les analyses contenues dans ce document, seules les données provenant de la base de données *Compustat* ont été utilisées et elles se sont limitées à tous les titres cotés à la bourse de New York, soit sur le *New York Stock Exchange* (ci-après « NYSE »). Pour chacun des titres cotés sur le NYSE, les données sous-jacentes aux variables suivantes ont été téléchargées ou calculées :

1. **Capitalisation boursière.** Il s'agit de la valeur boursière de la société cotée à une date donnée. La valeur fournie est en millions de dollars. À titre d'exemple, pour Alcoa inc. (cote boursière « AA »), au 31 mai 2016, la valeur fournie est de 12191,0445. Cette valeur devant être en millions, la valeur boursière est donc de 12 191 044 500 \$, soit 12,2 milliards de dollars.
2. **Volume.** Il s'agit du nombre d'actions qui ont changé de main durant la journée. La valeur fournie doit s'interpréter comme en millions d'actions. À titre d'exemple, la valeur pour Alcoa inc. (« AA ») donnée au 31 mai 2016 est de 21,09424. Le volume d'actions qui a été négocié pour la journée du 31 mai 2016 est donc de 21 094 424 actions.
3. **Valeur à la fermeture.** Il s'agit de la valeur du titre à la fermeture de la dernière journée du mois en question. Deux requêtes ont été effectuées. La première a été d'obtenir simplement les valeurs brutes (« *Closes* ») à la dernière journée du mois, sans aucun rajustement. La deuxième requête a été d'obtenir les valeurs ajustées des titres (« *Div_Adj_Price Close* ») à la dernière journée du mois, qui représentent la valeur du titre à la fermeture incluant l'effet des paiements de dividendes, ce qui a comme conséquence que le calcul du rendement fondé sur ces valeurs sera le rendement global du titre, soit les dividendes distribués et l'augmentation ou la diminution du prix du titre, c'est-à-dire le gain ou la perte en capital. Les données retenues pour ce document sont celles de la deuxième requête, soit la variable *Div_Adj_Price Close*. Les données comportaient quelques colonnes avec la mention « (*Invalid Identifier*) ». Une recherche plus approfondie révèle que les données pour ces titres ne sont pas disponibles. En effet, selon la base de données *Compustat*, aucune valeur ne peut être extraite. En effectuant la même recherche sur Yahoo Finance¹⁹ pour les titres ayant les valeurs manquantes, le site indique également qu'il n'est pas possible d'avoir les données historiques pour ces titres. Pour contourner ce problème, les valeurs brutes (« *Closes* ») à la fermeture pour le mois respectif ont été insérées en remplacement des données manquantes. Il est à noter que la différence entre les valeurs brutes et les valeurs ajustées est souvent minime et n'influencera pas de façon significative les conclusions des analyses effectuées. Aux fins d'identification des titres dont les valeurs brutes ont été utilisées dans le fichier

¹⁸ En ligne : https://en.wikipedia.org/wiki/S%26P_Capital_IQ.

¹⁹ En ligne : <http://finance.yahoo.com/>.

des valeurs ajustées, la série chronologique des titres dont les valeurs ajustées sont manquantes est présentée en rouge dans le chiffrier original. Les valeurs ajustées non disponibles et qui ont été remplacées par les valeurs ajustées sont :

Cote boursière	Colonne du chiffrier
CPE_PRA	WR
JGW	BBF
PEI_PRA	CAG
PSA_PRR	CED
PSA_PRT	CEF
PSA_PRU	CEG
PSA_PRV	CEH
PSA_PRW	CEI
PSA_PRX	CEJ
RBS_PRECL	CGG
RIT	CHZ
TRUP	CWK
TXTR	CXR
XUE	DGL

4. **Rendement en logarithme naturel.** Ces données n'ont pas été téléchargées de la base de données *Compustat*, mais calculées avec le chiffrier Excel. Il s'agit de la différence entre le logarithme naturel de la valeur du titre à la fin du mois en cours et le logarithme naturel de la valeur du titre à la fin du mois précédent. La transformation en logarithme naturel permet de scinder en mois le rendement des titres. L'addition des douze mois de rendement en logarithme naturel donnera la même valeur que le calcul du rendement annuel en logarithme naturel. Comme les données téléchargées de la valeur des titres commencent le 1^{er} janvier 2008, le rendement en logarithme naturel ne peut donc pas être calculé pour ce premier mois.

Une fois téléchargées de la base de données originale à un chiffrier Excel, les données de chacune des variables ci-dessus ont toutes un format standard à *S&P Capital IQ*, à savoir, chaque colonne du chiffrier Excel comporte la cote boursière du titre en question et chaque ligne représente les données à une date précise. La première colonne du chiffrier représente les dates, présentées sous forme ascendante.

En haut à gauche, dans le chiffrier, apparaissent les informations suivantes sur les données téléchargées :

start	2008-01-01
end	2016-05-31
freq	m
ascend/descend	asc
currency	USD

Les données ont une date de début du 1^{er} janvier 2008 et une date de fin du 31 mai 2015. La fréquence des données est mensuelle. Le classement est ascendant, en commençant par la date la plus éloignée en haut et en se rapprochant de la date actuelle à mesure que le chiffrier descend. Enfin, les données sont fournies en dollars américains.

3.2. MANIPULATION DES DONNÉES DANS EXCEL ET GRETL

Les logiciels utilisés pour cette recherche sont Excel et « *Gnu Regression, Econometrics and Time-series Library* » (ci-après « GRETL »²⁰). Ce logiciel économétrique est de type « *open-source* », gratuit à tous et téléchargeable dans la langue désirée. Il ne comporte aucune limite en matière de variables et en matière de nombre de données totales pouvant être analysées.

Toute la manipulation des données, ainsi que les régressions linéaires simples ont été effectuées avec Excel. Pour la régression multiple avec retards, le logiciel GRETL a été utilisé. Bien que les résultats de la régression multiple soient identiques avec Excel ou GRETL, ce dernier logiciel offre une capacité d'analyse beaucoup plus poussée des résidus et de certains tests de validation de la régression multiple. De plus, ce logiciel a l'avantage d'être offert dans plusieurs langues, dont le français.

Afin de reproduire certains tableaux contenus dans le présent document, les données ont dû être manipulées. Ces manipulations ont pu se faire avec le logiciel Excel. Pour ce faire, un nouveau fichier Excel a été créé en y copiant, pour chacune des dates mensuelles de l'échantillon, une ligne avec les entêtes (soit la mention « Date » et les cotes boursières comme telles) et les quatre lignes contenant chacune l'information suivante, et ce, dans cet ordre pour chacun des titres :

- 1) Capitalisation boursière;
- 2) Volume;
- 3) Valeur à la fermeture ajustée des dividendes;
- 4) Rendement en logarithme naturel;

Avec ce regroupement des données, chaque date mensuelle se voit attribuer toute l'information pertinente pour les quatre variables, et ce, pour chaque titre. Le fait de classer les données de cette façon permettra de trier les titres par ordre de grandeur de la capitalisation boursière, par exemple, tout en conservant toutes les données pour le mois en question pour ce titre.

Un des problèmes rencontrés était la présence des actions privilégiées dans l'échantillon. Pour un même titre, il pouvait y avoir la cote des actions dites « ordinaires », puis la ou les diverses séries d'actions privilégiées pour ce même titre. Chacune des catégories d'actions pour un même titre comportait la même valeur de capitalisation boursière. Dès lors, il devenait évident que toutes les actions privilégiées se devaient d'être éliminées, car les laisser aurait faussé indûment certains calculs, par exemple le calcul de la moyenne de capitalisation boursière (moyenne de la capitalisation boursière de tout l'échantillon divisé par le nombre total de titres).

Une fois les actions privilégiées retranchées de l'échantillon, l'examen de toutes les données recueillies a dénoté un autre problème, soit celui du manque d'information fondamentale pour certains titres. En effet, certains titres n'avaient que quelques informations, mais non les essentielles. À titre d'exemple, la capitalisation boursière et/ou le volume de certains titres pouvaient être manquants, et ce, pour toutes les dates de l'échantillon. Il devenait donc inutile de considérer ces titres dans l'échantillon. Tous les titres avec des données de base manquantes, comme la capitalisation boursière, le

²⁰ En ligne : <http://gretl.sourceforge.net/index.html>.

volume, la valeur boursière à la fin de la journée²¹ ont été éliminés de l'échantillon²². Dans le cas où la seule catégorie d'actions cotée en bourse en était une privilégiée, celle-ci n'a pas été supprimée de l'échantillon, ne procurant aucune valeur en double.

Enfin, comme l'échantillon s'étend de janvier 2008 à mai 2016, il arrive parfois que des titres nouveaux apparaissent ou d'anciens titres disparaissent de l'échantillon. Certaines sociétés feront leur apparition sur la bourse, alors que d'autres pourront faire faillite, redevenir privées, etc. Ces titres ont donc été tous conservés dans l'échantillon.

Afin de pouvoir importer ces données originales obtenues dans le logiciel GRET, le format de ces données doit être modifié. La première étape est donc de modifier le chiffrier pour rendre les données importables par GRET. D'abord, la cote de chacun des titres boursiers est originalement présentée sous la forme : « NYSE : COTE BOURSIÈRE ». Pour qu'une variable puisse être lue par GRET, la première rangée doit comporter un nom de variable valide. Pour être valide, une variable doit comporter un maximum de 31 caractères, commencer avec une lettre et ne contenir que des lettres, nombres ou le caractère du soulignement, soit « _ ». Toutes les variables ont donc été renommées afin d'enlever le mot NYSE, ainsi que les deux points. Pour les cotes ayant de l'information du type « AA.PRB » (titre AA, actions privilégiées B), la variable a été renommée comme « AA_PRB ».

Les dates étaient présentées sous la forme « 2008-01-31 », ce qui est illisible par GRET, car pour des données mensuelles, la date doit être sous le format « AAAA.MM » ou « AAAA:MM ». Chaque date a été donc été renommée sous le format « AAAA.MM ».

Enfin, chacune des cellules comportant les données du chiffrier était liée à la banque de données originales. À titre d'exemple pour une des cellules, la fonction suivante est sous-jacente à la cellule :

<code>=ciqfunctions.udf.CIQ(E\$7;"IQ_MARKETCAP";\$C10;\$B\$6)</code>
--

Sous cette forme, GRET ne peut lire la valeur associée à la formule de la cellule. Il faut donc copier le tableau de valeurs complet dans un nouveau chiffrier Excel en utilisant la fonction « Coller des valeurs » du menu d'options « Coller » d'Excel. Cette procédure s'assure de ne laisser que la valeur finale de la cellule et de supprimer tout lien avec d'autres fichiers ou base de données. Par souci de simplicité et pour faciliter le traitement des données, seules les colonnes de la date et des diverses variables ont été copiées dans le nouveau fichier Excel.

La deuxième étape est d'importer les données du nouveau fichier Excel dans le logiciel GRET. Pour ce faire, c'est la fonction « Fichier », puis « Ouvrir les données », puis « fichier usager... ». Une fois le répertoire du fichier à importer trouvé, il faut indiquer au

²¹ Le fait de ne pas avoir la valeur du titre ne permet pas de calculer le rendement de logarithme naturel, une autre donnée fondamentale.

²² Les titres présentant ce problème et qui ont été éliminés sont les suivants : ABW_PRA, AZUR, BMR_PRA, CCZ, DKM_CL, DKT, DTK, DXB, EHT, ELVY, FTB_PRA, FWF_CL, GPE_PRA, GPE_PRW, HJG_CL, HJJ_CL, HJN_CL, HJR_CL, IPL_PRD, KFI, MP_PRD, NCC_PRA, NCC_PRB, NCC_PRCL, PKJ, PLV_CL, SCE_PR, SCE_PRG, SOV_PR, USFD, WB_PRB, WB_PRCL, WB_PRCL et WCN. Certains titres avaient toutes les données de base, mais seulement les ratios « valeur aux livres-cours du titre » et/ou « rendement des capitaux propres » étaient manquants. Ceux-ci n'ont pas été éliminés de l'échantillon.

logiciel quel genre de fichier doit être importé en sélectionnant « fichier Excel (.xlsx) » en bas à droite du menu. En sélectionnant le fichier source Excel contenant toutes les données mises en forme pour GRETL, un menu apparaît, demandant « Débuter l'importation sur : ». Il faut alors spécifier la première colonne contenant des valeurs, soit celle de la date. Tous les fichiers Excel d'avant importation seront établis avec le même protocole, soit la colonne 1 comme étant les dates. Puis, il faut spécifier la ligne du début de l'importation. Il est important que cette ligne soit celle du nom des variables et non celle du début des données. Un menu apparaîtra pour indiquer s'il y a un problème dans l'importation, ou si tout s'est bien déroulé. Dans ce dernier cas, une petite ampoule allumée apparaît, signifiant que toutes les valeurs ont bel et bien été importées avec succès dans GRETL.

4. STATISTIQUES DESCRIPTIVES ET MÉTHODOLOGIE

À l'instar de l'article de Diether, Malloy et Scherbina (2002)²³, le tableau I-A a été reproduit avec les données recueillies du 1^{er} janvier 2008 au 31 mai 2016. Le tableau de l'article de Diether, Malloy et Scherbina présente un échantillon qui s'étend de janvier 1976 à décembre 2000 avec les statistiques descriptives présentées de façon triennale au 31 décembre. L'échantillon actuel étudié, quant à lui, est présenté chaque année, du 31 mai 2008 au 31 mai 2016.

Tableau I-A
Statistiques descriptives par année

Ce tableau démontre, au 31 mai de chaque année de l'échantillon, le nombre de firmes retenues qui se retrouvent dans l'échantillon, ainsi que la valeur de la capitalisation boursière moyenne (en millions) à chacune de ces dates.

Tableau I-A		
Date	Nombre de firmes	Moyenne capit. bours. (millions)
2008.05	1 829	11 784
2009.05	1 850	7 771
2010.05	1 915	8 835
2011.05	2 025	10 532
2012.05	2 107	8 875
2013.05	2 211	10 510
2014.05	2 364	11 299
2015.05	2 492	11 201
2016.05	2 565	10 022

Les statistiques descriptives colligées à partir de l'échantillon actuel comportent toutefois des différences avec celles de Diether, Malloy et Scherbina. D'abord, celles-ci ne se fondent que sur les titres transigés sur le NYSE, alors que celles de Diether, Malloy et Scherbina comportent les titres qui sont négociés sur le NYSE, l'AMEX et le NASDAQ. Pour être un titre admissible aux analyses de Diether, Malloy et Scherbina et présenté dans les divers tableaux, il doit comporter au moins une année financière de prévisions de bénéfices sur la base de données I/B/E/S, faire l'objet d'une analyse par au moins deux analystes professionnels et que le titre présente un prix de plus de 5,00 \$. Pour qu'un titre soit admissible et inclus dans l'échantillon du présent article, seule l'action dite « ordinaire », le cas échéant ou, dans le cas où il n'y a qu'une seule action privilégiée cotée en bourse pour un même titre, cette action privilégiée, a été retenue. Tel que mentionné précédemment, les actions privilégiées autres qu'ordinaires ont été éliminées de l'échantillon. De même, certaines actions où l'information était inaccessible ont aussi été éliminées. Enfin, les actions valant 5,00 \$ ou moins ont été conservées.

Le calcul de la moyenne de la valeur de la capitalisation boursière, pour chaque mois, se fonde sur le total de la capitalisation boursière de tous les titres retenus de

²³ Voir page 2119 de l'article.

l'échantillon et dont la capitalisation boursière est différente de zéro, divisé par le nombre total de titres dont la valeur boursière est différente de zéro. Cela assure donc une moyenne calculée de façon similaire et comparable à chaque mois.

Tableau I-B Statistiques descriptives par décile

Ces deux tableaux présentent les statistiques descriptives par décile fondé sur la capitalisation boursière, et ce, au 31 mai 2014 et 2016. Chacun des déciles présente le nombre total de titres qui y sont inclus, la moyenne de la capitalisation boursière, ainsi que le volume moyen des transactions pour les titres inclus dans le décile.

Titres admissibles (NYSE) - 31 mai 2014				Titres admissibles (NYSE) - 31 mai 2016			
Balises du NYSE	Nombre de titres	Moyenne capitalisation boursière (000 000)	Volume moyen (000 000)	Balises du NYSE	Nombre de titres	Moyenne capitalisation boursière (000 000)	Volume moyen (000 000)
Décile 1	2217	5 080,41	1,00	Décile 1	2400	4 302,12	1,38
Décile 2	78	61 113,29	4,01	Décile 2	92	51 095,90	4,30
Décile 3	30	100 684,24	4,13	Décile 3	29	90 305,28	5,04
Décile 4	15	147 697,71	6,70	Décile 4	14	129 000,09	5,72
Décile 5	13	190 686,69	7,82	Décile 5	11	159 532,63	12,04
Décile 6	6	232 364,20	4,33	Décile 6	10	200 371,86	17,63
Décile 7	3	274 382,03	13,47	Décile 7	5	238 572,11	14,44
Décile 8	1	316 118,88	2,92	Décile 8	1	277 984,80	30,30
Décile 9	0	0,00	0,00	Décile 9	1	309 969,62	10,10
Décile 10	1	431 713,52	11,49	Décile 10	2	358 179,42	9,87

Le tableau I-B ci-dessus présente l'échantillon des titres admissibles divisé en dix déciles, et ce, au 31 mai 2014 et au 31 mai 2016. Chacun des dix déciles comprend le nombre total de titres inclus dans celui-ci, la moyenne de la valeur de la capitalisation boursière des titres du décile, ainsi que le volume moyen de transactions sur les titres inclus dans le décile.

La méthodologie pour calculer les déciles de façon automatique pour chacun des mois est la suivante. D'abord, il s'agit de déterminer l'écart entre la valeur la plus grande de l'échantillon et la valeur la plus petite, autre que zéro, et ce, pour chacun de mois. La formule utilisée pour trouver la plus grande valeur dans la série de titres est la suivante :

=GRANDE.VALEUR(B12:CUG12;1)

Cette fonction indique à Excel de trouver la plus grande valeur dans la ligne représentant la capitalisation boursière de tous les titres²⁴, soit de B12 à CUG12 pour le mois de janvier 2008. L'argument « 1 » présenté dans la formule représente, dans la plage de la cellule retenue, la position de la valeur à renvoyer, déterminée à partir de la valeur la plus grande. Dans le cas présent, le chiffre « 1 » indique simplement à Excel de trouver la plus grande de la plage de données. À titre d'exemple, si le chiffre « 5 » était inscrit, la formule trouverait la cinquième plus grande valeur de la série, à commençant par la première valeur la plus grande.

²⁴ Les plages de valeurs utilisées dans les formules ci-dessus ne sont données qu'à titre indicatif seulement. Ces plages seront modifiées à chaque fois que les formules seront copiées au mois suivant.

La détermination de la valeur la plus petite, quant à elle, posait un problème, car pour tous les mois, il y a toujours au moins quelques titres qui ont un zéro d'inscrit comme valeur de capitalisation boursière (par exemple, un titre ayant fait faillite, ou un titre ayant fait son apparition sur le NYSE durant la période couverte par l'échantillon). La façon de résoudre ce problème est d'utiliser la formule de la plus petite valeur avec conditions, qui est la suivante :

```
=PETITE.VALEUR(B2:DHL2;NB.SI(B2:DHL2;0)+1)
```

Cette formule indique à Excel de trouver la plus petite valeur parmi toutes les valeurs de la plage de données qui sont différentes de zéro.

Une fois les valeurs extrêmes calculées, excluant les zéros, la valeur la plus petite est soustraite de la plus grande et le résultat est divisé par dix, ce qui donne, pour chacun des mois, les bornes des déciles.

Pour trouver le nombre de titres inclus dans le décile, la fonction suivante calculera le nombre de titres seulement à l'intérieur des bornes délimitant chacun des déciles respectivement.

```
=NB.SI.ENS(B12:CUG12;">"&CUN12;B12:CUG12;"<="&CUN13)
```

La fonction NB.SI.ENS représente le calcul du nombre de titres selon un ensemble de conditions. Dans le cas actuel, la formule calcule, dans la plage de référence B12:CUG12 (soit la capitalisation boursière), le nombre de titres qui se situeront dans un intervalle qui se situe entre une valeur supérieure à la borne inférieure du décile, mais inférieure ou égale à la borne supérieure du décile. Il est à noter que les signes logiques « > » ou « <= » sont calibrés pour que toutes les valeurs soient incluses dans l'addition qu'une seule fois et qu'aucune n'en soit exclue.

La moyenne des valeurs des capitalisations boursières pour chaque décile se calcule avec la formule conditionnelle de la moyenne, qui est la suivante :

```
=MOYENNE.SI.ENS(B12:CUG12;B12:CUG12;">"&CUN12;B12:CUG12;"<="&CUN13)
```

La logique de la formule du calcul de la moyenne sous conditions est quelque peu différente du calcul du nombre de titres sous conditions. À même la formule MOYENNE.SI.ENS, la première série est la série source, soit B12:CUG12, d'où les données doivent être puisées pour le calcul en tant que tel. Le reste de la formule représente l'ensemble des conditions à appliquer, avec la série précise sur laquelle la condition particulière doit être appliquée. Dans le cas présent, les conditions sont appliquées à la même série (B12:CUG12) que celle de la moyenne qui doit être calculée, mais segmentée par les bornes des déciles. En d'autres termes, la formule calcule pour chacun des déciles, la moyenne de la capitalisation boursière des titres inclus dans ce décile. Ce qui est très intéressant avec cette formule, c'est qu'elle permet de calculer la moyenne des valeurs de la capitalisation boursière, mais avec des conditions appliquées sur d'autres variables, par exemple le volume de transactions boursières. Si tel était le cas, la série précise devant la condition serait une autre plage de données que B12:CUG12, soit celle sur le volume.

Enfin, le calcul du volume moyen par décile est très similaire à celui de la moyenne des valeurs des capitalisations boursières, à la différence que la première série source d'où la moyenne doit se prendre est celle de la ligne des volumes de chacun des titres pour chacun des mois. Les séries représentant les conditions sont toujours les mêmes, soit celles des valeurs des capitalisations boursières, qui ont servi à délimiter les conditions pour déterminer les bornes des déciles chaque mois.

=MOYENNE.SI.ENS(B13:CUG13;B12:CUG12;">"&CUN12;B12:CUG12;"<="&CUN13)

Il est important de noter que lorsqu'un décile ne contient aucune valeur incluse entre les bornes, le chiffrier Excel indiquera alors « 0 » dans le total des titres et la mention « #DIV/0! » dans les cellules respectives de calcul de la moyenne des capitalisations boursières et du volume moyen. Aux fins de présentations des tableaux, toute mention « #DIV/0! » sera remplacée par zéro.

Tableau II-A Moyenne des rendements et dispersion

Ce tableau présente la moyenne des rendements de chacun des quintiles fondés sur le volume des transactions (en abscisse) et sur la capitalisation boursière (en ordonnée). En scindant en cinq quintiles fondés sur la capitalisation boursière permet d'isoler l'effet de la taille du titre de l'effet du volume des transactions. La dernière ligne de la partie sur la moyenne soustrait le quintile D5 du quintile D1. La portion du bas du tableau montre les écarts-types des rendements de chacun des quintiles.

Moyenne des rendements						
Quintile fondé sur volume	Quintile fondé sur capitalisation boursière					Tous les titres
	Petite		Grande			
	1	2	3	4	5	
D1 (petit)	0,00211860	0,00533745	0,00638456	0,00588937	0,00711976	0,00218876
D2	-0,02511713	-0,00193197	0,00959106	0,02517099	-0,01387721	-0,01240714
D3	-0,01202135	-0,03388520	-0,01148065	0,00240310	0,00845992	-0,00763618
D4	-0,01730260	-0,08852632	-0,00048186	0,00094773	0,00000000	-0,01548631
D5 (grand)	-0,04869495	0,00333828	0,00315051	-0,02345635	0,07287631	-0,00318493
D1 - D5	0,04657635	-0,00867573	-0,00953507	0,01756698	-0,07999607	0,00099617
Dispersion par rapport à la moyenne des rendements (Écart-type)						
Quintile fondé sur volume	Quintile fondé sur capitalisation boursière					Tous les titres
	Petite		Grande			
	1	2	3	4	5	
D1 (petit)	0,06013602	0,04812536	0,04054255	0,06387775	0,05832104	0,05964742
D2	0,15274053	0,10592276	0,05604559	0,08424100	0,04892113	0,11056463
D3	0,24149372	0,14624720	0,07874874	0,05831258	0,05157383	0,18213620
D4	0,18809660	0,10220169	0,11466478	0,05256602	0,00000000	0,12848227
D5 (grand)	0,23991817	0,12823190	0,08326219	0,04501252	0,05578881	0,16794751

Tableau II-B Évolution des rendements en fonction du volume

Ce tableau met en relief l'aspect visuel de l'évolution des rendements mensuels en fonction du volume des transactions, et ce, pour chacun des cinq quintiles et la moyenne de ceux-ci.

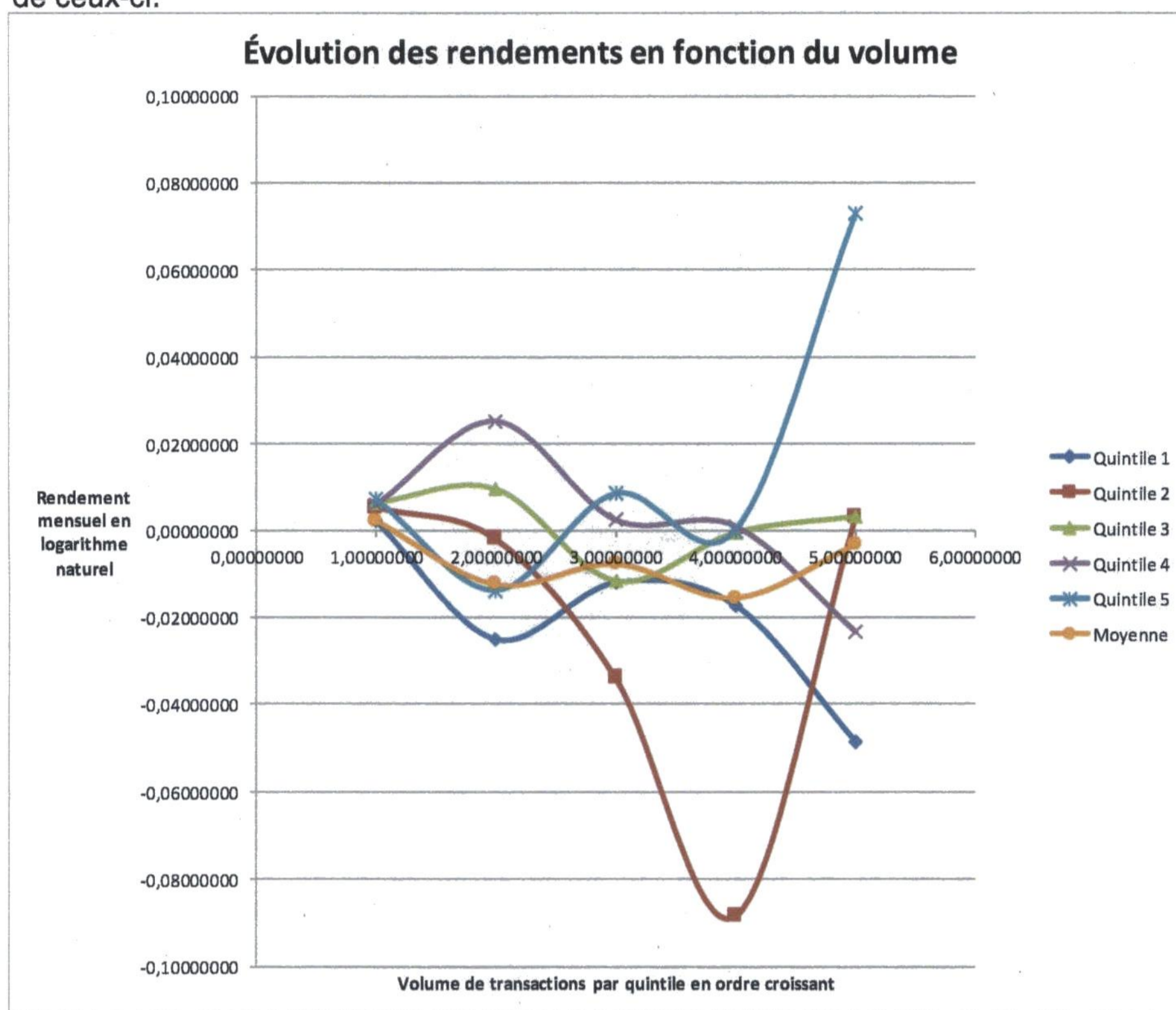
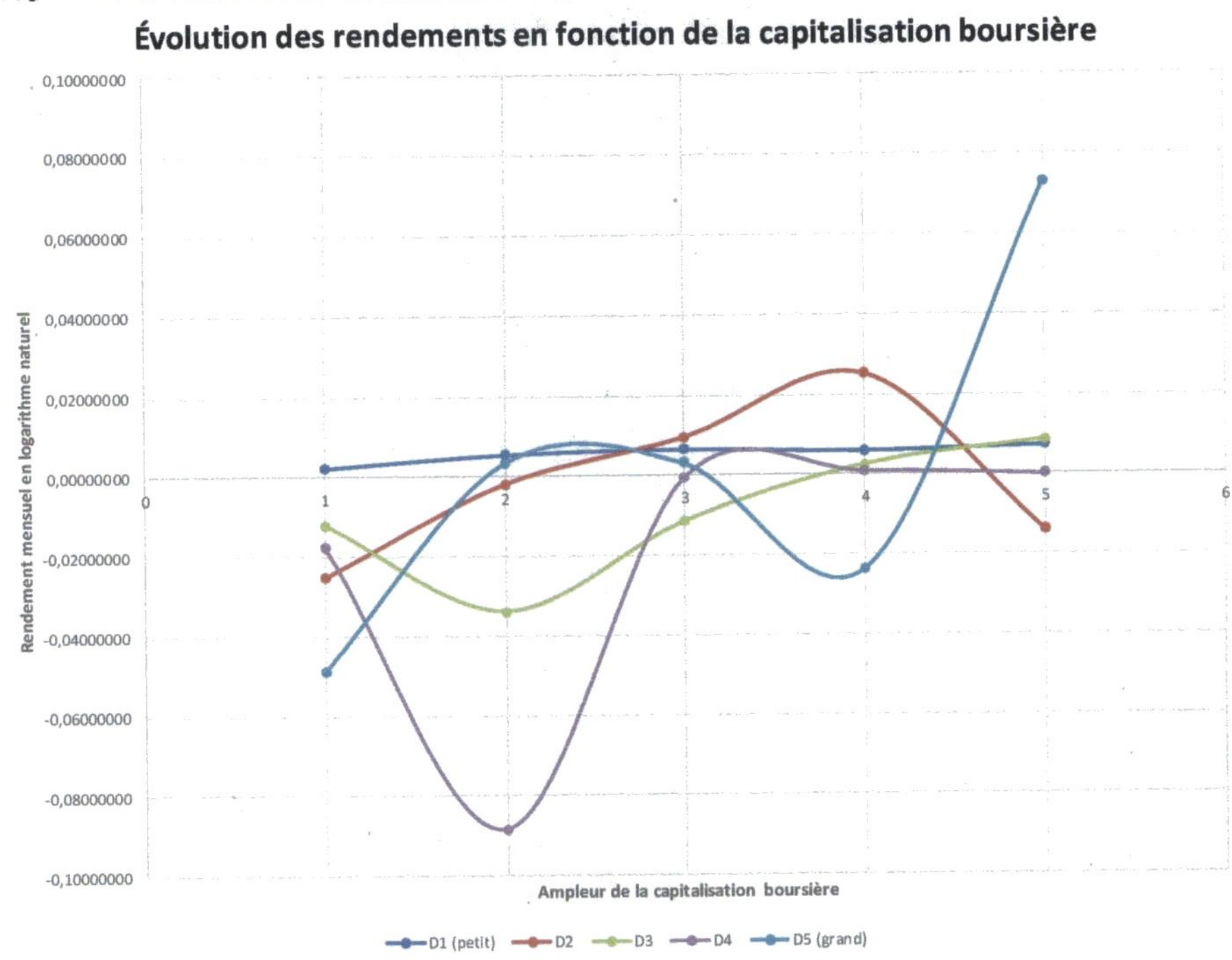


Tableau II-C

Évolution des rendements en fonction de la capitalisation boursière

Ce tableau met en relief l'aspect visuel de l'évolution des rendements mensuels en fonction de la capitalisation boursière, et ce, pour chacun des cinq quintiles et la moyenne de ceux-ci.



Le tableau II-A reprend le même format que celui de l'article de Diether, Malloy et Scherbina (2002)²⁵, à savoir qu'il présente en abscisse les quintiles fondés sur la valeur de la capitalisation boursière, du plus petit au plus grand. Il présente, en ordonnée, les quintiles fondés sur le volume, du plus petit au plus grand. Dans la portion supérieure du tableau, la moyenne des rendements de chacun des quintiles est présentée. La dernière colonne représente la moyenne de tous les titres, nonobstant le quintile auquel il appartient. La dernière ligne présente la différence entre le quintile D1 et le quintile D5.

L'objectif du tableau II-A est de donner un aperçu descriptif des différences dans les rendements en fonction de l'ampleur du volume. Le fait de scinder ces différences de rendements en quintiles fondés sur la capitalisation boursière permet de dégager l'effet de la taille sur le phénomène étudié. En d'autres mots, ce tableau permet de mettre en lumière si les rendements en fonction du volume varient aussi en fonction de la taille de l'entreprise. En théorie, si la taille de l'entreprise n'avait aucun effet sur les rendements, il ne devrait y avoir aucune différence statistiquement significative entre les quintiles fondés sur la valeur de la capitalisation boursière.

Pour calculer la moyenne des rendements à même un quintile, la formule suivante de la moyenne conditionnelle est utilisée pour chacun des mois :

```
=MOYENNE.SI.ENS(B15:CUG15;B12:CUG12;">="&CUK13;B12:CUG12;"<="&CUN12;B13:CUG13;">="&CUS13;B13:CUG13;"<="&CUV12)
```

²⁵ Voir page 2121 de l'article.

La plage de donnée B15:CUG15 représente la plage source sur laquelle la moyenne doit être calculée. Dans le cas présent, il s'agit du rendement mensuel en logarithme naturel de chacun des titres. Ensuite, quatre conditions doivent être imposées, soit les quatre bornes des deux quintiles qui s'entrecroisent. La plage B12:CUG12 pour les deux premières conditions délimite les bornes inférieure et supérieure des quintiles représentant la valeur de la capitalisation boursière. La plage B13:CUG13 pour les deux dernières conditions délimite les bornes inférieure et supérieure des quintiles représentant le volume de transactions. Chaque mois, un tableau des 25 quintiles est calculé en utilisant les formules appropriées. Ce tableau est copié pour chacun des mois de janvier 2008 à mai 2016. Une fois calculé pour chacun des mois, un tableau sommaire est construit pour insérer en ordonnée, toutes les dates de janvier 2008 à mai 2016, et ce, pour le quintile D1 seulement. En d'autres termes, ce tableau sommaire permet donc de visualiser tous les rendements pour les cinq quintiles fondés sur la valeur de la capitalisation boursière, mais que pour le quintile D1 du volume de transactions. La moyenne, la variance, ainsi que l'écart-type des rendements sont calculés pour chacun des cinq quintiles fondés sur la taille de la capitalisation boursière, et ce, en se limitant aux titres inclus dans le quintile D1 du volume des transactions boursières. La même procédure est effectuée pour les quatre autres quintiles D2 à D5. La moyenne, la variance et l'écart-type des rendements de tous les titres sont calculés en faisant abstraction des différents quintiles sur la valeur de la capitalisation boursière, mais en les séparant pour chacun des quintiles fondés sur le volume des transactions boursières. Ces données sont colligées dans un second tableau sommaire. Enfin, pour chacun des quintiles fondés sur la valeur de la capitalisation boursière, le rendement du quintile D5 est soustrait du rendement du quintile D1. Les moyennes et les écarts-types de chacun des quintiles sont reproduits dans le tableau II-A ci-dessus.

Le tableau II-B ci-dessus représente le tracé de la moyenne des rendements par quintile fondé sur le volume des transactions. Il apparaît évident, à l'œil, que plus le volume des transactions est élevé, moins grand est le rendement. Si une droite de régression se devait de traverser les points de chacun de ces quintiles, la pente serait, à l'œil, négative. Pour le dernier quintile fondé sur le volume des transactions boursières (le quintile le plus à droite sur le graphique), trois des cinq séries ont un rendement qui augmente drastiquement. Ces quintiles ne comptent que très peu de titres et cette remontée pourrait être due à toute autre raison que le volume des transactions. Les titres qui composent ce cinquième quintile ont peut-être joui d'une très grande popularité, ce qui a fait augmenter la valeur. Cependant, le faible nombre de titres inclus dans ces quintiles ne permet pas de conclure que cette remontée est représentative de l'ensemble des titres du marché.

Le tableau II-C, quant à lui, représente le tracé de la moyenne des rendements par quintile fondé sur la capitalisation boursière. Ici, l'effet est quelque peu différent du tableau II-B, en ce sens que les rendements semblent tous augmenter au fur et à mesure qu'augmente la valeur de la capitalisation boursière. Seuls les créneaux D2 et D4, au cinquième quintile (soit celui des titres ayant la plus grande capitalisation boursière sur le NYSE), subissent une diminution plutôt qu'une augmentation de rendement. L'ampleur de la capitalisation boursière semble donc avoir un effet sur le rendement des titres respectifs. Cependant, aux fins de ce document, les analyses de régressions ne se feront que pour les quintiles fondés sur le volume des transactions, car c'est ce dernier concept qui est le fondement de ce que ce document tente de démontrer.

5. RÉGRESSIONS LINÉAIRES ET MÉTHODOLOGIE

5.1. RÉGRESSION SIMPLE, AUCUN RETARD

La première partie de cette section examinera une régression linéaire simple dans laquelle la variable dépendante sera le rendement en logarithme naturel et la variable indépendante sera le volume des transactions. Il est bon de rappeler que le volume des transactions est mesuré à la dernière journée du mois en question et que le rendement est la différence entre le logarithme naturel de la valeur du titre à la fin du mois et le logarithme naturel de la valeur du titre au début du mois, découpant donc le rendement pour le mois qui vient de s'écouler. L'objectif de cette régression simple est de mesurer s'il existe une relation statistiquement significative entre l'effet du volume sur le rendement d'un titre et, le cas échéant, de mesurer l'ampleur de celle-ci et de combien le volume des transactions permet d'expliquer le rendement subséquent du titre. Comme pour toute régression, celle-ci ne fait qu'exprimer le degré de relation entre les variables étudiées, mais en aucun cas elle ne permet de déterminer la causalité du phénomène étudié. Pour ce faire, la moyenne des volumes de transactions des titres contenus dans chacun des quintiles a été calculée. À l'instar du calcul de la moyenne des rendements en logarithme naturel utilisé dans les tableaux II-A et II-B, la formule de la moyenne conditionnelle suivante a été utilisée dans le chiffrier Excel, et ce, pour chacun des mois :

```
=MOYENNE.SI.ENS(B13:CUG13;B12:CUG12;">="&CUK13;B12:CUG12;"<="&CUN12;B13:CUG13;">="&CUS13;B13:CUG13;"<="&CUV12)
```

La seule différence avec la formule utilisée pour calculer la moyenne des rendements est que la plage source (B13:CUG13), soit celle sur laquelle la moyenne est calculée en y appliquant les diverses conditions, est changée pour la série de données sur les volumes de transactions boursières, plutôt que sur les rendements en logarithme naturel.

Pour effectuer cette régression avec le chiffrier Excel, la colonne de la variable dépendante retenue est le rendement en logarithme naturel (« Rend_In ») et la colonne de la variable indépendante est le volume (« Volume »). Si les titres des variables sont inclus dans l'analyse de régression, ils doivent figurer sur la première ligne directement au-dessus de la ligne des données et la case « *Intitulé présent* », de l'option « *Régression linéaire* » du menu « *Utilitaire d'analyse* » de l'onglet « *Données* » doit être cochée afin qu'Excel reconnaisse ces valeurs comme du texte et non comme des valeurs numériques. Il est préférable de laisser les titres des variables présents, car lorsque les résultats de la régression seront présentés, les variables analysées porteront les noms exacts. Dans le cas contraire, Excel assignera un nom générique à chacune des variables analysées, ce qui pourrait porter à confusion dans l'interprétation.

Un deuxième élément est très important, soit celui des cases vides. Pour qu'Excel puisse faire une régression linéaire, aucune case ne doit être laissée en blanc. Dès lors, il y a quelques cases communes de données qui étaient vides, elles ont donc été éliminées. Ces cases vides, pour les quintiles D1 à D5, indiquent qu'il n'y a aucun titre dans le quintile qui répond aux conditions conjointes du volume de transactions et de la taille de la capitalisation boursière.

Comme le rendement en logarithme naturel ne peut être calculé pour la première date de l'échantillon, soit en janvier 2008, les données correspondantes au volume des transactions pour ce mois ont aussi été omises.

Enfin, en examinant les données recueillies pour chaque quintile, plusieurs quintiles ne comptent seulement que quelques titres qui répondent aux conditions pour pouvoir y être inclus. De plus, le fait d'effectuer une régression avec peu de titres dans un quintile pourra soit produire un résultat absurde, soit produire une régression fallacieuse, ou simplement produire une régression qui ne soit pas du tout représentative de l'ensemble du comportement des titres, parce qu'il est possible qu'il existe une explication précise du rendement d'un titre particulier qui ne soit pas nécessairement reliée avec le volume de ses transactions. Il s'agit d'une limite à cette analyse.

La régression linéaire effectuée dans cette section ne comporte aucun retard, c'est-à-dire que le volume de transactions d'un mois est comparé avec le rendement du même mois, donc cette régression n'admet aucun décalage entre le rendement du mois et le volume à la fin du même mois. L'intuition sous-jacente à cette première régression est que le marché réagit rapidement aux variations de volume et à d'autres facteurs, le cas échéant. De plus, la régression se fera en utilisant tous les titres à la fois, c'est-à-dire sans faire de distinction entre les quintiles fondés sur la valeur de la capitalisation boursière, mais en distinguant cependant les quintiles fondés sur le volume de transactions. Cette façon de procéder permettra d'avoir plus de titres par quintile fondé sur le volume. La colonne du rendement moyen mensuel pour tous les titres (sans égard à leur valeur de leur capitalisation boursière) pour le quintile D1 (fondé sur le volume des transactions) sera la variable dépendante et la colonne du volume de transactions de tous les titres pour ce même quintile D1 sera la variable indépendante. La même procédure sera répétée pour les autres quintiles fondés sur le volume des transactions. Les résultats de la régression simple obtenus pour les cinq quintiles sont les suivants :

Tableau III
Statistiques de la régression et analyse de variance – Aucun retard

Ce tableau présente les diverses statistiques de la régression et une analyse de la variance. Les coefficients de la variable « Volume », ainsi que leur statistique *t* respective sont présentés et ce, pour chacun des cinq quintiles.

	Statistiques de la régression et analyse de variance - Aucun retard				
	Quintile fondé sur le volume				
	Petit				Grand
	1	2	3	4	5
<i>Coefficients du volume</i>	(0,01068381)	0,00006578	0,00132657*	(0,00048637)	0,00002966
<i>Statistique t</i>	(0,77133410)	0,15049215	2,13366417	(1,17134541)	0,61197176
<i>Probabilité</i>	0,44236477	0,88069881	0,03618477	0,24960362	0,54197366
<i>SSR</i>	0,00212544	0,00027974	0,14419354	0,02241115	0,01063072
<i>SSE</i>	0,35009824	1,16105137	2,34382608	0,55535815	2,78179969
<i>SST</i>	0,35222368	1,16133111	2,48801962	0,57776930	2,79243040
<i>Erreur-type</i>	0,05976982	0,11113780	0,17797001	0,12780479	0,16848060
<i>R²</i>	0,00603435	0,00024088	0,05795514	0,03878910	0,00380698
<i>MSR</i>	0,00212544	0,00027974	0,14419354	0,02241115	0,01063072
<i>MSE</i>	0,00357243	0,01235161	0,03167333	0,01633406	0,02838571
<i>Statistique F</i>	0,59495629	0,02264789	4,55252278	1,37205006	0,37450944
<i>Valeur critique F</i>	0,44236477	0,88069881	0,03618477	0,24960362	0,54197366

Les quintiles D1 et D4 affichent un coefficient négatif pour la variable du volume des transactions, signifiant que le rendement a une relation inverse d'avec celui-ci. Il est difficile d'identifier une tendance dans le coefficient de la variable indépendante au fur et

à mesure que le volume des transactions augmente. Il faut toutefois porter un regard critique sur ces résultats, car ces résultats pourraient ne pas être significatifs.

Pour tester si la ou les variables (dans le cas d'une régression multiple) sont significatives, un test d'hypothèses est effectué.

H_0 : aucune relation significative entre y et x et par conséquent $\beta = 0$

H_1 : la relation est significative entre y et x et par conséquent $\beta \neq 0$

Ce test d'hypothèse se fonde sur le fait que pour qu'il y ait une relation entre la variable indépendante (le volume des transactions) et la variable dépendante (le rendement du titre), le coefficient de la variable indépendante (en l'occurrence le bêta (β)) doit être significativement différent de zéro. À l'inverse, si le coefficient de la variable indépendante n'est pas significativement différent de zéro, cela signifie qu'un changement dans le volume n'aura aucun effet sur le rendement d'un titre.

La statistique t est calculée en prenant le coefficient de la variable obtenu de la régression et en divisant celui-ci par son erreur type. La probabilité (« p -value ») associée à cette variable sert à soutenir ou non l'hypothèse que le ou les coefficients sont égaux à zéro et que la variable indépendante n'a aucun effet sur la variable dépendante. Comme l'hypothèse H_0 admet que le coefficient de la variable est de zéro, la valeur du coefficient obtenu de l'échantillon est, quant à elle, différente de zéro. La statistique t calculée (soit avec un signe positif ou négatif) indique si l'écart entre la valeur du coefficient réellement obtenue de l'échantillon et la valeur de zéro (hypothèse) est dû au fait que la variable indépendante a vraiment un effet sur la variable dépendante ou que ce résultat est dû à l'échantillon comme tel. Une faible valeur de la probabilité (< 0,05 ou 5 %) indique que l'on peut rejeter l'hypothèse nulle. En d'autres termes, l'ajout d'une variable explicative avec une faible probabilité dans le modèle sera bénéfique, car un changement dans la variable indépendante aura un effet sur la variable dépendante. À l'inverse, une grande valeur de la probabilité indique qu'un changement dans la variable indépendante n'est pas associé avec un changement dans la variable dépendante, mais plutôt dû à l'échantillon en question. Il est à noter que pour tous les tests effectués dans ce document, un niveau de confiance de 95 % sera utilisé et, par conséquent, une marge d'erreur de 5 % est acceptée.

Dans la régression ci-dessus, seul le coefficient à D3 est significatif. Pour tous les autres quintiles que D3, il est alors possible d'accepter l'hypothèse H_0 à l'effet que les coefficients sont égaux à zéro et n'ont donc aucun impact sur la variable dépendante.

Les acronymes SSR , SSE et SST sont empruntés de l'anglais et signifient respectivement « *sum of squares regression* », « *sum of squares errors* » et « *sum of squares total* ».

La variation expliquée par la régression (SSR) se calcule comme suit :

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

Il s'agit de la sommation de la différence entre les valeurs de la variable obtenues par la droite de régression et la moyenne des observations de la variable dépendante, au carré.

La variation résiduelle (*SSE*) est déterminée avec la formule suivante :

$$SSE = \sum (y_i - \hat{y}_i)^2$$

Pour analyser la variance des résultats obtenus par la régression, deux techniques sont particulièrement utiles. L'analyse de l'erreur type et du coefficient de détermination R^2 . L'erreur type (en anglais : *Standard Error*) représente la moyenne de la distance entre l'ensemble des valeurs observées et la droite de régression. Essentiellement, l'erreur type indique combien la droite de régression est représentative de l'ensemble des données en moyenne obtenue de l'échantillon. Plus la valeur est petite, meilleure sera la régression, car cela indique que les valeurs sont plus proches de la droite de régression. Contrairement au coefficient de détermination R^2 , l'erreur type obtenue peut être utilisée pour évaluer le degré de précision des prévisions qui seront effectuées. Environ 95 % des observations devraient se situer à deux erreurs types de la droite de régression, qui peut aussi s'interpréter comme étant une approximation à 95 % de l'intervalle qui servira à effectuer des prévisions.

Enfin, la somme totale des carrés (*SST*) représente l'ampleur de la variation totale des observations de la variable dépendante (rendement en logarithme naturel). Cette valeur est déterminée par la formule suivante :

$$SST = \sum (y_i - \bar{y})^2$$

Il s'agit essentiellement de la sommation de la différence entre chacune des observations de la variable pour l'échantillon et de la moyenne de celle-ci, au carré. Cette variation totale se décompose en deux, soit une variation expliquée par la régression (*SSR*) et une variation expliquée par les résidus (*SSE*), ou communément appelée variation résiduelle ou inexpliquée.

Le coefficient de détermination R^2 , quant à lui, indique la proportion de la variation totale dans la variable dépendante qui est expliquée par la droite de régression. Il s'exprime comme le rapport de la variation expliquée à la variation totale, comme suit :

$$R^2 = \frac{\text{Variation expliquée}}{\text{Variation totale}} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Ce ratio se situera entre 0 et 1 et ne compte aucune unité de mesure. Plus le ratio est proche de 1, plus la variation expliquée est proche de la variation totale et, par conséquent, plus les données observées seront proches de la droite de régression. À l'inverse, plus le ratio est proche de 0, plus les points seront dispersés de la droite de régression et moins la variation sera expliquée par le modèle.

Quelques constats sont à faire. D'abord, la valeur du R^2 peut être élevée sans toutefois que la forme linéaire soit la plus adéquate pour représenter le lien statistique entre deux variables. Un examen du diagramme de dispersion est donc très important pour

renseigner sur la forme du lien. Ensuite, le R^2 ne permet pas d'apprécier la proximité des valeurs prédites par les valeurs mesurées. Le R^2 est pertinent dans les cas où les prévisions précises sont nécessaires, alors qu'il ne peut servir à évaluer la précision.

À titre d'exemple pour le quintile D1, le R^2 calculé par Excel est de 0,0060343480. Pour le trouver, il suffit de diviser la somme des carrés de la régression par la somme totale des carrés, soit :

$$R^2 = \frac{0,00212544}{0,352223682} = 0,0060343480$$

L'interprétation à donner est que seulement 0,60343480 % de la variation totale est expliquée par la droite de régression, ce qui signifie que beaucoup de valeurs observées seront distantes de la droite de régression et que le pouvoir explicatif du modèle est très faible. En d'autres mots, cela signifie que la variable indépendante n'explique que 0,60 % du changement dans la variable dépendante, c'est-à-dire que le volume n'est pas vraiment une variable qui explique le rendement en logarithme naturel. Il appert donc que 99,40 % sont expliqués par la partie aléatoire des résidus. Il est donc évident que d'autres variables influent sur le rendement que le volume, et ce, dans un contexte où les données sont mesurées sans retard.

Pour les quintiles D1 à D5 du tableau ci-dessus, les R^2 respectifs sont très faibles. Ce qui signifie que peu importe dans quel quintile fondé sur le volume se situe le titre, le volume n'explique pas le rendement du titre pour une régression effectuée avec aucun retard.

Un autre test intéressant pour mesurer la validité du modèle est celui de la statistique F (*F-statistic*). Pour calculer celle-ci, la moyenne des carrés des erreurs (en anglais : « *mean square error* », ou « *MSE* ») et la moyenne des carrés de la régression (en anglais « *mean square regression* », ou « *MSR* ») devront être calculées. Elles se calculent respectivement comme suit :

$$MSE = \frac{SSE}{n - 2} = Se^2$$

$$MSR = \frac{SSR}{1}$$

Contrairement au SSR et SSE , les termes MSE et MSR ne peuvent s'additionner. On peut donc utiliser ces termes pour construire un test d'hypothèse identique à celui ci-dessus. Si la variable indépendante n'a aucun effet sur la variable dépendante, la variation expliquée par les résidus de la régression (SSE) sera très élevée par rapport à la variation expliquée par la régression. Cela pourrait aussi vouloir dire que d'autres variables explicatives n'ont pas été incorporées au modèle et pourraient très certainement mieux expliquer la relation entre ces dernières et la variable dépendante. La variation expliquée par la régression (SSR) et la variation expliquée par les résidus (SSE) donnent indirectement une bonne indication de l'ampleur de la relation entre la variable indépendante et la variable dépendante.

Le calcul de la statistique F s'effectue comme suit :

$$F_{(1, n-2 \text{ d.l.})} = \frac{MSR}{MSE}$$

Le critère de décision pour rejeter l'hypothèse H_0 , soit qu'il n'existe aucune relation entre la variable indépendante et la variable dépendante, est lorsque la statistique F est supérieure à la valeur critique de F .

Pour les quintiles 1, 3 et 4 du tableau ci-dessus, la valeur de la statistique F est plus grande que la valeur critique de F et donc, pour ces quintiles, nous pouvons rejeter l'hypothèse qu'il n'y a pas de relation entre le volume des transactions et le rendement d'un titre. Quant aux quintiles 2 et 5, nous pouvons admettre l'hypothèse qu'il n'y a aucune relation entre le volume et le rendement. Il est bon de se rappeler qu'à l'exception du quintile 1, le nombre de titres inclus dans les autres quintiles est assez limité, ce qui peut avoir un effet sur les résultats obtenus pour ces quintiles et l'interprétation à donner à ceux-ci.

En résumé, ce premier exercice statistique suggère une corrélation négative significative entre le volume et le rendement pour les quintiles 1 et 4, positive et significative pour le quintile 3 et non significative pour les quintiles 2 et 5. De ces résultats, il est important de se demander, intuitivement, s'il pouvait y avoir effectivement une relation entre le volume des transactions et le rendement, mais pour des périodes subséquentes. La prochaine section examinera l'effet du volume des transactions sur les rendements de périodes subséquentes.

5.2. RÉGRESSION SIMPLE, AVEC RETARDS

Il pourrait être intéressant d'analyser la même régression que celle de la section précédente, mais en examinant l'effet de retard entre la mesure du rendement du titre et le volume en question. L'intuition sous-jacente à l'utilisation de retards est qu'il est tout à fait possible que les effets du volume des transactions sur le rendement d'un titre ne se répercutent qu'après une certaine période de temps. Bien que tous les investisseurs pouvant transiger sur un titre aient l'information en même temps, les réactions de ceux-ci pourraient ne pas être instantanées.

Ces régressions se feront donc en comparant le volume des transactions d'un mois avec le rendement des titres du mois suivant et celui de deux mois suivants. Les équations des régressions et les résultats de celles-ci sont présentés dans les tableaux ci-dessous.

Tableau IV-A
Équations de la régression – Un retard

Ce tableau présente les cinq équations de régressions avec un retard d'une période. Le coefficient de la variable « Volume » diminue au fur et à mesure que celle-ci augmente.

Équations de la régression - Un retard	
Quintile fondé sur le volume	Équation
D1	Rend_In = 0,024420122 * Volume + €
D2	Rend_In = 0,001528382 * Volume + €
D3	Rend_In = 0,001582941 * Volume + €
D4	Rend_In = 0,000532492 * Volume + €
D5	Rend_In = 0,0000378671 * Volume + €

Tableau IV-B
Équations de la régression – Deux retards

Ce tableau présente les cinq équations de régressions avec un retard de deux périodes. Ici encore, le coefficient de la variable « Volume » diminue au fur et à mesure que celle-ci augmente.

Équations de la régression - Deux retards	
Quintile fondé sur le volume	Équation
D1	Rend_In = 0,034694393 * Volume + €
D2	Rend_In = 0,001463708 * Volume + €
D3	Rend_In = 0,001153827 * Volume + €
D4	Rend_In = 0,000587659 * Volume + €
D5	Rend_In = -0,00000047248 * Volume + €

Tableau V-A
Statistiques de la régression et analyse de variance – Un retard

Ce tableau présente les diverses statistiques de la régression et une analyse de la variance, et ce, avec un retard d'une période.

	Statistiques de la régression et analyse de variance - Un retard				
	Quintile fondé sur le volume				
	Petit				Grand
	1	2	3	4	5
SSR	0,01109719	0,15038169	0,20427551	0,02686314	0,01730693
SSE	0,34058097	1,01090177	2,27997449	0,53845364	2,77354980
SST	0,35167816	1,16128346	2,48425000	0,56531678	2,79085673
Erreur-type	0,05925491	0,10425887	0,17672727	0,12773714	0,16909553
R ²	0,03155496	0,12949611	0,08222824	0,04751874	0,00620130
MSR	0,01109719	0,15038169	0,20427551	0,02686314	0,01730693
MSE	0,00351114	0,01086991	0,03123253	0,01631678	0,02859330
Statistique F	3,16056263	13,83467445	6,54047330	1,64635089	0,60527940
Valeur critique F	0,07857096	0,00034126	0,01262132	0,20839473	0,43846430

Tableau V-B

Statistiques de la régression et analyse de variance – Deux retards

Ce tableau présente les diverses statistiques de la régression et une analyse de la variance, et ce, avec un retard de deux périodes.

Statistiques de la régression et analyse de variance - Deux retards					
	Quintile fondé sur le volume				
	Petit				Grand
	1	2	3	4	5
<i>SSR</i>	0,02239902	0,13780405	0,10853450	0,03260855	0,00000241
<i>SSE</i>	0,32904207	1,02347197	2,37567037	0,53011786	2,77013672
<i>SST</i>	0,35144109	1,16127602	2,48420487	0,56272641	2,77013913
<i>Erreur-type</i>	0,05854504	0,10547367	0,18164642	0,12870968	0,16986933
<i>R²</i>	0,06373477	0,11866606	0,04368983	0,05794743	0,00000087
<i>MSR</i>	0,02239902	0,13780405	0,10853450	0,03260855	0,00000241
<i>MSE</i>	0,00342752	0,01112470	0,03299542	0,01656618	0,02885559
<i>Statistique F</i>	6,53504761	12,38722006	3,28938044	1,96838045	0,00008364
<i>Valeur critique F</i>	0,01214512	0,00067422	0,07389635	0,17024864	0,99272206

L'analyse des équations de régressions est très révélatrice. Tant pour les régressions avec un retard que celles avec deux retards, les coefficients obtenus pour la variable indépendante diminuent au fur et à mesure que le volume des transactions augmente, suggérant que plus ce dernier augmente, plus le rendement obtenu diminuera. Il y a donc de toute évidence une tendance baissière du rendement au fur et à mesure que le volume des transactions augmente. En utilisant d'autres techniques mathématiques, *Diether, Malloy et Scherbina (2002)* constatent la même tendance, à la différence que les coefficients sont presque toujours négatifs, ce qui n'est pas le cas ici.

Contrairement aux résultats obtenus avec l'analyse sans retard, les R^2 des régressions avec un et deux retards sont respectivement plus élevés en général, ce qui indique clairement que la variation expliquée par le modèle s'améliore légèrement si la régression est effectuée avec un ou deux retards.

Dans le cas d'un retard, il apparaît clairement que la statistique F est supérieure à la valeur critique de F , et ce, pour tous les quintiles. Dans le cas de deux retards, la même conclusion est à faire, sauf pour le quintile 5. Cependant, la mise en garde précédente vaut encore ici, à savoir que le nombre de titres dans le cinquième quintile est très limité, ce qui limite la validité du résultat obtenu.

De l'analyse précédente se dessine une tendance, à savoir que plus le quintile du volume des transactions augmente, plus le rendement tend à diminuer. Il pourrait être intéressant de déterminer le moment où cette tendance diminue et se résorbe, afin de constater s'il existe une forme de généralité qui puisse se dégager et qui pourrait servir dans une stratégie de construction de portefeuille. En augmentant le nombre de retards, il sera possible d'analyser la tendance des résultats et l'effet sur les coefficients de la variable « volume ».

Les résultats obtenus en effectuant les analyses de régressions avec des retards de trois et de quatre mois complètent l'intuition qui commence à se dégager après les analyses avec un et deux retards. Les résultats sont présentés dans les tableaux VI-A et B et VII-A et B qui suivent.

Tableau VI-A
Équations de la régression – Trois retards

Ce tableau présente les cinq équations de régressions avec un retard de trois périodes. Le coefficient de la variable « Volume » diminue au fur et à mesure que celle-ci augmente.

Équations de la régression - Trois retards	
Quintile fondé sur le volume	Équation
D1	Rend _{ln} = 0,021254516 * Volume + E
D2	Rend _{ln} = 0,000645045 * Volume + E
D3	Rend _{ln} = 0,000481153 * Volume + E
D4	Rend _{ln} = -0,000124604 * Volume + E
D5	Rend _{ln} = -0,0000287361 * Volume + E

Tableau VI-B
Équations de la régression – Quatre retards

Ce tableau présente les cinq équations de régressions avec un retard de quatre périodes. La diminution du coefficient de la variable « Volume » semble s'arrêter dans les quintiles D2 à D4. L'effet de cette diminution s'estompe donc après quatre périodes.

Équations de la régression - Quatre retards	
Quintile fondé sur le volume	Équation
D1	Rend _{ln} = 0,005733195 * Volume + E
D2	Rend _{ln} = 0,000560568 * Volume + E
D3	Rend _{ln} = 0,000530844 * Volume + E
D4	Rend _{ln} = 0,0005522 * Volume + E
D5	Rend _{ln} = 0,0000227146 * Volume + E

Tableau VII-A
Statistiques de la régression et analyse de variance – Trois retards

Ce tableau présente les diverses statistiques de la régression et une analyse de la variance, et ce, avec un retard de trois périodes.

Statistiques de la régression et analyse de variance - Trois retards					
	Quintile fondé sur le volume				
	Petit				Grand
	1	2	3	4	5
SSR	0,00837213	0,02658927	0,01875639	0,00146337	0,00995359
SSE	0,34128414	1,12982817	2,32227638	0,54701071	2,67365943
SST	0,34965627	1,15641744	2,34103277	0,54847408	2,68361302
Erreur-type	0,05993717	0,11142573	0,18085386	0,13283639	0,16776109
R ²	0,01366962	0,01225645	(0,00595965)	(0,02950392)	(0,00677825)
MSR	0,00837213	0,02658927	0,01875639	0,00146337	0,00995359
MSE	0,00359246	0,01241569	0,03270812	0,01764551	0,02814378
Statistique F	2,33047099	2,14158513	0,57344757	0,08293173	0,35366936
Valeur critique F	0,13018608	0,14679936	0,45139805	0,77527966	0,55345711

Tableau VII-B**Statistiques de la régression et analyse de variance – Quatre retards**

Ce tableau présente les diverses statistiques de la régression et une analyse de la variance, et ce, avec un retard de quatre périodes.

Statistiques de la régression et analyse de variance - Quatre retards					
	Quintile fondé sur le volume				
	Petit				Grand
	1	2	3	4	5
<i>SSR</i>	0,00060915	0,02006776	0,02248571	0,02863681	0,00621706
<i>SSE</i>	0,34845599	1,13486277	2,31825887	0,50084555	2,67193950
<i>SST</i>	0,34906514	1,15493053	2,34074458	0,52948237	2,67815656
<i>Erreur-type</i>	0,06088496	0,11229241	0,18198347	0,12920856	0,16859682
<i>R²</i>	(0,00887464)	0,00645768	(0,00454226)	0,02255403	(0,00829221)
<i>MSR</i>	0,00060915	0,02006776	0,02248571	0,02863681	0,00621706
<i>MSE</i>	0,00370698	0,01260959	0,03311798	0,01669485	0,02842489
<i>Statistique F</i>	0,16432552	1,59146849	0,67895762	1,71530805	0,21871875
<i>Valeur critique F</i>	0,68612608	0,21037638	0,41274464	0,20023964	0,64110035

De façon sommaire, les tableaux VII-A et VII-B démontrent des R^2 faibles. Bien que la valeur de R^2 ne puisse être négative, c'est-à-dire qu'elle doit se situer entre 0 et 1, par définition, la valeur de R^2 ajustée quant à elle peut être négative. Cela est cependant un indicateur que le modèle explique faiblement les résultats obtenus. Il est à noter que ce sont toujours les valeurs des R^2 ajustés qui sont utilisées dans les tableaux du présent travail et non les R^2 ordinaires.

Pour les trois premiers quintiles du tableau VII-A, la statistique F est supérieure à la valeur critique F . Quant aux deux derniers quintiles, elle est inférieure à la valeur critique F , indiquant donc qu'on ne peut rejeter l'hypothèse qu'il n'existe aucune relation entre la variable indépendante et la variable dépendante. La même conclusion est à faire pour les quintiles un et cinq du tableau VII-B. Bien que les signes de ces variables soient tous positifs, les coefficients diminuent au fur et à mesure que le volume augmente. Ce qui suggère que les conclusions du modèle développé par Miller (1977) semblent s'appliquer ici, par le truchement de la diminution des coefficients de la variable « Volume » au fur et à mesure que celui-ci augmente.

Une analyse plus approfondie des coefficients de la variable « Volume » est cependant révélatrice. Le tableau VIII-A résume les coefficients de la variable « Volume » pour toutes les régressions effectuées. Le graphique VIII-B, quant à lui, présente visuellement ces coefficients pour toutes les régressions.

Tableau VIII-A**Coefficients de la variable « Volume »**

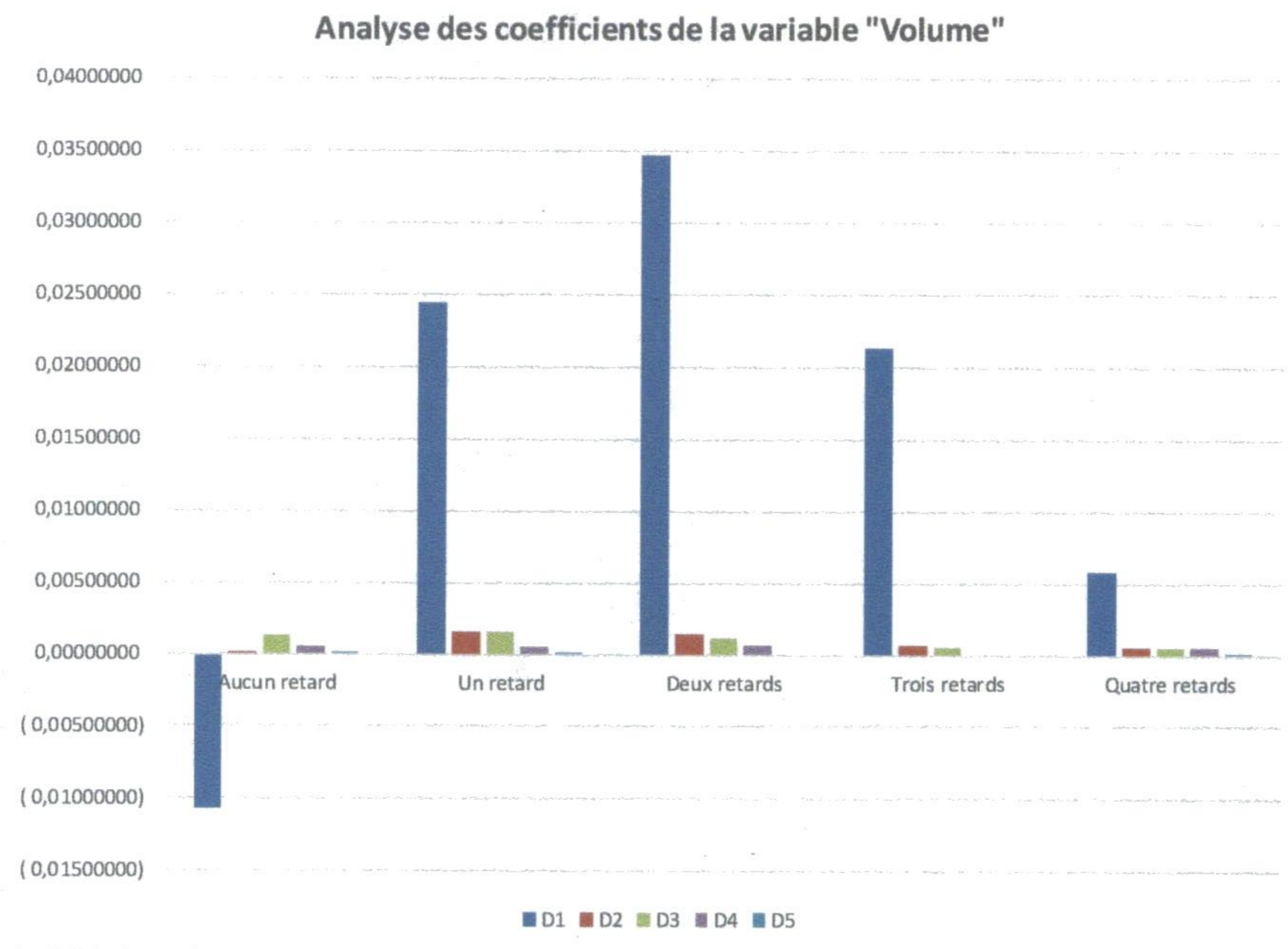
Ce tableau présente de façon comparative tous les coefficients de la variable « Volume » pour toutes les régressions simples effectuées.

Coefficients de la variable "Volume"					
	Aucun retard	Un retard	Deux retards	Trois retards	Quatre retards
D1	(0,01068381)	0,02442012	0,03469439	0,02125452	0,00573320
D2	0,00006578	0,00152838	0,00146371	0,00064505	0,00056057
D3	0,00132658	0,00158294	0,00115383	0,00048115	0,00053084
D4	0,00048637	0,00053249	0,00058766	(0,00012460)	0,00055220
D5	0,00002966	0,00003787	(0,00000047)	(0,00002874)	0,00002271

Tableau VIII-B

Évolution visuelle des coefficients de la variable « Volume »

Ce tableau présente, de façon comparative, l'évolution visuelle des coefficients de la variable « Volume » pour toutes les régressions simples effectuées.



De façon empirique, il apparaît clairement que les coefficients de la variable « volume » déclinent au fur et à mesure que le volume augmente. L'objectif des régressions précédentes était de tenter de faire ressortir qu'il existe effectivement une tendance entre le volume des transactions et le rendement futur. Intuitivement toutefois, bien que les résultats obtenus aient une certaine force explicative, ils démontrent clairement que d'autres variables indépendantes entrent en jeu et que le volume, à lui seul, n'est pas la seule explication d'un rendement qui décline avec une augmentation dans le volume des transactions.

5.3. DISCUSSION

Après l'analyse de ces régressions linéaires simples, il appert manifestement que le volume des transactions pris isolément a un effet significatif sur le rendement futur des titres. Cet effet n'est pas vraiment significatif lorsqu'il n'y a aucun retard, mais de façon générale, la situation s'améliore lorsque les régressions simples sont effectuées avec un ou deux retards et il ressort clairement qu'il y a une relation statistiquement significative entre le volume des transactions et le rendement futur. En divisant en quintiles, comme l'ont fait Diether, Malloy, et Scherbina (2002) et en prenant les titres que sur le NYSE, quelques quintiles se sont retrouvés avec moins de titres, car les critères conjoints de sélection étaient le volume des transactions et la valeur de la capitalisation boursière. Avec ces critères, certains quintiles se retrouvaient avec peu de titres.

Un autre constat à faire est que l'effet observé se résorbe en quatre périodes (quatre mois) environ. Sachant que les titres avec le plus grand volume de transactions auront un rendement qui tendra à diminuer plus rapidement, un investisseur qui souhaiterait tirer avantage de ce phénomène pourrait, dans la portion à court terme de son

portefeuille, choisir spécifiquement des titres ayant un faible volume de transactions, voire sous-évalués selon une analyse fondamentale. Après une période de quatre mois, cet investisseur pourrait disposer de ces titres et rebâtir un portefeuille de titres semblables avec les titres qui, à ce moment, montreront le plus faible volume de transactions. Cette stratégie permettrait, toute chose étant égale par ailleurs, de tirer profit de ce phénomène à court terme. Cette conclusion ne tient toutefois pas compte des frais de transactions, ainsi que de l'aspect fiscal des investisseurs. Le fait de connaître la psychologie des investisseurs qui influencent leurs décisions permet de tirer profit de tendances comme celle-ci.

5.4. RÉGRESSION MULTIPLE, AVEC RETARDS

Cette section présentera une analyse de régression multiple qui regroupera toutes les variables, soit celle sans retard, ainsi que celles comportant les quatre retards. Les résultats des régressions pour les cinq quintiles sont présentés dans les tableaux ci-dessous. Il est important de noter que les astérisques à droite des *p-value* indiquent les niveaux de confiance pour lesquels celles-ci sont significatives ou non. Un, deux ou trois astérisques indiquent respectivement que la variable est significative à un niveau de confiance de 90 %, 95 % et 99 %.

Tableau IX-A

Écran de sortie de la régression multiple du quintile D1

Ce tableau présente les résultats de la régression multiple effectuée avec les variables du quintile D1, soit celui comportant le plus faible volume de transactions parmi les cinq quintiles. La variable avec aucun retard est significative à 95 % et celle comportant deux retards l'est à 90 %.

Output régression D1:
MCO, utilisant les observations 2008:06-2016:05 (T = 96)
Variable dépendante: Rend_ln

	coefficient	erreur std.	t de Student	p. critique	
const	-0,0371818	0,0267781	-1,389	0,1684	
Vol_0retard	-0,0411624	0,0165709	-2,484	0,0148	**
Vol_1retard	0,0278372	0,0173283	1,606	0,1117	
Vol_2retards	0,0301763	0,0164687	1,832	0,0702	*
Vol_3retards	0,0234649	0,0173351	1,354	0,1793	
Vol_4retards	-0,0130979	0,0165601	-0,7909	0,4311	

Moy. var. dép.	0,001895	Éc. type var. dép.	0,060617
Somme carrés résidus	0,300000	Éc. type de régression	0,057735
R2	0,140563	R2 ajusté	0,092816
F(5, 90)	2,943940	p. critique (F)	0,016551
Log de vraisemblance	140,6614	Critère d'Akaike	-269,3228
Critère de Schwarz	-253,9367	Hannan-Quinn	-263,1035
rho	0,262750	Durbin-Watson	1,444970

Constante mise à part, la probabilité critique est la plus élevée pour la variable 6 (Vol_4retards)

Tableau IX-B Restrictions linéaires - Quintile D1

Ce tableau représente l'écran de sortie du test d'hypothèses requis pour la statistique F . Ce test pose l'hypothèse que le coefficient de chacune des variables indépendantes est égal à zéro. Une statistique F supérieure à la valeur critique de F indique le rejet de l'hypothèse de nullité des coefficients et l'acceptation qu'il y a au moins un coefficient statistiquement différent de zéro. Cette même statistique se retrouve dans le tableau IX-A ci-dessus.

Ensemble de restrictions
 1: b[Vol_0retard] = 0
 2: b[Vol_1retard] = 0
 3: b[Vol_2retards] = 0
 4: b[Vol_3retards] = 0
 5: b[Vol_4retards] = 0

Statistique de test: $F(5, 90) = 2,94394$, avec p. critique = 0,0165512

Estimations contraintes:

	coefficient	erreur std.	t de Student	p. critique
const	0,00189456	0,00618665	0,3062	0,7601
Vol_0retard	0,000000	0,000000	NA	NA
Vol_1retard	0,000000	0,000000	NA	NA
Vol_2retards	0,000000	0,000000	NA	NA
Vol_3retards	0,000000	0,000000	NA	NA
Vol_4retards	0,000000	0,000000	NA	NA

Écart type de la régression = 0,0606166

Tableau IX-C

Écran de sortie de la régression multiple du quintile D2

Ce tableau présente les résultats de la régression multiple effectuée avec les variables du quintile D2. Les variables avec aucun retard et un retard sont significatives à 99 % et celle comportant deux retards est significative à 95 %. Pour les variables comportant trois et quatre retards ne sont pas significatives.

Output régression D2:
 MCO, utilisant les observations 2008:06-2016:01 (T = 92)
 Variable dépendante: Rend_ln

	coefficient	erreur std.	t de Student	p. critique
const	-0,0664866	0,0236987	-2,805	0,0062 ***
Vol_0retard	-0,00144310	0,000513735	-2,809	0,0061 ***
Vol_1retard	0,00185068	0,000573331	3,228	0,0018 ***
Vol_2retards	0,00123064	0,000565287	2,177	0,0322 **
Vol_3retards	-0,000187433	0,000573585	-0,3268	0,7446
Vol_4retards	-0,000150336	0,000512170	-0,2935	0,7698

Moy. var. dép. -0,013526 Éc. type var. dép. 0,112657
 Somme carrés résidus 0,872022 Éc. type de régression 0,100697
 R2 0,244957 R2 ajusté 0,201059
 F(5, 86) 5,580162 p. critique (F) 0,000168
 Log de vraisemblance 83,75919 Critère d'Akaike -155,5184
 Critère de Schwarz -140,3877 Hannan-Quinn -149,4115
 rho 0,169307 Durbin-Watson 1,648242

Constante mise à part, la probabilité critique est la plus élevée pour la variable 6 (Vol_4retards)

Tableau IX-D

Écran de sortie de la régression multiple du quintile D3

Ce tableau présente les résultats de la régression multiple effectuée avec les variables du quintile D3. Selon cette régression, aucune des variables n'est significative.

Output régression D3

Fichier Édition Tests Sauvegarder Graphiques Analyse LaTeX

Output régression D3:
MCO, utilisant les observations 2008:06-2014:05 (T = 72)
Variable dépendante: Rend_ln

	coefficient	erreur std.	t de Student	p. critique	
const	-0,146279	0,0568269	-2,574	0,0123	**
Vol_0retard	0,000836486	0,000734234	1,139	0,2587	
Vol_1retard	0,00119347	0,000787648	1,515	0,1345	
Vol_2retards	0,000842107	0,000749540	1,123	0,2653	
Vol_3retards	-0,000433160	0,000786575	-0,5507	0,5837	
Vol_4retards	-0,000254348	0,000745350	-0,3412	0,7340	
Moy. var. dép.	-0,013742	Éc. type var. dép.		0,181572	
Somme carrés résidus	2,023035	Éc. type de régression		0,175077	
R2	0,135730	R2 ajusté		0,070255	
F(5, 66)	2,073010	p. critique (F)		0,079847	
Log de vraisemblance	26,43085	Critère d'Akaike		-40,86171	
Critère de Schwarz	-27,20171	Hannan-Quinn		-35,42362	
rho	0,033706	Durbin-Watson		1,906936	

Constante mise à part, la probabilité critique est la plus élevée pour la variable 6 (Vol_4retards)

Tableau IX-E

Écran de sortie de la régression multiple du quintile D4

Ce tableau présente les résultats de la régression multiple effectuée avec les variables du quintile D4. Selon cette régression, seule la variable ne comportant aucun retard est significative à 95 %. Toutes les autres sont non significatives.

Output régression D4

Fichier Édition Tests Sauvegarder Graphiques Analyse LaTeX

Output régression D4:
MCO, utilisant les observations 1-32
Variable dépendante: Rend_ln

	coefficient	erreur std.	t de Student	p. critique	
const	-0,0174563	0,0511734	-0,3411	0,7358	
Vol_0retard	-0,000986820	0,000438993	-2,248	0,0333	**
Vol_1retard	0,000677487	0,000456344	1,485	0,1497	
Vol_2retards	0,000639727	0,000463016	1,382	0,1788	
Vol_3retards	-0,000645198	0,000456366	-1,414	0,1693	
Vol_4retards	0,000516747	0,000435750	1,186	0,2464	
Moy. var. dép.	-0,003062	Éc. type var. dép.		0,130691	
Somme carrés résidus	0,375011	Éc. type de régression		0,120098	
R2	0,291741	R2 ajusté		0,155537	
F(5, 26)	2,141944	p. critique (F)		0,092059	
Log de vraisemblance	25,73855	Critère d'Akaike		-39,47710	
Critère de Schwarz	-30,68268	Hannan-Quinn		-36,56200	

Constante mise à part, la probabilité critique est la plus élevée pour la variable 6 (Vol_4retards)

Tableau IX-F

Écran de sortie de la régression multiple du quintile D5

Ce tableau présente les résultats de la régression multiple effectuée avec les variables du quintile D5, soit celui comportant le plus haut volume de transactions. Selon cette régression, aucune des variables n'est significative.

Output régression D5
MCO, utilisant les observations 2008:06-2016:05 (T = 96)
Variable dépendante: Rend_in

	coefficient	erreur std.	t de Student	p. critique
const	-0,0159857	0,0257428	-0,6210	0,5362
Vol_0retard	2,62601e-05	4,95925e-05	0,5295	0,5978
Vol_1retard	3,68383e-05	4,98387e-05	0,7392	0,4617
Vol_2retards	1,55053e-06	4,98476e-05	0,03111	0,9753
Vol_3retards	-3,22503e-05	4,98447e-05	-0,6470	0,5193
Vol_4retards	2,51853e-05	4,95618e-05	0,5082	0,6126
Moy. var. dép.	-0,005107	Éc. type var. dép.		0,167902
Somme carrés résidus	2,633567	Éc. type de régression		0,171061
R2	0,016649	R2 ajusté		-0,037981
F(5, 90)	0,304759	p. critique (F)		0,908875
Log de vraisemblance	36,39032	Critère d'Akaike		-60,78065
Critère de Schwarz	-45,39456	Hannan-Quinn		-54,56134
rho	0,189441	Durbin-Watson		1,595384

Constante mise à part, la probabilité critique est la plus élevée pour la variable 4 (Vol_2retards)

En effectuant une régression multiple avec des retards, les résultats diffèrent quelque peu d'avec ceux obtenus en effectuant des régressions simples avec chacun des retards. Il peut y avoir plusieurs explications à cette différence, notamment le fait que l'échantillon sur lequel ces régressions sont effectuées est relativement court (de janvier 2008 à mai 2016). De plus, quelques quintiles se retrouvent avec peu de titres, ce qui ajoute à la difficulté de conclure sur ces analyses. L'analyse des résidus pourra aussi être très révélatrice et les prochaines sections tenteront de mettre en évidence certains tests de spécificité du modèle.

5.4.1. TEST DE LA STATISTIQUE F

Un examen du test de la statistique F pour chacun des cinq quintiles révèle que pour les quintiles D1 à D4, il y a au moins un des coefficients des variables indépendantes qui est statistiquement différent de zéro. Seul le quintile D5, soit celui comportant très peu de titres, ne comporte aucun coefficient statistiquement différent de zéro. Dans une certaine mesure, il y a présence d'une forme quelconque de corrélation entre le volume des transactions et le rendement d'un titre.

5.4.2. TEST DE NORMALITÉ DES RÉSIDUS

Le test de Jarque-Bera est un test de spécification du modèle qui vise à déterminer si les résidus sont normalement distribués ou non. Les hypothèses sous-jacentes au test sont les suivantes :

H_0 : les résidus suivent une distribution normale

H_1 : les résidus ne suivent pas une distribution normale

La première étape pour effectuer ce test est de sauvegarder les résidus pour chacune des régressions obtenues pour les cinq quintiles. Une fois dans le fichier de sortie de la régression respective de chaque quintile, il faut cliquer sur « Sauvegarder », puis sur

« Résidus » dans le menu déroulant qui apparaît. Dans le menu qui apparaît, il est recommandé de laisser le nom « uhat1 », mais de donner une description de ce fichier. Ce fichier créé sera ajouté aux autres fichiers de chacune des variables. Il faut le sélectionner en cliquant une seule fois dessus afin de le mettre en surbrillance, puis il faut cliquer sur l'onglet « Variables » et sur « Test de Normalité » du menu déroulant qui apparaît. Le menu qui apparaît présente le résultat de différents tests qui visent à mesurer si les résidus suivent ou non une distribution normale. La règle de décision du test est que si la valeur critique du test (*p-value*) est supérieure à la valeur de la statistique obtenue en effectuant le test Jarque-Bera, par conséquent l'hypothèse H_0 ne peut être rejetée et il est alors possible de conclure que les résidus suivent une distribution normale. Il est souhaitable d'examiner les résultats des autres tests de normalité²⁶ pour valider la conclusion du test Jarque-Bera.

Tableau X

Test de spécificité de Jarque-Bera

Ce tableau résume les statistiques du test de Jarque-Bera, soit la statistique du test, ainsi que sa valeur critique associée. Une statistique du test plus élevée que la valeur critique indique un rejet de l'hypothèse H_0 .

Test de Jarque-Bera		
Quintile	Statistique JB	Valeur critique JB
D1	41,994	7,60533E-10
D2	5,19637	0,0744084
D3	22,7315	1,15858E-05
D4	4,02139	0,133896
D5	97,5259	6,64538E-22

Comme la statistique du test de Jarque-Bera est supérieure à la valeur critique calculée, et ce, pour tous les quintiles, il faut donc rejeter l'hypothèse H_0 et admettre l'hypothèse alternative. Les résidus ne suivent pas une distribution normale, et ce, pour tous les quintiles.

Un examen visuel des résidus est intéressant, car il permet de voir à l'œil nu la forme de la distribution des résultats. Le graphique de la distribution de fréquences des résidus pour les quintiles D1 et D2 est reproduit ci-dessous.

²⁶ GRETL fournit les résultats des tests de Doornik-Hansen, Shapiro-Wilk, Lilliefors, ainsi que le test de Jarque-Bera. Aux fins du présent document, seul le test de Jarque-Bera sera utilisé et commenté pour vérifier la normalité des résultats.

Tableau XI

Graphique de la distribution de fréquence des résidus – Quintile D1

Ce tableau présente la distribution de fréquence des résidus pour le quintile D1 en comparaison avec une courbe normale. La forme que prennent les résultats présente une déformation vers la droite, soit une asymétrie par rapport à la moyenne.

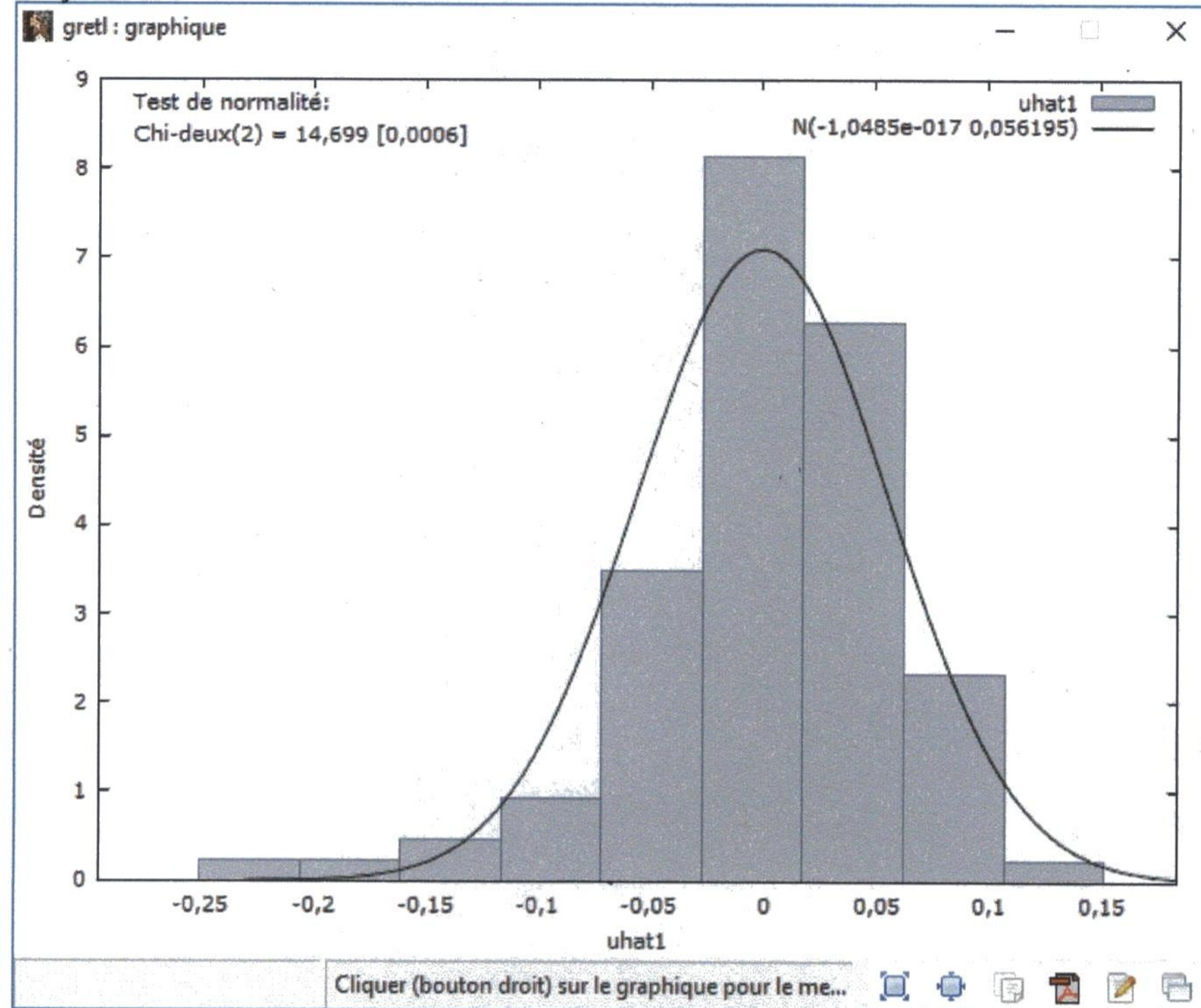
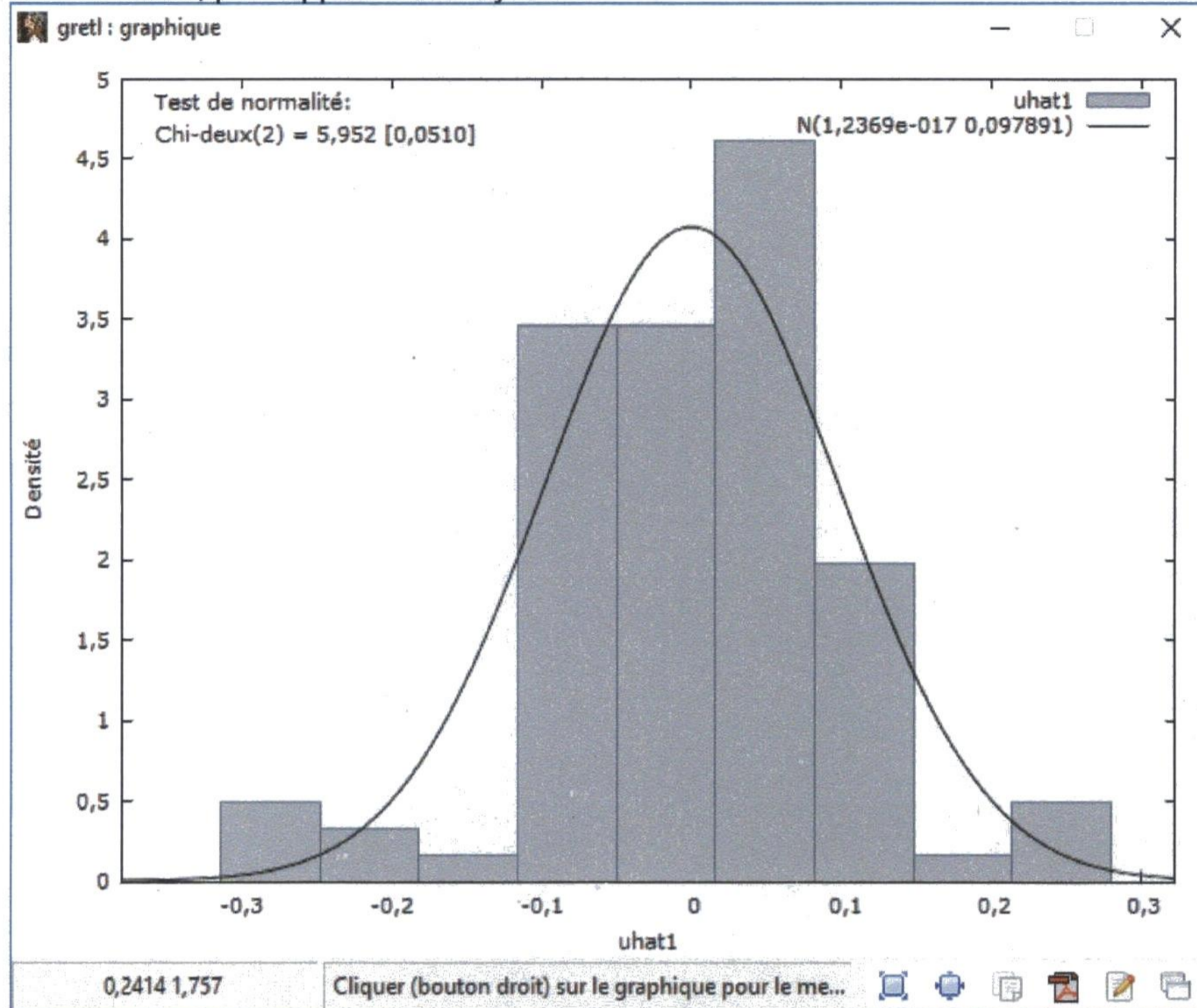


Tableau XII

Graphique de la distribution de fréquence des résidus – Quintile D2

Ce tableau présente la distribution de fréquence des résidus pour le quintile D2 en comparaison avec une courbe normale. La forme que prennent les résultats présente une concentration de ceux-ci entre -0,1 et 1,5, avec une légère asymétrie vers la droite, par rapport à la moyenne.



5.4.3. TEST DE WHITE - HÉTÉROSCÉDASTICITÉ

Le test de White mesure la présence d'hétéroscédasticité parmi les résidus, soit que la variance des résidus n'est pas constante pour toutes les observations. Les hypothèses pour ce test sont les suivantes :

H_0 : homoscélasticité (variance des résidus est constante pour toutes les observations)

H_1 : hétéroscédasticité (variance des résidus non constante pour toutes les observations)

Pour effectuer ce test, il suffit de se positionner sur l'écran de sortie de la régression multiple, puis de sélectionner l'onglet « Tests », puis « Hétéroscédasticité » et enfin « Test de White ». Sur l'écran de sortie, la statistique du test apparaît, ainsi que la valeur critique qui l'accompagne. La règle de décision du test est que si la valeur critique (*p-value*) est inférieure à la statistique du test, l'hypothèse H_0 ne sera pas rejetée. Dès lors, on admettra que les résidus sont homoscélastiques, soit qu'ils présentent une variance constante pour les observations.

Tableau XIII Test de spécificité de White

Ce tableau résume les statistiques du test de White pour mesurer l'hétéroscédasticité parmi les résidus. Lorsque la statistique du test de White est supérieure à sa valeur critique associée, il est alors possible d'admettre l'hypothèse H_0 et de conclure qu'il y a homoscedasticité dans les variances des observations.

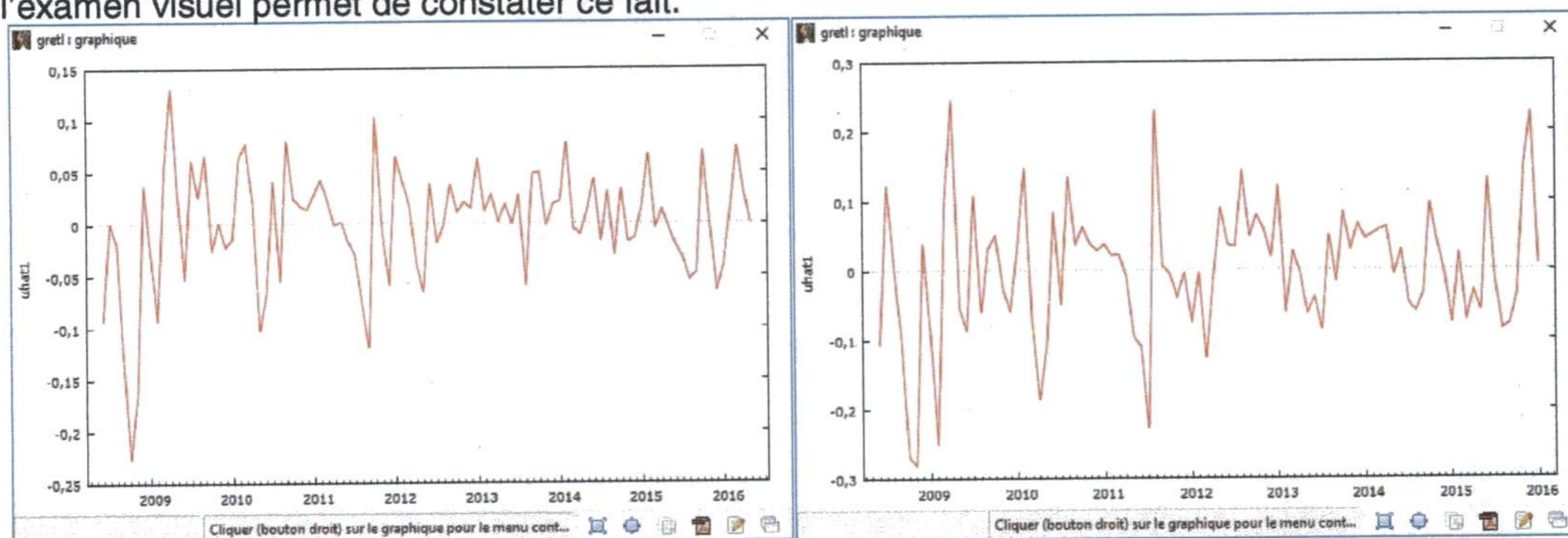
Test de White		
Quintile	Statistique White	Valeur critique White
D1	22,19891	0,329816
D2	22,101405	0,335034
D3	51,996037	0,000114
D4	13,458933	0,856837
D5	31,569124	0,048112

Pour chacun des cinq quintiles, la statistique du test de White est supérieure à sa valeur critique correspondante. Par conséquent, l'hypothèse H_0 ne peut donc être rejetée et la variance des résidus est donc homoscedastique. Bien que l'examen visuel du graphique des résidus ne remplace nullement l'analyse statistique des variances des résidus, il est toutefois intéressant d'en faire un examen. Les graphiques pour les quintiles D1 et D2 sont reproduits ci-dessous.

Tableau XIV

Graphique de la série temporelle des variances des résidus – Quintiles D1 et D2

Ces deux graphiques présentent la série temporelle des variances des résidus pour les quintiles D1 et D2. La conclusion est que statistiquement, la variance des résidus est homoscedastique et l'examen visuel permet de constater ce fait.



5.4.4. TEST DE DURBIN-WATSON – AUTOCORRÉLATION DE PREMIER ORDRE

Le test de Durbin-Watson est un test qui détecte l'autocorrélation de premier ordre dans les résidus. L'hypothèse nulle de ce test est qu'il y a absence d'autocorrélation de premier ordre. L'hypothèse alternative est qu'il y a présence d'autocorrélation de premier ordre.

H_0 : absence d'autocorrélation de premier ordre dans les résidus (aucun AR(1))

H_1 : présence d'autocorrélation de premier ordre dans les résidus (présence d'un AR(1))

En réalité, le test de Durbin-Watson pose deux hypothèses nulles et deux hypothèses alternatives afin de tenter de déceler la présence d'autocorrélation positive et négative.

Le calcul des valeurs critiques est assez complexe et dépend du nombre de variables dans le modèle et la taille de l'échantillon. Ce test est assez limité, car il ne peut être utilisé que pour une régression comportant une constante et qui ne comprend aucun retard pour la variable dépendante, ce qui est cependant le cas ici.

L'écran de sortie de la régression montre la statistique Durbin-Watson, mais ne montre pas la valeur critique (*p-value*). Pour obtenir cette dernière valeur, à l'écran de sortie de la régression, il faut cliquer sur l'onglet « Tests », puis sur le « P. critique de Durbin-Watson » dans le menu déroulant. L'écran apparaissant reprend la statistique DW et la compare à sa valeur critique. Le tableau suivant résume ces valeurs :

Tableau XV

Test de spécificité de Durbin-Watson

Ce tableau résume les statistiques du test de Durbin-Watson pour mesurer la présence d'autocorrélation entre les résidus. Lorsque la statistique du test de Durbin-Watson est supérieure à sa valeur critique associée, il faut rejeter l'hypothèse H_0 et de conclure qu'il y a présence d'autocorrélation de premier ordre dans les résidus.

Test de Durbin-Watson		
Quintile	Statistique DW	Valeur critique DW
D1	1,44497	0,00285022
D2	1,64824	0,0419179
D3	1,90694	0,326855
D4	N/A	N/A
D5	1,59538	0,023933

Pour tous les quintiles, la statistique Durbin-Watson est toujours supérieure à sa valeur critique, indiquant qu'il faut rejeter l'hypothèse H_0 et d'admettre l'hypothèse H_1 à l'effet qu'il y a présence d'autocorrélation de premier ordre dans les résidus. À noter que la faible quantité de résidus et le peu de résidus qui se suivent dans le quintile D4 ne permettent pas à GRETL de calculer ces valeurs, d'où la mention « N/A » dans le tableau ci-dessus.

5.4.5. TEST DE BREUSCH-GODFREY – AUTOCORRÉLATION

Le test classique de Durbin-Watson effectué à la section précédente ne fait que tester si des résidus qui se suivent ont une forme de relation statistique entre eux. Non seulement le test de Durbin-Watson doit satisfaire les conditions mentionnées précédemment, mais il comporte des lacunes quant à la détection de plusieurs formes d'autocorrélation dans les résidus. Le test de Breusch-Godfrey apporte des raffinements et des solutions à certaines lacunes du test de Durbin-Watson. Le test de Breusch-Godfrey pose sensiblement les mêmes hypothèses que le test de Durbin-Watson, mais à la différence que l'autocorrélation sera d'ordre P (un ou plusieurs retards). Les hypothèses du test sont les suivantes :

H_0 : absence d'autocorrélation de premier ordre dans les résidus (aucun AR(1))

H_1 : présence d'autocorrélation de premier ordre dans les résidus (présence d'un AR(1))

Plus précisément, le modèle sous-jacent à ce test est le suivant:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \dots + \rho_r u_{t-r} + v_t, v_t \sim N(0, \sigma_v^2)$$

Où :

u_t : terme d'erreur au temps t .

ρ_r : coefficient de corrélation affectant le résidu de la période en question par rapport au résidu au temps t .

v_t : terme d'erreur du modèle.

Et les hypothèses ci-dessus, exprimées en langage mathématique généralisé, sont les suivantes :

$$H_0: \rho_1 = 0 \text{ et } \rho_2 = 0 \text{ et } \dots \text{ et } \rho_r = 0$$

$$H_1: \rho_1 \neq 0 \text{ ou } \rho_2 \neq 0 \text{ ou } \dots \text{ ou } \rho_r \neq 0$$

Selon l'hypothèse H_0 , le terme d'erreur de la période courante est non corrélé avec aucune de ses valeurs précédentes.

Pour ce test, il sera intéressant de mesurer l'autocorrélation avec quatre retards. Pour ce faire, il suffit de cliquer sur l'onglet « Tests », à l'écran de sortie de la régression multiple, puis de cliquer sur « Autocorrélation ». Un écran apparaît pour poser la question « Ordre des retards pour le test », auquel le chiffre « 4 » sera entré. Le tableau résumant les résultats est présenté ci-dessous.

Tableau XVI
Test de spécificité de Breusch-Godfrey

Ce tableau résume les statistiques du test de Breusch-Godfrey pour mesurer la présence d'autocorrélation entre les résidus, avec spécification du nombre de retards à mesurer. Lorsque la statistique du test de Breusch-Godfrey (appelée ici « Statistique alternative ») est supérieure à sa valeur critique associée, il faut rejeter l'hypothèse H_0 et de conclure qu'il y a présence d'autocorrélation de premier ordre dans les résidus. Le tableau indique quel retard présente de l'autocorrélation, ainsi que le niveau de confiance de la mesure statistique.

Test de Breusch-Godfrey				
Quintile	Retard(s) significatif(s)	Retards non-significatifs	Statistique alternative	Valeur critique BG
D1	1 (***)	2, 3, 4	11,781777	0,019
D2	1 (*)	2, 3, 4	4,860081	0,302
D3	3 (**)	1, 2, 4	7,442054	0,114
D4	N/A	N/A	N/A	N/A
D5	1 (*)	2, 3, 4	5,075772	0,28

La règle de décision est que si la valeur de la statistique de Breusch-Godfrey est supérieure à la valeur critique (p -value), l'hypothèse nulle sera rejetée en faveur de l'hypothèse alternative, à savoir qu'il y a présence d'autocorrélation, et ce, pour tous les quintiles. Le nombre d'étoiles à droite des coefficients de retard indique la présence ou non d'autocorrélation. Une, deux ou trois étoiles indiquent le niveau de confiance de la relation statistique, soit respectivement de 90 %, 95 % et de 99 %. Aucune étoile signifie qu'il n'y a aucune relation statistique, ou dit autrement, que $\rho_r = 0$. Tout comme pour le test de Durbin-Watson, la faible quantité de résidus et le peu de résidus qui se suivent

dans le quintile D4 ne permet pas à GRETLM de calculer ces valeurs, d'où la mention « N/A » dans le tableau ci-dessus.

5.5. DISCUSSION

Une première conclusion à tirer de la régression multivariée, ainsi que de l'analyse des tests de spécificité est qu'il existe, dans une certaine mesure, une relation entre le volume des transactions et le rendement des titres. Il faut cependant faire preuve de jugement dans l'analyse des résultats afin d'en venir à cette conclusion.

D'abord, bien que les résidus ne suivent pas une distribution normale et, ce, pour tous les quintiles, la variance des résidus est cependant homosédastique pour tous les quintiles, donc présente une variance constante.

Le problème d'autocorrélation pour tous les quintiles est très révélateur et ne peut être ignoré. Les coefficients estimés pour les variables indépendantes dans le modèle de régression multiple demeurent toujours non biaisés, mais ne sont pas « efficaces », c'est-à-dire qu'ils ne sont pas « BLUE »²⁷, et ce, même pour de grands échantillons, de sorte que les erreurs standards estimées pourraient être erronées. Il y a donc des risques d'effectuer des inférences qui soient aussi erronées.

Il existe plusieurs façons de corriger l'autocorrélation et une des façons de le faire est d'appliquer la technique connue sur le nom anglais de « *Generalized Least Squares* » (GLS), mais cela requiert l'estimation d'un ρ . Toutefois, ces techniques dépassent l'objet du présent document.

Tel que mentionné plusieurs fois dans ce texte, un échantillon de janvier 2008 à mai 2016 est jugé assez court, ce qui peut occasionner certains problèmes. De plus, la division en quintile laisse certains de ceux-ci avec peu de titres, ce qui rend difficiles des analyses robustes sur lesquelles il est possible de bâtir un modèle qui aiderait dans les prévisions. Toutefois, les analyses effectuées dévoilent tout de même la présence d'une relation sous une forme quelconque entre le volume des transactions et le rendement futur des titres.

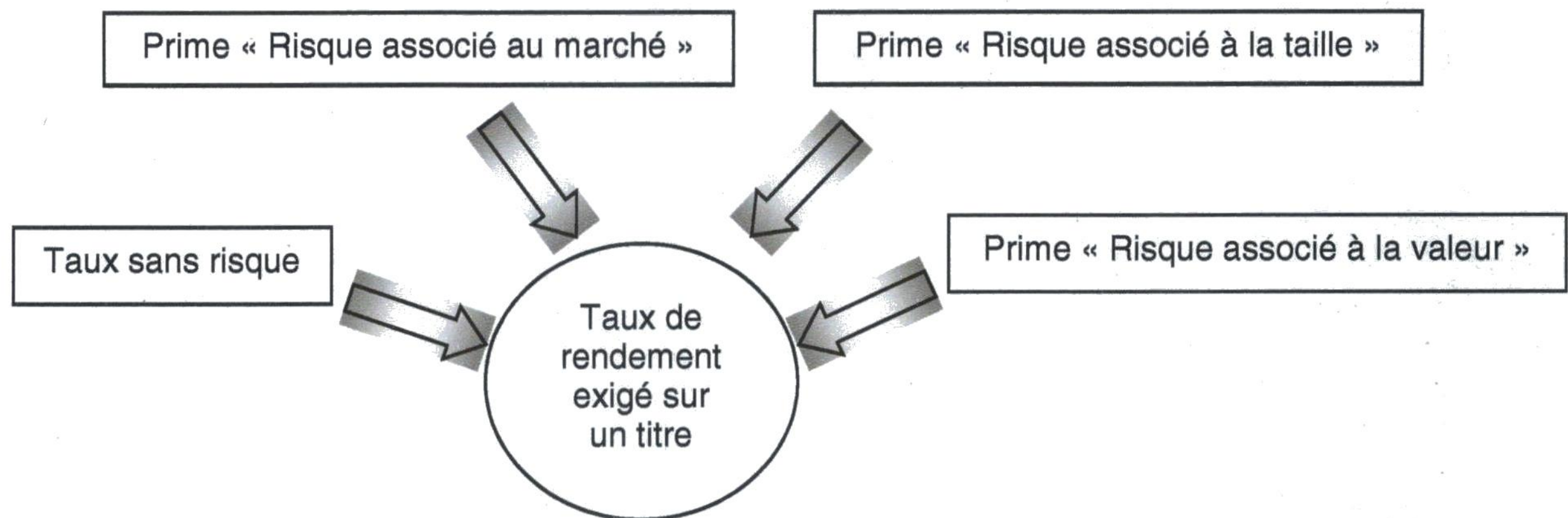
Pour tenter de soutenir la thèse que le volume des transactions a un effet sur le rendement futur des titres, Diether et Malloy (2002) suggèrent d'appliquer le modèle des trois facteurs de Fama-French aux cinq quintiles fondés sur le volume et de constater si les résultats des régressions permettent de conclure que le modèle explique toujours fortement le rendement futur en fonction des trois variables indépendantes retenues par Fama-French et ce, pour les cinq quintiles. Si le volume des transactions n'avait pas d'effet significatif sur le rendement futur, alors les résultats de l'application du modèle des trois facteurs de Fama-French devraient être sensiblement toujours les mêmes pour les cinq quintiles et expliquer fortement le rendement excédentaire, et ce, nonobstant le nombre de retards utilisés. C'est ce que la prochaine section tentera de mettre en lumière.

²⁷ L'acronyme anglais « BLUE » signifie « *Best Linear Unbiased Estimators* » et est utilisé pour décrire un modèle dont les résidus respectent les conditions associées aux régressions linéaires.

6. RÉGRESSION AVEC LE MODÈLE DES TROIS FACTEURS DE FAMA-FRENCH

6.1. PRÉSENTATION DU MODÈLE

Le modèle a été développé par deux économistes américains, soit Eugene Fama²⁸ et Kenneth French²⁹ (1993). L'intuition qui a servi de point de départ pour les auteurs du modèle est essentiellement de tenter de démontrer que d'autres facteurs que celui de la prime pour le risque associé au marché existent et influencent la valeur d'un titre. Voici le diagramme des trois facteurs de risque qui influencent le taux de rendement exigé sur un titre :



Le modèle d'évaluation des actifs financiers (« MÉDAF ») utilise seulement un facteur, soit le facteur « risque associé au marché ». Le modèle des trois facteurs de *Fama-French*, quant à lui, inclut deux autres facteurs pour calculer le taux de rendement exigé sur un titre. Le modèle Fama-French inclut donc trois facteurs au total, dont un est le facteur « risque associé au marché », qui est la prime calculée pour le MÉDAF. Le modèle inclut aussi la prime « risque associé à la taille » (« *SMB* »), parce que les conclusions des recherches antérieures de Fama et French ont démontré que les petites sociétés tendent à réaliser un meilleur rendement ou du moins, ont un taux de rendement exigé supérieur aux plus grandes sociétés, car elles sont plus risquées. Enfin, le dernier facteur est la prime « risque associé à la valeur » (« *HML* »), qui tient compte du fait que les sociétés qui démontrent un fort ratio valeur aux livres par rapport à la valeur au marché tendent à être sous-évaluées sur le marché et tendent à générer de meilleurs rendements. Tout comme pour le MÉDAF, le taux sans risque fait partie intégrante de la formule du calcul du taux de rendement exigé sur un titre.

²⁸ Eugene Francis « Gene » Fama (né le 14 février 1939 à Boston, Massachusetts) est un économiste américain auteur de la théorie d'efficience des marchés financiers. Ses travaux sur les prix des actifs lui ont valu le prix Nobel d'économie pour 2013 (Wikipedia).

²⁹ Kenneth R. French est un économiste américain né le 10 mars 1954. Il est professeur de finance à la Tuck School of Business du Dartmouth College. Il est célèbre pour son travail avec Eugene Fama sur l'évaluation des actifs. Ensemble, ils ont mené des travaux remettant en cause le modèle d'évaluation des actifs financiers et développé le modèle *Fama-French* à trois facteurs (Wikipedia).

6.2. ÉQUATION DU MODÈLE DE FAMA-FRENCH

$$r_i = R_F + \beta_i^{\text{marché}} RMRF + \beta_i^{\text{taille}} SMB + \beta_i^{\text{valeur}} HML$$

Où :

- r_i = taux de rendement exigé sur un titre individuel i .
- $RMRF$ = rendement du marché moins le rendement d'un titre sans risque
- SMB = rendement d'un portefeuille de petites sociétés moins le rendement d'un portefeuille de grandes sociétés. C'est une prime pour la taille de la société.
- β_i^{taille} = sensibilité du titre i aux changements de valeur du facteur SMB . C'est une forme de bêta.
- HML = rendement d'un portefeuille de titres démontrant un haut ratio valeur aux livres par rapport à la valeur au marché (titres de valeur) moins le rendement d'un portefeuille de titres démontrant un faible ratio valeur aux livres par rapport à la valeur au marché (titres de croissance). C'est une prime de valeur.
- β_i^{valeur} = sensibilité du titre i aux changements de valeur des titres de valeur. C'est comme un coefficient de corrélation pour la prime associée à la valeur.

Les bêtas sont calculés pour chacun des titres en utilisant une régression multivariée. Une fois les bêtas calculés, ils sont multipliés par leur facteur de risque respectif et le résultat des trois multiplications est additionné ensemble, auquel on ajoute le taux de rendement sans risque.

Exemple

Taux de rendement sans risque	3,0 %
Prime de risque associé au marché	5,0 %
Bêta	1,20
Prime de risque associé à la taille	2,2 %
Bêta pour le facteur taille du titre	0,12
Prime de risque associé à la valeur	3,8 %
Bêta pour le facteur valeur du titre	0,34

Avec ce modèle, le bêta correspondant au risque associé au marché sera, à toute fin pratique, celui utilisé par le MÉDAF, cependant, il ne sera pas calculé de la même manière. Pour calculer les trois bêtas, ce modèle utilise une régression multivariée qui inclura aussi les effets, ou degrés d'influence, des facteurs « taille » et « valeur au marché » dans les résultats de la régression. À titre d'exemple pour mieux en comprendre son interprétation, un bêta pour le facteur « valeur du titre » mesure la sensibilité du titre quant à ce facteur et, par conséquent, le risque associé au facteur « valeur du titre ».

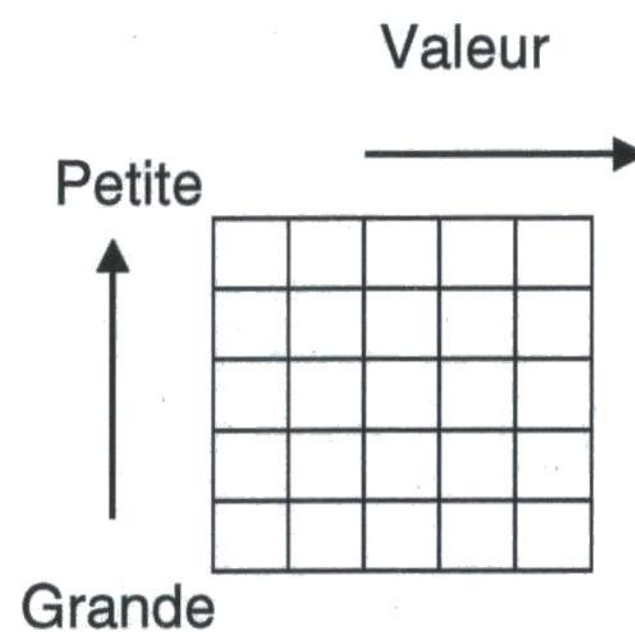
$$r_i = R_F + \beta_i^{\text{marché}} RMRF + \beta_i^{\text{taille}} SMB + \beta_i^{\text{valeur}} HML$$

$$r_i = 3 \% + 1,20(5 \%) + 0,12(2,2 \%) + 0,34(3,8 \%) = 10,56 \%$$

Pour le titre cité en exemple, le taux de rendement exigé par un investisseur pour celui-ci devrait être de 10,56 %.

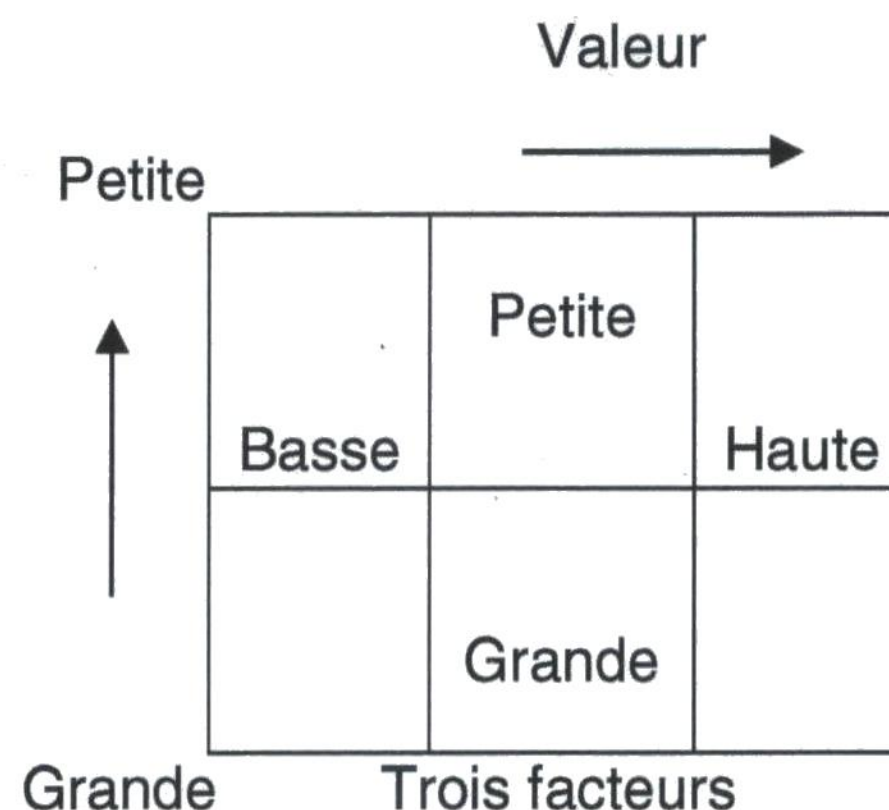
Essentiellement, l'excédent entre le rendement espéré d'un titre ou d'un portefeuille sur le taux de rendement sans risque est expliqué par une analyse de sa sensibilité par rapport aux trois facteurs. Les trois bêtas ci-dessus sont obtenus par une régression sur une série temporelle de données et représentent essentiellement la pente respective de chacun des facteurs. Bien que les données des paramètres ci-dessus soient fournies par le site internet de Kenneth French, il est intéressant d'expliquer comment Fama et French calculent leurs facteurs SML et HMB.

Les auteurs du modèle ont d'abord cherché à construire 25 portefeuilles démontrant un rendement excédentaire intéressant. Pour les construire, chaque mois de juin, ils déterminent la taille et le ratio de la valeur aux livres par rapport à la valeur au marché de chacun des titres. Chaque mois de janvier suivant, ils construisent des portefeuilles en incorporant les titres de petite taille et les titres dits de « valeur » (en anglais : *value stocks*) dans les portefeuilles. L'étape suivante est de diviser cet ensemble en 25 créneaux, ou portefeuilles en ayant à gauche la taille, déterminée par la capitalisation boursière du titre, soit le prix du titre multiplié par le nombre total de titres en circulation sur le marché. Les flèches indiquent le sens de la mesure. Plus la taille est petite, plus le titre sera placé en haut du cadre des 25 portefeuilles. Puis, les portefeuilles horizontaux sont classés en fonction de leur valeur, qui représente la valeur comptable d'un titre divisé par sa valeur boursière.



De façon générale, les titres de petite taille valent moins et les titres dits « de valeur » ont un prix sur le marché plus bas, par rapport à leur valeur comptable aux livres.

De façon similaire, ils ont formé une matrice des facteurs, qui ressemble à ce qui suit :



Pour le facteur lié au marché, tous les titres sont simplement répartis sur la grille précédente en fonction de leurs caractéristiques. Les portefeuilles correspondants aux facteurs SMB et HML sont répartis sur les mêmes dimensions dessinées ci-dessus en fonction de leurs caractéristiques. Pour les portefeuilles SMB, on prend le rendement des portefeuilles contenus dans la case « Petite » et on soustrait du rendement des portefeuilles contenus dans la case « Grande ». Quant aux portefeuilles HML, on prend le rendement des portefeuilles contenus dans la case « Haute », valeur de laquelle on soustrait le rendement des portefeuilles contenus dans la case « Basse »³⁰.

Ce que les auteurs cherchent à faire, c'est de démontrer l'existence ou non d'une différence significative dans les rendements entre les titres de valeur, les titres de croissance, les titres de petite taille et les titres de grandes tailles. Les auteurs concluent qu'il y a une différence statistiquement significative entre les rendements des différents portefeuilles cités précédemment.

Selon le tableau 1 de l'article Fama-French, en prenant la direction de grande à petite (facteur SMB), ou la direction de basse à haute (facteur HML) on voit de façon très claire les tendances à la hausse du rendement espéré par les investisseurs, lorsque les titres passent de « grande » à « petite », donc on s'attend à un rendement espéré plus grand pour les petites firmes que pour les grandes. De façon similaire, le même phénomène apparaît plus on se dirige vers la droite, soit la direction de basse à haute pour la valeur, le rendement espéré augmente. Cela est, selon les auteurs, un impact très significatif, démontrant que ces facteurs jouent donc un rôle clé dans leurs attentes de rendement. Le tableau 1 démontre donc que plus les titres tendent à être petits et plus les titres tendent à être des titres dits « de valeur », plus ils généreront un rendement moyen élevé. Intuitivement, partant de cette prémisse, les investisseurs devraient donc acheter ces titres avant tous les autres. C'est de là que les auteurs modélisent le fait que si ces titres ont un rendement moyen plus élevé, peut-être auront-ils aussi un bêta plus élevé. Donc, les auteurs rattachent un bêta pour chacun de ces facteurs, afin de bien mesurer la sensibilité de ces titres par rapport aux titres de la catégorie opposée. Avec cette intuition de départ, le modèle des trois facteurs a vu le jour.

L'avantage du modèle des trois facteurs de Fama-French sur le modèle MÉDAF réside dans le fait qu'il incorpore aussi l'effet des primes de risque associées à la taille et à la valeur du titre, ce que le modèle MÉDAF ne fait pas. Les diverses recherches de ces auteurs ont aussi démontré que le modèle des trois facteurs semble produire un calcul plus juste du taux de rendement à exiger sur un titre.

6.3. EXPLICATION DÉTAILLÉE DE LA MÉTHODOLOGIE UTILISÉE

Le site Internet³¹ de Kenneth French, propose une myriade de chiffriers comportant tous les calculs mensuels des facteurs de risques du modèle, et ce, pour de nombreuses années. La première étape est donc de télécharger le chiffrier approprié du site en question. Aux fins de cette recherche, le fichier avec données mensuelles « *Fama/French 3 Factors* », de la section « *U.S. Research Returns Data (Downloadable Files)* » a été utilisé. Bien que ce fichier soit le plus approprié pour les besoins de cette recherche, les données contenues dans celui-ci ont été colligées à partir des bourses du

³⁰ Le tableau 1 (*Table 1*), en page 11 de l'article de *Fama-French* de 1993 sur le modèle des trois facteurs montre le résultat de leurs analyses.

³¹ En ligne : http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

NYSE, AMEX et NASDAQ, alors que les données de la présente recherche ne se fondent que sur les titres cotés sur le NYSE. Bien qu'il puisse y exister des différences entre les résultats de l'analyse de régression ci-dessous et ceux d'une régression qui tiendrait compte des données compilées avec les trois bourses de valeurs ci-dessus, plutôt que juste sur le NYSE, la régression fera toutefois ressortir si la relation entre ces facteurs et le rendement est statistiquement significative. Le lecteur est prié de consulter le site internet³² de Kenneth French pour plus de détails sur la façon dont il a calculé les facteurs retenus. Il est important de mentionner que bien qu'il ait été possible de calculer les rendements des divers types de portefeuilles pour chacun des facteurs, ceux du site de Kenneth French ont été retenus.

Une fois le fichier de données téléchargé, il faut utiliser la fonction "Convertir" de la bande « Données » du chiffrier Excel pour convertir les données qui sont sous format CSV (ou format TXT) en colonnes appropriées et en chiffres manipulables. Les données pour les trois facteurs du mois de février 2008 au mois de mai 2016 sont extraites et copiées dans un fichier Excel. Il est important de mentionner que la date donnée sur le fichier des données du site de Kenneth French couvre le mois précédant cette date. À titre d'exemple, les rendements du mois de décembre 2015 sont en fait les rendements pour la période du 1^{er} novembre au 1^{er} décembre 2015. Il est donc important de s'assurer que toutes données téléchargées ou importées qui serviront dans la régression correspondent à la même période que les données du NYSE utilisées, ou à la période désirée, dans le cas d'une analyse avec retard(s). Le chiffrier servant de base à la régression multiple a comme dernière date³³ le mois de juin 2016 et a comme date du début celle du mois de mars 2008, qui représente en fait les rendements de février 2008. Cela permettra de faire un appariement des données de la recherche et celles importées du site de Kenneth French.

L'étape suivante est de soustraire le taux sans risque mensuel de chacun des rendements mensuels respectifs. Le résultat obtenu est le rendement mensuel excédentaire des titres sur le NYSE. Ce résultat représente la variable dépendante de la régression linéaire. Les trois variables indépendantes qui entreront dans l'équation de la régression sont les trois variables du modèle de Fama-French, soit RMRF (prime associée au risque de marché), SMB (prime de risque associée à la taille) et HML (prime de risque associée à la valeur).

Afin d'obtenir la régression pour chacun des quintiles associés aux volumes de transactions, dans la feuille de travail Excel, on sélectionne l'onglet du haut intitulé « Données », puis on sélectionne « Utilitaire d'analyse », de la section « Analyse » (généralement le dernier bouton à droite de la bande). Les plages respectives des cellules contenant les données pour la variable dépendante et celles contenant les variables indépendantes sont entrées dans le menu contextuel Excel. Il est important de cocher la case « Intitulé présent » du même menu, avant d'appuyer sur la touche « OK ». Les tableaux ci-dessous présentent respectivement un sommaire des résultats obtenus pour chacun des quintiles associés aux volumes de transactions des titres, et ce, avec aucun retard, un retard d'un mois et un retard de deux mois.

³² En ligne : http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_bench_factor.html.

³³ Cela représente en fait le mois complet de mai 2016, se terminant le 1^{er} juin 2016.

6.4. RÉGRESSION AVEC MODÈLE FAMA-FRENCH - AUCUN RETARD

L'analyse du tableau XVII a été effectuée avec aucun retard, c'est-à-dire que la régression est effectuée en prenant la variable dépendante et les trois variables indépendantes pour le même mois.

Tableau XVII

Statistiques de la régression et analyse de variance – Aucun retard

Ce tableau résume les statistiques de la régression effectuée pour la régression avec le modèle des trois facteurs de Fama-French, et ce, avec aucun retard. Selon les statistiques du modèle avec aucun retard, les R^2 sont extrêmement faibles. Ce modèle n'offre pas beaucoup de capacités explicatives.

Statistiques de la régression et analyse de variance - Aucun retard					
	Quintiles fondés sur le volume				
	Petit				Grand
	1	2	3	4	5
SSR	184,265421874	389,122615937	2882,978662637	503,743811013	1646,475239961
SSE	3343,671474463	11226,060925286	21975,438111374	5288,575019613	26285,407955721
SST	3527,936896337	11615,183541223	24858,416774010	5792,318830626	27931,883195682
Erreur-type	5,901687148	11,046375033	17,470386894	12,855659040	16,547094394
R^2 ajusté	0,022612561	0,001984937	0,079141621	0,001370764	0,029538153
MSR	61,421807291	129,707538646	960,992887546	167,914603671	548,825079987
MSE	34,829911192	122,022401362	305,214418214	165,267969363	273,806332872
F	1,763478722	1,062981364	3,148582866	1,016014200	2,004427999
Valeur critique de F	0,159327279	0,368802710	0,030151804	0,398451520	0,118491949
Constante:					
Coefficient	0,077205430	(1,431320236)	(0,290613626)	(1,865655500)	(0,812589818)
Erreur-type	0,597590064	1,138665852	2,011155521	2,152737118	1,675517347
Statistique t	0,129194634	(1,257015158)	(0,144500822)	(0,866643440)	(0,484978457)
Probabilité	0,897473937	0,211930079	0,885508727	0,392589432	0,628796466
Mkt-RF:					
Coefficient	0,127184014	0,223874126	0,696956385	0,230753832	0,467756155
Erreur-type	0,134965465	0,254098705	0,452548175	0,441348261	0,378414890
Statistique t	0,942344868	0,881051818	1,540071144	0,522838430	1,236093418
Probabilité	0,348381512	0,380585706	0,127926697	0,604688228	0,219439780
SMB:					
Coefficient	(0,031797765)	(0,171202867)	(0,151160828)	(1,135865088)	0,240982146
Erreur-type	0,276636268	0,531373827	0,904339854	0,991512567	0,775630142
Statistique t	(0,114944310)	(0,322189122)	(0,167150466)	(1,145588191)	0,310692085
Probabilité	0,908729340	0,748040407	0,867720227	0,260460986	0,756708283
HML:					
Coefficient	0,405698836	0,565017734	1,753350058	(0,935190812)	0,984615522
Erreur-type	0,221239203	0,416574670	0,725658315	0,901113084	0,620308377
Statistique t	1,833756541	1,356342030	2,416219895	(1,037817371)	1,587300057
Probabilité	0,069788396	0,178309721	0,018219730	0,307133968	0,115733100

D'abord, pour les cinq quintiles, le R^2 est très faible, ce qui va aussi de pair avec les résultats de la portion de la somme des carrés des résidus expliqués par la régression par rapport à la somme totale des carrés des résidus. Le modèle dont la régression tente de justifier n'explique, en fait, que très peu. Les trois facteurs agissant comme variables indépendantes n'expliquent donc pas l'excédent du rendement des titres cotés sur le NYSE sur le taux sans risque, ou simplement la prime pour le risque.

En portant une attention sur les variables indépendantes comme telles, les résultats obtenus vont aussi dans le sens des R^2 obtenus. D'abord, la raison pour laquelle la moyenne des carrés de la régression (*MSR*) et la moyenne des carrés des erreurs

(MSE) sont utilisées³⁴ est qu'elles indiqueront si l'hypothèse suivante sera acceptée ou rejetée :

$$H_0 : \beta_1 (Mkt-RF) = \beta_2 (SMB) = \beta_3 (HML) = 0$$

$$H_1 : \text{Au moins un des } \beta \neq 0$$

L'hypothèse H_0 indique que les trois coefficients obtenus de la régression seront statistiquement égaux à zéro. S'il est impossible de rejeter cette hypothèse, cela signifie que la régression effectuée est totalement inutile et ne sera, par conséquent, d'aucune utilité prédictive. Pour pouvoir faire cette détermination, l'analyse du test de Fisher (la statistique F) est utilisée et pour l'obtenir, il s'agit de diviser le MSR par le MSE, soit :

$$F_{(3,N)} = \frac{MSR}{MSE}$$

La façon d'interpréter la statistique F est que si celle-ci est supérieure à la valeur critique de F , l'hypothèse H_0 pourra être rejetée au profit de l'hypothèse H_1 . Un rejet de H_0 signifie qu'il y a au moins un des coefficients de régression qui est significatif avec la marge d'erreur retenue de 5 %, mais ce test ne donne aucune information pour déterminer lequel ou lesquels des coefficients sont significatifs.

Pour tous les cinq quintiles ci-dessus, la statistique F est supérieure à la valeur critique, ce qui signifie qu'il y a au moins un des coefficients de régression qui est significatif et cela permet de conclure que l'hypothèse H_0 est rejetée dans tous les cas.

L'étape suivante sera de déterminer si la contribution marginale de chaque variable explicative est significative ou non dans l'équation de régression. Il s'agit donc de déterminer si l'ajout d'une variable indépendante à la suite d'autres variables indépendantes déjà dans l'équation de régression apporte une contribution significative dans l'explication de la variable dépendante. Ce test permet donc de juger de la pertinence de chaque variable indépendante comme si elle était la dernière variable introduite dans l'équation. Ce test s'effectue à l'aide du T de Student et avec les hypothèses suivantes :

$$H_0 : \beta_1 (Mkt-RF) = 0, \beta_2 (SMB) = 0, \beta_3 (HML) = 0$$

$$H_1 : \beta_1 (Mkt-RF) \neq 0, \beta_2 (SMB) \neq 0, \beta_3 (HML) \neq 0$$

La règle de décision est que H_0 est rejetée si la probabilité $p < \alpha$ (seuil de signification). Dans ce cas, la variable indépendante étudiée a un effet significatif, ou offre une contribution marginale significative, à la droite de régression.

Pour les trois variables indépendantes ci-dessus, la probabilité est toujours plus élevée que le seuil de signification, à part le troisième quintile de la variable HML. Les

³⁴ Elles sont obtenues en divisant respectivement celles-ci par les degrés de liberté qui leur sont associés. À titre d'exemple, pour une régression à trois variables indépendantes, il y a trois degrés de liberté. Dans le tableau ci-dessus, il s'agit du SSR divisé par trois degrés de liberté, qui donnera le MSR. Le MSE sera obtenu en divisant le SSE par le résultat obtenu en soustrayant le nombre de degrés de liberté du nombre total d'observations.

régressions sont donc concluantes, pour la quasi-totalité, à l'effet que l'hypothèse H_0 doit être acceptée dans tous les cas, sauf pour le troisième quintile de la variable HML. Il faut se rappeler que le troisième quintile est celui qui comporte le moins de données et qu'il est fort possible que les résultats obtenus soient erronés à cause de ce fait.

Tout comme pour la régression simple, il sera intéressant de refaire ces régressions en appliquant une période d'un retard, puis deux périodes de retard. Les tableaux XVIII et XIX de la section suivante présenteront les résultats des régressions qui ont été effectuées en décalant le rendement excédentaire d'un mois, puis de deux mois.

6.5. RÉGRESSION AVEC MODÈLE FAMA-FRENCH – RETARD D'UN MOIS

Le tableau XVIII présente les résultats de la régression effectuée en décalant d'un mois le rendement excédentaire par rapport aux trois variables indépendantes. Dit autrement, ces régressions mesurent si les trois variables indépendantes ont un effet significatif sur le rendement du mois suivant, plutôt que celui sur le mois courant.

Tableau XVIII

Statistiques de la régression et analyse de variance – Retard d'un mois

Ce tableau résume les statistiques de la régression effectuée pour la régression avec le modèle des trois facteurs de Fama-French, et ce, avec un retard d'un mois. Pour les quintiles D1, D2 et D4 et D5, les R^2 sont nettement supérieurs à la régression effectuée avec aucun retard. Bien que la capacité de ce modèle appliqué à un retard soit forte, les résultats démontrent qu'il y a d'autres variables que celles du modèle de Fama-French qui expliquent ce qui influence le rendement, notamment pour les quintiles montrant un plus grand volume de transactions.

Statistiques de la régression et analyse de variance - Retard d'un mois					
	Quintiles fondés sur le volume				
	Petit				Grand
	1	2	3	4	5
SSR	3296,047866961	8282,134399791	1959,160529291	2022,557593580	11630,210581577
SSE	225,688326355	3332,336279928	22863,396123179	3642,015140714	16284,686471306
SST	3521,736193315	11614,470679719	24822,556652470	5664,572734294	27914,897052884
Erreur-type	1,541319757	6,051370093	17,944906344	10,839020222	13,092660246
R^2 ajusté	0,933891889	0,703628894	0,040008028	0,294833272	0,398208788
MSR	1098,682622320	2760,711466597	653,053509764	674,185864527	3876,736860526
MSE	2,375666593	36,619079999	322,019663707	117,484359378	171,417752330
F	462,473406605	75,389973387	2,027992646	5,738515902	22,615725663
Valeur critique de F	0,000000000	0,000000000	0,117704537	0,003041135	0,000000000
Constante:					
Coefficient	(0,689056739)	(2,498516438)	(0,575369930)	(0,683041130)	(2,001901445)
Erreur-type	0,156869273	0,627133991	2,079034002	1,839228814	1,332517854
Statistique t	(4,392553911)	(3,984023310)	(0,276748687)	(0,371373656)	(1,502344932)
Probabilité	0,000029113	0,000136375	0,782777171	0,712885445	0,136324901
Mkt-RF:					
Coefficient	1,111461176	1,668318099	0,932164929	1,112211282	2,298463044
Erreur-type	0,035252199	0,139213103	0,464841293	0,373901649	0,299447964
Statistique t	31,528846718	11,983915796	2,005340194	2,974609188	7,675667626
Probabilité	0,000000000	0,000000000	0,048739348	0,005639677	0,000000000
SMB:					
Coefficient	0,306165883	0,159850837	0,703842565	(0,860524517)	(0,268544795)
Erreur-type	0,072284852	0,291269960	0,929642488	0,835992796	0,614019907
Statistique t	4,235546902	0,548806464	0,757111012	(1,029344417)	(0,437355193)
Probabilité	0,000052713	0,584482325	0,451488640	0,311282641	0,662846413
HML:					
Coefficient	0,230034993	1,073713781	(0,553108302)	1,651045025	0,115087978
Erreur-type	0,057877863	0,228587638	0,746723460	0,765635366	0,491640490
Statistique t	3,974490068	4,697164685	(0,740713706)	2,156437777	0,234089706
Probabilité	0,000137428	0,000009306	0,461309735	0,038920904	0,815419077

Les résultats ci-dessus sont très révélateurs, et ce, à plusieurs niveaux. D'abord, la portion expliquée par la régression avec un retard est de beaucoup supérieure à celle obtenue avec la régression sans aucun retard, ce qui se reflète par des R^2 ajustés assez élevés. Le quintile numéro trois obtient toutefois la moins bonne performance des cinq quintiles, et ce, à plusieurs niveaux. Il faut rappeler que ce quintile est celui qui comporte le moins de titres et qu'il est possible que les résultats de la régression souffrent de ce manque de données.

C'est la tendance des R^2 ajustés qui est intéressante. À mesure que le volume des transactions augmente, les R^2 ajustés diminuent, ce qui signifie que le modèle perd de sa capacité explicative et que d'autres variables peuvent entrer en jeu pour expliquer la diminution du rendement. Le quintile numéro un, soit celui comportant les titres ayant le volume des transactions les plus faibles, obtient le R^2 ajusté le plus élevé, avec 93,39 %. Puis, les R^2 ajustés des quintiles suivants diminuent au fur et à mesure que le volume des transactions augmente, avec une remontée au cinquième quintile. Cette tendance aide à solidifier la thèse selon laquelle plus le volume des transactions augmente, moins le modèle des trois facteurs de Fama-French explique les résultats par les variables indépendantes qui sont incluses dans ce modèle. Ces résultats suggèrent donc qu'à mesure que le volume de transactions augmente, d'autres variables entreront en jeu pour venir expliquer le rendement obtenu.

Pour tous les quintiles, la statistique F est toujours supérieure à la valeur critique de F , indiquant que la régression est significative dans son ensemble. Quant à l'analyse de l'apport de chacune des variables indépendantes, les résultats sont très concluants pour la variable Mkt-RF. La probabilité est toujours inférieure au seuil de signification de 5 %, ce qui signifie que cette variable contribue significativement à expliquer la variation du rendement. Pour la variable SMB, seul le premier quintile est significatif, pour les autres quintiles, la probabilité est toujours supérieure au seuil de signification. Enfin, quant à la variable HML, celle-ci est significative pour les quintiles un, deux et quatre. La probabilité des quintiles trois et cinq est supérieure au seuil de signification, démontrant donc que pour ces quintiles, cette variable ne contribue pas à expliquer le rendement du mois suivant.

Les travaux de Diether, Malloy et Scherbina (2002) ont démontré que la constante de la régression devenait significativement négative pour les quintiles comportant de plus grands volumes de transactions. Dit autrement, l'intuition étant que plus le volume des transactions est élevé, plus le modèle de Fama-French semble devenir incomplet. Une des raisons possibles est qu'il manque de l'information sur l'importance de la dispersion des prévisions. Le résultat de la présente analyse démontre que ce sont les quintiles avec les plus faibles volumes qui sont les plus significatifs. Il est alors possible de penser que les investisseurs ont maintenant pris en compte les conclusions tirées par Diether, Malloy et Scherbina (2002) dans leurs décisions, mais détiennent maintenant trop de titres démontrant de faibles volumes de transactions, ce qui pourrait expliquer, dans une certaine mesure, leur surévaluation, d'où le fait que le modèle de Fama-French ne peut expliquer une part négative du rendement de ces titres.

6.6. RÉGRESSION AVEC MODÈLE FAMA-FRENCH – RETARD DE DEUX MOIS

Le tableau XIX présente les résultats de la régression effectuée en décalant de deux mois le rendement excédentaire par rapport aux trois variables indépendantes.

Tableau XIX

Statistiques de la régression et analyse de variance – Retard de deux mois

Ce tableau résume les statistiques de la régression effectuée pour la régression avec le modèle des trois facteurs de Fama-French, et ce, avec un retard de deux mois. Les R^2 tendent à diminuer significativement par rapport à la régression du modèle avec un mois de retard. La capacité explicative de ce modèle est faible avec deux mois de retard. Les R^2 diminuent cependant, au fur et à mesure que le volume des transactions augmente, ce qui sert à démontrer qu'il y a d'autres variables que celles du modèle de Fama-French qui expliquent ce qui influence le rendement.

Statistiques de la régression et analyse de variance - Retard de deux mois					
	Quintiles fondés sur le volume				
	Petit				Grand
	1	2	3	4	5
SSR	446,504541502	1378,651420192	617,046009047	604,474695832	1237,540429023
SSE	3072,321916770	10235,807783252	24204,812967754	5032,564253608	26465,371632241
SST	3518,826458272	11614,459203444	24821,858976801	5637,038949440	27702,912061264
Erreur-type	5,717016329	10,664482997	18,595242467	12,951916529	16,779347625
R^2 ajusté	0,099024978	0,089324678	(0,016932782)	0,017955929	0,014182657
MSR	148,834847167	459,550473397	205,682003016	201,491565277	412,513476341
MSE	32,684275710	113,731197592	345,783042396	167,752141787	281,546506726
F	4,553714101	4,040672068	0,594829641	1,201126633	1,465169933
Valeur critique de F	0,005037762	0,009594079	0,620475974	0,326213191	0,229062444
Constante:					
Coefficient	0,047702752	(1,649225440)	(0,960934543)	(1,370923094)	(0,275389692)
Erreur-type	0,584602934	1,110599359	2,168975156	2,237484061	1,715799866
Statistique t	0,081598551	(1,484986847)	(0,443036215)	(0,612707423)	(0,160502222)
Probabilité	0,935139447	0,141041578	0,659105715	0,544686839	0,872829892
Mkt-RF:					
Coefficient	0,435274302	0,779349855	(0,487226116)	0,512767164	0,492311334
Erreur-type	0,130876555	0,245575893	0,482689639	0,446919376	0,384120508
Statistique t	3,325838623	3,173560101	(1,009398331)	1,147337063	1,281658553
Probabilité	0,001258531	0,002061069	0,316260130	0,260310306	0,203114838
SMB:					
Coefficient	(0,448423617)	(0,490725460)	(0,205541619)	(0,098379776)	(0,601980090)
Erreur-type	0,268182681	0,513421033	0,963484607	0,998956443	0,787111698
Statistique t	(1,672082684)	(0,955795396)	(0,213331503)	(0,098482548)	(0,764796268)
Probabilité	0,097833641	0,341735936	0,831688511	0,922204224	0,446308662
HML:					
Coefficient	(0,419783562)	(0,802133760)	0,671193713	(1,599310988)	(1,103004218)
Erreur-type	0,215265282	0,403967863	0,775877902	0,938687327	0,631800014
Statistique t	(1,950075542)	(1,985637558)	0,865076466	(1,703773921)	(1,745812271)
Probabilité	0,054146529	0,050117173	0,389951867	0,098762242	0,084110873

Pour tous les quintiles, les R^2 ajustés sont très faibles, ce qui va dans le sens du ratio de la somme des carrés expliqués par la régression par rapport à la somme des carrés totaux. Ce modèle n'offre pas une grande capacité explicative lorsque la régression est effectuée avec deux mois de retard. Toutefois, il est intéressant de constater qu'au fur et à mesure que le volume des transactions augmente, la capacité explicative diminue. Cela démontre qu'il y a d'autres variables que celles du modèle de Fama-French qui expliquent ce qui influence le rendement et qu'il est fort possible que les investisseurs aient pris en compte les conclusions que suggèrent Diether, Malloy et Scherbina (2002).

La statistique F est toujours supérieure à la valeur critique de F , sauf pour le troisième quintile, celui comportant le moins de titres. Pour le troisième quintile, la régression est donc non significative et aucune des variables ne sera alors significative.

Enfin, la probabilité des trois variables est toujours plus élevée que le seuil de signification, sauf pour les quintiles un et deux de la variable Mkt-RF.

6.7. DISCUSSION

L'intuition sous-jacente à l'utilisation du modèle des trois facteurs de Fama-French est de constater si le modèle explique autant de la variable dépendante au fur et à mesure que le volume de transactions augmente. Les résultats obtenus ci-dessus démontrent que les meilleurs résultats sont obtenus lorsqu'il y a un mois de retard dans le rendement excédentaire. Pour les régressions avec un mois de retard, plus le volume de transactions augmente, moins le modèle tend à expliquer le rendement excédentaire des titres sur le NYSE, ce qui va dans le sens des conclusions du Diether, Malloy et Scherbina (2002) à l'effet que d'autres variables peuvent avoir un effet sur le rendement excédentaire à mesure qu'augmente le volume des transactions. Cette analyse ne permet toutefois pas de conclure que ce qui entre en jeu pour influencer la variable indépendante est nécessairement l'augmentation du volume des transactions. Cependant, cette analyse permet très certainement de démontrer que d'autres variables que celles de Fama-French peuvent influencer le rendement lorsque le volume des transactions augmente. Force est de constater qu'il est très difficile de mesurer l'effet réel du volume des transactions sur le rendement futur des titres.

CONCLUSION

Deux grandes écoles de pensée s'opposent dans les divers concepts analysés dans ce document. D'un côté, il y a l'école de pensée qui croit à une forme quelconque d'efficience des marchés et que toute l'information disponible se reflète d'une façon ou d'une autre dans le prix d'un titre. De l'autre côté, il y a une autre école de pensée, soit l'approche behaviorale, qui adopte la thèse que les prix des titres peuvent être influencés par des biais décisionnels, des erreurs de jugement des investisseurs, des engouements momentanés pour un titre ou une catégorie de titres quelconque, etc.

Hormis les investisseurs ayant accès à une information privilégiée, le reste des investisseurs ont accès en même temps à toute l'information. Ce qui différencie d'un investisseur à l'autre est les réactions de ceux-ci et les décisions différentes qu'ils prendront en fonction de leur interprétation des signaux et des différents biais décisionnels dont ils sont affligés. En effet, les perceptions, les sentiments et les goûts personnels des investisseurs peuvent aussi se rajouter et venir agir sur l'évaluation des titres, et ce, de façon très rapide.

Il ne serait pas exagéré de dire que le prix d'un titre peut être influencé par les deux écoles de pensée de façon concomitante. Lo (2005) suggère une théorie appelée « *Adaptive Markets Hypothesis* » qui rallie les deux écoles de pensée en édictant que compte tenu des conditions du marché, des conditions environnementales, ainsi que la façon personnelle propre à chacun de prendre les décisions, les investisseurs peuvent, à un moment donné, ne se retrouver ni dans le camp de l'efficience des marchés, ni dans le camp des behavioristes. Cette théorie se fonde sur les aspects biologiques de l'être humain, notamment en étudiant la neuropsychologie, et a comme postulat que les individus agissent dans leur propre intérêt personnel, qu'ils font des erreurs et que de leurs erreurs, ils apprennent et s'adaptent, que la concurrence pousse les individus à s'adapter et innover. De façon dynamique, ces forces interagissent par le truchement de la plupart des individus et c'est le résultat de leurs décisions financières qui configure le marché financier. Lo invoque que si l'on prend le marché dans son ensemble, qui inclut l'agrégat de tous les investisseurs, un investisseur pourra se fonder une opinion si le marché fournit des prix adéquats des titres ou non. Toutefois, chacun des investisseurs développe ses propres « règles du pouce » (c.-à-d. des modèles mentaux pour simplifier les décisions) et selon sa théorie, ces règles du pouce évoluent au fil du temps, elles s'adaptent au fur et à mesure que l'utilisation de celles-ci produit des succès ou des échecs. Selon cette théorie, la relation risque-rendement n'en est pas une qui est stable, les primes pour le risque varieront avec le temps et il existe des possibilités d'arbitrage de temps à autre. Les décisions des investisseurs sont donc le résultat d'un ensemble plus ou moins précis de règles du pouce et non le résultat d'une optimisation mathématique précise.

S'il existe un réel phénomène de lien entre le volume des transactions, qui est le reflet des sentiments, des perceptions que se font les investisseurs sur les titres, et le rendement subséquent, cet effet devait ressortir dans une analyse sur un échantillon de plus courte durée, comme celui utilisé dans ce document (de janvier 2008 à mai 2016) et sur la totalité des titres contenus sur une seule bourse de valeurs, soit le NYSE. Un tel phénomène se dira alors robuste s'il se vérifie dans divers échantillons. Le présent document a donc démontré qu'il existe une relation entre le volume des transactions, ou ce qui est convenu d'appeler une divergence d'opinions que se font les investisseurs d'un titre, et le rendement de ces titres. Plus le volume des transactions est élevé, moins

le rendement futur à court terme des titres sera grand. L'horizon temporel observé de ce phénomène et de sa résorption est de quatre mois. Ce phénomène observé va dans le sens des travaux de Miller (1977).

Ce qui est moins clair est l'explication de la raison exacte de ce phénomène. Il pourrait s'agir, tout simplement comme Miller le suggère, que les opinions favorables soient reflétées dans le prix des titres, alors que ceux-ci ne reflèteront pas toutes les opinions défavorables, à cause des coûts pour les ventes à découvert qui en découragent l'utilisation. Mais il pourrait aussi s'expliquer simplement par ce que Tversky et Kahneman ont convenu d'appeler la « conception erronée de la régression », à savoir que tant les titres à la hausse qu'à la baisse tendront à revenir vers la moyenne de rendement du marché. Se fondant sur ce principe, l'effet du phénomène observé ne peut être que temporaire et les rendements des titres tendront tous vers la moyenne.

Enfin, s'il existe toujours de l'incertitude quant aux causes exactes du phénomène observé dans ce document, une chose demeure certaine, soit qu'il peut être très intéressant et profitable, pour des investisseurs voulant un rendement plus grand à court terme, ou pour la gestion d'une portion à court terme d'un portefeuille diversifié, de construire un portefeuille de titres affichant des volumes plus faibles, de façon à profiter du phénomène de rendement subséquent plus élevé à court terme. Certes, les divers coûts de transactions et les impôts que les transactions pourraient générer pourraient rendre moins attrayante une telle stratégie. Les coûts de transactions sont inévitables, ne serait-ce que dans l'écart du prix acheteur-prix vendeur (en anglais, le « *bid-ask spread* »), ce dernier était supérieur pour les titres comportant de plus faibles volumes de transactions. Pour éliminer l'effet des impôts, ces stratégies pourraient être utilisées pour les portefeuilles de titres contenus dans des régimes de revenus différés, tels un régime enregistré d'épargne retraite ou un compte d'épargne libre d'impôt. L'analyse de l'effet des deux contraintes précédentes pourrait faire l'objet d'une recherche future qui les intégrerait dans l'implémentation d'une stratégie de construction d'un portefeuille à court terme fondée sur le choix de titres à plus faible volume de transactions.

BIBLIOGRAPHIE

- Basu, S., « Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. » *Journal of Financial Economics*, vol. 6, Juin-Septembre 1978, pp. 103-126.
- Beaver, William H. et Wayne R. Landsman, « Note on the Behavior of Residual Security Returns for Winner and Loser Portfolios. » *Journal of Financial Economics*, vol. 3, n° 3, décembre 1981, pp. 233-241.
- Bernoulli, Daniel, « Specimen Theoriae Novae de Mensura Sortis. » *Papers of the Imperial Academy of Sciences in Petersburg*, vol. V, 1738, pp. 175-192. Cet article a été traduit du latin à l'anglais par Louise Sommer et publié subséquemment sous « Exposition of a New Theory on the Measurement of Risk. » *Econometrica*, vol. 22, n° 1, janvier 1954, pp. 22-36.
- Cooley, L. Philip et Roenfeldt, L. Rodney, « A Comparative Multivariate Analysis of Factors Affecting Stock Returns, » *The Financial Review*, vol. 10, n° 1, février 1975, pp. 31-41.
- DeBondt, Werner F.M. et Richard Thaler, « Does the stock market overreact? » *Journal of Finance*, vol. 40, n° 3, juillet 1985, pp. 793-805.
- Diether, Karl B., Christopher J. Malloy et Anna Scherbina, « Differences of Opinion and the Cross Section of Stock Returns. » *Journal of Finance*, vol. 57, n° 5, octobre 2002, pp. 2113-2141.
- Fama, Eugene F. et Kenneth R. French, « Common Risk Factors on the Return on Stocks and Bonds. » *Journal of Financial Economics*, vol. 33, n° 1, février 1993, pp. 3-56.
- Fama, Eugene F. et Kenneth R. French, « Multifactor explanations of asset pricing anomalies. » *Journal of Finance*, vol. 51, n° 1, mars 1996, pp. 55-84.
- Gibbons, Michael R., Stephen A. Ross et Jay Shanken, « A test of efficiency of a given portfolio. » *Econometrica* vol. 57, 1989, pp. 1121-1152.
- Grether, David, « Bayes Rule as a Descriptive Model: The Representativeness Heuristic, » *Quarterly Journal of Economics*, vol. 95, n° 3, novembre 1980, pp. 537-557.
- Hong, Harrison et Jeremy C. Stein, « Disagreement and the Stock Market. » *Journal of Economic Perspective*, vol. 21, n° 2, printemps 2007, pp. 109-128.
- Kahneman, Daniel et Amos Tversky, « Subjective Probability: A Judgment of Representativeness Intuitive Prediction. » *Cognitive Psychology*, n° 3, 1972, pp. 430-454.
- Kahneman, Daniel et Amos Tversky, « Prospect Theory: An Analysis of Decision Making Under Risk. » *Econometrica*, vol. 47, n° 2, mars 1979, pp. 263-291.

Kahneman, Daniel et Amos Tversky, « Intuitive Prediction: Biases and Corrective Procedures. » du livre « Judgement Under Uncertainty: Heuristics and Biases » de D. Kahneman, P. Slovic et A. Tversky, Londres: Cambridge University Press, 1982.

Kleidon, A. W., « Stock Prices as Rational Forecasters of Future Cash Flows. » *Working Paper*, Graduate School of Business, Université de Chicago, novembre 1981.

Lee, Charles M.C. et Bhaskaran Swaminathan, « Price Momentum and Trading Volume. » *Journal of Finance*, vol. 55, n° 5, 2000, pp. 2017-2069.

Lintner, John, 1969, « The Aggregation of Investor's Diverse Judgments and Preferences in Purely Competitive Markets, » *Journal of Financial and Quantitative Analysis*, Vol. IV, n° 4, décembre 1969, pp. 347-400.

Lo, Andrew W., « Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. » *Journal of Investment Consulting*, vol. 7, n° 2, 2005, pp. 21-44.

Miller, Edward, « Risk, Uncertainty, and Divergence of Opinion. » *Journal of Finance*, 32(4), 1977, pp. 1151-1168.

Von Neumann, John et Oscar Morgenstern, « Theory of Games and Economic Behavior. » *Princeton University Press*, Princeton, New Jersey, 1944.

Piqueira, Natalia, « Trading Activity, Illiquidity Costs and Stock Returns. » Article non publié, Université de Houston.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id_874967, 2006.

Pratt, Shannon P., « Relationship Between Variability of Past Returns and Levels of Future Returns for Common Stocks, 1926-1960 », puisé de E. Bruce Fredrickson, édition *Frontiers of Investment Analysis*, (Scranton: International Textbook Co., 1971).

Shefrin, Hersh., 2008, « A Behavioral Approach to Asset Pricing. » *Academic Press*, deuxième édition, 2008.

Soldofsky, M. Robert et Miller, L. Roger, « Risk-Premium Curve for Different Classes of Long-Term Securities, 1950-1966 », *Journal of Finance*, Vol. XXIV, n° 3, juin 1969, pp. 429-446, Tableau 1.

Tversky, Amos et Daniel Kahneman, « Judgment under Uncertainty: Heuristics and Biases. » *Science*, New Series, vol. 185, n° 4157, 27 septembre 1974, pp. 1124-1131.