

UNIVERSITÉ DE MONTRÉAL

RÉPUTATION DES ENTREPRISES : UNE RECHERCHE EXPLORATOIRE
SUR L'UTILISATION DES DONNÉES ISSUES DE TWITTER

BERTRAND NEMBOT

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)

Juillet 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

RÉPUTATION DES ENTREPRISES : UNE RECHERCHE EXPLORATOIRE SUR
L'UTILISATION DES DONNÉES ISSUES DE TWITTER

présenté par : NEMBOT Bertrand

en vue de l'obtention du diplôme de : Maitrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. BOURGAULT Mario, Ph. D., président

Mme de MARCELLIS-WARIN Nathalie, Doct., membre et directrice de recherche

M. WARIN Thierry, Ph.D, membre et codirecteur de recherche

M. JOANIS Marcellin, Ph.D, membre

DÉDICACE

À ma mère d'abord, puis famille et amis qui ont contribué à mon développement durant mes années d'études et m'ont soutenu depuis mon tout premier jour à Polytechnique Montréal.

REMERCIEMENTS

Je souhaite remercier en premier lieu le Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), pour m'avoir offert un espace de travail adéquat pour conduire cette recherche. C'est un centre où se rassemblent des chercheurs et étudiants de très grands talents auprès de qui j'ai pu avoir de l'aide : je pense notamment à Carl St-Pierre pour son expertise en statistique, à Lisa Di Jorio pour ses conseils en analyse de sentiment, également à Marine Hadengue, William Sanger, Antoine Troadec pour les précieux échanges de connaissance faite en matière d'extraction et traitement de données sur Twitter.

Je souhaite aussi remercier les membres du jury, à savoir M. Mario Bourgault et M. Marcelin Joanis, pour le temps qu'ils ont pris afin d'évaluer mon mémoire.

Je souhaite enfin remercier ma directrice de recherche et co-directeur de recherche, respectivement Nathalie de Marcellis-Warin et Thierry Warin, pour leur support indescriptible tout au long de la réalisation de ce projet de recherche et également de mes études à Polytechnique Montréal. Je dois en grande partie la réalisation de ce mémoire à leur passion pour l'enseignement, le partage de leur expérience en recherche académique, et leur engouement pour s'attaquer à des problématiques d'actualité. Un grand merci.

Il est possible que j'oublie plusieurs personnes qui m'ont de près ou de loin aidé à compléter ce mémoire, je les prie de bien vouloir m'en excuser.

RÉSUMÉ

La réputation d'une entreprise constitue une importante, voire principale, partie de l'actif immatériel d'une compagnie (De Marcellis-Warin and Teodoresco 2012). La littérature donne de multiples définitions au concept de réputation d'entreprise. Par exemple on peut reprendre celle d'un construit combinant l'image d'entreprise, l'identité d'entreprise et l'identité désirée d'entreprise (Chun 2005). Depuis le début des années 2000, des organisations telles que l'Institut de la Réputation ou encore le magazine Fortune & Money proposent une mesure de la réputation d'entreprise et diffusent annuellement un classement des entreprises selon leur réputation. D'autres organisations s'intéressent à l'évaluation de la valeur de l'image d'entreprise (appelée équité de marque), telles que Brand-Finance, ou Interbrand. Ces dernières diffusent également un classement annuel des entreprises les plus valorisées au monde. Les résultats présentés dans ce mémoire indiquent que les classements produits par ces institutions sont difficilement interprétables et ne semblent pas présenter de lien direct avec l'évolution du cours de bourse d'une entreprise, pourtant la littérature nous indique que la réputation est positivement corrélée avec le cours de bourse d'une entreprise.

De plus, le processus menant à la conception de ces classements peut prendre plusieurs mois de compilation de données et d'analyse. C'est donc dans l'optique de réduire le temps écoulé entre l'enregistrement des performances des compagnies et la conception d'une mesure de la réputation que réside la présente opportunité de recherche : existe-t-il un autre moyen de fournir une mesure de la réputation d'entreprise à une fréquence plus rapprochée que celle offerte présentement (fréquence annuelle)? Serait-ce même possible d'en avoir une mesure en temps réel?

Les réseaux sociaux se présentent comme étant un canal de recherche afin d'améliorer la mesure de la réputation d'entreprise si les interactions entre les consommateurs et les entreprises peuvent être analysées, pas seulement qualitativement mais également quantitativement. Twitter de par sa nature, son nombre d'utilisateurs exponentiellement croissant, et l'accès plus ou moins libre aux données circulant sur sa plateforme, semble être le réseau social adéquat pour conduire une recherche de ce type. La nouvelle mesure de réputation présentée par ce travail de recherche est donc basée sur les données récoltées sur Twitter. Les résultats obtenus démontrent que le nombre de « followers » d'une entreprise, ainsi que le volume quotidien de « tweets », sont utiles à la création d'un « proxy » de la réputation d'entreprise. À l'inverse, l'analyse de sentiment sur les

tweets ne s'avèrent pas être utile dans la création d'un proxy pour la réputation d'entreprise. De plus, la mesure de réputation conçue à partir de Twitter ne peut capter toute l'essence de la réputation d'entreprise, puisqu'une partie non négligeable de la réputation d'entreprise ne peut être captée par les réseaux sociaux : comme par exemple la loyauté des consommateurs composante de l'image de marque, l'identité d'entreprise ou encore l'identité désirée d'entreprise. Cependant, la nouvelle mesure de réputation permet d'observer l'évolution de la réputation d'entreprise au travers du temps et particulièrement la notoriété d'une entreprise. Il s'agit donc d'un outil complémentaire aux évaluations de réputations existantes. .

ABSTRACT

The corporate reputation represents a major part of a company's intangible assets (De Marcellis-Warin and Teodoresco 2012). The literature review has several definitions of the concept of corporate reputation, such as (Chun 2005) who defines corporate reputation as a construct of corporate image, corporate identity, and corporate desired identity. Since the beginning of the 2000's, several institutions such as the Reputation Institute or Fortune & Money magazine offer a measure of corporate reputation and present each year rankings of company based on their corporate reputation. Other institutions, such as Brand-Finance or Interbrand, assess the value of a corporate brand (called brand equity) every year and produce also rankings for the company with the highest brand equity. The results presented in this master thesis indicate that all the rankings produced by the institutions mentioned previously, do not seem to be linked with the stock market. Moreover, the analytical process leading to the rankings can take several months because of the time it takes to gather the data from all the firms.

Thus, it is in the context of reducing the time between performance recording companies and the process leading to the measure of corporate reputation that lies this current research opportunity: is there a way to provide an assessment of corporate reputation at a lower frequency than the one currently provided (annual)? Would it be possible to have a real time measure of reputation? Social media are presented as a paging channel to improve the measurement of corporate reputation if the interactions between consumers and businesses can be analyzed not only qualitatively but also quantitatively. Twitter seems to be the adequate social media platform to conduct such a research, because of its exponential growing number of users, and more or less easy access to data diffused on it. The new measure of corporate reputation presented through this research is only based on the data Twitter can offer. The results presented in this master thesis indicate on one hand, that the number of followers of a company and the daily number of tweets are useful in creating a proxy of corporate reputation. On the other hand, the results indicate that sentiment analysis on tweets does not seem to provide useful information for the conception of a proxy of corporate reputation. In addition, the new measure of corporate reputation does not pretend to be able to capture the essence of corporate reputation, since a significant part of the corporate reputation cannot be captured by Twitter: including consumer loyalty component of branding, corporate identity or the desired identity of a company. However,

the new measure of corporate reputation enables us to observe the evolution of corporate reputation over time and especially the notoriety of a company. It is therefore a complimentary tool for the assessment of existing reputations, and it is absolutely not a self-sufficient tool.

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES TABLEAUX.....	XIII
LISTE DES FIGURES	XVI
LISTE DES SIGLES ET ABRÉVIATIONS	XVIII
LISTE DES ANNEXES	XIX
INTRODUCTION.....	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	4
1.1 Réputation de l’entreprise et image de marque	4
1.1.1 L’image de marque d’une entreprise.....	4
1.1.2 La réputation d’entreprise	10
1.2 Réputation d’entreprise et réseaux sociaux	14
1.2.1 Le “bouche-à-oreille”	14
1.2.2 Le bouche-à-oreille électronique.....	16
1.2.3 Observer la réputation d’entreprise au travers des réseaux sociaux.....	19
1.2.4 Résumé du lien existant entre réseaux sociaux et réputation d’entreprise	23
1.3 Analyse de sentiments	24
1.3.1 Définition du concept d’analyse de sentiments.....	24
1.3.2 L’analyse de sentiment appliquée sur les données issues de Twitter.....	29
1.3.3 Des illustrations concrètes de l’applications de l’analyse de sentiment avec Twitter.....	32

1.3.4	Développement d'outils en ligne d'analyse de sentiment.....	37
1.3.5	Résumé de la littérature sur le concept d'analyse de sentiments	39
CHAPITRE 2 PROBLÉMATIQUE ET CADRE DE RECHERCHE		40
2.1	Objectif général de la recherche.....	40
2.2	Mesures actuelles de la réputation d'entreprise	41
2.3	Une application des données de Twitter à la réputation d'entreprise	43
2.4	Cadre conceptuel : une représentation graphique des hypothèses	44
CHAPITRE 3 LES MESURES ACTUELLES DE LA RÉPUTATION D'ENTREPRISE		47
3.1	Présentation des mesures de la réputation d'entreprise et évaluation d'image de marque d'une entreprise.....	47
3.1.1	Interbrand	48
3.1.2	Brand-Finance	50
3.1.3	Le magazine Corporate Knights.....	51
3.1.4	Fortune	53
3.1.5	L'Institut de la Réputation.....	54
3.2	Informations financières sur les marchés	54
3.3	Remarques importantes sur la méthodologie	55
3.3.1	Difficulté concernant la construction de l'échantillon d'entreprises	55
3.3.2	Difficulté concernant le recensement de la valeur d'équité de marque	55
3.4	Composition de l'échantillon d'entreprises.....	55
3.5	Illustration des bases de données	57
3.6	Traitement des données	58
3.6.1	Définition du critère de comparaison	58
3.6.2	Modèle logit	60
3.7	Mesures actuelles de la réputation et cours de bourse : existe-t-il un lien?	60

3.7.1	Les mesures de la réputation	61
3.7.2	Les évaluations monétaires de l'image de marque.....	63
3.7.3	La mesure de responsabilité sociale	64
3.8	Discussions.....	65
3.8.1	Discussions sur les résultats obtenus.....	65
3.8.2	Discussion sur la méthodologie employée	67
3.9	Conclusion.....	68
CHAPITRE 4 DÉTERMINATION DES DONNÉES ISSUES DE TWITTER UTILES À LA CONCEPTION D'UNE MESURE DE LA RÉPUTATION		69
4.1	Retour sur la revue de littérature	69
4.2	Processus de récolte de données sur Twitter	70
4.3	Description des variables de la mesure de réputation	73
4.3.1	Notoriété de la marque	73
4.3.2	Qualité perçue de la marque.....	73
4.3.3	Associations	74
4.3.4	Loyauté des consommateurs	75
4.3.5	En résumé.....	75
4.4	Volume quotidien de tweets.....	76
4.5	Informations financières sur les marchés.....	77
4.6	Remarques importantes sur la méthodologie	77
4.7	Composition de l'échantillon d'entreprises.....	78
4.8	Illustration de la base de données.....	78
4.9	Traitement des données : régression linéaire	79
4.10	Analyse des résultats	80
4.10.1	Tests préliminaires	80

4.10.2	Régression sur l'ensemble de la base de données	82
4.10.3	Régression par industrie	84
4.10.4	Régression par niveau de prix de l'action	87
4.11	Discussions.....	91
CHAPITRE 5 PROPOSITION D'UNE MESURE DE LA RÉPUTATION D'ENTREPRISE		93
5.1	La mesure de la réputation d'entreprise à partir de Twitter	93
5.2	Traitement des données : modèle logit.....	94
5.3	Analyse des résultats	95
CHAPITRE 6 DISCUSSION, LIMITES ET AVENUES DE LA RECHERCHE		100
6.1	Contribution théorique et pratique de la recherche	100
6.2	Limites de la recherche	101
6.3	Avenues de la recherche.....	102
CONCLUSION		103
BIBLIOGRAPHIE		105
ANNEXES		114

LISTE DES TABLEAUX

Tableau 3.1 Description sommaire des données récoltées au travers des cinq sources d'évaluation de la réputation	56
Tableau 3.2 Modèle logit avec le magazine Fortune pour variable indépendante	61
Tableau 3.3 Modèle logit avec l'Institut de la Réputation pour variable indépendante	62
Tableau 3.4 Modèle logit avec Interbrand pour variable indépendante	63
Tableau 3.5 Modèle logit avec Brand-Finance pour variable indépendante	64
Tableau 3.6 Modèle logit avec Corporate Knight pour variable indépendante	65
Tableau 4.1 Proxys Twitter des indicateurs de l'équité de marque	76
Tableau 5.1 Résultats de la régression logistique du modèle 1	96
Tableau 5.2 Résultats de la régression logistique du modèle 2	97
Tableau 5.3 Résultats de la régression logistique du modèle 3	98
Tableau A1.1 Description des variables indépendantes et de la variable dépendante	115
Tableau A1.2 Skewness et Kurtosis avant	116
Tableau A1.3 Skewness et kurtosis après transformé logarithme naturel (ln)	117
Tableau A2.1 Matrice de corrélation globale	119
Tableau A2.2 Tableau de référence dans l'assignation d'une lettre à chacune des 26 industries, pour simplifier la nomination dans les tableaux subséquents	120
Tableau A2.3 Matrice de corrélation pour l'industrie « Airlines »	122
Tableau A2.4 Matrice de corrélation pour l'industrie « Apparel »	123
Tableau A2.5 Matrice de corrélation pour l'industrie « Beverages »	124
Tableau A2.6 Matrice de corrélation pour l'industrie « «Chemicals »	125
Tableau A2.7 Matrice de corrélation pour l'industrie « Computer software »	126
Tableau A2.8 Matrice de corrélation pour l'industrie « Computers »	127
Tableau A2.9 Matrice de corrélation pour l'industrie « Consumer Food Products »	128

Tableau A2.10 Matrice de corrélation pour l'industrie « Electric and Gas utilities ».....	129
Tableau A2.11 Matrice de corrélation pour l'industrie « Electronics »	130
Tableau A2.12 Matrice de corrélation pour l'industrie « Energy ».....	131
Tableau A2.13 Matrice de corrélation pour l'industrie « Entertainment ».....	132
Tableau A2.14 Matrice de corrélation pour l'industrie « Food and Drug Stores »	133
Tableau A2.15 Matrice de corrélation pour l'industrie « Food Services ».....	134
Tableau A2.16 Matrice de corrélation pour l'industrie « General Merchandisers »	135
Tableau A2.17 Matrice de corrélation pour l'industrie « Health Care : Insurance and Managed Care »	136
Tableau A2.18 Matrice de corrélation pour l'industrie « Insurance : Life and Health ».....	137
Tableau A2.19 Matrice de corrélation pour l'industrie « Insurance : Property and Casualty » ...	138
Tableau A2.20 Matrice de corrélation pour l'industrie « Megabanks »	139
Tableau A2.21 Matrice de corrélation pour l'industrie « Metals ».....	140
Tableau A2.22 Matrice de corrélation pour l'industrie « Mining, Crude-Oil Production ».....	141
Tableau A2.23 Matrice de corrélation pour l'industrie « Motor Vehicles »	142
Tableau A2.24 Matrice de corrélation pour l'industrie « Petroleum Refining »	143
Tableau A2.25 Matrice de corrélation pour l'industrie « Pharmaceuticals »	144
Tableau A2.26 Matrice de corrélation pour l'industrie « Soaps and Cosmetics »	145
Tableau A2.27 Matrice de corrélation pour l'industrie « Superregional Banks ».....	146
Tableau A2.28 Matrice de corrélation pour l'industrie « Telecommunications ».....	147
Tableau A3.1 Multicolinéarité initiale	148
Tableau A3.2 Multicolinéarité après ajustement (omission de tneutrelog)	149
Tableau A5.1 Régression linéaire globale associé au modèle 1	151
Tableau A5.2 Régression linéaire globale associée au modèle 2.....	153
Tableau A6.1 Régression linéaire avant ajustement du modèle 1 par industrie	155

Tableau A6.2 Régression linéaire après ajustement du modèle 1	158
Tableau A6.3 Régression linéaire avant ajustement du modèle 2.....	161
Tableau A6.4 Régression linéaire après ajustement du modèle 2.....	164
Tableau A6.5 Signe des coefficients des regressions linéaires pour les variables explicatives significatives au seuil $p < 0.20$ par industrie (modèle 1 et 2 combinés)	169
Tableau A7.1 Régression linéaire par niveau de prix de l'action, modèle 1	170
Tableau A7.2 Régression linéaire par niveau de prix de l'action avec un lag d'une journée, modèle 1	171
Tableau A7.3 Régression linéaire avec interaction entre le niveau de prix de l'action et le volume de tweet, modèle 1.....	172
Tableau A7.4 Régression linéaire par niveau de prix de l'action, modèle 2	173
Tableau A7.5 Régression linéaire par niveau du prix de l'action avec un lag d'une journée, modèle 2.....	174
Tableau A7.6 Régression linéaire avec interaction entre le niveau de prix de l'action et le nombre de followers, modèle 2	175
Tableau A7.7 Résumé de la significativité des variables indépendantes en dessous d'un seuil $p < 0.20$ (modèle 1 et 2 sans lag).....	176
Tableau A7.8 Résumé de la significativité des variables indépendantes en dessous d'un seuil $p < 0.20$ (modèle 1 et 2 avec lag)	177
Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie.	178
Tableau A8.2 Industries présentes dans l'échantillon	182
Tableau A8.3 Liste des 101 firmes composant l'échantillon.....	184
Tableau A9.1 Référence des compagnies sur Twitter.....	191
Tableau A10.1 Régression logistique avec interactions entre le modèle 2 de la réputation par Twitter et les industries	197

LISTE DES FIGURES

Figure 1.1 Le modèle d'équité de la marque (adapté de Aaker 2009)	7
Figure 1.2 Le modèle pyramidal (adapté de Tuominen 1999)	9
Figure 1.3 Le schéma de la réputation d'entreprise selon (adapté de Chun 2005)	11
Figure 1.4. Le Top 10 des compagnies avec la plus importante équité de marque (Oxford Metrica 2011).....	12
Figure 1.5 Modèle illustrant la relation entre l'image de marque et les micro-blogs adapté de (Esch et al. 2006, annoté par Jansen et al. 2009)	18
Figure 1.6 Modèle de la réputation d'entreprise en temps de crise (Warin et al. 2013)	20
Figure 1.7 La matrice de mesure et gestion de la réputation sur le web (adapté de Jones, Temperley, and Lima 2009)	21
Figure 1.8 La formule PMI-IR pour déterminer l'orientation sémantique des phrases (Turney 2002).....	24
Figure 1.9 Extraction de la subjectivité dans la classification de document réalisée par (Pang and Lee 2004).....	26
Figure 1.10 Architecture du système développé par Kim and Hovy 2004, pour la classification de la polarité au niveau de la phrase	27
Figure 1.11 Architecture du système développé par Hiu et Liu dans l'analyse des avis déposés par des consommateurs sur des produits (adapté de Hiu et Liu 2004).....	28
Figure 1.12 Le modèle graphique Tree Kernel appliqué au tweet: “@Fernando this isn't a great day for playing the HARP! :) “	30
Figure 1.13 Types d'options utilisées par (Agarwal et al. 2011) pour l'analyse de sentiment	31
Figure 1.14 Le modèle graphique d'Hashtag représentant les liens avec le hashtag #ipad (adapté de Wang et al. 2011)	32
Figure 2.1 Cadre conceptuel de la recherche	46
Figure 3.1 Modèle d'évaluation de l'équité de marque d'Interbrand (adapté d'Interbrand).....	48

Figure 3.2 Méthodologie d'évaluation de l'équité de marque de Brand-Finance (adapté de Brand-Finance).....	50
Figure 3.3 Illustration d'une portion de la base de donnée pour la première partie d'analyse	57
Figure 4.1 Démarche de réflexion dans la méthodologie de la seconde partie du mémoire.....	69
Figure 4.2 Illustration de la structure de requête au travers du REST API (adapté de Twitter)	71
Figure 4.3 Illustration de la structure de requête au travers du Streaming API (adapté de Twitter)	71
Figure 4.4 Aperçu de la base de donnée conçue avec les informations récoltées sur Twitter	79
Figure A4.1 Test d'Hauffman pour les régressions linéaires	150
Figure A11.1 Volume de tweets quotidien moyen des 101 entreprises composant l'échantillon	200

LISTE DES SIGLES ET ABRÉVIATIONS

API	<i>Application Programming Interface</i> ou Interface de programmation
BAO	Bouche-à-oreille
BAON	Bouche-à-oreille négatif
BAOP	Bouche-à-oreille positif
CIRANO	Centre Interuniversitaire de Recherche en Analyse des Organisations
eBAO	Bouche-à-oreille électronique
IR	Institut de la Réputation
POS	Part-of-speech
S&P	Standard and Poor's
SVM	Support Vector Machine
VIF	Variance inflator factor
WEKA	Waikato Environment Knowledge Analysis

LISTE DES ANNEXES

ANNEXE 1 – TEST DE NORMALITÉ	115
ANNEXE 2 – MATRICE DE CORRÉLATION	119
ANNEXE 3 – TEST DE MULTICOLINÉARITÉ	148
ANNEXE 4 – TEST D’HAUFFMAN	150
ANNEXE 5 – RÉGRESSION LINÉAIRE GLOBALE.....	151
ANNEXE 6 – RÉGRESSION LINÉAIRE PAR INDUSTRIE	155
ANNEXE 7 – RÉGRESSION LINÉAIRE PAR NIVEAU DE PRIX.....	170
ANNEXE 8 – RÉSUMÉ DES DONNÉES RÉCOLTÉES AU TRAVERS DES SOURCES DE MESURE DE RÉPUTATION	178
ANNEXE 9 – RÉFÉRENCE DES COMPAGNIES SUR TWITTER.....	191
ANNEXE 10 – NOUVELLE MESURE DE LA RÉPUTATION PAR TWITTER ET INDUSTRIES.....	197
ANNEXE 11– VOLUME DE TWEETS MOYEN POUR LES 101 ENTREPRISES DE L’ÉCHANTILLON.....	200

INTRODUCTION

Les entreprises sont toutes composées d'actifs matériels (ou tangibles) et d'actifs immatériels (ou intangibles). Pour les plus importantes firmes en terme de capitalisation boursière, regroupées dans le S&P500, la valeur des actifs immatériels surpasse la valeur des actifs matériels (Oxford Metrica 2011). Au sein des actifs immatériels, on distingue des éléments tels que les brevets, les droits d'auteur mais principalement la réputation de l'entreprise. La réputation d'entreprise sera définie dans la revue de littérature, cependant son importance dans la survie des opérations à long terme est à souligner : son principal atout est d'offrir à l'entreprise un avantage comparatif vis-à-vis de ses compétiteurs dans ses activités opérationnelles d'abord et activités financières ensuite. Les avantages qu'elle procure justifient la nécessité pour les entreprises d'en connaître sa valeur. Quelques institutions se sont penchées sur la tâche dont les deux sources les plus connues: l'Institut de la Réputation, et le magazine Fortune & Money. Cependant, les mesures de réputation qu'elles proposent sont évaluées sur une base annuelle, ce qui n'offre pas la possibilité aux entreprises de réagir rapidement en cas d'érosion de la valeur de leur principal actif intangible. Existe t-il un moyen d'estimer la valeur de la réputation d'entreprise à une fréquence plus rapprochée? Telle est l'interrogation à laquelle cette recherche exploratoire tentera d'apporter une contribution.

L'ère du Web 2.0 a introduit les réseaux sociaux au travers desquels il est aujourd'hui facile d'échanger de l'information (Madden 2007), de discuter de nos goûts, nos passions, nos inquiétudes, nos opinions sur divers sujets relevant du domaine politique, du domaine privé (rendu publique), du domaine financier etc. Les réseaux sociaux sont également une mine d'or regorgeant d'informations sur les perceptions ou goûts des consommateurs vis-à-vis de marques, de produits, ou de services et ceci de manière quotidienne. Les réseaux sociaux représentent peut-être un canal de recherche menant à la réponse à l'interrogation posée précédemment. L'un de ses réseaux sociaux se nomme Twitter où environ 241 millions d'utilisateurs actifs¹ (2013) ont accès à la plateforme sur leurs téléphones intelligents, tablettes ou ordinateurs. Twitter² a été fondé par

¹ <https://about.twitter.com/company>

² <https://about.twitter.com/company>

Jack Dorsey, Evan Williams, Biz Stone et Noah Glass en 2006. La compagnie a ensuite été incorporée le 19 avril 2007. Twitter inc. possède environ 2000 employés répartis au travers d'une vingtaine de bureaux au travers de le monde. La compagnie est récemment entrée en bourse le 7 Novembre 2013, atteignant une capitalisation boursière d'environ 30 milliards de dollar US (pour fin de comparaison Facebook a actuellement une capitalisation boursière de 115 milliards de dollar US). Il s'agit d'un réseau social offrant une plateforme d'échange d'information où chaque message posté (aussi appelé tweet) est limité par 140 caractères. Les quelques particularités de Twitter consistent en l'utilisation de hastags (#) faisant référence à un sujet, un thème, une tendance, ou une émotion particulière. L'utilisation du symbole @ sert à identifier un destinataire particulier pour un tweet. Il est possible de partager un tweet d'un utilisateur, cette action est appelé « retweet » et représenté par l'acronyme « RT ». Chaque utilisateur peut suivre le fil d'actualité d'un autre utilisateur. Dans ce cas l'utilisateur 1 qui suit un utilisateur 2 est caractérisé comme étant un « follower » du point de vue de l'utilisateur 2. Et d'un point de vue du l'utilisateur 1, l'utilisateur 2 est caractérisé comme étant un « followee ».

Twitter se présente être un réseau social propice à la conduite de la présente recherche dont la problématique et les concepts qui s'y rattachent seront définis un peu plus loin dans ce document. En effet, Twitter rend l'accès à ses données plus ou moins facile à obtenir. De plus, la compagnie a dédié un espace sur le web pour les chercheurs et développeurs : <https://dev.twitter.com/>.

Ainsi, le but de la recherche présentée au travers de ce mémoire est d'examiner la possibilité de concevoir une mesure d'évaluation de la réputation des entreprises en utilisant les données issues de Twitter.

Ce mémoire est organisé comme suit. Le chapitre 1 présente une revue de la littérature définissant le cadre de recherche dans lequel la présente étude s'insère. Plusieurs concepts sont rassemblés dans la conduite de la recherche, ce qui nécessite d'explicitement le lien entre les concepts, mais également de soulever les enjeux qui y sont liés. Les principaux concepts se rattachant à la recherche sont : la réputation, l'image de marque, les réseaux sociaux (Twitter) et l'analyse de sentiment. Nous avons préféré présenter la revue de la littérature avant de présenter la problématique de recherche, car celle-ci a été affinée en se basant sur les travaux recensés dans la littérature. Le chapitre 2 présente la question de recherche et les deux sous-questions de recherche associées, ainsi que les hypothèses découlant de la littérature qui s'y rattachent. Le

chapitre 3 présente les analyses effectuées afin de déterminer la meilleure source existante de mesure de la réputation. Le chapitre 4 présente les analyses effectuées afin de déterminer quelles informations offertes par Twitter sont pertinentes dans la construction d'une mesure de réputation pour entreprise. Le chapitre 5 détaille la conception de la mesure de réputation pour entreprises à partir de Twitter. La mesure est basée sur les résultats obtenus au chapitre 4. Une discussion sur les contributions théoriques et pratiques, ainsi que sur les limites de la recherche est apportée au chapitre 6. Finalement, nous concluons cette étude en incluant les travaux futurs nécessaires à l'amélioration du modèle présenté.

CHAPITRE 1 REVUE DE LITTÉRATURE

Ce chapitre présente la revue de littérature associée aux différents thèmes de la recherche. Les thèmes abordés regroupent la réputation d'entreprise, l'image de marque, les réseaux sociaux, et l'analyse de sentiment. La revue de la littérature est organisée comme suit : la première partie de la revue de la littérature définira les concepts connexes de réputation d'entreprise et d'image de marque. Ces deux concepts sont au cœur du travail de recherche qui est présenté dans ce mémoire d'où la nécessité de les présenter en premier. La seconde partie de la revue de la littérature présentera les travaux de recherche faisant le lien entre la réputation d'entreprise et les réseaux sociaux. Comme indiqué en introduction, nous utiliserons le réseau social Twitter pour tenter d'apporter une réponse quant à la faisabilité de créer un indice de réputation mesurable à une fréquence plus élevée qu'annuellement : d'où l'importance d'aborder le thème des réseaux sociaux dans la revue de littérature. Enfin, la troisième partie de la revue de la littérature s'intéressera à l'analyse de sentiment : une discipline nécessaire au traitement des données en format texte, telles que celles offertes par Twitter.

1.1 Réputation de l'entreprise et image de marque

Il est important de faire la distinction entre les concepts de « marque corporative » (également appelé « image corporative » ou « image de marque ») et de « réputation d'entreprise ». La revue de littérature est assez vaste sur le sujet, ce qui rend difficile d'obtenir une définition universelle pour chacun de ces deux concepts. Nous allons débiter par définir le concept d'image de marque avant de définir le concept de réputation d'entreprise, car comme nous allons le voir dans les prochaines sections l'image de marque est une partie intégrante de la réputation d'entreprise.

1.1.1 L'image de marque d'une entreprise

1.1.1.1 Définitions de l'image de marque

(King 1991; Knox and Bickerton 2003), distinguent deux dimensions à l'intérieur du concept de marque : la marque de produit et la marque corporative. La marque de produit fait référence à la perception des consommateurs envers un certain produit (ou service) et de son bénéfice allant au delà de ses performances fonctionnelles et de son prix. En ce qui concerne la marque corporative, il semble exister plusieurs définitions dans la revue de littérature. (Abratt 1989) mentionne que l'image corporative est « le point de contact entre les différents actionnaires et la compagnie ».

Ainsi l'image corporative peut être estimée au travers des expériences rencontrées par les actionnaires avec la compagnie. En d'autres termes, l'attitude des actionnaires envers la compagnie définit l'image corporative que possède la compagnie. Selon d'autres sources, la marque corporative peut être assimilée à l'image de l'entreprise ou encore à « comment est-ce que les autres nous perçoivent en tant que compagnie ? » (Chun 2005). La marque corporative est donc au niveau de l'organisation plutôt qu'au niveau des produits. Similairement à Chun, (Liou and Chuang 2010) définissent la marque corporative comme « la photographie mentale d'une compagnie soutenue par une audience, c'est-à-dire ce qui vient à l'esprit lorsque que quelqu'un voit ou entend le nom de l'entreprise ou son logo ».

Dans le livre intitulé « *Brand Culture* » (Schroeder and Morling 2005), John Balmer, professeur à l'université Brunel à Londres, a écrit un chapitre au sein duquel il caractérise la marque corporative selon quatre aspects. Le premier aspect est que les marques dénotent une propriété, c'est-à-dire une marque déposée possédée par une entité. Le deuxième aspect est que les marques servent d'outils à fabriquer des images dans l'esprit des consommateurs. Le troisième aspect est que les marques sont des symboles associés à des valeurs clés. Et finalement le quatrième aspect est que les marques sont un moyen par lequel les consommateurs construisent leur identité individuelle (i.e. les consommateurs s'approprient les marques dans le but de définir qui ils sont).

(Aaker 2004) enrichit la définition de la marque corporative en identifiant certains aspects que possède toute marque. Ces aspects sont les suivant : un héritage (les origines de la marque, ce qui la rendait initialement si spéciale), des actifs et capacités, des personnes (les employés sont au cœur de l'image corporative puisqu'ils représentent le point de contact entre la firme et ses clients), des valeurs et des priorités, une référence locale ou internationale (notoriété de la marque), des programmes de responsabilité sociale (Est-ce que la firme respecte l'environnement? Possède-t-elle une bonne diversité parmi ses employés? Est-ce que les employés sont de nature cordiale?), et finalement un historique de performance (est-ce que la compagnie connaît un succès sur le plan financier et opérationnel?).

1.1.1.2 Mesure de l'image de marque

Pour que la gestion de marque soit efficace, les entreprises ont besoin d'être en mesure d'évaluer la valeur de leur marque aussi appelée l' « équité de la marque ». (Aaker 2009) définit l'équité de

la marque comme étant l' « ensemble des actifs et passifs reliés à la marque (son nom, ses symboles) qui apportent de la valeur, ou retirent de la valeur à un produit ou un service. Les actifs incluent la loyauté vis-à-vis de la marque, la notoriété de la marque, la qualité perçue des consommateurs vis-à-vis de la marque et les associations faites avec la marque ». (Aaker 2009) soulève également plusieurs avantages de ces actifs. Une marque suscitant une forte loyauté au sein de son bassin de clientèle voit ses coûts marketing diminuer au fil du temps, attire davantage de nouveaux clients et augmente le délai maximal de temps de réponse lorsqu'une menace au niveau de sa compétition se présente. Une marque avec une forte notoriété conduit à une certaine familiarité parmi les consommateurs; il s'agit également d'un signal de substance et indique aux consommateurs que la marque est à considérer dans ses options d'achat. Plus une marque possède une excellente qualité perçue chez les consommateurs, plus la marque est en mesure de pouvoir se différencier de ses concurrents, et plus elle donne des arguments aux consommateurs pour acheter le produit (ou service) auprès de la marque. De plus, la marque peut charger à ses consommateurs un premium plus important sur le prix de vente de ses produits (ou services). Les associations d'une marque représentent les liens qu'une marque possède avec une idéologie, des valeurs, ou des causes particulières. Les associations d'une marque augmentent la différenciation d'une marque vis-à-vis de ses concurrents, elles permettent également à la firme de susciter des sentiments positifs concernant sa marque chez les consommateurs. Les associations d'une marque permettent aussi de donner des raisons supplémentaires aux consommateurs de choisir un produit (ou service) plutôt qu'un autre. Le modèle d'équité de la marque de D. A Aaker est présenté ci-après.

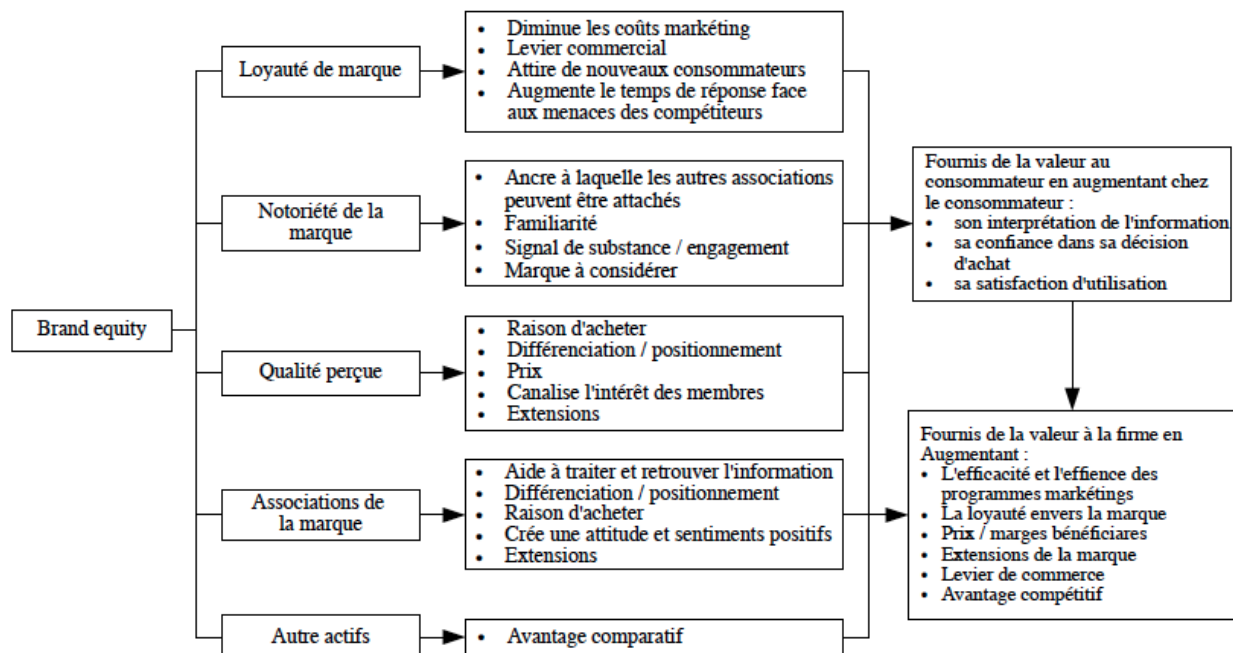


Figure 1.1 Le modèle d'équité de la marque (adapté de Aaker 2009)

Le concept d'équité de la marque a été étudié sous trois différentes perspectives: la perspective du client, la perspective de la firme, et la perspective financière. Sous la perspective client, les différentes variables conduisant à l'équité de la marque sont : la notoriété, les associations de la marque, l'attitude de la marque, l'attachement à la marque (loyauté), et l'activité de la marque (Aaker 2009). Sous la perspective de la firme, l'équité de la marque peut être mesurée au travers : de l'élasticité du prix et de la part de marché au sein de l'industrie, des premiums instaurés sur les prix, des réactions du public face aux campagnes publicitaires de la firme, et du degré d'acceptation que la firme possède face aux mauvaises critiques (Hoeffler and Keller 2003). Sous la perspective financière, il existe plusieurs approches dans l'évaluation de l'équité de la marque. La première est le quotient de Tobin, qui est le ratio de la valeur marchande d'une firme sur sa valeur comptable. (Larkin 2002) considère une approche assez similaire utilisant la différence entre la valeur comptable et la capitalisation boursière dans le but d'extraire la valeur de la réputation d'entreprise dont l'équité de la marque en constitue une partie intégrante. Les firmes avec une haute valeur d'équité de marque auront tendance à avoir un quotient de Tobin supérieur à 1, tandis que les firmes avec une faible valeur d'équité de marque auront tendance à avoir un quotient de Tobin proche ou inférieur à 1 (Lindenberg and Ross 1981). Une autre méthode

utilisée par (Aaker and Jacobson 1994) indique une corrélation positive entre la qualité perçue des consommateurs et les performances boursières de l'entreprise. Ainsi ils peuvent mesurer l'équité d'une marque en se basant sur la qualité perçue des consommateurs.

Il existe également deux sources qui offrent chaque année une approximation de la valeur de l'image corporative : Interbrand et Brand-Finance. Ces deux organisations sont décrites à la section 4.

1.1.1.3 Impact de l'image corporative

(K. L. Keller and Lehmann 2006) indiquent que les marques ont des impacts sur trois différents niveaux : le marché des consommateurs, le marché des produits, et les marchés financiers. La gestion d'une marque est importante car elle crée un avantage compétitif pour les firmes qui sont bien perçues par le public. En effet, les marques corporatives à fortes valeurs permettent à certaines firmes d'imposer un premium sur le prix de vente de leurs produits comparativement à leurs concurrents à plus faible valeur de marque corporative. De plus, une marque corporative forte donne la capacité à certaines firmes d'avoir une durée de vie plus longue dans la phase de déclin du cycle de vie d'un produit (Tuominen 1999).

Il est donc important pour les firmes de développer des nouvelles stratégies pour rehausser la valeur de leurs marques. Cette considération a conduit à l'introduction d'un nouveau champ de compétences : la gestion de marque (Gray and Balmer 1998). (Park, Jaworski, and MacInnis 1986) ont développé un concept de gestion de marque (« brand concept management » en anglais ou BCM) qui consiste en trois phases distinctes : la phase d'introduction, la phase d'élaboration et la phase de renforcement. À chaque phase correspond un ensemble de stratégies différentes pour développer la marque. Les stratégies suggérées par le BCM dépendent également du concept de marque sur lequel l'image de marque repose. En effet, il existe trois principaux concepts de marque : le concept de marque fonctionnelle, le concept de marque symbolique, et le concept de marque expérimentale. Le concept de marque fonctionnelle fait référence à une marque qui répond à un besoin fonctionnelle d'un client, par exemple la marque Kleenex qui est utilisée pour les mouchoirs qui est devenue une marque dite éponyme puisque utilisée pour nommer en général les mouchoirs dans la langue française. Le concept de marque symbolique fait référence à une marque qui possède un symbole particulier chez le consommateur, par

exemple la marque Louis-Vuitton (ou autres marques de luxe) suscite un symbole de prestige chez le consommateur. Le concept de marque expérimentale fait référence à une marque qui sait susciter des plaisirs sensorielles chez le consommateur : par exemple la marque Apple, dont ses produits au design simplifié procurent une expérience utilisateur qui se démarque de la compétition.

(Tuominen 1999) a décrit un modèle pyramidal de marque pour analyser le concept d'identité de marque. Le modèle est présenté ci-dessous (voir figure 1.2). Il est composé de trois niveaux : le cœur de la marque, le style de la marque qui « articule le cœur de la marque en termes de culture qu'elle émane, de personnalité, d'image personnelle », et enfin le niveau des thèmes de la marque qui font référence à comment est ce que la marque communique avec les consommateurs. Trois paramètres caractérisent le niveau des thèmes de la marque : le physique de la marque (son logo, ses couleurs), le reflet de la marque (le type de personne utilisée dans les campagnes publicitaires) et la relation avec les consommateurs.

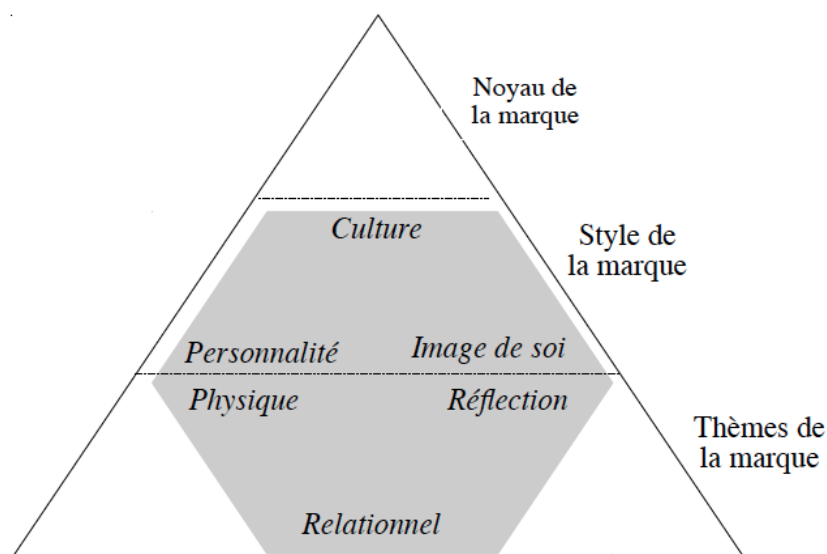


Figure 1.2 Le modèle pyramidal (adapté de Tuominen 1999)

1.1.1.4 Résumé de la revue de littérature sur le concept d'image corporative

Pour résumer, la marque corporative d'une entreprise est constituée des différentes perceptions entretenues par les consommateurs et les actionnaires résultant de la nature et qualité des activités entreprises par une compagnie. La marque corporative influence directement les performances

financières d'une entreprise aux niveaux par exemple des prix des produits, des élasticités de la demande de son bassin de clientèle, ou encore de l'acquisition de nouveaux clients. La gestion de marque permet à l'entreprise d'influencer positivement la perceptions des consommateurs et des actionnaires au travers de multiple stratégies : expérience cliente, associations symboliques, ciblage des besoins fonctionnels du client.

1.1.2 La réputation d'entreprise

1.1.2.1 Définitions de la réputation d'entreprise

(Fombrun 1996) mentionne que la réputation d'entreprise « représente les jugements cumulatifs du public au travers du temps ». (Balmer et al. 2002) sont également en accord avec la définition précédente; selon eux les notions de réputation d'entreprise et d'image corporative réfèrent toutes deux à la perception des individus envers la compagnie. Cependant, s'il s'agit d'une perception à un point donné du temps, la notion d'image corporative devrait être employée. Mais s'il s'agit d'une perception au travers du temps la notion de réputation d'entreprise est plus appropriée. (Weigelt and Camerer 1988) définissent la réputation d'entreprise comme « un ensemble d'attributs d'une firme, inférés de ses actions passées ». Pour (Chun 2005), la réputation d'entreprise est un construit composé de trois éléments distincts : (1) l'identité de l'entreprise, (2) l'image de l'entreprise et (3) l'identité désirée de l'entreprise. Le premier élément fait référence à « comment une compagnie se perçoit elle-même? ». Le deuxième élément fait référence à « comment est-ce que les autres (extérieurs à la compagnie) nous perçoivent? ». Le troisième élément fait référence à « comment est-ce que la compagnie veut que les autres la perçoivent? ». Le modèle de Chun est représenté à la figure suivante (figure 1.3).

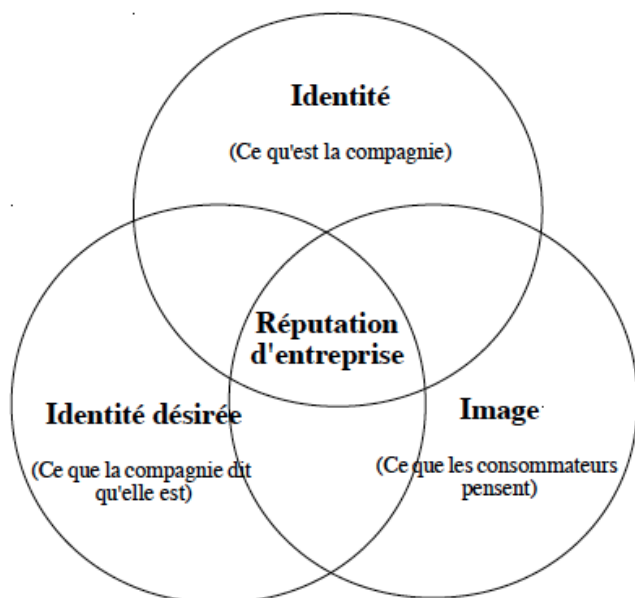


Figure 1.3 Le schéma de la réputation d'entreprise selon (adapté de Chun 2005)

(De Marcellis-Warin and Teodoresco 2012) apportent une autre définition de la réputation d'entreprise basée sur leurs recherches dans la revue de littérature. Ils résument la réputation d'entreprise comme étant « un actif intangible qui est construit au travers du temps et représente la valeur et la confiance que les actionnaires possèdent envers la compagnie ». La notion d'actif intangible attribué à la réputation d'entreprise a été pour la première fois mentionnée par (Hall 1992). De plus selon une étude d'(Oxford Metrica 2011), la valeur de cet actif intangible que représente la réputation d'entreprise, est pour certaines compagnies plus importante que la valeur comptable de l'entreprise et la valeur de l'équité de sa marque combinées. Dans la figure 1.4, la valeur de l'équité de la marque est mesurée par Interbrand³, et le premium (égale à la capitalisation boursière moins la valeur comptable, moins la valeur de l'équité de la marque) est considéré comme étant la réputation d'entreprise. Le premium inclut des actifs tels que le leadership, l'innovation, la propriété intellectuelle, l'expertise de l'équipe de gestion, la portée mondiale de la compagnie, et les compétences et expérience des employés.

³ <http://www.interbrand.com/fr/Default.aspx>

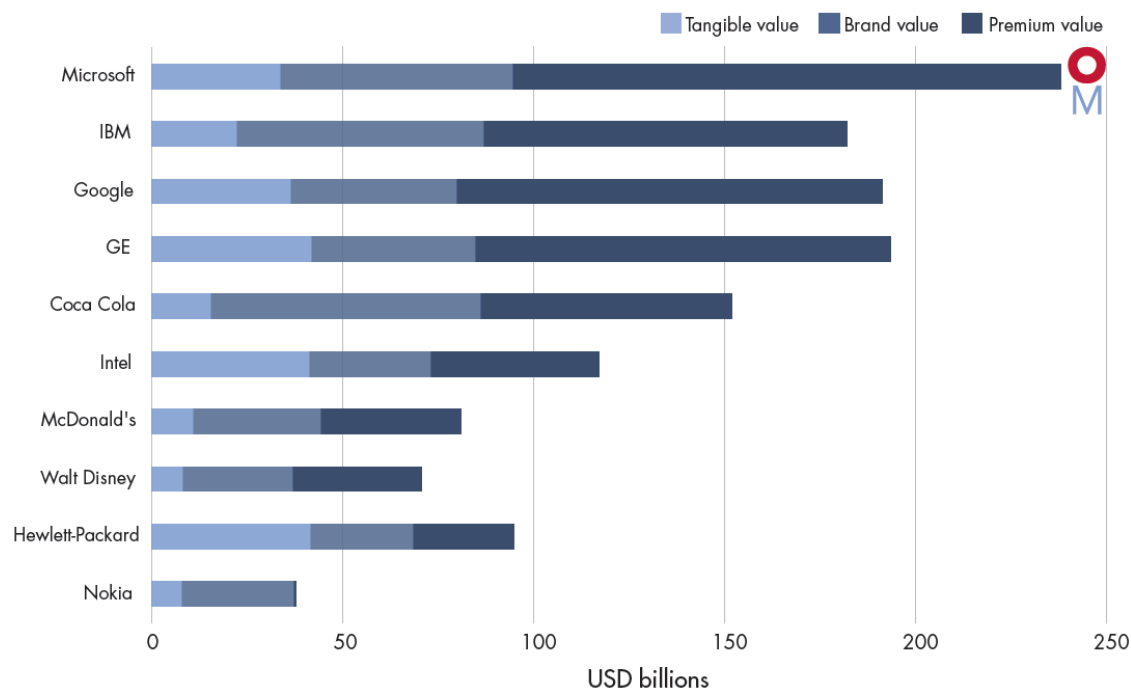


Figure 1.4. Le Top 10 des compagnies avec la plus importante équité de marque (Oxford Metrica 2011)

(Page and Fearn 2005) indiquent qu'une bonne réputation d'entreprise ne garantit pas le succès. De plus, les résultats de leur recherche les ont conduit à conclure que les préoccupations des consommateurs en matière de réputation d'entreprise sont portées sur la manière dont sont traités les consommateurs, le succès que connaît l'entreprise et son leadership dans son secteur. La responsabilité publique ne semble pas être une préoccupation majeure dans l'évaluation de la réputation d'entreprise par le public, par conséquent (Page and Fearn 2005) infèrent qu'insister beaucoup sur la responsabilité publique pour rehausser la réputation d'entreprise n'apportera pas un retour escompté significatif.

(Gabbioneta, Ravasi, and Mazzola 2007) ont conduit des travaux de recherche visant à déterminer les éléments clés de la réputation pour les analystes financiers. Les résultats de leur recherche indique les analystes financiers sont davantage préoccupés par la performance financière, le leadership, la structure de gouvernance, ainsi que la qualité des publications financières et des estimations futures des performances de l'entreprise.

1.1.2.2 Mesure de la réputation d'entreprise

Il existe aujourd'hui deux principales entités qui fournissent un proxy de la réputation d'entreprise : l'Institut de la Réputation (IR) et la liste du Fortune Magazine des entreprises les plus admirées. Les classements présentés par ces deux sources sont conduits annuellement.

L'IR a été fondé en 1997 par le Dr Charles Fombrun et le Dr Cees Van Riel. Le Dr. Charles Fombrun a été professeur au sein de Stern School of Business à New-York, et est également reconnu par ses pairs pour être un expert dans le champ d'étude de la réputation. Le Dr Van Riel est un professeur de l'Université Erasmus à Rotterdam (Pays-Bas). Aujourd'hui, l'IR couvre trente pays sur six continents. L'IR a développé une méthodologie/système nommé « RepTrack » qui mesure la réputation d'une firme au travers de sept dimensions : (1) produits et services, (2) innovation, (3) environnement de travail, (4) la responsabilité sociale, (5) la gouvernance, (6) le leadership, et (7) la performance. Les résultats sont présentés annuellement sous la nomination Forbes Published Global RepTrack 100 au sein du célèbre magazine américain Forbes.

La liste des compagnies les plus admirées offerte annuellement par le magazine Fortune, base ses classements au travers de neuf critères : (1) l'habileté à attirer et retenir du personnel employé talentueux, (2) la qualité de l'équipe de direction, (3) la responsabilité sociale envers les communautés locales ou l'environnement, (4) la capacité à innover, (5) la qualité des produits ou services offerts, (6) une utilisation sage des actifs d'entreprise, (7) la solidité financière de la firme, (8) la valeur des investissements à long-terme, (9) l'efficacité à faire des affaires mondialement.

1.1.2.3 Les impacts de la réputation d'entreprise sur les activités d'une entreprise

La réputation d'entreprise est connue pour avoir des impacts significatifs sur différents aspects des activités d'une firme : ses performances financières, sa compétitivité, la loyauté de sa clientèle et de ses employés, son habileté à survivre après une crise.

(Greyser 1999) a identifié plusieurs avantages d'une bonne réputation d'entreprise en commençant avec une meilleure préférence chez les consommateurs lorsque ceux-ci font face à des produits ou services identiques offerts par différentes compagnies. Deuxièmement, une bonne réputation offre également une sécurité lorsqu'une crise survient : une bonne réputation facilite la gestion de crise car le public maintiendra une certaine réserve dans son jugement envers la firme.

Troisièmement, (Greysier 1999) ajoute qu'une bonne réputation d'entreprise augmente la valeur de la compagnie sur les marchés financiers.

(Chun 2005) indique que la réputation d'entreprise est positivement corrélée avec la performance financière sur les marchés boursiers, la loyauté des consommateurs et de leur satisfaction (en fait la réputation est aussi une conséquence de la satisfaction des consommateurs), et avec la satisfaction et taux de rétention des employés.

(Bontis, Booker, and Serenko 2007) ont démontré que la réputation d'entreprise au sein de l'industrie bancaire joue le rôle de médiateur entre la satisfaction de la clientèle et la loyauté de la clientèle, et entre la satisfaction de la clientèle et la recommandation. Cela signifie que plus la réputation d'entreprise est forte, plus une compagnie est capable de transformer la satisfaction client en loyauté, et de transformer la satisfaction client en recommandations qui mèneront à l'acquisition de nouveaux clients. Et logiquement, l'acquisition de nouveaux clients mènent à une meilleure profitabilité.

1.1.2.4 Résumé du concept de réputation d'entreprise

Pour résumer, la réputation d'entreprise est intangible et ne peut être caractérisée unilatéralement. Elle regroupe les concepts d'image corporative, d'identité corporative et de d'identité désirée (Chun 2005). Elle s'observe sur le temps et non de manière ponctuelle (Balmer et al. 2002). On ne peut réellement offrir une évaluation de la réputation d'entreprise comme il est possible de le faire avec l'équité de marque, dont Interbrand, Brand-Finance et Oxford Metrica offrent une métrique avec chacun leur approche. Cependant, il est possible de fournir une mesure ou proxy de la réputation corporative comme le font le magazine Fortune et l'Institut de la Réputation.

1.2 Réputation d'entreprise et réseaux sociaux

1.2.1 Le "bouche-à-oreille"

Tel que discuté au travers des sections précédentes, la pérennité des marques corporatives dépend des opinions et évaluations des consommateurs. (E. Keller and Fay 2012) ont conduit un sondage auprès des consommateurs américains et ont estimé la distribution concernant les différents

moyens de communication employés par les consommateurs pour converser au sujet d'une marque. Selon leurs résultats, 75% du temps les consommateurs conversent à propos d'une marque en face à face, 15% du temps les consommateurs conversent au téléphone et les autres 10% du temps les conversations sur les marques se font sur le web.

L'impact du « bouche-à-oreille » (BAO) sur la décision d'achat des consommateurs a été grandement étudié dans la littérature. Le BAO est défini comme étant le processus de transmettre l'information de personne à personne (Jansen et al. 2009). (Bughin, Doogan, and Vetvik 2010) ont identifié trois formes de BAO : le BAO expérientiel, le BAO conséquent, et le BAO intentionnel. Le BAO expérientiel est la forme la plus commune de BAO et résulte d'une expérience client avec le produit ou service qui offre une performance au-delà des attentes exprimées par les clients. Le BAO expérientiel compte pour environ 50% à 80% de l'activité total en BAO. Le BAO conséquent est le résultat d'un consommateur qui a déjà été exposé à une campagne publicitaire d'un produit (ou service) et qui passe le message à propos du produit (ou service) publicisé. Le BAO intentionnel résulte lorsqu'une célébrité endosse un produit (ou service) et est utilisée afin de créer un buzz pour le lancement du produit (ou service) sur le marché. Un buzz est tout simplement un engouement général au sein du public.

(E. Keller and Fay 2012) affirment que le BAO entraîne les ventes à la hausse selon les résultats de leurs sondages indiquant que 58% des consommateurs trouvent fiable l'information obtenue au travers des conversations de BAO et 50% des consommateurs considèrent que les conversations de BAO constituent un facteur à forte influence dans leur décision d'acheter un produit (ou service) dans le cas d'un retour positif sur le produit (ou service). (Trusov, Bucklin, and Pauwels 2008) ont étudié les effets du marketing BAO sur l'acquisition de nouveaux membres au sein d'un réseau social sur le web. Leurs résultats indiquent que le BAO, comparé aux campagnes marketing traditionnelles, possède un plus fort et durable impact sur l'acquisition de nouveaux clients. En fait, leurs résultats suggèrent que l'élasticité à long-terme des nouvelles inscriptions pour le BAO, est 20 fois supérieure par rapport à celle liée au marketing traditionnel.

(East, Hammond, and Lomax 2008) mentionnent que le BAO est un processus efficace pour influencer la probabilité d'achat des consommateurs puisque les consommateurs bien informés peuvent fournir de l'information aux consommateurs moins bien informés. De plus, leurs

résultats indiquent que les consommateurs sont plus influencés par le BAO positif (BAOP) que par le BAO négatif (BAON).

(Anderson 1998) a montré que les consommateurs hautement non-satisfaits s'engagent davantage dans du BAO que les consommateurs hautement satisfaits. Cependant, il n'y a pas de différence significative dans le degré d'intensité du BAO entre les consommateurs satisfaits et les consommateurs non satisfaits : c'est –à-dire que qu'un consommateur satisfait exprimera sa satisfaction avec la même intensité qu'un consommateur non-satisfait exprimerait son insatisfaction. (Anderson 1998) mentionne également que le BAON possède un plus grand impact sur le sujet que le BAOP, contrairement à (East, Hammond, and Lomax 2008). (Williams, Buttle, and Biggemann 2012) ont conduit des recherches sur le lien existant entre le BAO généré par les consommateurs et la réputation d'entreprise. Ils ont conclut que le BAON a un impact plus important sur la réputation d'entreprise que le BAOP. Ces résultats vont dans le même sens que (Anderson 1998).

(Sundaram and Webster 1999) suggèrent que les impacts du BAO sur l'évaluation d'une marque dépendent de la familiarité de la marque avec les consommateurs. Le BAOP et le BAON ont un plus grand impact sur les marques non familières au public que les marques qui leur est familier; dans le sens où les marques non familières au public bénéficient davantage du BAOP que les marques familières, et que les marques non familières au public lorsqu'elles font face à du BAON sont sujet à une plus importante réduction de leur valeur que les marques familières au public.

1.2.2 Le bouche-à-oreille électronique

Avec l'introduction d'Internet, les consommateurs ont accès aux plateformes de partage d'opinions sur le web, où ils sont en mesure de partager publiquement leurs « opinions sur, et expériences avec, des biens et services avec une multitude d'autres consommateurs » (Hennig-Thurau et al. 2004). Ce processus est appelé le bouche-à-oreille électronique (eBAO).

(Cheung and Lee 2012) ont cherché à expliquer les raisons qui conduisent les consommateurs à propager l'eBAO au travers du web. Par ordre de significativité au sein du modèle qu'ils proposent les éléments clés sont: le sentiment d'appartenance à un groupe, l'engouement d'aider

les autres, et la réputation personnelle des consommateurs. Dans une moindre mesure, les autres éléments incluent : l'auto-efficacité (sentiment que possède un individu en sa capacité à accomplir un certain objectif) et l'obligation morale.

(Doh and Hwang 2009) ont conduit des travaux de recherche qui visaient à déterminer comment est-ce que les consommateurs évaluaient qu'un message eBAO puisse être considéré comme crédible. Leurs résultats de recherche indiquent que les consommateurs ne considèreraient pas un message d'eBAO provenant d'un site web comme une source fiable d'information si le ratio de messages positifs sur les messages négatifs était trop élevé (et vice versa). Le meilleur score de crédibilité concernant les messages eBAO a été obtenu pour un ratio de 8 :2, c'est à dire 8 messages positifs pour 2 messages négatifs.

En accord avec les travaux de (Williams, Buttle, and Biggemann 2012) concernant le BAO, (C. Park and Lee 2009) mentionnent que les messages négatifs du eBAO ont un impact supérieur aux messages positifs du eBAO sur le comportement des consommateurs vis-à-vis d'un sujet donné. De plus, ils indiquent que les biens d'expériences sont plus impactés par l'eBAO que les biens de recherche. Les biens d'expériences sont des biens dont les caractéristiques sont difficilement évaluables avant l'achat. À l'opposé les biens de recherche sont des biens dont les caractéristiques sont facilement évaluables avant l'achat.

(Lim, Chung, and Pedersen 2012) ont étudié l'impact de l'eBAO, et ont conclut que les consommateurs seraient plus influencés à acheter un produit en se référant à des messages d'avis de faible qualité, composés d'opinions subjectives et d'émotions, plutôt qu'en se référant à des messages d'avis d'excellente qualité composé d'opinions objectives. De plus, la qualité des messages provenant des agents de commercialisation n'affecte pas la décision d'achat des consommateurs. Ces découvertes veulent donc dire qu'un consommateur moyen construit sa décision objective sur un ensemble d'opinions subjectives provenant d'autres consommateurs, plutôt que sur des opinions objectives provenant d'experts.

La revue de littérature sur l'eBAO et les micro-blogs tel que Twitter, est encore limitée. Après une analyse sur plus de 150,000 messages sur des micro-blogs, (Jansen et al. 2009) ont trouvé qu'approximativement 19% des micro-blogs mentionnent une marque d'entreprise, et parmi ces derniers, 20% contiennent des expressions de sentiments à propos d'une marque. La moitié de ces expressions de sentiments était positive, et un tiers de ces expressions de sentiments fut

considéré comme critique pour une compagnie ou son produit. Un travail antérieur sur les micro-blogs et l'eBAO a été conduit par (Esch et al. 2006). Ils ont présenté un modèle (Esch et al. 2006 annoté par Jansen et al. 2009) pour illustrer comment la connaissance d'une marque et les relations d'une marque affectent les achats actuels et futures des consommateurs (voir figure 1.5). D'un côté, les achats actuels sont affectés directement par l'image de marque et indirectement par la notoriété de la marque. Et d'un autre côté, les achats futurs sont affectés par les relations d'une marque qui sont composées de la confiance dans la marque, de la satisfaction envers la marque, et de la loyauté envers la marque.

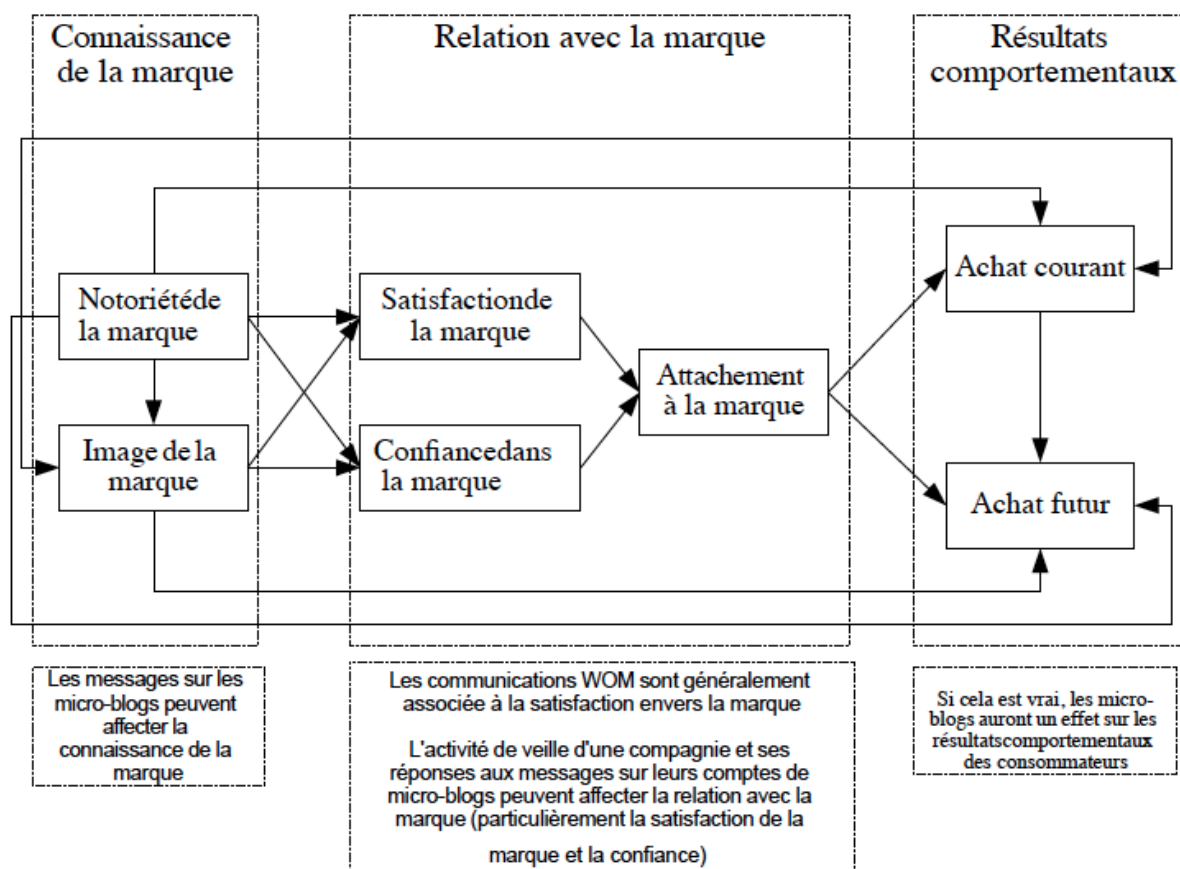


Figure 1.5 Modèle illustrant la relation entre l'image de marque et les micro-blogs adapté de (Esch et al. 2006, annoté par Jansen et al. 2009)

1.2.3 Observer la réputation d'entreprise au travers des réseaux sociaux

(Mangold and Faulds 2009) définissent les médias sociaux comme étant un regroupement de différentes formes de forums de BAO incluant en outre les « blogs », site web d'évaluation de produit ou de service, les espaces de « chats » au sein des compagnies, les « moblogs », les sites de réseaux sociaux, les sites de réseautage professionnel. Un réseau social est défini comme étant « une plateforme internet où un utilisateur peut créer un profil et construire un réseau personnel le connectant à d'autres utilisateurs » (Madden 2007).

(Warin et al. 2013) proposent un modèle économique évaluant l'impact d'une crise au sein des réseaux sociaux sur la réputation d'entreprise. Ce modèle est basé sur une approche de théorie des jeux dans un contexte d'un événement contraignant affectant la réputation d'une entreprise. Trois principaux paramètres sont identifiés pour la gestion de la réputation d'entreprise après une crise : (1) le temps de réaction après la crise, (2) le délai de recouvrement vers la réputation initiale (avant crise) et finalement (3) le niveau de réputation d'une firme. Ce modèle s'illustre selon deux axes : en abscisse le temps de recouvrement de la réputation, et en ordonnée le niveau de réputation de la firme. La réputation d'entreprise est définie comme une fonction linéaire à pente positive, où le degré de la pente dépend de la réputation de la firme : plus la pente est élevée plus le délai de recouvrement de la réputation suite à crise est faible et donc plus le niveau de réputation est important. À l'inverse, plus la pente est faible, plus le délai de recouvrement de la réputation suite à une crise est important et donc plus la réputation de l'entreprise peut être considérée comme faible. La ligne horizontale représente l'impact de l'évènement indésirable sur la réputation de la firme. Elle croise la droite de réputation en un point qui indique le temps de recouvrement nécessaire pour recouvrer la perte de réputation suite à la crise.

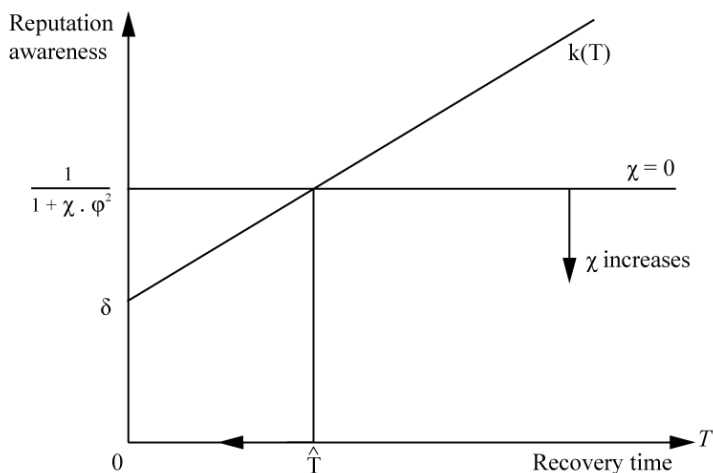


Figure 1.6 Modèle de la réputation d'entreprise en temps de crise (Warin et al. 2013)

(Jones, Temperley, and Lima 2009) ont analysé comment la réputation d'entreprise est gérée au sein des plateformes de réseaux sociaux, qui sont considérés comme des « outils donnant davantage de pouvoir aux consommateurs et citoyens ». Ils définissent la gestion de la réputation en ligne sous forme d'un processus en trois étapes : les firmes surveillent l'activité sur les réseaux sociaux, elles participent aux échanges sur les réseaux sociaux en interagissant avec les consommateurs, et elles mesurent les impacts de leur présence au sein des réseaux sociaux sur leur réputation et leurs marques. (Jones, Temperley, and Lima 2009) proposent également une matrice de contrôle de la réputation d'entreprise en ligne (voir figure 1.7).

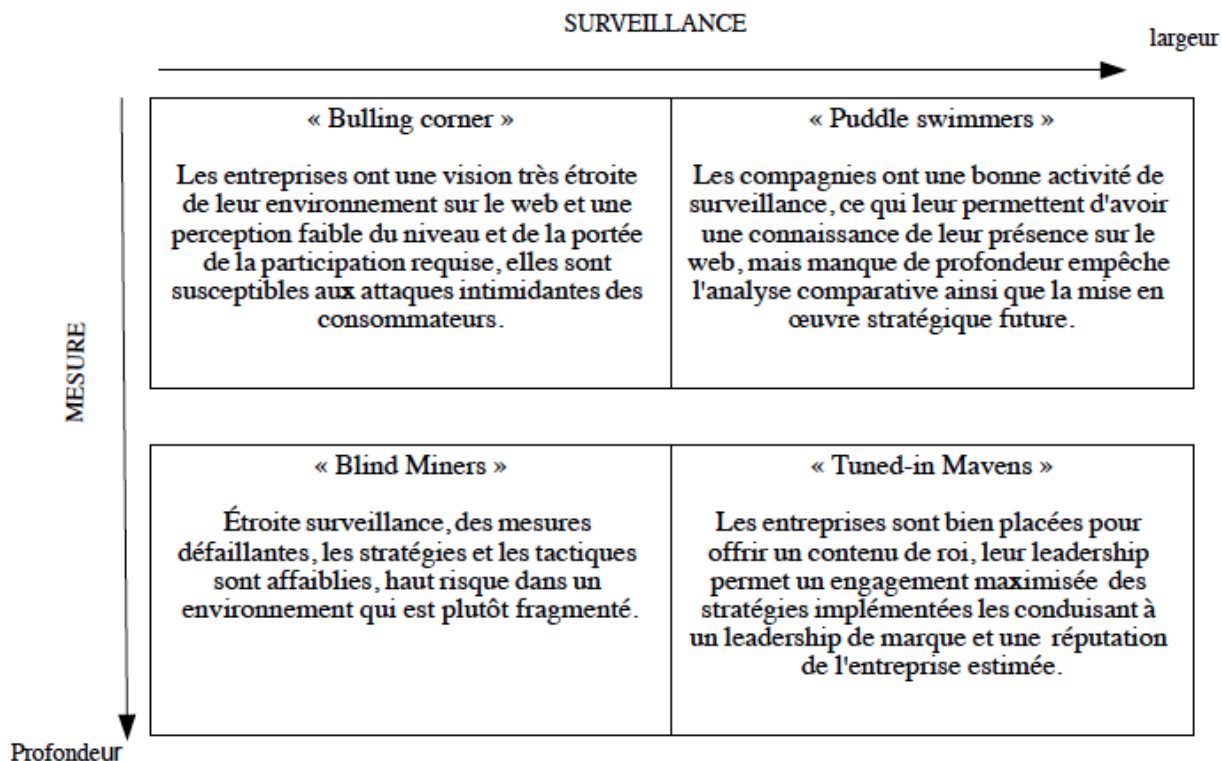


Figure 1.7 La matrice de mesure et gestion de la réputation sur le web (adapté de Jones, Temperley, and Lima 2009)

Les firmes de la catégorie “Tuned-in Mavens” sont celles qui ont un bon contrôle de leur présence sur les réseaux sociaux, c’est-à-dire qui ont une bonne connaissance de ce qui est dit en ligne à leur sujet. Elles peuvent mesurer de manière fiable les impacts de leurs activités au sein des réseaux sociaux sur leur réputation.

À l’opposé, les firmes de la catégorie “Bulling Corner” sont celles qui ont un contrôle limité sur leur présence au sein des réseaux sociaux. Elles ne conduisent aucune activité de veille sur les réseaux sociaux (ou conduisent de légères activités de veille), ou ne sont pas capables de mesurer de manière fiable les impacts de leur activités au sein des réseaux sociaux. Par conséquent, elles sont plus exposées aux attaques négatives des consommateurs sur le web qui peuvent affecter négativement leur réputation.

Les firmes appartenant à la catégorie “Blind Miners” sont des firmes qui conduisent des activités de veille restreintes sur les réseaux sociaux, mais qui développent des outils pour évaluer les impacts de leur présence au sein des réseaux sociaux sur leur réputation. Cependant leurs

activités de veille restreintes les conduisent inévitablement à des biais dans les mesures prises avec leurs outils.

Finalement, les firmes appartenant à la catégorie “Puddle Swimmers” sont les firmes qui entretiennent des activités de veille importantes mais qui ont des systèmes peu avancés pour mesurer leur présence sur les réseaux sociaux. Ainsi, elles possèdent une bonne idée de ce qui est dit en ligne à leur sujet mais ne peuvent pas vraiment capitaliser sur cette information de manière efficace. (Jones, Temperley, and Lima 2009) apportent également une liste de logiciels permettant d’avoir une mesure de la réputation en ligne et le buzz : TheBuzzMonitor⁴ (lancé par la Banque Mondiale en 2007, mais le projet semble avoir été abandonné au moment où ce document est écrit), Semonics⁵, Radian6⁶ (acquis par Salesforce.com), BuzzGain⁷ (acquis en 2010 par Meltwater).

(Grützmaier 2011) a examiné si les activités sur les réseaux sociaux affectaient la réputation d’entreprise. La recherche a été conduite au travers d’une étude de cas de la firme Nokia. Les découvertes de cette recherche indiquent que les consommateurs utilisant les réseaux sociaux forment un groupe homogène avec une faible différence entre eux dans leur utilisation. En effet, peu importe l’âge, la nationalité, le sexe, le fait d’être étudiant ou un professionnel, les consommateurs semblent tous utiliser les réseaux sociaux dans la même proportion. Deuxièmement, les utilisateurs des réseaux sociaux sont généralement au courant des activités sur les réseaux sociaux des produits ou services répondant à leurs besoins. Troisièmement, les réseaux sociaux ne semblent pas être déterminant pour la réputation de la firme Nokia : la plupart des sujets à l’étude étaient européens, ce qui suggèrent que les réseaux sociaux ont un rôle mineur

⁴ <http://buzzm.worldbank.org/>

<http://www.branded3.com/blogs/world-bank-buzzmonitor/>

⁵ <http://semonics.com/>

⁶ http://www.radian6.com/cms/about_us

<http://www.salesforcemarketingcloud.com/>

⁷ <http://www.buzzgrain.com/index.php?action=login>

<http://www.meltwater.com/products/meltwater-buzz-social-media-marketing-software/>

dans le rehaussement de la réputation d'entreprise du point des consommateurs européens en ce qui concerne la firme Nokia.

(Bughin, Chui, and Miller 2009) ont conduit un sondage visant à déterminer les bénéfices de l'utilisation des réseaux sociaux pour les entreprises dans leurs activités quotidiennes. Les résultats de leur recherche indiquent que les plus importants bénéfices de l'utilisation des réseaux sociaux liés à ce qui touche les consommateurs sont au niveau de l'augmentation de l'efficacité des activités de commercialisation. L'efficacité des activités de commercialisation est segmentée en quatre aspects : la notoriété, la considération, la conversion, et la loyauté (tous ces aspects sont impactés par la réputation d'entreprise tel que discuté dans les sections précédentes). Les chercheurs ont observé une amélioration entre 17 et 25% pour les quatre aspects mentionnés précédemment. Par conséquent, il est possible de conclure que l'utilisation des réseaux sociaux est actuellement une pratique importante pour les firmes dans la gestion efficace de leur réputation. Pourtant tel que vu dans la section précédente sur la réputation, le magazine Fortune et l'Institut de la Réputation ne considèrent pas actuellement l'activité sur les réseaux sociaux dans l'élaboration de leur mesure de la réputation.

1.2.4 Résumé du lien existant entre réseaux sociaux et réputation d'entreprise

En résumé les réseaux sociaux sont des plateformes de partage où l'information entre différents utilisateurs se propage plus rapidement que les moyens traditionnels de communications. Ce processus de bouche-à-oreille électronique (eBAO) joue un rôle clé dans le comportement des consommateurs vis-à-vis d'un produit ou d'un service. Avec l'avènement et la propagation d'internet, les consommateurs deviennent davantage sensibles aux opinions subjectives des autres consommateurs transigées sur les réseaux sociaux, par rapport aux opinions objectives transmises par des spécialistes ou agents commerciaux. Ainsi les réseaux sociaux peuvent jouer un rôle important dans l'amélioration ou la dégradation de la réputation d'entreprise.

1.3 Analyse de sentiments

1.3.1 Définition du concept d'analyse de sentiments

La première mention de l'expression « analyse de sentiment » dans la revue de littérature est attribué à (Nasukawa and Yi 2003). (Pang and Lee 2008) ont apporté une large définition de l'analyse de sentiment définie comme étant le « traitement informatique d'opinion, de sentiment et de subjectivité dans un texte ». Quelques années plus tard, (Liu 2012) a complété la définition du problème d'analyse de sentiment en définissant ce qu'était une opinion de manière plus précise. Une opinion est un quintuple composé de : (1) le nom de l'entité visé par l'opinion, (2) un aspect de l'entité qui est critiqué, (3) le sentiment exprimé sur l'aspect de l'entité, (4) un porteur d'opinion, et (5) le temps auquel l'opinion a été exprimée. Un sentiment peut avoir trois polarités distinctes : il peut être positif, négatif ou neutre. Il est important de mentionner que l'analyse de sentiment possède plusieurs terminologies dans la revue de littérature, avec le terme « extraction d'opinion » étant le synonyme le plus employé.

L'analyse de sentiment peut être réalisée à différents niveaux de granularité d'un texte : elle peut être conduite au niveau du document entier ou au niveau de la phrase. Au niveau du document (exemple un article de journal), l'objectif est d'évaluer le sentiment général du document. Parmi les premiers chercheurs sur le sujet, on compte (Turney 2002; Nasukawa and Yi 2003; Yu and Hatzivassiloglou 2003). (Turney 2002) a classifié les avis comme « recommandés » (thumbs up) ou non recommandés (thumbs down) au travers d'un processus en trois étapes. La première étape consiste à extraire les phrases de l'article individuellement et identifier les identifiants des parties du discours (« part-of-speech » ou POS tags) pour chacun des mots dans les phrases extraites. Un POS tags peut être un adverbe, un adjectif, un nom etc. La deuxième étape consiste à estimer l'orientation sémantique des phrases extraites en utilisant la formule suivante PMI-IR :

$$PMI(mot_1, mot_2) = \log_2 \left[\frac{p(mot_1 \& mot_2)}{p(mot_1) p(mot_2)} \right]$$

Figure 1.8 La formule PMI-IR pour déterminer l'orientation sémantique des phrases (Turney 2002)

Ici $p(\text{mot1} \ \& \ \text{mot2})$ est la probabilité qu'un mot1 et un mot2 apparaissent en même temps. Si les mots sont statistiquement indépendants, alors la probabilité qu'ils apparaissent est donnée par le produit de $p(\text{mot1})p(\text{mot2})$. Le ratio entre $p(\text{mot1} \ \& \ \text{mot2})$ et $p(\text{mot1}) p(\text{mot2})$ est donc une mesure du degré de dépendance statistique entre les mots. Le log de ce ratio est la quantité d'information que nous acquérons à propos de la présence d'un des mots lorsque nous observons l'autre mot.

Finalement, la troisième étape consiste à calculer l'orientation sémantique moyenne de toutes les phrases extraites et de classer de manière appropriée l'article comme recommandé (si la moyenne de l'orientation sémantique des phrases extraites est positive) ou non recommandé (si la moyenne de l'orientation sémantique des phrases extraites est négative).

(Yu and Hatzivassiloglou, 2003) ont présenté un modèle pour détecter les opinions des faits au niveau du document (également au niveau de la phrase dans la seconde partie de leur article). Ils ont utilisé trois méthodes pour accomplir cette tâche. La première méthode a été d'utiliser SIMFINDER, qui est un système pour mesurer la similarité des phrases, basé sur des mots, phrases de références et sur WordNet. WordNet est un dictionnaire de mots avec leur classification de polarité respective (exemple : « heureux, positif » ou « horreur, négatif »). La seconde méthode a été développée en utilisant un classificateur naïf bayésien. Le classificateur inclut des options telles que la présence de mots sémantiquement orientés, les tags POS, le nombre de mots positifs ou négatifs dans une phrase, que l'on retrouve dans la discipline du traitement de langage naturel (NLP pour « natural language processing »). La troisième méthode utilise plusieurs classificateurs de types Naïves Bayes reposant sur différents paquets d'options tel qu'énoncé précédemment (POS tags, présence de mots sémantiquement orientés etc.). Les résultats de leur recherche indiquent qu'au niveau du document, la seconde méthode (classificateur Naïve Bayes) obtient les meilleurs résultats avec une précision de 97%.

(Pang and Lee 2004) ont intégré une détection de la polarité du sentiment au niveau de la phrase dans un modèle de classification au niveau du document pour classer des textes. Leur modèle enlève les phrases objectives du texte en utilisant un classificateur commun conduisant la tâche d'analyse de sentiment au niveau de la phrase. La figure suivante représente les différentes étapes de leur méthode.

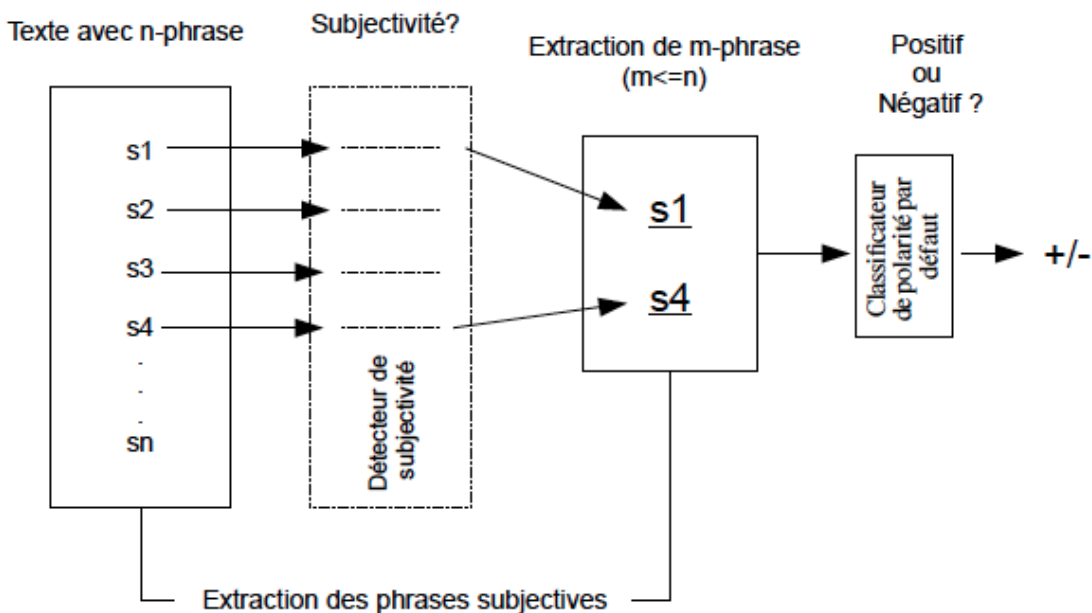


Figure 1.9 Extraction de la subjectivité dans la classification de document réalisée par (Pang and Lee 2004)

Au niveau de la phrase, l'objectif est d'évaluer le sentiment général de la phrase. (Kim 2004) soulève quelques difficultés avec la classification de l'analyse de sentiments au niveau de la phrase : par exemple, certains mots peuvent avoir une connotation positive et une connotation négative dépendamment du contexte, ou encore un porteur d'opinion peut exprimer deux opinions dans la même phrase. La figure suivante (figure 1.10) présente la structure architecturale de la méthode utilisée :

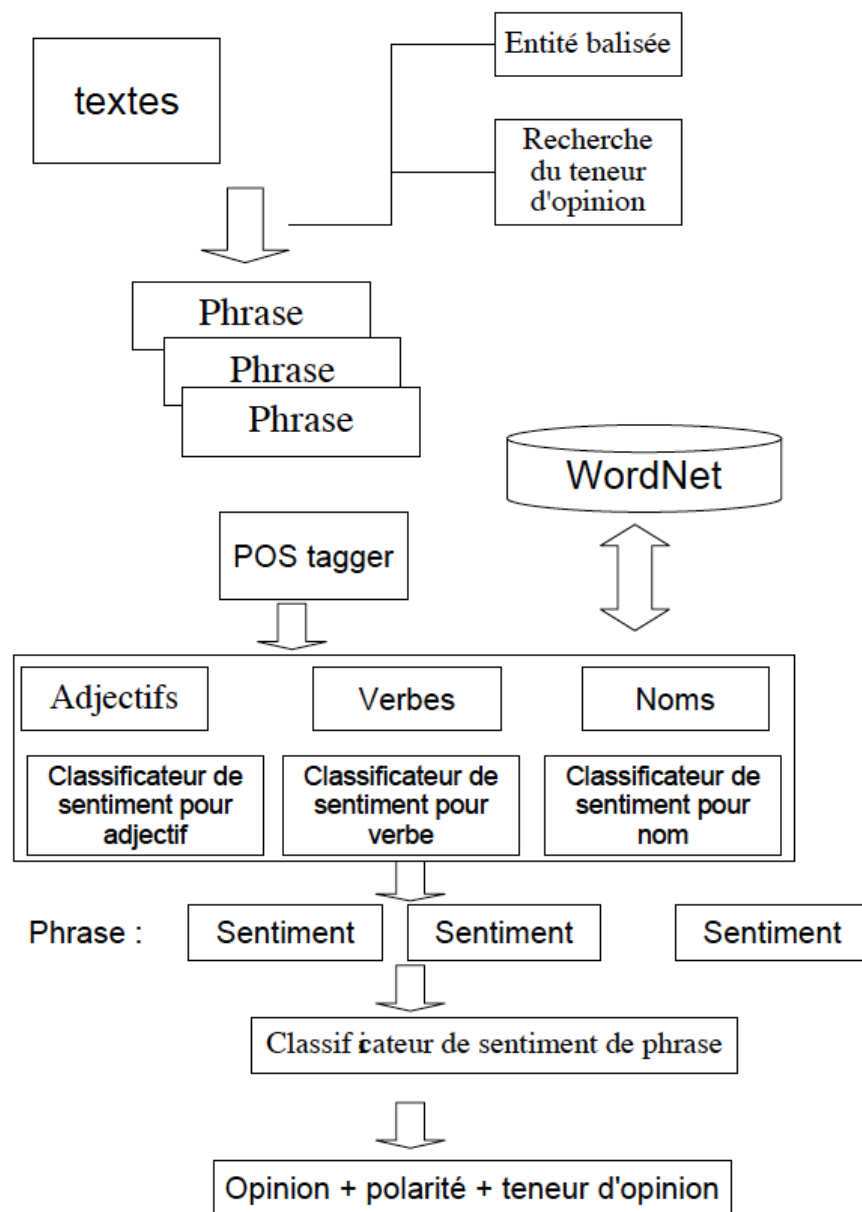


Figure 1.10 Architecture du système développé par Kim and Hovy 2004, pour la classification de la polarité au niveau de la phrase

(Hu and Liu 2004) ont développé une méthode pour analyser les avis des consommateurs à propos d'un produit. La méthode est basée sur un processus en trois étapes : la première étape identifie les caractéristiques d'un produit qui sont commentées; la seconde étape identifie l'opinion des phrases pour chaque avis recensé et détermine la polarité de ces avis (positif, négatif ou neutre); la troisième étape résume les résultats obtenus. Par exemple, considérons des

avis sur un vélo d'une firme X. La méthode va premièrement identifier les caractéristiques du vélo commentées (ex : sa vitesse, son guidon, sa scelle, son cadre, sa stabilité, son esthétisme etc.). Deuxièmement, les avis sur chaque caractéristique sont analysés et la polarité des avis est déterminée (positif, négatif ou neutre). Enfin la troisième étape consiste à présenter les résultats obtenus.

L'architecture de leur méthode est présentée sur la figure suivante (figure 1.11) :

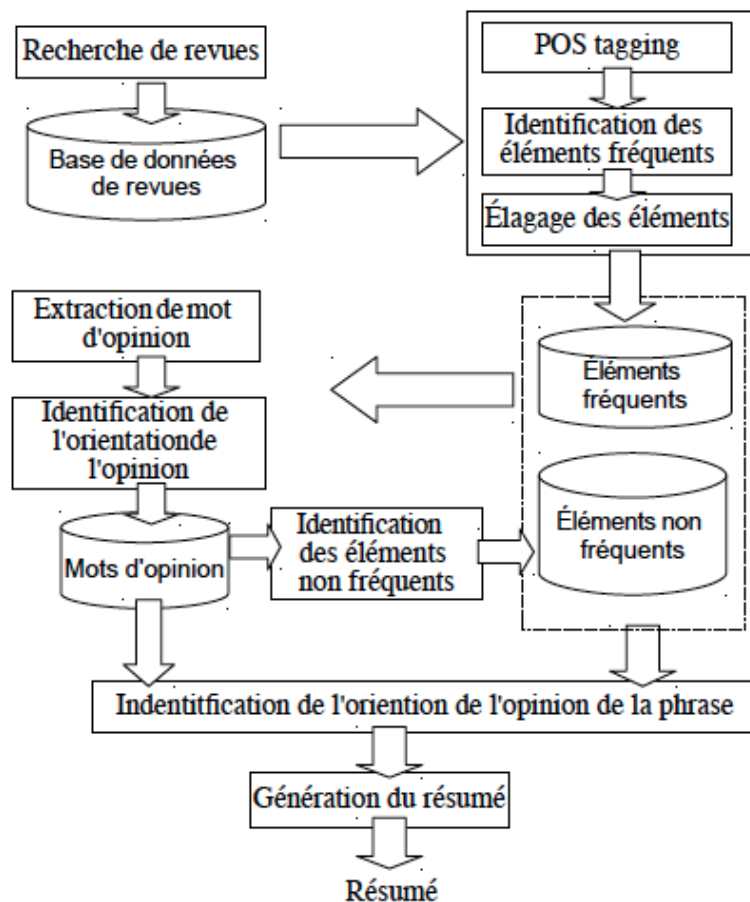


Figure 1.11 Architecture du système développé par Hiu et Liu dans l'analyse des avis déposés par des consommateurs sur des produits (adapté de Hiu et Liu 2004).

(Wilson, Wiebe, and Hoffmann 2005) ont présenté une nouvelle approche pour distinguer automatiquement la polarité préalable et la polarité contextuelle des mots. La polarité préalable d'un mot fait référence à la polarité d'un mot en dehors de tout contexte. La polarité contextuelle fait référence à la polarité d'un mot dans un contexte particulier. L'approche des chercheurs a

démontré une précision de 75,9% pour la classification entre les mots à polarité neutre et les mots à polarité non-neutre d'une part, et une précision de 65,7% pour la classification catégorisant les mots à polarité positive et négative.

(Agarwal, Biadsky, and Mckeown 2009) ont développé un classificateur avec WEKA dans le but de conduire une analyse de sentiment. Ils ont utilisé trois types d'outils pour cette tâche: (1) la polarité préalable d'un mot basé sur deux dictionnaires anglais indiquant la polarité des mots (le premier se nomme DAL et le second WordNet), (2) des outils lexicaux tels que les tags POS et les n-grams (qui sont des regroupement de « n » nombre de mots), et (3) une combinaison des scores de polarité provenant du dictionnaire DAL et d'autres outils syntaxiques.

Les deux niveaux de granularité de texte (document ou phrase) requièrent des approches plutôt différentes. Dans le contexte de recherche de cette étude, le niveau de granularité choisi sera le niveau de granularité de la phrase qui est le plus approprié pour les messages de Twitter (tweets) puisqu'ils contiennent au maximum 140 caractères.

1.3.2 L'analyse de sentiment appliquée sur les données issues de Twitter

(Go, Huang, and Bhayani 2009) furent les premiers (selon leur connaissance) à travailler sur l'analyse de sentiments sur les messages de Twitter. Ils ont comparé différents classificateurs (algorithmes qui performant une classification de sentiment sur un texte ou phrase donnée), basés sur différentes approches bien connue dans la revue de littérature de l'analyse sémantique telles que la méthode naïve bayésienne, l'entropie maximale (MaxEnt), le « Support Vector Machine » (SVM), et l'approche commune de base qui attribue une classification de polarité selon le nombre de mots contenant une certaine polarité (ex : si une phrase contient plus de mots positifs que de mots négatifs alors la phrase est positive). Leurs résultats indiquent que l'algorithme SVM performe mieux que les autres méthodes. À la suite de leur travaux, ils ont conçu une application nommée Sentiment 140 qui est considérée être la meilleure application publiquement disponible pour l'analyse de sentiment sur Twitter selon (Barbosa and Feng 2010).

(Bifet and Frank 2010) ont poursuivi le travail de (Go, Huang, and Bhayani 2009) en proposant une nouvelle mesure pour la tâche d'analyse de sentiment: la statistique Kappa. (Bifet and Frank 2010) ont obtenu de meilleurs résultats que leurs prédécesseurs en utilisant un modèle du gradient

stochastique (ou « *stochastic gradient descent* » en anglais). Ils ont construit leur algorithme en utilisant deux logiciels libre : WEKA développé par l'université de Waikato en Australie (voir section 2.3.4) et Massive Online Analysis (voir section 2.3.4) développé par (Bifet et al. 2010). Les deux logiciels sont basés avec le langage de programmation JAVA.

(Pak and Paroubek 2010) ont construit un classificateur basé sur un algorithme de type Naïves Bayes qui utilisent des options tels que les n-grams, les POS tags. Les Cependant, dans une autre étude (Efthymios Kouloumpis 2011) mentionnent les tags POS ne semblaient pas très utiles dans leur approche d'analyse de sentiment. L'objectif de recherche de (Pak and Paroubek 2010) était d'évaluer l'utilité d'options linguistiques tels que les tags POS, les hashtags ou les émoticônes dans la tâche de classification de la polarité des tweets. Ils indiquent que les émoticônes représentent une excellente option linguistique dans la classification de la polarité des tweets.

(Agarwal et al. 2011) ont introduit une nouveau modèle d'analyse de sentiment basé sur un « arbre kernel », qui semble avoir de meilleurs résultats que les modèles de base introduits par les précédents chercheurs tels que (Go, Huang, and Bhayani 2009; Pak and Paroubek 2010). Un exemple du modèle d' « arbre kernel » est représenté à la figure ci-après (voir figure 1.12).

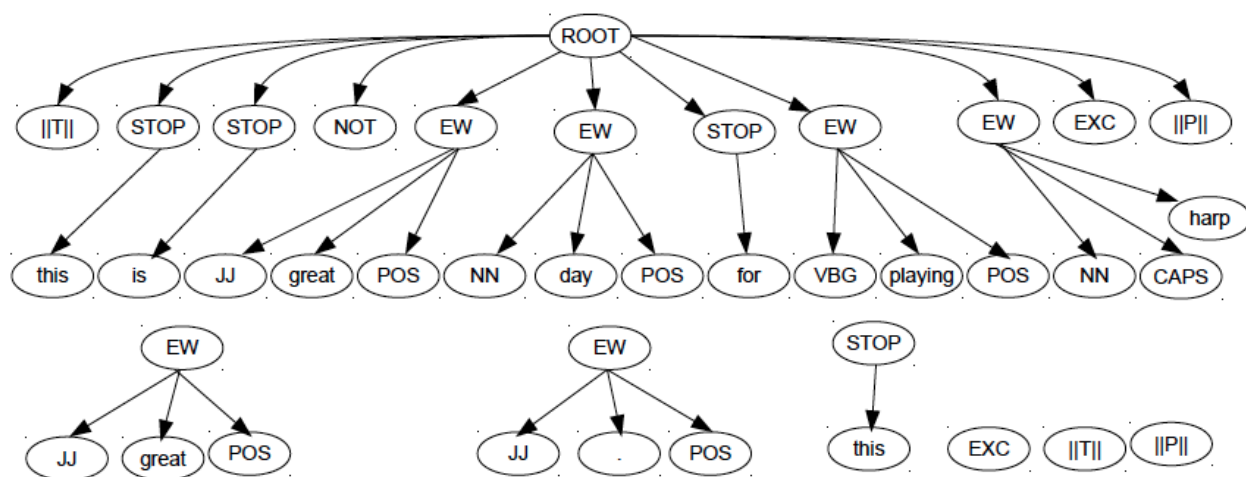


Figure 1.12 Le modèle graphique Tree Kernel appliqué au tweet: “@Fernando this isn't a great day for playing the HARP! :)”

(Agarwal et al. 2011) ont utilisé les outils suivants pour conduire leur analyse de sentiment:

\mathbb{N}	Polar	POS	# of (+/-) POS (JJ, RB, VB, NN)	f_1
		Other	# of negation words, positive words, negative words	f_2
			# of extremely-pos., extremely-neg., positive, negative emoticons	f_3
			# of (+/-) hashtags, capitalized words, exclamation words	f_4
	Non-Polar	POS	# of JJ, RB, VB, NN	f_5
		Other	# of slangs, latin alphabets, dictionary words, words	f_6
			# of hashtags, URLs, targets, newlines	f_7
\mathbb{R}	Polar	POS	For POS JJ, RB, VB, NN, \sum prior pol. scores of words of that POS	f_8
		Other	\sum prior polarity scores of all words	f_9
	Non-Polar	Other	percentage of capitalized text	f_{10}
\mathbb{B}	Non-Polar	Other	exclamation, capitalized text	f_{11}

Table 4: \mathbb{N} refers to set of features whose value is a positive integer. They are primarily count features; for example, count of number of positive adverbs, negative verbs etc. \mathbb{R} refers to features whose value is a real number; for example, sum of the prior polarity scores of words with part-of-speech of adjective/adverb/verb/noun, and sum of prior polarity scores of all words. \mathbb{B} refers to the set of features that have a boolean value; for example, presence of exclamation marks, presence of capitalized text.

Figure 1.13 Types d'options utilisées par (Agarwal et al. 2011) pour l'analyse de sentiment

(Wang et al. 2011) ont été les premiers à explorer une nouvelle approche dans l'analyse de sentiment des tweets en utilisant les hashtags des tweets. Leur modèle analyse les hashtags au travers d'une approche reposant sur la construction d'un graphe. L'idée est d'identifier la corrélation ou les liens entre les hashtags dans le but de déterminer la polarité d'un tweet auquel les hashtags sont associés. Ainsi, si un hashtag est généralement lié avec d'autres hashtag à polarité positive, alors le sentiment général du tweet contenant le hashtag sera considéré comme positif. Leur modèle ne performe pas mieux que le modèle de (Barbosa and Feng 2010) basé un algorithme utilisant la méthode SVM. Cependant, le modèle de (Wang et al. 2011) offre une bonne visualisation des liens qui existent entre les hashtags. Un exemple du modèle de graphe d'hashtags est présenté ci-après :

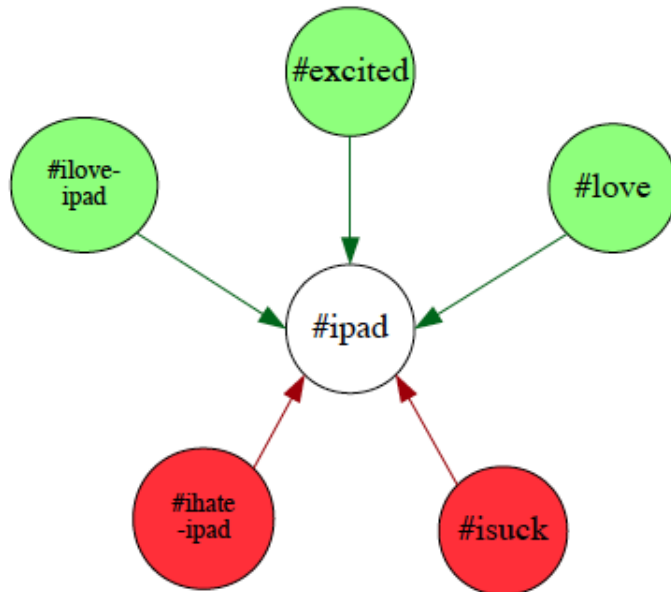


Figure 1.14 Le modèle graphique d’Hashtag représentant les liens avec le hashtag #ipad (adapté de Wang et al. 2011)

(Barbosa and Feng 2010) ont proposé un classificateur qui résulte de la combinaison de trois différentes applications conduisant des analyses de sentiment : Twendz, Tweetfeel, TwitterSentiment. La qualité individuelle de chacune de ces trois applications a été jugée raisonnable. De plus, le degré d’accord, entre les résultats suite à une analyse de sentiment sur un échantillon de tweets commun, s’est révélé être modéré. Ce fait indique que chaque application possède un biais qui lui est propre dans la détection de la polarité. Par conséquent, combiner ces trois sources améliore la performance de l’analyse de sentiment, comparativement à utiliser individuellement chaque application. Le classificateur construit par (Barbosa and Feng 2010), a été réalisé à l’aide de WEKA, un logiciel libre d’exploration de donnée (voir section 2.3.4).

1.3.3 Des illustrations concrètes de l’applications de l’analyse de sentiment avec Twitter

La recherche sur l’analyse de sentiment au niveau de granularité de la phrase appliquée aux messages de Twitter s’étend sur différents champs d’application tels que la politique, la finance de marché, ou l’industrie du cinéma,

(Tumasjan et al. 2010) ont conduit des travaux de recherche dans le contexte des élections nationales de 2009 en Allemagne. Le premier objectif de la recherche était d'évaluer si Twitter était utilisé seulement comme un moyen pour une personne d'exprimer son opinion politique ou si Twitter était aussi un moyen d'entretenir une conversation politique. Le second objectif de leur recherche était de déterminer si Twitter avait un certain pouvoir prédictif d'élections électorales en utilisant le volume des tweets reliés aux six partis en compétition durant ces élections nationales. Les résultats de leur recherche indiquent que Twitter est utilisé dans 30% des cas pour avoir une conversation politique et dans 70% des cas pour exprimer une opinion sur un parti. Pour la tâche d'analyse de sentiment, ils ont utilisé un logiciel nommé LIWC. Après avoir déterminé pour chaque parti politique le nombre de tweets positifs et le nombre de tweets négatifs publiés sur Twitter, ils ont compilé les scores en classifiant les partis selon le nombre de tweets publiés et le ratio de tweets positifs sur tweets négatifs. Leur résultats indiquent que le volume de tweets permet de prédire les résultats des élections avec une erreur absolue moyenne (MAE) de 1.65%, erreur similaire à celles des autres sources de prédictions électorales basées sur des sondages.

(Bermingham and Smeaton 2011) ont conduit une étude toujours en rapport avec des élections politiques dans le contexte des élections nationales de 2011 en Irlande. L'objectif de recherche était d'évaluer le pouvoir prédictif de Twitter dans la détermination des résultats suite aux élections. Les chercheurs ont collecté les tweets mentionnant les différents partis politiques représentés, dix-sept jours avant la tenue du vote électoral. Tous les tweets ont été analysés comme étant soit positifs, négatifs, neutres, mixtes, ou non-caractérisables, insignifiants ou encore ambigus. Tous les tweets caractérisés comme non-caractérisables, insignifiants ou ambigus ont été supprimés de l'échantillon de tweets. L'analyse de sentiment a été conduite en utilisant un classificateur Adaboost MNB. Selon les résultats de l'analyse de sentiment effectuée, ils ont calculé pour chaque partie la part du volume de tweets, la part du volume de tweets positifs, et la part du volume de tweets négatifs. Les pouvoirs prédictifs de ces variables ont été testés par le moyen d'un modèle de régression qui a obtenu un R^2 égal à 0,95! Cependant, les résultats de la prédiction ont une MAE égale à 10%, qui est beaucoup plus grande que les autres sources de prédictions électorales basées sur des sondages présentant une MAE inférieure à 3%. Les chercheurs ont alors conclu que le volume de tweets n'est pas suffisant pour prédire avec une

précision acceptable des élections politiques et que la méthode d'analyse de sentiment doit être également revue.

(O'Connor et al. 2010) indiquent que Twitter possède un potentiel de prédiction d'élections politique aux Etats-Unis. Leurs travaux ont été conduits après les élections présidentielles américaines opposant le démocrate et actuel président Barack Obama et le républicain John McCain. L'analyse de sentiment a été réalisée à l'aide d'OpinionFinder, une liste de mots avec leur polarité (négative ou positive), similaire à WordNet ou DAL mentionné précédemment. Ainsi chaque mot d'un tweet est comparé aux mots contenus dans OpinionFinder et caractérisé comme positif ou négatif selon la classification d'OpinionFinder. La polarité des tweets a été évaluée en comptant le nombre de mots positifs et négatifs présent dans un tweet. Si le nombre de mots positifs est supérieur au nombre de mots négatifs alors le tweet est positif. Par la suite, un score de sentiment a été mesuré chaque jour pour chacun des candidats en calculant le ratio de tweets positifs sur les tweets négatifs. Les résultats de l'analyse indiquent les données de twitter sont corrélés à ceux des sondages ($R^2=73.1\%$) et de plus que Twitter est un indicateur précurseur comparé aux sondages : la détection du sentiment de la population est détectable avant les sondages.

(Gayo-Avello, Metaxas, and Mustafaraj 2011) s'opposent aux affirmations mentionnant un pouvoir prédictif de Twitter en contexte électoral. Leur recherche a été menée durant les élections du Congrès américain de 2010. L'objectif de la recherche était d'évaluer le pouvoir prédictif de Twitter au travers de cette élection. Les chercheurs ont répété la même méthodologie que (Tumasjan et al. 2010; O'Connor et al. 2010) en utilisant le volume des tweets et l'analyse de sentiment. Cependant les résultats obtenus indiquaient que la méthode décrite par les précédents chercheurs n'était pas appropriée pour prédire les résultats d'une élection politique. Ainsi même si dans certains contextes la méthode a pu s'avérer être acceptable, puisqu'elle n'est pas en tout répétable, les auteurs infèrent que les résultats obtenus par les autres chercheurs ne résultent que du hasard.

Dans le domaine de la finance, l'analyse de sentiment a été utilisée pour développer des stratégies d'investissements sur les marchés boursiers. (Sprenger and Welppe 2010) ont analysé la polarité d'environ 250 000 tweets afin de détecter le caractère à la hausse de certains échanges sur des actions définies et en déduire le volume d'échange du jour suivant. Les chercheurs soulèvent la

difficulté de comprendre comment l'information est considérée au sein des microblogs tels que Twitter. Ils mentionnent que « choisir les bons tweets est aussi difficile que d'effectuer les bonnes transactions » sur les marchés financiers.

(Vincent and Armstrong 2010) ont montré que les points de rupture observés sur des marchés extrêmement liquides tels que le marché des changes, correspondent aux buzz sur Twitter qu'ils nomment les « alertes Twitter ».

(Mao, Counts, and Bollen 2011) ont comparé le pouvoir prédictif de différents indicateurs d'humeurs avec le Dow Jones Industrial Average (DJIA), le volume d'échange et l'indice de volatilité du marché (VIX). Ces indicateurs d'humeurs sont basés sur (1) des sondages auprès d'investisseurs (Investor Intelligence, Index quotidien de sentiment sur les marchés), les réseaux sociaux (volume de tweet et un indicateur du sentiment des investisseurs sur twitter), des nouvelles dans les médias, et des volumes obtenus dans les recherches Google (Google trends). Pour la partie d'analyse de sentiment sur Twitter, les chercheurs ont considéré les tweets contenant le terme « bullish » et les tweets contenant le terme « bearish ». Ils ont alors construit un indicateur du sentiment des investisseurs sur Twitter (TIS), calculé comme suit : $TIS = \text{nombre de termes bullish} / (\text{Nombre de tweets contenant le terme bullish} + \text{Nombre de tweets contenant le terme bearish})$. Ils ont également déterminé le volume de tweets de certains termes financiers considérés comme des proxy de l'humeur des traders par la revue de littérature. Ces termes financiers comportent par exemple les termes suivants : « récession », « faillite », « action à acheter », « à la hausse », « dow jones », « nouvelles financières ». Les résultats obtenus indiquent que les sondages ont un pouvoir prédictif faible comparé aux nouvelles des médias, à Google Trends et à Twitter. Ces trois derniers possèdent un pouvoir de prédiction significatif sur les marchés. Cependant, Twitter démontrait une meilleure performance dans la prédiction des tendances du DJIA comparativement à la performance Google Trends.

Afin de déterminer si Twitter était corrélé avec le DJIA, (Bollen, Mao, and Zeng 2011) se sont servis d'OpinionFinder pour déterminer la polarité des tweets. Ils ont aussi considéré un autre outil pour l'analyse de sentiment : GPOMS. GPOMS est un lexique composé de 964 termes qui sont annotés respectivement avec une des classes d'humeur suivantes : « calme », « alerte », « confiant », « vital », « gentil », « content ». Les résultats de leurs travaux indiquent que les variations de l'indice du DJIA peuvent être prédites 3 à 4 jours en avance par le changement

d'humeurs du public représenté par la classe d'humeur « calme ». De manière surprenante, les données d'humeur provenant d'OpinionFinder (humeur dénotée comme positive ou négative), n'avaient pas un bon pouvoir prédictif dans la détection des variations de l'indice DJIA .

(Zhang, Fuehres, and Gloor 2011) ont aussi conduit des recherches afin de prédire l'indice du Dow Jones, du NASDAQ, du S&P500 et la volatilité VIX, en utilisant l'analyse de sentiment sur les tweets. Pour la tâche d'analyse de sentiment, ils ont utilisé des termes d'humeurs tels que « peur, inquiétude, espoir etc., comme tags émotionnels d'un tweet ». Pour chaque tweet, la polarité a été déterminée en calculant le nombre de tags émotionnels positifs et négatifs. Si le nombre de tags émotionnels positifs était supérieur au nombre de tags émotionnels négatifs alors le tweet était catégorisé comme positif. Leur recherche indique que les gens ont globalement tendance à employer des mots à caractère émotionnel en temps d'incertitude économique, que le contexte soit positif ou non. Les résultats de leur recherche indiquent que les termes d'humeurs dont ils se sont servis pour conduire l'analyse de sentiment ne sont pas si efficaces qu'ils auraient pu penser, alors qu'ils ont obtenu des coefficients de corrélations très faible entre la polarité des tweets et les marchés boursiers (Dow Jones, NASDAQ, S&P500 et VIX).

(Brown 2012) a également tenté de démontrer l'existence d'une corrélation significative entre Twitter, les marchés boursiers, en se basant sur le sentiment des tweets, le volume de tweets, et l'impact de la réputation de certains utilisateurs sur le sentiment général au sein de Twitter et sur les cours boursiers. Pour réaliser l'analyse de sentiment, ils ont utilisé un algorithme de classification de type Naïve Bayes ainsi que la base de donnée conçue par (Hu and Liu 2004).

(Ruiz et al. 2012) ont développé une approche graphique pour représenter l'information de Twitter sur les messages au sujet d'une action en particulier. Ils ont extrait certaines caractéristiques de l'activité sur Twitter telles que par exemple le nombre de retweets, le nombre de tweets avec géolocalisation, le nombre de tweets mentionnant un utilisateur présent antérieurement dans le graphe développé. Leurs résultats indiquent que ces caractéristiques étaient corrélées dans une moindre mesure avec le prix de l'action, mais plutôt avec le volume d'échange de l'action.

(Smailovic, Grcar, and Znidarsic 2012) ont utilisé l'analyse de sentiment sur les tweets et le volume de tweets pour identifier d'important évènement et prédire le prix de l'action d'Apple. Pour la tâche d'analyse de sentiment, ils ont utilisé un classificateur de type « Supervised Vector

Machine » (SVM) fonctionnant avec la base de donnée de (Go, Huang, and Bhayani 2009). Leurs résultats indiquent qu'il est possible de prédire les changements du prix de l'action d'Apple deux jours à l'avance en se basant sur les variations du sentiment positif des tweets indiquées par l'analyse de sentiment.

(Asur and Huberman 2010) ont développé un modèle basé sur les messages de Twitter afin de prédire les ventes de billets de cinéma. Ils ont conçu un premier modèle basé sur le taux de tweets par minute portant sur un film donné présentement à l'affiche dans les salles de cinéma. Le second modèle construit, se base sur l'analyse de sentiment des tweets du film à l'étude. L'analyse de sentiment a été réalisée avec un classificateur DynamicLCM. Ils ont ensuite calculé pour un film donné, un ratio de polarité (PNratio) calculé comme le quotient du nombre de tweets positifs (numérateur) sur le nombre de tweets négatifs (dénominateur). Les résultats de leur analyse indique le ratio de tweet par minute est un meilleur indicateur afin de prédire les ventes de film que la polarité des tweets. Cependant, combiner le ratio de tweet par minute et la polarité des tweets dans le modèle améliore les performances par rapport à utiliser le ratio de tweet par minute individuellement.

À notre connaissance, aucun travail n'a été réalisé sur l'utilisation de twitter afin d'estimer la réputation d'entreprise. C'est en ce sens que le présent mémoire en tire son originalité.

1.3.4 Développement d'outils en ligne d'analyse de sentiment

Plusieurs applications en ligne ont été développées pour conduire des analyses de sentiments sur les messages de Twitter. Une des premières applications fut l'outil « The Summize », qui a par la suite été acquise en 2008 par Twitter. À notre connaissance, les applications publiquement disponibles sont : Tweetfeel⁸, SNTMNT⁹, Sentiment140¹⁰, et Social mention¹¹. L'algorithme de

⁸ <http://www.tweetfeel.com/>

⁹ <http://www.sntmnt.com>

¹⁰ <http://www.sentiment140.com>

¹¹ <http://socialmention.com/#>

classification de Sentiment 140 est basé sur la méthode d'entropie maximale. Le logiciel peut être retrouvé sur le site web du groupe du traitement de langage naturel de l'université Stanford¹².

D'autres applications retrouvées dans la revue de littérature ne sont pas encore disponibles au public (ou n'existe tout simplement plus), on retrouve Twisent développée par (Mukherjee et al. 2012), Twitratr mentionnée par (Efthymios Kouloumpis 2011), ou Twendz mentionnée par (Barbosa and Feng 2010) et développée par la firme Waggner Edstrom. Waikato Environment Knowledge Analysis¹³ (WEKA) est un logiciel libre développé par l'université de Waikato en Australie. WEKA offre une collection d'algorithmes d'apprentissage machine pour les tâches d'exploration de donnée ou de texte. WEKA peut fournir des outils de classification, régression, regroupement de donnée, de visualisation, etc. (Witten and Frank 2005) ont écrit un livre décrivant de manière détaillé cet outil. Pour plus d'information sur WEKA, (M. Hall et al. 2009) apportent une description sur l'histoire de la conception de l'outil en incluant également une discussion sur les récents développements obtenus par la communauté scientifique utilisant le logiciel. Dans la revue de littérature, WEKA a été utilisé par (Bifet and Frank 2010; Barbosa and Feng 2010) pour l'analyse de sentiment des messages de Twitter.

L'outil « Massive Online Analysis » (MOA) est également un autre logiciel libre offrant des algorithmes d'apprentissage machine pour l'exploration de données. MOA a été développé par (Bifet et al. 2010). Le logiciel peut performer des tâches similaires à WEKA. MOA est d'ailleurs compatible avec WEKA, ainsi il est possible pour les développeurs d'utiliser des classificateurs de WEKA vers MOA et MOA vers WEKA¹⁴.

Il existe aussi un autre logiciel libre permettant de réaliser des tâches en analyse de sentiment : il s'agit de R. Le logiciel a été conçu par Robert Gentleman et Ross Ihaka à la fin des années 1990s. C'est ce logiciel qui a été utilisé dans la conception de l'algorithme d'analyse de sentiment qui va être réalisée dans le cadre de la présente recherche. Le logiciel R est également connu pour la conduite d'analyses statistiques.

¹² <http://nlp.stanford.edu/software/classifier.shtml>

¹³ <http://www.cs.waikato.ac.nz/ml/index.html>

¹⁴ <http://moa.cms.waikato.ac.nz/>

1.3.5 Résumé de la littérature sur le concept d'analyse de sentiments

L'analyse de sentiment vise à déterminer la polarité subjective d'un texte (ou d'une phrase) en le caractérisant comme positif, négatif ou neutre. Il s'agit encore d'un nouveau champ de recherche dont les méthodes d'analyse sont encore en développement. La littérature n'est pas encore parvenue à un consensus sur les meilleures pratiques en analyse de sentiment. Cependant, son utilisation dans différents contextes a pu démontrer son utilité : en politique pour prédire les résultats des élections dans certaines régions, ou encore en finance pour prédire le comportement boursier de certains indexes ou titres boursiers. Ces applications indiquent qu'il est possible de recueillir de l'information supplémentaire auprès du public (consommateurs) à une fréquence plus élevée que celle proposée par les sondages. Des logiciels open source tels que WEKA, MOA ou R ont été développés rendant accessible au public l'utilisation d'algorithmes de classification.

CHAPITRE 2 PROBLÉMATIQUE ET CADRE DE RECHERCHE

2.1 Objectif général de la recherche

La revue de littérature fait mention de deux sources produisant une mesure de la réputation d'entreprise : le magazine Fortune et l'Institut de la Réputation. Elles apportent annuellement un classement des entreprises possédant les meilleures réputations sur la scène internationale. Toutefois, les méthodologies suivies par chacune de ces deux entités dans l'élaboration de leur mesure de réputation ne tiennent pas compte de l'activité sur les réseaux sociaux. Pourtant, la revue de littérature indique que les réseaux sociaux peuvent avoir un impact dans les décisions d'achat sur les consommateurs et donc indirectement sur la réputation d'entreprise. De plus, les réseaux sociaux possèdent des données qui sont accessibles sur une base quotidienne en temps réel, ce qui rend possible un suivi quotidien de l'opinion générale des utilisateurs (proxy de l'opinion publique) sur une entreprise.

Avec ses centaines de millions d'utilisateurs, Twitter est le réseau social de prédilection pour récolter des données en temps réel au travers de sa plateforme d'accès offerte pour les développeurs. Le principal objectif de cette recherche est de parvenir à la construction d'une mesure, proxy de la réputation d'entreprise à partir des messages textes de Twitter (tweets). Ce champ de recherche couvre plusieurs domaines d'études tels que la réputation, l'image de marque, le data-mining et l'analyse de sentiment. Il est à noter que ce travail est le premier à conduire une recherche afin de fournir une mesure de la réputation par Twitter.

L'objectif général de la recherche est d'offrir une mesure de la réputation d'une entreprise observable à une fréquence plus élevée que celle annuelle, voir à une fréquence proche du temps réel.

Il est important de préciser que lorsque l'on parle de « mesure » il s'agit d'un score comme peuvent le fournir le magazine Fortune et l'Institut de la Réputation et non une valeur monétaire comme peuvent l'offrir Interbrand ou Brand-Finance en ce qui a trait à l'équité de marque.

Nous tenterons d'atteindre l'objectif général énoncé en abordant le problème au travers de deux sous-questions de recherche définies ci-après.

2.2 Mesures actuelles de la réputation d'entreprise

Avant d'aborder la question principale de recherche qui vise à satisfaire l'objectif général de la recherche présenté précédemment, nous allons nous attarder sur une question de recherche d'un ordre inférieure. La littérature mentionne que la réputation est un actif intangible (De Marcellis-Warin and Teodoresco 2012), et que de plus la réputation est positivement corrélée avec les performances financières d'une entreprise (Chun, 2005). Ainsi la réputation est difficilement quantifiable, mais observe une corrélation avec la performance de l'entreprise qui est captée par le cours de bourse de l'entreprise. Le cours de bourse est une valeur tangible que nous pouvons mesurer, ainsi cette variable peut nous donner une indication quant à la précision des mesures actuelles de réputation dans un premier temps et dans un deuxième temps quant à la précision de la mesure de réputation qui sera conçue avec Twitter. Nous pourrions alors comparer les différentes mesures entre elles. Ainsi la première sous-question de recherche est la suivante :

Est-ce que les mesures de réputation proposées actuellement ont un impact sur le cours de bourse?

Cette question de recherche s'insère dans la continuité d'autres questions de recherche traitées dans la revue de littérature s'intéressant aux divers impacts de la réputation sur les résultats financiers d'une entreprise.

La revue de littérature indique que la réputation a un impact sur le cours de bourse (Chun, 2005), ainsi donc un changement dans les mesures de réputation devrait se traduire également par un changement dans le même direction du cours de bourse des entreprises : la relation est intuitivement positivement corrélée. Ainsi si les mesures de réputation sont à la hausse, on devrait observer un cours de bourse à la hausse également, et si les mesures de réputations sont à la baisse, on devrait observer un cours de bourse à la baisse. L'amplitude dans le changement de valeur des mesures de réputation n'est pas nécessairement conservée dans le changement de valeur du cours de bourses. On s'intéressera donc principalement à regarder les tendances à la hausse ou à la baisse sans nécessairement les quantifier.

Il serait intéressant de déterminer qui parmi le magazine Fortune et l'Institut de la Réputation, fournit la « meilleure » mesure de la réputation d'entreprise. L'objectif visé est de parvenir à mettre en perspective l'évolution du cours boursier et l'évolution de la mesure de réputation annuelle fournie par chacune des deux institutions. Dans cette section, les évaluations de l'image de marque fournies par Brand-Finance et Interbrand seront aussi comparées : ces dernières fournissent une évaluation monétaire des valeurs d'équité de marque des entreprises, ainsi il serait intéressant de voir si ces évaluations de l'image de marque semblent cohérentes avec l'évolution du cours de bourse des entreprises. Enfin dans un troisième temps, nous allons introduire la mesure de la responsabilité sociale fournie par le magazine Corporate Knight. Elle sera également analysée vis-à-vis du cours de bourse des entreprises.

Cette première partie nous permettra de mettre en perspective les mesures de réputation qui existent, de mieux en cerner leur signification, et surtout de déterminer la relation les liant avec le cours de bourse.

Tout ceci nous mène donc à nos quatre premières hypothèses de recherche.

Hypothèse 1 :

Les mesures actuelles de la réputation d'entreprise, d'évaluation monétaires de l'image de marque et de responsabilité sociale ont un impact significatif sur le cours de bourse.

Hypothèse 2 :

Les mesures actuelles de la réputation d'entreprise ont un impact plus significatif sur le cours de bourse que les évaluations monétaires de l'image de marque, car la réputation d'entreprise englobe davantage de caractéristiques d'une entreprise que l'image de marque.

Hypothèse 3 :

Les mesures actuelles de la réputation d'entreprise ont un impact plus significatif sur le cours de bourse que la mesure de responsabilité sociale car la réputation d'entreprise englobe davantage de caractéristiques d'une entreprise que la responsabilité sociale.

Hypothèse 4 :

Les évaluations monétaires de l'image de marque d'une entreprise ont un impact plus significatif sur le cours de bourse que la mesure de responsabilité sociale, car la valeur de la

marque est une mesure plus proche de la valeur d'entreprise que le degré de responsabilité sociale d'une entreprise.

2.3 Une application des données de Twitter à la réputation d'entreprise

Nous en venons donc à la principale question de recherche. :

Les données non structurées offertes par Twitter peuvent-elles conduire à la construction d'une mesure de la réputation?

Nous avons soulevé dans la revue de littérature que ni le Fortune, ni l'Institut de la Réputation ne prennent en compte les réseaux sociaux dans l'élaboration de leur mesure de la réputation. Pourtant, il s'avère que les réseaux sociaux ont un impact sur la réputation (Jones, Temperley and Lima, 2009 ; Grützmacher, 2011) .

Les données non-structurées font référence aux données qui ne sont pas organisées sous une forme directement analysable : par exemple du texte, des vidéos, des fichiers audio etc. Les données structurées font référence aux données directement analysables : par exemple des valeurs analogiques ou binaires. Ainsi les données non structurées que l'on retrouve sur les réseaux sociaux représentent une catégorie de données complémentaires aux données structurées existantes.

Il est donc ici question de déterminer la pertinence des informations récoltées sur Twitter dans la conception d'une mesure de la réputation d'entreprise basée sur des données non structurées. Intuitivement, l'analyse de sentiment des tweets d'une entreprise est utile à la détermination de la réputation d'une entreprise. Ainsi si la proportion de messages positifs d'une entreprise augmente au fil du temps, on pourrait penser que la qualité de la marque perçue par les consommateurs augmente, et donc la réputation de l'entreprise. De plus, (Aaker and Jacobson 1994) indiquent qu'une amélioration de la qualité perçue est positivement corrélée avec une augmentation du cours de bourse. À l'inverse, on s'attend à ce que la réputation d'entreprise diminue lorsque la proportion de messages négatifs augmente.

Une autre variable qui peut être intéressante à observer est la présence sur Twitter d'une compagnie. La présence sur Twitter d'une compagnie peut être définie par (1) le nombre d'utilisateurs qui suit quotidiennement les activités de l'entreprise sur Twitter, ainsi que par (2) le volume de messages qui traitent de la compagnie. Intuitivement lorsqu'une compagnie enregistre une présence de plus en plus forte sur les réseaux sociaux, cela signifie que son degré de notoriété augmente, tout en sachant que la notoriété est reliée à la réputation d'entreprise selon la revue de littérature.

Nous pouvons poser trois hypothèses de recherche supplémentaires :

Hypothèse 5 :

La proportion de tweets positifs est positivement corrélée avec le cours de bourse (et donc positivement corrélée avec la réputation d'entreprise).

Hypothèse 6 :

La proportion de tweets négatifs est négativement corrélée avec le cours de bourse (et donc négativement corrélée avec la réputation d'entreprise).

Hypothèse 7 :

Lorsqu'une compagnie enregistre une présence de plus en plus forte sur les réseaux sociaux, cela signifie que sa réputation augmente et donc son cours de bourse également.

2.4 Cadre conceptuel : une représentation graphique des hypothèses

Les liens entre les thèmes présentés dans la revue de littérature et les hypothèses énoncées dans les sections précédentes sont illustrés à la figure suivante. Les sept hypothèses sont indiquées par les encadrés H1 à H7. Une double flèche indique une interrelation. Une flèche simple indique qu'un thème a un impact, ou effectue une action sur un autre thème.

Ainsi au centre se retrouve la réputation dont l'évolution est suivie au travers de l'évolution du cours de bourse. Rappelons encore une fois que la réputation est intangible, on ne peut donc pas simplement prétendre pouvoir suivre l'évolution de la réputation sans se référer à un indicateur tangible qui s'avère être corrélé avec la réputation. Notre indicateur ici est le cours de bourse de

l'entreprise, dont son évolution dans le temps nous servira de proxy à l'évolution la réputation tel que mentionné précédemment.

À gauche des thèmes de la réputation et cours de bourse, se trouvent les trois thèmes de réputation d'entreprise, d'image de marque, et de responsabilité sociale avec les firmes qui fournissent une mesure ou une évaluation à chaque année. Les mesures de réputations, fournies par le magazine Fortune et l'Institut de la Réputation, sont censées refléter la réputation d'entreprise dans son ensemble. Tandis que les évaluations d'image de marque données par Interbrand et Brand-Finance, et la mesure de responsabilité sociale donnée par Corporate Knight, sont censées refléter la réputation en partie. En effet, l'image de marque et la responsabilité sociale ne sont qu'une partie de la réputation d'entreprise tel qu'observé dans la revue de la littérature. Les trois thèmes sont reliés au cours de bourse car nous nous attendons à ce qu'il y ait une relation avec l'évolution du cours de bourse (hypothèse 1). Nous nous attendons également à ce qu'une mesure de réputation influence davantage l'évolution du cours de bourse (donc de la vraie réputation d'entreprise) qu'une évaluation d'image de marque (hypothèse 2), car la seconde constitue une partie de la première. Nous nous attendons également à ce qu'une mesure de la réputation influence davantage l'évolution du cours de bourse de l'entreprise (donc de la vraie réputation d'entreprise) qu'une mesure de la responsabilité sociale (hypothèse 3). Ensuite nous nous attendons à ce qu'une évaluation de l'image de marque influence davantage l'évolution du cours de bourse, qu'une mesure de responsabilité sociale (hypothèse 4), car l'équité de marque est une valeur plus importante et tangible que la responsabilité sociale de l'entreprise, telle qu'observée dans la littérature. Tout ceci constitue la première partie analytique du mémoire.

À droite des thèmes de la réputation et du cours de bourse, se trouve le bloc Twitter. Nous allons nous servir des données non structurées offertes par Twitter, les traiter avec l'analyse de sentiment et d'autres méthodes afin de créer une mesure de la réputation d'entreprise. Nous nous attendons à ce que cette mesure ait une influence sur le cours de bourse de l'entreprise (et donc sur sa réputation). Ceci constitue la seconde partie analytique du mémoire, qui cherche à répondre à l'objectif général de la recherche.

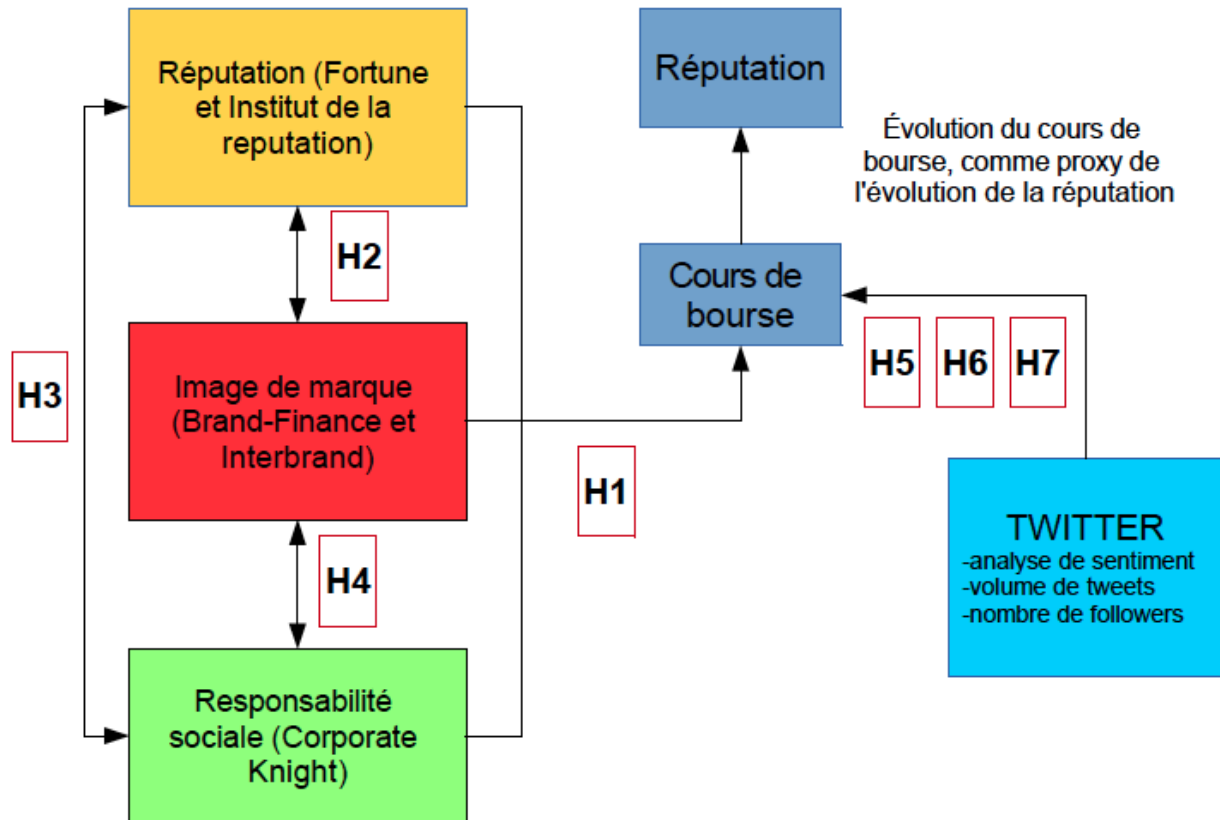


Figure 2.1 Cadre conceptuel de la recherche

CHAPITRE 3 LES MESURES ACTUELLES DE LA RÉPUTATION D'ENTREPRISE

Dans ce chapitre, il est question de comparer les différentes mesures existantes de la réputation d'entreprise (Fortune, Institut de la Réputation, Corporate Knight), ainsi que les évaluations de l'image de marque (Interbrand, Brand-Finance). Le but est de déterminer laquelle de ces mesures ou évaluations démontre une relation significative avec le cours de bourse: cette source sera alors considérée comme la mesure la plus fiable de la réputation. Le but ultime est de pouvoir comparer la meilleure mesure de réputation actuelle avec la mesure de réputation qui sera construite avec Twitter. Pour parvenir à déterminer la meilleure mesure de réputation actuellement disponible, cinq sources ont été recensées : Interbrand, Brand-Finance, Corporate Knight Magazine, Fortune & Money, l'Institut de la Réputation. Pour être plus précis, Interbrand et Brand-Finance fournissent une évaluation monétaire de l'équité de marque. Le magazine Corporate Knight fournit une mesure (score) de la responsabilité sociale des entreprises (partie intégrante de la réputation d'entreprise). La magazine Fortune & Money, ainsi que l'Institut de la Réputation fournissent une mesure de la réputation d'entreprise. Même si ces trois premières sources n'évaluent pas directement la réputation, elles l'évaluent indirectement en fournissant une valeur ou mesure d'une composante de la réputation d'entreprise telle que définie dans la revue de littérature. En effet, l'image de marque est une des trois composantes de la réputation selon (Chun 2005). C'est la raison pour laquelle elles ont tout de même été considérées dans le cadre de la recherche. Il peut être intéressant de les comparer aux mesures de la réputation fournit par l'Institut de la Réputation et Fortune & Money.

3.1 Présentation des mesures de la réputation d'entreprise et évaluation d'image de marque d'une entreprise

Dans cette section, les cinq sources de mesures de la réputation d'entreprise et évaluation de l'image de marque d'entreprise sont décrites.

3.1.1 Interbrand

Interbrand est une firme de conseil spécialisée dans la gestion de marque. La firme a été fondée en 1974 et possède aujourd'hui plus de trente bureaux au travers le monde. Chaque année depuis l'an 2000, Interbrand publie un classement des cent marques ayant l'équité de marque la plus importante. Le classement est publié à la même période chaque année soit vers la fin du mois de septembre ou début du mois d'octobre. Leur méthodologie fut la première des méthodologies certifiée ISO 10668 : évaluation d'une marque / exigences pour l'évaluation monétaire d'une marque. Leur méthodologie d'évaluation des marques est illustrée sur la figure suivante.

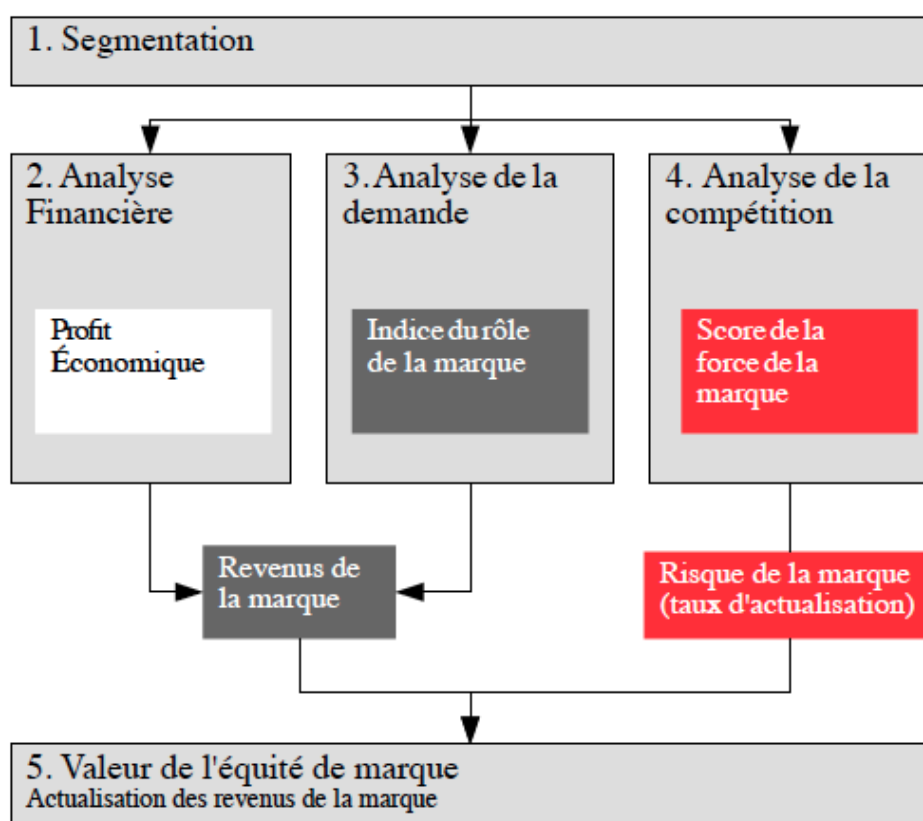


Figure 3.1 Modèle d'évaluation de l'équité de marque d'Interbrand (adapté d'Interbrand)

Selon Interbrand, l'évaluation d'une marque repose sur trois éléments essentiels : la performance financière, le rôle de la marque, et la force de la marque. La première étape consiste en une segmentation des différentes parties d'une compagnie afin de pouvoir analyser de manière plus précise l'impact des trois composants de l'évaluation d'une marque. L'analyse est réalisée au

niveau d'un segment d'une entreprise et non dans son ensemble : en effet une compagnie peut être composée de plusieurs marques formant les lignes de produit de la compagnie. La seconde étape consiste en une analyse financière de la marque. Plus précisément, il s'agit de mesurer le profit économique défini comme étant le bénéfice d'exploitation après taxes moins le coût en capital utilisé pour générer les revenus et les marges au travers de la marque. La troisième étape consiste à mesurer l'influence de la marque dans la décision d'achat du consommateur comparativement à d'autres facteurs tel que le prix, ou les caractéristiques du produit. Pour y parvenir, Interbrand se sert de trois méthodes définies ci-après par ordre de préférence : (1) recherche primaire statistique basée sur une modélisation de choix auprès du consommateur, (2) recherche existante combinée avec l'opinion d'Interbrand sur l'influence d'une marque, (3) évaluation qualitative basée sur le rapport de gestion des entreprises et l'expérience passée. Cette dernière méthode est généralement utilisée lorsqu'il n'y a pas eu de recherche effectuée par le passé. La quatrième étape consiste en une analyse compétitive de la force d'une marque défini comme l'habilité d'une marque à susciter de la loyauté chez les consommateurs, et ainsi de continuer à générer de la demande et des bénéfices dans le futur. La force d'une marque est évaluée au travers de 10 critères séparés en facteurs internes et facteurs externes. Les facteurs internes sont : la clarté de la marque au sein de l'organisation (son message, ses valeurs, sa clientèle cible, etc.), l'engagement de l'organisation dans la marque (support en terme d'investissements, de temps, et d'influence), la protection de la marque, l'habilité à s'adapter au changement du marché. Les facteurs externes sont : l'authenticité de la marque, la pertinence de la marque, la différenciation, la consistance de la marque, la présence de la marque, la compréhension de la marque. La cinquième étape consiste à actualiser les revenus générés au travers de la marque par un taux d'actualisation spécifique à la marque. Le résultat est alors égal à la valeur monétaire de la marque.

Pour illustration, la figure 3.3 montre que la compagnie correspondant au id = 2 (Adidas) a une évaluation monétaire égale à 3,65 milliards de dollars en 2001 et 7,535 milliards de dollars en 2013.

3.1.2 Brand-Finance

Brand-Finance est une firme de conseil spécialisée dans la gestion de marque et d'autres actifs intangibles. La firme a été fondée en 1996, et possède aujourd'hui plus de 20 bureaux au travers le monde. Chaque année Brand Finance établit un classement recensant la valeur monétaire de 500 marques internationales. Le classement est publié chaque année au début de l'année, durant le mois de janvier. Leur méthodologie d'évaluation de marque est basée sur trois méthodes différentes certifiées ISO : une approche par comparaison au marché, une approche par coût, et une approche par revenus générés. L'approche par comparaison au marché détermine la valeur d'une marque en la comparant avec les transactions passées (achats ou ventes) de marques similaires. L'approche par coût détermine la valeur d'une marque en calculant le coût pour recréer une marque ayant une utilité économique similaire.

L'approche par revenus générés détermine la valeur d'une marque en utilisant la méthode « Royalty Relief » définie comme l'actualisation des estimations des ventes futures en appliquant un taux d'actualisation appelée taux de royauté. Le taux de royauté est le taux exigé qu'une firme locateur demanderait à une seconde firme locatrice, pour l'utilisation de sa marque. La méthode est illustrée ci-dessous :

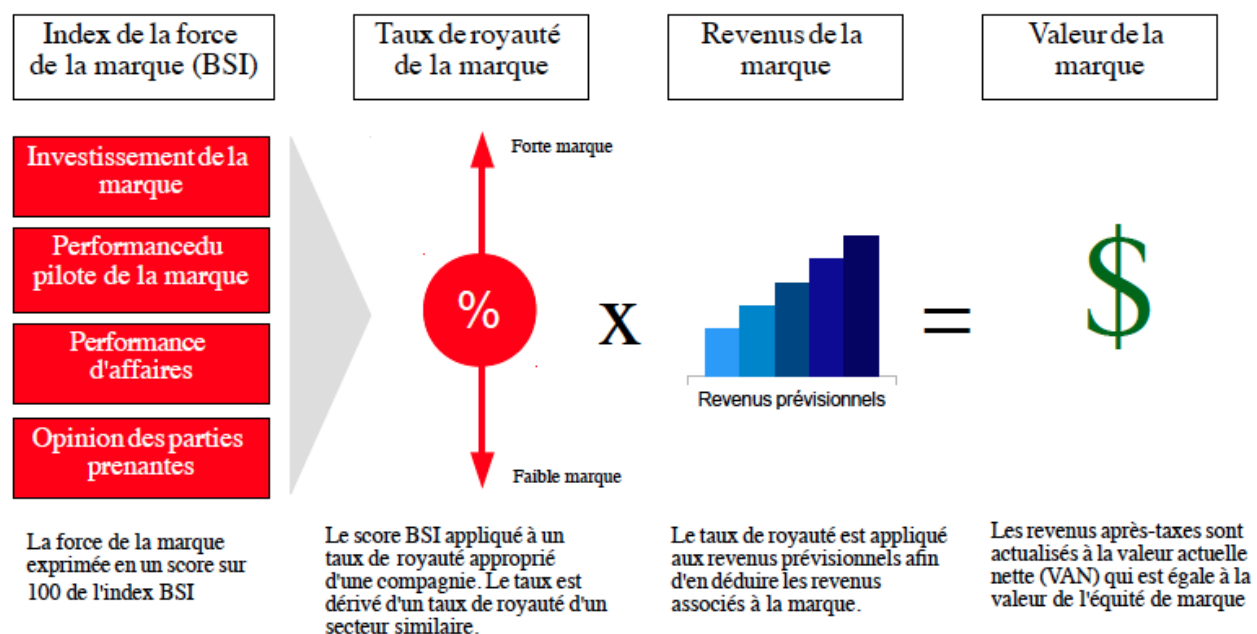


Figure 3.2 Méthodologie d'évaluation de l'équité de marque de Brand-Finance (adapté de Brand-Finance)

La méthode « Royalty Relief » est divisée en sept étapes. La première étape consiste à évaluer la force de la marque sur une échelle de 0 à 100 selon des critères tels que par exemple la connexion émotionnelle que suscite la marque, les performances financières actuelles de la marque, sa durabilité, etc. La seconde étape consiste à déterminer le taux de royauté de l'industrie au sein de laquelle la compagnie entretient ses activités. La tâche est réalisée en se basant sur les accords actuels de licence : les données sont fournies par des bases de données externes de Brand Finance. La troisième étape consiste à calculer le taux de royauté de la compagnie étudiée, de la manière suivante : le taux de royauté compagnie est égale au produit du taux de l'industrie (en %) multiplié par la force de la marque (score sur 100). Ensuite, la quatrième étape consiste à identifier les proportions des revenus générés par la marque parmi les revenus totaux de la compagnie mère à laquelle la marque appartient. La cinquième étape consiste à calculer les prévisions à long terme des revenus générés par la marque selon les données historiques des revenus générés par la marque. La sixième étape consiste à calculer la dépense en royauté pour l'utilisation de la marque en multipliant les prévisions de ventes futures (calculées à l'étape 4) au taux de royauté (calculé à l'étape 3). La septième étape consiste à finalement calculer la valeur monétaire de la marque en actualisant les dépenses après taxes en royautés calculées à l'étape 6. Le taux d'actualisation est le taux de royauté calculé à l'étape 4.

Pour illustration, la figure 3.3 indique que la compagnie correspondant au id=2 (Adidas) a une évaluation monétaire de sa marque égale à 5.26 milliards de dollars en 2007 et 5.79 milliards de dollars en 2013.

3.1.3 Le magazine Corporate Knights

Le **magazine Corporate Knights** est un magazine nord-américain promouvant un capitalisme vert définit comme « un système économique, dans lequel les prix tiennent compte des bénéfices et coûts de nature sociale, économique, et environnementale, et dans lequel les acteurs sont pleinement conscients des impacts de leurs actions »(Corporate Knight 2011). Le magazine a été fondé en 2002 et appartient au média Corporate Knights. Chaque année le magazine publie un

classement des entreprises adoptant les meilleures conduites en ligne avec le capitalisme vert. Le classement est publié en début de chaque année durant le mois de Janvier.

Leur méthodologie d'évaluation de la responsabilité sociale d'entreprise se décompose en six phases. La première phase consiste dans la détermination de l'univers des compagnies à classer : toutes les firmes au dessus d'une capitalisation boursière de 2 milliards de dollars sont considérées.

La deuxième phase consiste à évaluer la responsabilité sociale des entreprises selon 12 critères de performance : la productivité en énergie, la productivité en carbone, la productivité en eau, la productivité en déchets, la capacité d'innovation, le pourcentage de taxes payées, le ratio de la paie du CEO versus les employés, l'état du fond de pension, la performance en matière de sécurité, le ratio de roulement d'employés, la diversité au sein de l'équipe de direction, le lien entre le capitalisme vert et la rémunération de l'équipe de direction. Pour chaque secteur d'activité, sont définis des indicateurs prioritaires parmi les 12 indicateurs énoncés ci-dessus. Les firmes d'un même secteur n'affichant pas un score de 75% à tous les indicateurs sont éliminées de l'univers des possibilités.

La troisième phase consiste à évaluer les performances financières des firmes au travers de 9 tests : les profits doivent être positifs, les cash flows opérationnels doivent être positifs, la différence entre le ratio profit net / actif à la fin de l'année et le ratio profit net / actif au début de l'année doit être positive, les cash flows opérationnels doivent être supérieurs au profit net, le ratio de la dette à long terme / actifs n'a pas augmenté, le ratio actif à court terme / passif à court terme a augmenté, aucune augmentation d'actions ordinaires au courant de la dernière année, la marge brute a augmenté au courant de la dernière année, le ratio de rotation de l'actif a augmenté. Les firmes ne satisfaisant pas tous les tests sont éliminées de l'univers des possibilités.

La quatrième phase consiste en éliminant les firmes qui offrent des produits ou services à caractère litigieux tels que le tabac ou les armes.

La cinquième phase élimine les firmes qui ont été pénalisées de manière substantielle à cause de non-respect de norme(s) environnementale(s) ou loi(s) sociale(s), par rapport à leurs pairs de l'industrie dans laquelle elles conduisent leurs activités.

La sixième phase consiste à l'établissement d'un score général selon les 12 critères de performance énoncés ci-dessus. Les 100 compagnies obtenant les meilleurs scores sont conservées dans le classement. Les scores calculés sont sur une base maximale de 100%.

Pour illustration, la figure 3.3 indique que la compagnie correspondant au id=2 (Adidas) a un score de 58% en 2010 et un score de 68% en 2013.

3.1.4 Fortune

Fortune est un magazine américain couvrant l'actualité économique et le monde des affaires. Il a été fondé en 1930 et appartient au conglomérat américain médiatique Time Warner. Money est un magazine couvrant des sujets sur les finances personnelles (épargne, retraite, taxes, crédit, fonds pour payer les frais de scolarité universitaire, etc.). Money a été fondé en 1972 et appartient également à Time Warner. Depuis 1997, Fortune et Money se sont associées pour publier une liste des compagnies les plus admirées. Le classement est publié en début de chaque année approximativement vers le mois de février. Leur méthodologie suit une sélection à plusieurs paliers. Premièrement 1,400 compagnies sont répertoriées, il s'agit des 1000 compagnies américaines les plus importantes en taille et 400 compagnies internationales appartenant à la base de donnée Fortune 500 et possédant un chiffre d'affaires annuel au-dessus de 10 milliards de dollar US. Deuxièmement, au travers des résultats des sondages auprès d'exécutifs et managers séniors des industries représentées parmi les 1400 compagnies, Fortune & Money sélectionne par industrie, les 15 plus importantes firmes internationales et les 10 plus importantes firmes américaines. Chaque compagnie reçoit un score sur 10 au sein de son industrie selon neuf critères incluant la responsabilité sociale, la valeur marchande de la firme, etc. L'étude est conduite chaque année en partenariat avec la firme de consultation en gestion Hay Group.

Pour illustration, la figure 3.3 indique que la compagnie correspondant au id=2 (Adidas) a un score de réputation selon le magazine Fortune, égale à 6.45/10 en 2010, et à 6.71/10 en 2013.

3.1.5 L'Institut de la Réputation

L'Institut de la Réputation est une firme de conseil spécialisée dans la gestion de la réputation d'entreprise. La firme a été fondée en 1997 par Charles Fombrun (ancien professeur à l'université de Pennsylvanie) et Dr Cees van Riel (professeur à l'université Erasmus de Rotterdam). L'Institut de la Réputation a aujourd'hui une forte présence internationale avec plus de 30 bureaux. Chaque année la firme en association avec le magazine Forbes publie le Global RepTrack 100, un classement des 100 compagnies possédant les meilleurs réputations. Le classement est généralement publié durant le mois de septembre. Leur méthodologie d'évaluation de la réputation est basée selon sept dimensions : les produits ou services offerts, l'innovation, l'environnement de travail, l'implication sociale, la gouvernance de l'entreprise, le leadership, et la performance financière. Les scores obtenus sont sur une base maximale de 100.

Pour illustration, la figure 3.3 indique que la compagnie correspondant au id=2 (Adidas) a un score de réputation selon l'Institut de la Réputation égale à 76/100 en 2012 et 74.05/100 en 2013.

3.2 Informations financières sur les marchés

Les informations financières sur les marchés d'échanges des firmes qui composent l'échantillon d'analyse définit précédemment dans le document, sont nécessaires. Ces informations financières comportent : le cours d'ouverture, le cours de fermeture, le volume d'échange, le cours le plus haut et le cours le plus bas. Les cours de bourse peuvent être étudiés sur une échelle de temps continue puisque les données peuvent être fournis en temps réel par les compagnies d'information financières telles que Bloomberg ou Thomson Reuters. Nous avons choisis d'étudier les cours de bourses sur une base journalière plutôt qu'en temps réel, car les données sont accessibles gratuitement sur Yahoo Finance. Les cours de bourse journaliers des firmes composant l'échantillon, sont extraits pour la période couvrant les années recensées par les cinq sources d'évaluation de la réputation soit depuis le 1^{er} janvier 2000 au 31 décembre 2013.

Pour parvenir à réaliser cette tâche un programme informatique a été conçu : il permet également de stocker les données sous forme de panel, ce qui va être utile pour le traitement des données par la suite. Les données sont indiquées à la figure 3.3 dans la colonne intitulée « close ».

3.3 Remarques importantes sur la méthodologie

3.3.1 Difficulté concernant la construction de l'échantillon d'entreprises

La principale difficulté dans l'extraction des données des cinq sources décrites précédemment consistait à définir une nomenclature commune dans la nomination de toutes les firmes pour fin de concordance. Par exemple, l'entreprise Nissan pouvait être nommé « Nissan » par Fortune & Money, ou « Nissan Motor Co Ltd. » par Corporate Knight. Le nom Nissan a été désigné comme base commune pour les cinq sources de réputation. Cette démarche a du être entreprise pour toutes les firmes recensées.

3.3.2 Difficulté concernant le recensement de la valeur d'équité de marque

Nous avons vu que Interbrand et Brand-Finance évaluent la valeur monétaire d'une marque et non la reputation d'entreprise dans son ensemble. Une entreprise peut posséder plusieurs marques (en intégralité ou en partie) et voir chacune de ses marques listées dans le classement d'Interbrand et Brand-Finance. C'est le cas par exemple du groupe automobile Daimler qui possède plusieurs marques de voitures célèbres et à succès tels que : Mercedes-Benz, Chrysler, ou encore Maybach. Daimler possède aussi des parts dans le capital des fabricants automobiles Tesla Motors et Renault-Nissan. Comment donc considérer la valeur de la marque Daimler? La valeur de marque de Daimler sera tout simplement la somme des valeurs des marques qui composent le groupe. Cependant cette considération n'est pas toujours directe, en effet dans le cas de conglomérat tel que Mitsubishi, il n'est pas rare de voir l'entité mère listée dans les classements ainsi que ses filiales en même temps.

3.4 Composition de l'échantillon d'entreprises

L'échantillon d'entreprises a été construit suite à une analyse préliminaire détaillée ci-après. La première étape de l'analyse préliminaire constituait à faire la liste de toutes les firmes recensées au travers des classements de la réputation provenant de cinq sources distinctes : Interbrand, Brand-Finance, Fortune & Money, Institut de la Réputation, Corporate Knight Magazine. Au

total ce sont environ 1,395 compagnies appartenant à 71 industries différentes qui ont été recensées dans les classements couvrant les années de 2000 à 2013. L'extraction des données des différentes sources de réputation a été réalisée manuellement. Il s'agit d'un processus long et fastidieux qui a débouché sur la conception d'une base de données unique.

Dans le tableau ci-dessous, une description sommaire des données récoltées est présentée :

Tableau 3.1 Description sommaire des données récoltées au travers des cinq sources d'évaluation de la réputation

Sources	Nombre de firmes recensées par année	Années de disponibilités des données
Interbrand	100	2000-2013
Brand-Finance	500	2007-2013
Fortune & Money	362	2006-2013
Institut de la Réputation	100	2011-2013
Corporate Knights Magazine	100	2008-2011

Les 71 industries recensées dans cet échantillon préliminaire sont basées arbitrairement sur les industries définies par le Fortune & Money, et sont listées en ANNEXE 8 au tableau A8.1.

Par la suite, nous avons réduit l'échantillon à 101 firmes (nombre défini arbitrairement). Les firmes les plus recensées au travers des cinq sources d'évaluation de la réputation ont été sélectionnées selon la règle suivante : les firmes ayant une apparition dans les classements des cinq sources ont d'abord été sélectionnées (14 au total), ensuite celles qui étaient recensées au travers de 4 sources (23 au total), puis celles recensées dans 3 sources (30 au total), puis celles recensées dans 2 sources (32 au total) et enfin celles recensées dans 1 source (2 au total). Pour les firmes recensées dans 2 et 3 sources différentes, le bassin des possibilités étaient beaucoup plus importants que les autres catégories : ainsi le choix s'est porté sur les firmes possédant le plus d'apparition entre 2000 et 2013. Le choix d'obtenir un nombre de 101 firmes dans l'échantillon final a été déterminé arbitrairement : un échantillon de 101 firmes offre une certaine diversité

d'information et semble de taille raisonnable pour généraliser les résultats qui seront obtenus à la suite des analyses. Au final, ce sont 26 industries sur les 71 initiales qui sont représentées dans notre échantillon de 101 entreprises. Le tableau A8.2 en ANNEXE 8, présente la répartition des 101 compagnies selon les 26 industries. L'échantillon d'entreprises composé de 101 firmes est présenté dans le tableau A8.3 en ANNEXE 8.

3.5 Illustration des bases de données

Les données récoltées au travers des cinq sources citées précédemment (Interbrand, Brand Finance, Fortune & Money, Institut de la Réputation et le Corporate Knights Magazine) ont été regroupées dans une base de donnée dont une illustration est présentée ci-dessous.

id	Close	growth	Increase	Year	Fortune (score sur 10)	Interbrand (en milliards de dollars)	Brand (en milliards de dollars)	Replins (score sur 100)	CorKnight (score en pourcentage)
1	52.88			1	2000				
1	53.92	0.01966717		1	2001				
1	41.73	-0.2260757		0	2002				
1	42.8	0.02564103		1	2003				
1	42.74	-0.0014019		0	2004				
1	42.78	0.00093589		1	2005				
1	47.42	0.1084619		1	2006	6.68			
1	54.12	0.14129059		1	2007	6.75		2.33800	
1	56	0.03473762		1	2008	6.79		3.60500	
1	50.91	-0.0908929		0	2009	6.78		3.47200	
1	50.6	-0.0060892		0	2010	6.68		5.08400	
1	52.76	0.04268775		1	2011	6.6			
1	65.45	0.24052312		1	2012	6.39		69.89	
1	36.96	-0.4352941		0	2013	6.27		66.8	
2	16				2001		3.65000		
2	17.55	0.096875		1	2002		3.69000		
2	27.77	0.58233618		1	2003		3.67900		
2	41.14	0.48145481		1	2004		3.74000		
2	35.37	-0.1402528		0	2005		4.03300		
2	60.27	0.70398643		1	2006		4.29000		
2	79.95	0.32653061		1	2007		4.76700	5.26000	
2	50.39	-0.3697311		0	2008		5.07200	6.78300	
2	55.54	0.10220282		1	2009		5.39700	4.70000	
2	60.44	0.0882247		1	2010	6.45	5.49500	5.70200	58%
2	100.18	0.65751158		1	2011	6.38	6.15400	6.75400	41%
2	144.61	0.4435017		1	2012	6.64	6.69900	7.15000	76
2	236.78	0.63736948		1	2013	6.71	7.53500	5.79000	74.05
3	77.25				2000				
3	27.14	-0.6486731		0	2001				
3	24.47	-0.0983788		0	2002				
3	45.15	0.84511647		1	2003				
3	56.34	0.24784053		1	2004				
3	32.8	-0.4178204		0	2005				
3	37.76	0.15121951		1	2006	7.35			

Figure 3.3 Illustration d'une portion de la base de donnée pour la première partie d'analyse

Les cases grisées représentent les périodes auxquelles aucune mesure ou évaluation n'est disponible, car il n'y a pas eu de classement publié pour ces périodes. Les cases noires représentent les périodes auxquelles les compagnies n'ont pas été recensées dans les classements.

La base de données a été construite sous forme de panel et l'analyse économétrique en panel sera réalisée à l'aide d'un logiciel de traitement de données approprié (ici c'est STATA qui a été choisi). La première colonne représente les identifiants pour chaque compagnie. La deuxième colonne est le cours de bourse. La troisième colonne représente la croissance du cours de bourse d'une année à une autre. La quatrième colonne est la variable binaire créée à partir de la croissance du cours de bourse : elle prend la valeur 1 si la croissance du cours de bourse est positive et 0 si la croissance du cours de bourse est négative. La cinquième colonne représente l'année de l'observation. Les colonnes subséquentes sont les valeurs des évaluations de la réputation (Fortune et Institut de la Réputation) ou image de marque (Brand-Finance et Interbrand) ou responsabilité sociale (Corporate Knight).

3.6 Traitement des données

3.6.1 Définition du critère de comparaison

Nous allons procéder à une analyse statistique afin de déterminer qui parmi les mesures actuelles de réputation (Fortune et Institut de la Réputation), ou les évaluations monétaires de la marque (Interbrand et Brand-Finance), ou de responsabilité sociale (Corporate Knights), constitue le proxy le plus adapté à la réputation d'une entreprise.

La variable dépendante de l'analyse statistique est la croissance du cours de bourse sur deux années consécutives à un point donné dans l'année. En effet, la revue de littérature mentionne que la réputation a un impact sur les performances financières d'une entreprise et donc son cours de bourse de la manière suivante : plus la réputation d'une entreprise s'améliore plus l'entreprise voit sa valeur augmenter (Greyser 1999). Le cours boursier nous servira donc de base commune aux cinq sources de réputation pour fin de comparaison : la source de réputation qui statistiquement semble avoir un impact le plus significatif sur l'évolution du cours de bourse sera désignée comme étant le meilleur proxy actuel de la réputation d'entreprise. Ainsi les variables

indépendantes de l'analyse seront constituées des mesures de réputation, d'évaluation de la valeur monétaire des marques et de mesure de la responsabilité sociale présentées ci-avant.

La principale raison qui a conduit à analyser les observations selon un point fixe de l'année plutôt que de les observer sur une base journalière est la granularité des variables indépendantes mises à disposition. Les mesures de réputation ou d'évaluation de la valeur de marque sont des mesures annuelles et non journalières comme peut l'être la variable dépendante (la croissance du cours de bourse). Cela ne ferait donc pas vraiment de sens d'analyser une variable dépendante journalière avec une variable annuelle. Pour les observations, le point de référence choisi se situe un mois après la sortie du classement émis par chacune des sources : ainsi le point de référence pour le magazine Fortune & Money, Brand-Finance et Corporate Knight est le dernier jour ouvrable du mois de février et le point de référence pour l'Institut de la Réputation et Interbrand est le premier jour ouvrable du mois de Novembre. En effet, si l'on regarde la date de sortie des classements offerts par les cinq sources, certaines divulguent leur classement en début d'année (Brand-Finance, Corporate Knight et le magazine Fortune), et d'autres au début de l'automne (Interbrand et l'Institut de la Réputation). Ainsi, si l'on veut traiter de manière cohérente chaque source, il faut que le point de référence de chacune des sources soit à une distance similaire par rapport à la date de sortie du classement.

Une précision importante est que nous ne tentons pas de déterminer l'amplitude du taux de croissance annuelle de l'action à partir des variables indépendantes données. Nous savons que le taux de croissance de l'action dépend de plusieurs autres facteurs qui ne sont pas nécessairement agrégés dans la construction des mesures de la réputation. Ainsi, au lieu de tenter de déterminer, l'amplitude du taux de croissance du cours boursier (variable dépendante) à partir des cinq proxys de la réputation (variables indépendantes), nous allons plutôt analyser la propension à observer un changement similaire dans l'évolution des valeurs des proxys de la réputation avec le cours de bourse. Ainsi nous chercherons à savoir si lorsqu'une variable indépendante croît (ou décroît) d'une année (année $t-1$) à une autre (année t), observe-t-on également une croissance (ou décroissance) du cours de bourse l'année suivante (année t) ? La régression logistique permet de faire ressortir cette information.

Ainsi la croissance du cours de bourse sera analysée en tant que variable dépendante dichotomique. Une variable dichotomique est une variable pouvant prendre seulement deux

valeurs comme par exemple *oui* ou *non*. Ici, la croissance cours de bourse prendra pour valeur 1 indiquant une hausse du cours de bourse ou la valeur 0 indiquant une baisse du cours de bourse.

3.6.2 Modèle logit

Les données sont regroupées sous forme de panel, avec 101 compagnies et 14 observations (au maximum) par compagnie pour les années 2000 à 2013. Si l'on assume une distribution normale $N(0, \sigma_v^2)$ des variables indépendantes de l'analyse statistique, le modèle mathématique logit avec effets aléatoires v_i pour un panel est le suivant :

$$P(y_{i1}, \dots, y_{in_i} | x_{i1}, \dots, x_{in_i}) = \int_{-\infty}^{+\infty} \frac{e^{-v_i^2/2\sigma_v^2}}{\sqrt{2\pi}\sigma_v} \left[\prod_{t=1}^{n_i} F(y_{it}, z) \right] dv_i$$

où l'on a :

- y_{it} : Croissance du cours de bourse pour la $i^{\text{ème}}$ compagnie à l'année t , par rapport à l'année précédente $t-1$;
- x_{it} : Mesure de réputation ou évaluations de l'image de marque pour la $i^{\text{ème}}$ compagnie à l'année t par rapport à l'année $t-1$;

ainsi que

$$z = x_{it}\beta + v_i$$

$$F(y, z) = \begin{cases} \frac{1}{1 + \exp(-z)} & \text{si } y \neq 0 \\ \frac{1}{1 + \exp(z)} & \text{autrement} \end{cases}$$

3.7 Mesures actuelles de la réputation et cours de bourse : existe-t-il un lien?

Les résultats seront présentés en trois parties : d'une part nous aurons les résultats de l'analyse pour les mesures de réputation du magazine Fortune Money et l'Institut de la Réputation et d'autre part nous aurons les résultats de l'analyse des évaluations de l'image de marque fournies

par Interbrand et Brand-Finance, et finalement nous aurons les résultats de l'analyse pour la mesure de la responsabilité sociale fournie par le magazine Corporate Knight.

3.7.1 Les mesures de la réputation

Deux estimations ont été réalisées pour déterminer quelle mesure actuelle de la réputation semble avoir une influence sur le cours boursier en accord avec la revue de littérature. Pour chaque estimation la variable indépendante est la croissance du cours de bourse d'une année à une autre à un point donné dans le temps. Les estimations sont listées ci-après :

1. Modèle Logit avec pour variable indépendante la mesure donnée par le magazine Fortune (Fortune)
2. Modèle Logit avec pour variable indépendante la mesure donnée par l'Institut de la Réputation (RepIns)

Le premier modèle (tableau 3.2) indique que la mesure de réputation donnée par le magazine Fortune n'est pas significative à un seuil de 10% ($\Pr(> |z|) = 0.198 > 0.1$). Ainsi la mesure fournie par le magazine Fortune n'a à priori pas d'impact sur le cours de bourse selon les résultats obtenus. L'interprétation du signe du coefficient n'est pas requise puisqu'il n'y a pas de significativité apparente.

Tableau 3.2 Modèle logit avec le magazine Fortune pour variable indépendante

<i>Regression logistique- Effet aléatoire</i>		<i>Nombre d'obs = 603</i>			
<i>Variable de groupe: ID</i>		<i>Nombre de groupes = 99</i>			
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>		<i>Moy. obs. par groupe = 6.1</i>			
		<i>Wald $\chi^2(1) = 1.66$</i>			
<i>Log Likelihood = -412.702</i>		<i>Prob > $\chi^2 = 0.1976$</i>			
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>

Tableau 3.2 Modèle logit avec le magazine Fortune pour variable indépendante (suite)

Fortune	-.14655	.11375	-1.29	0.198	-.36950	.07639
_cons	.73657	.76445	0.96	0.335	-.76127	2.234873
/lnsig2u	-10.959	19.763			-49.693	27.77494
sigma_u	.00417	0.4121			1.62e-11	1074612
rho	5.29e-06	.00010			7.97e-23	1

Le deuxième modèle (tableau 3.3) indique que la mesure de réputation donnée par l'Institut de la Réputation n'est pas significative à un seuil de 10% ($\Pr(> |z|) = 0.463 > 0.1$). A priori, la mesure fournie par l'Institut de la Réputation ne possède pas d'impact sur le cours de bourse.

Tableau 3.3 Modèle logit avec l'Institut de la Réputation pour variable indépendante

<i>Régression logistique Effets aléatoires</i>			<i>Nombre d'obs. = 111</i>			
<i>Variable de groupe: ID</i>			<i>Nombre de groupes = 43</i>			
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>			<i>Moy. obs par groupe = 2.6</i>			
			<i>Wald $\chi^2(1) = 0.54$</i>			
<i>Log Likelihood = -71.0535</i>			<i>Prob > $\chi^2 = 0.4626$</i>			
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
RepIns	-.036785	.05008	-0.73	0.463	-.13494	.06137
_cons	3.3132	3.63224	0.91	0.362	-3.806	10.432
/lnsig2u	-13.485	604.78			-1198.8	1171.9
sigma_u	.00118	.35667			4.7e-261	2.9e+254
rho	4.23e-07	.00026			0	

3.7.2 Les évaluations monétaires de l'image de marque

Deux estimations ont été réalisées pour déterminer quelle évaluation actuelle de la valeur monétaire de l'image de marque semble avoir une influence sur le cours boursier en accord avec la revue de littérature. Pour chaque estimation la variable indépendante est la croissance du cours de bourse d'une année à une autre à un point donné dans le temps. Les estimations sont listées ci-après :

1. Modèle Logit avec pour variable indépendante l'évaluation donnée par Interbrand (Inter)
2. Modèle Logit avec pour variable indépendante l'évaluation donnée par Brand-Finance (Brand)

Le premier modèle (tableau 3.4) indique que l'évaluation monétaire de l'image de marque donnée par Interbrand n'est pas significative à un seuil de 10% ($\Pr(> |z|) = 0.460 > 0.1$). La mesure fournie par Interbrand ne possède donc pas d'impact sur le cours de bourse d'une entreprise selon les résultats obtenus.

Tableau 3.4 Modèle logit avec Interbrand pour variable indépendante

<i>Régression logistique Effets aléatoires</i>			<i>Nombre d'obs = 489</i>			
<i>Variable de groupe: ID</i>			<i>Nombre de groupes = 47</i>			
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>			<i>Moy. obs. par groupe = 10.4</i>			
			<i>Wald $\chi^2(1) = 0.55$</i>			
<i>Log Likelihood = -328.59</i>			<i>Prob > $\chi^2 = 0.4597$</i>			
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
Inter	-.00427	.00577	-0.74	0.46	-.015584	.00705
_cons	.47580	.12804	3.72	0.00	.22486	.72675
/lnsig2u	-20.912	811.10			-1610.65	1568.823
sigma_u	.00003	.01167			0	
rho	2.52e-10	2.04e-07			0	

Le deuxième modèle (tableau 3.5) indique que l'évaluation monétaire de l'image de marque donnée par Brand-Finance n'est pas significative à un seuil de 10% ($\Pr(> |z|) = 0.195 > 0.1$). La mesure fournie par Brand-Finance ne possède pas d'impact sur le cours de bourse.

Tableau 3.5 Modèle logit avec Brand-Finance pour variable indépendante

<i>Régression logistique Effets aléatoires</i>		<i>Nombre d'obs. = 576</i>				
<i>Variable de groupe : ID</i>		<i>Nombre de groupes = 96</i>				
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>		<i>Moy. d'obs. par groupe = 6.0</i>				
		<i>Wald $\chi^2(1) = 1.68$</i>				
<i>Log Likelihood = -396.0544</i>		<i>Prob > $\chi^2 = 0.1946$</i>				
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
Brand	0.10556	.008137	1.30	0.195	-.00539	.026504
_cons	-.31076	.13054	-2.38	0.017	-.56661	-.054907
/lnsig2u	-12.146	30.043			-71.030	46.737
sigma_u	.00230	.03461			3.77e-16	1.41e+10
rho	1.61e-06	.00005			4.32e-32	1

3.7.3 La mesure de responsabilité sociale

Les résultats présentés au tableau suivant (tableau 3.6) indique la mesure de responsabilité sociale d'entreprise donnée par Corporate Knight n'est pas significative à un seuil de 10% ($\Pr(> |z|) = 0.201 > 0.1$). Ainsi, la mesure fournie par Corporate Knight n'a pas d'impact sur le cours de bourse selon les résultats obtenus.

Tableau 3.6 Modèle logit avec Corporate Knight pour variable indépendante

<i>Régression logistique Effets aléatoires</i>			<i>Nombre d'obs. =71</i>			
<i>Variable de groupe: ID</i>			<i>Nombre de groupe = 33</i>			
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>			<i>Moy. obs par groupe = 2.2</i>			
			<i>Wald $\chi^2(1) = 1.63$</i>			
<i>Log Likelihood = -47.783</i>			<i>Prob > $\chi^2 = 0.2011$</i>			
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
CorKnight	-3.0926	2.4190	-1.28	0.201	-7.8337	1.6485
_cons	1.8503	1.2779	1.45	0.148	-.65424	4.3549
/lnsig2u	-30.543	2230328			-4371392	4371331
sigma_u	2.33e-07	.26003			0	
rho	1.65e-14	3.69e-08			0	

3.8 Discussions

3.8.1 Discussions sur les résultats obtenus

Lors de la revue de littérature, nous avons identifié des études qui indiquaient une relation linéairement positive entre la réputation des entreprises et les performances financières (Greyser, 1999; Chun, 2005).

Les mesures actuelles de réputation réalisées par l'Institut de la Réputation et le magazine Fortune présentent des résultats assez surprenants : prises individuellement comme variable indépendante du modèle logit, toutes les deux ne sont pas significatives à un seuil de 10%. Donc

les mesures de réputation de l'Institut de la Réputation et du magazine Fortune n'ont pas vraiment de lien avec le cours de bourse.

Plusieurs pistes de réflexions peuvent expliquer ces résultats: (1) les mesures de réputation ne sont pas calculées adéquatement, (2) la réputation d'une entreprise n'est pas le seul facteur influençant la hausse ou la baisse du cours de bourse, (3) les mesures de réputation ne se sont pas destinées à être interprétées prospectivement mais rétrospectivement, c'est-à-dire qu'elles servent simplement à décrire une situation passée de l'entreprise sans suffisamment tenir en compte les perspectives futures des entreprises. Ainsi par exemple, lorsque l'on observe une mesure de réputation dans un classement de 2013, cette mesure reflèterait en fait uniquement l'image de la compagnie sur l'année précédente (2012), année à partir de laquelle les données sont traitées en vue d'obtenir les mesures de réputation de 2013. Finalement, on ne peut pas trancher au sujet de quelle est la meilleure mesure de réputation entre les deux sources (Fortune ou Institut de la Réputation).

En ce qui concerne les évaluations actuelles de la valeur monétaire de l'image de marque, réalisées par Interbrand et Brand-Finance, les résultats sont aussi surprenants: en effet les évaluations de la valeur monétaire de l'image de marque réalisées par Interbrand et Brand-Finance ne s'avèrent pas être significatives à un seuil de 10%. Donc les évaluations de la valeur monétaire de l'image de marque réalisées par Interbrand et Brand-Finance n'ont pas de lien avec le cours de bourse. Finalement on ne peut pas trancher sur laquelle de ces deux évaluations est la meilleure.

Les facteurs potentiels mis en cause dans les résultats obtenus avec Interbrand comportent (1) la méthodologie d'évaluation de marque, (2) l'équité de marque d'une entreprise n'est pas le seul facteur influençant la hausse ou la baisse du cours de bourse, (3) les mesures de l'équité de marque ne se sont pas destinées à être interprétées prospectivement mais rétrospectivement, c'est-à-dire qu'elles servent simplement à décrire une situation passée de l'entreprise sans suffisamment tenir en compte les perspectives futures des entreprises.

En ce qui concerne la responsabilité sociale d'entreprise représentée par la mesure donnée par Corporate Knight, la mesure n'est pas significative à un seuil de 10%. Ceci confirme les attentes à l'encontre de la responsabilité sociale d'entreprise : elle ne consiste qu'en une petite partie de

ce qu'est la réputation d'entreprise dans son ensemble, on ne peut donc s'attendre à ce que la mesure soit significative dans l'évolution du cours de bourse.

3.8.2 Discussion sur la méthodologie employée

Une autre critique peut être faite concernant la méthodologie employée pour parvenir aux résultats présentés précédemment. Vouloir observer la pertinence d'une mesure de réputation uniquement par le niveau du cours de bourse c'est simplifier la complexité de ce que représente la mesure de réputation d'une entreprise. Le cours de bourse ne reflète pas nécessairement directement la valeur d'une entreprise : le cours de bourse peut être manipulé à la hausse ou à la baisse par une firme en rachetant des actions en circulation (pénurie d'action fait augmenter l'action) ou en émettant des actions (abondance d'action en circulation fait diminuer le cours de l'action). Cette remarque fait référence à l'hypothèse (2) énoncées dans la section précédente concernant les facteurs influençant à la hausse ou à la baisse le cours de bourse. Ces biais peuvent venir influencer grandement les résultats observés. Ainsi l'une des principales hypothèses implicitement faites en utilisant la valeur du cours de bourse comme approximation de la valeur d'entreprise est le volume d'action en circulation constant au travers des années. Tout en étant conscient de ces contraintes, la méthodologie utilisée fut la seule méthodologie exécutable avec les ressources possédées dans le cadre de cette recherche.

Nous rajouterons également que la base de données constitue également une limite dans le traitement statistique des données. Il manque plusieurs données à différents endroits du panel car les 101 firmes constituant l'échantillon n'ont pas nécessairement été recensées au sein des cinq sources de mesures de réputation, d'évaluation d'image de marque ou de mesure de responsabilité sociale.. Nous avons toutefois plusieurs centaines d'observations malgré les nombreuses données manquantes ou inexistantes.

3.9 Conclusion

La sous-question de recherche de cette première partie était : *Est-ce que les mesures de réputation proposées actuellement ont un impact sur le cours de bourse?*

La réputation est selon la littérature positivement corrélée avec les performances financières (Greyser, 1999; Chun, 2005). Cependant on ne peut pas affirmer que les mesures de réputation existantes, ainsi que la mesure de responsabilité sociale ont un impact sur le cours de bourse. En effet, d'après nos analyses ni la mesure de l'Institut de la Réputation, ni la mesure du magazine Fortune, ni la mesure de Corporate Knight ne démontrent de lien significatif avec la tendance à la hausse ou à la baisse du cours de bourse. L'hypothèse 1 énoncée dans le chapitre 3 est donc rejetée.

Ces résultats indiquent que nous ne sommes pas en position de pouvoir indiquer quel est le meilleur proxy de la réputation d'entreprise, parmi les cinq sources présentes dans l'étude puisqu'aucune d'entre elles ne démontrent aucune significativité. Ainsi nous ne pouvons pas statuer sur les hypothèses 2, 3, et 4 énoncées au chapitre 3.

Cependant, ne pas pouvoir se prononcer sur les hypothèses énoncées est également un résultat en soit. Une compagnie qui observe sa mesure de réputation à la baisse dans un de ces classements d'une année à l'autre, n'a pas vraiment de crainte à avoir sur l'impact que cela pourrait avoir sur son cours de bourse, puisqu'aucun lien n'a été détecté entre les classements et le cours de bourse. Similairement, une augmentation de la mesure de réputation d'une entreprise dans un des classements énoncés ci-dessus n'a pas d'impact sur le cours de bourse.

Dans la deuxième partie de la recherche, nous tenterons de déterminer ce que peut apporter Twitter dans la conception d'un indice de réputation. Twitter permettait-il d'obtenir une information supplémentaire, qui n'est pas encore captée par les mesures de réputation et évaluations de l'image de marque proposées actuellement?

CHAPITRE 4 DÉTERMINATION DES DONNÉES ISSUES DE TWITTER UTILES À LA CONCEPTION D'UNE MESURE DE LA RÉPUTATION

Pour la deuxième partie d'analyse, nous aborderons la deuxième sous-question de recherche : *Les données non structurées et structurées offertes par Twitter peuvent-elles conduire à la construction d'une mesure de la réputation?* Il est donc question de proposer une nouvelle mesure de la réputation à partir des informations récoltées sur Twitter. Avant de considérer la création de cette nouvelle mesure de la réputation (présenté au chapitre 5), nous allons d'abord déterminer quelles sont les informations pertinentes mises à disposition par Twitter.

La démarche de réflexion soutenant la méthodologie dans la seconde partie de ce mémoire est illustrée par le schéma suivant :

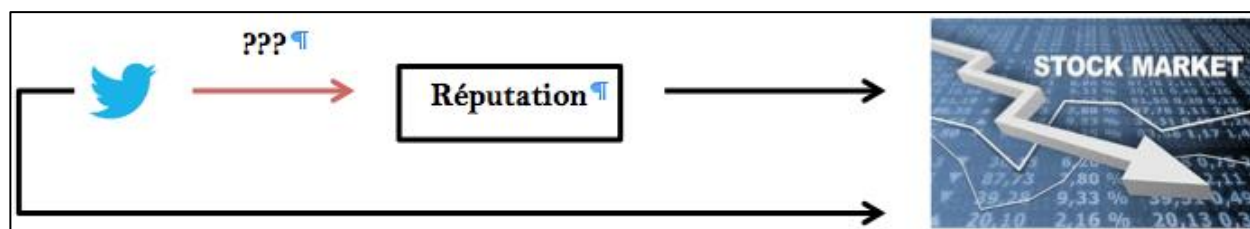


Figure 4.1 Démarche de réflexion dans la méthodologie de la seconde partie du mémoire

Si la revue de littérature indique que la réputation d'entreprise a un impact sur le cours de bourse, et que nous démontrons qu'en observons Twitter certaines informations nous permettent de suivre les tendances boursières, alors nous pourrions peut-être nous en servir afin d'en créer un indice de réputation d'entreprise.

4.1 Retour sur la revue de littérature

Tel que mentionné dans la revue de littérature, la réputation est un construit composé de trois concepts : l'identité d'entreprise, l'identité désirée de l'entreprise et son image (Chun 2005). L'identité d'entreprise, i.e. la perception des employés vis-à-vis de la firme, est difficilement mesurable au travers de Twitter. En effet, premièrement il faudrait qu'une proportion non négligeable des employés possède un compte Twitter et deuxièmement il faudrait être capable de

connaître le nom de ses employés afin de pouvoir extraire les tweets qu'ils produisent sur leur compagnie (données difficilement disponibles). La non disponibilité de ces données implique que le modèle présenté ne tiendra pas compte de l'identité d'entreprise dans la mesure de réputation. Il nous reste donc l'identité désirée de l'entreprise et l'image de l'entreprise.

En ce qui concerne, l'identité désirée de l'entreprise, où plus simplement « comment la firme désire être perçue ? », une fois de plus il est difficile d'extraire cette information à partir de Twitter. On peut modéliser une firme comme étant le comité de leadership qui la compose (CEO, CFO, COO, CMO, etc.). On pourrait donc suivre les comptes Twitter (si existant) de chacun des exécutifs de la compagnie, afin d'évaluer quotidiennement leur perception sur la compagnie. Mais c'est encore une piste fastidieuse qui ne sera pas poursuivie dans le cadre de la présente recherche.

Nous tiendrons donc en compte uniquement une évaluation de l'image d'entreprise pour approximer la réputation d'entreprise. L'image d'entreprise, correspond à la perception des consommateurs sur l'entreprise. Cette perception des consommateurs peut être évalué en partie au travers des opinions exprimées sur les réseaux sociaux tel que Twitter par exemple. Twitter représente une plateforme favorable à l'analyse de données non structurées : le nombre d'utilisateurs est assez vaste (230 millions d'utilisateurs dans le monde, soit 3% de la population mondiale) pour théoriquement avoir une bonne idée des tendances ou opinions qui circulent sur le web.

4.2 Processus de récolte de données sur Twitter

Les données non structurées proviendront des tweets récoltés au travers de l'*Application Programming Interface* (API) de Twitter. Les données sont récoltées sur la période du 26 février 2014 au 10 avril 2014. Il est important de mentionner que ce ne sont pas tous les tweets faisant mention des firmes composant l'échantillon qui sont récoltés sur ladite période, mais seulement des échantillons variant de 100 à 500 tweets dépendamment des firmes et des limites imposées par l'API de Twitter. Il existe en fait deux façons d'accéder aux tweets stockés dans la base de

donnée du réseau social : un accès au travers du REST API (figure 4.2) et un accès au travers du STREAM API¹⁵ (figure 4.3).

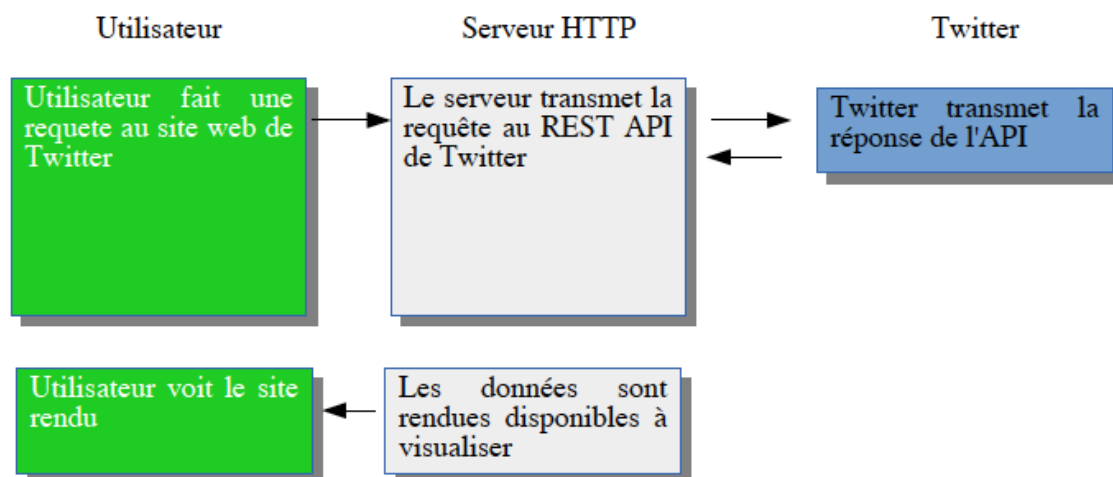


Figure 4.2 Illustration de la structure de requête au travers du REST API (adapté de Twitter)

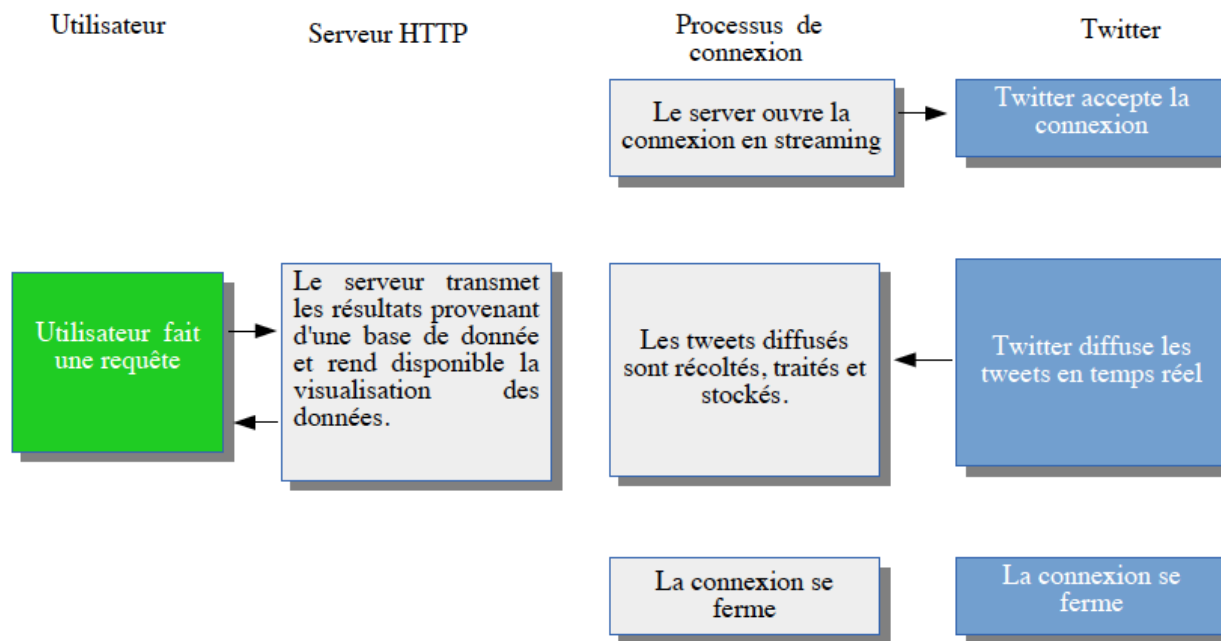


Figure 4.3 Illustration de la structure de requête au travers du Streaming API (adapté de Twitter)

¹⁵ <https://dev.twitter.com/docs/streaming-apis>

La première méthode consiste donc à passer au travers du REST API de Twitter. Ce sont au maximum 500 tweets par requête pour une compagnie en particulier, qui peuvent être extraits de manière rétrospective sur une plage de 7 jours en arrière au maximum. Cette contrainte dans la collecte maximale de tweets par requête au travers du REST API est imposée par Twitter, ce qui constitue un aspect limitant dans la méthodologie. On aimerait avoir un échantillon de tweets plus important afin d'obtenir une représentation plus précise de l'activité sur Twitter d'une compagnie donnée.

La deuxième méthode consiste à utiliser le Stream API de Twitter, qui permet d'accéder en temps réel aux tweets créés sur une compagnie en particulier. La récolte de tweets avec cette seconde méthode est prospective, c'est-à-dire que l'on ne se base pas sur des tweets antérieurs comme avec le REST API, mais des tweets postés en temps réel. La seule contrainte ici est le temps de réponse de Twitter pour récolter les tweets. Sachant que toutes les firmes n'ont pas la même notoriété sur Twitter, le temps de récolte de tweets peut être très long pour avoir un nombre suffisant de tweets par compagnie afin d'obtenir un sentiment général des tweets qui circule sur le web. Cependant, une alternative existe pour utiliser la deuxième méthode décrite précédemment : on pourrait disposer de 101 ordinateurs qui exécuteraient le code en même temps pour chaque compagnie. Le principal problème serait donc ici d'ordre opérationnel : posséder 101 ordinateurs a un coût financier qui n'est pas dans les moyens de notre recherche. De plus, un autre problème de nature technique se pose : il faudrait également posséder 101 comptes Twitter pour chacune des 101 requêtes de collecte de tweets car un compte Twitter individuel est requis pour chaque requête faite sur l'API de Twitter.

La première méthode a été choisie par soucis de temps d'exécution dans la récolte de données. En plus du temps de réponse de l'API de Twitter, la rapidité de la collecte des tweets dépend également de la rapidité de la connexion internet, de la fréquence des tweets par minute d'une entreprise, et de la rapidité d'exécution du script par l'ordinateur. Ainsi, le temps estimé pour récolter un échantillon de tweets pour chacune des 101 firmes jugé suffisamment grand (500 tweets par firmes) peut être très long soit plusieurs heures. Pour limiter l'introduction du biais dans la fenêtre de temps de collecte des tweets, la première méthode (REST API) est donc la plus adéquate dans le cadre de la présente recherche.

4.3 Description des variables de la mesure de réputation

Rappelons que selon la revue de littérature, l'image d'entreprise est quantifiée par l'équité de la marque (Aaker 2009). L'équité de la marque est segmentée en plusieurs variables la constituant dont: la loyauté des consommateurs, la notoriété de la marque, la qualité perçue de la marque, les associations faites avec la marque, et enfin les autres droits de propriété (ce dernier élément est exclu du modèle). Dans cette section chacune de ces variables seront décrites et évaluées par une variable proxy basée uniquement sur les informations récoltées sur Twitter.

4.3.1 Notoriété de la marque

La notoriété d'une marque, fait référence au degré de conscience de son existence parmi la population constituant son marché de consommateurs et potentiels consommateurs. Plus une marque possède une notoriété importante, plus le nombre de personnes connaissant la marque est important. La notoriété d'une marque impacte directement les revenus d'une firme, car elle place la firme dans l'univers des possibles pour un client en prospection. Dans le cas d'une firme possédant une notoriété de marque faible, la majorité des clients n'aura jamais eu connaissance de l'existence de cette firme et donc n'achètera pas de ses produits.

La notoriété de la marque sera ici approximée par les indicateurs suivant: le nombre de suiveurs (« followers ») sur Twitter et le volume de tweets par jour. L'hypothèse posée ici est que plus une firme voit sa notoriété augmentée plus elle possèdera un nombre important de suiveurs sur Twitter, et donc verra le volume de tweets par jour accroître (en référence à l'hypothèse 6 énoncée au chapitre 3).

4.3.2 Qualité perçue de la marque

La qualité perçue d'une marque, fait référence au degré de satisfaction des expériences clients. Plus un client est satisfait avec la qualité du produit ou service obtenu, plus il aura tendance à être incliné à recommander la marque à d'autres et également à entretenir une certaine loyauté en faveur dudit produit ou service (Aaker 2009).

La qualité perçue de la marque auprès des consommateurs sera approximée par les résultats de l'analyse de sentiment conduit sur les tweets. Pour un échantillon de tweets données, l'analyse de sentiment fait ressortir, la polarité de l'échantillon ainsi que les émotions auxquelles se rattachent les tweets.

La polarité de l'échantillon fait référence à la classification de tous les tweets de l'échantillon récolté parmi les catégories d'opinions suivantes : positif, négatif ou neutre. Pour chaque échantillon donné on obtient ainsi le pourcentage de tweets pour chacune des catégories et ainsi en extraire l'orientation générale des opinions exprimées par les consommateurs sur Twitter. Pour parvenir à réaliser cette tâche, le code se base sur les fonctions R du package « sentiment » disponible sur le site suivant : <http://cran.r-project.org/web/packages/sentiment/index.html>.

Les émotions auxquelles se rattachent les tweets sont définies par le package « sentiment » de R. Les tweets sont classifiés selon les émotions suivantes : « colère », « dégoût », « peur », « tristesse », « joie », « surprise » et « inconnue » dans le cas où il est difficile pour l'algorithme de classer le tweet parmi les émotions citées précédemment.

4.3.3 Associations

Les associations faites avec la marque représentent tout ce que la firme possède comme valeur et projette comme émotion auprès des consommateurs. Plus une marque noue des associations avec des entreprises, organismes, ou causes (sociales, humanitaires) qui sont perçues positivement par le public, plus la firme va être en mesure de susciter des émotions positives chez les consommateurs, leur donnant par moment des raisons d'acheter ses produits (Aaker 2009).

Les associations faites avec la marque peuvent être évaluées au travers du nombre de compte suivis sur Twitter par la firme. On suppose ici que si une firme décide de suivre un autre compte Twitter c'est qu'elle partage le désir de suivre les activités de cette autre page et donc d'associer sa marque d'une manière ou d'une autre au compte suivi.

Par exemple, le compte twitter du Groupe Dangote, conglomérat nigérien impliqué dans le secteur agroalimentaire et des matériaux principalement béton, suit le compte twitter de Coca-Cola, de Pepsico qui conduisent certaines activités similaires au Groupe Dangote. On en déduit que le Groupe Dangote souhaite associer sa marque à des marques telles que Coca-Cola ou Pepsi.

Autre exemple encore plus frappant, le groupe Coca-Cola suit sur Twitter des institutions qu'elles sponsorisent telles que le *Professional Golfers' Association* (PGA) Tour, ou des firmes avec qui elle fait affaire comme par exemple Target, ou encore des revues d'affaires auquel Coca-Cola s'identifie telle que la *Harvard Business Review*.

4.3.4 Loyauté des consommateurs

La loyauté des consommateurs fait référence au dévouement entretenu par les consommateurs envers une marque donnée. La loyauté est un facteur déterminant dans l'équité d'une marque, puisqu'elle permet à une firme de poser des barrières à l'entrée sur sa clientèle de base, elle permet également d'obtenir un premium sur le prix de vente de ses produits au sein de sa clientèle de base (Aaker 2009).

La loyauté des consommateurs ne peut pas être estimée à partir de Twitter, sauf si on pouvait analyser la fréquence à laquelle chacun des utilisateurs parlent d'une compagnie, et encore c'est une tâche compliquée qui ne donnerait probablement pas des résultats significatifs.

4.3.5 En résumé

Les paramètres de l'image de marque ainsi que leur proxy sur Twitter sont regroupés dans le tableau ci-après.

Tableau 4.1 Proxys Twitter des indicateurs de l'équité de marque

Indicateurs de l'équité de marque	Proxy sur Twitter
Loyauté des consommateurs	Aucun
Notoriété de la marque	Volume de tweets par jour Nombre de suiveurs sur Twitter
Qualité	Proportion de tweets positifs Proportion de tweets négatifs Proportion de tweets référant à la colère Proportion de tweets référant au dégoût Proportion de tweets référant à la peur Proportion de tweets référant à la tristesse Proportion de tweets référant à la joie Proportion de tweets référant à la surprise
Associations	Nombre de personnes suivies sur Twitter

4.4 Volume quotidien de tweets

Le volume quotidien de tweets est extrait à partir du site web www.peoplebrowsr.com. Le site web possède une entente avec Twitter pour un accès privilégié à la base de donnée du réseau social. Ainsi, Peoplebrowsr offre la possibilité d'extraire dans un fichier .csv, le nombre quotidien de tweets dans lequel est mentionné un terme précisé dans la boîte de recherche, et cela sur une période rétrospective de deux ans. Pour obtenir le volume de tweet quotidien pour une entreprise, il suffit d'indiquer le nom de la compagnie dans la boîte de recherche. Les noms employés pour chaque firme de notre échantillon sont les mêmes noms utilisés pour la recherche de tweets et correspondent à ceux indiqué au tableau A9.1 en ANNEXE 9 (colonne « Mot clé Twitter »).

4.5 Informations financières sur les marchés

Les informations financières sur les marchés d'échanges des firmes qui composent l'échantillon d'analyse défini précédemment dans le document, sont nécessaires. Les types d'informations financières récoltées sont identiques à la première partie d'analyse décrite précédemment: le cours d'ouverture, le cours de fermeture, le volume d'échange, le cours le plus haut et le cours le plus bas. Les cours de bourse journaliers des firmes composant l'échantillon, ont été extraits pour la période du 26 février 2014 au 10 avril 2014.

4.6 Remarques importantes sur la méthodologie

La principale remarque concernant la méthodologie utilisée pour analyser les données sur Twitter fait référence aux entreprises possédant des filiales qui possèdent un compte Twitter. Une entreprise telle que Time Warner, possède plusieurs entités tels que le magazine Time, la chaîne de télévision Home Box Office (HBO), le groupe de production télévisée Turner et le groupe de production cinématographique Warner Bros. Chacune de ces entités possède un compte Twitter en plus de la maison mère. Cet aspect rend l'analyse de l'activité du groupe Time Warner plus complexe qu'une entreprise ne possédant aucune filiale et interagissant avec le public au travers d'un seul compte Twitter. Deux méthodes sont suggérées pour analyser un conglomérat ou compagnie possédant plusieurs comptes Twitter associés à chacune de ses filiales. La première méthode consiste à analyser individuellement chaque compte twitter des filiales de la compagnie mère et de leur attribuer une pondération selon l'importance de la division dans la génération du revenu total de la maison mère. Ainsi par exemple, la filiale Time (magazine) représente 11% des revenus totaux du groupe Time Warner (données 2012), la filiale Warner Bros représente 41% des revenus, la filiale HBO représente 16%, et la filiale Turner représente 30%. Si l'on souhaite obtenir le sentiment général du public envers l'entreprise Time Warner on pourrait mesurer l'activité de chacune des filiales et pondéré le score selon la part des revenus de chaque filiale. La formule serait donc la suivante :

$$tp(Time Warner) = tp(Time Mag) * 12\% + tp(Warner Bros) * 41\% + tp(HBO) * 16\% + tp(Turner) * 30\%, \text{ avec } tp \text{ la proportion de tweets positifs}$$

La deuxième méthode consiste à tout simplement analyser les tweets portant sur la compagnie mère. Ainsi dans l'exemple précédent on s'intéressera uniquement aux tweets mentionnant Time Warner, c'est cette dernière méthode qui a été choisie car plus simple à réaliser.

4.7 Composition de l'échantillon d'entreprises

La composition de l'échantillon d'entreprises est la même que pour la première partie de l'analyse. Rappelons que le but est de comparer une mesure de réputation conçue à partir de Twitter avec les mesures de réputations actuelles. La comparaison doit donc être réalisée avec une base commune de compagnie pour limiter les biais de comparaison.

Le tableau avec les informations nécessaires à la récolte des tweets est présenté en ANNEXE 9 au tableau A9.1. La dernière colonne intitulée « mot clé twitter » contient les mots clés transmis à l'API de Twitter afin de récolter les tweets associés à la compagnie à laquelle le mot clé est lié. La colonne intitulée « compte twitter » contient les noms des comptes Twitter de chaque compagnie.

4.8 Illustration de la base de données

Les données récoltées sur Twitter ont également été regroupées dans une base de donnée dont un extrait est représenté ci-après.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
ID	Company	positif	negatif	neutre	anger	disgust	fear	sadness	joy	surprise	unknown	id_twitter	followers	followees	close	Volume	Date
1	Abbott	57	21	12	5	2	1	7	11	1	63	360054369	18902	3056	39,66	8752600	2014-02-26
1	Abbott	184	63	33	12	2	0	10	69	4	183	360054369	18950	3178	39,83	9189600	2014-02-27
1	Abbott	43	43	9	2	1	2	4	11	5	70	360054369	19124	2933	39,64	11006300	2014-03-03
1	Abbott	199	129	65	17	7	8	12	42	3	304	360054369	19171	3051	40,32	9460200	2014-03-04
1	Abbott	225	139	30	14	3	3	9	44	11	310	360054369	19235	3173	40,49	8285200	2014-03-05
1	Abbott	214	124	54	8	8	1	13	28	6	328	360054369	19279	3288	40,23	6442000	2014-03-06
1	Abbott	238	76	55	8	0	2	20	34	4	301	360054369	19314	3409	39,95	8630700	2014-03-07
1	Abbott	133	102	46	9	4	2	7	22	8	229	360054369	19390	3003	39,89	5576300	2014-03-10
1	Abbott	199	134	46	18	3	3	17	59	9	270	360054369	19551	3313	39,82	6964500	2014-03-13
1	Abbott	211	106	75	14	3	4	6	54	7	304	360054369	19758	3140	39,58	5768900	2014-03-17
1	Abbott	199	134	61	9	1	12	3	54	6	309	360054369	19864	3310	40,22	8679100	2014-03-19
1	Abbott	132	128	26	2	0	20	12	59	1	192	360054369	19961	3363	39,64	11396600	2014-03-20
1	Abbott	96	63	28	3	1	0	10	29	1	143	360054369	20212	3164	38,66	11078000	2014-03-24
1	Abbott	205	132	49	4	2	3	17	43	10	307	360054369	20314	3224	38,7	8209600	2014-03-25
1	Abbott	55	29	13	8	0	0	2	14	1	72	360054369	20405	3079	39,14	8174800	2014-03-26
1	Abbott	175	139	54	4	15	31	6	63	5	244	360054369	20543	3132	38,69	7404100	2014-03-27
1	Abbott	298	61	22	6	0	0	6	24	4	341	360054369	20691	3044	38,58	8038000	2014-03-31
1	Abbott	211	100	66	20	1	2	8	38	3	305	360054369	20746	3104	38,83	5631200	2014-04-01
1	Abbott	48	32	16	2	0	1	2	10	0	81	360054369	20778	3157	38,64	6407300	2014-04-02
1	Abbott	41	41	16	0	0	3	3	21	2	69	360054369	20815	3207	38,75	5874700	2014-04-03
2	Adidas	201	49	44	9	1	0	15	19	3	247	300114634	756451	93	84,94	688000	2014-02-26
2	Adidas	164	91	20	9	30	3	6	21	4	202	300114634	759496	93	84,86	1006100	2014-02-27
2	Adidas	225	103	45	13	3	5	16	37	8	291	300114634	768226	93	82,2	2108300	2014-03-03
2	Adidas	280	53	43	17	4	2	13	23	41	276	300114634	770308	93	83,45	1533600	2014-03-04
2	Adidas	58	18	12	0	1	0	5	2	13	67	300114634	772202	93	83,49	3363900	2014-03-05
2	Adidas	70	11	13	5	1	0	2	6	10	70	300114634	774971	93	80,92	2561800	2014-03-06
2	Adidas	64	14	5	3	0	1	5	9	13	52	300114634	777130	94	81,46	1622500	2014-03-07
2	Adidas	56	18	17	2	0	2	15	11	2	59	300114634	783326	94	80,42	1040400	2014-03-10
2	Adidas	192	48	26	17	3	2	8	20	1	215	300114634	792888	94	78,06	1145300	2014-03-13
2	Adidas	51	21	13	0	0	0	7	5	5	68	300114634	810091	95	79,68	1227100	2014-03-17
2	Adidas	178	9	4	0	0	0	5	13	2	171	300114634	818873	96	79,63	1180400	2014-03-19
2	Adidas	71	15	6	2	0	1	3	6	1	79	300114634	822754	96	77,85	930900	2014-03-20
2	Adidas	63	9	117	1	0	0	2	10	16	160	300114634	837995	96	76,44	1224400	2014-03-24
2	Adidas	83	27	74	1	0	1	4	15	16	147	300114634	843398	96	76,41	931000	2014-03-25
2	Adidas	218	53	74	4	0	5	12	36	9	279	300114634	847692	96	78,5	1556300	2014-03-26
2	Adidas	322	48	25	3	1	0	12	20	70	289	300114634	851189	97	78,78	1032500	2014-03-27

Figure 4.4 Aperçu de la base de donnée conçue avec les informations récoltées sur Twitter

4.9 Traitement des données : régression linéaire

Avant de procéder à l'analyse de la nouvelle mesure de réputation présentée au chapitre 6, nous allons d'abord estimer la significativité de chacune des variables indépendantes au travers d'une régression linéaire, dans l'explication de la variation du cours de bourse qui constitue la variable dépendante de la régression linéaire. On veut ainsi ressortir les relations linéaires (si existantes) entre le cours de bourse et les indicateurs de twitter énoncées dans les sections précédentes.

Les données sont regroupées sous forme de panel, avec 101 compagnies et 24 observations par compagnie enregistrées entre le 26 février 2014 et le 10 avril 2014. Le modèle mathématique de régression linéaire est le suivant :

$$y_{it} = x_{it} \beta + \varepsilon_{it}, \text{ où}$$

- y_{it} est le cours de bourse de la i ème compagnie au temps t ;
- x_{it} est la matrice des variables indépendantes pour la i ème compagnie du panel au temps t ;
- β est la matrice des coefficients associés à chacune des variables indépendantes;
- ε_{it} est le terme d'erreur du modèle pour la i ème compagnie au temps t

La variable dépendante et les variables indépendantes sont listées en ANNEXE 1 au tableau A1.1.

4.10 Analyse des résultats

4.10.1 Tests préliminaires

L'analyse des résultats s'intéresse à l'utilité de chacune des variables indépendante listées au tableau A1.1 en ANNEXE 1, dans l'explication de la variation du prix de l'action d'une entreprise (variable dépendante). Pour y parvenir, nous allons procéder à une régression linéaire de la variable dépendante avec les variables de contrôle mentionnées précédemment. Cependant avant de réaliser une régression linéaire plusieurs contraintes sont à respecter : (1) les variables indépendantes doivent observées une distribution normale (test de normalité), et (2) elles doivent être indépendantes entre elles (matrice de corrélation et test de multicolinéarité).

4.10.1.1 Test de Normalité

Afin de respecter le test de normalité, la distribution d'une variable indépendante doit avoir un « skewness » égale à 0 ± 1.5 et un « kurtosis » de 3 ± 1.5 .

La description de la distribution de chacune des variables indépendantes est en ANNEXE 1 au tableau A1.2. Ce tableau indique que seules les variables suivantes ont une distribution quasi-normale: « tp » et « tn », correspondant respectivement à la proportion quotidienne de tweets positifs et la proportion quotidienne de tweets négatifs.

Pour obtenir une distribution quasi-normale, une transformation en logarithme naturel (ln) a été effectuée sur toutes les variables indépendantes. La description de la distribution des variables est indiquée dans le tableau A1.3 en ANNEXE 1. Premièrement on constate que la transformation en logarithme naturelle de la variable « tp » éloigne davantage d'une distribution normale, nous conserverons donc la variable « tp » sans transformation pour notre régression. Concernant la variable « tn », on observe une très faible amélioration du « skewness » après transformation, et une très faible augmentation du « kurtosis », nous allons donc conserver également la variable tn sans transformation. Concernant les autres variables de contrôle, la transformation en logarithme naturel a rapproché la distribution des variables vers celle d'une distribution normale, cependant

pour la grande majorité la distribution des variables n'est pas encore normale. Un compromis a été choisis : nous considérerons uniquement les variables qui ont un skewness (en valeur absolu) inférieur à 3 et un kurtosis (en valeur absolu) inférieur à 16. Les seules variables qui respectent ce nouveau critère sont : « tp », « tn », « tneutre », « tjoy », « Volumetweet », « followers », « followees », « Ctn » et « Ctneutre ». Toutes les autres variables indépendantes ont une distribution qui est très loin d'une distribution normale, elles seront donc ignorées de nos régressions linéaires ultérieures.

L'introduction d'un biais dans la distribution des variables indépendantes va introduire un biais dans la régression linéaire. C'est un choix à prendre sous peine ne pas avoir suffisamment de variables indépendantes à tester.

4.10.1.2 Test de multi-colinéarité

La seconde contrainte à respecter est l'indépendance des variables indépendantes. Le test de multi-colinéarité permet d'analyser l'indépendance des variables indépendantes. Le test de multi-colinéarité détermine si une variable indépendante x_1 est une expression linéaire des autres variables indépendantes de telle sorte que l'on puisse exprimer :

$$x_1 = x_2 \beta_2 + x_3 \beta_3 + \dots + x_n \beta_n + \varepsilon$$

Le facteur calculé au travers du test de multi-colinéarité est le « *variance inflator factor* » (VIF). Un VIF supérieur à 10 indique qu'une variable est une expression linéaire d'autres variables indépendante.

Les résultats du test sont présentés en ANNEXE 3 au tableau 3.1. On constate que « tp » et « tn » ont un VIF supérieur à 10. En fait c'est un résultat auquel on peut s'attendre puisque la relation suivante existe :

$$tp + tn + tneutre = 100\%$$

On va donc procéder à une suppression de variable : la variable « tneutre » qui représente la proportion de tweet à polarité neutre, sera supprimée de notre liste de variables indépendantes que l'on va pouvoir utiliser dans les régressions subséquentes. Les tweets à polarité neutre sont moins intéressants à évaluer a priori puisqu'ils ne comportent pas une information signifiante

quant à l'expression sentimentale de l'auteur du tweet. Il s'agit en fait d'une absence de prise de position vis-à-vis d'une compagnie, qui est moins intéressante à analyser qu'un tweet positif ou négatif d'une entreprise. De plus « Ctneutre » qui représente la croissance quotidienne des tweets neutres par rapport à la veille est conservée : on obtient donc de l'information sur l'impact des tweets neutres même si l'on omet la variable tneutre de l'analyse.

Ainsi la liste des variables indépendantes à considérer est réduite à : « tp », « tn », « tjoy », « Volumetweet », « followers », « followees », « Ctn » et « Ctneutre ». Le tableau A3.2 de l'ANNEXE 3 regroupe les nouveaux résultats du test de multicollinéarité après omission de la variable tneutrellog (transformation en logarithme naturel de la variable tneutre).

4.10.1.3 Matrices de corrélation

Une matrice de corrélation globale est présentée en ANNEXE 2. La matrice de corrélation permet de distinguer les variables corrélées deux à deux. Les variables corrélées deux à deux ne peuvent être mises ensemble dans une régression linéaire commune.

De plus, une matrice de corrélation pour chaque industrie a été réalisée. Les 26 matrices de corrélation sont listées en ANNEXE 2.

4.10.2 Régression sur l'ensemble de la base de données

La matrice de corrélation globale présente à l'ANNEXE 2 au tableau A2.1, indique que les paires corrélées sont « tp » / « tn » et « Volumetweet » / « followerslog ». Ainsi ces paires ne peuvent se retrouver dans un même modèle linéaire pour respecter la contrainte d'indépendance des variables. Pour observer l'influence de toutes les variables indépendantes, nous allons procéder à la régression de deux modèles :

$$y = tp * \beta_1 + tjoylog * \beta_2 + Volumetweetlog * \beta_3 + followeeslog * \beta_4 + ctnlog \beta_5 + ctneutrellog \beta_6 \quad (1)$$

$$y = tn * \beta_7 + tjoylog * \beta_2 + followerslog * \beta_8 + followeeslog * \beta_4 + ctnlog \beta_5 + ctneutrelog \beta_6 \quad (2)$$

Une première régression linéaire sans « lag » a été effectuée sur l'ensemble des compagnies présentes dans la base de donnée pour chacun des deux modèles décrits ci-dessus. Une seconde régression a été réalisée avec un « lag » d'une journée et une troisième régression a été réalisée avec un « lag » de deux journées. Les résultats sont présentés en ANNEXE 5 au tableau A5.1 pour le modèle 1 et au tableau A5.2 pour le modèle 2.

Les résultats indiquent qu'aucune des variables indépendantes ne sont significatives en dessous d'un seuil de $p < 0.20$. Le R^2 est environ de 5% pour les deux modèles. Plusieurs éléments sont à considérer pour expliquer de tels résultats à la grandeur de la base de donnée. Premièrement, la base de donnée regroupe des compagnies d'industries différentes qui ne réagissent pas de la même manière vis-à-vis des réseaux sociaux. Les compagnies de chaque industrie possèdent une approche totalement différente avec les consommateurs vis-à-vis des réseaux sociaux et de l'impact qu'elles peuvent avoir sur leur activité. Prenons l'exemple de l'industrie des vêtements (« Apparel ») qui regroupe entre autre Adidas (id 2) et Nike (id 62). La présence sur les réseaux sociaux de ces deux entreprises est plutôt importante comme on peut en juger le nombre de tweet moyen journalier sur la période pour chacune des entreprises (VOIR ANNEXE 11). Tandis que la présence sur des réseaux sociaux d'entreprise de l'industrie du secteur pharmaceutique avec les entreprises telles que Abbott Laboratories (id 1) et GlaxoSmithKline (id 40) est moins importante. L'intérêt auprès des consommateurs dépend donc de l'industrie, ainsi on peut se douter que l'impact des réseaux sociaux sur une compagnie dépend également de l'industrie à laquelle elle est associée. Deuxièmement, la base de donnée est composée d'entreprises avec un prix d'action d'ordre très différent selon les entreprises : on parle d'action de dizaine de dollars à des actions de plus 500 dollars et des actions cotés en won (monnaie coréenne) à plus de 1,000,000 de won (Samsung). Là encore, il peut y avoir un important effet de taille qui vient nuire aux résultats de la régression linéaire. Pour toutes ces raisons, nous allons plutôt procéder à une analyse industrie par industrie et selon le niveau de prix de l'action.

4.10.3 Régression par industrie

Une régression linéaire par industrie a été effectuée pour chacun des deux modèles décrits précédemment. Les résultats sont présentés en ANNEXE 6. Les tableaux A6.1 et A6.3 représentent les résultats pour les modèles 1 et 2 respectivement avant ajustement. L'ajustement dont il est ici question, concerne les nouvelles paires de variables explicatives corrélées pour chaque industrie (les matrices de corrélation pour chaque industrie sont définies en ANNEXE 2). L'exercice de l'indépendance doit encore être effectuée pour chacune des régressions linéaires associées à chaque industrie. D'ailleurs on retrouve des aberrations dans certains résultats de ces deux tableaux avec certaines régressions possédant un R^2 égale à 1.

Les tableaux A6.2 et A6.4 en ANNEXE 6 regroupent les résultats des régressions effectuées après ajustement en s'assurant de l'indépendance des variables explicatives. On remarquera que la variable `followeeslog` (la transformée en logarithme naturel de la variable `followees` correspondant au nombre de compte twitter suivis par une compagnie) a été retirée de la plupart des modèles après ajustement. Cette variable était souvent fortement à plusieurs autres variables explicatives et également au cours de bourse. C'est la seule variable qui est censée être un proxy des associations faites avec l'entreprise : cependant on peut douter de la pertinence de cette variable. En effet, ce n'est pas parce qu'une entreprise suit plusieurs compte twitter appartenant à d'autres entreprises que l'on peut tirer de quelconque conclusion.

Le tableau A6.5 résume les résultats concernant les signes des coefficients des variables explicatives significatives à un seuil de $p < 0.20$. Les variables non significatives en dessous du seuil indiqué ($p < 0.20$) ne sont pas analysées.

Les résultats indiquent que dans 9 cas sur 26 la variable « `tp` » est significative. Les industries concernées par la significativité de la variable « `tp` » sont : « Apparel », « Beverages », « Computer software », « Consumer Food Products », « Electric and Gas utilities », « Entertainment », « Mining, Crude-Oil Production », « Petroleum Refining », « Telecommunications ». Sur ces industries, dans 56% des cas la variable « `tp` » a un coefficient négatif, et un coefficient positif dans 44% des cas (résultat proche de la parité 50-50). Intuitivement, on s'attendrait à ce qu'une augmentation du ratio de tweet positif quotidien conduise à une augmentation du cours de bourse (en référence à l'hypothèse 5 énoncée au

chapitre 3). Cependant on ne peut pas l'affirmer avec les résultats obtenus. On ne peut donc rien conclure sur l'impact de la variable « tp ».

Les résultats indiquent que dans 9 cas sur 26 la variable « tn » est significative. Les industries concernées par la significativité de la variable « tn » sont : « Computer Software », « Energy », « Health Care : Insurance and Managed Care », « Insurance : Life and Health », « Metals », « Mining, Crude-Oil Production », « Petroleum Refining », « Pharmaceuticals », « Telecommunications ». Sur ces industries, dans 22% des cas la variable « tn » a un coefficient négatif, et un coefficient positif dans 78% des cas. Intuitivement, on s'attendrait à ce qu'une augmentation du ratio de tweet négatif quotidien conduise à une diminution du cours de bourse (en référence à l'hypothèse 6 énoncée au chapitre 3). Cependant on observe tout le contraire expérimentalement avec un coefficient positif dans la majorité des industries où la variable « tn » est significative : un résultat non logique, qui indique que le contexte dans lequel se trouvaient les compagnies de l'échantillon sur la période où les données ont été récoltées est probablement à prendre en considération.

On observe également que dans 7 cas sur 26 la variable tjoylog est significative. Les industries concernées par la significativité de la variable tjoylog sont : « Apparel », « Energy », « Entertainment », « Megabanks », « Mining, Crude-Oil Production », « Superregional Banks », « Telecommunications ». Sur ces industries, dans 71% des cas la variable tjoylog a un coefficient négatif, et un coefficient positif dans 29% des cas. Intuitivement, on s'attendrait à ce qu'une augmentation du ratio de tweet relié à une émotion de joie conduise à une augmentation du cours de bourse (coefficient positif). Ce n'est pas ce que nous observons expérimentalement, au contraire le coefficient est négatif dans la majorité des industries où la variable tjoylog est significative. Encore un autre signe qui indique que le contexte est certainement un aspect qui aurait dû être pris en compte pour pouvoir aller plus loin dans l'interprétation.

La variable « Ctnlog » est significative dans 6 cas sur 26. Les industries concernées par la significativité de la variable « Ctnlog » sont : « Consumer Food Products », « Entertainment », « Petroleum Refining », « Pharmaceuticals », « Soaps and Cosmetics », « Telecommunications ». Sur ces industries, dans 67% des cas la variable « Ctnlog » a un coefficient négatif, et un coefficient positif dans 33% des cas. Intuitivement, on s'attendrait à ce qu'une augmentation de la croissance du ratio des tweets négatifs conduise à une diminution du cours de bourse. C'est ce

que nous observons expérimentalement pour la majorité des industries où la variable « Ctnlog » est significative.

La variable « Ctneutrellog » est significative dans 2 cas sur 26. Les industries concernées par la significativité de la variable « Ctneutrellog » sont : « Chemicals », « Insurance : Life and Health ». Dans les deux cas elle a un impact négatif sur le cours de bourse.

La variable « Volumetweetlog » est significative dans 12 cas sur 26. Les industries concernées par la significativité de la variable « Volumetweetlog » sont : « Apparel », « Beverages », « Computers », « Consumer Food Products », « Food and Drug Stores », « Food Services », « Insurance : Property and Casualty », « Megabanks », « Metals », « Motor Vehicles », « Petroleum Refining », « Soaps and Cosmetics ». Sur ces industries, dans 58% des cas la variable Volumetweetlog a un coefficient négatif, et un coefficient positif dans 42% des cas. C'est un résultat intéressant, il semble donc que dans la majorité des cas où la variable « Volumetweetlog » est significative, une augmentation du volume de tweets conduit à une diminution du cours de bourse. C'est une observation qui pourrait être expliquée par une observation mentionnée dans la revue de littérature, au sujet du fait que les consommateurs non-satisfaits s'engagent davantage dans du BAO que les consommateurs satisfaits (Anderson 1998; Williams, Buttle, and Biggemann 2012).

La variable « followerslog » est significative dans 20 cas sur 26. Les industries concernées par la significativité de la variable followerslog sont : « Apparel », « Beverages », « Computer Software », « Consumer Food Products », « Electric and Gas Utilities », « Energy », « Entertainment », « Food and Drug Stores », « Food Services », « General Merchandisers », « Health Care : Insurance and Managed Care », « Insurance : Property and Casualty », « Megabanks », « Metals », « Mining, Crude-Oil Production », « Motor Vehicles », « Petroleum Refining », « Pharmaceuticals », « Superregional Banks », « Telecommunications ». Sur ces industries, dans 35% des cas la variable followerslog a un coefficient négatif, et un coefficient positif dans 65% des cas. Ces résultats montrent que la variable followerslog est très pertinente dans la création d'un indice de réputation. En effet, intuitivement on s'attendrait à ce qu'une augmentation du nombre de followers entraîne une augmentation du cours de bourse (en référence à l'hypothèse 7 énoncée au chapitre 3). C'est ce que l'on observe dans la majorité des

industries où la variable *followerslog* est significative. De plus la variable est significative dans 77% des industries de l'échantillon.

En conclusion, les tweets positifs (variable *tp*), tweets négatifs (variable *tn*), et les tweets rattachés à une émotion de joie (variable *tjoylog*) ne démontrent pas des résultats concluants : ils sont significatifs respectivement dans 35%, 35%, 27% seulement des industries présentes dans l'échantillon. De plus, leur coefficient des régressions linéaires ne se comporte pas comme l'on pourrait s'y attendre intuitivement. Même si la croissance de tweets négatifs voit le signe de son coefficient de régression s'aligner avec les attentes, la variable est significative dans seulement 23% des industries présentes dans l'échantillon. La croissance du ratio de tweets neutres (*Ctneutrelog*) est significative dans seulement 11,5% des industries présentes dans l'échantillon : elle ne donne pas vraiment d'information sur le cours de bourse, ainsi les tweets neutres peuvent être éliminés dans la considération d'un indice de réputation pour entreprise. Les seules variables qui s'avèrent être pertinentes sont le volume de tweets (*Volumetweetlog*) et le nombre de followers (*followerslog*). Nous allons pouvoir vérifier ces premières conclusions en analysant les données selon une autre segmentation de notre base de données dans la section suivante : nous avons décidé de segmenter notre échantillon de données par niveau de prix de l'action.

4.10.4 Régression par niveau de prix de l'action

Une régression linéaire par niveau de prix de l'action a été réalisée pour chacun des deux modèles décrits précédemment. Quatre niveaux de prix du cours de bourse ont été définis (choix arbitraire) : (1) niveau 1 pour un prix d'action inférieure à 50 unités (unité pour indiquer que nous ne nous intéressons pas à la monnaie d'échange : dollars, won, Livres, euros), (2) niveau 2 pour un prix d'action compris entre 50 et 100 unités, (3) niveau 3 pour un prix d'action compris entre 100 et 200 unités, et (4) niveau 4 pour un prix d'action supérieur à 200 unités.

Les résultats des régressions sont présentés en ANNEXE 7. Les tableaux A7.1 et A7.4 sont associés aux régressions linéaires des modèles 1 et 2 respectivement, sans lag. Le tableau A7.7 résume la significativité des variables explicatives avec les régressions linéaires sans lag. Les variables « *tp* » et « *tn* » sont significatives pour les niveaux 2 et 3. La variable « *tjoylog* » est

significative pour le niveau 1. La variable « Volumetweetlog » est significative pour le niveau 2. La variable « followerslog » est significative pour le niveau 1 et 3. La variable « Ctnlog » est significative pour le niveau 2. Et « Ctneutrelog » ne démontre aucune significativité selon les quatre niveaux du prix de l'action.

Si nous nous intéressons maintenant au signe des coefficients des variables explicatives lorsqu'elles sont significatives au seuil $p < 0.20$, pour les régressions linéaires sans lag. La variable « tp » a un coefficient positif pour le niveau 2 et négatif pour le niveau 3 : ce qui indiquerait que lorsque la proportion de tweets positifs augmente on observe une augmentation du cours de bourse pour les entreprises du niveau 2 et une diminution du cours de bourse pour les entreprises du niveau 3. Cette observation laisse toute conclusion difficile. La variable « tn » a un coefficient négatif pour le niveau 2 et un coefficient positif pour le niveau 3 : ce qui indiquerait que lorsque la proportion de tweets négatifs augmente on observe une diminution du cours de bourse pour les entreprises du niveau 2 et une augmentation du cours de bourse pour les entreprises du niveau 3. La conclusion est encore difficile à faire. Le coefficient de la variable « tjoylog » est négatif pour les entreprises du niveau 1 : ce qui indique que plus la proportion de tweets associés au sentiment de « joie » est importante plus le cours de bourse diminue pour les entreprises du niveau 1. C'est un résultat contraire à la logique qui ne nous permet pas de tirer de conclusion. La variable « Volumetweetlog » possède un coefficient négatif pour les entreprises du niveau 2 : ce qui indique que plus le volume de tweets d'une entreprise augmente plus le cours de bourse diminue. La variable « followerslog » possède un coefficient positif pour les entreprises du niveau 1 et 3. Ainsi, lorsque le nombre de « followers » augmente on observe une augmentation du cours de bourse. La variable « Ctnlog » possède un coefficient positif pour les entreprises du niveau 2 : ainsi lorsque la croissance des tweets négatifs augmentent on observe une augmentation du cours du bourse. C'est encore un résultat contraire à la logique, qui ne nous permet pas de tirer de conclusion.

Les tableaux A7.2 et A7.5 sont associés aux régressions linéaires des modèles 1 et 2 respectivement avec une journée de lag. Le tableau A7.8 résume la significativité des variables explicatives avec les régressions linéaires avec lag. La variable « tp » ne démontre aucune significativité selon les niveaux de prix de l'action. La variable « tn » est maintenant uniquement

significative pour le niveau 2. La variable « tjoylog » est significative pour le niveau 1 et 2. La variable « Volumetweetlog » est significative pour le niveau 3. La variable « followerslog » est significative pour le niveau 1 et 3. La variable « Ctnlog » est significative pour le niveau 2. La variable « Ctneutrelog » ne démontre toujours aucune significativité.

Si nous nous intéressons maintenant au signe des coefficients des variables explicatives lorsqu'elles sont significatives au seuil $p < 0.20$, pour les régressions linéaires avec lag. La variable « tn » possède un coefficient négatif pour les entreprises du niveau 2 : ce qui indique lorsque la proportion de tweets négatifs augmente une journée avant, l'action diminue. La variable « tjoy » possède un coefficient négatif, ce qui indique que lorsque la proportion de tweets associés au sentiment de « joie » la journée précédente augmente, le cours de bourse des entreprises du niveau 1 et 2 diminue. C'est une observation contraire à la logique qui ne nous permet pas de conclure. La variable « Volumetweetlog » possède un coefficient négatif pour les entreprises du niveau 3 : ce qui indique que lorsque le nombre de tweets la journée précédente augmente, le cours de bourse diminue. La variable « followerslog » possède un coefficient positif pour les niveaux 1 et 3 : ce qui indique que lorsque le nombre de followers la journée antérieure augmente, l'action augmente également. La variable Ctnlog possède un coefficient positif pour les entreprises du niveau 2 : ce qui indique que lorsque la croissance de la proportion du nombre de tweets négatifs augmente la journée antérieure, l'action augmente. C'est un résultat contraire à la logique qui ne nous permet pas de tirer de conclusion.

Les tableaux A7.3 et A7.6 en ANNEXE 7 présentent les résultats de régressions linéaires des modèle 1 et 2 avec respectivement une interaction du niveau de prix de l'action avec le volume de tweets (modèle 1) et une interaction du niveau de prix de l'action avec le nombre de followers. L'équation théorique du modèle avec interaction est la suivante :

$$y = tp * \beta_1 + tjoylog * \beta_2 + Volumetweetlog * \beta_3 + followeeslog * \beta_4 + ctnlog \beta_5 + ctneutrelog \beta_6 + \beta_7 * niveau2 + \beta_8 * niveau3 + \beta_9 * niveau4 + \beta_{10} * niveau2 * Volumetweetlog + \beta_{11} * niveau3 * Volumetweetlog + \beta_{12} * niveau4 * Volumetweetlog$$

$$\begin{aligned}
y = & \textit{tn} * \beta_7 + \textit{tjoylog} * \beta_2 + \textit{followerslog} * \beta_8 + \textit{followeeslog} * \beta_4 + \textit{ctnlog} \beta_5 \\
& + \textit{ctneutrelog} \beta_6 + \beta_7 * \textit{niveau2} + \beta_8 * \textit{niveau3} + \beta_9 * \textit{niveau4} + \beta_{13} \\
& * \textit{niveau2} * \textit{followerslog} + \beta_{14} * \textit{niveau3} * \textit{followerslog} + \beta_{15} * \textit{niveau4} \\
& * \textit{followerslog}
\end{aligned}$$

Le niveau 1 est omis de la régression par défaut, il s'agit du niveau de référence auquel seront comparés les autres niveaux de prix au travers de l'interaction. Ainsi les coefficients β_{10} , β_{11} , β_{12} , β_{13} , β_{14} , β_{15} , sont exprimés comparativement au niveau 1. Selon les résultats du tableau A7.3, lorsque le volume de tweets pour les entreprises du niveau 1 augmente, le prix de l'action augmente davantage que lorsque le volume de tweets augmente chez les entreprises avec un plus haut prix d'action (les interactions ont toutes un coefficient négatif). Selon les résultats du tableau A7.6, lorsque le nombre de followers pour les entreprises du niveau 1 augmente, le prix de l'action augmente davantage que lorsque le nombre de followers augmente pour les entreprises du niveau 2 et 3 (coefficient négatif pour les interaction avec le niveau 1 et 2), mais l'action augmente moins que lorsque le nombre de followers des entreprises du niveau 4 augmente (coefficient positif pour l'interaction avec le niveau 4).

En conclusion, nous ne pouvons tirer de conclusion au sujet de l'impact de la proportion de tweets positifs (tp), et de la proportion de tweets associés à l'émotion de joie (tjoylog) dans la création d'un indice de réputation d'entreprise. Ainsi l'hypothèse 5 énoncée au chapitre 3 ne peut être confirmée. La proportion de tweets neutres ne s'avère pas pertinente dans la création d'un indice de réputation d'entreprise. La proportion de tweets négatifs peut s'avérer utile si elle est observée avec un « lag » d'une journée ou la journée même seulement pour les entreprises du niveau 2 (prix de l'action entre 50 et 100 unités). Ainsi, l'hypothèse 6 énoncée au chapitre 3 est confirmée. La croissance de la proportion de tweets négatifs n'est pas utile dans ce cas, puisque le coefficient est positif pour les entreprises du niveau 2 (niveau auquel la variable est significative), ce qui est contraire à la logique. Encore une fois, le nombre de followers et le volume de tweets présentent des résultats intéressants. Le signe du coefficient des régressions linéaires est consistant avec celui obtenu dans la section précédente, ce qui confirme encore que

ces deux variables sont consistantes dans leurs résultats et dans la création d'un indice de réputation. L'hypothèse 7 énoncée au chapitre 3 est confirmée.

4.11 Discussions

La première remarque importante à soulever est que Twitter présente plusieurs limites dans l'interprétation de la réputation. Par exemple, advenant le cas d'un buzz à caractère négatif (scandale, accident, etc.) ciblant une compagnie qui est reconnue mondialement et ayant une excellente réputation bâtie au fil du temps. La crise soudaine créée par ce buzz aura certes un impact sur la réputation de l'entreprise, mais un impact que l'on peut raisonnablement qualifier de limitée pour les raisons suivantes : une firme à forte réputation n'est pas susceptible de voir s'effondrer sa réputation à cause d'un seul incident, mais plutôt par une succession d'incidents. La dynamique soudaine des signaux perçus sous Twitter (tel que par exemple l'augmentation du nombre de tweets ou l'élévation du ratio de tweets négatifs) est une indication d'une crise, mais son impact sur la réputation est difficilement analysable simplement avec Twitter. Le suivi de l'activité sur les réseaux sociaux présente aussi des limites advenant le cas d'une fausse rumeur, ou d'une information qui s'avèrera être erronée post-diffusion. Dans ces contextes, on observera ponctuellement sur Twitter des commentaires négatifs, sans pour autant qu'il y ait un impact significatif sur la compagnie par la suite.

La deuxième remarque est que le contexte commercial dans lequel se trouvait chacune des compagnies de l'échantillon n'a pas été évalué : est-ce que les compagnies subissaient une crise ou un boost de popularité durant la période de prélèvement des tweets? Le contexte commercial est un angle mort de l'expérimentation qui a été conduite et qui aurait permis de tirer de meilleures conclusions lorsque les observations concernant l'impact des variables issues de l'analyse de sentiment sur le cours de bourse étaient contraire à la logique.

Il est donc important de garder en mémoire que Twitter se positionne ici comme étant un signal de la perception du public sur une entreprise, mais n'a pas la prétention d'être auto-suffisant dans l'évaluation de la réputation d'entreprise : ce qui est confirmé par les résultats obtenus. Ils indiquent que l'analyse de sentiment ne se démontrent pas utile dans l'explication de la variance du cours de bourse. C'est également un résultats que (Zhang, Fuehres, and Gloor 2011) ont remarqué dans la revue de littérature. Seule la proportion de tweets négatifs est significative et

cela pour un type particulier d'entreprise (celle appartenant au niveau 2 c'est à dire avec un prix d'action entre 50 et 100 unités). Pour ce qui est du nombre de followers d'une entreprise on pouvait s'y attendre : lorsqu'une entreprise voit son nombre de followers augmenter c'est qu'elle gagne en popularité et notoriété, c'est évident. De l'autre côté, on observe que le volume quotidien de tweets agit donc comme un signal fort sur l'activité d'une entreprise : on peut observer une forte croissance du nombre de tweets lorsqu'une crise survient (Warin et al. 2013). Les résultats indiquent une relation inversement proportionnelle du cours de bourse et du volume de tweets. C'est un aspect paradoxal car plus le nombre d'utilisateurs augmente, plus le nombre de tweets quotidien augmente également intuitivement. On aurait pensé observer un comportement similaire du volume de tweets avec le cours de bourse à celle du nombre de followers avec le cours de bourse.

CHAPITRE 5 PROPOSITION D'UNE MESURE DE LA RÉPUTATION D'ENTREPRISE

Dans ce chapitre, nous allons maintenant nous pencher sur la création d'une mesure de la réputation d'entreprise par Twitter. Plusieurs modèles seront proposés, basés sur les résultats obtenus au chapitre 5. Chacun des modèles sera testé à l'aide d'une régression logistique tout comme avec les mesures de réputation, d'image de marque et de responsabilité sociale, présentées au chapitre 4. Le modèle de réputation qui offrira les résultats les plus significatifs sera considéré comme le modèle le plus adéquat.

5.1 La mesure de la réputation d'entreprise à partir de Twitter

La revue de littérature indique que la réputation a un impact sur le cours de bourse d'une entreprise. Elle indique également que Twitter peut être utilisé dans certains modèles financiers pour prédire l'évolution d'indices ou les points de rupture sur les médias (Bollen, Mao, and Zeng 2010; Vincent and Armstrong 2010). La revue de littérature mentionne aussi un lien entre les réseaux sociaux et la réputation d'entreprise (Jones, Temperley and Lima, 2009). Ainsi donc, nous tenterons de concevoir un indice de réputation à partir de Twitter.

La mesure de la réputation d'entreprise à partir de Twitter proposée est un construit des variables qui ont démontré une significativité et une cohérence dans leur impact sur le cours de bourse tel qu'analysé dans le chapitre 5 soit : le volume de tweet, le nombre de followers et le ratio de tweets négatifs. Cette mesure cherche à refléter l'activité d'une entreprise sur Twitter. Une fois la mesure définie pour chaque observation de la base de donnée, elle sera analysée similairement aux cinq sources de mesures de la réputation et d'évaluation de l'image de marque présentées dans le cadre de la première sous-question de recherche (chapitre 4). Ainsi cette mesure constituera la variable indépendante du modèle logit, et la croissance du cours de bourse restera notre variable dépendante. Les résultats obtenus nous permettront de déterminer si Twitter présente une information supplémentaire utile dans la conception d'une mesure de la réputation d'entreprise.

Il est important de mentionner que dans la démarche présentée précédemment, un biais favorisant les résultats peut exister : en effet, nous utilisons des variables qui ont déjà démontré une significativité avec le cours de bourse pour concevoir un indice de réputation. Cet indice devrait donc démontrer une significativité à un seuil acceptable.

La mesure de la réputation proposée, notée « eRep » sera calculée comme suit :

1. Modèle 1 : $eRep_1 = \ln \left[\frac{followers}{Volumetweet} \right]$
2. Modèle 2 : $eRep_2 = \frac{followerslog}{Volumetweetlog}$
3. Modèle 3 : $eR_3 = \frac{followerslog}{Volumetweetlog} * (1 - tn)$

Nous avons observé dans le chapitre 5 que le cours de bourse observe une relation proportionnelle avec le nombre de followers et inversement proportionnelle avec le volume de tweets. De sorte que lorsque le nombre de followers augmente on observe une augmentation du cours de bourse et lorsque le volume de tweet augmente on observe en général une diminution du cours de bourse. Ces observations expliquent pourquoi les numérateurs des modèles de mesures de la réputation sont composés du nombre de followers et le dénominateur est composé du volume de tweet. Concernant le modèle 3, le terme « 1 – tn », représente la proportion de tweets non négatifs.

5.2 Traitement des données : modèle logit

Une fois les modèles de réputation d'entreprise à partir de Twitter conçus, tels que définis dans la section 6.1, nous allons analyser leur influence sur le cours de bourse à l'aide d'une régression logistique similairement au traitement de données effectué avec les mesures de la réputation dans la première partie du mémoire.

Les données sont regroupées sous forme de panel, avec 101 compagnies et 24 observations par compagnie enregistrées entre le 26 février 2014 et le 10 avril 2014. Si l'on assume une

distribution normale $N(0, \sigma_v^2)$ des variables indépendantes de l'analyse statistique, le modèle mathématique logit avec effets aléatoires v_i pour un panel est le suivant :

$$P(y_{i1}, \dots, y_{in_i} | x_{i1}, \dots, x_{in_i}) = \int_{-\infty}^{+\infty} \frac{e^{-v_i^2/2\sigma_v^2}}{\sqrt{2\pi}\sigma_v} \left[\prod_{t=1}^{n_i} F(y_{it}, z) \right] dv_i$$

où l'on a :

- y_{it} : Croissance du cours de bourse pour la $i^{\text{ème}}$ compagnie à l'année t , par rapport à l'année précédente $t-1$;
- x_{it} : Croissance de la mesure de réputation ou évaluations de l'image de marque pour la $i^{\text{ème}}$ compagnie à l'année t par rapport à l'année $t-1$;

ainsi que

$$z = x_{it}\beta + v_i$$

$$F(y, z) = \begin{cases} \frac{1}{1 + \exp(-z)} & \text{si } y \neq 0 \\ \frac{1}{1 + \exp(z)} & \text{autrement} \end{cases}$$

5.3 Analyse des résultats

Les résultats de l'analyse logistique donnée par STATA sont présentés aux tableaux 5.1, 5.2 et 5.3 ci-après. La croissance du cours de bourse est définie par la variable *closebin*, qui prend la valeur 1 lorsque la croissance est positive (augmentation du cours de bourse) et 0 lorsque la croissance est négative (diminution du cours de bourse).

Tableau 5.1 Résultats de la régression logistique du modèle 1

<i>Régression logistique Effets aléatoires</i>			<i>Nombre d'obs. = 2418</i>			
<i>Variable de groupe: ID</i>			<i>Nombre de groupes = 101</i>			
<i>Random effects u_i ~ Gaussian</i>			<i>Moy. obs. par groupe = 23.9</i>			
			<i>Wald chi2(1) = 2.25</i>			
<i>Log Likelihood = -1649.192</i>			<i>Prob > chi2 = 0.1338</i>			
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
eRep1	.03176	.02119	1.50	0.134	-.00976	.07329
_cons	.18546	.08267	2.24	0.025	.02342	.34750
/lnsig2u	-15.341	18.531			-51.661	20.979
sigma_u	.00047	.00432			6.05e-12	35931.25
rho	6.61e-08	1.22e-06			1.11e-23	1

Tableau 5.2 Résultats de la régression logistique du modèle 2

<i>Régression logistique</i>		<i>Nombre d'obs. = 2418</i>				
<i>Variable de groupe: ID</i>		<i>Nombre de groupes = 101</i>				
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>		<i>Moy. obs. par groupe = 23.9</i>				
		<i>Wald $\chi^2(1) = 7.43$</i>				
<i>Log Likelihood = -1646.184</i>		<i>Prob > $\chi^2 = 0.0064$</i>				
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
eRep2	.21882	.08028	2.73	0.006	.06148	.37617
_cons	-.04299	.12934	-0.33	0.740	-.29649	.21050
/lnsig2u	-15.372	18.520			-51.672	20.927
sigma_u	.00046	.00425			6.02e-12	35015.59
rho	6.41e-08	1.19e-06			1.10e-23	1

Tableau 5.3 Résultats de la régression logistique du modèle 3

<i>Régression logistique Effets aléatoires</i>		<i>Nombre d'obs. = 2418</i>				
<i>Variable de groupe: ID</i>		<i>Nombre de groupes = 101</i>				
<i>Effets aléatoires $u_i \sim \text{Gaussian}$</i>		<i>Moy. obs. par groupe = 23.9</i>				
		<i>Wald $\chi^2(1) = 3.56$</i>				
<i>Log Likelihood = -1648.44</i>		<i>Prob > $\chi^2 = 0.0593$</i>				
<i>Increase</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>z</i>	<i>P> z </i>	<i>[95% Conf. Interval]</i>	
eRep3	.15462	.08199	1.89	0.059	-.00608	.31532
_cons	.11648	.10199	1.14	0.253	.08341	.31637
/lnsig2u	-15.367	27.372			-69.015	38.282
sigma_u	.00046	.00630			1.03e-15	2.05e+08
rho	6.44e-08	1.76e-06			3.23e-31	1

Les résultats indiquent que la mesure de réputation la plus adéquate est le modèle 2. En effet, premièrement il s'agit du modèle qui présente le niveau de significativité le plus bas (0,6%), de plus son coefficient (égal à son effet marginal) est le plus élevé comparativement aux autres modèles 1 et 3. Le coefficient égale à $0.219 > 0$, indique que lorsque la mesure de réputation associée au modèle 2 augmente, la probabilité d'avoir d'une hausse du cours de bourse (closebin = 1) augmente. L'effet marginal d'avoir une augmentation d'une unité de la mesure de réputation du modèle 2 indique qu'il est 21.9% plus probable que le cours de bourse augmente comparativement à lorsque celui-ci diminue.

Nous avons également effectué des interactions entre le modèle 2 de la réputation et les industries, afin de déterminer quelles industries en particulier sont plus sensibles à la nouvelle mesure de la réputation par Twitter. L'industrie de référence par rapport à laquelle les coefficients des interactions sont comparés est l'industrie « Airlines » (industrie A) : c'est un choix arbitraire qui n'a pas vraiment d'importance, car nous cherchons simplement à voir ce qui ressort en général des interactions avec les industries. Les résultats sont présentés en ANNEXE 10. L'industrie la plus impactée par la nouvelle mesure de réputation par Twitter est l'industrie des « Computer Software » composée dans notre échantillon de quatre compagnies : Adobe, Microsoft, Oracle, SAP. L'effet marginal est égale à 2.99 ce qui signifie qu'une augmentation de la nouvelle mesure d'une unité, fait augmenter la probabilité d'une hausse du cours de bourse de 300% comparativement à l'industrie « Airlines ». Tandis que l'industrie la moins impactée par cette nouvelle mesure de la réputation est l'industrie des « Beverages » composée dans notre échantillon de : Coca-Cola et Heineken. L'effet marginal est égale à -1.20, ce qui signifie qu'une augmentation de la nouvelle mesure de réputation d'une unité, fait diminuer la probabilité d'une hausse du cours de bourse de -120% comparativement à l'industrie « Airlines » (voir remarque ci-dessous).

Remarque : L'effet marginal doit en général être compris entre -1 (-100%) et 1 (100%). Cependant l'effet marginal traduit le changement de probabilité que la variable dépendante y prenne la valeur 1 lorsque la variable indépendante x augmente d'une unité : elle ne traduit pas une probabilité en elle même mais un taux de variation.

CHAPITRE 6 DISCUSSION, LIMITES ET AVENUES DE LA RECHERCHE

Dans ce chapitre, nous présenterons dans un premier temps les contributions de la recherche, ensuite nous reviendrons sur les limites associées à la recherche et enfin nous présenterons les avenues de la recherche.

6.1 Contribution théorique et pratique de la recherche

D'un point de vue théorique, nous avons pu soulever un doute concernant le lien entre les mesures de réputation et l'appréciation du cours de bourse d'une entreprise. En effet, si nous faisons l'hypothèse que les marchés financiers sont efficients, alors le cours de bourse est supposé capter parfaitement la valeur d'une entreprise. Par conséquent, une augmentation de la réputation devrait conduire à une augmentation de la valeur de l'entreprise et donc du cours de bourse : ce qui n'est pas le cas avec les résultats obtenus, qui ne permettent pas de montrer un lien entre mesure de réputation et cours de bourse. Existe-t-il alors un lien clair entre performances d'entreprises et réputation? Est-ce que les mesures de réputations existantes sont à améliorer? Telles sont les questions que nous pouvons nous poser à l'obtention des résultats observés.

Une autre contribution théorique de ce mémoire est qu'il s'agit de la première fois qu'une recherche a été conduite dans la conception d'une mesure de l'évaluation de la réputation d'entreprise uniquement à partir des données recueillies sur Twitter. Nous avons ainsi pu apporter de l'information supplémentaire sur le lien existant entre les réseaux sociaux et la réputation d'entreprise.

D'un point de vue pratique, nous avons réalisé une base de données unique regroupant 1395 compagnies appartenant à 71 industries différentes, en y associant leur score de réputation (Fortune et Institut de la Réputation), leur évaluation d'image de marque (Brand-Finance et Interbrand) et leur mesure de responsabilité sociale (Corporate Knight), sur une période s'étendant de l'année 2000 à 2013. Il sera ainsi possible pour d'autres recherches de conduire

d'autres types d'analyses, comme par exemple étudier les relations existantes entre les différentes mesures de réputation ou évaluations de l'image de marque.

6.2 Limites de la recherche

Plusieurs limites sont présentes dans le cadre de cette recherche. La première limite est d'ordre technique et concerne la dépendance à un seul outil d'analyse de sentiment fourni par le module « sentiment » de R. La disponibilité de la précision de l'outil dans l'analyse de sentiment n'est pas disponible. Le choix sur R s'est porté ici par une contrainte budgétaire et aussi par soucis de simplification de la tâche puisque R permet également de récolter des tweets : le but était d'effectuer toutes les analyses de données avec un nombre d'outils analytiques restreint.

La deuxième limite est encore d'ordre technique : l'analyse de sentiment a été performée avec une quantité de tweets limités de 100 à 500 tweets extraits quotidiennement par compagnie, ce qui est une quantité de tweets limitée par rapport à la quantité de tweets moyenne disponible pour chaque compagnie. Ainsi, l'analyse de sentiment n'est pas aussi précise que si elle était basée sur un plus large échantillon de tweets. Puisque c'est Twitter qui impose cette contrainte, une des solutions envisagées est la création de plusieurs comptes Twitter tout en effectuant des requêtes multiples sur plusieurs ordinateurs en parallèles. Cette proposition est cependant très coûteuse.

Nous pouvons également mentionner une limite d'ordre méthodologique : l'évolution de la réputation d'entreprise a été simplifiée par l'évolution du cours de bourse de l'entreprise. Toutefois même si une relation existe entre les deux, la réputation n'est pas uniquement constituée par le cours de bourse d'une entreprise. Cet aspect nous rappelle que les résultats obtenus sont donc à mettre en perspective par rapport à cette considération méthodologique.

Nous pouvons également soulever que la présente étude ne présente pas d'analyse sur le lien existant entre le type de marché auquel les firmes composant l'échantillon s'adressent et la propension d'utiliser Twitter comme moyen de propager une opinion sur les entreprises et leurs produits. Cette analyse aurait également permis d'avoir un meilleur aperçu des spécificités de l'utilisation de Twitter selon les secteurs d'activités.

6.3 Avenues de la recherche

Dans la section précédente, nous avons abordé l'utilisation d'un seul outil pour conduire l'analyse de sentiment (le logiciel libre R). Il est donc suggéré dans une recherche ultérieure d'utiliser d'autres outils d'analyse de sentiment afin de déterminer si les résultats concernant l'analyse de sentiment vont dans le même sens que ceux présentés dans le cadre de cette recherche.

Étudier la réputation d'entreprise au travers des réseaux sociaux est encore un nouveau champ de recherche. Nous nous sommes ici penchés sur Twitter, mais il existe encore d'autres réseaux sociaux populaires dans le monde : Facebook, Vk (Russie), Weibo (Twitter chinois), Google +, Baidu space (Google + chinois). Une recherche similaire pourrait être conduite au travers de ces différents réseaux sociaux pour avoir une meilleure idée de l'impact que les réseaux sociaux ont sur la réputation d'entreprise.

Les résultats obtenus en lien avec la deuxième sous-question de recherche, nous laisse suggérer que les différents secteurs d'activités n'ont pas le même lien avec les réseaux sociaux. Il pourrait également être intéressant de chercher à définir le profil des consommateurs qui émettent des tweets pour chaque industrie afin de mieux comprendre la relation qui existe entre une industrie et l'activité sur les réseaux sociaux.

CONCLUSION

La recherche visait donc à concevoir une mesure ou indice de réputation à partir de Twitter qui pourrait offrir une évaluation de la réputation d'entreprise avec une fréquence plus grande que celle offerte par les mesures actuelles de la réputation. Nous avons premièrement vu que les mesures actuelles de la réputation (Fortune & Money, l'Institut de la Réputation, Interbrand, Brand-Finance, Corporate Knight) sont difficilement interprétables et ne se comportent pas de manière cohérente en relation avec le cours de bourse d'entreprise (qui est une approximation de la valeur d'entreprise en supposant un volume d'action en circulation constant au travers des années) telle que vue dans la littérature. En effet, la littérature indique que la réputation d'entreprise et le cours de bourse sont positivement corrélés (Greyser, 1999; Chun, 2005). Ainsi à la première sous-question de recherche, *est-ce que les mesures de réputation proposées actuellement ont un impact sur le cours de bourse?* Nous ne pouvons affirmer que ces mesures de réputation ont un impact sur le cours de bourse à la vue des résultats des régressions logistiques. Les résultats suggèrent qu'il ne semble pas y avoir de lien entre ces classements de la réputation et le cours de bourse. Ainsi avec les résultats obtenus, les hypothèses suivantes n'ont pu être validées :

- Les mesures actuelles de la réputation d'entreprise, d'évaluation monétaires de l'image de marque et de responsabilité sociale ont un impact significatif sur le cours de bourse.
- Les mesures actuelles de la réputation d'entreprise ont un impact plus significatif sur le cours de bourse que les évaluations monétaires de l'image de marque, car la réputation d'entreprise englobe davantage de caractéristiques d'une entreprise que l'image de marque.
- Les mesures actuelles de la réputation d'entreprise ont un impact plus significatif sur le cours de bourse que la mesure de responsabilité sociale car la réputation d'entreprise englobe davantage de caractéristiques d'une entreprise que la responsabilité sociale.
- Les évaluations monétaires de l'image de marque d'une entreprise ont un impact plus significatif sur le cours de bourse que la mesure de responsabilité sociale, car la valeur de la marque est une mesure plus proche de la valeur d'entreprise que le degré de responsabilité sociale d'une entreprise.

La seconde sous-question de recherche fut : *Les données non structurées offertes par Twitter peuvent-elles conduire à la construction d'une mesure de la réputation?* Les résultats des analyses indiquent que les seules informations significatives sont le volume quotidien de tweets, le nombre de followers et la proportion de tweets négatifs avec une journée de lag. Ainsi les hypothèses suivantes sont confirmées :

- La proportion de tweets négatifs est négativement corrélée avec le cours de bourse (et donc négativement corrélée avec la réputation d'entreprise).
- Lorsqu'une compagnie enregistre une présence de plus en plus forte sur les réseaux sociaux, cela signifie que sa réputation augmente et donc son cours de bourse également.

Ainsi, Twitter apporte quelques informations utiles dans la construction d'un indice de réputation d'entreprise au travers des réseaux sociaux. À partir de ces trois variables, nous avons construit une mesure dont la significativité est au seuil de 1% et l'effet marginal est de 21.9% indiquant qu'une augmentation de l'indice conduit à ce qu'il soit 21.9% plus probable qu'une augmentation du cours de bourse soit observée.

Au-delà de ces résultats, en prenant du recul, nous pouvons affirmer que Twitter donne un sentiment général de l'évolution de la réputation d'entreprise mais ne prétend pas en capter toute l'essence. Rappelons-le, la réputation d'entreprise est un construit composé de l'identité d'entreprise, l'identité désiré d'entreprise et l'image de marque (Chun 2005). Twitter ne peut capter qu'une partie de l'image de marque (la loyauté des consommateurs ne peut être captée par Twitter, ni les autres droits de propriétés tels que définis par (Aaker 2009)), et par conséquent Twitter ne peut capter qu'une partie encore plus restreinte de la réputation d'entreprise. Nous indiquerons donc que la nouvelle mesure de réputation par Twitter suggérée doit plutôt être perçue comme une variable de contrôle supplémentaire dans les créations d'indice de réputation existantes. D'ailleurs ces dernières cherchent à inclure davantage de paramètres de la réputation d'entreprise dans leur modèles d'évaluation. En fait la nouvelle mesure de la réputation par Twitter, telle que définie, se démontre être un bel outil pour déterminer la popularité d'une entreprise parmi le public.

BIBLIOGRAPHIE

Aaker, David A. 2004. "Leveraging the Corporate Brand." *California Management Review* (May). <http://hbr.org/product/leveraging-the-corporate-brand/an/CMR281-PDF-ENG>.

———. 2009. *Managing Brand Equity*. Simon and Schuster.

Aaker, David A., and Robert Jacobson. 1994. "The Financial Information Content of Perceived Quality." *Journal of Marketing Research* 31 (2) (May): 191. doi:10.2307/3152193.

Abratt, Russell. 1989. "A New Approach to the Corporate Image Management Process." *Journal of Marketing Management* 5 (1): 63–76.

Agarwal, Apoorv, Fadi Biadisy, and Kathleen R. Mckeown. 2009. "Contextual Phrase-level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-grams." In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 24–32. EACL '09. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1609067.1609069>.

Agarwal, Apoorv, Boyi Xie, Ilya Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. "Sentiment Analysis of Twitter Data." In *Proceedings of the Workshop on Languages in Social Media*, 30–38. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=2021109.2021114>.

Anderson, Eugene W. 1998. "Customer Satisfaction and Word of Mouth." *Journal of Service Research* 1 (1) (August 1): 5–17. doi:10.1177/109467059800100102.

Asur, Sitaram, and Bernardo A. Huberman. 2010. "Predicting the Future with Social Media." In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, 492–499. WI-IAT '10. Washington, DC, USA: IEEE Computer Society. doi:10.1109/WI-IAT.2010.63. <http://dx.doi.org/10.1109/WI-IAT.2010.63>.

Balmer, John M. T., Stephen A. Greyser, John M. T. Balmer, Stephen A. Greyser, and Richard P. Chapman Professor. 2002. "Managing the Multiple Identities of the Corporation." *California Management Review*: 71–86.

Barbosa, Luciano, and Junlan Feng. 2010. "Robust Sentiment Detection on Twitter from Biased and Noisy Data." In *Proceedings of the 23rd International Conference on Computational*

Linguistics: Posters, 36–44. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1944566.1944571>.

Birmingham, Adam, and Alan F. Smeaton. 2011. *On Using Twitter to Monitor Political Sentiment and Predict Election Results*. In : Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP), 13th November 2011, Chiang Mai, Thailand.

Bifet, Albert, and Eibe Frank. 2010. “Sentiment Knowledge Discovery in Twitter Streaming Data.” In *Proceedings of the 13th International Conference on Discovery Science*, 1–15. DS'10. Berlin, Heidelberg: Springer-Verlag. <http://dl.acm.org/citation.cfm?id=1927300.1927301>.

Bifet, Albert, Geoff Holmes, Richard Kirkby, and Bernhard Pfahringer. 2010. “MOA: Massive Online Analysis.” *J. Mach. Learn. Res.* 11 (August): 1601–1604.

Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. 2011. “Twitter Mood Predicts the Stock Market.” *Journal of Computational Science* (March): Pages 1–8. doi:10.1016/j.jocs.2010.12.007.

Bontis, Nick, Lorne D. Booker, and Alexander Serenko. 2007. “The Mediating Effect of Organizational Reputation on Customer Loyalty and Service Recommendation in the Banking Industry.” *Management Decision* 45 (9) (October 23): 1426–1445. doi:10.1108/00251740710828681.

Brown, Eric. 2012. “Will Twitter Make You a Better Investor? A Look at Sentiment, User Reputation and Their Effect on the Stock Market.” *SAIS 2012 Proceedings* (January 1). <http://aisel.aisnet.org/sais2012/7>.

Bughin, Jacques, Michael Chui, and Andy Miller. 2009. “How Companies Are Benefiting from Web 2.0: McKinsey Global Survey Results.” *McKinsey Quarterly*. http://www.mckinsey.com/insights/business_technology/how_companies_are_benefiting_from_web_20_mckinsey_global_survey_results.

Bughin, Jacques, Jonathan Doogan, and Ole Jorgen Vetvik. 2010. “A New Way to Measure Word-of-mouth Marketing | McKinsey & Company.” *McKinsey Quarterly*. http://www.mckinsey.com/insights/marketing_sales/a_new_way_to_measure_word-of-mouth_marketing.

Cheung, Christy M.K., and Matthew K.O. Lee. 2012. “What Drives Consumers to Spread

Electronic Word of Mouth in Online Consumer-opinion Platforms.” *Decision Support Systems* 53 (1) (April): 218–225. doi:10.1016/j.dss.2012.01.015.

Chun, Rosa. 2005. “Corporate Reputation: Meaning and Measurement.” *International Journal of Management Reviews* 7 (2): 91–109. doi:10.1111/j.1468-2370.2005.00109.x.

Doh, Sun-Jae, and Jang-Sun Hwang. 2009. “How Consumers Evaluate eWOM (Electronic Word-of-Mouth) Messages.” *CyberPsychology & Behavior* 12 (2) (April): 193–197. doi:10.1089/cpb.2008.0109.

East, Robert, Kathy Hammond, and Wendy Lomax. 2008. “Measuring the Impact of Positive and Negative Word of Mouth on Brand Purchase Probability.” *International Journal of Research in Marketing* 25 (3) (September): 215–224. doi:10.1016/j.ijresmar.2008.04.001.

Efthymios Kouloumpis, Theresa Wilson. 2011. “Twitter Sentiment Analysis: The Good the Bad and the OMG!”. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*, The AAAI Press, (2011).

Esch, Franz-Rudolf, Tobias Langner, Bernd H. Schmitt, and Patrick Geus. 2006. “Are Brands Forever? How Brand Knowledge and Relationships Affect Current and Future Purchases.” *Journal of Product & Brand Management* 15 (2) (February 1): 98–105. doi:10.1108/10610420610658938.

Fombrun, Charles J. 1996. *Reputation: Realizing Value from the Corporate Image*. Harvard Business Press.

Gabbioneta, Claudia, Davide Ravasi, and Pietro Mazzola. 2007. “Exploring the Drivers of Corporate Reputation: A Study of Italian Securities Analysts.” *Corporate Reputation Review* 10 (2): 99–123. doi:10.1057/palgrave.crr.1550048.

Gayo-Avello, Daniel, Panagiotis Metaxas, and Eni Mustafaraj. 2011. “Limits of Electoral Predictions Using Twitter.”

Go, A, L Huang, and R Bhayani. 2009. “Twitter Sentiment Classification Using Distant Supervision.” *CS224N Project Report*, Stanford edition.

Gray, Edmund R., and John M.T. Balmer. 1998. “Managing Corporate Image and Corporate Reputation.” *Long Range Planning* 31 (5) (October 12): 695–702. doi:10.1016/S0024-

6301(98)00074-0.

Greyser, Stephen A. 1999. "Advancing and Enhancing Corporate Reputation." *Corporate Communications: An International Journal* 4 (4) (December 1): 177–181. doi:10.1108/13563289910299292.

Grützmacher, Antje. 2011. "Reputation 2.0: The Role of Social Media in Corporate Reputation - Case Nokia". Finland: Aalto University School of Economics.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. "The WEKA Data Mining Software: An Update." *SIGKDD Explor. Newsl.* 11 (1) (November): 10–18. doi:10.1145/1656274.1656278.

Hall, Richard. 1992. "The Strategic Analysis of Intangible Resources." *Strategic Management Journal* 13 (2): 135–144. doi:10.1002/smj.4250130205.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler. 2004. "Electronic Word-of-mouth via Consumer-opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?" *Journal of Interactive Marketing* 18 (1): 38–52. doi:10.1002/dir.10073.

Hoeffler, Steve, and Kevin Lane Keller. 2003. "The Marketing Advantages of Strong Brands." *Journal of Brand Management* 10 (6) (August 1): 421–445. doi:10.1057/palgrave.bm.2540139.

Hu, Minqing, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. KDD '04. New York, NY, USA: ACM. doi:10.1145/1014052.1014073. <http://doi.acm.org/10.1145/1014052.1014073>.

Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. "Twitter Power: Tweets as Electronic Word of Mouth." *Journal of the American Society for Information Science and Technology* 60 (11): 2169–2188. doi:10.1002/asi.21149.

Jones, Brian, John Temperley, and Anderson Lima. 2009. "Corporate Reputation in the Era of Web 2.0: The Case of Primark." *Journal of Marketing Management* 25 (9-10): 927–939. doi:10.1362/026725709X479309.

Keller, Ed, and Brad Fay. 2012. "Word-of-Mouth Advocacy: A New Key to Advertising

- Effectiveness.” *Journal of Advertising Research* 52 (4). http://www.kellerfay.com/wp-content/uploads/2012/12/ARF-JAR-WOM-Advocacy-A-New-Key-to-Advertising-Effectiveness_12-10-2012.pdf.
- Keller, Kevin Lane, and Donald R. Lehmann. 2006. “Brands and Branding: Research Findings and Future Priorities.” *Marketing Science* 25 (6) (November 1): 740–759. doi:10.1287/mksc.1050.0153.
- Kim, Soo-min. 2004. “Determining the Sentiment of Opinions.” In *In Proceedings of COLING*, 1367–1373.
- King, Stephen. 1991. “Brand- building in the 1990s.” *Journal of Marketing Management* 7 (1): 3–13. doi:10.1080/0267257X.1991.9964136.
- Knox, Simon, and David Bickerton. 2003. “The Six Conventions of Corporate Branding.” *European Journal of Marketing* 37 (7/8): 998–1016.
- Larkin, Judy. 2002. *Strategic Reputation Risk Management*. Basingstoke: Palgrave Macmillan. <http://www.palgraveconnect.com/pc/doifinder/10.1057/9780230511415>.
- Lim, Choong Hoon, Jinwook Jason Chung, and Paul M. Pedersen. 2012. “Effects of Electronic Word - of - Mouth Messages.” *Choregia* 8 (1) (April 30): 55–76. doi:10.4127/ch.2012.0064.
- Lindenberg, Eric B., and Stephen A. Ross. 1981. “Tobin’s q Ratio and Industrial Organization.” *The Journal of Business* 54 (1): 1–32.
- Liou, James, and Mei-Ling Chuang. 2010. “Evaluating Corporate Image and Reputation Using Fuzzy MCDM Approach in Airline Market.” *Quality & Quantity: International Journal of Methodology* 44 (6): 1079–1091.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Madden, Mary. 2007. “Social Networking Websites and Teens.” *Pew Research Center’s Internet & American Life Project*. <http://www.pewinternet.org/2007/01/07/social-networking-websites-and-teens/>.
- Mangold, Glynn, and David Faulds. 2009. “Social Media: The New Hybrid Element of the Promotion Mix.” *Business Horizons* 52 (4) (July): 357–365. doi:10.1016/j.bushor.2009.03.002.
- Mao, Huina, Scott Counts, and Johan Bollen. 2011. “Predicting Financial Markets: Comparing

Survey, News, Twitter and Search Engine Data.” *arXiv:1112.1051* (December 5). <http://arxiv.org/abs/1112.1051>.

De Marcellis-Warin, Nathalie, and Serban Teodoresco. 2012. “Corporate Reputation: Is Your Most Strategic Asset at Risk?” *CIRANO Burgundy Report 2012RB-02* (April).

Mukherjee, Subhabrata, Akshat Malu, Balamurali A.R., and Pushpak Bhattacharyya. 2012. “TwiSent: A Multistage System for Analyzing Sentiment in Twitter.” In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2531–2534. CIKM ’12. New York, NY, USA: ACM. doi:10.1145/2396761.2398684. <http://doi.acm.org/10.1145/2396761.2398684>.

Nasukawa, Tetsuya, and Jeonghee Yi. 2003. “Sentiment Analysis: Capturing Favorability Using Natural Language Processing.” In *Proceedings of the 2Nd International Conference on Knowledge Capture*, 70–77. K-CAP ’03. New York, NY, USA: ACM. doi:10.1145/945645.945658. <http://doi.acm.org/10.1145/945645.945658>.

O’Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.” <http://cs.wellesley.edu/~cs315/Papers/From%20Tweets%20to%20Polls.pdf>.

Oxford Metrica. 2011. “Reputation Review.” www.oxfordmetrica.com.

Page, Graham, and Helen Fearn. 2005. “Corporate Reputation: What Do Consumers Really Care About?” *Journal of Advertising Research* 45 (03): 305–313.

Pak, Alexander, and Patrick Paroubek. 2010. “Twitter as a Corpus for Sentiment Analysis and Opinion Mining.” In European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf.

Pang, Bo, and Lillian Lee. 2004. “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.” In *Proceedings of the ACL*, 271–278.

———. 2008. “Opinion Mining and Sentiment Analysis.” *Found. Trends Inf. Retr.* 2 (1-2) (January): 1–135. doi:10.1561/1500000011.

Park, C. Whan, Bernard J. Jaworski, and Deborah J. MacInnis. 1986. “Strategic Brand Concept-image Management.” *The Journal of Marketing*: 135–145.

Park, Cheol, and Thae Min Lee. 2009. "Information Direction, Website Reputation and eWOM Effect: A Moderating Role of Product Type." *Journal of Business Research* 62 (1) (January): 61–67. doi:10.1016/j.jbusres.2007.11.017.

Ruiz, Eduardo J., Vagelis Hristidis, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. 2012. "Correlating Financial Time Series with Micro-blogging Activity." In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, 513–522. WSDM '12. New York, NY, USA: ACM. doi:10.1145/2124295.2124358. <http://doi.acm.org/10.1145/2124295.2124358>.

Schroeder, Jonathan, and Miriam Salzer Morling. 2005. *Brand Culture*. Routledge.

Smailovic, Jasmina, Miha Grcar, and Martin Znidarsic. 2012. "Sentiment Analysis on Tweets in a Financial Domain." <http://ipssc.mps.si/papers/smailovic-paper.pdf>.

Sprenger, Timm, and Isabell Welp. 2010. "Tweets and Trades: The Information Content of Stock Microblogs". SSRN Scholarly Paper ID 1702854. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=1702854>.

Sundaram, D.S., and Cynthia Webster. 1999. "The Role of Brand Familiarity on the Impact of Word-Of-Mouth Communication on Brand Evaluations." *NA - Advances in Consumer Research* 26: 664–670.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels. 2008. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site." doi:10.2139/ssrn.1129351.

Tumasjan, A, TO Sprenger, PG Sandner, and IM Welp. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment." In , 178–185. http://scholar.google.de/scholar.bib?q=info:mc319eHjea8J:scholar.google.com/&output=citation&hl=de&as_sdt=0&ct=citation&cd=28.

Tuominen, Pekka. 1999. "Managing Brand Equity." *The Finnish Journal of Business Economics* 1999 :1, 65-100.

Turney, Peter D. 2002. "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews." In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 417–424. ACL '02. Stroudsburg, PA, USA:

Association for Computational Linguistics. doi:10.3115/1073083.1073153.
<http://dx.doi.org/10.3115/1073083.1073153>.

Vincent, Arnaud, and Margaret Armstrong. 2010. “Predicting Break-Points in Trading Strategies with Twitter”. SSRN Scholarly Paper ID 1685150. Rochester, NY: Social Science Research Network.

<https://docs.google.com/viewer?url=http://www.fianzaonline.com/forum/attachments/forex/1499783d1319814776-oibo-ii-la-disfatta-twitter-ts.pdf&chrome=true>.

Wang, Xiaolong, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. “Topic Sentiment Analysis in Twitter: a Graph-based Hashtag Sentiment Classification Approach.” In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 1031–1040. CIKM '11. New York, NY, USA: ACM. doi:10.1145/2063576.2063726.
<http://doi.acm.org/10.1145/2063576.2063726>.

Warin, Thierry, Nathalie de Marcellis-Warin, William Sanger, Bertrand Nembot, and Venus Hosseinali Mirza. 2013. “Corporate Reputation and Social Media: A Game Theory Approach”. CIRANO Working Paper 2013s-18. CIRANO. <http://ideas.repec.org/p/cir/cirwor/2013s-18.html>.

Weigelt, Keith, and Colin Camerer. 1988. “Reputation and Corporate Strategy: A Review of Recent Theory and Applications.” *Strategic Management Journal* 9 (5): 443–454. doi:10.1002/smj.4250090505.

Williams, Martin, Francis Buttle, and Sergio Biggemann. 2012. “Relating Word-of-mouth to Corporate Reputation | VTLS VITAL 5.3, Relating Word-of-mouth to Corporate Reputation.” <http://www.researchonline.mq.edu.au/vital/access/manager/Repository/mq:24559>.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. “Recognizing Contextual Polarity in Phrase-level Sentiment Analysis.” In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347–354. HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1220575.1220619. <http://dx.doi.org/10.3115/1220575.1220619>.

Witten, Ian H., and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Morgan Kaufmann.

Yu, Hong, and Vasileios Hatzivassiloglou. 2003. “Towards Answering Opinion Questions:

Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences.” In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 129–136. EMNLP '03. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1119355.1119372. <http://dx.doi.org/10.3115/1119355.1119372>.

Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. 2011. “Predicting Stock Market Indicators Through Twitter ‘I Hope It Is Not as Bad as I Fear’.” *Procedia - Social and Behavioral Sciences* 26: 55–62. doi:10.1016/j.sbspro.2011.10.562.

ANNEXES

ANNEXE 1 – TEST DE NORMALITÉ

Tableau A1.1 Description des variables indépendantes et de la variable dépendante

Variable	Description
close	Cours de bourse (variable dépendante)
tp	Proportion de tweets positifs
tn	Proportion de tweets négatifs
tneutre	Proportion de tweets neutres
tanger	Proportion de tweets associés au sentiment de colère
tdisgust	Proportion de tweets associés au sentiment de dégoût
tfear	Proportion de tweets associés au sentiment de peur
tsadness	Proportion de tweets associés au sentiment de tristesse
tjoy	Proportion de tweets associés au sentiment de joie
tjoylog	Transformée en logarithme naturel de la variable tjoy
tsurprise	Proportion de tweets associés au sentiment de surprise
Volumetweet	Nombre de tweets quotidiens
Volumetweetlog	Transformée en logarithme naturel de la variable Volumetweet
Ctweet	Croissance du nombre de tweets quotidiens sur deux jours ouvrables consécutifs
followers	Nombre quotidien de followers (suiveurs) d'une entreprise

Tableau A1.1 Description des variables indépendantes et de la variable dépendante (suite)

followerslog	Transformée en logarithme naturel de la variable followers
followees	Nombre quotidien de followees (compte twitter suivis) d'une entreprise
followeeslog	Transformée en logarithme naturel de la variable followees
cfwers	Croissance du nombre de followers sur deux jours consécutifs
cfwees	Croissance du nombre de followees sur deux jours consécutifs
Ctp	Croissance de la proportion de tweets positifs sur deux jours consécutifs
Ctn	Croissance de la proportion de tweets négatifs sur deux jours consécutifs
Ctnlog	Transformée en logarithme naturel de la variable Ctn
Ctneutre	Croissance de la proportion de tweets neutre sur deux jours consécutifs
Ctneutrellog	Transformée en logarithme naturel de la variable Ctneutre

Tableau A1.2 Skewness et Kurtosis avant

Variable	Skewness	Kurtosis
tp	-1.2538	4.9771
tn	1.55	6.06
tneutre	3.27	23.68
tanger	5.75	58.95
tdisgust	16.5	411.67
tfear	8.79	122.72

Tableau A1.2 Skewness et Kurtosis avant (suite)

tsadness	7.82	110.41
tjoy	3.19	20.86
tsurprise	10.27	160.7658
Volumetweet	3.5	16.26
Ctweet	9.21	125.47
followers	3.76	17.45
followees	7.41	63.94
cfwers	5.05	37.77
cfwees	28.64	1078.213
Ctp	18.19	397.9
Ctn	18.38	509.26
Ctneutre	12.64	260.051

Tableau A1.3 Skewness et kurtosis après transformé logarithme naturel (ln)

Variable	Skewness	Kurtosis
tp	-3.65	21.47
tn	-0.72	6.19
tneutre	-0.60	5.49
tanger	4.93	45.09

Tableau A1.3 Skewness et kurtosis après transformé logarithme naturel (ln) (suite)

tdisgust	13.23	266.77
tfear	7.43	89.41
tsadness	5.73	63.81
tjoy	2.3	13.21
tsurprise	8.62	118.24
Volumetweet	-0.16	2.79
Ctweet	8.47	107.46
followers	-1.05	7.49
followees	-0.02	3.14
cfwers	5.05	37.75
cfwees	28.56	1073.947
Ctp	5.31	63.046
Ctn	2.12	15.86
Ctneutre	1.19	11.01

ANNEXE 2 – MATRICE DE CORRÉLATION

Tableau A2.1 Matrice de corrélation globale

Global	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.0341	1								
tn	-0.0535	-0.8967	1							
tneutrelog	0.0152	-0.3934	0.0157	1						
tjoylog	-0.046	0.045	-0.159	0.3254	1					
Volumetwee~g	0.1813	-0.0927	0.0294	0.1464	0.1064	1				
followerslog	0.1057	0.0188	-0.0457	0.0515	0.1091	0.6186	1			
followeeslog	0.2303	0.0929	-0.0813	-0.0185	0.1164	0.1919	0.4084	1		
ctnlog	0.0163	-0.2712	0.3563	-0.0757	-0.0724	0.0205	0.0148	-0.0079	1	
ctneutrelog	0.0001	-0.1522	-0.0746	0.4924	0.0816	0.0157	0.0182	0.0065	-0.0465	1

Tableau A2.2 Tableau de référence dans l'assignation d'une lettre à chacune des 26 industries, pour simplifier la nomination dans les tableaux subséquents

Lettre de réf	Industrie
A	Airlines
B	Apparel
C	Beverages
D	Chemicals
E	Computer Software
F	Computers
G	Consumer Food Products
H	Electric and Gas Utilities
I	Electronics
J	Energy
K	Entertainment
L	Food and Drug Stores
M	Food Services
N	General Merchandisers
O	Health Care: Insurance and Managed Care
P	Insurance: Life and Health
Q	Insurance: Property and Casualty

Tableau A2.2 Tableau de référence dans l'assignation d'une lettre à chacune des 26 industries, pour simplifier la nomination dans les tableaux subséquents (suite)

R	Megabanks
S	Metals
T	Mining, Crude-Oil Production
U	Motor Vehicles
V	Petroleum Refining
W	Pharmaceuticals
X	Soaps and Cosmetics
Y	Superregional Banks
Z	Telecommunications

Tableau A2.3 Matrice de corrélation pour l'industrie « Airlines »

A	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.5872	1								
tn	-0.5607	-0.9691	1							
tneutrelog	-0.2633	-0.3367	0.1137	1						
tjoylog	0.2159	0.225	-0.2718	0.0236	1					
Volumetwee~g	0.7791	0.2913	-0.2558	-0.2214	0.2153	1				
followerslog	0.696	0.6138	-0.5642	-0.3746	0.0832	0.641	1			
followeeslog	-0.3683	-0.0414	0.0656	-0.0886	-0.2411	-0.1353	0.3974	1		
ctnlog	-0.0307	-0.3517	0.4143	-0.3272	0.0211	-0.0063	-0.0317	0.0028	1	
ctneutrelog	0.0632	0.0917	-0.2125	0.4865	0.22	0.0307	0.0563	-0.013	-0.4794	1

Tableau A2.4 Matrice de corrélation pour l'industrie « Appareil »

B	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.1965	1								
tn	-0.013	-0.7011	1							
tneutrelog	-0.2992	-0.6806	0.1159	1						
tjoylog	-0.0742	0.0346	0.008	0.0125	1					
Volumetwee~g	-0.7677	-0.2986	0.2098	0.2379	0.1731	1				
followerslog	-0.4804	-0.2737	0.276	0.1443	0.253	0.9029	1			
followeeslog	0.1882	-0.1597	0.3037	-0.066	0.2503	0.4469	0.7688	1		
ctnlog	0.0195	-0.4381	0.5357	0.1979	0.0628	-0.0143	-0.0463	-0.0281	1	
ctneutrelog	0.0467	-0.3055	-0.1241	0.5858	-0.0935	-0.083	-0.0988	-0.0782	0.0299	1

Tableau A2.5 Matrice de corrélation pour l'industrie « Beverages »

C	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.0925	1								
tn	-0.0242	-0.7398	1							
tneutrelog	-0.1281	-0.6128	0.0672	1						
tjoylog	0.4125	0.1727	-0.0694	-0.2116	1					
Volumetwee~g	-0.9486	-0.0149	-0.0574	0.1156	-0.4142	1				
followerslog	-0.9636	-0.052	0.0097	0.0638	-0.4515	0.9442	1			
followeeslog	-0.9679	-0.0472	0.0051	0.0654	-0.4433	0.9538	0.9986	1		
ctnlog	0.0132	-0.5105	0.7285	0.0342	-0.1491	-0.0772	0.0279	0.0175	1	
ctneutrelog	0.0376	-0.2142	-0.1386	0.5587	-0.0271	-0.0307	-0.0492	-0.048	0.1431	1

Tableau A2.6 Matrice de corrélation pour l'industrie « Chemicals »

D	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.0937	1								
tn	-0.1033	-0.9505	1							
tneutrelog	0.0816	-0.2969	-0.0007	1						
tjoylog	0.1022	0.1186	-0.1837	0.1797	1					
Volumetwee~g	0.38	-0.3994	0.3655	0.2127	0.2818	1				
followerslog	0.6476	-0.2767	0.2376	0.2479	0.3231	0.8631	1			
followeeslog	0.9675	0.0343	-0.0554	0.1291	0.1272	0.4721	0.7418	1		
ctnlog	0.0191	-0.3725	0.4061	-0.0071	-0.1384	0.0342	0.0128	0.0255	1	
ctneutrelog	-0.0722	-0.0302	-0.1505	0.561	0.3265	-0.0676	-0.0398	-0.0476	-0.0243	1

Tableau A2.7 Matrice de corrélation pour l'industrie « Computer software »

E	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.6114	1								
tn	0.6001	-0.9487	1							
tneutrelog	0.2711	-0.3747	0.0874	1						
tjoylog	0.1831	-0.133	-0.0118	0.4295	1					
Volumetwee~g	-0.2937	0.1263	-0.0947	-0.1564	-0.2366	1				
followerslog	-0.6575	0.498	-0.4669	-0.2563	-0.2666	0.8368	1			
followeeslog	-0.3981	0.8514	-0.8348	-0.2741	-0.0879	0.1909	0.5092	1		
ctnlog	0.0101	-0.1116	0.1922	-0.2401	-0.187	0.0017	0.0094	0.0476	1	
ctneutrelog	-0.034	-0.1458	-0.1047	0.7296	0.2487	0.011	0.0146	-0.0172	-0.3006	1

Tableau A2.8 Matrice de corrélation pour l'industrie « Computers »

F	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.2367	1								
tn	0.2929	-0.918	1							
tneutrelog	-0.1836	-0.1259	-0.2623	1						
tjoylog	0.0011	0.2814	-0.3506	0.244	1					
Volumetwee~g	0.8229	-0.1044	0.1577	-0.167	0.0344	1				
followerslog	0.5706	-0.2435	0.2577	-0.06	0.0011	0.0643	1			
followeeslog	-0.6182	0.0334	-0.092	0.165	-0.0127	-0.9178	0.1489	1		
ctnlog	0.0598	-0.6725	0.6965	-0.1092	-0.2078	0.0606	0.0077	-0.0555	1	
ctneutrelog	-0.024	-0.1901	-0.0915	0.6913	0.0153	-0.0126	-0.0232	0.0221	0.0237	1

Tableau A2.9 Matrice de corrélation pour l'industrie « Consumer Food Products »

G	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.1631	1								
tn	-0.2325	-0.8203	1							
tneutrelog	-0.0447	-0.3478	-0.1763	1						
tjoylog	0.2816	-0.0209	-0.0634	0.1961	1					
Volumetwee~g	0.4713	-0.1727	0.0907	0.0042	0.0949	1				
followerslog	0.3752	-0.2681	0.1881	0.0757	0.0879	0.7396	1			
followeeslog	0.315	-0.2251	0.1232	0.1012	0.1346	0.8394	0.8967	1		
ctnlog	-0.0369	-0.4505	0.5591	-0.1283	-0.0406	-0.0521	-0.0118	-0.0221	1	
ctneutrelog	0.0373	-0.2188	-0.1492	0.6017	0.0086	0.048	0.0571	0.0514	-0.2329	1

Tableau A2.10 Matrice de corrélation pour l'industrie « Electric and Gas utilities »

H	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.1608	1								
tn	-0.3977	-0.9396	1							
tneutrelog	0.5809	-0.4353	0.1094	1						
tjoylog	0.8055	0.1272	-0.351	0.5098	1					
Volumetwee~g	-0.761	-0.2682	0.4744	-0.46	-0.6403	1				
followerslog	0.9831	0.1185	-0.3747	0.6439	0.8304	-0.7546	1			
followeeslog	0.9803	0.117	-0.3762	0.6532	0.8328	-0.7585	0.9997	1		
ctnlog	0.0167	-0.333	0.3617	0.029	0.159	-0.0492	0.0333	0.0301	1	
ctneutrelog	0.0188	-0.2615	0.1309	0.381	-0.0576	0.0961	0.0126	0.0105	0.184	1

Tableau A2.11 Matrice de corrélation pour l'industrie « Electronics »

I	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.0748	1								
tn	-0.0514	-0.7522	1							
tneutrelog	0.1186	-0.5278	-0.0558	1						
tjoylog	-0.1427	-0.1366	0.1016	0.203	1					
Volumetwee~g	0.5532	-0.1331	0.0309	0.0797	0.1409	1				
followerslog	0.4061	-0.217	0.1642	0.0723	0.3604	0.8291	1			
followeeslog	0.7273	-0.1151	0.1025	-0.0057	-0.0276	0.3324	0.4739	1		
ctnlog	0.0596	-0.3982	0.5382	0.0372	-0.0018	0.0867	0.0598	0.0354	1	
ctneutrelog	-0.0121	-0.3025	-0.0732	0.592	0.0929	0.0422	0.0625	-0.0282	0.1423	1

Tableau A2.12 Matrice de corrélation pour l'industrie « Energy »

J	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.3827	1								
tn	0.3776	-0.9163	1							
tneutrelog	0.0854	-0.5158	0.135	1						
tjoylog	-0.0466	-0.1177	0.0742	0.2309	1					
Volumetwee~g	-0.4572	0.097	-0.1572	0.0858	-0.0419	1				
followerslog	-0.9659	0.4066	-0.4068	-0.0888	0.0527	0.5826	1			
followeeslog	-0.9375	0.4084	-0.4119	-0.0884	0.0622	0.6289	0.9947	1		
ctnlog	0.0548	-0.5982	0.6702	0.0201	0.1284	0.0245	-0.0578	-0.0541	1	
ctneutrelog	-0.0101	-0.2585	0.0145	0.5843	0.1512	0.0237	0.0193	0.0274	0.1805	1

Tableau A2.13 Matrice de corrélation pour l'industrie « Entertainment »

K	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.0505	1								
tn	-0.4472	-0.4721	1							
tneutrelog	0.1772	-0.6119	-0.2224	1						
tjoylog	0.3235	-0.2449	-0.3904	0.5183	1					
Volumetwee~g	0.9669	0.1541	-0.4513	0.0834	0.2609	1				
followerslog	-0.9858	-0.0917	0.4461	-0.1194	-0.2598	-0.9831	1			
followeeslog	-0.984	-0.0905	0.4455	-0.1211	-0.2496	-0.9803	0.9994	1		
ctnlog	0.2202	0.0737	0.1855	-0.1448	-0.2619	0.2233	-0.2143	-0.2174	1	
ctneutrelog	0.2373	-0.4666	-0.2629	0.7109	0.4914	0.1926	-0.2348	-0.2327	-0.1818	1

Tableau A2.14 Matrice de corrélation pour l'industrie « Food and Drug Stores »

L	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.0323	1								
tn	0.0734	-0.7756	1							
tneutrelog	-0.0309	-0.575	-0.0155	1						
tjoylog	-0.0766	-0.1023	-0.0158	0.1129	1					
Volumetwee~g	0.6444	-0.0701	-0.0275	0.1606	-0.1855	1				
followerslog	0.873	0.0206	0.0285	-0.053	-0.099	0.5399	1			
followeeslog	0.5632	0.1467	0.0425	-0.2811	0.0816	0.2449	0.3719	1		
ctnlog	-0.0356	-0.4006	0.5726	-0.0362	-0.083	0.0093	-0.0387	0.0218	1	
ctneutrelog	-0.0079	-0.2798	-0.0805	0.5875	0.0044	0.0458	-0.0006	0.0335	-0.081	1

Tableau A2.15 Matrice de corrélation pour l'industrie « Food Services »

M	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.1389	1								
tn	0.2322	-0.8588	1							
tneutrelog	-0.1729	-0.3894	-0.1254	1						
tjoylog	-0.4125	-0.0905	-0.0924	0.342	1					
Volumetwee~g	-0.9415	0.0577	-0.1611	0.1974	0.3719	1				
followerslog	-0.9917	0.1525	-0.2382	0.1525	0.4059	0.9522	1			
followeeslog	-0.9924	0.1506	-0.2365	0.1539	0.4058	0.9534	0.9999	1		
ctnlog	0.0729	-0.452	0.5791	-0.1647	-0.1217	-0.025	-0.0621	-0.0609	1	
ctneutrelog	-0.0067	-0.286	-0.1496	0.8129	0.162	0.0436	0.0053	0.0052	-0.1669	1

Tableau A2.16 Matrice de corrélation pour l'industrie « General Merchandisers »

N	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.0882	1								
tn	0.045	-0.7958	1							
tneutrelog	0.0645	-0.6001	0.056	1						
tjoylog	0.0284	-0.0024	-0.0779	0.1604	1					
Volumetwee~g	0.5608	-0.0833	0.0649	0.1028	-0.0489	1				
followerslog	0.2631	-0.0917	0.0597	0.1226	0.0384	0.8873	1			
followeeslog	-0.147	0.2502	-0.2243	-0.1783	0.3513	-0.589	-0.4146	1		
ctnlog	-0.0461	-0.3907	0.6076	-0.1251	0.0435	-0.0399	-0.0264	0.0459	1	
ctneutrelog	0.0051	-0.3344	-0.044	0.6419	0.2636	-0.0635	-0.0151	0.0549	-0.0228	1

Tableau A2.17 Matrice de corrélation pour l'industrie « Health Care : Insurance and Managed Care »

O	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.0871	1								
tn	0.4113	-0.7845	1							
tneutrelog	-0.4792	-0.5368	-0.4906	1						
tjoylog	0.0774	0.634	-0.6697	-0.1281	1					
Volumetwee~g	-0.0712	-0.6381	0.4762	0.3039	-0.5029	1				
followerslog	0.564	-0.1364	0.3581	-0.2462	-0.1137	-0.4243	1			
followeeslog	0.5946	-0.0803	0.3713	-0.3304	-0.1514	-0.3663	0.9664	1		
ctnlog	-0.1088	-0.6531	0.6865	-0.1198	-0.6086	0.5728	-0.1091	-0.1211	1	
ctneutrelog	0.1022	0.3617	-0.5107	0.375	0.5933	-0.2832	-0.0302	-0.0907	-0.5869	1

Tableau A2.18 Matrice de corrélation pour l'industrie « Insurance : Life and Health »

P	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.1984	1								
tn	0.4322	-0.8505	1							
tneutrelog	-0.4638	-0.3268	-0.1823	1						
tjoylog	-0.6728	0.0864	-0.3331	0.5018	1					
Volumetwee~g	0.3757	-0.326	0.5065	-0.3168	-0.3623	1				
followerslog	0.8793	0.0712	0.1272	-0.4473	-0.5203	0.0771	1			
followeeslog	-0.934	0.3379	-0.5539	0.395	0.6647	-0.5227	-0.6631	1		
ctnlog	-0.0591	-0.3104	0.3586	-0.0125	0.0823	0.024	-0.0158	0.0984	1	
ctneutrelog	0.0641	-0.1373	-0.0226	0.3719	-0.0992	0.0006	0.0756	-0.0592	0.037	1

Tableau A2.19 Matrice de corrélation pour l'industrie « Insurance : Property and Casualty »

Q	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.3371	1								
tn	-0.2361	-0.8788	1							
tneutrelog	-0.2183	-0.3949	-0.0621	1						
tjoylog	-0.0636	-0.2717	0.1687	0.3179	1					
Volumetwee~g	-0.3649	-0.1919	0.1352	0.0339	-0.2204	1				
followerslog	0.2942	0.0078	-0.0401	0.0192	0.0644	0.0635	1			
followeeslog	-0.9993	-0.3364	0.2361	0.2161	0.0601	0.3611	-0.2813	1		
ctnlog	-0.0342	-0.6362	0.7408	-0.0652	-0.0853	0.1632	0.1122	0.0323	1	
ctneutrelog	-0.026	-0.0142	-0.3586	0.7661	0.1701	-0.0803	-0.0256	0.0274	-0.256	1

Tableau A2.20 Matrice de corrélation pour l'industrie « Megabanks »

R	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.1231	1								
tn	-0.0675	-0.8931	1							
tneutrelog	-0.1539	-0.3686	-0.026	1						
tjoylog	-0.07	0.0765	-0.1654	0.2703	1					
Volumetwee~g	-0.0917	-0.1324	0.059	0.1325	0.1387	1				
followerslog	0.4781	0.3424	-0.2346	-0.3222	-0.114	-0.0253	1			
followeeslog	0.1074	0.2103	-0.092	-0.2818	-0.0369	-0.0743	0.634	1		
ctnlog	-0.008	-0.2825	0.3939	-0.1888	-0.0882	-0.0061	0.015	-0.0242	1	
ctneutrelog	0.0335	-0.1386	-0.0915	0.498	0.0976	0.0736	0.007	0.0285	-0.1648	1

Tableau A2.21 Matrice de corrélation pour l'industrie « Metals »

S	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.1963	1								
tn	0.258	-0.9615	1							
tneutrelog	-0.1088	-0.0188	-0.2252	1						
tjoylog	0.3385	-0.165	0.0233	0.604	1					
Volumetwee~g	-0.2701	-0.0816	0.0947	-0.2555	-0.3379	1				
followerslog	0.6584	0.2452	-0.1373	-0.2324	-0.061	0.0002	1			
followeeslog	-0.9122	-0.0284	-0.0589	0.171	-0.1502	0.1348	-0.9013	1		
ctnlog	-0.0064	-0.3146	0.3858	-0.2682	-0.1968	0.2297	0.0175	0.0021	1	
ctneutrelog	-0.0607	-0.2257	0.151	0.3483	0.2143	-0.0376	-0.1199	0.0989	-0.2031	1

Tableau A2.22 Matrice de corrélation pour l'industrie « Mining, Crude-Oil Production »

T	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.2448	1								
tn	-0.2339	-0.9665	1							
tneutrelog	0.0462	0.1363	-0.3637	1						
tjoylog	-0.3863	-0.1942	0.1184	0.2582	1					
Volumetwee~g	0.2378	-0.3778	0.3472	-0.0055	-0.0379	1				
followerslog	-0.3283	0.0239	0.1032	-0.512	-0.208	-0.6408	1			
followeeslog	0.0927	-0.4073	0.495	-0.5065	-0.3422	0.5308	0.9613	1		
ctnlog	0.0483	-0.1477	0.1909	-0.0419	-0.157	-0.0534	-0.0141	-0.0446	1	
ctneutrelog	-0.1892	-0.2573	0.1323	0.3638	0.1267	-0.0501	0.0105	0.1015	-0.2917	1

Tableau A2.23 Matrice de corrélation pour l'industrie « Motor Vehicles »

U	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.0177	1								
tn	0.0458	-0.8426	1							
tneutrelog	-0.0936	-0.4488	-0.0222	1						
tjoylog	-0.0184	-0.012	-0.2019	0.4557	1					
Volumetwee~g	-0.2759	-0.1642	0.1517	0.0798	0.1421	1				
followerslog	-0.1595	-0.1642	0.126	0.1354	0.243	0.907	1			
followeeslog	-0.1772	0.0858	-0.038	-0.1772	-0.175	-0.0486	-0.0254	1		
ctnlog	0.0026	-0.481	0.6023	-0.078	-0.0825	-0.0467	-0.0275	0.0358	1	
ctneutrelog	0.004	-0.2092	-0.1145	0.5606	0.1293	-0.0504	-0.0511	0.0792	-0.0609	1

Tableau A2.24 Matrice de corrélation pour l'industrie « Petroleum Refining »

V	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.3953	1								
tn	0.2051	-0.9051	1							
tneutrelog	0.6822	-0.5862	0.2271	1						
tjoylog	0.7157	-0.2037	0.0156	0.5907	1					
Volumetwee~g	0.6237	-0.8422	0.7315	0.6188	0.4373	1				
followerslog	-0.027	0.7603	-0.6796	-0.4099	-0.1141	-0.6586	1			
followeeslog	-0.9693	0.1851	-0.0083	-0.5812	-0.6915	-0.4437	-0.2044	1		
ctnlog	-0.0855	-0.1301	0.2984	-0.2445	-0.087	-0.0466	0.021	0.0794	1	
ctneutrelog	-0.0727	-0.0315	-0.1685	0.3588	-0.0474	-0.0414	0.0143	0.0601	-0.3224	1

Tableau A2.25 Matrice de corrélation pour l'industrie « Pharmaceuticals »

W	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.2342	1								
tn	-0.2994	-0.8023	1							
tneutrelog	0.0288	-0.4767	-0.0246	1						
tjoylog	0.373	0.0157	-0.1209	0.4558	1					
Volumetwee~g	-0.6969	-0.4204	0.3946	0.1869	-0.2084	1				
followerslog	0.8661	0.1011	-0.1531	-0.0917	0.1	-0.7151	1			
followeeslog	-0.1244	0.1376	-0.1172	0.1796	0.4075	0.3519	-0.5575	1		
ctnlog	0.0007	-0.3699	0.4757	-0.014	-0.0763	0.0012	0.0377	-0.0708	1	
ctneutrelog	-0.0383	-0.2415	0.0304	0.3807	0.0348	-0.0707	-0.0057	-0.0396	0.3801	1

Tableau A2.26 Matrice de corrélation pour l'industrie « Soaps and Cosmetics »

X	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.9113	1								
tn	-0.9398	-0.9645	1							
tneutrelog	0.1404	-0.078	-0.1728	1						
tjoylog	0.6632	0.7042	-0.724	0.1706	1					
Volumetwee~g	0.341	0.29	-0.3241	0.1256	0.2369	1				
followerslog	-0.995	-0.9063	0.9412	-0.1683	-0.6712	-0.3795	1			
followeeslog	-0.9831	-0.9023	0.9268	-0.1397	-0.6641	-0.3989	0.9815	1		
ctnlog	0.0519	-0.1352	0.1108	0.0576	-0.3158	0.0917	-0.0673	-0.0661	1	
ctneutrelog	0.0163	-0.1162	-0.0534	0.6427	0.0128	-0.1374	-0.0147	-0.0006	-0.1607	1

Tableau A2.27 Matrice de corrélation pour l'industrie « Superregional Banks »

Y	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	-0.156	1								
tn	-0.0995	-0.8504	1							
tneutrelog	0.6114	-0.0803	-0.3816	1						
tjoylog	0.4741	0.0436	-0.2277	0.5592	1					
Volumetwee~g	0.8936	-0.3632	0.0966	0.5484	0.3945	1				
followerslog	0.1493	0.4641	-0.5042	0.2591	0.4403	-0.0962	1			
followeeslog	-0.667	0.4547	-0.3014	-0.2726	-0.068	-0.7838	0.6152	1		
ctnlog	-0.0299	-0.2483	0.3803	-0.2773	-0.1149	0.0239	0.0424	0.0452	1	
ctneutrelog	-0.009	-0.2407	-0.0478	0.3774	-0.0378	0.121	-0.0893	-0.0389	-0.3464	1

Tableau A2.28 Matrice de corrélation pour l'industrie « Telecommunications »

Z	close	tp	tn	tneutr~g	tjoylog	Volume~g	fo~rslog	fo~eslog	ctnlog	ctneut~g
close	1									
tp	0.247	1								
tn	-0.0258	-0.8487	1							
tneutrelog	-0.3575	-0.7007	0.2478	1						
tjoylog	-0.5175	-0.5728	0.3464	0.5938	1					
Volumetwee~g	0.6228	0.5036	-0.3224	-0.4362	-0.6898	1				
followerslog	-0.7795	-0.4428	0.2464	0.4035	0.5525	-0.5896	1			
followeeslog	-0.3211	0.5039	-0.4385	-0.3279	-0.2925	0.1986	-0.1799	1		
ctnlog	-0.0231	-0.3379	0.4945	0.0155	0.1337	-0.0059	-0.0249	0.0649	1	
ctneutrelog	-0.04	-0.2118	-0.031	0.4827	0.1045	0.0013	-0.0019	0.0934	0.1717	1

ANNEXE 3 – TEST DE MULTICOLINÉARITÉ

Tableau A3.1 Multicolinéarité initiale

Variable	VIF	1/VIF
tp	20.28	0.049313
tn	17.41	0.057433
tneutrelog	4.21	0.23781
followerslog	2.14	0.468083
Volumetwee~g	1.9	0.526874
ctneutrelog	1.4	0.715253
followeeslog	1.23	0.810464
tjoylog	1.22	0.820652
ctnlog	1.17	0.858049
Mean VIF	5.66	

Tableau A3.2 Multicolinéarité après ajustement (omission de tneutrelog)

Variable	VIF	1/VIF
tn	8.02	0.124728
tp	7.73	0.129444
followerslog	2.14	0.46783
Volumetwee~g	1.9	0.526688
ctneutrelog	1.36	0.736252
followeeslog	1.23	0.810573
ctnlog	1.17	0.857252
tjoylog	1.12	0.894411
Mean VIF	3.08	

ANNEXE 4 – TEST D’HAUFFMAN

```
. hausman fixed ., sigmamore
```

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) .		
tp	954.8402	966.4311	-11.59088	30.656
tn	513.1723	527.9752	-14.80291	46.03209
tjoylog	-114.1859	-112.8365	-1.349452	2.683397
Volumetweevg	-29.26946	-3.053884	-26.21558	16.29843
followerslog	935.2238	1201.919	-266.6951	957.6561
followeeslog	2450.221	4167.596	-1717.375	980.6917
ctnlog	72.82851	70.36797	2.460535	3.908665
ctneutrellog	-134.1738	-132.6601	-1.513607	2.842802

b = consistent under Ho and Ha; obtained from xtreg

B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned}
 \text{chi2}(8) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\
 &= \mathbf{16.30} \\
 \text{Prob}>\text{chi2} &= \mathbf{0.0383}
 \end{aligned}$$

Figure A4.1 Test d'Hauffman pour les régressions linéaires

ANNEXE 5 – REGRÉSSION LINÉAIRE GLOBALE

Tableau A5.1 Régression linéaire globale associé au modèle 1

Variable	Sans lag	1 Jour de lag	2 jours de lag
tp	572.74531 0.5686		
tjoylog	-120.13367 0.5126		
Volumetwee~g	-35.783279 0.8665		
followeeslog	2708.8334 0.2806		
ctnlog	109.25187 0.7826		
ctneutrelog	-174.35229 0.5318		
Tp L1.		152.56094 0.9086	
Tp L2.			-410.29813 0.7666
Tjoylog L1.		-142.47813 0.5559	
Tjoylog L2.			-216.08629 0.3988

Tableau A5.1 Régression linéaire globale associé au modèle 1 (suite)

Volumetwee~g L1.		31.074685	
		0.9073	
Volumetwee~g L2.			37.442267
			0.898
Followeeslog L1.		2799.6316	
		0.3855	
followeeslog L2.			2613.6169
			0.4464
Ctnlog L1.		134.58407	
		0.7954	
Ctnlog L2.			116.7881
			0.8216
Ctneutrelog L1.		-328.14529	
		0.3861	
Ctneutrelog L2.			-129.11512
			0.74
_cons	-2869.9739	-4886.0599	-3004.5781
	0.8659	0.8219	0.896
N	2258	1368	976
r2_o	0.05337016	0.04754851	0.05174966
r2_b	0.05575385	0.05680735	0.05818055

Tableau A5.2 Régression linéaire globale associée au modèle 2

Variable	Sans lag	1 Jour de lag	2 jours de lag
tn	-400.60268 0.7208		
tjoylog	-124.07343 0.5007		
followerslog	1039.4856 0.6951		
followeeslog	2426.8155 0.3565		
ctnlog	85.0816 0.839		
ctneutrelog	-227.21724 0.3948		
tn L1.		213.61812 0.8866	
tn L2.			-44.192393 0.9774
Tjoylog L1.		-139.90844 0.5659	
Tjoylog L2.			-222.54401 0.3899
Followerslog L1.		1119.1646 0.7457	

Tableau A5.2 Régression linéaire globale associée au modèle 2 (suite)

Followerslog L2.			1009.219
			0.7795
Followeeslog L1.		2488.6082	
		0.4611	
Followeeslog L2.			2363.8259
			0.5088
Ctnlog L1.		68.578655	
		0.9001	
Ctnlog L2.			191.43832
			0.7276
Ctneutrelog L1.		-337.16175	
		0.3534	
Ctneutrelog L2.			-98.369442
			0.7918
_cons	-12417.188	-15021.816	-12698.396
	0.6739	0.6973	0.7549
N	2258	1368	976
r2_o	0.05216034	0.04495156	0.04911781
r2_b	0.05467556	0.05394382	0.05544639

ANNEXE 6 – RÉGRESSION LINÉAIRE PAR INDUSTRIE

Tableau A6.1 Régression linéaire avant ajustement du modèle 1 par industrie

Variable	IA	IB	IC	ID	IE	IF	IG	IH	II
tp	-0.38945763	4.1381163	3.4044694	-1.5007712	-5.4458434	2.5822933	-1.8375018	-0.3663874	11890.112
	0.5854	0.1653	0.0339	0.4382	0.1486	0.6252	0.1698	0.71	0.3696
tjoylog	-0.08495337	-0.79705171	0.00278032	-0.03397393	-0.18595159	0.22450571	0.16516884	-0.03132602	-1839.2318
	0.5856	0.0476	0.9937	0.951	0.7243	0.7631	0.5678	0.8903	0.4009
Volumetwee~g	0.08474683	0.85005535	-0.58884131	-0.25798166	-0.4681664	0.94681802	-0.42715303	0.01137986	-437.61489
	0.4987	0.1649	0.1701	0.5387	0.4373	0.2062	0.0496	0.9858	0.8585
followeeslog	0.33030655	-82.226575	18.658997	-11.341142	7.6421512	-3.4884815	32.76729	27.015802	94501.72
	0.8996	0.0025	0.0001	0.0009	0.0287	0.8639	0.44	0	0.0538
ctnlog	0.20492345	-0.70914766	0.59497489	-0.37268481	-0.59209566	0.05644464	-0.8793302	-0.56416645	2340.3502
	0.6815	0.4023	0.2777	0.5783	0.6979	0.9756	0.1579	0.467	0.5635
ctneutrelog	0.29145459	0.46309084	0.21647279	-1.1230495	-0.58543674	0.25506049	-0.30928451	0.05912319	-1704.762
	0.2833	0.4057	0.5463	0.1215	0.5539	0.8438	0.3339	0.9036	0.5839
_cons	19.149151	492.10342	-102.22156	150.63156	12.656388	138.32889	-187.60219	-119.41616	-606288.47
	0.3613	0.0002	0.0047	0	0.624	0.2507	0.5651	0	0.1196
N	91	92	43	68	92	114	115	45	205
r2_o	0.09554802	0.04634423	0.93204449	0.93444143	0.08608261	0.44388829	0.09790209	0.96074599	0.52775093

Tableau A6.1 Régression linéaire avant ajustement du modèle 1 par industrie (suite)

r2_b	0.11728406	0.04683234	1	0.97794847	0.09259532	0.43269463	0.09858968	1	0.53844754
Variable	IJ	IK	IL	IM	IN	IO	IP	IQ	IR
tp	-0.43448519	-1.8288582	1.6625601	4.4269216	-1.3164064	-1.5856709	-2.6207411	2066.3052	0.86194463
	0.5662	0.1552	0.359	0.2706	0.3097	0.4859	0.2323	0.7837	0.436
tjoylog	-0.14025273	0.99546272	0.03355911	0.49498027	-0.21882938	0.51984937	-0.02826915	-272.07678	-0.45323275
	0.1814	0.0017	0.913	0.5331	0.4329	0.2377	0.9174	0.8124	0.0104
Volumetwee~g	0.10995954	0.36721341	0.47550683	1.3417196	-0.63275315	1.1866418	0.38553923	-1141.1974	0.36400705
	0.3246	0.4349	0.0288	0.0684	0.0469	0.2485	0.435	0.2064	0.1255
followeeslog	190.59941	40.918388	18.917426	42.044965	11.437434	44.884866	-7.1587224	38038.679	14.358839
	0.0002	0.3143	0.0006	0.185	0.8231	0.0534	0.0259	0.6101	0.1534
ctnlog	0.01809735	0.53430811	0.29620198	1.1822866	-0.36216596	-0.49482978	0.07678333	36.199022	0.29072096
	0.9469	0.1821	0.5755	0.3155	0.4474	0.6851	0.9271	0.9889	0.5549
ctneutrelog	-0.12231945	-0.44711127	-0.2169281	-0.06782127	0.1631653	0.14087284	-1.350575	317.92675	0.28529372
	0.5457	0.2649	0.5977	0.9409	0.6312	0.8927	0.0634	0.8026	0.3223
_cons	-1099.1236	-139.05729	-90.661853	-370.92502	-29.467875	-156.51002	85.210543	-32049.459	-28.685857
	0.0002	0.5118	0.0271	0.2654	0.947	0.1811	0	0.8993	0.5896
N	64	46	92	46	92	22	65	46	226

Tableau A6.1 Régression linéaire avant ajustement du modèle 1 par industrie (suite)

r2_o	0.88585545	0.96544553	0.32033228	0.98384877	0.03341381	0.38319469	0.84664698	0.99882565	0.01186435
r2_b	0.8809665	1	0.32137151	1	0.03432199	.	0.86535446	1	0.01158114
Variable	IS	IT	IU	IV	IW	IX	IY	IZ	
tp	-17.503363	4.9077929	0.8515793	-4.1735263	0.5063071	-0.26077097	0.57555838	-7.2902974	
	0.2196	0	0.7502	0.0039	0.6749	0.926	0.7173	0.0005	
tjoylog	3.1912216	-0.19773572	-0.39272344	-0.16363379	-0.14973702	-0.38364972	0.57678588	-0.81692374	
	0.3284	0.1727	0.3522	0.5407	0.5211	0.271	0.1655	0.0488	
Volumetwee~g	-7.0967461	-0.17386302	-1.6005723	-0.2724102	-0.25957639	-1.1556603	0.0824993	0.24807871	
	0.107	0.5207	0.0109	0.1862	0.4201	0.0637	0.8239	0.3357	
followeeslog	28.639564	-4.8841883	-25.249908	-13.174213	-6.9568816	-48.840742	4.1540559	29.739233	
	0.5909	0.6336	0.2596	0.0586	0.2584	0.0021	0.0002	0.0001	
ctnlog	3.1569027	-0.02517414	0.3384013	-1.0902483	0.66865892	-2.0002298	-0.71238814	-1.4297189	
	0.6809	0.9486	0.7133	0.0793	0.0995	0.1735	0.4751	0.031	
ctneutrelog	-1.2626157	0.35407245	0.12657817	-0.70410288	-0.00016493	-0.12400437	0.35488533	-0.27262583	
	0.7742	0.3401	0.8398	0.096	0.9996	0.8461	0.6183	0.5966	
_cons	21.050756	58.177901	295.74922	136.64268	109.39677	449.82748	39.747972	-184.28187	
	0.9532	0.2604	0.1152	0.002	0.0088	0.0001	0	0.001	
N	69	65	160	92	102	45	69	92	

Tableau A6.1 Régression linéaire avant ajustement du modèle 1 par industrie (suite)

r2_o	0.59040056	0.00203486	0.03639073	0.95392921	0.1287973	0.94693827	0.38806867	0.10955363	
r2_b	0.6511698	0.00040545	0.0366941	0.95551116	0.02248294	1	0.41355903	0.11498959	

Tableau A6.2 Régression linéaire après ajustement du modèle 1

Variable	IA	IB	IC1	ID1	IE	IF1	IG	IH1	II1
tp	-0.38945763	4.1381163	2.6559357	-1.3962706	-5.4458434	2.4394579	-1.8375018	2.718471	10603.436
	0.5854	0.1653	0.1216	0.5074	0.1486	0.5288	0.1698	0.1757	0.4263
tjoylog	-0.08495337	-0.79705171	-0.25620655	0.3948073	-0.18595159	0.22154015	0.16516884		-1879.3331
	0.5856	0.0476	0.5487	0.5021	0.7243	0.7627	0.5678		0.3941
Volumetwee~g	0.08474683	0.85005535	-1.1249793	-0.14592986	-0.4681664	0.9586908	-0.42715303	-1.1629412	-862.29149
	0.4987	0.1649	0.0249	0.7484	0.4373	0.1943	0.0496	0.378	0.7263
followeeslog	0.33030655	-82.226575			7.6421512		32.76729		
	0.8996	0.0025			0.0287		0.44		
ctnlog	0.20492345	-0.70914766		-0.33608229	-0.59209566		-0.8793302	-0.04498009	2265.6508
	0.6815	0.4023		0.6451	0.6979		0.1579	0.9747	0.5787
ctneutrelog	0.29145459	0.46309084	0.09988953	-1.3577231	-0.58543674	0.23249867	-0.30928451	0.83495846	-2191.7536
	0.2833	0.4057	0.8201	0.0848	0.5539	0.8535	0.3339	0.4225	0.4831
_cons	19.149151	492.10342	51.203536	85.024692	12.656388	117.96297	-187.60219	41.112942	144887.77

Tableau A6.2 Régression linéaire après ajustement du modèle 1 (suite)

	0.3613	0.0002	0	0	0.624	0	0.5651	0	0
N	91	92	43	68	92	114	115	46	205
r2_o	0.09554802	0.04634423	0.89537764	0.00736661	0.08608261	0.63871658	0.09790209	0.31022397	0.0871057
r2_b	0.11728406	0.04683234	1	0.92852604	0.09259532	0.66646096	0.09858968	1	0.13311992
Variable	IJ1	IK1	IL1	IM1	IN	IO	IP1	IQ1	IR
tp	-0.63129934	-1.7875014	1.1355137	4.5438131	-1.3164064	-1.5856709	-5.1041342	1852.2031	0.86194463
	0.3632	0.1643	0.5554	0.2629	0.3097	0.4859	0.0112	0.7146	0.436
tjoylog	-0.18575083	0.97158223	0.0023646	0.00285233	-0.21882938	0.51984937	-0.11758597	-162.75659	-0.45323275
	0.1077	0.0021	0.9942	0.9968	0.4329	0.2377	0.6742	0.8744	0.0104
Volumetwee~g	-0.03682558	0.19513572	0.30697756	0.87774199	-0.63275315	1.1866418	0.89728281	-984.52407	0.36400705
	0.7506	0.6554	0.1714	0.174	0.0469	0.2485	0.0521	0.2347	0.1255
ctneutreglog	-0.15956886	-0.57559181	-0.23612445	0.25439443	0.1631653	0.14087284	-1.439128	266.06381	0.28529372
	0.4692	0.1325	0.5906	0.7755	0.6312	0.8927	0.0561	0.8218	0.3223
ctnlog		0.43751361	0.18809914	1.2022773	-0.36216596	-0.49482978	-0.43655495		0.29072096
		0.2586	0.7385	0.3123	0.4474	0.6851	0.6025		0.5549
followeeslog					11.437434	44.884866			14.358839
					0.8231	0.0534			0.1534

Tableau A6.2 Régression linéaire après ajustement du modèle 1 (suite)

_cons	21.30119	74.982339	52.384747	71.915913	-29.467875	-156.51002	53.01853	97482.45	-28.685857
	0	0	0	0	0.947	0.1811	0	0	0.5896
N	65	46	92	46	92	22	65	46	226
r2_o	0.14080175	0.46000896	0.30687303	0.84632095	0.03341381	0.38319469	0.09378023	0.17346885	0.01186435
r2_b	0.60179772	1	0.99237201	1	0.03432199	.	0.24785523	1	0.01158114
Variable	IS1	IT	IU	IV1	IV2	IW	IX	IY	IZ
tp	-18.672205	4.9077929	0.8515793		-3.8298227	0.5063071	-0.26077097	0.57555838	-7.2902974
	0.183	0	0.7502		0.005	0.6749	0.926	0.7173	0.0005
tjoylog	3.1876995	-0.19773572	-0.39272344	-0.49700953	-0.2102636	-0.14973702	-0.38364972	0.57678588	-0.81692374
	0.326	0.1727	0.3522	0.065	0.4391	0.5211	0.271	0.1655	0.0488
Volumetwee~g	-7.6381824	-0.17386302	-1.6005723	-0.0156762		-0.25957639	-1.1556603	0.0824993	0.24807871
	0.0735	0.5207	0.0109	0.9387		0.4201	0.0637	0.8239	0.3357
ctnlog	3.285396	-0.02517414	0.3384013	-0.21209092	-1.0946438	0.66865892	-2.0002298	-0.71238814	-1.4297189
	0.6667	0.9486	0.7133	0.7076	0.0798	0.0995	0.1735	0.4751	0.031
ctneutrellog	-1.3593525	0.35407245	0.12657817	-0.07597305	-0.52360686	-0.00016493	-0.12400437	0.35488533	-0.27262583
	0.7558	0.3401	0.8398	0.8528	0.2137	0.9996	0.8461	0.6183	0.5966
followeeslog		-4.8841883	-25.249908			-6.9568816	-48.840742	4.1540559	29.739233
		0.6336	0.2596			0.2584	0.0021	0.0002	0.0001

Tableau A6.2 Régression linéaire après ajustement du modèle 1 (suite)

_cons	213.35836	58.177901	295.74922	48.768162	52.016505	109.39677	449.82748	39.747972	-184.28187
	0	0.2604	0.1152	0	0	0.0088	0.0001	0	0.001
N	69	65	160	92	92	102	45	69	92
r2_o	0.15888342	0.00203486	0.03639073	0.51096564	0.10178142	0.1287973	0.94693827	0.38806867	0.10955363
r2_b	0.3210076	0.00040545	0.0366941	0.92284259	0.1235199	0.02248294	1	0.41355903	0.11498959

Tableau A6.3 Régression linéaire avant ajustement du modèle 2

Variable	JA	JB	JC	JD	JE	JF	JG	JH	JI
tn	0.52471853	-0.60653512	-5.715333	1.9366276	3.4555408	0.15179259	-0.92007134	0.3011757	-9724.6206
	0.4168	0.8802	0.032	0.3514	0.3202	0.9788	0.4127	0.7886	0.5703
tjoylog	-0.03324711	-0.98848179	0.10110156	-0.0250081	0.20792122	0.56572119	-0.29754984	-0.00454853	-1697.3239
	0.822	0.0146	0.7775	0.9638	0.6751	0.4659	0.203	0.9839	0.4323
followerslog	5.1297375	-28.131443	-5.2052631	-9.2803346	-33.601838	5.772123	37.992163	3.6192527	5522.0058
	0.3347	0.0162	0.529	0.5732	0.0041	0.2558	0	0.3392	0.8961
followeeslog	-4.8985187	8.9164665	21.988867	-10.899883	13.059325	-23.401153	-49.416842	16.187713	90403.01
	0.4084	0.8438	0.0001	0.0015	0.0005	0.3615	0.1536	0.1661	0.0852
ctnlog	0.12226861	-1.1273263	1.5264582	-0.49524503	0.34356626	-0.62033596	-0.13001082	-0.72049132	2447.5129
	0.7985	0.2127	0.0411	0.474	0.8071	0.746	0.7862	0.3579	0.5971

Tableau A6.3 Régression linéaire avant ajustement du modèle 2 (suite)

ctneutreglog	0.28698373	-0.03970974	-0.26341358	-0.95120399	0.21551989	0.26457177	-0.11029627	0.07990209	-2794.7143
	0.259	0.9388	0.498	0.1816	0.7847	0.836	0.6228	0.8641	0.3596
_cons	-3.011886	430.07573	-69.573229	239.33066	402.08529	197.84775	43.81449	-83.085543	-631257.23
	0.9245	0.0013	0.3896	0.1518	0.0046	0.1105	0.8638	0.0463	0.2219
N	91	92	43	68	92	114	115	45	205
r2_o	0.95342642	0.29562029	0.93319096	0.85235814	0.41275467	0.52218358	0.02138584	0.96288393	0.5339539
r2_b	0.95951298	0.29635931	1	0.88579239	0.41631824	0.51103974	0.02189806	1	0.54464412
Variable	JJ	JK	JL	JM	JN	JO	JP	JQ	JR
tn	1.2664771	1.7753495	-1.8518148	-0.56967651	0.81985965	4.8518719	9.4075433	3836.847	-1.0611956
	0.1119	0.1774	0.3638	0.8845	0.5975	0.0144	0.0006	0.6052	0.3652
tjoylog	-0.12624074	0.80882609	0.29841327	-0.54219802	-0.39607901	0.62731939	0.04658228	-232.8716	-0.44777678
	0.1662	0.0034	0.3158	0.4451	0.1259	0.1228	0.8491	0.7951	0.0107
followerslog	13.181787	-108.54697	-9.0405361	142.92531	23.244753	-20.162301	-3.554725	226127.39	-11.811547
	0.0002	0.0001	0.0001	0.0027	0	0.5897	0.3562	0	0.0007
followeeslog	2.8185441	138.61963	23.068779	-105.9038	-10.117857	49.156593	-2.4488104	-485663.16	26.128961
	0.9621	0.0011	0	0.0252	0.8292	0.3871	0.4743	0.0002	0.0135
ctnlog	-0.15820781	0.69260535	0.5023379	1.0221478	-0.22083123	-0.94654954	-1.2876105	-1948.8271	0.44473541

Tableau A6.3 Régression linéaire avant ajustement du modèle 2 (suite)

	0.5418	0.0447	0.3463	0.3638	0.6633	0.3869	0.1281	0.4037	0.389
ctneutrelog	-0.02122018	0.08812647	-0.24385642	0.20569052	0.31227554	0.64776247	-1.0437387	18.173017	0.19153544
	0.9067	0.7757	0.4925	0.7894	0.2806	0.4824	0.107	0.9868	0.4649
_cons	-109.18827	1017.0431	-10.859039	-969.11393	-151.74395	8.7060242	98.301477	-246803.54	37.817904
	0.739	0.0026	0.7852	0.0163	0.7098	0.9355	0.0015	0.2418	0.4649
N	64	46	92	46	92	22	65	46	226
r2_o	0.92319157	0.96671132	0.1474722	0.98721659	0.05996913	0.50207785	0.46131514	0.99908233	0.00016524
r2_b	0.92296011	1	0.14776407	1	0.06033736	.	0.52196399	1	0.00014972
Variable	JS	JT	JU	JV	JW	JX	JY	JZ	
tn	3.8352297	-4.2103687	-3.1327015	-2.0874766	-3.5859158	2.0686229	0.71140321	8.1130109	
	0.7684	0	0.3115	0.0794	0.0075	0.545	0.6034	0.0001	
tjoylog	0.41210374	-0.23288622	-0.33960823	-0.06387581	-0.15788578	-0.29678149	0.26438551	-0.75438307	
	0.8923	0.0761	0.4279	0.7598	0.4305	0.4222	0.4713	0.0402	
followerslog	190.78087	12.668719	48.151223	30.293437	-10.224502	41.802907	31.527926	-19.097905	
	0	0.0004	0.0015	0	0.0361	0.099	0.0006	0.0006	
followeeslog	-10.08679	-52.609627	-37.450489	-19.917734	-0.80855225	-51.514422	2.8216879	49.006496	
	0.8278	0.0017	0.0961	0.0003	0.9019	0.0016	0.0077	0	

Tableau A6.3 Régression linéaire avant ajustement du modèle 2 (suite)

ctnlog	0.3478882	0.27864992	1.0158645	0.46869576	1.3094831	-2.2980448	-1.0837977	-1.8078467
	0.9588	0.4349	0.2986	0.36	0.0031	0.1037	0.2227	0.0078
ctneutrelog	0.13354168	0.10175793	-0.01208773	-0.18332784	-0.1288073	0.12788022	0.29239201	0.53869516
	0.9719	0.7387	0.9833	0.5463	0.6721	0.8263	0.5999	0.2412
_cons	-1493.542	187.8521	-206.861	-171.68458	175.97199	1.9176166	-247.8139	-112.44
	0.0007	0.002	0.3429	0.0011	0.0006	0.9939	0.0028	0.0418
N	69	65	160	92	102	45	69	92
r2_o	0.44281758	0.00216749	0.00108873	0.34686345	0.82826775	0.96317956	0.00942077	0.01580128
r2_b	0.44501772	0.00561447	0.00125439	0.34688966	0.84078038	1	0.01071523	0.01675685

Tableau A6.4 Régression linéaire après ajustement du modèle 2

Variable	JA1	JA2	JB	JC1	JD	JE1	JE2	JF1	JG1
tjoylog	-0.03189813	-0.04679188	-0.98848179	-0.03339723	-0.0250081	0.04270413	0.08776912	0.47045807	-0.33239197
	0.8183	0.7345	0.0146	0.9397	0.9638	0.9361	0.8683	0.5343	0.1552
followerslog	1.0409153		-28.131443	17.200976	-9.2803346	-13.672705	-14.942189	2.9783939	36.140592
	0.6545		0.0162	0.0424	0.5732	0.2127	0.1719	0.4577	0
ctnlog	0.40333709	0.11912875	-1.1273263		-0.49524503	0.73565205	-0.46918434		-0.16102648

Tableau A6.4 Régression linéaire après ajustement du modèle 2 (suite)

	0.2643	0.8011	0.2127		0.474	0.5578	0.7523		0.7379
ctneutreslog	0.33470582	0.30468897	-0.03970974	-0.09477259	-0.95120399	0.01869152	0.13903067	0.11161759	-0.10955671
	0.1826	0.2267	0.9388	0.8318	0.1816	0.9824	0.8691	0.929	0.6268
tn		0.59986004	-0.60653512	-1.5734903	1.9366276		5.4359559	-0.92443199	-0.84226005
		0.3417	0.8802	0.45	0.3514		0.1404	0.8201	0.455
followeeslog			8.9164665		-10.899883				
			0.8438		0.0015				
_cons	9.206496	22.174398	430.07573	-158.72032	239.33066	232.38139	247.88648	93.580187	-316.87047
	0.7515	0	0.0013	0.1066	0.1518	0.1029	0.081	0.0493	0
N	91	91	92	43	68	92	92	114	115
r2_o	0.47728645	0.22911556	0.29562029	0.92473422	0.85235814	0.43168881	0.447791	0.3207116	0.1414964
r2_b	0.4872194	0.56923282	0.29635931	1	0.88579239	0.43610342	0.45217389	0.32196595	0.14193914
Variable	JH1	J11	JJ1	JK1	JL	JM1	JN	JO1	JP1
tn	-0.23624693	-4354.7408	1.2686902	0.6739258	-1.8518148	1.3715547	0.81985965	2.4498366	10.994948
	0.8141	0.7537	0.1075	0.6393	0.3638	0.6844	0.5975	0.06	0
followerslog	8.8972655	33223.686	13.300244	-56.912376	-9.0405361	57.644473	23.244753	15.20698	-2.6627836
	0	0.3991	0	0.0199	0.0001	0.0335	0	0.1672	0.2755

Tableau A6.4 Régression linéaire après ajustement du modèle 2 (suite)

ctnlog	-0.58645216		-0.15964311	0.42354615	0.5023379		-0.22083123		-1.5578492
	0.3727		0.5316	0.2587	0.3463		0.6633		0.0481
ctneutrelog	0.06365214	-2724.9189	-0.02082925	-0.31244899	-0.24385642	-0.16083122	0.31227554	1.2968297	-0.9871378
	0.8895	0.3604	0.9075	0.342	0.4925	0.8393	0.2806	0.1314	0.1124
tjoylog		-1659.6865	-0.12602951	0.82645358	0.29841327	0.0389778	-0.39607901		
		0.4437	0.1625	0.0077	0.3158	0.9542	0.1259		
followeeslog					23.068779		-10.117857		
					0		0.8292		
_cons	-27.326302	-242718.03	-93.65119	946.47309	-10.859039	-786.00798	-151.74395	-64.901362	80.857253
	0	0.6009	0	0.0119	0.7852	0.0537	0.7098	0.5038	0.0008
N	46	205	64	46	92	46	92	22	69
r2_o	0.96614188	0.17459351	0.93650968	0.97750905	0.1474722	0.9832967	0.05996913	0.35156729	0.15306199
r2_b	1	0.17604621	0.93825429	1	0.14776407	1	0.06033736	.	0.24948773
Variable	JQ1	JR	JS1	JS2	JT	JU	JV1	JV2	JW1
tn	-2129.0059	-1.0611956		22.296169	-4.2103687	-3.1327015		2.8723458	-3.5961595
	0.6961	0.3652		0.1232	0	0.3115		0.0322	0.0069

Tableau A6.4 Régression linéaire après ajustement du modèle 2 (suite)

tn	0.7557964		0.71140321	8.1130109
	0.8419		0.6034	0.0001
tjoylog	-0.39734989		0.26438551	-0.75438307
	0.3245		0.4713	0.0402
ctnlog	-2.2221869	-1.5062357	-1.0837977	-1.8078467
	0.1561	0.2325	0.2227	0.0078
ctneutrelog	-0.03972734	-0.05470186	0.29239201	0.53869516
	0.9512	0.9304	0.5999	0.2412
followerslog		17.576014	31.527926	-19.097905
		0.4754	0.0006	0.0006
followeeslog			2.8216879	49.006496
			0.0077	0
_cons	99.980385	-92.546904	-247.8139	-112.44
	0	0.7314	0.0028	0.0418
N	45	46	69	92
r2_o	0.57144498	0.98300949	0.00942077	0.01580128
r2_b	1	1	0.01071523	0.01675685

ANNEXE 7 – RÉGRESSION LINÉAIRE PAR NIVEAU DE PRIX

Tableau A7.1 Régression linéaire par niveau de prix de l'action, modèle 1

Variable ind.	N1	N2	N3	N4
tp	-0.00056662	0.85752268	-12.12104	22694.346
	0.9986	0.1228	0.09	0.4177
tjoylog	-0.10887766	-0.00656736	-0.24473458	-3733.2106
	0.0562	0.9479	0.8482	0.4713
Volumetwee~g	0.06417299	-0.19135135	-1.7081437	-4597.9472
	0.2759	0.1352	0.2721	0.5671
ctnlog	-0.01285909	0.19226087	-1.1907916	4039.8209
	0.9218	0.3616	0.6492	0.6513
ctneutrelog	-0.01918376	-0.01097853	0.38149792	-5459.4988
	0.8285	0.9426	0.8301	0.4816
_cons	29.641855	71.739782	190.81987	402830.64
	0	0	0	0
N	1027	970	192	91
r2_o	8.25E-06	0.02129462	0.23429324	0.13173972
r2_b	0.00013168	0.00249346	0.2667434	0.15317842

Tableau A7.2 Régression linéaire par niveau de prix de l'action avec un lag d'une journée, modèle 1

Variable ind.	N11	N21	N31	N41
Tp L1.	0.05070674	0.64487016	-3.0843824	-4154.0789
	0.9023	0.3287	0.7228	0.932
Tjoylog L1.	-0.09616486	-0.18642906	-1.6369407	-10971.067
	0.1887	0.1176	0.3446	0.229
Volumetwee~g L1.	0.0005653	-0.17537507	-2.8474297	-3130.6074
	0.9938	0.2272	0.1351	0.7729
Ctnlog L1.	-0.00506739	0.29567056	-0.31910543	-3499.7244
	0.9764	0.2217	0.9246	0.8361
Ctneutrelog L1.	-0.04576825	0.20201487	0.46577097	-18955.238
	0.6925	0.2819	0.8435	0.2274
_cons	30.173803	71.224294	189.8408	385312.59
	0	0	0	0.0019
N	616	594	118	54
r2_o	0.00314183	0.01919076	0.19764749	0.0796831
r2_b	0.02995821	0.00859139	0.22683621	0.14488548

Tableau A7.3 Régression linéaire avec interaction entre le niveau de prix de l'action et le volume de tweet, modèle 1

Variable ind.	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
tp	663.8244	1005.973	0.66	0.509	-1307.845	2635.494
tjoylog	-118.9935	183.4707	-0.65	0.517	-478.5895	240.6025
Volumetweetlog	13.02211	288.2017	0.05	0.964	-551.8428	577.887
ctnlog	115.1887	395	0.29	0.771	-658.9971	889.3745
ctneutrelog	-175.8604	278.6049	-0.63	0.528	-721.916	370.1953
level2	179.71	3525.225	0.05	0.959	-6729.604	7089.024
level3	519.674	6540.669	0.08	0.937	-12299.8	13339.15
level4	414409.9	53662.57	7.72	0	309233.2	519586.6
niveau#c.Volumetweetlog						
2	-22.4869	425.1769	-0.05	0.958	-855.8183	810.8445
3	-53.00888	792.4839	-0.07	0.947	-1606.249	1500.231
4	-2286.347	1565.92	-1.46	0.144	-5355.494	782.8004
_cons	-683.6891	10615.69	-0.06	0.949	-21490.06	20122.68
sigma_u	100917.63					
sigma_e	4512.9695					
rho	0.99800417	(fraction	of variance due	to	u_i)	

Tableau A7.4 Régression linéaire par niveau de prix de l'action, modèle 2

Variable ind.	N21	N22	N32	N42
tn	0.25742107	-1.7185308	10.098846	-27264.53
	0.4111	0.0078	0.2024	0.4529
tjoylog	-0.11519496	-0.01816167	-0.13614794	-3562.5797
	0.0377	0.8564	0.906	0.4866
followerslog	5.0934826	-0.4227104	103.01504	123900.07
	0	0.8235	0	0.2392
ctnlog	-0.07560631	0.36881881	-1.2245671	5836.5387
	0.5623	0.1025	0.6368	0.5633
ctneutrelog	-0.00904692	-0.10360283	1.6303451	-8142.3807
	0.9128	0.4717	0.302	0.2927
_cons	-26.015864	75.809773	-997.78264	-1079444.2
	0.0001	0.0003	0	0.3839
N	1027	970	192	91
r2_o	1.46E-03	0.02079057	0.00418063	0.22544086
r2_b	0.00007841	0.00872678	0.03407831	0.23410838

Tableau A7.5 Régression linéaire par niveau du prix de l'action avec un lag d'une journée, modèle 2

Variable	N211	N221	N321	N421
tn L1.	0.05708319	-1.5159013	-0.93170126	13305.426
	0.8879	0.059	0.9205	0.8585
Tjoylog L1.	-0.07909788	-0.19324593	0.43925018	-10464.71
	0.2716	0.1056	0.7682	0.2359
Followerslog L1.	4.4267959	-2.4305233	123.76444	91307.732
	0	0.2874	0	0.5073
Ctnlog L1.	-0.03618526	0.45428565	1.4592166	-5179.2436
	0.8314	0.0791	0.6546	0.7938
Ctneutrelog L1.	-0.05817102	0.11875204	1.0162432	-15592.184
	0.5977	0.5022	0.6106	0.2343
_cons	-18.588929	97.507002	-1225.8524	-722306.05
	0.0253	0.0001	0.0001	0.6546
N	616	594	118	54
r2_o	0.00223882	0.02245605	0.00543506	0.20460889
r2_b	0.00005103	0.01044657	0.02250508	0.23360444

Tableau A7.6 Régression linéaire avec interaction entre le niveau de prix de l'action et le nombre de followers, modèle 2

close	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
tn	-574.7877	1104.041	-0.52	0.603	-2738.668	1589.093
tjoylog	-126.916	181.6054	-0.7	0.485	-482.856	229.024
followerslog	56.98494	2211.533	0.03	0.979	-4277.539	4391.509
ctnlog	120.1847	412.5691	0.29	0.771	-688.4358	928.8052
ctneutrelog	-226.3634	263.3799	-0.86	0.39	-742.5785	289.8516
level2	133.5833	7753.783	0.02	0.986	-15063.55	15330.72
level3	646.8097	27160.54	0.02	0.981	-52586.87	53880.49
level4	-1035512	193858.4	-5.34	0	-1415468	-655556.9
niveau#c.followerslog						
2	-13.58453	742.2885	-0.02	0.985	-1468.443	1441.274
3	-60.80387	2934.736	-0.02	0.983	-5812.781	5691.174
4	120830	15841.2	7.63	0	89781.84	151878.2
_cons	-606.5178	26433.81	-0.02	0.982	-52415.83	51202.79
sigma_u	99764.747					
sigma_e	4483.652					
rho	0.99798427	(fraction	of variance due	to	u_i)	

Tableau A7.7 Résumé de la significativité des variables indépendantes en dessous d'un seuil $p < 0.20$ (modèle 1 et 2 sans lag)

Sans Lag	Bas Prix (0-50)	Prix Moyen (50-100)	Prix Élevé (100-1000)	Prix très élevé (1000+)
tp		✓	✓	
tn		✓	✓	
tjoylog	✓			
Volumetweetlog		✓		
Followerslog	✓		✓	
Ctnlog		✓		
Ctneutrelog				

Tableau A7.8 Résumé de la significativité des variables indépendantes en dessous d'un seuil $p < 0.20$ (modèle 1 et 2 avec lag)

Avec Lag 1 jour	Bas Prix (0-50)	Prix Moyen (50-100)	Prix Élevé (100-1000)	Prix très élevé (1000+)
tp				
tn		✓		
tjoylog	✓	✓		
Volumetweetlog			✓	
Followerslog	✓		✓	
Ctnlog		✓		
Ctneutrelog				

ANNEXE 8 – RÉSUMÉ DES DONNÉES RÉCOLTÉES AU TRAVERS DES SOURCES DE MESURE DE RÉPUTATION

Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie

No	Industries	Nombre de compagnies recensées
1	Aerospace and Defense	15
2	Airlines	22
3	Apparel	48
4	Automotive Retailing, Services	8
5	Beverages	43
6	Building Materials, Glass	4
7	Chemicals	22
8	Computer Peripherals	3
9	Computer Software	14
10	Computers	19
11	Consumer Credit Card and Related Services	9
12	Consumer Food Products	11
13	Delivery	10

Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie (suite)

14	Diversified Outsourcing Services	20
15	Electric and Gas Utilities	36
16	Electronics	35
17	Energy	34
18	Engineering, Construction	32
19	Entertainment	41
20	Financial Data Services	11
21	Food and Drug Stores	23
22	Food Production	32
23	Food Services	16
24	Forest and Paper Products	14
25	Furniture	1
26	General Merchandisers	19
27	Health Care: Insurance and Managed Care	10
28	Health Care: Medical Facilities	11
29	Health Care: Pharmacy and Other Services	13

Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie (suite)

30	Home Equipment and Furnishings	19
31	Homebuilders	8
32	Hotels, Casinos, Resorts	12
33	Household and Personal Products	5
34	Industrial and Farm Equipment	14
35	Information Technology Services	19
36	Insurance: Life and Health	33
37	Insurance: Property and Casualty	28
38	Internet Services and Retailing	18
39	Medical Equipment	12
40	Megabanks	59
41	Metals	24
42	Mining, Crude-Oil Production	39
43	Mortgage Services	10
44	Motor Vehicle Parts	19
45	Motor Vehicles	41
46	Network and Other Communications Equipment	12

Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie (suite)

47	Oil and Gas Equipment, Services	10
48	Packaging, Containers	12
49	Petroleum Refining	27
50	Pharmaceuticals	24
51	Pipelines	8
52	Publishing: Newspapers and Magazines	9
53	Railroads	6
54	Real Estate	19
55	Scientific, Photographic and Control Equipment	2
56	Securities	12
57	Semiconductors	15
58	Soaps and Cosmetics	22
59	Specialty Retailers: Apparel	13
60	Specialty Retailers: Diversified	14
61	Superregional Banks	61
62	Telecommunications	84

Tableau A8.1 Description de l'échantillon préliminaire d'entreprises recensées par industrie (suite)

63	Temporary Help	6
64	Tobacco	15
65	Trading	17
66	Transportation and Logistics	4
67	Trucking, Transportation, Logistics	23
68	Wholesalers: Diversified	18
69	Wholesalers: Electronics and Office Equipment	10
70	Wholesalers: Food and Grocery	7
71	Wholesalers: Health Care	7

Tableau A8.2 Industries présentes dans l'échantillon

Industries	Nombre de compagnies
Airlines	4
Apparel	4
Beverages	2
Chemicals	3
Computer Software	4

Tableau A8.2 Industries présentes dans l'échantillon (suite)

Computers	5
Consumer Food Products	5
Electric and Gas Utilities	2
Electronics	9
Energy	3
Entertainment	2
Food and Drug Stores	4
Food Services	2
General Merchandisers	4
Health Care: Insurance and Managed Care	1
Insurance: Life and Health	3
Insurance: Property and Casualty	2
Megabanks	10
Metals	3
Mining, Crude-Oil Production	4
Motor Vehicles	7
Petroleum Refining	4
Pharmaceuticals	5

Tableau A8.2 Industries présentes dans l'échantillon (suite)

Soaps and Cosmetics	2
Superregional Banks	3
Telecommunications	4

Tableau A8.3 Liste des 101 firmes composant l'échantillon

ID	Compagnie	Ticker (Yahoo! Finance)	Industrie (Fortune & Money)
1	Abbott Laboratories	ABT	Pharmaceuticals
2	Adidas	ADS.DE	Apparel
3	Adobe Systems	ADBE	Computer Software
4	Aetna	AET	Health Care : Insurance and Managed Care
5	Air France-KLM Group	AF	Airlines
6	Allianz	ALV	Insurance : Property and Casualty
7	Apple	AAPL	Computers
8	ArcelorMittal	MT	Metals
9	AT&T	T	Telecommunications
10	Audi	NSU	Motor Vehicles
11	AXA	CS	Insurance : Life and Health

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

12	Bank of America	BAC	Megabanks
13	Barclays	BCS	Megabanks
14	BASF	BAS.DE	Chemicals
15	Bayer	BAYN.DE	Chemicals
16	Berkshire Hathaway	BRK-A	Insurance: Property and Casualty
17	BG Group	BG	Mining, Crude-Oil Production
18	BHP Billiton	BHP	Mining, Crude-Oil Production
19	BMW	BMW.DE	Motor Vehicles
20	BP	BP	Petroleum Refining
21	Canon	CAJ	Computers
22	Citi	C	Megabanks
23	Coca-Cola	KO	Beverages
24	Comcast	CMCSA	Telecommunications
25	Credit Suisse	CS	Megabanks
26	CVS	CVS	Food and Drug Stores
27	Daimler	DAI.DE	Motor Vehicles

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

28	Danone	BN.PA	Consumer Food Products
29	Delta Air Lines	DAL	Airlines
30	Deutsche Bank	DB	Megabanks
31	DuPont	DD	Chemicals
32	E.ON	EOAN.DE	Energy
33	EMC	EMC	Computers
34	Emerson Electric	EMR	Electronics
35	EnCana	ECA	Mining, Crude-Oil Production
36	Exelon	EXC	Electric and Gas Utilities
37	Ford Motor	F	Motor Vehicles
38	GDF Suez	GSZ.PA	Energy
39	General Electric	GE	Electronics
40	GlaxoSmithKline	GSK	Pharmaceuticals
41	Goldman Sachs Group	GS	Megabanks
42	Heineken	HEIA.AS	Beverages
43	Hewlett-Packard	HPQ	Computers
44	Honda Motor	HMC	Motor Vehicles
45	HSBC Holdings	HSBC	Megabanks

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

46	J.P. Morgan Chase	JPM	Megabanks
47	Johnson & Johnson	JNJ	Pharmaceuticals
48	Kellogg	K	Consumer Food Products
49	Kohl's	KSS	General Merchandisers
50	Kraft Foods	KRFT	Consumer Food Products
51	Kroger	KR	Food and Drug Stores
52	L'Oréal	OR.PA	Soaps and Cosmetics
53	LG Electronics	066570.KS	Electronics
54	Lufthansa Group	LHA.DE	Airlines
55	LVMH	MC.PA	Apparel
56	McDonald's	MCD	Food Services
57	Merck	MRK	Pharmaceuticals
58	MetLife	MET	Insurance: Life and Health
59	Microsoft	MSFT	Computer Software
60	Morgan Stanley	MS	Megabanks
61	Motorola	MSI	Electronics
62	Nike	NKE	Apparel
63	Nordstrom	JWN	General Merchandisers

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

64	Novartis	NVS	Pharmaceuticals
65	Oracle	ORCL	Computer Software
66	PepsiCo	PEP	Consumer Food Products
67	Petrobras	PBR	Petroleum Refining
68	PNC Financial Services Group	PNC	Superregional Banks
69	Polo Ralph Lauren	RL	Apparel
70	POSCO	PKX	Metals
71	Procter & Gamble	PG	Soaps and Cosmetics
72	Prudential Financial	PRU	Insurance: Life and Health
73	Royal Dutch Shell	RDS-A	Petroleum Refining
74	Royal Philips Electronics	PHG	Electronics
75	RWE	RWE.DE	Energy
76	Safeway	SWY	Food and Drug Stores
77	Samsung Electronics	005930.KS	Electronics
78	SAP	SAP	Computer Software
79	Siemens	SI	Electronics
80	Sony	SNE	Electronics
81	Southern	SO	Electric and Gas Utilities

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

82	Southwest Airlines	LUV	Airlines
83	Starbucks	SBUX	Food Services
84	State Street Corp.	STT	Superregional Banks
85	Target	TGT	General Merchandisers
86	Tata Steel	TATASTEEL.NS	Metals
87	Time Warner	TWX	Entertainment
88	Toshiba	TOSYY	Electronics
89	Total	TOT	Petroleum Refining
90	Toyota Motor	TM	Motor Vehicles
91	U.S. Bancorp	USB	Superregional Banks
92	Unilever	UL	Consumer Food Products
93	Vale	VALE	Mining, Crude-Oil Production
94	Verizon Communications	VZ	Telecommunications
95	Vodafone	VOD	Telecommunications
96	Volkswagen	VOW3.DE	Motor Vehicles
97	Wal-Mart Stores	WMT	General Merchandisers
98	Walgreens	WAG	Food and Drug Stores
99	Walt Disney	DIS	Entertainment

Tableau A8.3 Liste des 101 firmes composant l'échantillon (suite)

100	Wells Fargo	WFC	Megabanks
101	Xerox	XRX	Computers

ANNEXE 9 – RÉFÉRENCE DES COMPAGNIES SUR TWITTER

Tableau A9.1 Référence des compagnies sur Twitter

ID	Compagnie	Compte Twitter	Mot clé Twitter
1	Abbott Laboratories	@AbbottNews	Abbott
2	Adidas	@adidas	Adidas
3	Adobe Systems	@adobe	Adobe
4	Aetna	@aetna	Aetna
5	Air France-KLM Group	@airfrance	Air France
6	Allianz	@allianz	Allianz
7	Apple	@applenws	Apple
8	ArcelorMittal	@arcelormittal	ArcelorMittal
9	AT&T	@att	AT&T
10	Audi	@audi	Audi
11	AXA	@axa	AXA
12	Bank of America	@bankofamerica	Bank of America
13	Barclays	@barclays	Barclays
14	BASF	@basf	BASF
15	Bayer	@bayer	Bayer
16	Berkshire Hathaway	@brk_b	Berkshire

Tableau A9.1 Référence des compagnies sur Twitter (suite)

17	BG Group	@bggroup	BG Group
18	BHP Billiton	@nsftawards	BHP
19	BMW	@bmw	BMW
20	BP	@bp_america	BP
21	Canon	@canonusa	Canon
22	Citi	@citi	Citi
23	Coca-Cola	@cocacolaco	Coca
24	Comcast	@comcast	Comcast
25	Credit Suisse	@creditsuisse	Credit Suisse
26	CVS	@cvs_extra	CVS
27	Daimler	@daimler	Daimler
28	Danone	@danonegroup	Danone
29	Delta Air Lines	@delta	Delta
30	Deutsche Bank	@deutschebank	Deutsche Bank
31	DuPont	@dupont_news	DuPont
32	E.ON	@eonenergyuk	Eon
33	EMC	@emccorp	EMC
34	Emerson Electric	@emerson_news	Emerson Electric

Tableau A9.1 Référence des compagnies sur Twitter (suite)

35	EnCana	@encana	EnCana
36	Exelon	@exelon	Exelon
37	Ford Motor	@ford	Ford Motor
38	GDF Suez	@gdfsuez	GDF
39	General Electric	@generalelectric	GE
40	GlaxoSmithKline	@gsk	GlaxoSmithKline
41	Goldman Sachs Group	@goldmansachs	Goldman Sachs
42	Heineken	@heineken	Heineken
43	Hewlett-Packard	@hp	HP
44	Honda Motor	@honda	Honda
45	HSBC Holdings	@hsbc_uk_press	HSBC
46	J.P. Morgan Chase	@jpmorgan	JP Morgan
47	Johnson & Johnson	@jnjnews	Johnson & Johnson
48	Kellogg	@kelloggs_us	Kellogg
49	Kohl's	@kohls	Kohl
50	Kraft Foods	@kraftfoods	Kraft
51	Kroger	@kroger	Kroger
52	L'Oréal	@Loreal	Oreal

Tableau A9.1 Référence des compagnies sur Twitter (suite)

53	LG Electronics	@lgus	LG
54	Lufthansa Group	@lufthansa_usa	Lufthansa
55	LVMH	@lvmh	LVMH
56	McDonald's	@mcdonalds	McDonald
57	Merck	@merck	Merck
58	MetLife	@metlife	MetLife
59	Microsoft	@microsoft	Microsoft
60	Morgan Stanley	@morganstanley	Morgan Stanley
61	Motorola	@motorola	Motorola
62	Nike	@nike	Nike
63	Nordstrom	@nordstrom	Nordstrom
64	Novartis	@novartis	Novartis
65	Oracle	@oracle	Oracle
66	PepsiCo	@pepsico	PepsiCo
67	Petrobras	@petrobras	Petrobras
68	PNC Financial Services Group	@pncnews	PNC
69	Polo Ralph Lauren	@ralphlauren	Ralph Lauren
70	POSCO	@POSCO_E_n_C	POSCO

Tableau A9.1 Référence des compagnies sur Twitter (suite)

71	Procter & Gamble	@proctergamble	Procter & Gamble
72	Prudential Financial	@prudentialbyc	Prudential Financial
73	Royal Dutch Shell	@shell	Shell
74	Royal Philips Electronics	@philips	Philips
75	RWE	@rwe_group	RWE
76	Safeway	@safeway	Safeway
77	Samsung Electronics	@samsung	Samsung
78	SAP	@sap	SAP
79	Siemens	@siemensusa	Siemens
80	Sony	@sony	Sony
81	Southern	@southerncompany	Southern
82	Southwest Airlines	@southwestair	Southwest
83	Starbucks	@starbucks	Starbucks
84	State Street Corp.	@statestreet	State Street
85	Target	@target	Target
86	Tata Steel	@tatacompanies	Tata Steel
87	Time Warner	@time	Time Warner
88	Toshiba	@toshibausa	Toshiba

Tableau A9.1 Référence des compagnies sur Twitter (suite)

89	Total	@total	Total
90	Toyota Motor	@toyota	Toyota
91	U.S. Bancorp	@usbank	US Bancorp
92	Unilever	@unilever	Unilever
93	Vale	@valeglobal	Vale
94	Verizon Communications	@verizon	Verizon
95	Vodafone	@vodafoneuk	Vodafone
96	Volkswagen	@vw	Volkswagen
97	Wal-Mart Stores	@walmart	WalMart
98	Walgreens	@walgreens	Walgreens
99	Walt Disney	@disney	Walt Disney
100	Wells Fargo	@wellsfargo	Wells Fargo
101	Xerox	@xeroxcorp	Xerox

ANNEXE 10 – NOUVELLE MESURE DE LA RÉPUTATION PAR TWITTER ET INDUSTRIES

Tableau A10.1 Régression logistique avec interactions entre le modèle 2 de la réputation par Twitter et les industries

closebin	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Ind2						
B	-0,9005986	2,392467	-0,38	0,707	-5,589747	3,78855
C	1,727317	3,215322	0,54	0,591	-4,574599	8,029233
D	-0,9299554	2,574631	-0,36	0,718	-5,97614	4,116229
E	-4,055732	3,97496	-1,02	0,308	-11,84651	3,735047
F	-0,0882522	2,152865	-0,04	0,967	-4,30779	4,131285
G	0,8333208	2,622662	0,32	0,751	-4,307002	5,973644
H	0,5191596	2,103972	0,25	0,805	-3,60455	4,642869
I	0,099639	2,071645	0,05	0,962	-3,960711	4,159989
J	0,6599665	3,143503	0,21	0,834	-5,501185	6,821118
K	0,0007524	2,358127	0	1	-4,621092	4,622597
L	0,0134467	3,203853	0	0,997	-6,265989	6,292883
M	-2,615009	3,769255	-0,69	0,488	-10,00261	4,772596
N	0,1288727	2,24287	0,06	0,954	-4,267072	4,524818
O	1,601603	4,038	0,4	0,692	-6,312732	9,515937
P	-1,338791	3,277015	-0,41	0,683	-7,761623	5,084041
Q	-1,263959	3,998355	-0,32	0,752	-9,10059	6,572672
R	-1,931199	2,213193	-0,87	0,383	-6,268978	2,406581
S	-0,8036366	2,298291	-0,35	0,727	-5,308203	3,70093

Tableau A10.1 Régression logistique avec interactions entre le modèle 2 de la réputation par Twitter et les industries (suite)

T	0,5048871	2,056977	0,25	0,806	-3,526713	4,536488
U	-0,146564	2,646977	-0,06	0,956	-5,334544	5,041416
V	1,50918	2,361968	0,64	0,523	-3,120192	6,138552
W	-1,244281	2,102101	-0,59	0,554	-5,364323	2,875761
X	0,8556418	3,214596	0,27	0,79	-5,444851	7,156135
Y	-0,4357037	2,153538	-0,2	0,84	-4,656561	3,785154
Z	-0,9556822	2,406041	-0,4	0,691	-5,671437	3,760072
eRep2	0,1229852	1,319679	0,09	0,926	-2,463538	2,709508
Ind2#c.eRep2						
B	0,4913722	1,563201	0,31	0,753	-2,572446	3,555191
C	-1,202927	2,336141	-0,51	0,607	-5,781679	3,375826
D	0,4322818	1,680052	0,26	0,797	-2,86056	3,725124
E	2,990183	2,832153	1,06	0,291	-2,560734	8,5411
F	0,1276138	1,406897	0,09	0,928	-2,629853	2,88508
G	-0,3786203	1,788065	-0,21	0,832	-3,883164	3,125923
H	-0,0121872	1,331392	-0,01	0,993	-2,621668	2,597294
I	-0,0215961	1,342945	-0,02	0,987	-2,653719	2,610527
J	-0,5965764	2,272237	-0,26	0,793	-5,05008	3,856927
K	0,0383169	1,490768	0,03	0,979	-2,883534	2,960168
L	-0,0622298	2,12458	-0,03	0,977	-4,226331	4,101871
M	1,617736	2,473293	0,65	0,513	-3,229829	6,4653

Tableau A10.1 Régression logistique avec interactions entre le modèle 2 de la réputation par Twitter et les industries (suite)

N	-0,0315852	1,447472	-0,02	0,983	-2,868578	2,805408
O	-0,7874816	2,390467	-0,33	0,742	-5,47271	3,897747
P	1,216194	2,237052	0,54	0,587	-3,168347	5,600734
Q	1,07969	3,137588	0,34	0,731	-5,06987	7,22925
R	1,3318	1,45008	0,92	0,358	-1,510305	4,173905
S	0,7378677	1,445985	0,51	0,61	-2,096211	3,571946
T	-0,1388763	1,328294	-0,1	0,917	-2,742285	2,464532
U	0,0310395	1,845727	0,02	0,987	-3,586519	3,648598
V	-1,125099	1,655662	-0,68	0,497	-4,370136	2,119939
W	0,5175788	1,347552	0,38	0,701	-2,123575	3,158733
X	-0,5938081	1,921351	-0,31	0,757	-4,359588	3,171972
Y	0,3137365	1,377826	0,23	0,82	-2,386752	3,014225
Z	0,5740728	1,606256	0,36	0,721	-2,574131	3,722277
_cons	0,063376	2,02687	0,03	0,975	-3,909217	4,035969
/lnsig2u	-14,31848	20,03349			-53,58341	24,94644
sigma_u	0,0007776	0,0077895			2,31E-12	261246,9
rho	1,84E-07	3,68E-06			1,63E-24	1

**ANNEXE 11– VOLUME DE TWEETS MOYEN POUR LES 101
ENTREPRISES DE L'ÉCHANTILLON**

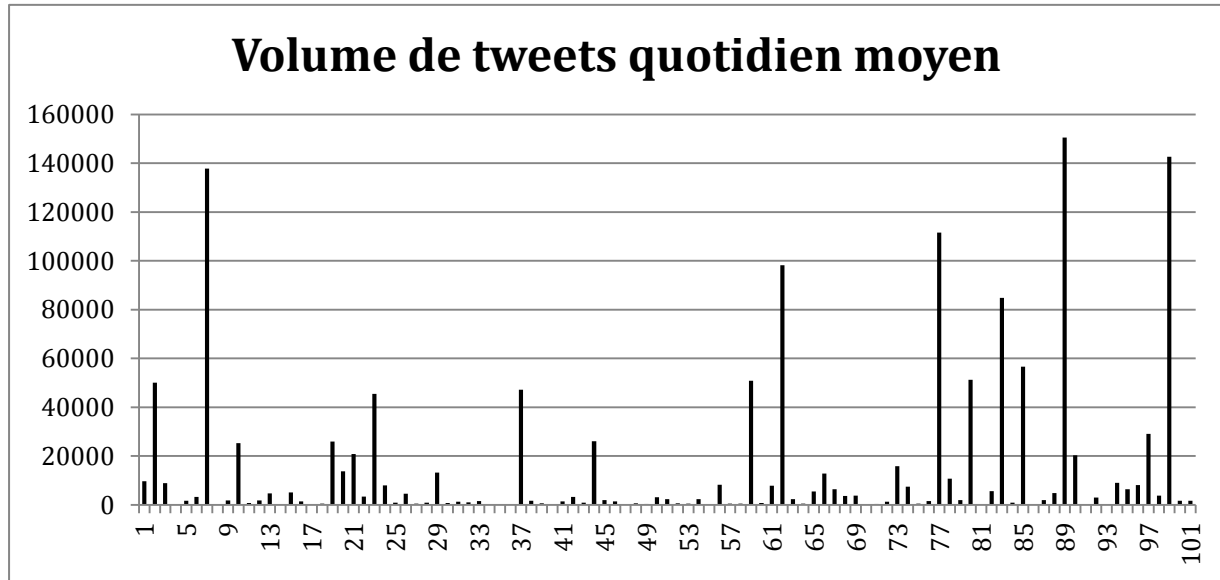


Figure A11.1 Volume de tweets quotidien moyen des 101 entreprises composant l'échantillon