

Utiliser Facebook pour prévoir l'évolution des marchés financiers

Vous avez aimé il y a quelques mois l'article "Utiliser Twitter pour prévoir l'évolution des marchés !" ? Et bien c'est reparti pour un tour, mais nous allons parler aujourd'hui de Facebook ! Le principe général est le même : utiliser les réseaux sociaux et la masse d'information incroyable publiée chaque jour sur le web pour calculer en temps réel un "indice de confiance", en fonction des statuts Facebook, tweets et autres informations que vous et moi postez chaque jour sur le net. L'idée est ensuite de comparer ces nouveaux indicateurs avec l'évolution des marchés financiers, pour voir s'il existe un lien de causalité entre le sentiment moyen de la population et l'évolution d'un indice financier (CAC40, Dow Jones, S&P...).

En supposant que l'humeur et le sentiment des investisseurs peuvent avoir un impact sur le prix des actifs financiers, on remet alors en cause l'hypothèse classique qui suppose que les marchés sont parfaitement rationnels. En effet, une décision rationnelle d'investissement ne doit être faite qu'en fonction des prévisions actualisées de cash-flows et du risque ; si votre humeur joue sur vos décisions, alors vous n'êtes plus parfaitement rationnel. Et pourtant, principalement en ce qui concerne les investisseurs individuels, mais aussi dans une moindre mesure pour les traders expérimentés, les études empiriques montrent qu'il existe un lien entre l'humeur des investisseurs et les décisions d'investissement. Si votre équipe de foot adorée a perdu un match la veille ou bien si votre bébé a pleuré toute la nuit, et bien vous aurez tendance à vendre vos actions le lendemain. Cela vous paraît peut-être un peu tiré par les cheveux, mais il existe de nombreuses études très sérieuses mêlant psychologie et finance comportementale, et les résultats sont très encourageants.

Pour reprendre une phrase issue d'un papier de référence dans le domaine, la question actuelle n'est plus de savoir, comme c'était le cas il y a quelques dizaines d'années, si l'humeur des investisseurs a un impact sur le prix des actifs financiers, mais plutôt de définir comment est-il possible de mesurer l'humeur des investisseurs et comment en quantifier les effets (source : "Investor Sentiment in the Stock Market" - Baker & Wurgler, Journal of Economic Perspectives, 2007)

"Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects" - Baker & Wurgler

Il y a une quinzaine d'années, mesurer le sentiment ou l'humeur d'une population était loin d'être évident. La méthode la plus classique était le sondage d'opinion, qui était (et est toujours) réalisé environ une fois par mois sur un panel de seulement quelques milliers d'individus. Le problème de cette méthode provient principalement de la disponibilité à faible fréquence des données (une fois par mois) et le délai entre le sondage et la publication du chiffre (temps de collecte des données + temps d'analyse des données). Le développement d'internet et plus particulièrement des réseaux sociaux permet à la fois d'avoir des données en très haute-fréquence (chaque jour, voire même chaque heure ou minute), basées sur un échantillon bien plus important et disponibles en temps réel.

"Mais Captain', les statuts Facebook sont privés, donc comment est-il possible d'utiliser cette information ?". Contrairement aux tweets qui sont accessibles publiquement et peuvent être extraits via une API (interface de programmation, même si cela demande tout de même un bon petit niveau en geekage), les statuts Facebook sont en effet privés et il est impossible pour un chercheur d'avoir accès aux statuts des utilisateurs. Impossible pour vous et moi, certes, mais pas impossible pour tout le

monde ! La "Facebook Data Team" a en effet développé en interne un algorithme de calcul de l'humeur des utilisateurs de Facebook, en analysant chaque jour des dizaines de millions de statuts en attribuant une note à chaque statut en fonction du nombre de mots à caractère positif ("happy", "good" ...) et du nombre de mots à caractère négatif présents dans chaque statut Facebook (source : "An Unobtrusive Behavioral Model of "Gross National Happiness"). Vous pouvez vous amuser à voir l'émotion contenue dans une phrase directement sur le site du "Linguistic Inquiry and Word Count", qui est le logiciel utilisé par la Facebook Data Team (source : "Testing LIWC Online"). Voici par exemple les résultats pour la phrase "I am happy. I love Captain Economics" ; une phrase pleine d'émotion positive !

Officiellement, tout est anonymisé et automatisé, de telle sorte qu'aucun chercheur ne puisse avoir accès à vos statuts. Officiellement ...

"This approach to modeling GNH is based on Facebook status updates, of which there are over 40 million posted per day. I then narrowed the focus to users from United States who had selected English as the language in which they preferred to view the website; all such users' updates were anonymously analyzed beginning September 7, 2007. Status updates are short-format (m=9 words per update) notes that are broadcast to some or all of the user's friends." [...] "To protect users' privacy, data was piped directly from Hive into TAWC, so that no update was ever seen by a researcher during analysis."

Tout cela permet à Facebook de calculer en interne un indice journalier de "Bonheur National Brut" ("Gross National Happiness"). Cela donne quelque chose de ce style là, comme expliqué dans un article par un chercheur de la Team Facebook ("How Happy Are We?" par Adam Kramer qui est l'auteur du papier précédemment cité).

Bon bref, quel est le lien entre cet indicateur et les marchés financiers ? Dans un papier de 2011 intitulé "Can Facebook Predict Stock Market Activity?", un chercheur a étudié les liens entre cette variable "Facebook", le volume d'activité sur les marchés et le rendement journalier du marché action sur une période allant de septembre 2007 à septembre 2011. Pour les adeptes de l'économétrie, le modèle utilisé est un modèle VAR (Vector AutoRegressive) tout bête, avec "z" le vecteur avec nos 3 variables (sentiment, volume et rendement) et "x" des variables de contrôle (dummy pour le jour de la semaine, autres indicateurs d'humeur...).

En analysant plus de 1000 jours de trading, l'auteur montre que la variable sentiment "Facebook" retardée d'une période a un impact significatif sur le rendement du marché action à une date t. Pour le dire autrement, lorsque les investisseurs sont de bonne humeur un jour, alors le marché action a tendance à augmenter le lendemain (et inversement). Une hausse d'une standard-deviation de l'indice GNH prévoit une hausse de 11 points de base de l'indice boursier le lendemain.

De plus, en faisant une distinction entre les petites entreprises et les grosses entreprises, l'auteur montre que cette relation est plus forte en ce qui concerne les petites entreprises, ce qui est en accord avec la théorie expliquant que pour les petites capitalisations, il est plus difficile et plus risqué pour un "investisseur sophistiqué" d'aller trader contre les "investisseurs irrationnels" car le pourcentage d'investisseur individuel est plus élevé et que le mouvement irrationnel peut s'auto-alimenter pendant une longue période (lire "Il y a des idiots. Regardez autour de vous !" - Divergence de la valeur d'un actif par rapport aux fondamentaux" sur les travaux de Brad DeLong et al. (1990)).

Conclusion : Le développement d'internet et l'accès au "big-data" permettent aux chercheurs de mieux

comprendre les comportements des individus en relâchant l'hypothèse de parfaite rationalité. Comme expliqué par G. King dans "Ensuring the Data-Rich Future of the Social Sciences", internet pourrait être aux sciences sociales et à l'économie ce que le microscope a été à la biologie, en donnant la possibilité d'analyser l'économie à un niveau "micro" pour mieux comprendre l'univers "macro". Se posent alors de nombreuses questions sur le respect de la vie privée et la protection des données sensibles, qu'il est très important de ne pas négliger pour ne pas que le potentiel incroyable créé par cette information nouvelle ne se retrouve perdu à cause de certaines utilisations frauduleuses ou douteuses de ces nouvelles données.