

Nos discours et nos recherches Google prédisent les tendances économiques

Les discours des décideurs économiques et les requêtes effectuées sur les moteurs de recherche permettent de déceler les sentiments qui animent la société. Ces nouvelles données élargissent les possibilités en matière de prévisions conjoncturelles.

Thomas Chuffart

Abrégé Les économistes ont de plus en plus de données à disposition. Celles issues des recherches sur Google ont ainsi été ouvertes en 2006 ; de leur côté, les banques centrales retranscrivent leurs discours et leurs décisions. Différents outils ont été développés pour analyser toutes ces informations. Ils sont notamment utiles pour effectuer des prévisions en temps réel ou des analyses économiques du rôle des médias dans la formation des prix des actifs financiers. Si ces nouvelles techniques constituent une aubaine pour les économistes qui s'attachent à prédire certains agrégats économiques, les maîtriser face à des données toujours plus hétéroclites n'est pas chose aisée.

L'incertitude, la peur, la défiance... Autant de sentiments qui s'avèrent défavorables à l'économie : l'investissement et la consommation diminuent, le chômage augmente. Si la mesure de ces différentes variables est encore nouvelle, elle est de plus en plus étudiée par les économètres afin de soutenir les théories des économistes. L'accès à une masse de données importante et à des ressources informatiques plus performantes a permis l'émergence de nouvelles pistes de recherche.

Dans la finance, des textes d'actualités financières et issus des réseaux sociaux concernant une entreprise sont par exemple utilisés pour prévoir les fluctuations des prix des actifs et étudier l'impact que pourraient avoir de nouvelles informations sur leur évolution future. En macroéconomie, le texte est utilisé pour prévoir les variations de l'inflation et du chômage ou encore estimer les effets de l'incertitude politique. Néanmoins, en sciences économiques, la notion de « sentiment » est définie de manière particulière. Par exemple, sur les marchés financiers, une mesure de l'incertitude est donnée par la volatilité. L'incertitude porte sur les décisions de politique économique qui seront prises à la suite

des discours et actions des banquiers centraux¹. On appelle ce genre d'objets des « variables latentes », pour lesquelles un proxy doit être utilisé afin de mesurer le phénomène étudié. Parmi les plus célèbres en économie, citons la volatilité des rendements financiers, modélisée à l'aide de séries temporelles. On observe chaque jour un rendement, mais pas sa volatilité inhérente. Pour y parvenir, plusieurs variables proxys peuvent alors être utilisées, comme le carré des rendements ou un modèle statistique qui infère cette variable latente.

Internet, miroir de nos préoccupations

Les économistes utilisent donc généralement des variables quantitatives. Pourtant, en tant que science sociale, l'économie s'intéresse également aux variables qualitatives. Historiquement, l'utilisation de questionnaires pour quantifier ces variables a été prédominante, comme dans le cas de l'indice de confiance des consommateurs de l'université du Michigan. Cette enquête téléphonique est menée chaque mois auprès d'environ 500 ménages pour connaître leur opinion (sous forme de réponses qualitatives) concernant leur situation financière actuelle, les conditions financières attendues d'ici un an ou encore les conditions commerciales générales attendues au cours des douze prochains mois. L'indice de confiance des consommateurs représente la moyenne des réponses à ces questions.

Ces dernières années, toutefois, il est devenu de plus en plus courant d'analyser le sentiment inhérent aux données textuelles, devenues

¹ Baker et al. (2016).

largement disponibles grâce à la numérisation des supports de communication. Un nouveau volet de recherche économétrique s'est ainsi développé: il étudie la meilleure manière de transformer de grands volumes de données qualitatives liées aux sentiments en variables quantitatives². Les méthodes ainsi développées permettent par exemple de mieux cerner les tendances dans les discours des banquiers centraux ou les articles sur la politique monétaire des journaux spécialisés: un corpus de textes peut aller d'une vingtaine de documents à des milliers d'articles.

De nombreuses approches existent sur l'utilisation de données textuelles en économétrie³. Les données qualitatives sont transformées en nombres à l'aide d'outils bien connus des linguistes (traitement automatique du langage naturel, etc.). Autrement dit, les sentiments sont quantifiés. Deux branches se distinguent dans ce domaine: les algorithmes d'apprentissage automatique (« machine learning » en anglais) et les méthodes basées sur le lexique. Les résultats obtenus restent toutefois largement dépendant du choix de la méthode: une méthode de quantification des sentiments doit donc être sélectionnée avec soin en fonction du but recherché.

Il est également possible d'utiliser des sondages ou des proxys⁴, tels que les données de recherche sur Internet répertoriées dans Google Trends. Ces dernières sont publiquement disponibles, au contraire des enquêtes, coûteuses et difficiles à reproduire. Si chacune des recherches effectuées sur Google l'a été pour une raison propre, l'agrégation des données permet de répondre à certaines questions de recherche présentées plus loin. Les requêtes des utilisateurs peuvent en effet représenter les intérêts, les préoccupations ou les intentions du public.

L'impact de la communication des banques centrales

Cette quantification des données de recherche via Google Trends et des données textuelles (qui servent à mesurer des variables latentes de sentiment) est utile pour améliorer la prévision dans de nombreux domaines économiques, comme l'impact de la communication des banques cen-

trales sur les marchés financiers. Une étude⁵ a ainsi utilisé le contenu des déclarations du Comité fédéral américain d'open market (FOMC) pour prévoir les fluctuations des titres du Trésor américain. Il s'est avéré que les changements dans le contenu des déclarations faites dans les rapports du FOMC constituent le principal facteur de la variation des taux d'intérêt, plus que les déviations non anticipées du taux directeur. Une autre recherche⁶ a élargi cette idée pour étudier l'effet du sentiment donné par les banques centrales sur les rendements et la volatilité des marchés boursiers. Les rapports de stabilité financière et les discours prononcés par les gouverneurs des banques centrales ont permis de construire un indice de sentiment de stabilité financière. Résultat: les rapports optimistes ont tendance à faire monter les cours des actions et à réduire la volatilité des marchés au cours du mois suivant.

La prévision du prix des actifs peut également être améliorée grâce à l'analyse des données disponibles sur Internet. L'utilisation de dictionnaires spécifiques regroupant les termes optimistes et pessimistes a par exemple permis de montrer que si les articles de presse sont pessimistes, les rendements prédits seront négatifs, mais que cet effet est transitoire. Ainsi, les articles de presse n'apportent pas d'informations fondamentales et ont un impact transitoire sur les prix, qui retrouvent leur équilibre quelques jours après. Néanmoins, les méthodes d'apprentissage s'avèrent plus performantes que les indices basés sur les dictionnaires afin de prévoir le prix des actions⁷.

Des estimations en temps réel

Des variables importantes telles que le chômage et le produit intérieur brut (PIB) sont mesurées à une faible fréquence et les estimations publiées avec un retard important. En France, l'Institut national de la statistique et des études économiques (Insee) publie par exemple l'indice mensuel de consommation sous un délai d'un mois. D'autres variables, telles que les préjugés raciaux ou la corruption des autorités locales, ne sont quant à elles pas du tout prises en compte par les mesures standard. Les textes produits en ligne (recherches

² Pour davantage de détails, voir Algaba et al. (2019).

³ Pour une revue détaillée, voir notamment Gentzkow et al. (2018).

⁴ Variable observable qui remplace une variable non observable ou non mesurable.

⁵ Lucca et Trebbi (2009).

⁶ Born et al. (2014).

⁷ Jegadeesh et Wu (2013).



Les recherches effectuées sur Internet reflètent les grandes tendances au sein de la société: un nombre élevé de requêtes contenant le mot «grippe» dans un moteur de recherche laisse augurer une épidémie.

Google, publications sur les réseaux sociaux, listes sur des sites Web d'emploi, etc.) peuvent être utilisés pour créer des estimations alternatives en temps réel des valeurs actuelles de ces variables. Contrairement à l'exercice standard de prévision des variables futures, ce processus utilise plusieurs sources de données pour estimer les données actuelles.

Le projet Google Flu Trends offre un premier exemple édifiant. Il prévoit la progression de la grippe en temps réel à partir des requêtes des utilisateurs. Au niveau sanitaire, le volume de recherches d'informations relatives à une maladie peut ainsi être un puissant prédicteur de sa prévalence⁸. En économie, on l'utilise principalement pour prévoir le taux de chômage avec des termes sur la recherche d'emploi. Les données de Google Trends peuvent aussi être utilisées pour la prévision immédiate de variables socioéconomiques⁹. Certaines catégories de recherche sur Google permettent ainsi de prédire les ventes d'automobiles et de maisons, mais aussi la demande touristique et le chômage. Il est également possible d'utiliser les résultats des

requêtes sur le Web pour expliquer le rôle de la corruption. En Russie, le nombre de recherche du mot-clé «Navalny», membre du parti politique d'opposition Labloko et auteur d'un blog sur la corruption, est par exemple utilisé pour décrire les améliorations de la gouvernance des sociétés corrompues¹⁰.

Ces nouvelles techniques sont donc une aubaine pour les économistes qui s'attachent à prédire certains agrégats économiques. Néanmoins, de grands pouvoirs impliquent une grande responsabilité. Maîtriser ces outils dans une jungle des données de plus en plus dense n'est pas chose aisée. Une recherche personnelle sur Google peut, en effet, être le fait d'un individu voulant simplement vérifier l'orthographe du mot «chômage» ou d'un étudiant en médecine qui cherche à vérifier ses connaissances sur la grippe.

Thomas Chuffart

Maître de conférences en économétrie, Centre de recherche sur les stratégies économiques (Crese), université de Franche-Comté, Besançon (France)

⁸ Voir notamment Zeng et Wagner (2002).

⁹ Choi et Varian (2012).

¹⁰ Enikolopov et al (2018).

Bibliographie

Algaba A., Ardia D., Bluteau K., Borms S. et Boudt K. (2019). *Econometrics meets sentiment: An overview of methodology and applications*.
Baker S. R., Bloom N. et Davis S. J. (2016). «Measuring economic policy uncertainty». *The Quarterly Journal of Economics*, 131, 1593–1636.
Born B., Ehrmann M. et Fratzscher M. (2014). «Central bank communication on financial stability». *The Economic Journal*, 124(577), 701–734.

Choi H. et Varian H. (2012). «Predicting the present with Google Trends». *Economic Record*, 88, 2–9.
Enikolopov R., Petrova M. et Sonin K. (2018). «Social media and corruption». *American Economic Journal: Applied Economics*, 10 (1): 150–74.
Gentzkow M., Kelly B.T. et Taddy M. (2017). «Text as Data». *NBER Working Paper*, N° 23276, mars.

Jegadeesh Narasimhan et Wu Di (2013). «Word power: A new approach for content analysis». *Journal of Financial Economics*, 110, 712–729.
Lucca D. O. et Trebbi F. (2009). «Measuring central bank communication: An automated approach with application to FOMC statements». *NBER Working Paper*, N° 15367, septembre.
Zeng X. et Wagner M. (2002). «Modeling the effects of epidemics on routinely collected data». *Journal of the American Medical Informatics Association*, vol. 9 (Supplement_6), S17–S22.