

Les messages issus des articles de presse peuvent contribuer à prédire l'activité économique en temps réel

La plupart des indicateurs conjoncturels d'activité économique utilisés habituellement dans les *Notes de conjoncture* sont mensuels ou trimestriels, et disponibles seulement à la fin du mois ou du trimestre écoulé. Néanmoins, connaître l'évolution de l'activité de manière encore plus précoce peut être crucial, en particulier pendant la crise sanitaire du Covid-19 qui a occasionné des mouvements économiques soudains et de grande ampleur. L'information qualitative contenue dans les articles de la presse économique française peut être mobilisée en ce sens. Elle permet notamment de calculer un indicateur mesurant la tonalité de l'opinion médiatique sur l'activité économique. Cet indicateur donne des indications en temps réel sur l'économie française souvent concordantes avec l'évolution du PIB mesurée *ex post*.

L'indicateur de sentiment médiatique vise à synthétiser l'information qualitative collectée dans des articles de presse

Les articles de presse contiennent une quantité d'information qualitative abondante sur le contexte économique courant, abordent des thématiques variées et sont disponibles deux à trois semaines avant les indicateurs conjoncturels habituels, comme les enquêtes mensuelles de conjoncture (disponibles seulement en fin de mois). À l'aide de techniques d'analyse textuelle, de collecte automatisée de données en ligne (*web scraping*) et d'apprentissage supervisé (*machine learning*), un indicateur de sentiment médiatique sur l'économie française a été élaboré à partir des articles numériques du quotidien *Les Échos*. Pour cela, les mots qui apparaissent dans un article sont catégorisés et comptabilisés en « positif » (respectivement « négatif »), au sens où leur présence traduit une opinion optimiste (respectivement « pessimiste ») sur l'activité économique française le jour de publication de l'article¹. Un indicateur, qui compare les occurrences des mots « positifs » relativement à celles des mots « négatifs », peut alors être calculé : il mesure donc la tonalité générale du journal sur

la situation économique du pays un jour donné. En fonction du nombre de termes positifs et négatifs de l'article, un « score de sentiment » est attribué : la valeur de l'indicateur de sentiment médiatique un mois donné est la moyenne des « scores » des articles du mois. L'indicateur est ensuite centré autour d'une moyenne de 100 et réduit à un écart-type de 10 sur l'ensemble de la période depuis 2000. Par ailleurs, les articles contenant des indicateurs d'instituts économiques comme l'Insee ne sont pas pris en compte dans l'analyse afin d'éviter tout biais de circularité. Cependant, cette restriction n'élimine toutefois pas entièrement l'éventuelle circularité puisque les publications de l'Insee peuvent aussi influencer la tonalité d'articles qui ne les citent pas explicitement.

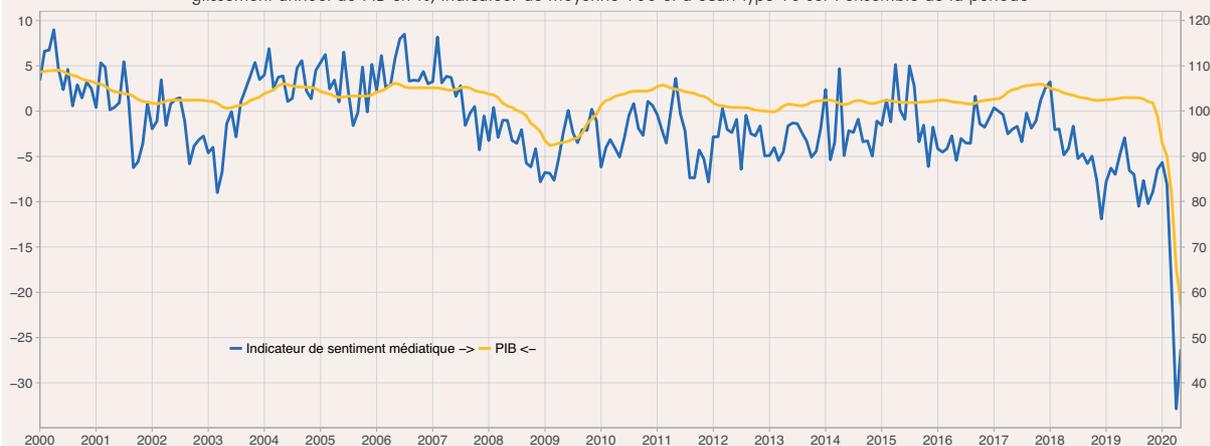
L'indicateur de sentiment médiatique permet d'anticiper les fluctuations de l'activité économique

Calculé sur longue période, l'indicateur de sentiment reflète souvent les évolutions du glissement annuel du PIB, notamment lors des épisodes de contraction économique importante (*graphique 1*). Ainsi, entre 2007 et 2008, la valeur de l'indicateur a nettement

1. La méthodologie utilisée est proche de celle décrite dans l'article de C. Bortoli, S. Combes, T. Renault, Prévoir la croissance du PIB en lisant le journal, *Économie et Statistiques* n°505-506, 2018.

1 - Indicateur de sentiment médiatique et glissement annuel du PIB français

glissement annuel du PIB en %, indicateur de moyenne 100 et d'écart-type 10 sur l'ensemble de la période



Lecture : en mai, l'indicateur était inférieur de plus de 5 écarts-types à sa moyenne de long terme (100 entre 2000 et 2020).

Source : Insee, quotidien *Les Échos*. Calculs Insee. Dernier point : mai 2020 pour l'indicateur de sentiment médiatique, prévision du deuxième trimestre pour le PIB

baissé, tandis qu'en 2008 le PIB en volume a reculé de 2,2 %. Depuis le mois de février 2020, l'indicateur de sentiment médiatique a fortement chuté, bien en deçà de ses pires niveaux antérieurs. Cette forte dégradation de l'opinion médiatique sur la situation conjoncturelle est cohérente avec les estimations de pertes d'activités économiques estimées dans les *Points de conjoncture* depuis le début de la crise actuelle.

L'indicateur suggère une diminution de l'activité économique avant même la mise en place du confinement, qui s'est accentuée avec son entrée en vigueur effective

L'indicateur peut être calculé de façon quotidienne, mais sa volatilité ne permet pas une bonne interprétation. Un arbitrage entre fréquence et volatilité conduit à préférer le calculer de façon

hebdomadaire. L'indicateur obtenu montre une forte baisse de l'activité à partir de la deuxième semaine de mars, à l'issue de laquelle le confinement général de la population a été annoncé puis instauré (*graphique 2*). Cette chute s'est ensuite accentuée jusqu'à fin avril, avant un rebond qui reflète la tonalité moins pessimiste des articles de presse à l'approche du déconfinement. Cette tendance est cohérente avec les niveaux d'activité courante estimés dans les précédents *Points de conjoncture*. Cependant, l'indicateur médiatique a un peu tardé à atteindre son point bas au regard d'une chute d'activité qui a été beaucoup plus subite durant la dernière quinzaine de mars. Depuis mi-mai, la reprise économique se confirmerait : l'indicateur semble en effet revenir vers les niveaux observés juste avant le confinement, mais reste néanmoins en-dessous de son niveau de février dernier. ■

2 - Indicateur de sentiment médiatique hebdomadaire et perte d'activité mensuelle



Lecture : au début du mois de mars, l'indicateur était inférieur de 15 % à sa moyenne de long terme (100 entre 2000 et 2020). Pendant le confinement, il a chuté jusqu'à 35, soit plus de 6 écarts-types en deçà de cette moyenne.
 Source : *Insee, quotidien Les Échos*.