

Les données « haute fréquence » sont surtout utiles à la prévision économique en période de crise brutale

L'ampleur et la soudaineté du choc engendré par la pandémie de Covid-19 ont amoindri la pertinence et le pouvoir prédictif des indicateurs conjoncturels mobilisés usuellement pour mesurer et prévoir l'activité économique. Le suivi conjoncturel en cette période a donc privilégié l'utilisation de nouvelles sources de données, à plus haute fréquence que mensuelle ou trimestrielle. En temps normal, ces indicateurs sont en général – à l'exception de nouvelles données mobilisées pour le suivi de l'activité française depuis la crise de la Covid-19, mais en-dehors du champ de cette étude qui se concentre sur les comparaisons internationales – relativement peu utiles pour la prévision et sont parfois plus volatils que les principaux agrégats économiques. Toutefois, pour les quatre principales économies de la zone euro, les États-Unis et le Royaume-Uni, ces nouvelles données expliquent une part non négligeable de la variation d'indicateurs traditionnels de production et de consommation. Ainsi, dans l'attente de ces résultats mensuels des enquêtes, les données haute fréquence s'avèrent donc utiles à l'analyse et à l'estimation de l'activité. Ensuite, en période de crise comme actuellement, les indicateurs haute fréquence apportent une information complémentaire aux enquêtes de conjoncture permettant de mieux appréhender la perte d'activité à très court terme.

Le pouvoir prédictif des indicateurs usuels issus des enquêtes mensuelles de conjoncture se dégrade à l'approche d'une crise et durant celle-ci

En période normale, l'analyse conjoncturelle et la prévision à court terme réalisées par l'Insee repose en grande partie sur l'exploitation des enquêtes de conjoncture. Une des méthodes mobilisées pour prévoir l'activité économique – production ou consommation par exemple – consiste en des étalonnages¹ utilisant l'information nouvelle apportée chaque mois par les enquêtes auprès des entreprises ou auprès des ménages. Les enquêtes de conjoncture sont en majorité publiées mensuellement, comme les autres indicateurs (ventes au détail, immatriculations) utilisés pour prévoir les grands agrégats économiques, alors que les variables économiques prévues sont pour la plupart mesurées trimestriellement. Dans les modèles d'étalonnage, les enquêtes sont exploitées jusqu'à leur disponibilité la plus récente, par exemple jusqu'à l'enquête de mai pour une prévision du 2^e trimestre. Hors période de crise, cette méthode fournit des prévisions de bonne qualité (Dubois, 2006).

Toutefois, en période de crise majeure ou de grande instabilité économique, ces méthodes sont moins adaptées. La crise de 2008 en fournit un exemple, comme le montre un des éclairages du *Point de conjoncture* du 9 avril dernier : le cadre opérationnel décrit précédemment n'avait permis de rendre compte que très progressivement de l'ampleur du choc. La crise actuelle en est un nouvel exemple : la disponibilité mensuelle et la publication parfois relativement tardive des indicateurs usuels, au regard du choc inédit par sa soudaineté, et les relations économétriques perturbées en ces circonstances par l'ampleur de ce choc, ont conduit à se tourner vers un nouveau mode de suivi conjoncturel, privilégiant

l'estimation en temps réel de l'activité à l'aide de sources alternatives de données.

Les récents *Points de conjoncture* de l'Insee ont ainsi mobilisé des indicateurs à haute fréquence afin de témoigner des conséquences économiques de la crise sanitaire. Le principal avantage des indicateurs à haute fréquence réside, par définition, dans leur actualisation hebdomadaire voire journalière, permettant de suivre presque instantanément la situation des économies et de les comparer. Par exemple, le nombre de requêtes Google relatives au chômage, disponibles dans les *Google Trends*², peut être utilisé comme un indicateur des perspectives d'emploi, voire du nombre de demandeurs d'emploi ; les *Google Trends* relatifs aux centres commerciaux peuvent constituer un indicateur avancé de la fréquentation des commerces de détail et donc de la consommation des ménages. D'autres indicateurs haute fréquence, comme la consommation d'électricité et la concentration de l'air en dioxyde d'azote peuvent aussi témoigner de l'activité globale d'une économie (*tableau 1*).

Cet éclairage vise à évaluer la qualité de ces indicateurs haute fréquence en tant que signaux avancés de l'activité économique et, à ce titre, analyse leur adéquation avec les indicateurs mensuels traditionnels, telles que l'indice de production industrielle pour la production, les ventes au détail pour la consommation. Pour accroître le nombre de points d'identification, seuls les indicateurs haute fréquence disponibles sur une période suffisamment longue et à une fréquence au moins hebdomadaire ont été considérés. Enfin, l'approche retenue dans les modèles économétriques mobilisés, relativement simples, consiste davantage à comparer le pouvoir explicatif

1. Les étalonnages sont des régressions économétriques reliant la variable économique que l'on cherche à prévoir, par exemple la production, aux données mensuelles d'enquêtes de conjoncture ou d'indicateurs avancés, par exemple les ventes au détail ou les immatriculations.

2. Les *Google Trends* sont issues des recherches sur le moteur de recherche Google et indiquant la popularité au cours du temps de certains sujets ou termes de recherches à partir du nombre de requêtes des internautes.

des indicateurs haute fréquence que de rechercher les meilleurs modèles prédictifs. Par ailleurs, les modèles retenus ne reflètent pas nécessairement d'une part les pratiques habituellement utilisées pour la prévision – par exemple, ici la confiance des consommateurs est utilisée dans les modèles pour prévoir les ventes au détail en France dans un but de comparaison, alors qu'elle est très peu mobilisée dans la pratique effective, ni d'autre part les pratiques utilisées actuellement pour la France, reposant sur les données de transactions par cartes bancaires ou encore les données de caisse issues des grandes enseignes de la distribution. Le choix des modèles s'est donc essentiellement fait par souci de comparaison des indicateurs entre les différents pays avancés. En France, les données de transaction par carte bancaire et de caisse ont été extrêmement précieuses pour estimer la consommation des ménages. Elles ne sont malheureusement pas rendues disponibles par les Instituts Statistiques Nationaux à un niveau aussi détaillé pour les autres pays, ce qui explique pourquoi elles n'ont pas été intégrées dans ces comparaisons.

En période « normale » du cycle économique, les données à haute fréquence apportent une information limitée par rapport aux indicateurs usuels

En période « normale » du cycle économique, c'est-à-dire avec des variations limitées de l'activité et excluant par conséquent les périodes de crise comme celle de 2008-2009 ou la crise sanitaire actuelle, les indicateurs à haute fréquence n'améliorent pas significativement la prévision à court terme des agrégats macroéconomiques.

En premier lieu, les indicateurs à haute fréquence, comme les requêtes Google, la consommation

d'électricité ou la pollution de l'air, sont très volatils (même corrigés de facteurs climatiques dans le cas de la consommation d'électricité), bien davantage que ne le sont en temps normal les agrégats macroéconomiques (*graphiques 1 et 2*). Hors période de crise, l'utilisation des indicateurs haute fréquence risque donc de contenir un bruit statistique important, brouillant l'information conjoncturelle.

Bortoli et Combes (2015) l'ont vérifié en utilisant les données *Google Trends* afin de prévoir la consommation mensuelle des ménages. Des recherches Google, comme celles relatives à certains produits, peuvent en effet témoigner du volume de ventes de ces produits. Les auteurs montrent toutefois que les *Google Trends* ne permettent pas d'améliorer significativement la prévision de la consommation agrégée des ménages, seulement celle de certains postes spécifiques, comme par exemple l'habillement-chaussures.

De façon plus systématique, on peut chercher à mesurer la capacité des indicateurs haute fréquence, par rapport aux indicateurs usuels, à rendre compte de la variabilité des variables macroéconomiques (IPI, ventes au détail, immatriculations de véhicules neufs, etc.). Pour ce faire, on compare le pouvoir explicatif (via le R^2 ajusté, part de la variance de la variable endogène expliquée par les variables exogènes, ajustée du nombre de variables introduites dans le modèle) de deux modèles linéaires multiples, l'un modélisant la variable d'intérêt à l'aide uniquement des indicateurs usuels (enquêtes de conjoncture) et l'autre en y ajoutant des indicateurs haute fréquence parmi ceux cités dans le *tableau 1*. La comparaison de ces deux modélisations permet d'appréhender l'apport d'informations des données haute fréquence orthogonales à celui des enquêtes.

Tableau 1 - indicateurs usuels et haute fréquence utilisés dans cet éclairage pour estimer l'activité économique de différents pays pendant la crise

| Agrégat macroéconomique | Indicateurs usuels mensuels | Disponibilité des indicateurs usuels | Indicateurs haute fréquence |
|-------------------------|--|---|--|
| Production | PMI Enquêtes de conjoncture IPI | PMI : disponible à partir du 20 du mois Enquêtes de conjoncture : disponible à partir du 25 du mois IPI : disponible environ 40 à 50 jours après la fin du mois | Consommation d'électricité Concentration en NO2 de l'air <i>Google Trends</i> « Chômage », « Crises » et « Consommation » Indicateur de fret routier (Allemagne) |
| Consommation | PMI Confiance des consommateurs Ventes au détail | PMI : disponible à partir du 20 du mois Indicateur de confiance : disponible à partir du 25 du mois | Consommation d'électricité ; concentration en NO2 de l'air <i>Google Trends</i> « Consommation », « Centre commercial », « Crésit », « Chômage », <i>Google Trends</i> sur le thème de l'achat de véhicules |
| Emploi | Statistiques de l'emploi taux de chômage | Indicateur de perspective d'emploi : disponible à partir du 25 du mois | <i>Google Trends</i> « Chômage » |

Notes :

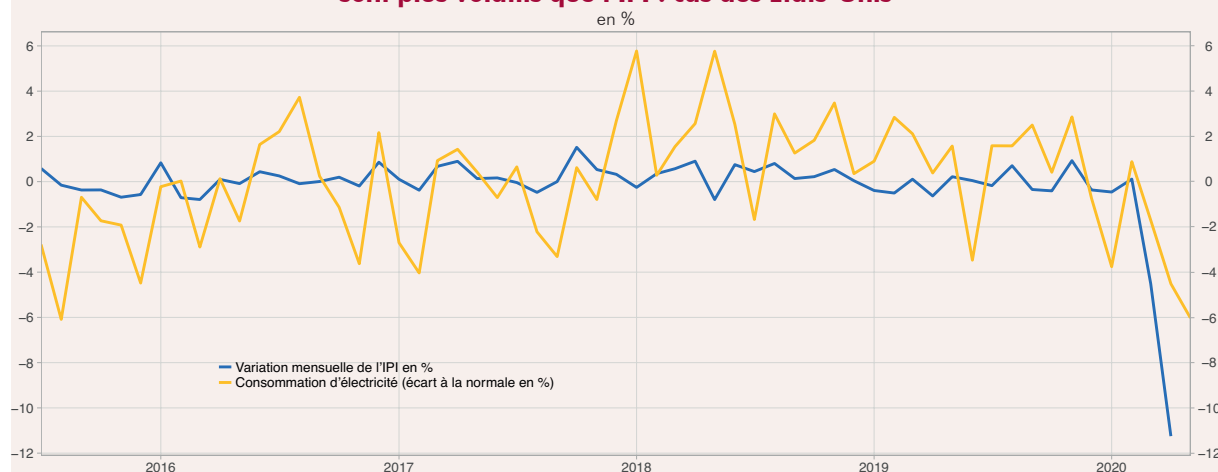
- Seuls sont listés ici les indicateurs à haute fréquence utilisés dans les modèles de prévision présentés plus bas dans cet éclairage. Ces modèles étant homogènes dans les différents pays, certains indicateurs spécifiquement disponibles en France et mobilisés par ailleurs dans ce *Point de conjoncture* (par exemple les transactions par carte bancaire) ne sont pas rappelés dans ce tableau. En effet, nous n'en disposons pas pour les autres pays suivis.
- Les données sur le nombre de requêtes « chômage » sur *Google* ont aussi été utilisées dans les régressions concernant la consommation et la production pour des raisons économiques. Le nombre de demandeurs d'emploi (potentiellement reflété par ces requêtes) est fortement corrélé à l'évolution de la production. Par ailleurs, une augmentation du nombre de demandeurs d'emploi peut affecter négativement les dépenses de consommation des ménages et favoriser une épargne de précaution.

Afin de faciliter la comparaison entre les modèles « avec » et « sans » indicateurs à haute fréquence, les modèles ont été estimés sur la même période, qui est restreinte par la disponibilité de ces indicateurs : nous disposons des données sur la consommation d'électricité à partir de 2015, donc les modèles pour prévoir l'indice de production industrielle sont estimés à partir de 2015. Pour la consommation et le chômage, la période d'estimation exclut la crise de 2008-2009 et commence en 2012. Afin de mesurer le gain moyen pour la prévision, l'erreur de prévision (RMSFE, *root mean square forecast error*) est calculée pour les deux modèles selon une méthode par fenêtre glissante³. À chaque date, le modèle est estimé jusqu'à la dernière donnée disponible, puis la prévision est calculée pour la date suivante et comparée avec l'indicateur effectivement observé pour obtenir l'erreur de prévision « hors échantillon ».

En période normale, l'utilisation des indicateurs haute fréquence ne permet d'améliorer que modestement la qualité d'ajustement d'un modèle linéaire aux données. S'agissant de la modélisation de l'emploi par exemple, bien que les requêtes *Google Trend* relatives au chômage puissent représenter près de la moitié des variations du taux de chômage des quatre principaux pays de la zone euro ou de l'emploi mensuel des autres pays avancés, une fois l'indice PMI et les perspectives d'emploi issues de l'enquête DG ECFIN disponibles, les données à haute fréquence n'apportent que peu d'informations supplémentaires. Ainsi, la hausse du R^2 ajusté suite à l'ajout de ces indicateurs est comprise entre seulement 1 % et 10 % pour un modèle explicatif du taux de chômage français, allemand, italien et espagnol. Toutefois, ce dernier résultat peut témoigner de la présence d'un sur-ajustement du modèle aux données. L'amélioration est aussi

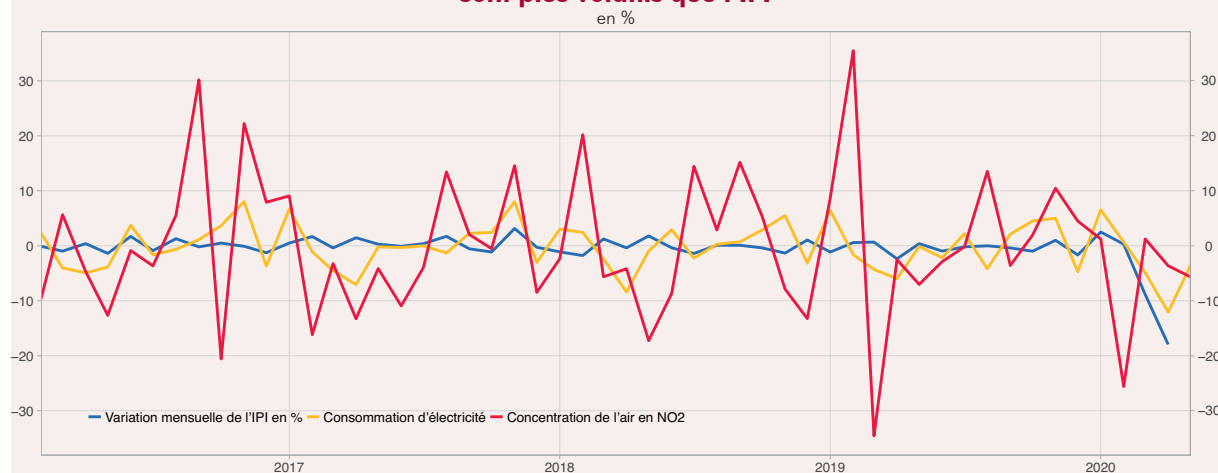
3. Les modèles avec ou sans indicateurs à haute fréquence (HF) sont estimés sur une période allant jusqu'à T , puis sont utilisés pour prévoir le point en $T+1$. Les modèles sont ensuite estimés jusqu'à $T+1$ puis utilisés pour prévoir $T+2$ et ainsi de suite.

1 - Hors période de crise, les indicateurs à haute fréquence (ici la consommation d'électricité) sont plus volatils que l'IPI : cas des États-Unis



Source : Réserve fédérale, Energy Information Administration, calculs Insee

2 - De même qu'aux États-Unis, en Allemagne les indicateurs haute fréquence sont plus volatils que l'IPI



Source : ENTSO-E, EEA, calculs Insee

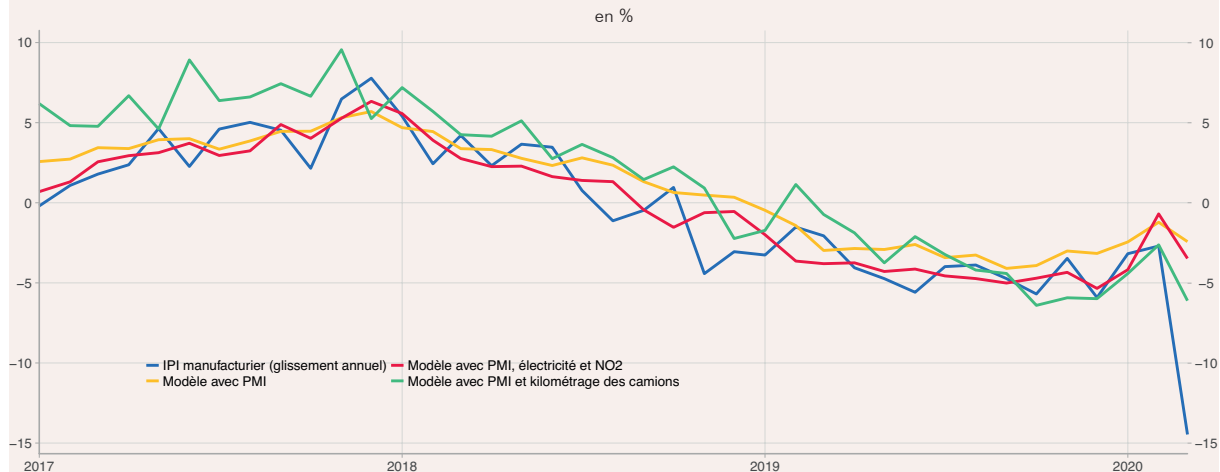
minime pour un modèle de la consommation des ménages dans les pays de la zone euro, mesurée par les ventes au détail, et où le R^2 ajusté baisse de près de 2 % en Italie, et n'augmente que de 2 % et 3 % en Espagne et en Allemagne respectivement. Les indicateurs haute fréquence semblent apporter une information plus significative dans un modèle expliquant la production industrielle, notamment en Allemagne, en Espagne et en France, avec une hausse du R^2 ajusté comprise en 8 % et un peu plus de 40 %. Là encore, une telle hausse de 40 % peut indiquer un phénomène de sur-ajustement. L'ajout des indicateurs haute fréquence peut avoir comme conséquence qu'un modèle ne soit pas suffisamment généralisable pour obtenir de bonne prévision sur de nouvelles observations. Autrement dit, le modèle peut capter à tort une partie de l'aléa du processus générateur des données. C'est pourquoi, afin de mesurer la qualité des informations apportées par les indicateurs haute fréquence, un autre critère doit être utilisé. Ce dernier doit permettre d'évaluer la capacité d'un modèle à correctement prévoir

une nouvelle observation, exclue de l'échantillon d'estimation. Ce critère est le RMSFE.

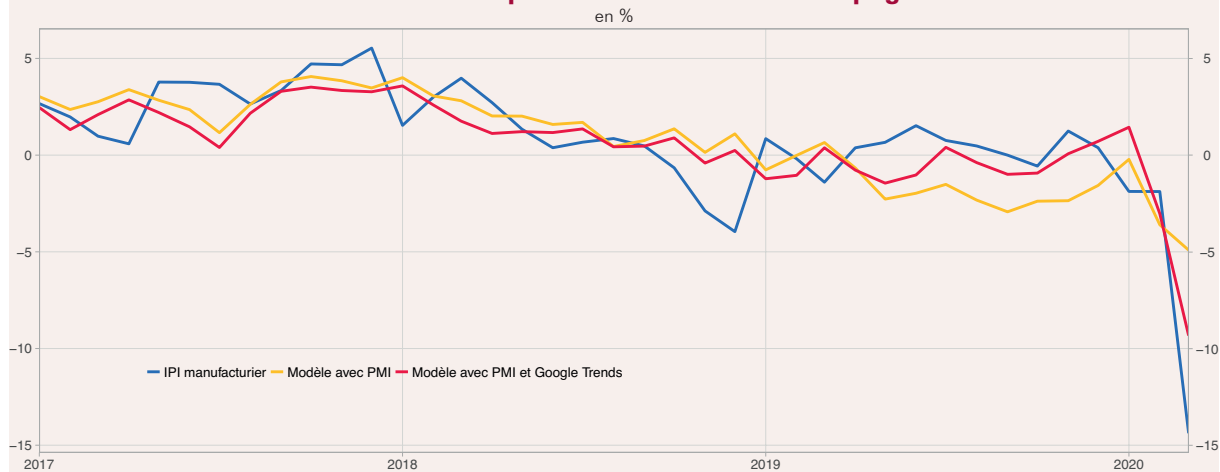
De manière générale, l'utilisation des indicateurs haute fréquence ne permet pas d'améliorer la qualité des prévisions à court terme en période « normale » i.e. hors période de crise. Ainsi, l'erreur de prévision de la production industrielle française, allemande et italienne s'accroît légèrement tandis que les baisses de l'erreur de prévision restent très modestes comme celle de la production industrielle espagnole. Les données de consommation d'électricité sont toutefois beaucoup plus utiles à un niveau fin.

Aux États-Unis, où le détail de la consommation mensuelle est disponible 30 jours après la fin du mois et que l'on cherche donc à prévoir à la place des ventes au détail, l'ajout des *Google Trends* « centre commercial » et « chômage » explique une certaine part des variations, mais n'améliore pas la prévision de la consommation mensuelle aux États-Unis. Au Royaume-Uni non plus, ces indicateurs à haute fréquence n'améliorent pas la prévision de

3 - Les données sur le kilométrage des camions, les Google Trends et la concentration de l'air en NO2 apportent un complément significatif aux PMI dans la modélisation de l'activité industrielle allemande



4 - De même qu'en Allemagne, les indicateurs haute fréquence rendent mieux compte de la chute de la production industrielle en Espagne



la consommation des ménages, même si certains d'entre eux permettent d'améliorer la prévision de postes spécifiques de la consommation, comme par exemple les immatriculations de véhicules.

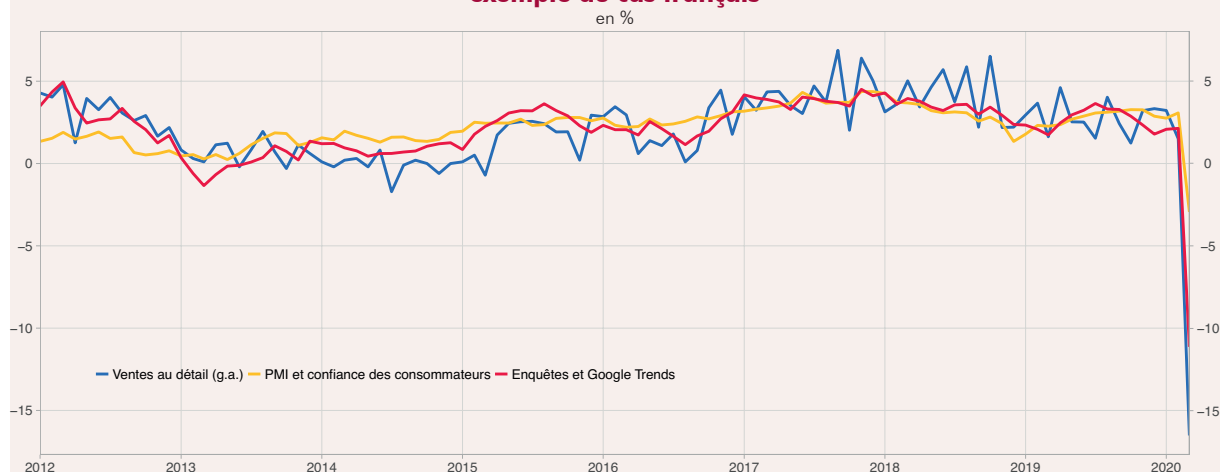
Enfin, aux États-Unis, les statistiques d'emploi relatives à un mois donné sont publiées le premier vendredi du mois suivant (sauf lorsque celui-ci est férié ou tombe le 1^{er} du mois). Ces chiffres sont donc rapidement disponibles, si bien que les indicateurs haute fréquence apportent bien moins d'information que pour la France et l'Allemagne et sont donc moins intéressants que pour la prévision de la production industrielle ou de la consommation des ménages.

C'est en période de crise que certains indicateurs à haute fréquence permettent de mieux appréhender la perte d'activité

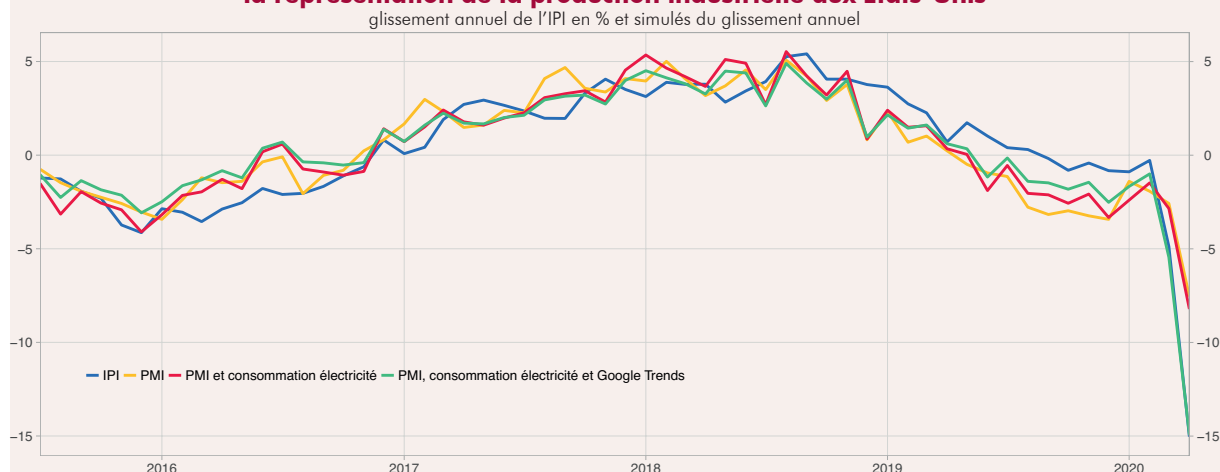
En période de crise, les indicateurs haute fréquence améliorent l'estimation de l'ampleur du choc, par rapport aux indicateurs usuels. Ainsi, à l'aide des

seuls indices PMI, l'estimation de l'effet de la crise sanitaire sur la production industrielle en mars 2020 s'élèverait à -2,4 % et -7,5 % en Allemagne et en France respectivement, contre -6 % et -12 % avec l'ajout d'indicateurs haute fréquence, et en comparaison à une baisse effective d'un peu plus de 14 % et 19 % respectivement en Allemagne (*graphique 3*) et en France. En Espagne les indicateurs haute fréquence permettent aussi de mieux capter la baisse d'activité effective en mars 2020, mais dans une moindre mesure en Italie. La chute de l'activité industrielle observée en mars en Espagne s'élève à -13 % et celle estimée via l'ajout des indicateurs haute fréquence (Google Trends notamment) est de l'ordre de -9,3 % (contre -4,9 % à l'aide uniquement des indices PMI, *graphique 4*). Toutefois, l'écart entre l'ampleur estimée et effective du choc reste élevé, d'environ 4 à 7 points de pourcentage selon les pays. Par conséquent, malgré l'apport des indicateurs haute fréquence, les modèles économétriques parviennent difficilement à rendre compte de l'ampleur de la baisse d'activité actuelle.

5 - Les Google Trends ont sensiblement amélioré l'estimation de la baisse des ventes au détail : exemple du cas français



6 - De même qu'en Allemagne, les indicateurs haute fréquence améliorent la représentation de la production industrielle aux États-Unis



S'agissant de la baisse de la consommation, les indicateurs haute fréquence permettent un gain de prévision significatif. Par exemple, tandis que les ventes au détail ont diminué de plus de 16 % en France en mars, l'amplitude du choc estimée par des indicateurs standards ne s'élève qu'à -3 %, contre une baisse prévue de -11 % avec l'introduction des données de *Google Trends* (*graphique 5*). En revanche, ces gains de prévisions sont moins significatifs pour l'emploi et la consommation en Allemagne, en Espagne et en Italie.

Au Royaume-Uni, la consommation d'électricité n'apporte pas d'information supplémentaire. En revanche, l'ajout de la pollution apporte une information significative : le R^2 (en échantillon) augmente de 90 %, à 64 %. Aux États-Unis, la consommation d'électricité et les recherches *Google* sur le thème du chômage apportent une information supplémentaire aux PMI pour l'indice de production industrielle (*graphique 6*) : le R^2 augmente de 20 % lorsque l'on ajoute ces deux indicateurs haute fréquence dans la régression aux côtés de l'indice PMI de l'*Institute Supply of Management*. Hors échantillon néanmoins, c'est-à-dire lorsque, pour prévoir chaque point, le modèle est estimé sur les données disponibles à cette date (donc jusqu'à la date $T-1$), l'erreur moyenne de prévision ne diminue que très légèrement. L'ajout des indicateurs haute

fréquence, comme la consommation d'électricité ou les *Google Trends* sur le chômage, améliore beaucoup la prévision de la baisse de l'activité aux États-Unis (*graphique 3*). Hors échantillon, l'amélioration de la prévision est moindre mais toujours sensible.

En définitive, si dans la plupart des cas les indicateurs à haute fréquence n'apportent pas d'information significative supplémentaire à celle des enquêtes de conjoncture et ne permettent d'améliorer les prévisions économiques que de façon limitée pendant les périodes « normales », en revanche en période de crise et de variations brutales et de grande ampleur de l'activité économique, les modèles usuels se trouvent inadaptés pour prévoir l'activité économique. Les indicateurs à haute fréquence peuvent alors être mobilisés afin d'améliorer un peu la prévision. Toutefois, certaines données à haute fréquence mobilisées dans le cas particulier de la France mais hors du champ de cette étude, comme les données de caisse ou les données sur les transactions par carte bancaire, constituent une source d'information précieuse pour le suivi conjoncturel.

L'expertise et l'analyse du prévisionniste sont nécessaires pour ajuster et modifier les modèles économétriques afin de mieux appréhender l'évolution de l'activité. ■

Bibliographie

C. Bortoli et S. Combes, (2010) « Apports de *Google Trends* pour prévoir la conjoncture française : des pistes limitées », *Note de conjoncture* de mars 2015

E. Dubois, (2006) « Étalonnages à l'aide d'enquêtes de conjoncture: de nouveaux résultats ». *Économie & prévision*, n° 172, janvier 2006

Insee, (2020) « Mesure et prévision en temps de crise : une comparaison avec la période 2008-2009 », *Point de conjoncture* du 9 avril 2020. ■