



COMMISSION « SYSTÈME FINANCIER ET FINANCEMENT DE L'ÉCONOMIE »

COMPTE RENDU DE LA RÉUNION DE
LA COMMISSION SYSTÈME FINANCIER ET FINANCEMENT DE L'ÉCONOMIE
- 15 mai 2025 -

Présidente : Anne ÉPAULARD, professeure des universités, Université Paris Dauphine – PSL et directrice du département Économie et environnement, OFCE

Rapporteurs : Marie LECLAIR, cheffe du Département des comptes nationaux (DCN) à l'Insee
Emmanuel GERVAIS, chef du Service d'Études et statistiques des opérations financières (SESOF) à la Banque de France

Responsable de la commission : Céline CRAVATTE (01 87 69 57 03)

***Tous les documents sont disponibles sur le site cnis.fr à la [page de la commission](#)
En particulier l'ordre du jour, une liste de références à consulter, les supports de présentation, le suivi d'avis***

RAPPEL DE L'ORDRE DU JOUR

| | |
|---|----|
| LISTE DES PARTICIPANTS..... | 3 |
| SUIVI DE L'AVIS DE MOYEN TERME..... | 5 |
| PROGRAMME DE TRAVAIL DES PRODUCTEURS..... | 7 |
| POINT D'INFORMATION : LA RÉVISION DE LA NOMENCLATURE A17..... | 10 |
| SOLLICITATION D'AVIS D'OPPORTUNITÉ (enquête crédit-bail)..... | 11 |
| THÈME CENTRAL : QUELLES DONNÉES RELATIVES POUR LES STATISTIQUES DE COURT TERME ?..... | 14 |
| 1. Un panorama de l'utilisation des données « alternatives » pour l'analyse économique..... | 14 |
| 2. Utilisation des données privées par le Service statistique public : quel cadre juridique ?..... | 18 |
| 3. Une approche alternative pour estimer le nombre de Voyageurs étrangers en France et leurs dépenses..... | 20 |
| 4. Quel apport des données alternatives pour le nowcasting ?..... | 22 |
| 5. Les apports de l'utilisation des données du Crédit Mutuel : l'expérience du CAE..... | 25 |
| DOCUMENTS PRÉPARATOIRES ET COMPLÉMENTAIRES..... | 30 |
| AVIS D'OPPORTUNITÉ..... | 31 |

LISTE DES PARTICIPANTS

Commission

| | | |
|----------------------------|-------------------|---|
| EPAULARD | Anne | Présidente – Université Paris Dauphine – PSL/ Observatoire Français des conjonctures économiques - OFCE |
| GERVAIS LECLAIR | Emmanuel Marie | Rapporteur – Banque de France (BdF) Rapporteur – Insee |

Intervenants

| | | |
|-----------------------|-----------|---|
| DE ALBUQUERQUE | Sarita | Banque de France (BdF) |
| DOZ | Catherine | Ecole d'économie de Paris (PSE-Paris School of Economics) |
| MARTIN | Tanguy | Insee |
| PARIS | Hélène | Conseil d'analyse économique (CAE) |
| PRADELLE | Félix | Banque de France (BdF) |
| REDOR | Patrick | Insee |

Autres participants

| | | |
|----------------------|---------------|---|
| ANDRE | Julien | Banque de France (BdF) |
| BARADJI | Eva | Ministère de l'éducation nationale et de la jeunesse - Direction de l'évaluation, de la prospective et de la performance (DEPP) |
| BASSO | Florian | Particulier |
| BEN HASSINE | Haithem | France Stratégie |
| BESSEC | Marie | Université Paris Dauphine |
| BILLOT | Sylvain | Institut national des statistiques et des études économiques (Insee) |
| BONFILS | Louise | Caisse centrale mutualité sociale agricole (CMSA) |
| BOSSY | Fabien | Société Générale |
| BOURSIER | Nathalie | Chambre de commerce et d'industrie (CCI) |
| BRETNACHER | Gilles | France Assureurs |
| BURON | Maël | Institut national des statistiques et des études économiques (Insee) |
| CARON | Nathalie | Institut national des statistiques et des études économiques (Insee) |
| CHAPUIS-MIROL | Maëlle | Chambre des métiers et de l'artisanat France (CMA France) |
| CRISAN | Mihaela | Institut national des statistiques et des études économiques (Insee) |
| DEVINEAU | Jeanne | Institut national des statistiques et des études économiques (Insee) |
| DIOGO | François | Région Grand Est |
| DUNAND | Olivier | Institut national des statistiques et des études économiques (Insee) |
| EGUIENTA | Janine | Institut national des statistiques et des études économiques (Insee) |
| FIGARO | Hélène | Institut national des statistiques et des études économiques (Insee) |
| GASPAROTTO | Jean-philippe | Confédération générale du travail (CGT) |
| GUANNEL | Boris | Institut national des statistiques et des études économiques (Insee) |
| GUENE | Stéphane | Institut national des statistiques et des études économiques (Insee) |
| HAUSEUX | Yaëlle | Institut national des statistiques et des études économiques (Insee) |
| HEAM | Jean Cyprien | Institut national des statistiques et des études économiques (Insee) |
| HILLION | Mélina | Institut national des statistiques et des études économiques (Insee) |
| LACAILLE | Yves | Union nationale des professions libérales (UNAPL) |
| LAVAUD | Catherine | Institut national des statistiques et des études économiques (Insee) |

| | | |
|-------------------------|------------|--|
| LEMARCHAND | David | Institut national des statistiques et des études économiques (Insee) |
| LONGEAUD-D. | François | Particulier |
| MENAA | Marwan | Banque de France (BdF) |
| MOISSON | Francoise | Confédération générale du travail (CGT) |
| MORELLO | Emmanuel | Institut national des statistiques et des études économiques (Insee) |
| MUTRICY | Philippe | Banque Publique d'Investissement (Bpifrance) |
| PEPIN-STOLIAROFF | Anne-Marie | Institut national des statistiques et des études économiques (Insee) |
| PLAYOUST | Barbara | Institut national des statistiques et des études économiques (Insee) |
| RAHARIVOHITRA | Désiré | Fédération des industries mécaniques |
| REBUFA | Bérénice | Ubiquis |
| ROCHE | Corinne | Institut national des statistiques et des études économiques (Insee) |
| SILBERMAN | Roxane | Centre national de la recherche scientifique (CNRS) |
| STUDER | Nicolas | Institut national des statistiques et des études économiques (Insee) |
| TCHA | Stephanie | Ministère de l'Intérieur - Service statistique ministériel de la sécurité intérieure (SSMSI) |
| VESSILLIER | Delphine | Fédération française du bâtiment |

Secrétariat Général du Cnis

| | | |
|----------------------------|-----------|-----------------------------|
| CRAVATTE | Céline | Secrétariat général du Cnis |
| DE RAPHELIS | Magali | Secrétariat général du Cnis |
| GUILLAUMAT-TAILLIET | François | Secrétariat général du Cnis |
| LAGARENNE | Christine | Secrétariat général du Cnis |
| SEKOURI | Mohamed | Secrétariat général du Cnis |

SUIVI DE L'AVIS DE MOYEN TERME

FACILITER L'UTILISATION DE DONNÉES ALTERNATIVES POUR AMÉLIORER LES PRÉVISIONS ET DIAGNOSTICS ÉCONOMIQUES

Qu'est-ce que le suivi d'avis ?

Chaque séance de commission traite d'un sujet central faisant référence à un avis prévu par le programme de [moyen terme](#) du Cnis. Ces avis, qui sont révisés tous les cinq ans, traduisent les attentes de la société en matière d'informations statistiques, sur des sujets qu'elle considère comme majeurs. Ils prennent la forme de recommandations adressées à la statistique publique. À l'issue de chaque séance de commission, un suivi d'avis est adopté afin de faire le point sur les réponses apportées aux besoins mentionnés par l'avis de moyen terme, et rappeler les besoins restant à couvrir.

Avis de moyen terme étudié lors de cette séance :

Avis n° 10 : La crise sanitaire de 2020 a confirmé, s'il en était besoin, l'utilité de la prévision à très court terme (nowcasting) pour aider les décideurs publics à s'adapter au mieux à la situation économique et sociale.

Le Conseil salue les investigations menées par la statistique publique pendant la crise sanitaire pour utiliser de nouvelles sources de données afin d'améliorer le diagnostic rapide de certaines grandeurs économiques. Il encourage la consolidation des travaux menés sur l'utilisation des données à haute fréquence en complément des données d'enquêtes et des sources administratives. Il recommande que ces données soient systématiquement rapprochées des statistiques publiques et de leurs concepts et que des études de validation soient menées, pour celles qui présentent un potentiel plus pérenne.

Le conseil encourage par ailleurs le rapprochement entre producteurs, une meilleure articulation de leurs sources respectives et la poursuite des études croisées permettant d'enrichir les analyses macroéconomiques par des données individuelles. Il souhaite que ces études et leurs résultats, une fois consolidés, soient rendus publics et facilement accessibles. Il encourage une réflexion sur la façon dont ces données pourront être mises à disposition des chercheurs

Suivi d'avis adopté à l'issue de la séance :

La commission Système financier et financement de l'économie a pris connaissance de travaux mobilisant des données alternatives pour améliorer les prévisions et les diagnostics économiques et aider les décideurs publics à s'adapter au mieux à la situation économique et sociale. Elle avait déjà abordé ce thème au printemps 2021, en lien avec les innovations mises en place dans le cadre de la crise Covid. Par données alternatives, on entend celles qui diffèrent des enquêtes statistiques et des bases administratives : données liées au « big data », données privées accessibles à un niveau plus ou moins granulaire, données de formats variés (web scraping, données textuelles, images en particulier satellitaires). Ces données sont rapidement disponibles, de haute fréquence, avec une granularité très fine, et leur potentiel diffère selon chaque situation et chaque source de données.

Les données alternatives peuvent être utilisées pour alléger la charge d'enquêtes sur les répondants, faire face à des situations exceptionnelles, mesurer des phénomènes mal documentés auparavant, réduire les délais de production, et effectuer de la prospective. Pour la production statistique courante, elles ont permis un remplacement partiel d'enquêtes et des gains d'informations (données de caisse pour l'indice des prix à la consommation et les indices de chiffre d'affaires (ICA), données de transaction bancaire pour les ICA provisoires, données de téléphonie mobile pour la ligne dépenses des voyageurs de la balance des paiements, par exemple). Pour la prospective, les pistes explorées montrent que ces données peuvent permettre parfois d'anticiper les résultats des enquêtes mais ne remplacent des enquêtes que dans des situations exceptionnelles où elles ne peuvent avoir lieu. Pour l'analyse, elles permettent d'explorer finement avec des délais réduits et d'éventuels appariements des domaines peu documentés (capacité de remboursement des très petites entreprises, dépenses énergétiques des ménages grâce aux données de comptes bancaires appariées au diagnostic de performance énergétique du logement, variations infra-mensuelles des dépenses des ménages).

La commission encourage la poursuite des travaux de la statistique publique permettant d'explorer et de mobiliser ces sources prometteuses. Elle note que la révision récente du règlement statistique européen 223/2009 ouvre des perspectives d'extension du droit d'accès aux données privées, sous la contrainte de finalités de statistiques européennes, et nécessite une adaptation nationale de la loi de 1951. Elle rappelle les conditions nécessaires à leur utilisation par la statistique publique en complément des outils traditionnels qui restent centraux : accès pérenne aux données, possibilité de contrôler le champ, sa représentativité et ses évolutions, cohérence des données avec les concepts usuels. Dans le cas de données privées, des échanges étroits et continus avec les détenteurs des bases de données sont nécessaires, dans le respect de la confidentialité. La commission rappelle donc le caractère essentiel de la mise en place de mutualisations et de partenariats entre services statistiques, mais aussi avec les fournisseurs de données privés et les chercheurs au niveau national et européen.

La séance est ouverte à 9 heures 35 sous la présidence de Anne EPAULARD.

Anne EPAULARD présente l'ordre du jour.

PROGRAMME DE TRAVAIL DES PRODUCTEURS

Marie LECLAIR, Insee

Je vais vous présenter le programme du service statistique ministériel des finances publiques (SSM) de la Direction générale des finances publiques (DGFIP) du Pôle statistique publiques, en trois blocs.

Le premier concerne l'acquisition de nouvelles bases de données ou des travaux sur l'accessibilité et la maniabilité des bases de données existantes. Un travail est réalisé sur la base de données patrimoniales permettant d'extraire un certain nombre de données individuelles sur les successions, les ventes, et les donations. Un autre travail consiste à construire une base sur les taux de détention des entreprises via une approche individuelle incluant les personnes physiques qui détiennent ces entreprises. Des travaux sur les fichiers de données foncières sont également réalisés afin de les rendre plus accessibles et en extraire des données sur les transactions immobilières et sur le foncier bâti, permettant de recalculer des taxes foncières théoriques. Outre les travaux réalisés sur ces bases de données, l'obtention de nouvelles sources comme les données européennes sur les plateformes d'économie collaborative est également au programme.

Le deuxième axe est le renforcement des travaux d'analyse, avec une section créée l'année dernière pour l'évaluation de la fraude fiscale, en étendant les travaux sur la TVA à l'impôt sur les sociétés. Par ailleurs, une publication infra-annuelle sur les recettes fiscales sera mise en place et la date de publication annuelle usuelle sera avancée.

Le troisième axe couvre des travaux plus classiques de production et de diffusion de données statistiques. Le SSM consacrera un temps important à la migration vers Python, à la mise en conformité des traitements de données personnelles avec la Commission nationale de l'informatique et des libertés (CNIL), et au développement d'outils de visualisation personnalisée sur via des cubes de données structurés via *data.economie.gouv*, tout en respectant le secret statistique.

Par ailleurs, le département des comptes nationaux de l'Insee, dont je suis la cheffe, a publié en nouvelle base les comptes l'année dernière, mais nous nous projetons déjà vers le prochain changement de base prévu pour 2029, voire 2030. Ce changement sera particulièrement lourd, car, outre une revue des méthodes et sources réalisée à chaque changement de base, deux évolutions majeures auront lieu : un changement de nomenclature et un nouveau système européen des comptes (SEC). Ces changements sont rares, les derniers datant de 2008 et 2010 respectivement.

Nous travaillons dès maintenant avec Eurostat pour établir ce futur système européen des comptes qui n'est pas encore établi bien que la commission statistique de l'ONU ait adopté un système de comptabilité nationale au printemps. Un travail important de discussion est mené pour établir ces nouvelles règles et concepts en parallèle du programme de transmission qui inclura à la volonté d'Eurostat de nouveaux produits tels que les comptes des ménages distribués, le tableau des entrées-sorties (TES) digital ou les comptes multinationaux étrangers.

Par ailleurs, un rapport de l'Inspection générale de l'Insee, qui est un audit de performance des comptes nationaux, a mis en place un certain nombre d'axes de travail sur lesquels nous avançons, notamment se confronter davantage aux utilisateurs via cette commission, mais aussi via un club utilisateur. Ce rapport préconisait également de renforcer les investissements et notre présence internationale, d'augmenter les comptes dans leurs dimensions environnementale et sociale en allégeant les travaux courants et en incitant les parcours professionnels dans ce domaine. Nous nous attellerons l'année prochaine à rendre ces productions régulières.

Anne EPAULARD, Présidente de la commission

Merci, j'ai appris des choses sur les activités de la DGFIP, notamment sur les bases de données. Tout cela sera-t-il versé avec le Centre d'accès sécurisé aux données (CASD) ?

Marie LECLAIR

Dans leur programme de travail, certaines choses seront effectivement versées avec le CASD. Pour d'autres, ils sont encore dans des travaux exploratoires. Ils cherchent d'abord à construire ces bases de données avant de les verser. Tout n'est pas explicitement indiqué comme devant être prochainement versé au CASD. Je vous invite à consulter le programme de travail qui a été diffusé.

Anne EPAULARD

Savoir que c'est dans leur champ et qu'ils y pensent systématiquement est déjà très rassurant, car l'accès aux données DGFIP a toujours été compliqué.

Emmanuel GERVAIS, Banque de France

Je vais vous parler de trois points concernant la Banque de France : les projets concernant la balance des paiements, les statistiques bancaires, et les comptes financiers.

Concernant la balance des paiements, la refonte de la partie recette de ligne voyage sera présentée par mes collègues. Du côté des investissements directs, un règlement européen, European Business Statistics (EBS), est mis en œuvre, visant à publier des informations sur le commerce international et la manière de repenser les services qui sont offerts. Cela sera estimé en croisant diverses sources, dont des données provenant de l'Insee et des collectes de la balance des paiements.

Nous travaillons également sur l'estimation de l'empreinte carbone des investissements directs, dans le cadre des mises en œuvre des recommandations de la Data Gap Initiative. En l'absence de bilan carbone publié, nous croisons des données privées avec nos données existantes dont nous disposons par ailleurs. D'un point de vue opérationnel, nous utilisons une grande enquête menée avec l'aide des succursales de la Banque de France sur les investissements directs français à l'étranger, exploitant des données comptables pour le recensement des investissements directs. Une expérimentation est en cours avec l'Insee pour automatiser l'extraction de ces données grâce à l'intelligence artificielle. Ce projet est bien avancé et devrait passer à l'industrialisation dans l'année 2025.

Sur les statistiques monétaires, un important projet européen appelé IREF est toujours en cours. À l'échelle de la zone euro, il vise à harmoniser les États déclarés par les établissements de crédit, au moins pour la partie utilisée pour la collecte de statistiques monétaires centralisées par la Banque centrale européenne (BCE). L'idée est aussi de créer une plateforme de collecte, avec des enjeux informatiques majeurs. Le démarrage du projet est prévu pour 2029.

Côté comptes financiers, nous participons aux discussions avec la BCE pour mettre à jour l'Orientation organisant la collecte de données sur les comptes financiers, afin de tenir compte du nouveau règlement sur les comptes nationaux qui sera adopté par EUROSTAT suite à l'adoption par les Nations Unies du SNA 2025. Ces discussions influencent également le programme de travail du Système européen de banques centrales (SEBC) en matière de comptes financiers. Les changements concernent notamment les ventilations sectorielles et les nouveaux instruments, avec un suivi plus précis des cryptomonnaies.

Du point de vue pratique de la production des comptes financiers, nous avons affiné le calcul des partages entre les produits d'assurance-vie et les droits à pension, qui sont deux instruments différents dans les comptes financiers.

Roxane SILBERMAN, CNRS

Merci beaucoup pour ces présentations. J'ai une question pour la Banque de France. Un pilote est lancé au niveau européen avec la BCE et quelques banques centrales pour permettre l'accès aux données. La BdF n'y est pas incluse, mais je n'ai pas compris si chaque banque centrale participe avec son propre mode d'accès.

Emmanuel GERVAIS

Ce dispositif est un complément de ce qui existe déjà au niveau national. Pour les données européennes, il faut s'adresser à la BCE, la BdF ne pouvant fournir que des données françaises.

Roxane SILBERMAN

Il s'agit d'accès aux données de chaque pays.

Emmanuel GERVAIS

L'idée est peut-être de créer un outil commun permettant à tous d'accéder à toutes les données, mais nous sommes vraiment au début de ce projet. Chaque pays a ses propres données et certaines ne sont collectées que dans certains pays. Les systèmes bancaires sont assez différents, et les différences dans les collectes reflètent les différences entre pays. C'est donc très difficile de créer un modèle commun. Pour la collecte, c'est déjà très compliqué nécessitant de longues discussions. Pour le partage des données nationales, la BdF a déjà des outils assez bien développés.

Anne EPAULARD

L'intérêt pour les chercheurs est d'avoir accès non seulement aux données françaises via le CASD, mais aussi de pouvoir les comparer, même s'il faut ajuster les définitions. Je comprends que c'est cet environnement qui est en développement.

Emmanuel GERVAIS

C'est un projet pilote, sachant que pour les séries macro, la BCE met déjà énormément de données en ligne.

Anne EPAULARD

Je crois que la question portait plutôt sur les données individuelles.

Roxane SILBERMAN

Je re-précise donc ma question. S'agit-il d'un système central BCE ? L'idée est-elle de construire un mode d'accès unique, ou bien chaque banque conservera t-elle son propre mode d'accès (pour la France, via le CASD) ?

Emmanuel GERVAIS

À ma connaissance, il n'existe pas de projet de création d'une plateforme informatique commune pour accéder à toutes les données granulaires. Pour le moment, cela n'existe pas, même pour la collecte de données entre les banques centrales nationales. Le jour où nous aurons une collecte harmonisée avec un partage des données, cela se fera plus naturellement.

Hélène PARIS, Conseil d'analyse économique (CAE)

J'ai une question pour le département des comptes nationaux. J'ai été assez intriguée par cette notion de TES digital. Pourriez-vous en dire quelques mots ?

Marie LECLAIR

L'idée est de pouvoir distinguer dans le TES ce qui est acheté sur plateforme de ce qui est acheté dans des magasins physiques, ainsi que les services numériques utilisés par les entreprises. Nous nous retrouvons avec des TES où les équilibres ressources-emploi sont éclatés selon le mode d'acquisition des produits, et dans la partie comptes de branche, où l'on examine la production et les consommations intermédiaires, nous pourrions identifier la consommation digitale et non digitale. Cet outil est plus expérimental et analytique.

POINT D'INFORMATION : LA RÉVISION DE LA NOMENCLATURE A17

Un [document](#) est projeté.

Marie LECLAIR

Ce point d'information vise à recueillir les avis de nos utilisateurs sur une partie du changement de nomenclature, uniquement au niveau d'agrégation A17 pour le moment.

Ce travail sur la structure de la nomenclature d'activité française est déjà achevé et a été réalisé dans le cadre d'un groupe de travail du Cnis et d'autres commissions. Les nomenclatures sont d'abord discutées au niveau international de l'ONU, puis déclinées au niveau européen. Au niveau national, nous pouvons ajouter un niveau de détail, ce qui a été fait par le groupe de travail et validé par Eurostat. Le niveau français détaillé n'intéresse pas directement les macro-économistes, mais a des impacts sur les niveaux plus agrégés.

a. Les principaux changements

Désormais, la réparation de véhicules automobiles sera classée dans les services de réparation et non plus dans le commerce. De plus, les donneurs d'ordre qui sous-traitent la production industrielle seront classés en industrie même s'ils fournissent des intrants immatériels, alors qu'auparavant ils étaient classés en commerce. Par exemple, si une marque demande à un sous-traitant de produire entièrement son produit, ce donneur d'ordre deviendra industriel. Nous aurons donc des mouvements entre commerce et industrie. La nomenclature permet également de repérer les activités de service d'intermédiation classées selon le produit concerné, comme Doctolib dans le domaine de la santé. Enfin, les activités d'information et de communication, classées jusqu'ici dans la même section J, sont éclatées en deux sections distinctes.

Cette nomenclature ne sera appliquée que progressivement dans les statistiques, en commençant dans les répertoires statistiques, puis dans les statistiques structurelles, et enfin dans les comptes nationaux qui basculeront en 2029.

Les nomenclatures les plus agrégées ne sont cependant pas encore décidées. Les nomenclatures naturelles sont structurées en sections, divisions et sous-classes. Certains niveaux d'agrégation intermédiaire ne respectent pas cette structure, notamment le niveau A17 qui est le niveau de publication des comptes trimestriels, mais aussi des publications conjoncturelles comme l'emploi au niveau français, le niveau A10 (regroupement de sections) pour l'internationale et le A21 (niveau naturel des sections).

Le niveau A17 a été créé, car le niveau A10, établi au niveau international, n'était pas assez détaillé dans l'industrie et la nomenclature supérieure était trop détaillée dans les services. Le niveau A17 est donc un niveau intermédiaire qui permet d'avoir plus de détails dans l'industrie.

Les réflexions sur le changement de nomenclature sont menées en parallèle des discussions internationales puisque le niveau A10 n'a pas encore été repensé à l'international, notamment l'impact que l'éclatement de la section J pourrait avoir sur cette nomenclature.

b. Les questions spécifiques à la France

La branche C2 « cokéfaction et raffinage » a un poids trop faible et nous envisageons soit de la regrouper avec la branche D2 pour créer une branche énergie, soit de l'intégrer aux autres produits industriels. À l'inverse, la branche C5 « fabrication des autres produits industriels » est très importante et nous voudrions la subdiviser, soit selon la destination finale des branches, soit selon leur intensité énergétique. Nous avons également des questions de taxinomie pour nommer cette nouvelle branche sur la classification.

Nous avons tout de même un certain nombre de contraintes dans l'établissement de ces nomenclatures. Nous souhaitons emboîter la future A17 dans les nomenclatures amont et aval au maximum afin qu'elle nous permette de répondre aux principes de diffusion d'Eurostat en nous assurant la visibilité pour les utilisateurs et que ce nouveau niveau nous permette d'apporter une nouvelle information trimestrielle. De plus, le niveau A17 permet d'avoir des séries annuelles très longues (implication pour les comptes annuels sur les séries 1949-1979, il faut donc avoir la possibilité de rétopoler les séries en nouvelle nomenclature).

Nous vous invitons à nous contacter si vous avez des suggestions.

Céline CRAVATTE, Secrétariat général du Cnis

Les présentations et le compte rendu détaillé seront disponibles sur le site du Cnis.

Anne EPAULARD

Merci. Je suis impressionnée par toutes les réflexions menées avant de procéder à des changements. En tant qu'utilisatrice, j'ai souvent l'impression que les choses n'ont pas été suffisamment pensées, mais je constate la complexité de satisfaire tout le monde, et ces questions de rétopolation sont effectivement très importantes.

SOLLICITATION D'AVIS D'OPPORTUNITÉ (enquête crédit-bail)

François GUILLAUMAT-TAILLIET, Secrétariat général du Cnis

Toute enquête qui a vocation à devenir une enquête de la statistique publique doit être présentée en opportunité au Cnis, en commission publique. Si la commission juge l'enquête opportune compte tenu des informations disponibles à ce stade, sans tous les détails ni le questionnaire final, le droit de l'instruire sera ouvert, puis elle passera devant le Comité du Label qui décidera de lui accorder le label ou non, avec caractère obligatoire ou non, selon qu'il s'agit d'une enquête auprès des ménages ou des entreprises. Si le visa est donné, elle sera publiée au Journal officiel, ce qui permettra d'aller sur le terrain.

Christine LAGARENNE, Secrétariat général du Cnis

Cette pratique relève du décret du Cnis qui demande de donner un avis, notamment sur le programme des enquêtes dans le cadre des missions de la commission.

François GUILLAUMAT-TAILLIET

Je souhaite ajouter que le Cnis émet un avis et que l'autorité administrative décidera ensuite.

Anne EPAULARD

Il ne faut pas négliger l'importance de notre avis. L'expérience récente du recensement, où une question avait fait débat dans la presse et au niveau politique, montre que le fait que l'enquête et le questionnaire aient suivi tout ce processus de discussion publique au Cnis a été très utile pour sa défense par l'Insee.

Bien que notre avis ne soit que consultatif, il est important que nous jouions ce rôle de façon éclairée, en sachant qu'il peut avoir des conséquences. J'imagine cependant que pour le crédit-bail, les enjeux ne sont pas aussi importants que pour le recensement et la question de l'origine des parents.

Tanguy MARTIN, Insee, Département des comptes nationaux

Un [document](#) est projeté

Je sollicite le renouvellement de l'avis d'opportunité pour l'enquête « crédit-bail ».

Le crédit-bail est un mode de financement hybride entre location et crédit existant depuis 1962, défini comme tout contrat de location comportant une option d'achat. On parle parfois (avec abus) de leasing. Ce dispositif est avantageux fiscalement puisqu'il est possible de déduire la TVA sur les loyers assimilés à des charges et permet aux particuliers de changer facilement de véhicule.

Dans ce système, le crédit-preneur (entreprise, ménage ou administration) fait appel à un crédit-bailleur (banque ou société financière) qui achète le bien et le lui loue. Le preneur verse des redevances pendant la durée du contrat. À l'échéance, deux options se présentent : soit acquérir le bien en payant une valeur résiduelle, soit le restituer au bailleur.

Notre enquête cible un secteur très réglementé, car les crédits-bailleurs ont besoin d'un agrément. Une centaine de sociétés sont habilitées à proposer du crédit-bail en France et la plupart sont membres de l'Association des sociétés financières (ASF) et donc connues de la BdF avec qui nous communiquons beaucoup sur notre méthodologie et les intérêts de cette enquête. Sur les 112 entreprises recensées, certaines ne gèrent que des contrats existants, et plusieurs étant des comptes financiers, elles peuvent

mutualiser leurs réponses, ce qui nous amène à 80 contacts effectifs. L'enquête est exhaustive et annuelle, avec une collecte d'avril à juin en N+1. Nous n'avons donc pas d'incertitude due au sondage et peu de redressements. Le nombre faible d'interlocuteurs permet de nouer avec eux une relation de confiance entre enquêteurs et enquêtés, simplifiant la collecte de données qualitatives. De plus, nous utilisons un système de dépôt-retrait via le portail Coltrane où les entreprises complètent directement un fichier Excel.

Nous les interrogeons sur les nouveaux contrats, les montants investis, les types de produits loués, les secteurs d'activité des clients pour les entreprises, et les investissements ventilés par région. Nous nous intéressons également aux loyers perçus et au montant des achats.

Récemment, nous avons optimisé le questionnaire en supprimant les questions obsolètes, prenant en compte la charge de réponse pour les enquêtés. Nous posons notamment des questions sur le chiffre d'affaires qui est en général calculé par un autre service que celui qui répond au questionnaire, complexifiant la collecte. La simplification du questionnaire a également permis de passer d'une ventilation départementale à régionale, répartie selon deux grands types de crédits-bails (mobilier et immobilier) pour répondre à la demande des entreprises. Enfin, nous avons ajouté des questions en raison du développement du crédit-bail chez les ménages, notamment pour l'électroménager. De plus, nous améliorons la qualité de nos redressements en demandant le partage entre intérêts et remboursement du principal.

Nous publions une information rapide en début d'année qui détaille les résultats de l'enquête. Ce document est très attendu par beaucoup de répondants qui les utilisent pour des études de marché afin de se positionner sur le territoire. Par ailleurs, les nouveaux contrats de crédit-bail représentent 44,5 milliards d'euros.

L'enquête répond à un règlement européen qui impose au département des comptes nationaux certains retraitements spécifiques fondés sur la distinction entre propriété juridique et propriété économique. Au lieu d'identifier des banques comme les possesseurs des biens, nous allons considérer ces actifs comme appartenant à leur propriétaire économique, c'est-à-dire ceux qui les utilisent quotidiennement. La TVA est également récupérée et transmise au Trésor qui s'en sert pour actualiser son modèle TVA.

Les comptes nationaux se servent de cette enquête pour obtenir et caractériser l'investissement des entreprises et administrations publiques, ainsi que la consommation des ménages en crédit-bail. Cette information est difficile à identifier sans l'enquête, car elle est absente des liasses fiscales notamment. Nous en avons également besoin pour corriger les loyers déclarés dans les sources fiscales et administratives. Le crédit-bail est également l'objet d'une correction spécifique pour calculer le déficit des administrations publiques, calculer la dette au sens de Maastricht et récupérer la TVA.

Pour résumer, bien que cette enquête soit peu conséquente, elle nous impacte fortement. C'est pour cette raison que le département des comptes nationaux sollicite aujourd'hui le renouvellement d'avis d'opportunité pour les cinq années à venir.

Emmanuel GERVAIS

J'ai une question concernant le périmètre. Couvrez-vous uniquement le crédit-bail au sens strict, ou incluez-vous également la location financière, où la valeur actualisée des loyers couvre au moins 90 % de la valeur du bien, mais sans option d'achat ?

Tanguy MARTIN

Cet élément est exclu de l'enquête. Je précise que nous nous intéressons uniquement aux contrats avec option d'achat. Le terme « leasing » peut porter à confusion, car il désigne parfois la location financière ou la location longue durée, qui ne font pas partie de notre champ.

Emmanuel GERVAIS

En comptabilité privée, ces notions sont souvent mélangées.

Roxane SILBERMAN

Merci pour cette présentation. Je souhaite d'abord souligner l'importance des avis d'opportunité, qui permettent d'impliquer les utilisateurs pour améliorer les enquêtes, plusieurs ayant été significativement modifiées suite à ce processus.

J'ai ensuite plusieurs questions. Premièrement, cette enquête étant utilisée par un certain nombre d'utilisateurs, ont-ils été consultés sur les évolutions envisagées ? Deuxièmement, les données européennes sont alimentées différemment selon les pays. Existe-t-il des équivalents à l'enquête française dans d'autres pays européens ? Enfin, l'enquête permet-elle de situer en volume ce mode d'acquisitions pour un type de bien donné au regard de l'ensemble des modes d'acquisitions de ces biens ?

Tanguy MARTIN

Sur les consultations, notre approche est différente des enquêtes classiques. Nous nous appuyons principalement sur l'ASF, qui regroupe une grande partie des entreprises crédits-bailleuses, ainsi que sur l'expertise de la Banque de France et en particulier du Service d'études et statistiques des opérations financières (SESOF). Nous les avons consultés directement sur les évolutions envisagées, leur faisabilité et leur intérêt. Récemment, l'ASF nous a confirmé par écrit qu'ils apprécient beaucoup la ventilation régionale et sont satisfaits qu'un organisme public s'intéresse au crédit-bail.

Concernant la dimension européenne, notre travail s'inscrit dans le cadre du Système européen des comptes (SEC) pour harmoniser notre comptabilité nationale avec celle des autres pays. Chaque pays doit cependant faire ce genre de retraitement selon son propre système d'information, sans nécessairement avoir une enquête spécifique sur le crédit-bail. Il me semble qu'il en existe une au Royaume-Uni, mais je n'en suis pas certain.

Par ailleurs, la section « consommation des ménages » s'intéresse à la catégorisation du volume d'achats de voitures en crédit-bail ou par d'autres moyens et suit les données de crédit-bail pour le leasing automobile, ainsi que les achats classiques de voitures, en distinguant bien les deux.

Céline CRAVATTE, responsable de la commission

Le Comité du Label avait souligné lors de l'avis de conformité, cinq ans auparavant, l'importance de concerter avec les principales organisations patronales. En novembre, nous avons donc reconsulté, via les membres du Cnis, le Mouvement des entreprises de France (MEDEF), la Confédération des petites et moyennes entreprises (CPME) et l'Union des entreprises de proximité (U2P) pour recueillir leurs réactions et suggestions d'évolution. L'U2P nous a demandé davantage d'informations par taille et catégorie juridique. L'ASF nous a également confirmé par courrier l'importance pour eux qu'une autorité indépendante produise ces données sur le marché.

Plus généralement, les fiches descriptives des enquêtes sont diffusées très en amont avec une large communication. Tout utilisateur peut nous contacter pour réagir, pour qu'en cas d'omission de certains utilisateurs pertinents, ils puissent se saisir de cet outil afin que leurs demandes soient relevées.

Pour cette enquête, nous n'avons pas reçu d'autres demandes spécifiques, hormis celle de l'U2P. Concernant la CPME et le MEDEF, je ne sais pas s'ils ont fait un retour.

Tanguy MARTIN

Ils ont indiqué qu'ils feraient potentiellement un retour, mais ne sont pas revenus vers moi.

Emmanuel GERVAIS

Je trouve cette enquête essentielle. Du côté monétaire et bancaire, le crédit-bail est traité comme une opération financière et n'est pas enregistré comme en comptabilité privée. Cette enquête assure donc une bonne harmonisation entre les différentes données pour ce type d'opération.

La commission Système financier et financement de l'économie émet à l'unanimité un avis d'opportunité favorable sur l'enquête crédit-bail.

THÈME CENTRAL : QUELLES DONNÉES RELATIVES POUR LES STATISTIQUES DE COURT TERME ?

1. Un panorama de l'utilisation des données « alternatives » pour l'analyse économique

Un [document](#) est projeté.

Marie LECLAIR

Ce panorama a été réalisé à la demande de la commission, en essayant de l'élargir au-delà de l'Insee, même si mes connaissances pouvaient être limitées et que je n'avais pas de compétences spécifiques. Je me suis donc beaucoup appuyée sur l'important travail de préparation du secrétariat de cette commission et sur un dossier à paraître fin mai dans le Courrier des statistiques concernant les données privées. J'ai également utilisé mon expérience personnelle, ayant été chef de la division des prix à la consommation lors de la mise en place des données de caisse et du webscraping, ainsi que mes expériences antérieures dans la statistique d'entreprise où nous utilisions déjà certaines données alternatives.

a. Qu'entend-on par données alternatives ?

Je souhaite d'abord souligner un glissement sémantique. Dans les années 2010, nous parlions plutôt de « big data » que de « données alternatives ». Ce changement lié au développement de l'économie numérique qui multiplie les traces numériques des agents économiques et une capacité croissante à stocker et analyser ces traces, apparaissant comme un matériau riche que les statisticiens pourraient exploiter.

Les big data se caractérisent par le volume de ces données largement supérieur à celui des données traditionnelles, mais aussi par des formats plus complexes et variables, contrairement aux bases de données classiques des statisticiens comprenant des observations et des variables bien définies. Dans les années 2010, le terme de « big data » était beaucoup utilisé, car les enjeux étaient principalement techniques et informatiques avec la question de savoir comment traiter ces données.

Pour l'Insee, le premier exemple fut l'utilisation des données de caisse, qui étaient extrêmement volumineuses, mais sans format particulièrement complexe et variable. Ce fut le premier moment où l'Insee a été capable de traiter des données aussi volumineuses.

Aujourd'hui, nous parlons moins de big data, avec un premier glissement sémantique vers la notion de « données privées ». La difficulté technique ayant rapidement été surmontée, elle réside désormais dans l'apparition d'un nouveau type de données qui s'ajoute aux sources traditionnelles des statisticiens (enquêtes, recensements et données administratives). L'accent est donc mis sur la propriété des données plutôt que sur la difficulté technique.

Les données auxquelles nous faisons référence sont des données qui ont déjà été prétraitées ou préagrégées par le producteur de données privées, n'impliquant aucune difficulté technique de traitement par le statisticien. Ces données peuvent être par exemple des données bancaires du Crédit Mutuel et de La Banque Postale, des données de transactions bancaires du Groupement d'intérêt économique (GIE) Carte bancaire qui permettent de savoir quelles transactions ont eu lieu, des données de téléphonie mobile qui enregistrent la localisation des personnes, des données de plateformes de location courte durée comme Airbnb collectées au niveau européen qui apportent des informations intéressantes sur le tourisme, et données de consommation d'électricité utilisées notamment pendant la crise sanitaire et qui sont d'accès libre.

Aujourd'hui, nous ne parlons plus de données privées, mais de données alternatives, car cette notion va au-delà des données possédées par des acteurs privés. Il s'agit de nouvelles possibilités techniques permettant de collecter directement l'information et de créer des bases de données à partir de tout ce qui est disponible « librement ».

Les notions de base de données et de data mettent en avant le fait que pendant longtemps, les bases de données ne représentaient pas un actif, donc n'était pas un investissement pour les comptes nationaux. Le SEC actuel a permis de commencer à considérer les bases de données comme des actifs et donc un investissement potentiel des entreprises et des institutions. Le futur SEC introduira la notion de data, au-delà des bases de données, comme des phénomènes observables enregistrés, pas forcément directement exploitable, mais potentiellement valorisable en tant qu'actif. Ces éléments seront à nouveau présentés à la réunion de la commission du mois d'octobre.

Parmi les exemples de données alternatives nous entendons également les data :

- Le webscraping réalisé à l'aide de robots pour collecter de la data sur internet.

Cette technique est notamment utilisée par l'Indice des prix à la consommation (IPC) concernant les prix des trains ou des avions. Les travaux de la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) sont également réalisés à partir d'une collecte d'information sur les sites d'offres d'emploi. Bien que ces données soient juridiquement reconnues comme des données alternatives, leur méthode de collecte est quelque peu différente.

- L'analyse d'images

La BdF utilise par exemple des images satellitaires pour détecter l'activité des cimenteries et en déduire la production de ciment. Un projet de l'Insee et de la BdF travaille également sur l'exploitation de documents administratifs comme les comptes sociaux d'entreprises via reconnaissance automatique.

- L'analyse textuelle

L'analyse d'articles de presse, de recherches Google ou de tweets permet d'évaluer les anticipations économiques, notamment sur l'inflation.

La BdF travaille également sur les conventions collectives grâce aux analyses textuelles qui permettent de repérer les augmentations prévues de salaire afin de produire une estimation des hausses de salaire.

b. Quelle utilisation des données alternatives pour la production de statistiques de court terme ?

Le premier intérêt de ces données alternatives, précisé par le cadre légal, est qu'elles peuvent remplacer des collectes de données traditionnelles pour réduire la charge d'enquête. Cela permet de convaincre des entreprises de donner accès à leurs données, comme avec les données de caisse pour l'IPC et les données de transaction qui ont remplacé l'enquête sur le calcul des chiffres d'affaires dans le commerce. Ensuite, ces données alternatives permettent de produire des statistiques qui continuent à être représentatives face à l'évolution du contexte, notamment pendant la crise sanitaire, ou pour suivre des champs en transformation comme le *yield management* sur le prix des trains. Cela a également été le cas pour la location touristique où le statisticien ne peut pas enquêter auprès de tous les propriétaires de logement, l'accès aux données des plateformes permettant de couvrir cette activité. Les offres d'emploi qui étaient suivies par la DARES via Pole emploi, n'étaient pas non plus représentatives des offres en raison des nombreux sites qui en publient.

Ces données permettent également d'apporter une granularité et une fréquence qui serait impossible à produire avec des statistiques traditionnelles sans un coût excessif pour le service public et les personnes interrogées. La taille importante des échantillons permet d'étudier des sous-populations spécifiques, comme dans le cas des cartes bancaires. Nous bénéficions également d'une plus grande fraîcheur des données.

c. Les points d'attention quand on utilise des données alternatives

Ces données alternatives présentent néanmoins plusieurs points d'attention. D'abord, l'accès doit être pérenne et sécurisé. Le cadre juridique, qui sera présenté par Patrick Redor, doit être complété par des conventions et partenariats avec les producteurs privés pour garantir la transmission des données sécurisée dans un format exploitable. De manière plus générale, la pérennité des structures privées est également un enjeu, car la disparition de ces structures, un changement de format ou de méthode de collecte pourraient remettre en cause la production statistique. De plus, la suppression d'une taxe par exemple pourrait impacter le système statistique.

Le second point d'attention est la question du champ et des concepts. Ces données ne sont pas toujours représentatives. Sur les données bancaires, si nous n'avons pas accès aux données de plusieurs banques, il se peut qu'elles ne soient pas suffisamment représentatives. L'Insee n'ayant qu'accès qu'à celles du Crédit Mutuel, a étendu son enquête à celles de la Banque Postale. Le constat est le même pour la téléphonie mobile selon les opérateurs couverts.

Les concepts ne correspondent pas toujours à ceux des statisticiens. Une transaction bancaire n'est pas exactement une consommation. Sur les unités statistiques, nous observons des regroupements de comptes plutôt que des ménages. Nous retrouvons ces éléments sur les données administratives. Il faut donc

disposer de métadonnées et échanger avec les producteurs de données pour comprendre le cadre de production et l'utilisation première comme donnée de gestion.

Un autre problème concerne la stabilité et la comparabilité de ces données dans le temps et l'espace. La relation entre ce qui est observé et la variable que veut mesurer le statisticien n'est pas forcément stable dans le temps ou l'espace. Par exemple, une augmentation des paiements par carte n'équivaut pas forcément à une hausse de la consommation. Cela peut aussi être le cas pour les données mobiles en fonction de la couverture géographique de l'opérateur. La saisonnalité et la volatilité de ces données à haute fréquence sont également problématiques hors période de crise majeure.

Le traitement de ces données a un coût réel dû au besoin de *data scientists* et d'infrastructures pour recevoir des données de formats différents. La question de la confidentialité soulève également des interrogations sur l'acceptation sociale de la transmission de ces données. Bien que les statisticiens publics sachent traiter la confidentialité des données, la réticence de certaines personnes peut limiter leur capacité à les mettre à disposition de tiers tels que des chercheurs.

d. Conclusion

En conclusion, la crise Covid a été un catalyseur facilitant l'accès à certaines données privées et poussant les statisticiens à prendre plus de risques dans leurs productions. De nombreuses études ont été poursuivies afin de comprendre les limites de l'interprétation de ces données. Plusieurs publications ont résulté de ces travaux, notamment des blogs de l'Insee et le *Courrier des statistiques*. Ces données sont donc utiles, mais doivent être vues comme un complément aux sources traditionnelles. Les travaux européens sont par ailleurs essentiels pour partager les expériences et développer des méthodologies communes pour exploiter ces données alternatives. Le niveau européen facilite également l'accès des données à la fois de manière réglementée et dans les négociations avec les plateformes pour avoir accès à leurs données.

Anne EPAULARD

Merci beaucoup pour cet exposé très clair.

Céline CRAVATTE

Je signale que les références évoquées sont disponibles sur la page de la Commission. Le recensement n'est pas exhaustif, mais déjà conséquent.

Roxane SILBERMAN

Récupérez-vous les données quand vous travaillez sur les données privées ? Par exemple, quand vous travaillez sur les données des opérateurs mobiles ou sur les données de caisse, récupérez-vous les données ou travaillez-vous sur leurs systèmes ? Cela change la nature du problème, pas seulement pour la transmission, mais peut-être aussi pour leur traitement.

Marie LECLAIR

Je pense que c'est très variable selon les cas. Pour les données de caisse, elles sont récupérées par l'Insee, mais agrégées par établissement et par jour. Concernant la téléphonie mobile, je ne me souviens plus exactement. Parfois nous travaillons dans les locaux des entreprises pour les données privées. Pour le GIE, ils récupèrent des données agrégées.

Mélina HILLION, Insee

Je suis en charge notamment des données alternatives, surtout en transaction par carte bancaire et téléphonie mobile. Concernant les données de téléphonie mobile, leur utilisation a beaucoup évolué depuis 2016. Les premiers partenariats, de nature recherche, ont été établis avec Orange, Eurostat et l'Insee. Nous pouvions accéder aux données d'Orange uniquement sur place, lors d'une journée dédiée. Ces données très agrégées concernaient les connexions au réseau pour chaque appel et SMS. Nous n'avions pas accès à ces données sur notre infrastructure. Ce partenariat s'est terminé en 2020.

Pendant la période Covid, nous avons bénéficié d'un accès exceptionnel à des données agrégées au niveau départemental sur les connexions au réseau de trois opérateurs (Orange, SFR et Bouygues). Plus

récemment, l'Insee a accès à des données très agrégées issues d'une offre commerciale d'Orange, dans le cadre d'un partenariat de recherche financé par l'Agence Nationale de Recherche. Ce partenariat est limité à un périmètre géographique restreint sur le mois de septembre 2022, et uniquement pour des travaux de recherche exploratoires. Ces données ne sont absolument pas au niveau individuel, mais sont des données de comptage de populations présentes par tranches horaires et zones agrégées.

Pour les transactions par carte bancaire, plusieurs sources existent à l'Insee. Dans le cadre du partenariat avec le GIE Carte bancaire et de la chaire Sciences Digitales, nous avons accès aux données individuelles des transactions, mais seul le doctorant encadré par l'Insee et Télécom peut y accéder pour ses travaux. Il existe aussi un partenariat entre le Crédit Mutuel, la Banque Postale et l'Insee, qui donne accès au département d'études économiques à des données agrégées de comptes pour un échantillon d'environ 200 000 clients.

Anne EPAULARD

Merci beaucoup. Hélène Paris complétera ces questions lors de sa présentation.

Jean-Philippe GASPAROTTO, CGT

J'ai une question concernant les analyses textuelles des conventions collectives en vue de la prévision des évolutions salariales. La CGT est sceptique sur ce point. Que considère-t-on comme convention collective ? Les accords NAO sont-ils inclus ? Beaucoup de salariés et d'entreprises ne sont pas couverts par des accords résultant de négociations collectives, ce qui rend la corrélation avec la prévision des évolutions de masse salariale questionnable.

Concernant les offres d'emploi, la CGT a réalisé plusieurs analyses empiriques qui ont révélé de nombreuses pratiques frauduleuses. Dans le cadre de cette démarche, est-ce qu'une analyse qualitative des offres d'emploi est prévue pour détecter ce type de problèmes ?

Emmanuel GERVAIS

La Banque de France a publié un billet en mars sur son blog exploitant les résultats de ces travaux sur les conventions collectives, incluant les NAO. Notre intérêt est de mesurer le lien avec l'inflation et ses effets réciproques entre salaires et prix. Nous souhaitons aussi comparer les pratiques au sein de la zone euro, où les rythmes de mise à jour sont différents (annuel en France, triennal en Allemagne). C'est un indicateur qui ne remplace pas les chiffres officiels, mais permet d'obtenir des résultats plus rapidement.

Philippe MUTRICY, Banque Publique d'Investissement (Bpifrance)

Existe-t-il une liste de toutes ces données alternatives quelque part ?

Marie LECLAIR

Non, le travail préparatoire du secrétariat de la Commission a justement été d'essayer de recenser toutes les collectes de données alternatives. J'ai annoncé des statistiques qui relevaient plutôt des données privées, notamment les transactions par carte bancaire et les données de téléphonie.

Anne EPAULARD

Pour répondre à une des questions de Monsieur Gasparotto, la DARES consacre énormément de temps au retraitement des données pour identifier les fausses offres d'emploi et repérer les doublons, bien plus que ce qu'ils avaient imaginé. Votre remarque est donc tout à fait pertinente. Ces nouvelles données posent des problèmes de qualité que nous n'avons pas avec les anciennes. Des avantages et des inconvénients sont remarqués.

Céline CRAVATTE

Je voulais signaler une question sur la conversation en ligne concernant l'accord du règlement européen à laquelle Monsieur Redor répondra et à propos de l'accueil du grand public au partage de ses données avec l'Insee.

Marwan MENAA, Banque de France

J'ai une question sur les offres d'emploi en ligne et leurs plateformes. Le travail de la DARES est considérable, mais je m'interroge sur la pérennité du webscraping étant donné que les sites évoluent constamment. Vous avez également évoqué une base à l'échelle européenne et je voudrais savoir quel est le nom de cette base de données. Par ailleurs, qu'en est-il des problèmes liés à l'évolution des conditions générales d'utilisation (CGU) des sites et à la légalité du webscraping ?

Marie LECLAIR

Je vais répondre selon mon expérience du webscraping en tant qu'ancienne chef de division des prix à la consommation. Dans le cadre du Cnis et du Label, nous avons présenté notre pratique de webscraping comme mode de collecte pour données relatives aux trains et avions. Nous envoyons un courrier aux différents sites pour les prévenir et leur demander de ne pas nous bloquer, tout en leur rappelant que nous sommes ouverts à un autre mode de collecte et leur obligation de répondre à l'enquête de l'indice des prix à la consommation.

Nous nous accordons tous sur le fait que cette méthode n'est pas très pérenne. Dans le cadre de l'IPC, elle a été choisie, car la solution précédente était encore moins acceptable. Des travaux ont été menés pour anticiper les blocages et trouver des alternatives si nous ne pouvions plus collecter par robot cette information. Surtout, nous établissons des négociations avec les détenteurs de données pour trouver des solutions moins intrusives, comme l'accès par API (Application Programming Interface). La Direction générale de l'aviation civile (DGAC) accède notamment aux API qui permettent de mettre à disposition les informations sur les prix des vols, ce qui est plus stable qu'un site internet. Nous vision davantage ce type de collecte en conservant le webscraping comme solution transitoire quand nous n'avons pas d'autre solution et que la collecte actuelle est insatisfaisante.

Patrick REDOR, Insee, Unité affaires juridiques et contentieuses (UAJC)

Selon moi, ce n'est pas aussi simple et immédiat. Rien dans la loi 51 ne prévoit ce mode de collecte. Dériver une obligation d'ouverture d'accès aux sites sur la base d'une obligation de réponse à une enquête n'est pas une évidence juridique.

Marie LECLAIR

La subtilité est que nous leur proposons dans ce cas un autre mode de collecte, qui serait de nous fournir directement l'information sur les prix qu'ils pratiquaient.

Patrick REDOR

La formulation utilisée doit être très précise pour proposer un nouveau mode de collecte sans l'imposer. La loi est claire : interroger une entreprise ou accéder à ses données via un site n'est pas équivalent. Cette distinction est importante. Il faut également être vigilant, car l'obligation de réponse est assortie de sanctions possibles. Cependant, une entreprise ne peut pas être sanctionnée pour son refus d'accès à ses données publiées sur internet.

Récemment, le Comité du Label a reconnu le dispositif de webscraping comme méthodologie de collecte alternative, mais sur la base du volontariat. Nous avons dû travailler avec une équipe, car des entreprises n'avaient pas répondu à une enquête. Nous leur avons alors demandé l'accès aux données qu'ils auraient dû nous transmettre via l'enquête par le biais de leur site. Notre travail consistait à délier la demande dans des formes qui ne donnent pas l'impression d'imposer ce webscraping sur la base de dispositions de la loi 51. Il faut être vigilant, car la notion d'obligation assortie au webscraping n'existe pas actuellement.

2. Utilisation des données privées par le Service statistique public : quel cadre juridique ?

Un [document](#) est projeté.

Patrick REDOR

a. Cadre juridique national

Le droit de la statistique repose sur deux textes fondamentaux : un texte en droit français et un texte au niveau européen. En droit français, la loi 51-711 sur la coordination, l'obligation et le secret en matière de statistique établit de nombreux éléments importants pour du service statistique public (SSP). Elle définit ce service en référence à l'Insee, mais également aux informations qu'il produit. Elle définit également l'obligation de réponse aux enquêtes statistiques d'intérêt général et le secret statistique sous deux formes : le secret associé aux enquêtes et le secret associé aux données collectées auprès d'autres administrations.

L'article 7bis est important, car il délie ces administrations de leurs obligations à l'égard du secret professionnel, sauf exception ou demande formulé par le SSP. Par ailleurs, depuis 2016 et la loi pour la République numérique, des dispositions ont été introduites, bien que spécifiques à l'accès aux données du secteur privé.

L'article 3bis prévoit que le SSP puisse exiger l'accès à des bases de données, mais sous la condition que ces bases se substituent à une enquête statistique obligatoire. Des préoccupations du législateur concernent le contrôle de la charge statistique, visant à l'alléger notamment pour les données de caisse.

Les autres conditions concernent le secret statistique qui s'attache à ces données d'origine privée. Tout un processus de concertation préalable est imposé au SSP avec une étude d'opportunité, un avis du Cnis et une décision ministérielle sous forme d'arrêté.

Ces informations collectées, à la différence des autres collectes du SSP, ne peuvent pas être rendues disponibles aux chercheurs notamment. Les dispositions du code du patrimoine ne peuvent pas s'appliquer à ces données. Les seuls accès possibles concernent des données agrégées qui ne permettent pas de réidentifier les organismes auprès desquels elles ont été collectées.

Le régime de sanctions spécifique est plus sévère que le régime courant des enquêtes statistiques (150 à 2 500 euros) s'élevant à 25 000 à 50 000 euros après avis du Comité du contentieux des enquêtes statistiques obligatoires.

b. Cadre européen

Jusqu'à présent, le règlement européen autorisait l'accès du système statistique européen (SSE), et donc des Instituts nationaux de statistique (INS), à des données privées. Une disposition évoquait les données privées comme une des sources qui pouvaient être mobilisées sans donner plus de précision. Le règlement 223 n'imposait pas d'obligation concernant ces données à la différence des données administratives en libre accès pour les INS des États membres.

La dernière révision du règlement européen 223 en décembre 2024 l'a élargi aux données privées. Ce règlement définissant les dispositions fondamentales en matière de statistiques européennes a introduit une obligation de permettre aux INS d'accéder à des données privées pour les États membres sous plusieurs conditions. La première est leur contribution à la production de statistiques européennes. Cette obligation ne s'applique pas si elles servent des intérêts statistiques nationaux dans des secteurs tels que la Défense. Ensuite, le recours à ces données doit être justifié par un besoin sans autre alternative et leur intérêt par rapport à la réduction de la charge statistique.

Eurostat est directement intéressé par cette disposition et le règlement prévoit qu'il puisse exiger d'obtenir des données auprès d'opérateurs privés, avec un régime de sanctions directement prévu dans le règlement. Une coordination entre Eurostat et les INS est prévue, notamment pour les opérateurs multinationaux. Eurostat pourrait intervenir dans certains cas à la place des INS. Il est cependant important de souligner que les seuls organismes autorisés par le règlement 223 à exiger l'accès à des données privées sont les INS, pas les autres autorités nationales comme les services statistiques ministériels français. L'article 7bis prévoit que ces derniers passent par l'intermédiaire de l'INS.

La phase de concertation sur ces évolutions a beaucoup profité de l'expérience française impliquant que de nombreux éléments soient en miroir de ceux de l'article 3bis. Nous retrouvons le principe suivant lequel les données doivent être nécessaires et proportionnées. Le règlement 223 reprend également des éléments du Règlement général de la protection des données (RGPD) concernant les traitements des données à finalité statistique.

c. Enjeux d'adaptation du droit national

Le règlement 223 prévoit que les États membres doivent prendre des mesures appropriées pour faciliter l'accès aux données publiques, telles que les sanctions administratives. Elles sont aujourd'hui prévues pour Eurostat, mais pas en droit national. L'article 3bis n'est pas complètement cohérent avec le dispositif de collecte du règlement 223, qui n'impose pas la condition de substitution à une enquête statistique obligatoire existante.

Il faut revoir la loi 51 pour introduire un régime de sanctions compatible avec l'élargissement prévu par le règlement 223. Ce règlement permet désormais de collecter des données privées sans la condition préalable d'une enquête obligatoire existante. Une révision de l'article 3 bis est prévue, portée par les prochaines lois Ddadue (loi portant dispositions diverses d'adaptation au droit européen). Nous espérons pouvoir, d'ici la fin de l'année, modifier la loi 51 pour l'adapter à ce nouveau règlement européen.

d. Conclusion

Ce nouveau règlement européen offre une perspective d'expansion, sous contraintes de finalité de statistiques européennes, pour exiger l'accès à de nouvelles sources. Les intérêts de la statistique européenne rejoignent en grande partie ceux de la statistique nationale, ce qui va enrichir cette dernière, avec toutefois la limite des domaines statistiques particuliers échappant à l'application du droit européen, comme la justice.

La loi 51 doit donc être complétée par un régime de sanctions avant de lancer les démarches. De nombreux projets pourront ainsi se développer. L'expérience des données de caisse montre que le processus d'entente ou de négociation est très lourd. Tout un travail reste à faire sur les conventions. Cette disposition juridique ne signifie pas que l'application concrète et les résultats tangibles, en termes d'élargissement des dispositifs statistiques, seront immédiats.

Anne EPAULARD, présidente de la commission

Je vous remercie pour la clarté de l'exposé. Nous n'imaginons jamais toutes ces subtilités juridiques en tant qu'économistes, nous avons tendance à nous jeter sur les données sans réfléchir à toutes les conséquences.

3. Une approche alternative pour estimer le nombre de Voyageurs étrangers en France et leurs dépenses

Un [document](#) est projeté.

Sarita de ALBUQUERQUE, Responsable de la ligne voyage à balance des paiements de la Banque de France

Nous allons vous présenter une nouvelle approche pour estimer le nombre de voyageurs étrangers en France et leurs dépenses dans le cadre de la ligne voyage de la balance des paiements. Cette dernière mesure le solde commercial lié au voyage, un élément très surveillé. En 1945, l'Office d'échange a publié la première balance des paiements, et en 1960, la Banque de France en est devenue responsable. De 1963 à 2001, les enquêtes aux frontières permettaient de suivre les flux des voyageurs étrangers en France. En 2004, suite à l'extension de l'espace Schengen, nous avons dû revoir cette méthode en créant l'enquête auprès des visiteurs étrangers, reposant sur des comptages manuels aux frontières et une enquête sur les dépenses de ces personnes étrangères en France ainsi que des informations qualitatives sur leur comportement.

Dès 2015, des expérimentations ont été réalisées sur les données de téléphonie mobile, mais nous n'étions pas suffisamment matures pour les conserver pour ce sujet. En 2022, l'enquête E.V.E s'est arrêtée après le Covid, ayant atteint ses limites. Notre objectif est aujourd'hui d'utiliser de nouvelles données comme celles de la téléphonie mobile et certaines du trafic aérien. La qualité des données de téléphonie mobile a beaucoup progressé depuis 2015 et suscite notre intérêt. D'autres pays comme l'Italie, l'Espagne et l'Estonie s'y intéressent également. Nous travaillons avec Orange, qui a construit un business autour de ces données.

Ces données sont intéressantes, car 95 % de la population mondiale vit dans une zone couverte par un réseau mobile, et 70 % étaient équipés d'un smartphone en 2023. De plus, les accords de roaming entre pays se multiplient et les opérateurs disposent de plus en plus d'informations dans leurs bases.

Pour commencer, nous ferons un bref rappel des anciennes. Nous vous parlerons ensuite des défis que représente l'estimation des dépenses des voyageurs étrangers en France et nous vous montrerons le potentiel des nouvelles bases de données.

Notre présentation s'intéresse uniquement aux dépenses des voyageurs étrangers en France, couvrant tout type de voyage (personnel ou professionnel, court ou avec nuitées), incluant l'hébergement, la restauration et les achats, mais excluant le transport international. La population des voyageurs étrangers est très hétérogène et vaste, ce qui la rend difficile à identifier. Le périmètre de la ligne voyage est plus large que celui de la statistique publique. Grâce aux données de téléphonie mobile, nous pouvons capter davantage qu'avec des vacations de comptage aux frontières et une enquête.

En 2023, la France était la première destination touristique mondiale en nombre d'arrivées avec 98 millions de touristes étrangers et 140 millions de voyageurs excursionnistes à la journée. Elle occupait le troisième rang mondial en recettes de voyages avec 66 milliards d'euros alors que les dépenses des voyageurs français à l'international s'élevaient à 52 milliards. Le poste voyage contribue positivement au solde des paiements courants de la balance des paiements de la France (14 milliards d'euros). L'importance du tourisme justifie donc l'existence d'un dispositif statistique proportionné.

La méthodologie datant de 2004 reposait sur trois types de données : des données de trafic sortant (résident et non résident) qui ne permettaient pas d'estimer le nombre de résidents, nécessitant des vacations de comptage aux frontières dont nous pouvions obtenir un échantillon d'environ 60 000 personnes et procéder à une enquête aux frontières pour récupérer les dépenses de ces voyageurs et des informations sur le comportement et les préférences de ces personnes.

Cette méthode présentait plusieurs problèmes : une population cible inconnue et difficile à atteindre, des efforts de collecte coûteux et lourds, une complexité du trafic routier aux frontières, des perturbations récentes (Covid, grèves, gilets jaunes) dégradant les résultats, des difficultés de recrutement d'enquêteurs, et les particularités géographiques de la France (énormément de frontières terrestres, des côtes maritimes...). Cette méthodologie a atteint ses limites, nous obligeant à revoir entièrement notre processus, soit l'estimation de la population cible et des dépenses.

Félix PRADELLE, Banque de France (BdF), Direction de la balance des paiements

Nous travaillons sur une nouvelle façon d'estimer ces recettes. L'enquête reste indispensable pour collecter certaines informations, mais nous remplaçons les vacations de comptage par de nouvelles bases de données pour identifier notamment combien d'Allemands ou d'Espagnols sont présents dans les aéroports. À partir de ces bases, nous construisons un plan de sondage pour mettre en place une enquête plus précise et ciblée. En 2026, nous disposerons des résultats des premières enquêtes et de l'extrapolation pour arriver un chiffre final plus précis.

Nous utilisons déjà des données de trafic routier, ferroviaire, portuaire et aérien, qui sont très intéressantes, mais manquent de précision, car elles fournissent seulement des nombres de véhicules ou d'avions, et sont agrégées mensuellement, ce qui est insuffisant pour construire un plan de sondage. Nous disposons également de données de paiement bancaire, mais elles sont trop imprécises et agrégées et ne captent pas les paiements en espèces ou par d'autres moyens. Nous utilisons également les données miroirs d'autres pays et les résultats de la précédente enquête qui, malgré ses imprécisions, offrent des ordres de grandeur de comparaison assez intéressants.

Parmi les nouvelles bases de données, la base de données la plus solide et innovante dont nous disposons est celle de téléphonie mobile, issue d'un partenariat entre la Banque de France et Orange. Son fonctionnement repose sur des antennes situées près des corridors d'entrée ou de sortie du territoire (frontières routières, aéroports, ports et Eurotunnel). Orange capte ainsi les cartes SIM franchissant ces points, nous fournissant des données précises sur les entrées et sorties. Nous obtenons également la nationalité de la carte SIM (assimilée à celle du visiteur), la date de sortie, le lieu de séjour sur le territoire et le corridor de sortie emprunté (avion, bateau, voiture...). C'est une base exceptionnelle, mais l'inconvénient principal est qu'elle compte des cartes SIM et non des voyageurs. Un travail de redressement est nécessaire, utilisant des informations sur la part de marché Orange et le taux d'utilisation des cartes SIM par touriste. Malgré cette difficulté, elle représente la meilleure base dont nous disposons.

La deuxième base dont nous disposons provient de ForwardKeys, un des leaders mondiaux de l'agrégation de données de voyageurs aériens. Elle complète les informations de la DGAC, fournissant des données relatives à l'estimation des nationalités des voyageurs, les dates d'entrée et de sortie, la durée de séjour et le motif du voyage (absent des données Orange). Nous savons également si une personne fait une correspondance à Paris ou reste sur le territoire. Cette base complète efficacement celle d'Orange pour couvrir la population cible des visiteurs étrangers en France.

Sarita de ALBUQUERQUE

Cette analyse nous a conduits à une stratégie combinant données de téléphonie mobile et données de trafic aérien de ForwardKeys pour estimer la population cible. Certaines informations comme le pays de résidence figurent dans toutes les bases, tandis que le motif du voyage n'est disponible que via ForwardKeys, et uniquement pour le transport aérien. Nous travaillons sur l'ensemble des modes de transport. L'estimation de la population cible par les données de téléphonie mobile semblait pertinente, mais nécessitait la vérification et le calibrage de ces données. Une enquête était donc indispensable pour obtenir des informations qualitatives comme le motif de voyage, les activités et les types de dépenses.

Félix PRADELLE

Notre approche consiste à utiliser la base Orange, très fiable malgré le biais du redressement entre cartes SIM et voyageurs. Nous isolons les voyageurs aériens dans Orange (identifiés par leur dernière localisation à un aéroport) et les comparons avec les données ForwardKeys pour mieux calibrer le redressement. Par exemple, pour une carte SIM brésilienne convertie en sept personnes, les données Forward Keys nous ont confirmé que le ratio d'un pour quatre était plus réaliste que celui d'un pour sept suggéré par Orange. Le calibrage de ces redressements grâce à Forward Keys, nous permet d'optimiser la base Orange pour tous les types de trafic. L'objectif est donc d'utiliser la base Orange en l'améliorant avec les autres bases à notre disposition afin de réduire les biais.

Les courbes annuelles des bases Orange et de trafic aérien sont très corrélées. De même, une forte corrélation existe entre les dépenses et les flux de voyageurs observés via Orange, confirmant la qualité de cette base de données.

Sarita de ALBUQUERQUE

Pour estimer la population des voyageurs étrangers en France, les données de téléphonie mobile constituent donc un outil précieux. Elles offrent un potentiel considérable sans toutefois remplacer les méthodes traditionnelles comme les enquêtes. Nous prévoyons une enquête en 2026. Ce travail nécessite des vérifications, des calibrages et une collaboration permanente avec Orange pour redéfinir certains concepts et mieux comprendre ces données.

Anne EPAULARD

Merci beaucoup pour cette présentation impressionnante. Il faut beaucoup réfléchir avant d'utiliser ces données et je suis convaincue de l'efficacité des redressements que vous avez présentés.

4. Quel apport des données alternatives pour le nowcasting ?

Un [document](#) est projeté.

Catherine Doz, Paris School of Economics (PSE) et Université Paris I, porteuse de la Chaire de Mesures de l'économie, Nowcasting

J'anime l'axe nowcasting de la chaire Mesurer l'économie à Paris School of Economics, un partenariat entre PSE, l'Insee et la Société Générale. Cette chaire comporte deux axes : le nowcasting et mesurer au-delà du PIB, en lien avec les questions de comptes augmentés à l'Insee.

Le terme « nowcasting » est une contraction de « now forecasting » et recouvre des pratiques de prévision conjoncturelle connues depuis longtemps. Il s'agit de prévisions réalisées pour une variable avant sa publication et pendant la période sur laquelle elle est mesurée. Ce terme est apparu dans la littérature académique dans les années 2000, au moment de l'émergence de nouvelles techniques de prévision, puis a été repris par les acteurs de la prévision à court terme. Pour certains, le mot nowcasting est réservé à des

prévisions qui sont obtenues par ces techniques économétriques nouvelles et automatisées, mais il n'est en fait pas nécessairement réservé à cela.

Le premier exemple de nowcasting est concerné par les prévisions mensuelles d'une grandeur trimestrielle. Si nous faisons des prévisions du taux de croissance du PIB du trimestre Q pendant les mois M1, M2, M3 de ce trimestre, nous faisons du nowcasting. Par exemple, le point de conjoncture de l'Insee du 7 février 2024 donnait une prévision du taux de croissance du PIB au premier trimestre 2024, correspondant à l'information disponible au premier mois. Cette prévision est actualisée dans la note de conjoncture du 14 mars 2024 (l'information connue est principalement de l'information de M2). Ces prévisions intègrent des approches variées et ne résultent pas de l'application d'un seul modèle.

Le deuxième exemple concerne les prévisions hebdomadaires ou infrahebdomadaires d'une grandeur trimestrielle, pratique plus courante aux États-Unis. Nous pouvons ainsi faire du nowcasting du PIB à une fréquence hebdomadaire chaque semaine du trimestre Q, ou plus fréquemment.

La Fed de New York publie un «GDP Nowcast» hebdomadaire, tandis que celle d'Atlanta publie un «GDPNow» qui est actualisé chaque fois qu'une des séries utilisées pour construire l'indicateur est mise à jour. Ces nowcasts s'accompagnent d'avertissements précisant qu'ils ne constituent pas des prévisions officielles, car ils sont fréquemment mis à jour et reposent sur des méthodes purement statistiques (modèles à facteurs dynamiques) sans intégration de jugements d'experts ni confrontation avec des prévisions utilisant d'autres méthodes. D'autres institutions peuvent utiliser le même type de techniques comme « benchmark » interne, sans les publier. Nous pouvons également faire du nowcasting en faisant des prévisions hebdomadaire ou infra-hebdomadaire d'une grandeur mensuelle.

Il faut noter qu'on peut aussi faire du « backcasting » pour les variables macroéconomiques qui sont publiées avec des délais variables suivant les variables. Par exemple, le PIB en France est publié à la fin du premier mois du trimestre suivant et l'Indice de production industrielle (IPI) mensuel est publié au début du 2e mois suivant le mois concerné. On peut donc continuer à faire des prévisions du PIB du trimestre Q pendant le premier mois du trimestre suivant, jusqu'à sa publication.

Les travaux académiques sur le nowcasting concernent le plus souvent le taux de croissance du PIB, mais les techniques peuvent évidemment être utilisées pour le nowcasting de toute autre grandeur macroéconomique (taux d'inflation, consommation des ménages, activité dans le secteur immobilier, commerce international...). Toutes les institutions publiques qui font de la prévision conjoncturelle (Instituts de Statistique, Banques Centrales, etc.) sont donc concernées par le nowcasting. Cependant, il est également pratiqué dans le secteur privé, par certaines entreprises qui en font pour leur usage interne, ou par d'autres qui vendent leurs prévisions.

Ce que je vais présenter est principalement centré sur ce qui est fait dans le secteur public, en France ou à l'étranger.

a. Données et méthodes utilisées

Les données généralement utilisées pour le Nowcasting sont les enquêtes de conjoncture, les indicateurs macroéconomiques nationaux et internationaux, ainsi que les indicateurs permettant de décrire les marchés financiers. Ces indicateurs n'ont pas tous les mêmes délais de publication ni la même fréquence de publication, mais les enquêtes de conjoncture qui sont mensuelles et publiées très rapidement jouent un rôle très important dans les prévisions, notamment en début de trimestre, en France et à l'international.

Quelle que soit la méthode employée, la prévision conjoncturelle vise à intégrer toute l'information disponible à la date où elle est réalisée. La prévision du PIB par l'Insee repose sur une prévision de ses composantes (consommation des ménages, production manufacturière, etc.), qui est basée sur les enquêtes de conjoncture et sur les données déjà publiées. Elle repose aussi sur des méthodes dites d'étalonnage, c'est-à-dire sur des modèles économétriques permettant de relier directement le taux de croissance du PIB aux résultats des enquêtes de conjoncture, à l'IPI et à la consommation de biens.

Il existe aussi en Nowcasting de approches purement statistiques et automatisées. Ces dernières reposent sur le même type de données, à la différence qu'elles ne hiérarchisent pas les indicateurs utilisés. Les modèles couramment utilisés sont les Modèles à Facteurs dynamiques (DFM), qui visent à extraire l'information commune d'un grand nombre d'indicateurs et à relier le taux de croissance du PIB à ces fluctuations communes. Ces modèles permettent de traiter facilement des données n'ayant pas la même fréquence ni les mêmes délais de publication. De plus, leur mise en œuvre est rapide permettant, aux Fed

par exemple, de publier fréquemment des prévisions. Les années récentes ont vu aussi se développer des travaux fondés sur les méthodes de machine learning, encore expérimentales.

b. Utilisation de nouvelles données

Le développement de modèles économétriques capables de prendre en compte un grand nombre de variables, l'apparition de nouvelles données liées à la digitalisation de nombreuses activités économiques, et l'essor des techniques de machine learning, conduisent à s'interroger sur l'apport potentiel de ces nouvelles données au nowcasting, mais aussi sur leurs limites.

La période du Covid, pendant laquelle les données classiques étaient mesurées de façon incomplète a conduit les Instituts statistiques à mobiliser d'autres types de données pour le suivi conjoncturel et a amorcé l'utilisation de ces nouvelles données. Depuis janvier 2020, l'Insee utilise les données de caisse des hypermarchés et supermarchés pour le calcul de l'IPC, mais, pendant la période du Covid, l'Insee a utilisé d'autres données haute fréquence (données de cartes bancaires, de consommation électrique, Google Trends, etc.) qui ont contribué à l'évaluation de l'activité. Certaines de ces données sont toujours fournies à l'Insee.

Par ailleurs, le département conjoncture utilise beaucoup les données de caisse des supermarchés et hypermarchés, en particulier pour en déduire la consommation alimentaire des ménages. La relation entre la consommation alimentaire des ménages et leur consommation totale étant très stable, la consommation des ménages peut ensuite être prévue puis intégrée dans la prévision du PIB. L'une des modalités de la prévision du PIB est en effet de prévoir ses composantes.

En revanche, les données de carte bancaire se sont avérées moins utiles après la période Covid pour le département de la conjoncture. Elles ont une forte volatilité. On ne connaît pas assez précisément le champ couvert (couvrent environ 75 % du marché, mais ce chiffre fluctue). Les données CB seraient plus facilement utilisables que si elles couvraient l'ensemble du champ.

c. Nouvelles données de nowcasting

Les travaux de nowcasting utilisant ces type de nouvelles données sont encore au stade exploratoire, car ils ont commencé juste avant la période du Covid. De nombreux travaux ont cependant été réalisés et il serait intéressant d'en faire un bilan, afin de prioriser les données à collecter et de faire aussi un bilan sur les méthodes utilisées pour les traiter (Machine Learning vs DFM par exemple).

Des travaux intéressants ont notamment été menés à la BdF, à la BCE, au KOF (Centre de Recherches conjoncturelles, Zurich), à European Commission Joint Research Center, à l'ESCOE (UK). Une analyse de ces travaux permet de tirer quelques conclusions :

- Les données Google Trends semblent utiles pendant les quatre premières semaines du trimestre, avant la publication de l'enquête de conjoncture. Ensuite, leur intérêt diminue face aux données plus conventionnelles. (Il faut cependant noter que, dans l'étude concernée, comme dans de nombreuses autres, les données hebdomadaires sont traitées en écart avec les données de la même semaine de l'année antérieure, et qu'il conviendra de mener des études en utilisant d'autres outils de désaisonnalisation)
- Les travaux du KOF sur le PIB de la Suisse montrent aussi un apport important des données haute fréquence en période de crise, mais pas dans les périodes normales. Ces données concernent l'énergie, le trafic routier et aérien, les données CB dans le commerce de détail, des données de recherche Google.
- Des travaux de nowcasting de la croissance du PIB mondial avec des données haute fréquence ont aussi montré que ces données n'ont un apport qu'en période de crise. Les travaux de la BdF utilisant les données haute fréquence de trafic maritime montrent cependant que ces données sont utiles pour le nowcasting du commerce mondial.

D'autres résultats concernent l'analyse textuelle. La BdF a mené des travaux sur l'analyse textuelle des appréciations jointes à l'enquête de conjoncture et a développé un indicateur de sentiments basés sur les commentaires des chefs d'entreprise dans ses enquêtes de conjoncture, permettant d'estimer l'impact d'événements (comme par exemple les gilets jaunes) sur le PIB, et potentiellement mobilisables pour le nowcasting.

Par ailleurs, un certain nombre de travaux se basent sur des données textuelles pour construire des indicateurs de sentiment basés sur des articles de presse (mais il faut noter malgré la lourdeur du traitement de données dans ce cas). Par exemple, laBdF a construit un indicateur de perception de l'inflation, qui contribue efficacement à la prévision de l'inflation ; le "joint research center" de la commission européenne (le JRC) a construit un indicateur de sentiment qui est utilisé pour prévoir le PIB européen.

d. Conclusion

Il serait intéressant de faire un bilan complet des travaux qui ont été menés en utilisant des données alternatives pour le nowcasting. Cela pourrait orienter le choix de données à collecter. Cependant, pour le nowcasting du PIB, il semble que ces données n'améliorent vraiment la prévision qu'en début de trimestre (avant la sortie des enquêtes de conjoncture), ou en période de crise. De plus, les questions de saisonnalité des données hebdomadaires ou journalières doivent être traitées.

Il faudra aussi explorer si les approches issues du machine learning surpassent les modèles à facteurs dynamiques, y compris dans les versions qui ont été proposées après le Covid pour traiter les questions de volatilité. Un travail comparatif reste à faire entre ces deux types de méthodes.

Mais toute amélioration dans la collecte de données nouvelles et pertinentes pour prévoir les blocs constitutifs du PIB bénéficiera au nowcasting, quelle que soit la méthode employée. Il y a cependant aujourd'hui un problème d'accès à ces données pour les chercheurs, et des projets comme notre chaire sont intéressants précisément parce qu'ils mettent en relation chercheurs et instituts statistiques.

Anne EPAULARD, présidente de la commission

Merci beaucoup.

5. Les apports de l'utilisation des données du Crédit Mutuel : l'expérience du CAE.

Un [document](#) est projeté.

Hélène PARIS

Je vais vous présenter très rapidement vu le temps imparti notre expérience dans l'exploitation des données de comptes bancaires grâce au partenariat avec le Crédit Mutuel Alliance Fédéral.

Ce partenariat est né pendant le premier confinement, grâce aux relations interpersonnelles entre l'ancien directeur général du Crédit Mutuel et notre président délégué de l'époque, Philippe Martin. Nous avons rapidement construit, avec les équipes informatiques, un échantillon de ménages de comptes de particuliers pour être opérationnels dès la sortie du confinement. Très vite, l'Insee nous a rejoints dans l'accès à ces données. Dans les mois suivants, nous avons évoqué l'intérêt d'étendre notre accès aux données de comptes d'entreprises et un plan de sondage rigoureux a été co-construit avec l'Insee, centré sur les très petites entreprises (TPE) et les petites et moyennes entreprises (PME), sachant que pour des raisons de confidentialité le Crédit Mutuel ne pouvait pas nous donner accès aux comptes d'entreprises de plus grande taille.

Ces données bancaires ne sont pas des données statistiques en soi, de même que les données administratives ne le sont pas. Nous cherchons donc à extraire l'intelligence de ces données pour approcher au mieux des catégories statistiques. Par exemple, pour les particuliers, nous essayons d'identifier ce qui s'apparente le plus au revenu disponible, et pour les entreprises, ce qui pourrait correspondre au chiffre d'affaires.

Notre motivation au CAE n'est ni de faire de la statistique publique, ni de la conjoncture, ni de la prévision, mais de réaliser des analyses économiques en exploitant cette nouvelle source de données, particulièrement là où il existe des lacunes dans la statistique publique.

Trois exemples d'utilisation de ces données :

- L'étude des petites entreprises face aux chocs des prix de l'énergie : nous avons pu mesurer précisément leur exposition hétérogène à ce choc.

- L'analyse des Prêts Garantis par l'État (PGE) : notre échantillon étant centré sur les TPE-PME, les principales bénéficiaires des PGE, nous avons pu compléter les études de la Banque de France et du Trésor en nous concentrant sur les entreprises ayant étalé au maximum leurs échéances de remboursement, pour évaluer leur risque de défaut.
- Une étude sur les comptes de particuliers, réalisée en collaboration avec le Crédit Mutuel, qui a apparié leurs données clients avec la base de l'Agence de la transition écologique (ADEME) sur les diagnostics de performance énergétique (DPE). Cela nous a permis de comparer la consommation énergétique théorique des logements avec leur consommation réelle, révélant des écarts significatifs et un enjeu important sur ce que nous pouvons attendre des opérations de rénovation des logements.

La liste d'une douzaine de publications s'appuyant sur l'exploitation de ces données bancaires et illustrant leur potentiel vous est affichée.

Anne EPAULARD

Merci beaucoup. Je vous invite à consulter les notes et focus du CAE pour plus de détails.

Hélène PARIS

Sur la question de la représentativité, le Crédit Mutuel est une banque qui représente assez bien la population moyenne française. Pour les entreprises, nous avons vérifié en comparant avec les données de statistique publique que notre échantillon reflète bien le tissu productif des petites entreprises par secteur d'activité et en termes d'exposition aux dépenses d'énergie. Cette représentativité nous conforte dans l'idée que ces données font sens.

Anne EPAULARD

J'avais insisté pour que cette présentation soit faite, car cette utilisation de ces données est assez différente et complète bien les autres. Nous avons vu que ces données peuvent être utilisées pour la balance des paiements ou le nowcasting, notamment pendant les périodes difficiles. Ce qui fait au CASD ajoute une dimension de l'utilisation de ces données qui peuvent servir à surveiller des politiques économiques ciblées sur des populations en temps réel. Au-delà de la prévision, nous pouvons mesurer l'efficacité des interventions de politiques publiques, ce qui illustre bien l'ampleur de ce qu'on peut tirer d'une utilisation raisonnée de ces données.

Merci beaucoup. J'ouvre maintenant la discussion sur les présentations précédentes.

Céline CRAVATTE

À propos de la question qui attendait, il était demandé si des réactions citoyennes manifestaient un désaccord ou une inquiétude vis-à-vis de ces partenariats.

Patrick REDOR

Dans le cadre des conventions établies, les données communiquées ne sont pas considérées comme des données à caractère personnel, car si des données individuelles sont communiquées, elles sont pseudonymisées de façon à éliminer tout élément d'identification comme les noms ou prénoms. Les variables sont agrégées pour limiter les risques de réidentification. Pour un organisme bancaire soumis au secret bancaire, qui communique des données suffisamment pseudonymisées pour qu'elles ne soient plus considérées comme données à caractère personnel, est dispensé d'en informer les personnes concernées. Dans les conventions, nous rappelons que chaque partie doit respecter ses obligations en matière de protection des données personnelles. La communication de données à caractère personnel constitue un traitement de données soumis au RGPD. Je n'ai pas connaissance de clients s'étant manifestés concernant ces conventions avec des organismes bancaires. Pour tous ces traitements de données, une information est disponible sur le site de l'Insee, accessible à tous. Toute personne peut contacter le contact du RGPD de l'Insee, c'est-à-dire moi-même, mais je n'ai jamais reçu de demande concernant ces opérations.

Roxane SILBERMAN

J'ai une remarque et un petit ajout à propos de l'intervention de Patrick Redor. Il me semble que le règlement européen 223 voit l'ouverture aux données privées (et aussi administratives) pas seulement comme pouvant remplacer des données d'enquêtes mais comme une source d'enrichissement pour la statistique publique. Les présentations d'aujourd'hui montrent bien que ces données vont au-delà du simple remplacement et que de nombreux acteurs autour de la statistique publique utilisent déjà ces données.

Deuxièmement, je voudrais signaler que l'article 23 de ce règlement européen 223 indique que les données administratives comme privées désormais dès lors qu'elles ont été mobilisées pour la statistique publique peuvent être également utilisées par les chercheurs. C'est une disposition qu'avait demandé avec insistance le European Statistical Advisory Committee (Esac), un peu l'équivalent du Cnis au niveau européen et qui a finalement été introduite. J'espère que cette disposition sera reprise dans les régimes nationaux, notamment dans la loi 51.

Anne EPAULARD

Merci pour cette intervention. Le dernier point que vous mentionnez m'intéresse particulièrement, car nous voulons favoriser l'usage par la statistique publique, mais aussi par les chercheurs.

J'ai par ailleurs une question pour la Banque de France concernant l'établissement des comptes du tourisme : la méthode présentée est-elle déjà mise en place ou s'agit-il d'un projet ?

Sarita de ALBUQUERQUE

C'est une méthode qui est en train d'être mise en place. Nous avons récupéré les nouvelles données, nous les analysons pour identifier les biais et les angles morts, ainsi que le recouvrement entre les différentes bases. Notre objectif est de valider ces données pour estimer notre population cible. Parallèlement, nous travaillons sur un plan de sondage représentatif de cette population, avec l'intention de lancer une enquête au deuxième trimestre 2026. Nous envisageons soit un simple sondage stratifié basé uniquement sur les données de téléphonie mobile, soit des méthodes plus complexes utilisant plusieurs bases de données et leurs chevauchements pour mieux estimer les recettes touristiques.

Félix PRADELLE

Pour compléter, l'enquête précédente, E.V.E, s'est terminée en 2022, avec 2019 comme dernière année vraiment fiable en raison du Covid. La prochaine enquête basée sur les nouvelles méthodes est prévue pour mi-2026, avec des résultats fin 2026. En attendant, nous conservons le cadre de l'ancienne enquête et nous appliquons des facteurs d'actualisation via les données bancaires pour continuer à fournir des chiffres officiels jusqu'à la prochaine enquête. Nous avons reçu les données d'Orange et de ForwardKeys depuis quelques mois et commençons à obtenir des résultats intéressants, tout en travaillant avec des prestataires sur le plan de sondage pour l'enquête de 2026.

Nicolas STUDER

Nous parlons toujours de données de partenariat et de convention, mais il faut savoir que ces données sont achetées. Le secteur privé a l'obligation de nous fournir ces données, mais pas gratuitement, et le prix est fixé par le fournisseur. Si ce prix est prohibitif, l'accès devient impossible. Je ne sais pas comment vous fonctionnez avec Orange, mais cet élément est important.

Félix PRADELLE

Nous achetons ces données.

Emmanuel GERVAIS

E.V.E coûtait déjà très cher. Nous devons comparer l'ancien coût et le nouveau afin d'optimiser les frais.

Nicolas STUDER,

Je comprends très bien que les entreprises doivent être rémunérées pour les traitements qu'elles réalisent, mais parfois il s'agit simplement d'une transmission d'informations.

Anne EPAULARD

Dans le cadre du Crédit Mutuel, je pense que l'accès à ces données est gratuit, mais je pense que c'est un point important qui était dans l'angle mort de notre discussion.

Catherine DOZ

C'est aussi ce qui empêche les chercheurs d'accéder à ces données, leurs institutions n'ayant pas les moyens de les acheter.

Roxane SILBERMAN

Cette question des prix est en effet centrale et a fait l'objet de beaucoup de discussions pour le règlement 223.

Anne EPAULARD, présidente de la commission

Merci pour cette précision.

François GUILLAUMAT-TAILLIET donne lecture du texte suivant :

« La Commission Système financier et financement de l'économie a pris connaissance de travaux mobilisant les données alternatives pour améliorer les prévisions et les diagnostics économiques et aider les décideurs publics à s'adapter au mieux à la situation économique et sociale. Elle avait déjà abordé ce thème au printemps 2021, en lien avec les innovations mises en place dans le cadre de la crise COVID.

Par données alternatives, on entend celles qui diffèrent des enquêtes statistiques ou des bases administratives, données liées au big data, données privées et accessibles à un niveau plus ou moins granulaire, données de format varié, webscraping, données textuelles, images en particulier satellites. Ces données sont rapidement disponibles, de haute fréquence, avec une granularité très fine, un coût de collecte faible, mais un coût de traitement élevé, et leurs potentiels diffèrent selon chaque situation et chaque source de données.

Les données alternatives peuvent être utilisées pour alléger la charge d'enquête sur les répondants, faire face à des situations exceptionnelles, mesurer des phénomènes mal documentés auparavant, réduire les délais de production et effectuer de la prospective.

Pour la production statistique courante, elles ont permis un remplacement partiel d'enquête et des gains d'informations (données de caisse pour l'indice des prix à la consommation et les indices de chiffre d'affaires, données de transactions bancaires pour les indices de chiffre d'affaires provisoires, données de téléphonie mobile pour la ligne dépenses des voyageurs de la balance des paiements). Pour la prospective, les pistes explorées montrent que ces données peuvent permettre parfois d'anticiper les résultats des enquêtes, mais ne remplacent des enquêtes que dans des situations exceptionnelles où elles ne peuvent avoir lieu. Pour l'analyse, elles permettent d'explorer finement, avec des délais réduits et d'éventuels appareillages, des domaines peu documentés (capacité de remboursement des très petites entreprises, dépenses énergétiques déménagent grâce aux données de compte bancaire appareillées au diagnostic de performance énergétique du logement, variation infra-annuelle des dépenses des ménages).

La Commission encourage la poursuite des travaux de la statistique publique permettant d'explorer et de mobiliser ces sources prometteuses. Elle note que la révision récente du règlement statistique européen 2023 ouvre des perspectives d'extension du droit d'accès aux domaines privés sous la contrainte de finalité statistique européenne et nécessite une adaptation nationale de la loi de 1951. Elle rappelle les conditions nécessaires à leur utilisation par la statistique publique en complément des critiques traditionnelles qui restent centraux : accès terrain aux données, possibilité de contrôler le champ, la représentativité et ses évolutions, cohérence des données avec les concepts usuels.

Dans le cas de données privées, des échanges étroits et continus avec les détenteurs des bases de données sont nécessaires dans le respect de la confidentialité et du secret des affaires. La Commission rappelle donc le caractère essentiel de la mise en place de mutualisations et de partenariats entre le service statistique, mais aussi avec les fournisseurs de données privées et des chercheurs au niveau national et européen. »

Marie LECLAIR

Vu la précision apportée par Nicolas, il me semble que la formulation sur les coûts convient toujours pas. Enfin, je me demande même s'il faut parler des coûts, parce que c'est très discutable.

François GUILLAUMAT-TAILLIET

D'accord, on supprime " un coût de collecte faible, mais un coût de traitement élevé"

Roxane SILBERMAN

A la dernière phrase, je préférerais "les chercheurs" et non "des chercheurs"

François GUILLAUMAT-TAILLIET

C'est corrigé

Christine LAGARENNE

Je me demande si on pourrait reprendre la recommandation de Catherine Doz de faire un bilan complet des études sur le nowcasting, qui permettrait d'identifier les données auxquelles on souhaite accéder de manière privilégiée.

Anne EPAULARD

Cela évolue tellement vite que le bilan risque d'être rapidement caduque et devrait être fréquemment renouvelé. Je préfère donc ne pas graver cette obligation dans le marbre *et expliciter les bases des données dont on a besoin, dont le champs va par définition s'élargir.*

Jean-philippe GASPARETTO

Je suis un peu surpris par la référence au secret des affaires dans l'avant dernière phrase car le respect de la confidentialité suffit à couvrir l'ensemble des précautions nécessaires.

François GUILLAUMAT-TAILLIET

C'est un bon point, merci.

Le texte final est approuvé suite aux modifications apportées par les membres de la commission.

Anne EPAULARD

Je rappelle que vous pouvez retrouver les documents sur le site de la commission. La prochaine réunion aura lieu le 6 novembre après midi. Je vous remercie tous de votre participation.

La séance est levée à 12 heures 44.

DOCUMENTS PRÉPARATOIRES ET COMPLÉMENTAIRES

[Quelques travaux existants à propos des sources de données alternatives pour les diagnostics de court terme.](#)

[Bilan 2024 des producteurs de la commission](#)

[Programme de travail 2025 des producteurs de la commission](#)

[Enquête Crédit-Bail - Fiche descriptive.](#)

[La révision de la nomenclature d'activités A17 Leclair Insee.](#)

[L'enquête Credit-Bail Martin Insee](#)

[Un panorama de l'utilisation des données alternatives pour l'analyse économique Leclair Insee.](#)

[L'accès aux données privées par le SSP : cadre juridique Redor Insee](#)

[Nouvelles sources de données : une approche alternative pour estimer le nombre de voyageurs étrangers et leurs dépenses en France De Albuquerque Pradelle Banque de France.](#)

[Quel apport des données alternatives pour le Nowcasting Doz PSE.](#)

[L'apport de l'utilisation des données du Crédit Mutuel Paris CAE.](#)

AVIS D'OPPORTUNITÉ

[Avis d'opportunité Enquête Crédit-Bail](#)