

Commission *Système financier et financement de l'économie* du 18 novembre 2020

Présidente : Anne-Laure Delatte

Synthèse

À situation exceptionnelle, ordre du jour exceptionnel : la séance commence par une information sur **les conséquences de la crise sanitaire sur le programme de travail 2020 (PAT 2020) de la statistique publique** concernant la commission **et sur la gestion des avis d'opportunité (AO)**.

La crise sanitaire débutée au printemps a contraint à aménager le PAT 2020 des producteurs de la statistique publique qui avait été mis en ligne à la page de la commission au printemps et à innover. Le dialogue resserré entre la Banque de France et les banques ont permis le respect des échéances et le maintien de la qualité de l'information. Une page dédiée à la crise sanitaire, « Suivi de crise Covid-19 », a été créée sur le site de la Banque de France, une note mensuelle de statistiques sur la situation financière des ménages et des entreprises a été publiée. Dans la volonté de maintenir la fiabilité de la cotation des entreprises, la Cotation 2020 a été différée et une enquête auprès des entreprises sur l'impact de la Covid-19 sur l'évolution des agrégats financiers et de la liquidité, les mesures de soutien utilisées, etc., a été lancée en juillet. Enfin, des articles du *Bulletin* ont fait le point sur la situation d'avant-crise et les billets du blog ont suivi l'actualité. L'ACPR, l'Autorité de contrôle prudentiel et de résolution, a intensifié l'exploitation de ses bases de données et les échanges avec les principaux organismes pour analyser de façon très rapprochée l'impact de la crise sur les banques et les assurances. Avec les autorités de supervision européennes, elle a assoupli les conditions de reportage des remises prudentielles lors du premier confinement. De leur côté, face à une crise inédite, les comptables nationaux de l'Insee ont dû s'adapter afin de continuer à produire une mesure robuste du PIB : trouver d'autres indicateurs de court terme, s'appuyer sur les travaux du département de la conjoncture de l'Insee pour la première estimation à T+30 jours, adapter la méthodologie, mobiliser des données administratives pour les comptes d'agents, en particulier pour les activités non marchandes qui n'avaient pas de suivi infra-annuel. La comparabilité entre les États membres de l'Union européenne a d'ailleurs été rendue plus difficile par cette prise en compte du secteur non marchand diverse selon les pays, malgré la production d'un guide méthodologique par Eurostat. Points et notes de conjoncture réguliers, articles dans le blog de l'Insee ont permis de maintenir le niveau d'information publique sur ces impacts de la crise sanitaire sur l'économie.

Le Cnis est revenu sur les actions liées à cette crise sanitaire qui ont dû être menées en urgence afin de poursuivre la concertation malgré l'annulation des commissions de printemps et accompagner au mieux les producteurs dans les modifications du système d'information. Ainsi, une consultation électronique des membres des commissions a été mise en œuvre pour les examens d'AO : pour la commission *Système financier*, cela a concerné l'enquête sur le crédit-bail. La même procédure a été appliquée pour les demandes d'accès à des sources administratives au titre de l'article 7bis de la loi de 1951. Pour les enquêtes originales suscitées par la pandémie, enquêtes dites « Covid » (par exemple, l'enquête de l'Insee Impact sur les entreprises), le Cnis a mis en œuvre une procédure d'urgence en conformité avec les textes réglementaires avec une communication particulière sur son site pour chaque opération.

Une des grandes innovations accélérée par la crise sanitaire a été **le recours aux données massives pour le suivi conjoncturel lors du confinement par l'Insee**.

Pendant la crise, les enquêtes mensuelles de conjoncture ont continué à être collectées, même si le taux de réponse a diminué notamment en avril. Mais ces enquêtes ne sont pas adaptées lorsque la situation devient extrêmement évolutive, comme au mois de mars. Le recours aux données massives est vite devenu indispensable pour compléter l'information. Les données utilisées résultent de l'accroissement des traces numériques des activités humaines, dans des champs de plus en plus étendus, elles sont donc multiformes : données de caisse de la grande distribution, transactions par cartes bancaires, connexions sur les réseaux des opérateurs de téléphonie mobile, requêtes des moteurs de recherche sur le web, services de géolocalisation présents sur nos smartphones, consommation d'électricité, trafic routier, images satellites, etc. Elles sont disponibles rapidement, « homogènes » d'un pays à l'autre, avec un coût marginal proche de zéro et une charge de collecte faible voire insensible pour les particuliers ou les entreprises concernées. Avant la crise, certaines de ces données étaient déjà utilisées par la statistique publique pour la production statistique, notamment les données de caisse de la grande distribution pour la mesure de l'inflation, et, pour les conjoncturistes, elles constituent un substitut utile aux données traditionnelles... tant que celles-ci ne sont pas encore disponibles. Le suivi de deux agrégats macroéconomiques a été privilégié : le PIB et la consommation des ménages. Les données massives ont contribué à un diagnostic conjoncturel assez fin de l'activité et de la consommation par secteur économique, permettant de différencier les deux chocs de mars et de novembre. Au final, certains outils ont aidé à parer à l'urgence, d'autres ont sans doute un potentiel plus pérenne, en étant attentifs à leur processus de production et à leur représentativité. L'enjeu est ardu : tenter de mesurer les phénomènes en temps quasi réel ET en rendant compte de l'hétérogénéité... tout en donnant un cadre cohérent à la mesure de l'activité économique.

Le sujet central de la réunion a traité des **évolutions récentes de l'accès aux données granulaires financières**. L'exploitation statistique du gisement d'informations individuelles bancaires et financières produit à des fins de régulation et d'évaluation des politiques a constitué une « révolution » dans les années post-crise de 2008-2009. Ces données granulaires au niveau des ménages ou des firmes sont devenues un outil clé dans l'analyse empirique, car elles permettent d'identifier les mécanismes sous-jacents aux grandes relations économiques et de mieux cerner la distribution des risques afin d'appréhender les enjeux de stabilité financière. Si un travail important de « normalisation » de ces données a été fait, les règles générales pour descendre au niveau micro sont encore à édicter. Depuis quelques années, des travaux sont en cours tant au niveau national qu'au niveau européen pour développer le recours à ces données. Leur mise à disposition implique deux enjeux pour la statistique publique : un enjeu d'accès et un enjeu d'homogénéisation au niveau international.

La Banque de France et le CASD (Centre d'accès sécurisé aux données) présentent leurs innovations dans l'accès et l'appariement des données granulaires françaises. Dans la suite de l'*Open Data Room* « physique » ouverte en 2016, et qui avait été alors présentée dans une commission, la Banque de France met en place une *Open Data Room* « virtuelle » qui permettra un accès distant sécurisé à partir d'un simple poste personnel (projet ODR 2.0). La capacité d'accueil sera ainsi accrue. Cet accès sera complètement étanche au système d'information de la Banque de France et se fera sans téléchargement des données. Une réflexion est en cours sur l'anonymisation des données et sur la facturation de frais de la mise à disposition. Par ailleurs, le CASD, groupement d'intérêt public à but non lucratif, propose aux chercheurs 300 sources de données. Deux accès différenciés aux données quant au dispositif, à la tarification ou au hachage ou non des identifiants, sont ainsi possibles, ce qui pose plusieurs problèmes : un besoin de données CASD et Banque de France pour les chercheurs peut être servi *via* deux accès séparés/ parallèles aux données ; les recherches impliquant des appariements et croisements entre les données CASD et les données Banque de France sont plus difficiles ; des chercheurs s'autocensurent parfois sur la source « minoritaire » de données. Le projet de collaboration actuel entre les deux organismes prend en compte les deux aspects juridique et technique des solutions à trouver. Chaque jeu de données ayant ses spécificités juridiques, il faudrait passer d'un alignement sur les contraintes maximales à un principe plus souple de règles générales de base impliquant une possibilité d'exploitation des données dans l'un ou l'autre des environnements, avec une anonymisation des données Banque de France ou CASD pour un chercheur « lambda ». Le projet ODR 2.0 va permettre de réduire les différences structurelles et faciliter les projets d'échanges croisés de données.

Sur le plan européen et international, l'évolution générale est à une ouverture croissante de l'accès aux données confidentielles, de plus en plus à distance, et sur la base d'un « cercle de confiance » constitué par le chercheur, le projet de recherche, les données, la sécurité de l'accès et le contrôle de l'anonymisation des résultats. Deux projets sont présentés par la Banque de France. Le projet INEXDA (*International Network for Exchanging Experience on Statistical Handling of Granular Data*), porté par plusieurs banques centrales dont la Banque de France, est révélateur de cette orientation commune. Il se veut un lieu d'échange sur les pratiques des membres concernant le partage de données (à des fins de recherches, mais pas uniquement), non un réseau de partage de données entre les membres à destination des chercheurs. Plusieurs groupes de travail réfléchissent sur les bonnes pratiques et techniques à partager pour définir une accréditation (catalogue des données « partageables », définition harmonisée des métadonnées), mettre les données à disposition (modélisation des processus d'autorisation, anonymisation), contrôler les sorties. La crise de 2008-2009 a montré la nécessité de mettre en commun les registres de crédits des différents États membres de l'Union européenne. Il en a résulté le projet AnaCredit (*Analytical Credit Datasets*) de l'UE, fruit d'une collaboration intense entre les banques centrales et la profession bancaire de chaque pays. Ce registre rassemble les informations prêt par prêt sur l'exposition des banques de la zone euro à toutes les personnes morales y compris les PME, soit des données sur 25 millions de prêts transmises chaque mois depuis novembre 2018. Ces données servent à de multiples domaines d'analyse : conduite de la politique monétaire, gestion des risques, stabilité financière, supervision microprudentielle, développement de nouveaux agrégats et indicateurs statistiques.

Dans les échanges qui suivent les présentations, l'accent est mis sur les obstacles techniques et juridiques au rapprochement des données Banque de France et CASD, avec une demande d'assouplir les règles d'accès aux chercheurs. Si l'axe de la coopération est bien un accroissement de l'accès aux données avec pseudonymisation et définition d'un cadre juridique de confiance, il n'est pas envisagé de construire une plateforme commune avec un accès global hors procédure ODR. La présidente de la commission rappelle que ce sujet de l'accès aux données est un enjeu fort du moyen terme du Cnis, dans la suite des groupes de travail sur l'accès aux données bancaires et financières de 2015 (rapport Capelle-Blancard) et sur l'accès des chercheurs aux données de 2016 (rapport Bozio). Tout en ne préjugant pas de la nécessité d'un nouveau groupe de travail, elle souhaite être informée au nom de la commission de la progression des travaux sur ce sujet, notamment sur l'instruction des questions juridiques relativement à l'accès aux données par les chercheurs et à la protection de la confidentialité dans leur usage. Un avis est adopté en ce sens.