

L'économie racontée par les données bancaires

Ce que nos relevés de comptes disent de nous



Odran Bonnet* et Tristan Loisel**

Depuis la crise sanitaire en 2020, l'Insee bénéficie d'un accès à des échantillons anonymisés de comptes bancaires de La Banque Postale et du Crédit Mutuel Alliance Fédérale. Ces données, riches en informations, ont permis de documenter l'évolution de la situation financière des ménages en temps réel lors de la crise inédite de la Covid-19, puis lors de l'épisode inflationniste en 2022. Elles ont également permis de documenter la situation quotidienne des ménages, mettant en lumière des épisodes de découvert en fin de mois, et d'évaluer une politique publique en mesurant les impacts financiers, distributifs et environnementaux de la remise à la pompe instaurée en 2022 à la suite de la hausse des prix du carburant.

Les données bancaires sont une mine d'informations précieuses, mais leur utilisation soulève de nombreux défis. Après quatre années d'utilisation, un premier bilan de leur exploitation est proposé dans cet article. Comment construire des concepts qui ont un sens économique à partir de ces données ? Comment s'assurer de leur représentativité ? Quels sont les apports de ces données pour la statistique publique ? Si elles ne permettent pas de remplacer les données d'enquêtes et fiscales, elles complètent les analyses conjoncturelles. En outre, elles permettent de répondre à des questions anciennes sur la consommation et l'épargne des ménages auxquelles les sources traditionnelles ne pouvaient apporter de réponses. Dans cet article, sont détaillés les enjeux et les difficultés pour utiliser des données qui proviennent d'acteurs privés et n'ont pas été produites à des fins statistiques, mais à des fins de gestion par les banques.

 Since the health crisis in 2020, the National Institute of Statistics and Economic Studies (INSEE) has had access to anonymized samples of bank accounts from La Banque Postale and Crédit Mutuel Alliance Fédérale. These data, rich in information, have enabled the analysis of the real-time evolution of households' financial situation during both the unprecedented COVID-19 crisis and the inflationary episode in 2022. They have also paved the way for the analysis of households' daily situation, highlighting instances of overdrafts at month-end, and for the evaluation of a public policy by measuring the financial, distributive, and environmental impacts of the fuel subsidy implemented in 2022 following the rise in fuel prices.

Bank data are a goldmine of valuable information, but their use raises many challenges. After four years of use, a first assessment of their exploitation is presented in this paper. How can we build economically meaningful concepts from these data? How can we ensure their representativeness? What are the benefits of these data to public statistics? While they do not replace survey and tax data, they complement cyclical analyses. Moreover, they allow us to address long-standing questions about household consumption and savings that traditional sources could not answer. This paper details the challenges and difficulties in using private data that were not produced for statistical purposes but for the management of bank operations.

* Chargé d'études, Département des études économiques, Dese, Insee.
odran.bonnet@insee.fr

** Chargé d'études, Département des études économiques, Dese, Insee.
tristan.loisel@insee.fr

Les relevés de comptes bancaires en disent beaucoup sur nos vies. Lorsque les fins de mois sont difficiles, les soldes des comptes sont au plus bas et les découverts plus fréquents. Lorsque nous perdons notre emploi, un virement régulier de salaire peut être remplacé par un virement de France Travail. Même les départs en vacances ne passent pas inaperçus ; les dépenses de transport et d'hébergement augmentent, de même que les dépenses réalisées à l'étranger. Les données bancaires, très riches en information, permettent donc des exploitations variées.

Dans cet article, nous reviendrons d'abord sur le contexte qui a permis l'accès de l'Insee à ces données ; le mouvement d'ouverture vers des données d'origine privée a été accéléré par la crise sanitaire. Ensuite, nous expliciterons comment ces données sont produites, depuis les transactions des clients jusqu'aux graphiques publiés dans les collections de l'Insee. Enfin, nous détaillerons les défis à relever pour les exploiter et expliquerons les nouvelles informations qu'elles apportent à la statistique publique : comment complètent-elles le diagnostic conjoncturel ? À quelles nouvelles questions économiques permettent-elles de répondre ? Enfin, nous reviendrons sur leurs limites, notamment en matière de représentativité.

► Une aubaine pour décrire la situation financière des ménages pendant la crise sanitaire

En 2020, la crise sanitaire et le confinement à grande échelle de la population ont provoqué un choc sur l'économie aussi soudain qu'inédit. À l'Insee, le suivi de la conjoncture en a été bouleversé, conduisant notamment au recours accru à des données nouvelles, plus rapidement disponibles que les données traditionnelles (Pouget, 2020). Parmi les nouvelles sources mobilisées, les données de comptes bancaires se sont avérées précieuses car très détaillées et rapidement disponibles. Pour l'Insee, ces



Parmi les nouvelles sources mobilisées, les données de comptes bancaires se sont avérées précieuses car très détaillées et rapidement disponibles.



données offraient une réelle opportunité pour conforter et enrichir les informations provenant des sources traditionnelles d'enquête avant que les données de sources fiscales ne soient disponibles. En outre, elles permettaient d'éclairer la situation financière des ménages sous un jour nouveau, grâce à des indicateurs atypiques pour la statistique publique mais très parlants pour le grand public, comme les montants présents sur les comptes en fin de mois ou la proportion de ménages à découvert.

L'intérêt de l'Insee pour ces données s'est concrétisé lorsque certaines banques, dans le contexte de la crise sanitaire, ont souhaité contribuer à l'analyse de la situation économique par la statistique publique. Crédit Mutuel Alliance Fédérale (CMAF) a ainsi permis un accès sécurisé à des données anonymisées de comptes bancaires d'un échantillon de sa clientèle dès 2020¹. Des premières études sur l'évolution de la situation

¹ Le Conseil d'analyse économique (CAE) a demandé et obtenu un accès à des données anonymisées de Crédit Mutuel Alliance Fédérale. À la suite de ce partenariat, l'Insee a formulé la même demande.

financière des ménages pendant la crise sanitaire ont résulté de cette collaboration (Insee Note de conjoncture, 2021). Cependant, utiliser des données provenant d'un seul réseau bancaire pouvait être insuffisant pour décrire l'ensemble de la population. La clientèle de CMAF étant plus aisée que la moyenne, ces données pouvaient notamment s'avérer inadéquates pour documenter la situation financière des plus fragiles. Or, le sujet était brûlant à l'époque, en plein cœur de la crise sanitaire. L'Insee a pris contact avec un second réseau : La Banque Postale (LBP). Il s'agit de l'établissement le plus à même de fournir des informations sur les clients bancarisés les plus fragiles, car le législateur lui a confié une mission d'accessibilité bancaire. Les informations de cette banque ont permis de montrer la stabilité des revenus, malgré la crise, y compris pour les plus modestes. Cela a conforté les estimations obtenues par le modèle de microsimulation Ines² qui concluait à un taux de pauvreté stable. Ce constat, à rebours des discours publics³, rendait le recours à des données externes particulièrement précieux pour convaincre de la solidité des résultats obtenus avec les méthodes usuelles (Tavernier, 2022). Ainsi, la publication sur les données LBP a été programmée le même jour que la publication annuelle portant sur le taux de pauvreté, soit une dizaine de jours avant la journée mondiale de la pauvreté.

La crise inflationniste succédant à la crise sanitaire, les données bancaires ont de nouveau été mobilisées afin d'éclairer l'évolution de la situation financière des ménages dans ce nouveau contexte en 2022 et 2023.

► Une nouvelle utilisation de « big data » privées dans la statistique publique

Si l'accès aux données bancaires a été accéléré par la crise sanitaire, il s'inscrit dans un contexte plus large de nouvelles données massives d'origine privée, comme les données de caisse.

Si l'accès aux données bancaires a été accéléré par la crise sanitaire, il s'inscrit dans un contexte plus large d'utilisation par les instituts nationaux de statistiques de nouvelles données massives d'origine privée, comme les données de caisse (Leclair, 2019). Ces données, appelées communément « big data », peuvent inclure des informations aussi variées que celles d'annonces de logements, de téléphonie mobile ou de compteurs de gaz et d'électricité. Ces données procurent de nombreux avantages par rapport aux sources usuelles comme la haute fréquence ou la granularité (Blanchet et Givord, 2018). Mais il existe également de nombreux obstacles dans leur utilisation. Outre les enjeux de

représentativité, ces données sont peu structurées, et leurs exploitations nécessitent donc d'importants investissements.

² <https://www.insee.fr/fr/information/2021951>.

³ Pendant la crise sanitaire de la Covid, le chiffre d'un « million de pauvres supplémentaires » était repris dans la presse (Le Monde, 2020). Ce chiffre provenait d'associations caritatives. Les outils déployés par la statistique publique pour mesurer le taux de pauvreté (finalement stable en 2020) sont détaillés par la suite (Tavernier, 2022).

Avant d'être transformée en information utile au débat public, la donnée bancaire subit de nombreux traitements, du terminal de paiement aux tableaux ou graphiques des publications, en passant par les systèmes d'information des banques. Ce long chemin n'est pas propre aux données bancaires ou aux données privées et fait écho aux traitements réalisés lors d'utilisation de données administratives (Cotton et Haag, 2023), par exemple celles de propriétés immobilières (André et Meslin, 2022).

Nous allons décrire ce parcours de la donnée et les retraitements réalisés par la banque puis par le statisticien.

► Parcours de la donnée bancaire : du système de gestion privé à une meilleure compréhension des comportements économiques

L'exploitation des données bancaires repose sur des partenariats étroits avec les banques. Le statisticien public ne se contente pas de récupérer et d'exploiter une base préexistante, il doit la construire progressivement et conjointement avec le partenaire bancaire. L'objectif est de transformer des données issues des processus de gestion de comptes bancaires en des tables de données individuelles agrégées au niveau ménage. Cette entreprise nécessite de nombreux échanges avec les banques. D'un côté, le statisticien doit expliciter ses besoins, de l'autre la banque doit détailler le contour des données qu'elle peut mettre à disposition. Les données s'enrichissent au fur et à mesure du partenariat au gré des besoins du statisticien et des contraintes des banques (**encadré 1**). Le parcours de la donnée depuis les opérations réalisées par le client jusqu'à la publication des analyses statistiques se réalise en quatre grandes étapes.

► Encadré 1. Des partenariats étroits avec les banques qui facilitent la montée en puissance de l'utilisation des données bancaires à des fins statistiques

Des échanges réguliers avec les banques ont permis à l'Insee d'améliorer significativement l'exploitabilité des données depuis le début du partenariat. Ces collaborations étroites se sont d'ailleurs traduites par des co-publications avec les banques (de l'Insee avec La Banque Postale et du Conseil d'Analyse Économique (CAE) avec Crédit Mutuel Alliance Fédérale).

Les partenariats avec les banques ont également conduit à des projets d'expériences aléatoires sur certains clients. L'Insee mène actuellement une expérimentation en collaboration avec La Banque Postale, dans le but d'évaluer un dispositif d'accompagnement budgétaire, qui s'inscrit dans le cadre de la mission publique d'accessibilité bancaire de La Banque Postale, destiné aux clients en situation de fragilité financière. Le CAE

a également estimé des propensions marginales à consommer, c'est-à-dire les montants dépensés par les ménages qui bénéficient d'une hausse de revenu, à partir d'une expérience sur des clients du CMAF (Boehm et alii, 2023).

Apparier à d'autres sources (tout en préservant l'anonymat) peut ouvrir d'autres perspectives d'études : grâce à un appariement avec des données publiques de diagnostic de performance énergétique (DPE)*, le CAE a étudié la performance énergétique des logements (Astier et alii, 2024). Au Danemark, des chercheurs ont également étudié la réaction des ménages lors d'une perte d'emploi en appariant les données bancaires des clients d'une grande banque danoise avec les données sur leurs employeurs** (Andersen et alii, 2023).

* Appariement réalisé par la banque à partir de l'adresse des clients.

** L'appariement repose sur une identification de l'employeur par le compte bancaire utilisé pour verser les salaires.

Les deux premières étapes correspondent aux traitements effectués par la banque elle-même. La première consiste à récolter et construire les informations sur ses clients à des fins de gestion à partir des différents types d'opérations réalisées chaque jour sur les comptes de la banque. La seconde étape vise à réaliser les traitements nécessaires afin de mettre à disposition un sous-ensemble de données exploitable par les statisticiens. La table finale doit contenir les variables pertinentes pour le statisticien, sur un champ défini à l'avance, et doit également respecter les critères d'anonymat de la clientèle.

Les deux étapes suivantes sont effectuées par le statisticien. Il va d'abord effectuer des retraitements spécifiques à la source de données afin de constituer une base mobilisable pour une étude. La dernière étape n'est pas spécifique aux données bancaires : une fois les retraitements opérés, le statisticien peut réaliser calculs et graphiques comme il le fait à partir de n'importe quelle autre source.

► Du côté de la banque : des données de gestion privées aux bases de données statistiques

D'où viennent les données bancaires ?

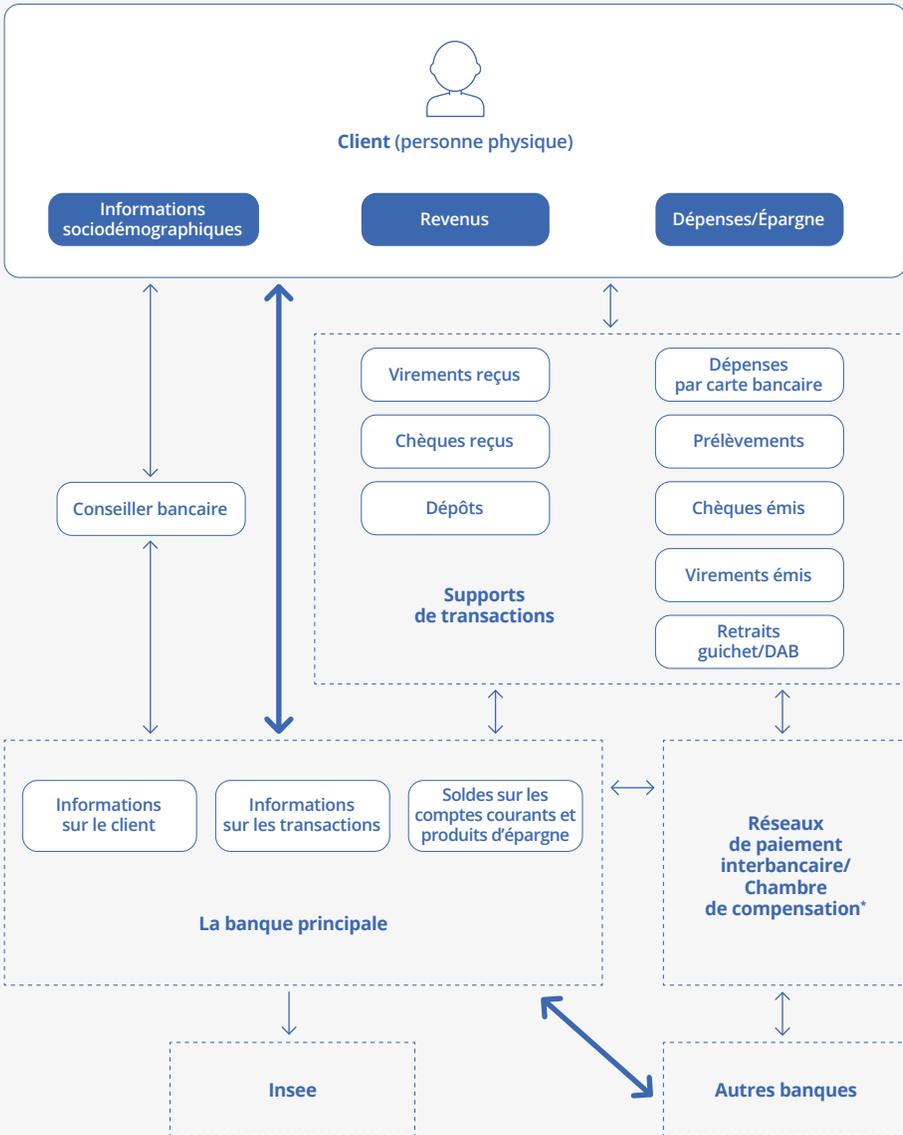
Chaque jour, des millions d'opérations bancaires ont lieu. Toutes ces opérations sont enregistrées par le système d'information bancaire afin que l'argent circule entre les comptes des ménages et des entreprises. Une transaction bancaire génère un flux de données entre la banque émettrice et la banque acquéreur, en passant par plusieurs intermédiaires (*figure 1*). Par exemple, lors d'un paiement par carte chez un commerçant, le traitement du paiement se déroule généralement en trois étapes : l'autorisation, la compensation et le règlement. La phase d'autorisation a lieu au moment où le client insère sa carte dans le terminal de paiement du magasin. La banque du commerçant (banque acquéreur) interroge alors la banque du client (banque émettrice) via le réseau de cartes pour savoir si la transaction est approuvée. La banque émettrice renvoie ensuite une réponse, positive ou négative, qui parvient jusqu'au terminal de paiement et qui permet de valider ou refuser l'opération. En fin de journée, le commerçant transfère à sa banque un fichier récapitulatif

Une transaction bancaire génère un flux de données entre la banque émettrice et la banque acquéreur, en passant par plusieurs intermédiaires.

toutes les transactions enregistrées (début de l'étape de compensation). Celle-ci regroupe les informations pour l'ensemble de ses clients et les transmet à la banque émettrice, par l'intermédiaire du réseau de cartes, afin de préparer l'étape de règlement. Lors de cette dernière étape, trois règlements ont lieu (de manière indépendante, et pas nécessairement dans cet ordre) : la banque acquéreur reçoit les fonds de la banque émettrice, verse le montant dû au commerçant, tandis que la banque émettrice débite le compte du client.

Au-delà de ces flux, les banques collectent également des variables sociodémographiques afin de mieux connaître leur clientèle (profession, nombre d'enfants par exemple). Cela leur permet de proposer des services adéquats. Ces

► **Figure 1 - Représentation des flux de données : du client de la banque jusqu'au statisticien**



Légende

↔ Échanges d'informations

↔ Flux monétaires

* Les chambres de compensation sont des infrastructures privées, qui déterminent pour chaque banque le solde à livrer ou à recevoir en fonction de l'ensemble des transactions produites. Ce mécanisme permet d'éviter la multiplication des flux entre les banques.

informations sont traditionnellement collectées lors des rendez-vous avec les conseillers en agence mais également de plus en plus via des questionnaires envoyés par courriel.

Sélection et transformation des données pour répondre aux besoins de la statistique publique

De manière itérative, à la suite de nombreux échanges visant à clarifier les données disponibles et l'expression des besoins de l'Insee, la banque sélectionne les données pertinentes, les anonymise et les transforme selon les étapes suivantes.

Tout d'abord, la banque opère une **sélection des tables et des variables**. Au sein de la masse de données stockées au sein de la banque à des fins de gestion (des centaines, voire des milliers de tables contenant chacune quelques dizaines de variables), la banque sélectionne les informations sur les transactions, les soldes, les crédits, les incidents bancaires, ainsi que les caractéristiques sociodémographiques de ses clients. Les données sont **figées** un mois donné, il arrive donc qu'il manque certaines transactions remontant avec un délai, notamment celles effectuées les derniers jours du mois.

Ensuite, la banque procède au **tirage d'un échantillon**. Une fois ces tables et variables rassemblées, la banque tire un échantillon de clients à mettre à disposition selon une méthodologie préalablement définie. Cette opération d'échantillonnage qui est monnaie courante dans la statistique publique est rare dans le monde bancaire. Le tirage se fait uniquement sur les clients dont la banque estime héberger les revenus et la majorité des opérations de consommation.

Puis la banque opère l'**anonymisation des données**. L'anonymisation (*figure 2*) implique la suppression de variables identifiantes comme le(s) nom(s) et adresse des clients mais également le(s) nom(s) d'émetteurs et les libellés des virements. Or, ces dernières variables sont utiles pour les analyses puisqu'elles permettent de repérer les salaires perçus par les clients, mais aussi les prestations sociales, les allocations chômage, les pensions de retraites, etc. Avant la suppression, la banque extrait donc les informations utiles et non identifiantes de ces variables (type d'émetteur, régularité de l'émetteur, libellé contenant des mots-clés comme « salaire », etc.). Il s'agit d'un compromis qui permet de garantir l'anonymat des clients, tout en limitant la perte d'information (Redor, 2023).

Enfin, la banque opère le **transfert des données**. Elle transmet les données et les métadonnées à l'Insee dans des environnements sécurisés, au format CSV via le Centre d'Accès Sécurisé aux Données (CASD) pour LBP, et sous la forme d'une base de données relationnelle via une machine virtuelle pour CMAF.

► Du côté du statisticien : de la base de données à la publication d'une étude économique

Les données transmises comportent de riches informations contenues dans de multiples tables (*encadré 2*) ; le statisticien doit les retraiter afin de réaliser des études économiques.

► **Figure 2 - L'anonymisation des données**

Identifiant anonyme du client bénéficiaire (clé d'appariement)	Compte émetteur	Compte bénéficiaire	Nom de l'émetteur	Nom du bénéficiaire	Libellé	Date	Montant	Catégorie
1	Compte émetteur 1	Compte bénéficiaire 1	France Travail	Baron Danglars	Allocation	01/02/2024	1256,34	Chômage
2	Compte émetteur 2	Compte bénéficiaire 2	CAF	La marquise de Saint-Méran	Prestation CAF Marseille	05/02/2024	607,75	Prestations sociales
3	Compte émetteur 3	Compte bénéficiaire 3	CNAV	Louis Dantès	Pension	09/02/2024	1482,73	Retraite
4	Compte émetteur 4	Compte bénéficiaire 4	Gaspard Caderousse	Fernand Mondego	Remboursement dettes de jeu	27/02/2024	443,87	Inconnu
5	Compte émetteur 5	Compte bénéficiaire 5	Armateur Morel	Edmond Dantès	Salaire février	27/02/2024	2978,36	Salaire
5	Compte émetteur 6	Compte bénéficiaire 6	Mercédès Herrera	Edmond Dantès	Vide	28/02/2024	200,00	Inconnu

↓ Anonymisation

Identifiant anonyme du bénéficiaire	Date	Montant	Catégorie
1	01/02/2024	1256,34	Chômage
2	05/02/2024	607,75	Prestations sociales
3	09/02/2024	1482,73	Retraite
4	27/02/2024	443,87	Inconnu
5	27/02/2024	2978,36	Salaire
5	28/02/2024	200,00	Inconnu

Légende

- Variables créées par la banque pour anonymiser les données tout en limitant la perte d'information
- Informations sensibles ou variables identifiantes à supprimer avant la mise à disposition des données

Transformer les données transmises en bases statistiques facilement utilisables

Une fois les données reçues, la première étape consiste en des **macro-contrôles internes**. Cela consiste à produire une série de vérifications sur la volumétrie des données afin de s'assurer que le dernier rafraîchissement ne comporte pas d'erreurs grossières. Concrètement, cela revient à vérifier dans les différentes tables que les évolutions du nombre de clients et du nombre de transactions ne sont pas erratiques. Le statisticien ne maîtrisant pas le processus amont, les macro-contrôles sont primordiaux. Il se peut par exemple que les dernières données transmises au statisticien soient incomplètes (comme des transactions manquantes en fin de mois).

► Encadré 2. Au cœur des finances des clients : que contiennent les données transmises par la banque ?

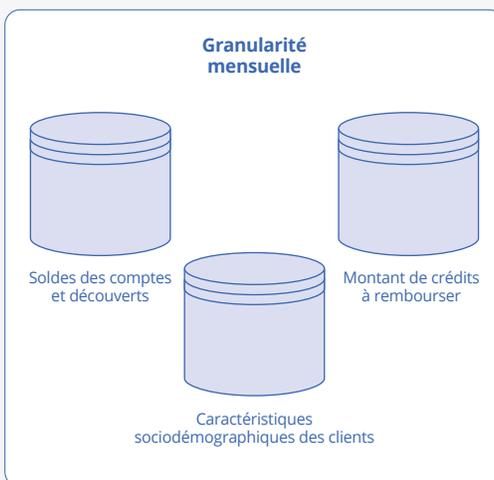
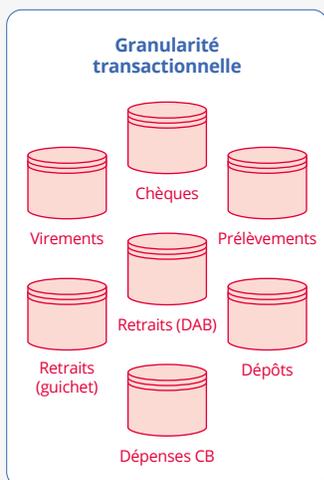
Les données transférées se trouvent dans plusieurs tables selon leurs origines (*figure encadré*) : une table contient le solde des comptes bancaires, d'autres contiennent les transactions par type de mouvement (carte, prélèvement, virement, retraits) et d'autres les caractéristiques des individus titulaires des comptes. Elles couvraient initialement des panels anonymisés de 300 000 clients/groupes familles*.

Pour les deux réseaux bancaires, les données contiennent peu ou prou l'équivalent anonymisé de ce que chacun peut retrouver sur son relevé de compte. Plus précisément, elles incluent : les soldes des comptes en fin de mois pour différents types de comptes (comptes courants individuels ou joints, livrets, assurances-vie et comptes-titres), toutes les transactions effectuées (montants et dates des opérations par carte bancaire, chèques, virements, prélèvements, retraits et dépôts). En outre, elles contiennent quelques informations sociodémographiques connues de la banque : âge, sexe, département, statut marital, type d'habitat et catégorie socioprofessionnelle**.

Cependant, pour des questions de respect de l'anonymat (cf. supra), les données mises à la disposition de l'Insee ne sont pas aussi riches que celles présentes sur nos relevés personnels. Afin de pallier cette perte d'information, les

banques mettent à disposition une catégorisation des opérations. Les dépenses par carte sont catégorisées selon une nomenclature du système bancaire*** qui identifie le type d'établissement récipiendaire du paiement (tel que supermarché, station essence, boulangerie, etc.). Cette nomenclature de 650 postes permet notamment de savoir comment les consommateurs ont modifié leur panier de consommation pendant la crise sanitaire. Les catégorisations de prélèvement permettent d'identifier également les consommations d'énergie et de mesurer l'impact de la crise énergétique sur le portefeuille des ménages français. De plus, depuis début 2020 dans les données de CMAF et mi-2020 dans les données de LBP, les virements entrants provenant de France Travail, des Caisses d'Allocations Familiales (CAF) et des caisses de retraites sont identifiés ; cela permet de repérer les versements de prestations de chômage, de solidarité et les pensions de retraite. Des virements entrants correspondant à des salaires sont également isolés.

Au total, ces informations riches impliquent une volumétrie élevée : de l'ordre d'une dizaine de millions d'opérations mensuelles. Cela requiert d'adapter les traitements à ces données volumineuses.



* La banque regroupe les clients qu'elle considère appartenir à un même ménage dans ce qu'elle nomme un groupe famille. Ces groupes sont constitués à partir de la détention de comptes joints et de l'information collectée par les conseillers.

** Les catégorisations de professions reposent sur des nomenclatures proches de celles utilisées à l'Insee, mais elles ne les recourent pas parfaitement au niveau le plus fin à quatre positions.

*** La nomenclature internationale se nomme Merchant Category Code (MCC).

Ensuite, le statisticien construit à partir des données bancaires **des variables économiques** : les variables correspondant aux concepts à étudier. Par exemple, le patrimoine financier brut d'un groupe famille/foyer est reconstitué à partir de la somme des soldes sur ses comptes.

Les données sont structurées selon une **représentation rectangulaire** familière au chargé d'étude, dans laquelle une ligne correspond à une observation, c'est-à-dire un groupe famille/foyer un mois donné (ou un jour donné selon la fréquence étudiée). Les colonnes contiennent les variables d'intérêt : revenus, dépenses, patrimoine financier et caractéristiques sociodémographiques.

L'échantillon est ensuite **restreint** à la population d'intérêt et un **calage**⁴ est mis en place. Différents filtres peuvent être appliqués afin de conserver uniquement l'échantillon de clients souhaité. De plus, l'échantillon peut être calé sur plusieurs marges (âge quinquennal, sexe, population par département) afin de le rendre davantage représentatif de la population française.

Enfin de nouveaux **macro-contrôles externes** sont réalisés sur les données finales. L'objectif est de confronter les informations et les messages issus des données bancaires à ceux provenant d'autres sources. Idéalement, les données bancaires doivent être confrontées à des données portant sur les mêmes concepts, pour permettre des comparaisons plus directes, et surtout de même fraîcheur. Une première confrontation naturelle et fondamentale est donc la comparaison des données des deux banques. Mais d'autres sources peuvent également être mobilisées, comme les indicateurs de la Banque de France (publiés mensuellement), ou encore les données du Groupement d'Intérêt Économique des Cartes Bancaires (GIE CB), qui permettent de tracer des évolutions agrégées de dépenses par carte bancaire.

Au bout du compte, des données comme les autres ?

Les données bancaires sont alors traitées par le statisticien de la même manière que n'importe quelle autre source de données. Il peut donc effectuer les opérations statistiques et économétriques nécessaires pour répondre à sa problématique. La dernière étape consiste à extraire les résultats agrégés du CASD (pour LBP) ou de la banque (pour CMAF) après vérification de leur conformité (notamment l'absence de données individuelles).

Ces données permettent de répondre à plusieurs problématiques, du suivi conjoncturel à l'évaluation de politiques publiques.

⁴ Le calage sur marges est une technique statistique visant à améliorer la précision des enquêtes par sondage. Elle consiste à modifier les poids de sondage des individus de l'échantillon afin que les totaux pondérés sur l'échantillon de certaines variables correspondent aux totaux connus pour ces variables sur l'ensemble du champ d'observation (la population, le parc de logements, les entreprises). Voir aussi « Miscellanées sur le calage » : <https://www.insee.fr/fr/information/2387498>.

► Voir vite et net : des données transformées pour améliorer le suivi conjoncturel

Pour éclairer la conjoncture, ces données ont le quadruple avantage de la fraîcheur, de la granularité fine, de la grande taille de l'échantillon et de la diversité des informations fournies.

Pour éclairer la conjoncture, ces données ont le quadruple avantage de la **fraîcheur**, de la **granularité fine**, de la grande **taille** de l'échantillon et de la **diversité** des informations fournies. En effet, elles sont mobilisables rapidement : les données du mois M sont disponibles à la fin, voire au milieu du mois M+1. Les opérations sont disponibles au niveau journalier, permettant une analyse en temps réel des réactions des clients face à un choc soudain, comme la baisse des taxes sur le carburant. En outre, les échantillons de clients sont ici suffisamment grands pour focaliser l'étude sur des populations particulières, comme les

bénéficiaires du RSA⁵ pendant la crise sanitaire ou les grands consommateurs de carburant au moment du déclenchement de la guerre en Ukraine. Avec les enquêtes traditionnelles, ces populations ne sont souvent pas suffisamment nombreuses dans les échantillons pour être étudiées spécifiquement. Cet avantage pour la statistique publique a également été souligné lors de l'utilisation de données administratives exhaustives qui permettent d'aller au-delà des décompositions habituelles en dixième de patrimoine et permettent de zoomer jusqu'au centième, voire jusqu'au millième (André et Meslin, 2022). Enfin, les variables sont suffisamment riches pour à la fois mesurer les concepts habituellement mobilisés à l'Insee comme les revenus, la consommation ou l'épargne mais également des grandeurs parlantes pour le grand public comme le solde à la fin du mois.

Ces atouts ont été pleinement exploités lors de la crise sanitaire et de l'épisode inflationniste qui a suivi. Tout d'abord, les données bancaires ont permis de documenter la hausse de l'épargne du fait de la chute de la consommation pendant les confinements (Insee, 2021). Ensuite, elles ont permis de montrer que les revenus de la plupart des clients ont été affectés seulement de manière limitée et temporaire par la crise, mais que certaines populations en marge du marché de l'emploi, comme les allocataires du RSA, ont été davantage pénalisées (Bonnet et alii, 2021). Par ailleurs, plusieurs indicateurs ont été construits permettant de conclure que la précarité financière n'a pas augmenté significativement au cours de la crise sanitaire (**figure 3**). Au début de la période inflationniste en 2022 et de nouveau en 2023, deux études ont montré que la précarité financière, mesurée par quelques indicateurs sur les découverts bancaires, augmente depuis début 2021, mais reste inférieure à son niveau d'avant-crise sanitaire (Adam et alii, 2022, Bonnet et alii, 2023).

En dehors de l'Insee, d'autres exploitations conjoncturelles des données bancaires ont été publiées en France. Par exemple, les données CMAF ont été mobilisées par le Conseil d'Analyse Économique pour étudier l'évolution de la situation financière des ménages (Astier et alii, 2024, Fize et alii, 2022a, Fize et alii, 2022b) mais aussi des entreprises (Epaillard et alii, 2021, Fize et alii, 2022b).

⁵ RSA : Le revenu de solidarité active est une prestation de protection sociale française, qui complète les revenus d'une personne démunie ou aux ressources faibles, afin de lui garantir un revenu minimal.

► **Figure 3 - Précarité financière sur le panel de clients de La Banque Postale entre janvier 2019 et juin 2021**



Note : Les bandes vertes correspondent aux périodes de confinement.

Lecture : Fin mars 2020, 44 % des clients ont moins de 3 000 euros sur leurs comptes (précarité en patrimoine). Au cours de ce mois, 6,3 % des clients ont cumulé au moins 5 jours débiteurs hors découvert autorisé (précarité en découvert), 27,6 % ont perçu moins de 1 000 euros (précarité en revenus) et les revenus sont de 1 740 euros en moyenne.

Champ : France, échantillon de clients bancarisés principaux de La Banque Postale présents sur toute la période janvier 2019-juin 2021 après filtrage des comptes inactifs.

Sources : La Banque Postale ; Insee, calculs des auteurs.

► Mesurer la diversité des ménages : des données massives qui complètent les enquêtes sans pouvoir les remplacer

En Espagne, des chercheurs soulignent l'importance de ces données, qui permettent d'augmenter la précision des mesures d'inégalités de consommation, lesquelles reposent sinon sur des enquêtes. Leur apport est affirmé dans un document de travail de l'Université de Cambridge : « Une fois organisées selon les principes de la comptabilité nationale, [les données bancaires peuvent] reproduire les statistiques officielles actuelles sur la consommation globale au niveau national avec un haut degré de précision. En raison de la richesse des données de transaction, elles peuvent en outre produire de nouveaux comptes distributionnels de la consommation, qui révèlent des inégalités de consommation plus importantes que ne le suggèrent les enquêtes. » (Buda et alli, 2022).

Elles permettent également d'aller au-delà des limites des données d'enquêtes de consommation (comme l'enquête [Budget de Famille](#)⁶), grâce à la possibilité de suivre les ménages en panel et d'étudier certaines populations spécifiques, qui, sinon, ne seraient pas en effectifs suffisants. De plus, ces enquêtes ne recensent la consommation de certains biens que sur une courte période, ce qui limite la connaissance de la distribution de la consommation. Par exemple, si un ménage répondant n'a pas consommé un bien pendant la période d'interrogation, il est impossible de savoir s'il ne le consomme jamais ou seulement rarement. Disposer d'informations très précises et détaillées est ainsi fondamental pour mesurer la diversité des pertes liées à une hausse des prix du carburant ou de l'énergie, qui peuvent varier très fortement au cours de l'année.

Enfin, les données bancaires peuvent partiellement remédier aux limites liées aux erreurs de déclarations, comme la sous-déclaration du patrimoine financier.

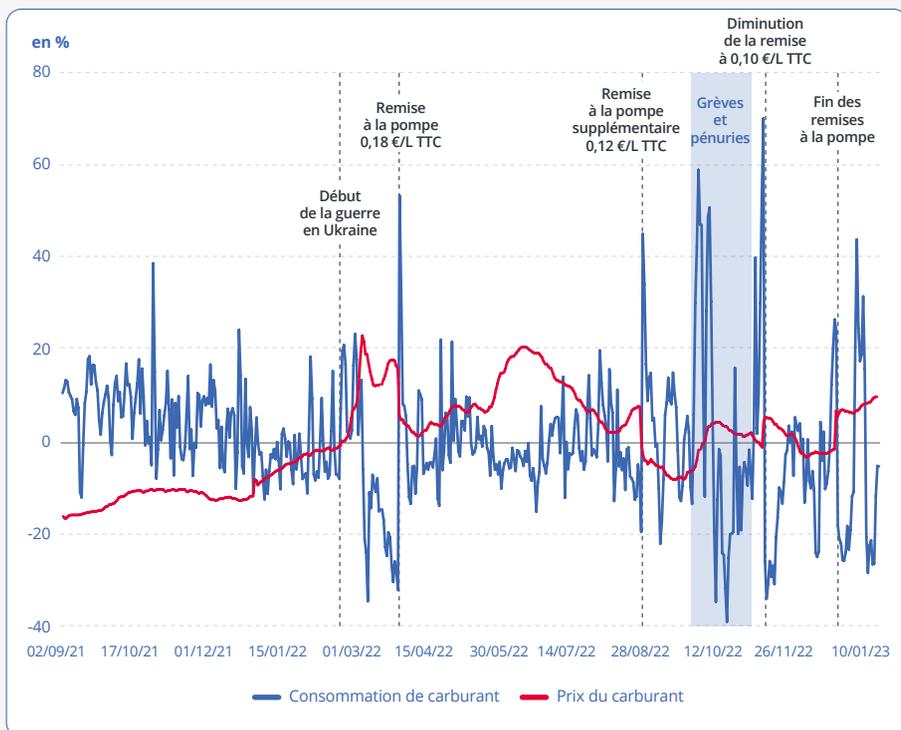
► Analyser l'évolution de la situation financière des ménages à la suite de chocs de prix ou de revenus

Les données bancaires ont l'avantage décisif de contenir simultanément des informations quotidiennes sur le revenu, la consommation et l'épargne d'un panel de ménages observé sur plusieurs années.

Pour les études économiques, les données bancaires ont l'avantage décisif de contenir simultanément des informations quotidiennes sur le revenu, la consommation et l'épargne d'un panel de ménages observé sur plusieurs années. L'apport de ces données est souligné pour mesurer la réaction des ménages à des chocs de revenus (Baker et alli, 2020), qu'ils soient permanents (hausse de salaire, passage à la retraite) ou temporaires (épisode de chômage, allocation exceptionnelle, prime inflation). Il est possible de savoir si les ménages peuvent puiser dans leur épargne ou doivent se résoudre à diminuer leur consommation à la suite

6 <https://www.insee.fr/fr/metadonnees/source/serie/s1194>.

► **Figure 4 - Prix et consommation de carburant entre septembre 2021 et janvier 2023, variations quotidiennes par rapport à la moyenne sur la période**



Lecture : Le 2 septembre 2021, les prix du carburant étaient inférieurs de 16,4 % à leur valeur moyenne sur la période (de septembre 2021 à janvier 2023) et les quantités achetées supérieures de 10,6 % aux quantités achetées moyennes sur la même période.

Champ : Échantillon de clients actifs et automobilistes Crédit Mutuel Alliance Fédérale.

Source : Crédit Mutuel Alliance Fédérale, calculs Insee.

d'un choc de revenu. Elles participent donc à l'évaluation de politiques publiques. À ce titre, ces données ont déjà permis d'évaluer les impacts financiers, distributifs et environnementaux des remises sur le carburant mises en place par le gouvernement à la suite des fortes variations de prix consécutives à la guerre en Ukraine (Adam et alii, 2023, *figure 4*).

Ces données ont également permis d'étudier l'évolution de la situation financière des ménages au jour le jour (Bonnet et alii, 2023), illustrant les contraintes de liquidité qui pèsent sur ceux-ci au fil du mois : parmi eux, combien sont dans le rouge la veille du jour de paie (*figure 5*) ? Enfin, elles permettent de mesurer les achats à l'étranger des résidents français, et donc d'étudier les questions de tourisme, d'achats transfrontaliers et d'évitement des taxes pour des biens tels que le tabac (Hillion, 2024) ou le carburant (Adam et alii, 2024).



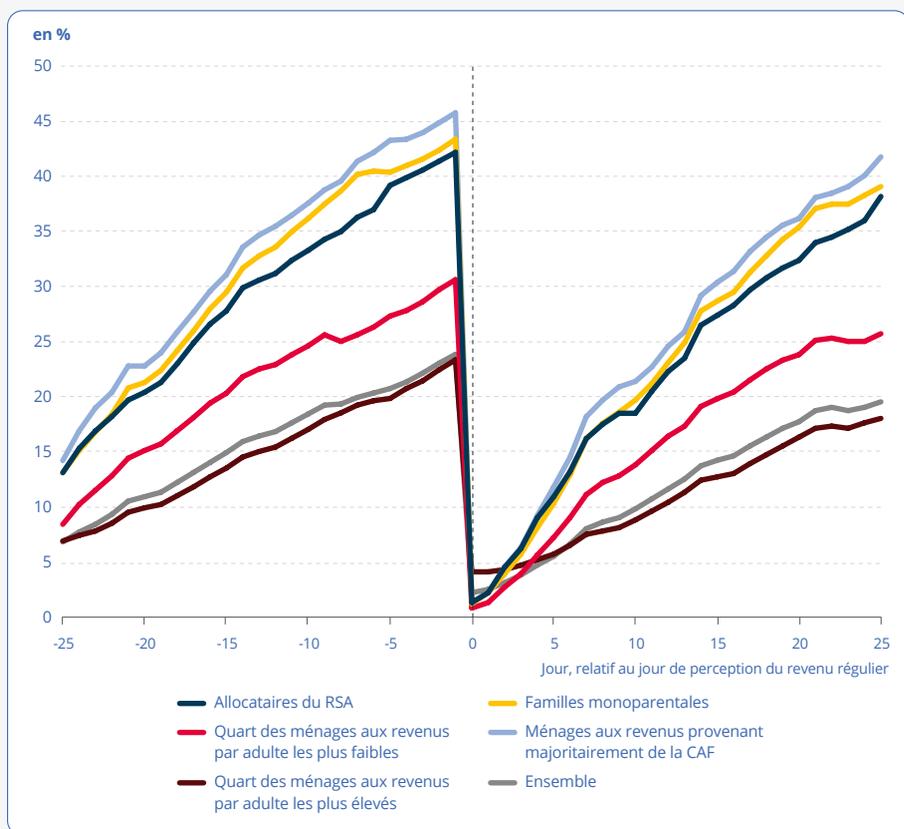
Les données bancaires sont donc une source d'information riche, et ont le grand avantage, par rapport aux enquêtes, d'épargner au statisticien public (et aux comptes publics) les coûts liés au processus de collecte.



Les données bancaires sont donc une source d'information riche, et ont le grand avantage, par rapport aux enquêtes, d'épargner au statisticien public (et aux comptes publics) les coûts liés au processus de collecte. L'inconvénient principal est la perte de la maîtrise du processus (Rivière, 2018). Tout comme les données administratives, les données bancaires existent à des fins de gestion, indépendamment des besoins statistiques. Par conséquent, le champ et les

variables de ces données ne répondent pas nécessairement aux besoins du statisticien. La construction conjointe des bases avec les banques ne permet que partiellement de résoudre ce problème du fait de la nature des données d'origine.

► **Figure 5 - Part des ménages à découvert au fil du mois**



Lecture : 20,7 % des ménages de La Banque Postale percevant un revenu régulier les trois derniers mois sont à découvert 5 jours avant la perception de leur revenu régulier.

Champ : France métropolitaine, échantillon de clients bancarisés principaux à La Banque Postale entre janvier 2019 et août 2023, après filtrage des comptes inactifs et repérage d'un revenu régulier. Période : Mai 2022 – août 2023.

Source : La Banque Postale, calculs Insee.

► La représentativité : étudier la santé financière des Français ou simplement celle d'une clientèle particulière ?

Le champ étudié dans les données bancaires correspond aux clients de la banque et n'est donc pas représentatif de la population française. Il exclut par définition les personnes non titulaires d'un compte bancaire, mais chaque banque peut aussi se spécialiser différemment sur certains segments de la population. Avoir les données de deux banques aux clientèles différentes est un atout pour répondre à cet enjeu. Des comparaisons avec des sources externes permettent également de quantifier l'ampleur du biais. En termes de revenus, de patrimoine et de consommation, la clientèle de CMAF apparaît en moyenne un peu plus aisée que la population générale et celle de LBP un peu moins. Cependant les clientèles des deux banques présentent une diversité importante qui semble couvrir une large partie du spectre économique des ménages français. L'analyse des catégories socioprofessionnelles disponibles dans les fichiers bancaires pointe une surreprésentation des étudiants et une sous-représentation des retraités ; cela peut résulter d'un biais de représentativité ou d'un délai de mise à jour de la situation professionnelle.

Afin de corriger ce biais potentiel, l'échantillon peut être calé sur l'âge quinquennal, le sexe et le département. Ces variables sont les seules pour lesquelles il est certain que la définition est la même entre l'échantillon bancaire et la source de référence. Le calage contribue à tendre vers un échantillon plus représentatif de la population ; mais faute d'avoir les données de l'ensemble des banques, l'échantillon ne peut pas respecter les standards de représentativité des enquêtes. Il est donc impossible de produire des chiffres « officiels » sur l'évolution des revenus, de l'épargne ou de la consommation à partir de ces données bancaires.

► Un texte à trou : les comptes bancaires, une vision partielle du patrimoine des ménages

Outre la question de la représentativité, les informations sur les ménages et leur situation économique provenant des données bancaires sont incomplètes (Baker et alii, 2020). Certains clients peuvent disposer de plusieurs comptes dans différentes banques, ce qui peut conduire à sous-estimer leurs revenus, leurs dépenses, mais surtout leur patrimoine financier (notamment en cas de détention d'assurances-vie). L'enquête *Histoire de Vie et Patrimoine*⁷ mesure l'ampleur du phénomène qui se concentre surtout chez les ménages aux plus hauts revenus : les 10 % de ménages aux plus hauts revenus détiennent 60 % de leur patrimoine financier⁸ au sein de leur banque principale, contre 90 % pour les 10 % aux revenus les plus faibles. De plus, le patrimoine immobilier est inconnu. Pour réduire ce problème de complétude, des chercheurs utilisent parfois des données d'application qui agrègent les comptes des clients des différentes institutions bancaires (Olafsson et Pagel, 2018). Cependant, les utilisateurs d'une telle application bancaire sont probablement moins représentatifs de la population générale que les clients d'une banque dans leur ensemble.

7 <https://www.insee.fr/fr/metadonnees/source/serie/s1005>.

8 La proportion s'élève à 75 % en se restreignant au patrimoine détenu en banque.

► Des statistiques éloquentes, mais éloignées des mesures usuelles de la statistique publique

Contrairement aux enquêtes, les données bancaires ne permettent pas une mesure directe des concepts usuels de la statistique publique. Alors que le statisticien souhaite mesurer les revenus, l'épargne et la consommation, la banque enregistre des flux entrants ou sortants par type de moyen de paiement. Et alors que le statisticien souhaite calculer des statistiques par ménage, les données bancaires permettent, dans le meilleur des cas, un regroupement des membres du ménage bancarisés au sein de l'établissement bancaire considéré.

Précisément, la notion de **ménage**⁹ est approchée par celle de groupe famille, qui repose sur des informations transmises au conseiller et sur la détention de comptes joints. La taille des ménages est donc sous-estimée, car le conjoint et les enfants ne sont pas toujours intégrés dans le ménage, notamment s'ils n'ont pas de compte ouvert dans cette banque¹⁰. Les **revenus** mensuels sont mesurés à partir des flux entrants, mais cette mesure inclut des transferts qui ne correspondent pas tous à des revenus, tels que des transferts entre comptes d'un même individu dans des banques différentes. La catégorisation des virements (pension de retraite, allocation de chômage, prestation sociale, salaire, etc.) distingue certains types de revenus au sein des virements entrants. Mais si le repérage des revenus de remplacement semble fiable, comme les revenus d'allocations chômage qui proviennent de France Travail, celui des revenus d'activité l'est moins, tout comme celui des revenus du capital ; par exemple, les loyers versés entre particuliers ne peuvent être distingués de transferts réguliers entre membres d'une même famille n'appartenant pas à la même banque. Les **dépenses** mensuelles peuvent être mesurées par la somme des dépenses par carte¹¹, des retraits (au distributeur ou au guichet), et de certains prélèvements¹². Le fait d'exclure les chèques et les virements émis conduit à sous-estimer les dépenses. Cependant, inclure tous les virements, chèques et prélèvements conduirait à surestimer les dépenses, tel qu'en cas de virement à soi-même ou de prélèvement

de remboursement de crédit. La nature des dépenses n'est pas toujours facile à identifier. La catégorisation des dépenses par carte ne permet qu'imparfaitement d'identifier la structure du panier de consommation. Seul le type de vendeur est connu mais pas le détail des biens achetés. Or, un supermarché, par exemple, vend des biens très divers.

Les exploitations des données bancaires sont encore récentes. Elles se sont révélées précieuses lors de la période exceptionnelle de la crise sanitaire, mais la statistique publique doit encore prouver sa capacité à les utiliser dans le cadre d'un usage conjoncturel pérenne.

Les exploitations des données bancaires sont encore récentes. Elles se sont révélées précieuses lors de la période exceptionnelle de la crise sanitaire, mais nous devons encore prouver notre capacité à les utiliser dans le cadre d'un usage conjoncturel pérenne.

9 <https://www.insee.fr/fr/metadonnees/definition/c1879>.

10 En outre, à CMAF, un nouveau groupe client est automatiquement créé quand un enfant atteint l'âge de 18 ans, même s'il réside toujours chez ses parents.

11 La nomenclature MCC identifie le type d'établissement récipiendaire de la transaction (station essence, boulangerie, mais aussi supermarché, etc.) : le détail des biens consommés au sein de ces établissements n'est pas observé.

12 Les prélèvements relatifs aux crédits ou aux impôts sont exclus, car il ne s'agit pas de dépenses de consommation.

► **En temps de crise les évolutions écrasent le bruit des données, mais que peut-on mesurer par temps calme ?** —

Les questions de volatilité des indicateurs, de représentativité partielle, ou de complétude peuvent apparaître secondaires dans l'étude d'un choc aussi violent que la crise sanitaire ou la crise inflationniste. Le bruit provenant de dynamiques saisonnières et d'imprécisions inhérentes à ces données est alors de deuxième ordre par rapport aux variations subies par l'économie. En période de moins grandes turbulences, ces limites sont plus problématiques et pourraient conduire à tirer des enseignements erronés sur la conjoncture. L'enjeu est d'autant plus important que la profondeur temporelle de ces données demeure faible. À terme, avec des séries de plus longue période, il sera possible de neutraliser une partie du bruit en désaisonnant¹³. La désaisonnalisation pourrait même être journalière, plutôt que mensuelle, afin de bénéficier de la grande fréquence de ces données.

► **Des études ponctuelles ou un partenariat pérenne ?** —

Un autre inconvénient d'utiliser des données privées par rapport à des données d'enquêtes est l'absence de garantie sur la pérennité des partenariats : les conventions précisent que les deux partenaires peuvent mettre fin à la collaboration à tout moment.

Au-delà de ces incertitudes, la méthode d'échantillonnage doit également permettre un rafraîchissement de la clientèle étudiée. Le premier échantillonnage avait été pensé pour des études ponctuelles sur la crise sanitaire. Les premières publications de l'Insee à partir des données bancaires reposaient sur un échantillon de clients issus de la clientèle de 2019-2020 de chaque banque. Pour une exploitation pérenne, il est cependant nécessaire de rafraîchir l'échantillon par l'ajout de nouveaux clients, pour éviter un biais d'attrition¹⁴ (certains clients disparaissent) et de vieillissement (avec un risque de décalage de la distribution des revenus observés). Ces deux biais s'aggravant avec le temps, une nouvelle méthodologie d'échantillonnage inspirée du plan de sondage de l'échantillon démographique permanent (EDP) a été mise en place. Cette nouvelle méthode permet un rafraîchissement régulier, mensuel, en ajoutant au fil du temps les clients qui entrent dans le champ du plan de sondage. Ce nouvel échantillon ouvre des perspectives pérennes à l'utilisation de ces données.

► **Un fort potentiel pour les études économiques et des limites potentiellement surmontables** —

Les études réalisées depuis quatre ans à partir des données bancaires ont permis de mieux cerner leur potentiel et leurs limites. Elles renseignent en temps réel sur les revenus, la consommation et le patrimoine de centaines de milliers de ménages et permettent ainsi de documenter l'évolution de la situation financière des ménages à

¹³ Désaisonnaliser consiste à appliquer un traitement statistique pour éliminer les effets dus aux phénomènes saisonniers.

¹⁴ Biais d'attrition : perte de représentativité de l'échantillon à cause de la disparition de certains clients de l'échantillon.

la suite de chocs conjoncturels ou individuels. En revanche, malgré leur richesse, elles ne peuvent pas remplacer les enquêtes. Les principales limites sont les suivantes : la clientèle d'une banque n'est jamais totalement représentative de la France entière (certains ménages ne détiennent d'ailleurs aucun compte en banque), les différents types de biens achetés dans une enseigne sont inconnus des banques, les données d'une seule banque ne donnent qu'une vision partielle des flux et du patrimoine des clients multi-bancarisés. Par ailleurs, pour l'instant, le statisticien ne dispose pas de suffisamment de recul pour désaisonnaliser ces données et donc pour exploiter tout le potentiel qu'offre leur très grande fréquence.

À terme, certaines limites peuvent cependant être partiellement surmontées. Ainsi, pour mieux connaître leurs clients, les banques enrichissent continûment leurs données. Chaque amélioration de la catégorisation des flux sur les comptes est un pas supplémentaire vers une identification des revenus et des dépenses tels qu'usuellement définis dans la statistique publique. Chaque information collectée par le conseiller améliore également le regroupement des clients d'un même ménage. Nourrir les partenariats existants, et en développer de nouveaux, permettra de gagner en représentativité et d'acquérir une vision plus complète des comptes des personnes multi-bancarisées. Enfin, avec des partenariats plus anciens, la profondeur temporelle des données augmente : de deux ans à la naissance des partenariats, l'historique disponible des clients est désormais de plus de cinq ans.

L'Insee remercie La Banque Postale et Crédit Mutuel Alliance Fédérale pour leur disponibilité et pour avoir permis l'accès à des données de comptes bancaires dans un cadre garantissant l'anonymat des clients.

Nos partenaires souhaitent rappeler les éléments suivants :

Pour La Banque Postale : ce partenariat contribue pleinement à la réalisation des objectifs environnementaux et sociaux que La Banque Postale a défini dans ses statuts en tant qu'entreprise à mission. Les données bancaires, communiquées anonymement, fournissent des points de vue inédits et complémentaires sur les situations financières des ménages, permettant d'enrichir les outils de statistique publique, et ainsi plus largement le débat public.

Pour Crédit Mutuel Alliance Fédérale : Crédit Mutuel Alliance Fédérale, première banque à adopter le statut d'entreprise à mission, participe à ces études dans le cadre des missions qu'elle s'est fixées :

- Contribuer au bien commun en œuvrant pour une société plus juste et plus durable : pour Crédit Mutuel Alliance Fédérale, participer à l'information économique c'est contribuer au débat démocratique ;
- Protéger l'intimité numérique et la vie privée de chacun : Crédit Mutuel Alliance Fédérale veille à la protection absolue des données de ses clients. Toutes les analyses réalisées dans le cadre de cette étude ont été effectuées sur des données strictement anonymisées et sur les seuls systèmes d'information sécurisés du Crédit Mutuel et hébergés en France.

► Bibliographie

- ADAM, Marine, BONNET, Odran et LOISEL, Tristan, 2022. Avec l'inflation, une précarité financière en légère hausse, mais inférieure en août 2022 à son niveau d'avant-crise sanitaire. In : *Insee Analyses*. [en ligne]. 13 octobre 2022. Insee. N° 76. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/6542263>.
- ADAM, Marine, BONNET, Odran, FIZE, Étienne, RAULT, Marion, LOISEL, Tristan et WILNER, Lionel, 2023. How does fuel demand respond to price changes? Quasi-experimental evidence based on high-frequency data. In : *Documents de travail*. [en ligne]. 19 juillet 2023. Insee. N° 2023-17. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/en/statistiques/7644836>.
- ADAM, Marine, BONNET, Odran, FIZE, Étienne, RAULT, Marion, LOISEL, Tristan et WILNER, Lionel, 2024. Cross-border shopping for fuel at the France-Germany border. In : *Documents de travail*. [en ligne]. 7 mai 2024. Insee. N° 2024-08. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/en/statistiques/8183838>.
- ANDERSEN, Asger L., JENSEN, Amalie S., JOHANNESSEN, Niels, KREINER, Claus T., LETH-PETERSEN, Soren et SHERIDAN, Adam, 2023. How Do Households Respond to Job Loss? Lessons from Multiple High-Frequency Datasets. In : *American Economic Journal: Applied Economics*. Octobre 2023. Vol. 15, N° 4, pp. 1-29.
- ANDRÉ, Mathias et MESLIN, Olivier, 2022. Patrimoine immobilier des ménages : enseignements d'une exploitation de sources administratives exhaustives. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 107-125. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6035944?sommaire=6035950>.
- ASTIER, Jeanne, FACK, Gabrielle, FOURNEL, Julien, MAISONNEUVE, Flavie et SALEM, Ariane, 2024. Performance énergétique du logement et consommation d'énergie : les enseignements des données bancaires. In : *Focus*. [en ligne]. Janvier 2024. CAE. N° 103. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/performance-energetique-du-logement-et-consommation-d-energie-les-enseignements-des-donnees-bancaires>.
- BAKER, Scott R., FARROKHANIA, Robert A., MEYER, Steffen, PAGEL, Michaela et YANNELIS, Constantine, 2020. How Does Household Spending Respond to an Epidemic? Consumption during the 2020 COVID-19 Pandemic. In : *The Review of Asset Pricing Studies*. [en ligne]. Décembre 2020. Volume 10, N° 4, pp. 834-862. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://academic.oup.com/raps/article/10/4/834/5874450>.
- BLANCHET, Didier et GIVORD, Pauline, 2018. Les Big Data : quelles perspectives pour la statistique publique ? In : *Enjeux numériques*. [en ligne]. Juin 2018. N° 2. Big Data : économie et régulation. [Consulté le 29 juillet 2024]. Disponible à l'adresse : https://www.anales.org/enjeux-numeriques/2018/en_02_06_18.html.

- BOEHM, Johannes, FIZE, Etienne et JARAVEL, Xavier, 2023. Five Facts about MPCs: Evidence from a Randomized Experiment. In : *site github de Johannes Boehm*. [en ligne]. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://jmboehm.github.io/Helicopter.pdf>.
- BONNET, Odran, LOISEL, Tristan et OLIVIA, Tom, 2021. Impact de la crise sanitaire sur un panel anonymisé de clients de La Banque Postale – Les revenus de la plupart des clients ont été affectés de manière limitée et temporaire. In : *Insee Analyses*. [en ligne]. 3 novembre 2021. Insee. N° 69. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5760458>.
- BONNET, Odran, DIOP, Ouleye, LOISEL, Tristan, OLIVIA, Tom et WILNER, Lionel, 2023. La situation financière des ménages au jour le jour. In : *Insee Analyses*. [en ligne]. 5 décembre 2023. Insee. N° 90. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/7721885>.
- BUDA, Gergely, CARVALHO, Vasco M., HANSEN, Stephen, ORTIZ, Alvaro, RODRIGO, Tomasa et RODRIGUEZ MORA, José V., 2022. National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. In : *Cambridge Working Papers in Economics*. [en ligne]. Juillet 2022. 2244. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://ideas.repec.org/p/cam/camdae/2244.html>.
- COTTON, Franck et HAAG, Olivier, 2023. L'intégration des données administratives dans un processus statistique – Industrialiser une phase essentielle. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 104-125. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635829?sommaire=7635842>.
- EPAULARD, Anne, FIZE, Étienne, LE CALVÉ, Titouan, MARTIN, Philippe, PARIS, Hélène, PARRA RAMIREZ, Kevin et SRAER, David, 2021. La situation financière des PME/TPE en août 2021 au vu de leurs comptes bancaires. In : *Focus*. [en ligne]. Septembre 2021. CAE. N° 65. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/la-situation-financiere-des-pme-tpe-en-aout-2021-au-vu-de-leurs-comptes-bancaires>.
- FIZE, Étienne, PARIS, Hélène et RAULT, Marion, 2022. Quelle situation financière des entreprises et des ménages deux ans après le début de la crise Covid ? In : *Focus*. [en ligne]. Mars 2022. CAE. N° 83. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/quelle-situation-financiere-des-entreprises-et-des-menages-deux-ans-apres-le-debut-de-la-crise-covid>.
- FIZE, Étienne, MARTIN, Philippe, PARIS, Hélène et RAULT, Marion, 2022. La situation financière des ménages en début de crise énergétique. In : *Focus*. [en ligne]. Juillet 2022. CAE. N° 88. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/la-situation-financiere-des-menages-en-debut-de-crise-energetique>.
- HILLION, Mélina, 2024. Une évaluation des achats transfrontaliers de tabac et des pertes fiscales associées en France. In : *Documents de travail*. [en ligne]. 16 avril 2024. Insee. N° 2024-06. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8172202>.

- INSEE, 2021. En 2020, la chute de la consommation a alimenté l'épargne, faisant progresser notamment les hauts patrimoines financiers : quelques résultats de l'exploitation de données bancaires. In : *Note de conjoncture*. [en ligne]. 11 mars 2021. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5232043>.
- LECLAIR, Marie, 2019. Utiliser les données de caisses pour le calcul de l'indice des prix à la consommation. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 61-75. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254225?sommaire=4254170>.
- LE MONDE, 2020. Covid-19 : la crise sanitaire a fait basculer un million de Françaises et de Français dans la pauvreté. In : *Le Monde*. [en ligne]. 6 octobre 2020. [Consulté le 29 juillet 2024]. Disponible à l'adresse : https://www.lemonde.fr/societe/article/2020/10/06/un-million-de-nouveaux-pauvres-fin-2020-en-raison-de-la-crise-due-au-covid-19_6054872_3224.html.
- OLAFSSON, Arna et PAGEL, Michaela, 2018. The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software. In : *The Review of Financial Studies*. [en ligne]. 25 avril 2018. Volume 31, N° 11, pp. 4398-4446. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://ideas.repec.org/a/oup/rfinst/v31y2018i11p4398-4446..html>.
- POUGET, Julien, 2020. Nouvelles données pour suivre la conjoncture économique pendant la crise sanitaire : quelles avancées ? Quelles suites ? In : *Le blog de l'Insee*. [en ligne]. 28 juillet 2020. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://blog.insee.fr/nouvelles-donnees-pour-suivre-la-conjoncture-economique-pendant-la-crise-sanitaire-quelles-avancees-quelles-suites/>.
- REDOR, Patrick, 2023. Confidentialité des données statistiques : un enjeu majeur pour le service statistique public. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 46-63. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635823?sommaire=7635842>.
- RIVIÈRE, Pascal, 2018. Utiliser les déclarations administratives à des fins statistiques. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 14-24. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3647013?sommaire=3647035>.
- TAVERNIER, Jean-Luc, 2022. Crise Covid et mesure de la pauvreté. In : *Constructif*. [en ligne]. Juin 2022. N° 62 [Consulté le 29 août 2024]. Disponible à l'adresse : http://www.constructif.fr/bibliotheque/2022-6/crise-covid-et-mesure-de-la-pauvrete.html?item_id=6819.