

Un panorama de l'utilisation des données alternatives pour l'analyse économique

Commission CNIS « Système financier et financement de l'économie » du 15 mai 2025



- 1 DÉFINITION DES DONNÉES ALTERNATIVES**
- 2 INTÉRÊT DE CES DONNÉES ALTERNATIVES**
- 3 LIMITES DE CES DONNÉES ALTERNATIVES**

01 QU'ENTEND-ON PAR DONNÉES ALTERNATIVES ?

AU DÉPART, LE BIG DATA

- Multiplication des traces numériques, capacité croissante à les stocker ou à les analyser
- Données beaucoup plus volumineuses mais aussi avec des formats plus complexes et variables
- Des enjeux techniques, informatiques à traiter ces données

Premier exemple à l'Insee : l'utilisation des données de caisses pour les indices de prix à la consommation.

PUIS LES DONNÉES PRIVÉES

- Une nouvelle manière de collecter les données à côté des enquêtes et recensements d'une part et des données administratives d'autre part
- Un focus mis sur la propriété des données et l'accès à ces données par le statisticien plutôt que sur la difficulté technique à les traiter
- De fait, ces données peuvent déjà être pré-agrégées

Les données bancaires, du GIE CB, les données de téléphonie mobile, données des plateformes de location de courte durée mais aussi la consommation d'électricité font partie de ces exemples.

ET ENFIN LES DONNÉES ALTERNATIVES

- Les nouvelles techniques de traitement des données ouvrent des possibilités de collecter l'information elle-même et de créer des bases de données
(la notion de bases de données et de data dans le futur système européen des comptes)
- Les possibilités offertes par le web scraping
 - Les prix des trains, des avions pour l'IPC, la base de données JOCAS de la Dares pour les emplois vacants
- L'analyse d'image
 - L'utilisation d'images satellitaires : prédiction du rendement à la parcelle, production de ciment pour la Banque de France
 - L'extraction de données : les comptes sociaux des entreprises
- L'analyse textuelle
 - Analyse des conventions collectives pour déterminer les augmentations de salaire de la Banque de France
 - Analyse des google trends, tweet, articles de presses sur les anticipations des agents économiques

02

QUELLE UTILISATION DES DONNÉES ALTERNATIVES POUR LA PRODUCTION STATISTIQUE DE COURT TERME ?

REEMPLACER DES COLLECTES DE DONNÉES TRADITIONNELLES POUR LIMITER LA CHARGE D'ENQUÊTE

- Ce qui était prévu au départ dans le cadre légal
- Une manière également de convaincre les entreprises de nous donner accès à leurs données
- Le cas d'usage des données de caisses ou des données de transaction : remplacement de collecte de prix pour l'IPC et de l'enquête EMAGSA pour le calcul des chiffres d'affaires dans le commerce

CONTINUER À PRODUIRE DES STATISTIQUES REPRÉSENTATIVES FACE À UN CONTEXTE EN ÉVOLUTION

- Des champs qui deviennent/sont mal couverts par les statistiques traditionnelles

Indices de prix des trains avec yield management, location touristique avec le développement de plateformes de locations de courte durée, les offres d'emploi

Utilisation possible des données bancaires pour compléter budget des familles (question de la fenêtre d'observation)

- L'exemple de la crise sanitaire

Collecte de données traditionnelles altérée (entreprises fermées...)

OBTEINIR DES DONNÉES AVEC UNE GRANULARITÉ OU UNE FRÉQUENCE PLUS IMPORTANTE POUR PRODUIRE DE NOUVELLES STATISTIQUES

- **Grande taille des échantillons qui permet d'étudier des sous-populations**
études sur l'impact sur la consommation ou la situation financière des ménages à partir des données bancaires
- **Haute fréquence et fraîcheur des données qui permet d'évaluer rapidement l'impact d'un phénomène économique**
Mesure de l'impact à très court terme d'un phénomène ponctuel comme les différents confinements, la hausse des prix de l'énergie...
Nowcasting

03

LES POINTS D'ATTENTION QUAND ON UTILISE DES DONNÉES ALTERNATIVES

ACCÉDER À CES DONNÉES ALTERNATIVES DE MANIÈRE PÉRENNE ET SÉCURISÉE

- Cadre juridique, convention, partenariat
- Pérennité des structures privées produisant ces données et des données elles-mêmes

LA QUESTION DU CHAMP ET DES CONCEPTS

- Un problème de représentativité : des données sur un champ partiel

Exemple des données bancaires ou de la téléphonie mobile

- Des concepts qui ne correspondent pas toujours aux concepts des statisticiens

Exemple des données bancaires : revenu, consommation, ménage

- Comme pour les données administratives, nécessité d'avoir des métadonnées et d'échanger avec les producteurs de données pour comprendre le cadre de production de la donnée et son utilisation première comme données de gestion

LA STABILITÉ ET LA COMPARABILITÉ DE CES DONNÉES DANS LE TEMPS ET DANS L'ESPACE

- Exemple des données de transaction par CB et augmentation des paiements en CB
- Exemple des données de téléphonie mobile quand zones mal couvertes
- Problème de saisonnalité de ces données à haute fréquence

LE COÛT DE TRAITEMENT DE CES DONNÉES

LA CONFIDENTIALITÉ DES DONNÉES

4

CONCLUSION

– Où en est-on ?

La crise Covid a été un catalyseur ; des études ont été poursuivies pour comprendre les limites et les apports de ces données

Mais des limites demeurent ;

Ces données doivent être vues comme un complément de sources traditionnelles

– Le rôle important des travaux européens

- S'inspirer de nos voisins sur les types de données alternatives qu'ils mobilisent, les traitements méthodologiques à leur apporter
- Négocier au niveau européen l'accès à nos données

Retrouvez-nous sur

insee.fr



Marie Leclair

Cheffe du département des comptes nationaux

Marie.leclair@insee.fr