



Covid-19 et prix immobiliers au Royaume-Uni : quels enseignements des données extraites d'Internet (*web-scraping*) ?

Les statistiques officielles sur le marché immobilier fournissent en général des données agrégées et/ou retardées. En complément, de nombreuses informations granulaires, librement accessibles en temps réel sur les sites d'annonces immobilières, peuvent être récupérées grâce à des techniques de *web-scraping* (récupération automatisée de données sur Internet). En s'appuyant sur l'exemple du Royaume-Uni, cet article illustre l'intérêt d'une base de données quotidiennes, granulaires et disponibles en temps réel, notamment pour l'analyse du comportement des vendeurs. Ainsi, ces données confirment, pendant la première période de confinement, un gel de l'activité et un attentisme des vendeurs. L'approche granulaire signale en outre de grands écarts entre les régions : alors que les prix affichés ont été stables, voire ont augmenté après le premier confinement dans les régions plus rurales, ils ont enregistré un déclin continu à Londres.

Jean-Charles BRICONGNE, Nicolas CHATELAIS, Baptiste MEUNIER * et Sylvain POUGET **

Direction générale des Statistiques, des Études et de l'International

Direction de l'Économie et de la Coopération internationales

Codes JEL
E01, R30

Jérôme COFFINET * et Etienne KINTZLER

Direction générale du Système d'information

Direction des Données et des Services analytiques

* Unité de rattachement au moment de la réalisation de cette étude.

** En stage à la Banque de France au moment de la réalisation de cette étude.

Cet article présente le résultat de travaux de recherche menés à la Banque de France. Les idées exposées dans ce document reflètent l'opinion personnelle de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Les éventuelles erreurs ou omissions sont de la responsabilité des auteurs.

80%

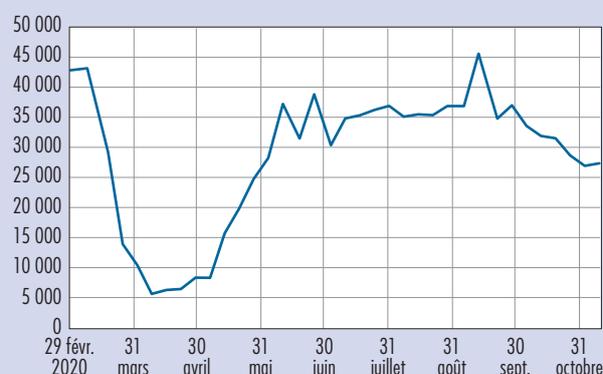
au moins de baisse du nombre de nouvelles offres immobilières avec le premier confinement au Royaume-Uni

90%

des offres encore disponibles après un mois sur Internet au plus haut du premier confinement, contre environ les deux tiers en dehors

Nouvelles offres immobilières hebdomadaires

(en unités)



Sources : Zoopla et calculs des auteurs.



1 Les données issues des sites immobiliers : un complément utile des statistiques immobilières officielles

Recourir à des données alternatives permet de construire des indicateurs avec davantage de rapidité (indices quotidiens en temps réel vs mensuel/trimestriel avec délai) **et de granularité** (au niveau du code postal par exemple) que ce que permettent les statistiques officielles. Obtenir un niveau de granularité plus élevé permet, dans le cas présent, de détecter des déséquilibres sur les marchés locaux du logement. Dans une perspective plus large, cet article reflète l'effort récent des économistes pour concevoir des indicateurs à haute fréquence (par exemple Lewis *et al.*, 2020) – dont la Covid-19 et les soudaines fluctuations d'activité ont rendu la disponibilité d'autant plus pressante (cf. par exemple Bricongne *et al.*, 2020). En effet, tirer parti de ces informations en temps réel est encore plus essentiel pendant les épisodes de crise tels que la pandémie de la Covid-19 dans la mesure où les délais de publication des statistiques traditionnelles ne permettent de saisir les retournements conjoncturels qu'avec retard.

Ainsi, de nombreuses informations sont disponibles publiquement et en temps réel sur les sites de ventes immobilières. Cela est particulièrement vrai pour le marché résidentiel au Royaume-Uni où 92 % des agences immobilières publient des annonces sur Internet ¹. La littérature économique comprend des exemples d'utilisation de données alternatives pour suivre le marché immobilier, comme Kulkarni *et al.* (2009) avec les *Google Trends* ou, de manière plus récente, de recours au *web-scraping* (récupération automatisée de données sur Internet) pour analyser le marché du logement (par exemple Hanson et Santas, 2014 ou Bricongne *et al.*, 2019). Le présent article se concentre sur le point de vue des vendeurs (ce que permet précisément de réaliser le téléchargement automatique de données sur les sites immobiliers, où les données sont postées par ceux-ci, par opposition avec les données de transactions finales

qui résultent de l'interaction entre vendeurs et acheteurs) et dérive des indicateurs de leur point de vue. Il couvre le marché britannique dans cette période particulière marquée par la Covid-19 et le Brexit – bien que cette approche puisse être étendue de manière transparente à d'autres zones géographiques.

Les statistiques officielles sur l'immobilier résidentiel sont relativement lacunaires et disponibles avec retard.

Au Royaume-Uni, les principaux indicateurs (indices de prix, enquêtes sur le secteur de la construction, indice de la construction de nouveaux logements, données détaillées sur les transactions, etc.) ont une fréquence en général mensuelle et sont publiés au moins dix à quinze jours environ après la fin du mois, voire plus selon les sources et les types de données. Ce délai est en général supérieur dans la plupart des autres pays de l'OCDE, ce qui renforcerait l'intérêt de disposer de statistiques alternatives. En outre, au Royaume-Uni, les statistiques ne sont pas toujours disponibles à un niveau infranational très détaillé, alors que des écarts entre les zones urbaines – notamment Londres – et rurales ont été documentés dans la littérature (par exemple Poon et Garratt, 2012). Les données de transactions individuelles sont en revanche disponibles, ce qui ne concerne encore que quelques pays dont la France ².

2 Deux innovations importantes : un suivi en temps réel et du point de vue du vendeur

À l'aide des données extraites du *Web*, **des statistiques quotidiennes** sur le marché du logement sont disponibles en avance par rapport aux statistiques officielles. Fait intéressant, les données de localisation peuvent être également utilisées pour construire ces indicateurs à un niveau très granulaire. Cela permet de suivre plusieurs indicateurs habituels (par exemple prix, nombre de nouveaux projets de construction) avec une rapidité améliorée et une granularité maximale par rapport aux statistiques officielles.

¹ Source : *National Association of Realtors*, disponible sous ce lien : <https://www.nar.realtor/sites/default/files/documents/2018-real-estate-in-a-digital-world-12-12-2018.pdf>
S'agissant de la couverture des sites utilisés, si elle peut varier selon les régions, elle reste toutefois suffisamment élevée dans tous les cas pour assurer une bonne représentativité au niveau infranational.

² À ce sujet, cf. : <https://app.dvf.etalab.gouv.fr/>



ENCADRÉ

Une base de données journalières très granulaire

Les données sont récupérées par le biais d'un téléchargement automatique quotidien des données sur Internet des sites majeurs de petites annonces immobilières au Royaume-Uni : Rightmove, Zoopla et OnTheMarket. Pour améliorer la couverture sur des régions plus spécifiques, PropertyPal – un site immobilier spécialisé dans l'Irlande du Nord – et S1Homes, son *alter ego* pour l'Écosse, sont également couverts. Grâce à ces sites *Web*, plus de 1,5 million d'offres immobilières sont téléchargées en moyenne chaque jour, correspondant au stock total des offres disponibles. Environ les deux tiers de ces dernières sont des offres de vente (cf. tableau). Les données pour Zoopla sont téléchargées depuis début mars 2020, alors que les données d'autres sites *Web* sont disponibles depuis juillet 2020.

Nombre moyen d'offres *web-scrapées* par jour

(par URL unique)

Nom du site	Résidentiel		Commercial	
	Vente	Location	Vente	Location
Zoopla	300 000	200 000	15 000	35 000
Rightmove	500 000	200 000	20 000	25 000
OnTheMarket	250 000	100 000	5 000	10 000

Note : L'URL, pour *Uniform Resource Locator*, désigne communément une adresse Internet, correspondant en l'occurrence à une page d'annonce.

Sources : Sites Zoopla, Rightmove et OnTheMarket ; calculs des auteurs.

Pour chaque offre, les informations décrivant en détail le logement, le type de transaction et le prix, et surtout son emplacement, sont récupérées à un niveau très granulaire. Les informations peuvent différer selon les sites, mais pour le fournisseur principal Zoopla, les informations suivantes sont disponibles pour les offres immobilières résidentielles : ville, code postal, adresse, superficie, type de transaction (vente, location), type de bâtiment (logement neuf ou ancien), nombre de chambres, nombre de salles de bain, nombre de pièces à vivre, description générale (par exemple « bel appartement avec terrasse dans un quartier sûr »), et le type de bien (par exemple, appartement, duplex, propriété). Avoir des informations détaillées, non seulement objectives, mais éventuellement aussi plus subjectives, permet de mieux identifier les éléments qui peuvent affecter le prix d'une offre (emplacement, type de bien, quartier, etc.). En particulier, récupérer la description générale permet de vérifier s'il contient des mots-clés faisant référence à des installations supplémentaires (par exemple « balcon », « terrasse », « garage ») qui peuvent avoir un impact à la hausse sur le prix.

Une fois les données récupérées, leur qualité est vérifiée. S'agissant des surfaces, les données sont manquantes pour une proportion élevée d'offres qui concernent certains types de bien. La surface est en général renseignée pour les logements de qualité supérieure susceptibles d'intéresser les acheteurs internationaux, comme les grands appartements à Londres. Du fait de ce biais de sélection, les auteurs choisissent de ne pas utiliser les données de surface et de se concentrer sur les prix absolus des logements plutôt que rapportés aux surfaces, ce que l'on retrouve d'ailleurs dans la plupart des statistiques officielles. On peut alors supposer implicitement, lorsqu'on analyse les évolutions de prix absolus, que les surfaces moyennes évoluent peu au cours du temps, même si certains épisodes atypiques comme ceux des confinements peuvent conduire à des effets de composition éventuellement non négligeables. Les doublons qui peuvent exister sur un même site sont également éliminés. Pour les logements partageant la même adresse, la même description et le même prix, une seule offre est conservée.

.../...



Cependant ce tri est opéré uniquement pour les logements anciens. Dans le cas des logements neufs, une offre apparemment similaire peut correspondre à différentes propriétés car plusieurs logements présentant des caractéristiques comparables peuvent être construits simultanément sur un même site.

Les données sont également filtrées en fonction de leur catégorie car le choix a été fait de se focaliser sur l'immobilier résidentiel et de ne pas traiter l'immobilier commercial, même si cela est faisable et utile, au moins sur certains segments, compte tenu des enjeux de stabilité financière. L'immobilier commercial a en effet des caractéristiques et des cycles distincts et comporte moins d'observations. Les biens hors logement tels que les parkings, les mobile homes, ou les terrains, sont également exclus. Par ailleurs, un certain nombre de biens immobiliers sont négociés par enchères au Royaume-Uni : dans ce cas, le prix affiché est en fait le prix de réserve. Par conséquent, ce type d'offres est écarté. Les valeurs extrêmes restantes sont supprimées en effectuant une winsorisation¹ au seuil de 1 %.

¹ Procédure statistique visant à réduire l'impact de données extrêmes ou aberrantes peu représentatives dans l'estimation d'un paramètre de distribution de type moyenne ou écart-type.

De plus, construire des données à partir des offres affichées/mises en ligne **donne le point de vue des vendeurs** – ce qui n'est pas le cas des statistiques habituelles fondées sur les données de transaction issues de la confrontation de l'offre et de la demande. Cette approche innovante permet, par exemple, de suivre le **nombre de nouvelles offres** qui, corrigé des effets saisonniers, indique si les vendeurs sont désireux de proposer leurs biens sur le marché³. Un écart par rapport à sa valeur « normale » signifierait que les vendeurs anticipent (ou sont confrontés à) un choc sur le marché immobilier. Ces données permettent également de suivre **l'évolution du prix d'une offre dans le temps**. Un écart à la tendance « normale » des corrections (car même dans des conditions économiques stables, les vendeurs peuvent ajuster leurs prix au fil du temps pour corriger leurs évaluations initiales) pourrait signaler un épisode de tension. Cette nouvelle approche repose sur des offres, donc à un stade précoce du processus d'une vente immobilière, ce qui devrait *a priori* permettre une détection plus rapide des retournements du marché. À l'inverse, les statistiques officielles sont établies à partir des transactions réalisées. À l'extrême, si les vendeurs ont plus de difficultés à vendre leurs logements, les données de transactions pourraient se tarir alors qu'une information abondante sur les révisions à la baisse des biens postés serait toujours disponible. Il convient toutefois de signaler qu'une partie de l'ajustement des

prix peut se faire au moment de la transaction, traduisant la marge de négociation, qui informe sur l'état du marché et que l'on peut reconstruire en appariant, autant que possible, offres postées et transactions (cf. *infra*).

Les statistiques calculées du point de vue du vendeur ne doivent donc pas nécessairement être comparées aux statistiques officielles dans la mesure où elles offrent plutôt un point de vue complémentaire⁴.

Par ailleurs, les innovations du suivi en temps réel et du point de vue du vendeur ne sont pas propres au cas du Royaume-Uni, puisqu'elles seraient transposables à la plupart des autres pays.

3 Illustration avec le marché britannique du logement résidentiel pendant la crise de la Covid-19

Un premier indicateur illustratif est le nombre de nouvelles offres – qui indique si les vendeurs trouvent le moment opportun pour vendre leur propriété. Le graphique 1 *infra* affiche le nombre hebdomadaire (qui présente l'avantage de neutraliser les effets saisonniers au sein d'une semaine) de nouvelles offres. Ce nombre a enregistré une forte variation, ayant été cinq à six fois moins important lors du (premier) confinement qu'après ou avant cette période.

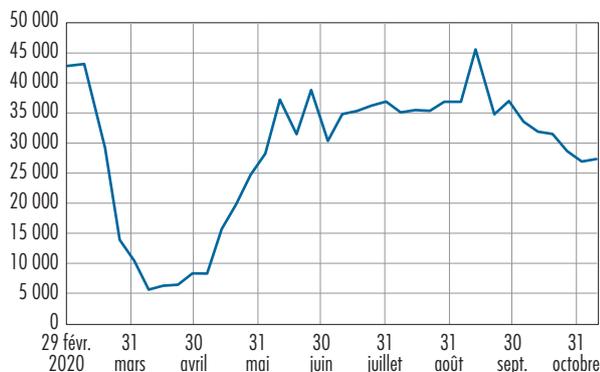
³ Il est également possible d'appréhender le point de vue des acheteurs à travers par exemple les recherches qui sont effectuées sur Internet.

⁴ Pour une éventuelle comparaison entre les deux approches, il faut par ailleurs garder en tête que les chiffres publiés au Royaume-Uni décrivent en général des évolutions médianes plutôt que moyennes, ce qui peut induire des écarts importants surtout dans des zones comme Londres avec une forte proportion de biens onéreux. Un calcul des prix médians à partir des données issues du *web-scraping* permet d'obtenir des ordres de grandeur plus proches des chiffres officiels agrégés.



G1 Nouvelles offres immobilières hebdomadaires

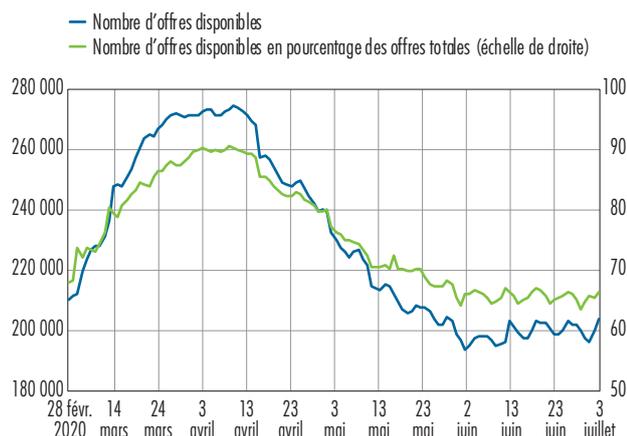
(en unités)



Sources : Zoopla et calculs des auteurs.

G2 Offres encore disponibles après un mois sur Internet

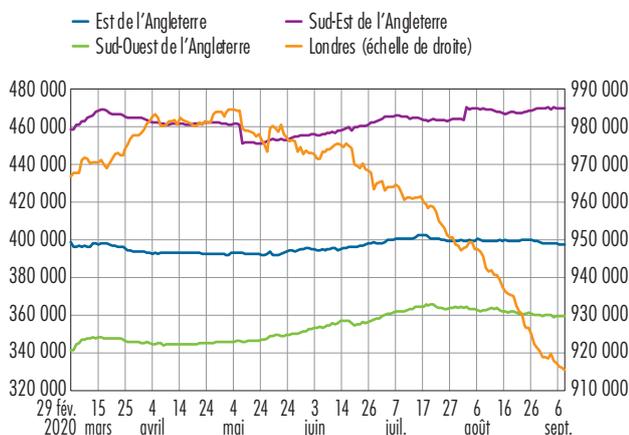
(en unités et en %)



Sources : Zoopla et calculs des auteurs.

G3 Prix de vente moyen affiché par région

(en livres sterling par logement)



Sources : Zoopla et calculs des auteurs.

Un autre indicateur est le nombre d'offres encore disponibles après un mois sur Internet (cf. graphique 2) : l'effet du confinement est visible avec environ 90% des offres publiées fin mars/début avril qui étaient encore disponibles un mois plus tard. Après le confinement, la proportion diminue progressivement et se stabilise autour de son niveau de pré-confinement soit environ les deux tiers du total.

La granularité des données permet d'analyser les prix au niveau local. Le graphique 3 illustre l'évolution des prix dans quatre régions de l'Angleterre : l'Est de l'Angleterre, le Sud-Est, le Sud-Ouest et Londres. Alors que les tendances sont relativement similaires pour les trois premières avec une certaine stabilité voire une légère hausse tendancielle, le cas de Londres se démarque avec une baisse continue depuis mai. Comme il s'agit de prix absolus et non pas au mètre carré, cette baisse peut refléter un prix au mètre carré posté qui diminue, un effet de composition avec des biens vendus qui sont de surface moindre ou de moindre qualité, ou une combinaison des deux effets. Il est cependant raisonnable de penser que ces effets de composition restent d'ampleur relativement limitée sur des périodes assez courtes, en tout état de cause inférieure aux évolutions spectaculaires identifiées dans le graphique 3.

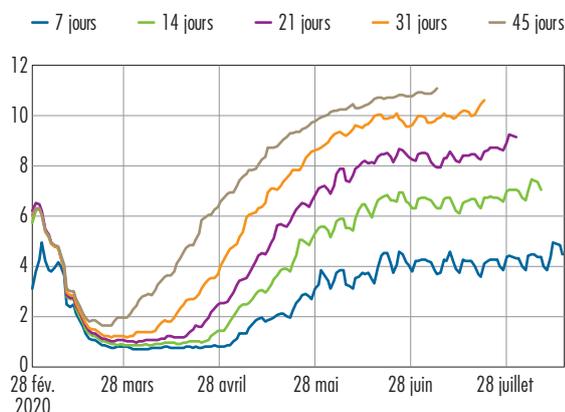
Les prix affichés sur le graphique 3 sont des prix moyens alors que les données officielles publiées se rapprochent en général davantage des prix médians. L'objet de la présente étude n'est pas de reproduire les statistiques officielles, mais plutôt de les compléter. Il n'en demeure pas moins qu'un calcul de prix médians donnerait des ordres de grandeur proches des statistiques publiées. Il est à noter que les tendances de prix sont identiques en utilisant les prix médians ou moyens. De même, pour pallier les effets de composition entre différents types de biens (studios, deux-pièces, etc.), des indices de prix par types de biens sont calculés : durant la crise de la Covid-19, les tendances sont aussi comparables entre catégories de biens.

À l'aide des données extraites du Web, il est également possible d'analyser la façon dont le prix fluctue pour chaque offre. Ainsi, pour chaque offre publiée un jour donné, les auteurs vérifient si elle est toujours disponible



le lendemain et si son prix a changé ; le même processus est ensuite répété jour après jour. En comparant les offres pour lesquelles le prix évolue sur le nombre total d'offres, il semble que seule une minorité (entre 8 et 10 %) connaît des changements de prix, 90 % de ces changements étant des révisions à la baisse. Le graphique 4 offre un aperçu de la part des offres avec des changements de prix après 7, 14, 21, 31 et 45 jours sur le marché. Ce graphique illustre qu'au début du confinement, les vendeurs n'étaient pas encore prêts à modifier leurs prix et ont préféré adopter une approche attentiste : une part nettement inférieure des offres a subi des changements de prix par rapport à la situation de pré-confinement, même après une période prolongée. Cependant, après le premier confinement, la part des offres avec une variation de prix (essentiellement des baisses) a augmenté rapidement et s'est stabilisée autour de niveaux supérieurs, par rapport à la période de pré-confinement, surtout pour les offres étant restées plus longtemps sur le marché.

G4 Part des offres dont le prix est révisé, selon le nombre de jours (en %)



Note : Les révisions de prix peuvent être à la hausse ou à la baisse, même si les baisses sont largement prépondérantes (environ 90 % des révisions), les hausses de prix concernant majoritairement les logements neufs (environ 10 % de l'échantillon des logements). L'inclusion des logements neufs (ainsi que leur poids), est cohérente avec le périmètre retenu par l'ONS (*Office for National Statistics*, Bureau des statistiques nationales du Royaume-Uni) pour le calcul de l'indice des prix immobiliers.

Sources : Zoopla et calculs des auteurs.

4 En conclusion, une vision plus fine et de nouveaux indicateurs potentiels de tension sur les marchés immobiliers

Comme l'illustre cet article, les données issues du *web-scraping* offrent une vision plus fine et granulaire que les statistiques traditionnelles, et permettent une analyse du point de vue du vendeur, particulièrement utile pendant les épisodes particuliers comme la pandémie de Covid-19.

En outre, ces données faciliteraient la construction de nouveaux indicateurs de tension. En appariant les données de la transaction ⁵ et de l'annonce immobilière, il serait possible de calculer l'écart entre le prix observé et le prix offert ⁶, et ainsi de mesurer la marge de négociation des acheteurs, selon les zones géographiques et la période. Une augmentation de cette marge peut être le signe d'un retournement du marché.

Enfin, les mêmes méthodes de *web-scraping* peuvent être utilisées pour d'autres variables que les prix des logements : loyers, prix des terrains à bâtir, etc. qui offrent à leur tour la possibilité d'établir d'autres indicateurs utiles au suivi des marchés immobiliers, comme les ratios loyers sur prix, qui fournissent une indication des rendements.

⁵ Les informations disponibles sur les deux sources n'étant pas exactement identiques, on procède à un appariement par l'algorithme des plus proches voisins en identifiant les biens les plus proches qui ont la plus forte probabilité de correspondre, compte tenu de leur localisation, de leurs caractéristiques et de la période.

⁶ Les calculs effectués montrent une marge de négociation qui est inférieure à Londres par rapport aux autres régions.



Bibliographie

Bricongne (J.-C.), Coffinet (J.), Delbos, (J.-B.), Kaiser (V.), Kien (J.-N.), Kintzler (E.), Lestrade (A.), Meunier (B.), Mouliom (M.) et Nicolas (T.) (2020)

« Le suivi économique en période de Covid-19 : l'apport d'indicateurs à haute fréquence », *Bulletin de la Banque de France*, n° 231/5, septembre-octobre.

[Télécharger le document](#)

Bricongne (J.-C.), Turrini (A.) et Pontuch (P.) (2019)

« Assessing house prices: insights from "Houselev", a dataset of price level estimates », *European Commission Discussion Paper*, n° 101.

Hanson (A.) et Santas (M.) (2014)

Field experiment tests for discrimination against Hispanics in the US rental housing market, *Southern Economic Journal*, vol. 81, n° 1, p. 135-167.

Kulkarni (R.), Haynes (K.), Stough (R.) et Paelinck (J.) (2009)

« Forecasting housing prices with Google Econometrics », *GMU School of Public Policy Research Paper*, n° 2009-10.

Lewis (D.), Mertens (K.) et Stock (J.) (2020)

« Monitoring real activity in real time: the weekly economic index », *Federal Reserve Bank of New York Liberty Street Economics*, n° 20200330b.

Poon (J.) et Garratt (D.) (2012)

« Evaluating UK housing policies to tackle housing affordability », *International Journal of Housing Markets and Analysis*, vol. 5, n° 3, p. 253-271.

Éditeur

Banque de France

Secrétaire de rédaction

Nelly Noulin

Directeur de la publication

Gilles Vaysset

Réalisation

Studio Création

Direction de la Communication

Rédaction en chef

Françoise Drumetz

ISSN 1952-4382

Pour vous abonner aux publications de la Banque de France

<https://publications.banque-france.fr/>

Rubrique « Abonnement »

