

Améliorer l'observation des marchés grâce aux fichiers fonciers

Le cas des sociétés civiles immobilières

Jusqu'à une période récente, les propriétaires immobiliers en France étaient connus au travers de statistiques partielles¹ ou limitées à un territoire restreint². C'était en particulier le cas de ceux regroupés en sociétés civiles immobilières (SCI). À l'exception de quelques travaux concernant leur présence dans les grandes villes françaises en 2002³ (De Moncan) et les logiques sociales associées à ces acteurs de marché dans la région lyonnaise (Ruz, 2005), les SCI sont des acteurs méconnus qui suscitent beaucoup de questions.

L'ouverture des bases de données statistiques localisées sur la propriété cadastrale rend désormais possible l'appréhension globale, fine et quantifiée de ces phénomènes. Or, l'exploration des fichiers fonciers met en évidence qu'il n'est pas possible de déterminer de manière univoque une caté-

gorie de propriétaires. L'information les concernant y apparaît en effet composite, en raison notamment de son caractère multi-sources (elle provient de différents services de la DGFIP). Il n'existe pas, alors, de variable discriminante pour déterminer les contours de différentes propriétés comme le démontre ici l'exemple des SCI. Dès lors, comment observer un marché dans ce contexte incertain ?

La question apparaît d'autant plus cruciale que les données de la DGFIP s'imposent peu à peu auprès des acteurs, notamment publics, comme la source de référence pour l'observation des parcs et des marchés fonciers-immobiliers. La qualité, la fiabilité et l'utilisabilité de ces données nécessitent en conséquence d'être étudiées pour affiner la connaissance de leurs conditions d'exploitation.

À travers l'exemple des propriétaires de logements regroupés en SCI, nous interrogeons la possibilité d'objectiver un acteur de marché à partir des fichiers fonciers, la source de référence pour cette information. La proposition méthodologique mise en place consiste à consolider les observations issues de cette source et se fonde sur un principe de réduction d'incertitude. Les résultats contribuent également à la compréhension des sources d'erreurs potentielles liées aux différentes interprétations de l'objet juridico-social que sont les SCI.

Les fichiers fonciers : la source de référence pour l'observation de la propriété immobilière

La majorité des bases de données qui existent sur la propriété immobilière sont renseignées à l'occasion des mutations du parc de logement : les données sur les ventes foncières-immobilières (Perval-Bien,

DVE, EPTB) informent sur les nouveaux acquéreurs et les prix, les données sur la construction neuve (Sit@del2) fournissent des repères sur les grands acteurs qui supportent la production de logements et enfin, les données sur le parc de logements (BD logement, ENL) en livrent un portrait sous l'angle de leur mode d'occupation (propriétaires occupants ou non). La seule et unique source d'information exhaustive, localisée et régulièrement actualisée est celle issue du fichier des impôts, Majic. Depuis 2012, ces données sont consolidées et enrichies par le Cerema pour le compte de la DGALN, qui les a intégrées dans les fichiers fonciers. Ce reconditionnement, réalisé dans des formes proches des bases de données géographiques, présente l'avantage de faciliter l'exploitation des données. Le Cerema a également mis à disposition de nombreuses ressources qui ont facilité leur utilisation (modèles conceptuels de données, indices de confiance concernant les variables, tables géographiques exploitables sous SIG, etc.). L'observation des parcs et des marchés fonciers-immobiliers en a été considérablement facilitée.

Dans les fichiers fonciers, les informations sur la propriété cadastrale sont désormais groupées et spatialisées à la parcelle (ou à l'unité foncière), ce qui est par exemple utile pour identifier des gisements fonciers (Demarquet et Toujas, 2018 ; outils de prospection développés par le Cerema et l'Inra⁴ et par le Cerema et l'Agence foncière de Loire Atlantique⁵). Une autre utilité de cette donnée est d'offrir une lecture de la structure de la propriété foncière-immobilière.

⁴ <https://urbansimul.paca.inra.fr/urbansimul/>

⁵ https://www.cerema.fr/fr/system/files/documents/2017/12/171208-plaquette_GIRAF.pdf

¹ La répartition des propriétaires bailleurs et des propriétaires occupants livrée par l'Insee par exemple.

² Dans le cadre d'études préalables à une opération d'aménagement par exemple.

³ D'après l'auteur, les SCI détiennent 2,6 % du patrimoine immobilier de ces grandes villes.

Remerciements

Les auteurs adressent leurs remerciements à l'établissement public foncier Provence-Alpes-Côte d'Azur et en particulier à sa directrice générale Claude Bertolino pour son soutien dans les diverses activités de recherche engagées sur l'observation foncière et immobilière.

Laure Casanova Enault
Mathieu Coulon
Guilhem Boulay
Université d'Avignon
UMR Espace 7300

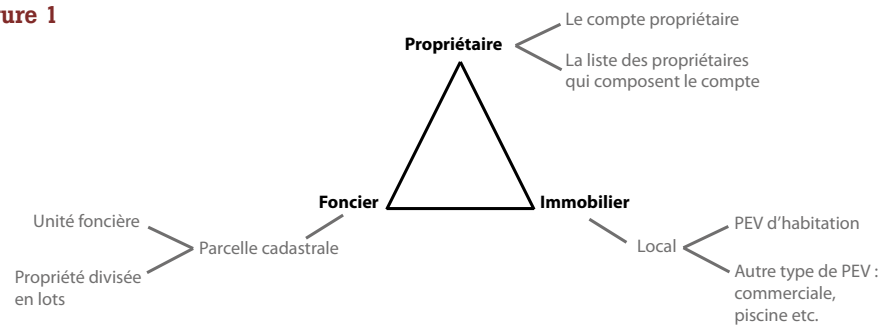
Ce type d'approche invite à des approfondissements de connaissance sur le lien entre les caractéristiques économiques, sociales et démographiques des propriétaires et sur le fonctionnement des marchés (Le Goix *et al.*, 2019 ; Desponds et Bergel, 2013). Il devient donc possible d'analyser finement des marchés de propriétaires occupants, des marchés concernés par des investisseurs privés ou publics (Bonneval, Pollard, 2017), des profils sociodémographiques des propriétaires (Casanova Enault, 2017), la localisation des patrimoines, etc.

Il n'en reste pas moins que ces données sont formatées par leur origine fiscale. Cela a deux conséquences :

- La première est l'obligation préalable de s'approprier certaines notions de fiscalité immobilière. En effet, plusieurs objets et catégories structurantes des fichiers fonciers sont délimités de ce point de vue : un logement est par exemple désigné par un local avec une partie d'évaluation (PEV) principale d'habitation, près de trois cent formes juridiques peuvent être associées à des droits de propriété sur un local (cf. variable DFORM) etc.
- La seconde est la nécessité pour obtenir une information sur la propriété immobilière, de relier plusieurs des six fichiers de base (relatifs à la propriété bâtie, à la propriété divisée en lots, à la propriété non bâtie, aux propriétaires etc.) qui composent les fichiers fonciers sans correspondre à une base de données relationnelle⁶ au sens strict. Les six fichiers proviennent chacun de services indépendants de la DGFIP (service fiscalité immobilière, centre des impôts fonciers etc.), qui qualifient les rapports de propriété selon un prisme particulier, lequel est fonction de la mission principale du service : enregistrer un acte notarié, appliquer l'impôt, etc. (Casanova *et al.*, 2017). Il est donc difficile de définir et d'identifier une catégorie de propriétaires immobiliers de manière univoque. L'exigence de rendre transparents les choix et les critères de sélection de l'information en est d'autant plus grande.

⁶ Guidé du modèle conceptuel de données fourni par le Cerema, l'utilisateur doit ainsi générer les tables de correspondances requises (notamment entre les comptes-propriétaires et les personnes propriétaires qui le composent) et questionner l'intégrité référentielle du modèle. Celle-ci peut ne pas être vérifiée si l'information est lacunaire. C'est le cas par exemple d'un local ne disposant pas de référence parcellaire valide.

Figure 1



La création d'une base de données resserrée autour des propriétaires immobiliers de type SCI

Afin de reconstituer la relation triangulaire entre le foncier, l'immobilier et le(s) propriétaire(s), trois fichiers de base doivent au minimum être liés : ceux des propriétés non bâties, des propriétés bâties et des propriétaires, qui caractérisent respectivement les notions/objets fiscaux de parcelle cadastrale, de local⁷ et de compte-propiétaire⁸. Or, la relation triangulaire « standard » (un compte-propiétaire composé d'une unique personne détenant un local d'habitation positionné sur une parcelle) n'est qu'une combinatoire parmi de nombreuses autres possibilités. Il existe différentes configurations de relations dites « de plusieurs à plusieurs »⁹, c'est-à-dire de relations multilatérales entre ces trois composantes (figure 1).

Par ailleurs, les variables¹⁰ qui décrivent les rapports de propriété du territoire sont éclatées dans ces trois fichiers. De manière paradoxale, les informations qu'ils contiennent peuvent à la fois apparaître en redondance

⁷ « Un local dans les fichiers fonciers correspond (hors locaux industriels) soit à une "propriété", s'il s'agit d'une "maison individuelle" comportant un ensemble de sols, terrains et bâtiments [...] destinés, en raison de leur agencement à être utilisés par un même occupant, soit par une "fraction de propriété" s'il s'agit d'un "appartement" situé dans un immeuble collectif. »

⁸ Le compte-propiétaire correspond au « compte communal » dans Majic. Il désigne « l'ensemble des personnes (dans la limite de six) exerçant des droits concurrents sur un ou plusieurs biens d'une commune » (DGFIP, 2011).

⁹ Un compte-propiétaire peut détenir un ou plusieurs locaux positionnés sur une ou plusieurs parcelles. Ce même compte-propiétaire peut être composé d'une ou plusieurs personnes dans la limite de six, lesquelles peuvent se retrouver dans plusieurs comptes propriétaires. Du côté des locaux, il n'y a pas de correspondance systématique entre un local et un bâtiment, de même la présence de plusieurs locaux sur une parcelle ne signifie pas nécessairement qu'il s'agit de collectifs d'habitation.

¹⁰ Le dictionnaire des variables du Cerema est accessible sur Internet à l'adresse : <http://doc-datafoncier.cerema.fr/ff/docff/>

(répétition d'une même modalité dans plusieurs variables) et ne pas se recouper (une même requête réalisée sur des variables différentes renvoie des résultats différents).

Cette situation met l'utilisateur face à une première alternative : doit-il déterminer a priori la variable la plus pertinente ou doit-il croiser toutes les variables susceptibles de renseigner la requête ? Dans le cas des fichiers fonciers où les variables sont complémentaires, l'expérience montre que c'est la deuxième option qui doit être retenue. L'utilisateur est alors confronté à une seconde alternative : retenir les individus statistiques à l'intersection des modalités des variables qui renvoient l'information recherchée (par exemple les propriétaires qui sont des SCI) ou hiérarchiser les variables selon le niveau de finesse de leur spécification (logique de l'entonnoir). La première option, fondée sur une logique d'intersection des modalités d'intérêt, est trop restrictive comme l'illustre l'exemple reporté dans le tableau 1. De plus, le recouvrement entre les résultats des deux requêtes est partiel et inconnu.

Quant à la seconde option, elle n'est tout simplement pas applicable car le niveau de raffinement de l'information porté par chaque variable n'est pas connu a priori.

Reste la possibilité d'appliquer un filtre en fonction du niveau de précision des modalités, donc de certitude quant aux résultats.

Retrouver ses petits... : un protocole pour réduire l'erreur de caractérisation des propriétaires immobiliers de type SCI

Un protocole méthodologique (tableau 2) est mis en place afin de consolider les requêtes relatives aux propriétaires immobiliers dans les fichiers fonciers. Celui-ci se fonde sur un *scoring* utilisant treize variables qui renseignent trois registres d'informations : l'identification des personnes morales et de leur type (ex. CCOGRM), la →

→ dénomination des propriétaires (ex. DDENOM) et leur identifiant d'appariement avec les bases Insee (ex. DSIREN). Ce *scoring* est appliqué aux modalités des variables en fonction du niveau de certitude de leurs spécifications (par exemple dans notre cas, « SCI » vs « autre personne morale »).

Le résultat (tableau 3) est la constitution de deux sous-populations statistiques d'intérêt : celle dont il est certain qu'il s'agit de propriétaires immobiliers de type SCI (A) et celle pour laquelle il y a une forte suspicion, mais aucune certitude (B).

Ce sont donc un peu plus de 9 700 comptes-propriétaires de type SCI détenant des droits de propriété sur des logements qui sont identifiés avec certitude dans le Vaucluse en 2015 et près de 33 200 pour le département des Bouches-du-Rhône la même année. Ils représentent 3 % du nombre total de comptes-propriétaires et 24 % des comptes propriétaires de type personne morale dans le 84. Ils représentent respectivement 5 % et 25 % dans les Bouches-du-Rhône.

Ces propriétaires détiennent des droits sur 27 400 logements en Vaucluse (soit 8,7 % du parc) et près de 66 700 dans les Bouches-du-Rhône (soit 6,3 % du parc). Les chiffres sont plus élevés que ceux avancés par De Moncan en 2002 et confirment dans le même temps, le constat établi par l'auteur concernant la répartition géographique inégale du phénomène (cartes 1 et 2).

Quelle marge d'erreur potentielle ?

D'après le tableau 3 restituant les résultats du protocole de réduction d'incertitude, la marge d'erreur potentielle est estimée à 15 % dans le cas du Vaucluse et à 23 % dans le cas des Bouches-du-Rhône, soit 19 % en moyenne sur deux départements.

Un autre enseignement de ces résultats est la forte variation du taux d'erreur d'un département à l'autre puisqu'on a 0,6 % d'erreur dans le Vaucluse et 10,4 % dans les Bouches-du-Rhône. De même, la marge d'erreur potentielle peut varier de manière significative d'une commune à une autre.

D'autres sources d'erreurs peuvent être introduites en lien avec l'activité de la société. Par exemple, certaines peuvent ne pas être actives ou en sommeil (Ruz, 2005). Il est nécessaire pour cela de vérifier les informations à partir des données d'entreprise accessibles par exemple via les Kbis.

Le test réalisé sur un échantillon de quatre

Tableau 1. Exemple de deux requêtes sur le nombre de propriétaires en SCI à partir de deux variables distinctes (Vaucluse, 2015).

Variables/ modalités de la requête	Résultat de la requête (en nombre de comptes propriétaires)
Variable : CCOGM Modalité : personnes morales représentant des sociétés	17 000
Variable : DDCOM Modalités : SCI, société civile immobilière etc.	3 200

Source : fichiers fonciers, 2015, DGFIP. Réalisation : les auteurs.

Tableau 2. Protocole de sélection des propriétaires immobiliers de type SCI dans les fichiers fonciers.

Étape du protocole	Méthode	Variables
1. Constitution de la population statistique « les propriétaires immobiliers personnes morales »	Liaison des locaux d'habitation et des comptes propriétaires impliquant des personnes morales	LOGT GTOPER DNATPR CCOGRM
2. Sélection des variables susceptibles de spécifier des SCI propriétaires de logements	Huit variables sélectionnées qui précisent la sélection établie lors de l'étape 1	TYPPROPRO CODGRM DDNOM CODNOM DFORMJUR DFORM DSIREN DSGLPM
3. Reclassement des modalités des variables en fonction de leur niveau de certitude	Les modalités sont codées selon quatre niveaux de certitude : -2 : les modalités qui ne spécifient pas des SCI 2 : celles qui spécifient des SCI 1 : celles qui sont susceptibles de spécifier des SCI mais pour lesquelles il y a une incertitude (exemple : « autre personne morale ») 0 : celles dont on ne peut pas qualifier le niveau de certitude d'information (exemple : ce qui est non renseigné)	Même liste de huit variables qu'à l'étape 2
4. Éviction des combinaisons de modalités incohérentes	Par exemple, co-présence d'une modalité codée 2 et d'une modalité codée -2 pour qualifier un même propriétaire	
5. Éviction des personnes morales associées à des droits de gestion		TYPDROIT
6. <i>Scoring</i> et délimitation des sous-populations de propriétaires de type SCI	Trois sous-populations statistiques sont obtenues en résultat : • SCI avec certitude : désignées par les modalités « 2 » exclusivement • SCI avec incertitude : désignées par les modalités « 1 » exclusivement • Ne sont pas des SCI avec certitude : désignées par au moins une modalité « -2 »	

Tableau 3. Résultats du protocole de réduction d'incertitude.

Fichiers fonciers 2015	Nombre de comptes propriétaires (Vaucluse)	Nombre de comptes propriétaires (Bouches-du-Rhône)
Population statistique de départ : les propriétaires de logements		
Total des propriétaires (personnes morales et privées confondues). On entend par « propriétaire », un compte-propriétaire.	310 304	691 394
Personnes morales	41 328	133 475
Dont personnes morales privées	40 170	45 432
Dont personnes morales privées susceptibles d'être des SCI	11 450	42 971
Résultats du protocole de réduction d'incertitude		
A. Propriétaires de type SCI avec certitude	9 722	33 149
B. Propriétaires de type SCI avec incertitude	1 656	5 348
C. Hors champ	72	4 474

Source : fichiers fonciers, 2015, DGFIP. Réalisation : les auteurs.

communes de la côte bleue en 2015 met en évidence d'importants écarts :

	Fichiers fonciers (2015)	Kbis (2015)
Nombre de propriétaires de logements de type SCI	611	470

77 % des propriétaires de logements constitués en SCI et identifiés dans les fichiers cadastraux sont réellement enregistrés aux greffes du tribunal de commerce du secteur. Le non-recouvrement des informations (23 %) s'explique majoritairement par le fait que certaines SCI ne sont pas actives/enregistrées aux greffes. Et pour les autres, par le fait qu'elles sont enregistrées dans d'autres greffes.

À côté de l'erreur de caractérisation du type de propriétaire existent d'autres sources de variation possibles concernant l'identification de cette catégorie de propriétaire. Par exemple, des requêtes similaires réalisées sur des fichiers d'entreprises, telle que la BD Siren, renverraient des résultats différents. En effet, dans ces fichiers, les statistiques sont établies en fonction du siège social de la société et non en fonction de l'adresse de l'établissement principal (qui peut d'ailleurs être ou non un logement). C'est ainsi qu'en 2009, la base de données indique que 30 % des sociétés immatriculées en France sont des SCI.

De manière paradoxale, il existe donc différentes manières « d'objectiver » certaines

catégories de propriétaires immobiliers à l'image de ceux constitués SCI.

Quelle est la nature de l'erreur ?

Il est nécessaire de différencier trois sources d'erreur : l'inconnu, l'erroné et l'incertain.

La première source d'erreur – l'inconnu – vient du fait que les variables des fichiers fonciers présentent des taux de renseignement fluctuants et inconnus localement (on ne dispose que de statistiques nationales). Dans le cas des treize variables considérées

ici, les taux de renseignement s'échelonnent de 10 % à 20 %.

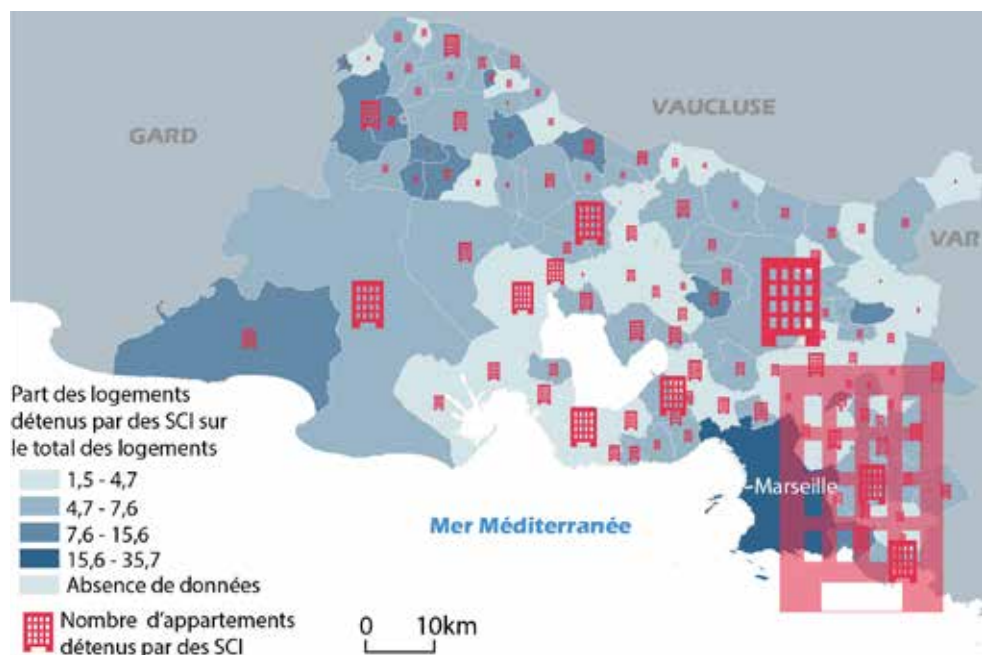
La deuxième source d'erreur – l'erroné – s'explique par le fait que certaines combinaisons de modalités sont incohérentes. Il s'agit par exemple de fondations ou encore de syndicats mixtes propriétaires mais aucunement constitués en SCI. La méthode de sélection basée sur un croisement de variables permet de détecter ces erreurs de caractérisation.

La troisième source d'erreur – l'incertain – est liée à la construction de la qualification (Desrosières, 1995). L'ambiguïté de la définition de certaines modalités (par exemple « autre personne morale privée », « société », etc.), génère une incertitude sur la nature de l'information : des propriétaires « mal » nommés, « mal » caractérisés dans les fichiers fonciers renvoient à des SCI. Parmi les cas observés, on retrouve fréquemment ces quatre configurations :

- La référence au terme « SCI » n'apparaît pas dans les variables de désignation.
- Le propriétaire est désigné comme « SCI » mais ne dispose pas de n° Siren ou n'est pas qualifié par la forme juridique adéquate.
- Le propriétaire n'est pas désigné comme SCI mais présente des caractéristiques proches : il est dénommé « copropriétaire » par exemple et présente un numéro Siren.
- Enfin, plusieurs cas de propriétaires dénommés « Les associés de SCI », « SARL », « SA », « société en nom collectif », « société civile foncière », « foncière » correspondent bien, malgré ce →

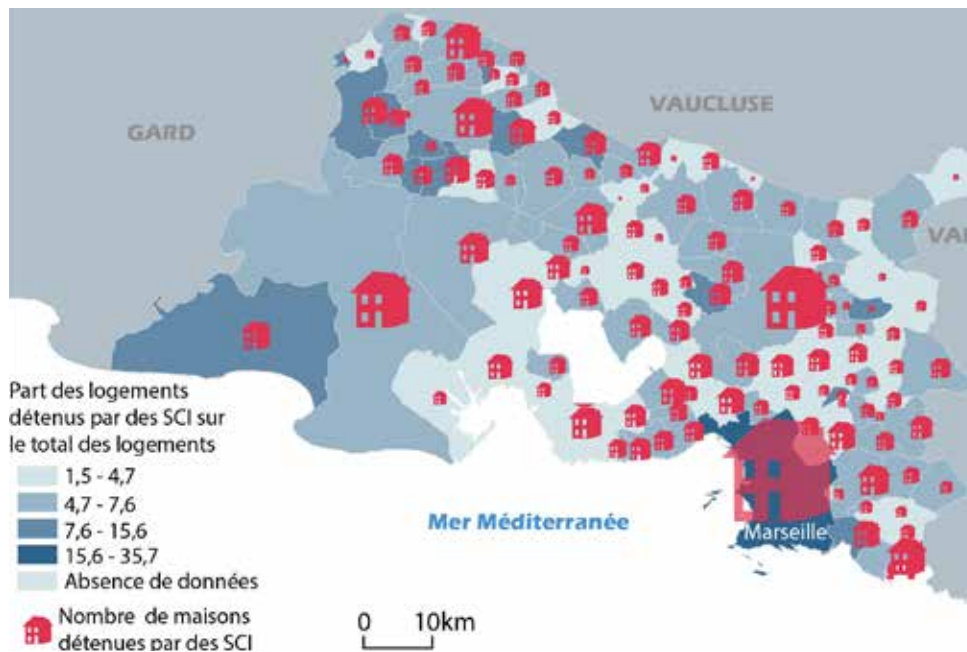
Carte 1. Appartements détenus par des SCI dans les Bouches-du-Rhône en 2015.

Sources : fichiers fonciers 2015, Cerema DGFIP Espace. Réalisation : auteurs.



Carte 2. Maisons détenues par des SCI dans les Bouches-du-Rhône en 2015.

Sources : fichiers fonciers 2015, Cerema DGFIP Espace. Réalisation : auteurs.



→ que laisse penser leur dénomination, à des propriétaires immobiliers de type SCI. Il s'agit le plus souvent d'une évolution de l'activité d'une société sans que celle-ci n'ait modifié ses statuts.

Chacune de ces configurations nécessite un traitement particulier. Généralement, elles relèvent en effet, moins d'erreurs que de libertés prises avec les délimitations juridiques (Bühl, 2013) définies dans la loi : elles sont l'expression des usages

sociaux de la catégorie juridique définie « société civile immobilière ».

L'observation de catégories de propriétaires immobiliers au sein des fichiers fonciers, par exemple les SCI, montre qu'il existe plusieurs manières d'interroger ces bases, et tout autant d'estimations possibles du phénomène. Comparer des statistiques issues des mêmes fichiers des impôts ne signifie pas, pour autant, qu'elles sont comparables. Il apparaît donc crucial de connaître le mode de construction de la donnée qui conditionne en amont la définition des catégories d'analyse et détermine en aval les observations sur les parcs et marchés fonciers-immobiliers (Desrosières, 2014). Cela est d'autant plus important que c'est sur ces observations que vont se fonder, et dans certains cas même se légitimer, les politiques publiques définies en matière de logement.

La solution méthodologique proposée ici pour réduire cette incertitude, résulte de la com-

binaison d'un protocole formalisé d'analyse des données et d'un savoir expert. Les observations sont donc consolidées et fiabilisées afin de former un ensemble cohérent, et de tendre vers l'objectivation des catégories de propriétaires. Les fichiers fonciers imposent ainsi de réformer nos modes d'observation des marchés et leur analyse selon des pratiques abductives. D'ailleurs certains outils offrent désormais de réelles possibilités pour la détection de *pattern* comme le « *machine learning* » dont il faudra suivre les résultats avec attention. ■

Repères bibliographiques

- Bonneval L. et Pollard J., « Promoteurs immobiliers, bailleurs sociaux, collectivités locales : des acteurs aux frontières des marchés du logement », *Métropoles* [En ligne], 20 | 2017, URL : <http://journals.openedition.org/metropoles/5423>
- Bühl M., 2013, *Sociétés civiles immobilières : création, gestion, évolution*, éd. Delmas, 391 p.
- Casanova E. L., Boulay G., Gérard Y., Yahi L., 2017, « Deux bases de données, aucune référence de prix. Comment observer les prix immobiliers en France avec Dvf et Perval ? », *Revue d'économie régionale & urbaine*, 2017/4, p. 711-732. DOI : 10.3917/revu.174.0711.
- Casanova Enault L., 2017, « Des acteurs de poids sur le littoral : les petits propriétaires immobiliers. Application au littoral provençal et azuréen », *Cybergeo : European Journal of Geography* [En ligne], DOI : 10.4000/cybergeo.28394.
- De Moncan P., 2002, *À qui appartient la France ? : histoire de la propriété urbaine de 1789 à nos jours et état actuel de la propriété immobilière en ville*, Paris, Éditions du Mécène.
- Desponds et Bergel, 2013, « Transactions immobilières et substitutions socio-résidentielles en Seine-Saint-Denis », *L'Espace géographique*, 2013/2 (tome 42), p. 115-127.
- Desrosières A., 1995, « Classifier et mesurer : les deux faces de l'argument statistique », *Réseaux*, 1995/3 (n° 71), p. 11 à 29.
- Desrosières A., 2014, *Prouver et gouverner. Une analyse politique des statistiques publiques*, Paris, La Découverte, coll. « Sciences humaines », 256 p.
- Démarquet M. et Toujas D., 2018, « Comment mesurer les gisements fonciers ? », *La revue Foncière*, n° 23.
- Geniaux G., Podjleski C. et Leroux B., 2009, « Les données Majic et leur valorisation au service de l'observation foncière », *Études foncières*, n° 139, p. 28-32.
- Le Goix R, Giraud T, Cura R, Le Corre T, Migozzi J, 2019, « Who sells to whom in the suburbs ? Home price inflation and the dynamics of sellers and buyers in the metropolitan region of Paris, 1996–2012 ». *PLoS ONE* 14(3): e0213169. <https://doi.org/10.1371/journal.pone.0213169>
- Ruz S., 2005, « Pour une sociologie des sociétés civiles immobilières. Logiques patrimoniales, dynamiques familiales et professionnelles, identités socio-juridiques. Le cas de la région lyonnaise (1978-1998) ». Thèse. Université Lumière Lyon 2.

Exemple de Kbis de SCI dite « familiale » enregistré aux greffes du tribunal de commerce.