



HAL
open science

Des chiffres et des droits. Le data mining ou la statistique au service du contrôle des allocataires

Vincent Dubois, Morgane Paris, Pierre-Edouard Weill

► To cite this version:

Vincent Dubois, Morgane Paris, Pierre-Edouard Weill. Des chiffres et des droits. Le data mining ou la statistique au service du contrôle des allocataires. *Revue des politiques sociales et familiales*, 2018, 126 (1), pp.49-60. 10.3406/caf.2018.3263 . halshs-02420230

HAL Id: halshs-02420230

<https://shs.hal.science/halshs-02420230>

Submitted on 14 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Des chiffres et des droits

Le *data mining* ou la statistique au service du contrôle des allocataires

Vincent Dubois	Professeur de sociologie et de science politique à l'Université de Strasbourg, Sage (UMR 7363).
Morgane Paris	Doctorante au laboratoire Sage, Université de Strasbourg.
Pierre-Édouard Weill	Maître de conférences en sociologie à l'Université de Bretagne occidentale, Lab-Ex et Sage.

Depuis 2011, la branche Famille de la Sécurité sociale a recours à l'outil statistique du *data mining* pour la sélection des dossiers d'allocataires à contrôler. Cet outil participe au renforcement de la politique de contrôle dont il constitue, en retour, un élément central. Le mode de raisonnement probabiliste et prédictif qu'il met en œuvre est en affinité avec la pensée managériale en termes de « maîtrise des risques », à partir de laquelle le contrôle des allocataires a été redéfini. Cet article restitue l'histoire de l'importation et de l'usage de cette technique, et en révèle les enjeux au plan organisationnel (centralisation de la politique de contrôle et réduction des initiatives des agents de terrain) et au plan socioéconomique, avec la focalisation de fait du contrôle sur les allocataires précaires à laquelle il conduit.

Jusqu'au milieu des années 1990, en France, le contrôle des bénéficiaires de prestations sociales et la lutte contre la fraude demeuraient relativement peu présents dans le débat public, et ne formaient pas une orientation stratégique des politiques sociales. Ces questions ont en revanche été promues en enjeux majeurs au cours de la période qui a suivi, en France comme dans la plupart des autres pays européens. Elles sont devenues omniprésentes tant dans les médias que dans le discours politique, des présidents de conseils départementaux affirmant la nécessité d'un contrôle renforcé du revenu de solidarité active (RSA) jusqu'au sommet de l'État — ce que le discours du Président Nicolas Sarkozy à Bordeaux le 15 novembre 2011 a illustré de manière particulièrement marquante⁽¹⁾. Sous l'effet de ces changements externes mais aussi selon des logiques qui leur sont propres, les organismes de protection sociale se sont dotés d'une politique de contrôle de plus en plus systématique, avec une organisation renforcée, des moyens juridiques et humains plus importants et des outils techniques nouveaux.

C'est à l'outil devenu central dans la mise en œuvre de cette politique qu'est consacré cet article : le *data mining*. Le terme générique « *data mining* », exploration ou fouille de données, désigne un ensemble de méthodes statistiques qui consistent à traiter un grand nombre de données pour constituer des modèles permettant, en retour, de systématiser les observations sur des données de même nature. Contrairement à d'autres méthodes qui reposent sur un raisonnement de type hypothético-déductif (partir d'hypothèses qui sont ensuite testées empiriquement et validées ou non), le *data mining* procède par induction, en partant de l'observation exploratoire des données pour dégager des régularités et des corrélations entre variables puis, dans un second temps, élaborer sur cette base des modèles prédictifs, qui sont, dans un troisième temps, opérationnalisés pour traiter des données à grande échelle ; ici, les très nombreuses informations consignées dans les dossiers de près de treize millions d'allocataires des caisses d'allocations familiales (Caf).

L'usage de cette technique à des fins de contrôle par un organisme de protection sociale illustre quatre évolutions contemporaines dans la conduite des politiques publiques. Il constitue, en premier lieu, un exemple d'importation dans le secteur public d'une technologie au départ mobilisée dans des organisations privées, où le *data mining* est utilisé pour modéliser les comportements de consommation et garantir un meilleur ciblage à des fins de marketing, pour prospecter des donateurs potentiels dans les campagnes de *fund raising* (collecte de fonds) et, surtout, pour maîtriser les risques associés aux comportements des clients, comme dans le cas du calcul actuariel des taux dans la banque et l'assurance, ou l'identification des mauvais payeurs potentiels dans des activités de services comme la téléphonie mobile (Lazarus, 2012). Le *data mining* est désormais de plus en plus utilisé dans le secteur public, aux États-Unis notamment (Federal Agency Data Mining Reporting Act of 2007). Il sert à la recherche de criminels, fraudeurs aux impôts, aux douanes, et autres contrevenants potentiels. En France, le *data mining* est en particulier promu en tant qu'instrument permettant d'améliorer le ciblage pour la lutte contre la fraude, par la Caisse nationale des allocations familiales (Cnaf) en 2010, les douanes et Pôle emploi à partir de 2013, l'administration fiscale en 2014 (rapport de la Délégation nationale à la lutte contre la fraude, 2014).

L'usage du *data mining* révèle ainsi une seconde tendance des politiques publiques, consistant à mettre en œuvre une gestion prédictive des populations. L'essor de la « justice prédictive » aux États-Unis, qui

consiste à anticiper les délits avant qu'ils ne surviennent en estimant la probabilité de récidive sur la base des actes passés et sert d'outil « d'aide à la décision » pour fixer la période d'incarcération ou le montant de la caution en fonction des risques de récidive, l'illustre bien (Harcourt, 2014). En matière de contrôle des bénéficiaires de prestations sociales, le *data mining* vise à prévoir les risques d'irrégularité, et c'est cette prévision qui préside au déclenchement d'une enquête. Le *data mining* s'inscrit à cet égard dans la tendance beaucoup plus vaste de recours croissant au *big data* dans l'action publique, dans la mesure où il repose sur la collecte et l'analyse de grands volumes de données disponibles servant des organisations, publiques ou privées, dans leur recherche d'informations sur des individus (Harcourt, 2014 ; Ollion, 2015). La branche Famille de la Sécurité sociale dispose en l'occurrence d'un matériau très riche : la connaissance de près de treize millions de foyers allocataires regroupant près de trente millions de personnes (avec les conjoints et les enfants) avec, pour chacun d'entre eux, un ensemble considérable d'informations qui concernent non seulement les prestations reçues mais aussi les situations familiales, professionnelles, les ressources, l'âge et la scolarisation des enfants, le logement, etc. Enfin, et l'article est particulièrement focalisé sur ce point, le *data mining* est mis au service d'une politique de ciblage qui doit permettre de rendre plus systématique et efficace l'ajustement de l'action publique à la diversité des situations à traiter, ici par l'identification des risques d'irrégularités ou d'erreurs dans les dossiers des allocataires.

Importation de technologies managériales privées dans le secteur public, gestion prédictive des populations reposant sur la manipulation de grandes masses de données et mise en œuvre au travers du ciblage, cet article revient sur ces orientations à partir des usages du *data mining* dans la politique de contrôle de la branche Famille. L'objectif est de montrer comment le modèle managérial de la maîtrise des risques facilite la rencontre entre une orientation politique en faveur du renforcement du contrôle des allocataires et un outil statistique qui en renforce l'efficacité, mais conduit aussi *de facto* au ciblage des populations les plus précaires.

Méthodologie

La recherche s'est déployée à trois niveaux :

- l'espace national des relations politiques et des pratiques institutionnelles au sein duquel les politiques de contrôle sont directement élaborées ou indirectement influencées (Caisse nationale des Allocations familiales (Cnaf), administrations centrales, Délégation nationale à la lutte contre la fraude, Cour des comptes, etc.) a été étudié sur la base d'entretiens au sein des organismes et administrations concernées ($n = 13$), et de la collecte de documents internes ;
- l'espace localisé de mise en œuvre des politiques de contrôle a été étudié sur la base d'études locales dans deux caisses d'allocations familiales (Caf) : l'une en milieu urbain fortement défavorisé, l'autre dans un département dans une situation socio-économique plus favorable, avec une ville-centre de plus de cent mille habitants, plusieurs villes moyennes et des zones rurales. Dans les deux cas, des entretiens ont été réalisés avec le personnel concerné ($n = 22$: contrôleurs, agents comptables, personnels du service contentieux, cadres dirigeants) ; des entretiens complémentaires avec d'autres organismes (conseils départementaux) ; des observations de tournées de contrôle à domicile, de réunions de commissions fraudes et autres concernant l'objet de la recherche ; la collecte de documents et de statistiques internes ;
- la population contrôlée a été appréhendée lors de l'observation des contrôles sur place, et étudiée sur la base des données statistiques fournies par les deux Caf et, au plan national, par la direction des statistiques, des études et de la recherche de la Cnaf. Conjuguant les méthodes qualitatives et quantitatives, l'enquête procède à la fois d'une attention aux analyses statistiques produites par la Cnaf et d'un retraitement de données secondaires produites par cette même institution. La perspective multiniveau adoptée permet de reconstituer les logiques qui conduisent au ciblage des populations, des sommets de l'État jusque dans les pratiques des *street-level bureaucrats* (Lipsky, 1980), en l'espèce, les contrôleurs en relations directes avec les « assistés ».

C'est dans les conditions de son importation par cet organisme qu'il faut chercher les raisons de la mobilisation du *data mining* comme outil de détection des irrégularités dans les dossiers, promu depuis 2010 comme l'élément central d'une politique systématisée et rationalisée de « maîtrise des risques », qui passe par un renforcement du contrôle des allocataires. Ce recentrement du dispositif de contrôle autour du *data mining* est allé de pair avec des transformations des relations entre les acteurs institutionnels impliqués aux niveaux national et local, et a contribué à modifier les pratiques de contrôle en même temps qu'à en renforcer l'efficacité. Ces évolutions produisent enfin des effets sur la vie quotidienne des administrés qui perçoivent des prestations sociales et familiales, en particulier pour ceux d'entre eux qui en sont le plus fortement dépendants, et font l'objet d'un contrôle accentué.

Importation et usages d'un instrument statistique au service du contrôle

La réussite d'une technique statistique implique « *qu'elle corresponde à de nouvelles façons de penser et d'organiser les rapports sociaux* » et « *qu'un réseau d'innovateurs trouve suffisamment d'alliés pour la diffuser* » (Desrosières, 2014, p. 81). En l'occurrence, si la branche Famille a été précurseur en la matière et traite à grande échelle les dossiers de ses allocataires avec le *data mining* depuis 2011, l'injonction croissante à garantir l'exactitude des paiements et à réduire les dépenses injustifiées a donné lieu à la construction indissociablement intellectuelle, institutionnelle et technique d'une nouvelle stratégie globale dite de « *maîtrise des risques* ». Les conditions de l'importation de cet outil lui ont conféré, au moins initialement, un sens et des usages tournés préférentiellement vers le contrôle des allocataires, et en particulier vers la détection des indus (soit les montants versés à tort, en raison de retards, d'erreurs

ou d'omissions, volontaires ou non, dans les déclarations des allocataires ou le traitement des informations par l'organisme) et des fraudes (ceux qui, parmi les indus, sont considérés comme résultant de pratiques intentionnelles des allocataires visant à percevoir des sommes auxquelles ils n'ont pas droit).

Un outil de « maîtrise des risques »

L'usage du *data mining* s'inscrit dans le contexte plus large de promotion de la « fraude sociale » et, plus généralement, du contrôle des prestations sociales dans le débat public et les politiques gouvernementales, et constitue en ce sens la réponse technique à un problème politique. Cette promotion tient d'abord à des incitations plus ou moins directement contraignantes, formulées par les gouvernements successifs et différentes administrations, sur la Cnaf. Une intervention plus directe du gouvernement dans l'organisation et les comptes des organismes de Sécurité sociale est rendue possible à partir de la réforme de 1996, qui donne notamment lieu à la signature de conventions d'objectifs et de gestion (COG), dans lesquelles est consigné un ensemble d'engagements pluriannuels des organismes. Ces engagements concernent notamment la lutte contre la fraude, et les outils qui permettent de la combattre. Ainsi, la COG pour 2001-2004 prévoit l'établissement d'un « référentiel général des risques » parmi lesquels figurent les déclarations des allocataires, et préconise « l'analyse des risques et de leurs modalités de sécurisation » afin de procéder à « l'établissement des normes de contrôle ». De la même manière, « la prévention et l'amélioration du recouvrement des indus » doit faire l'objet d'un plan d'action qui repose sur des analyses permettant d'« identifier leurs causes », ainsi que leurs perspectives de recouvrement. Cette démarche plus analytique intègre l'importance nouvelle accordée au ciblage, en réduisant la part des contrôles aléatoires ou déclenchés individuellement, et en mobilisant l'analyse statistique pour mieux définir les cibles et en augmenter l'efficacité. Les conventions suivantes témoignent d'une incitation constante à déployer de nouveaux outils techniques permettant d'identifier les sommes indûment versées, puis d'un suivi de leurs résultats. Ces incitations externes à renforcer les outils du contrôle proviennent également, de manière plus indirecte, de la Délégation nationale à la lutte contre la fraude (DNLF), structure interministérielle créée en 2008, qui contribue entre autres à promouvoir l'usage d'une technique telle que le *data mining* ainsi qu'à mettre en visibilité ses résultats. La Cour des comptes, occasionnellement par le biais de ses rapports annuels ou thématiques, et de façon continue *via* sa mission de certification des comptes depuis 2006, joue quant à elle un rôle très direct à cet égard.

L'efficacité du contrôle est constituée comme un critère de validation des comptes, et les auditeurs de la Cour formulent des recommandations en matière notamment de « maîtrise des risques », avec une incitation forte à l'usage du *data mining*, ce qui a été déterminant dans les efforts consentis dans ce domaine.

C'est largement en réponse à ce qui est souvent décrit comme des « pressions externes » critiquant une supposée trop faible activité en matière de contrôle et de lutte contre la fraude que les activités de la branche Famille en la matière ont été réorganisées et renforcées. Les modalités et les outils de cette évolution, avec la maîtrise des risques et le *data mining*, constituent en l'occurrence moins l'application directe d'injonctions à mobiliser des instruments précis que la traduction organisationnelle, en fonction des relations et des ressources internes à l'institution, et selon une logique du « bon élève », d'une incitation multiforme à améliorer l'efficacité du contrôle.

À l'ancienne séparation entre « contrôle interne », soit le contrôle des opérations des agents des Caf, et « contrôle externe », soit celui des « données entrantes » et donc des allocataires, se substitue la notion de « maîtrise des risques », supprimant la distinction entre les deux types de contrôle, à travers notamment la création au sein de la Cnaf du département qui porte ce nom en 2005. Cette réorganisation et cette nouvelle manière de penser ont contribué à augmenter la place occupée par le contrôle des allocataires dans l'activité de l'institution. De plus en plus d'acteurs au sein des Caf y interviennent, dans la perspective de la « maîtrise des risques » externes d'un point de vue financier et comptable. Les agents comptables, autrefois étrangers à ce contrôle, en sont désormais des opérateurs majeurs. Dans le même temps, l'appellation générique, et qui, à ce titre, concerne l'ensemble des agents, de « maîtrise des risques » constitue un quasi-synonyme du contrôle des allocataires, dans la mesure où il est essentiellement tourné vers les risques externes tenant à la fiabilité des données entrantes, c'est-à-dire des informations fournies par les bénéficiaires. D'une compétence sectorielle liée aux prestations, et relativement subalterne, le contrôle des allocataires est ainsi devenu une préoccupation transversale qui participe plus largement aux nouveaux modes de gestion de l'organisme. Or, il existe une affinité directe entre la pensée organisationnelle en termes de maîtrise des risques et un outil probabiliste et prédictif tel que le *data mining*. Ce dernier ne permet pas d'identifier des anomalies, comme le font les contrôles de cohérence par rapprochement d'informations en provenance de différentes sources, par exemple. Il

évalue la probabilité (le niveau de risque) qu'une anomalie survienne. Cette statistique prédictive, en amont de la logique administrative de vérification de la conformité aux règles, mais toujours au service de celle-ci, sert une logique gestionnaire tournée avant tout vers un objectif de rigueur financière.

Cependant, cet instrument ne détermine pas en lui-même ses usages possibles. Il s'agit d'un outil technologique en apparence « neutre » au sens où celui-ci peut être utilisé pour des opérations diverses et à des fins qui le sont tout autant. Les « cibles » des modèles de *data mining* peuvent être des dossiers « à risque » financier, mais également des publics éligibles aux prestations dispensées par l'organisme. Il serait ainsi concevable pour la Cnaf d'utiliser l'instrument *data mining* à l'inverse de son utilisation actuelle, soit pour cibler des allocataires potentiels. Cet instrument permettrait dans cette mesure de lutter non pas (seulement) contre la fraude, mais contre le non-recours – ces deux usages n'étant pas mutuellement exclusifs. Mais, tout comme la politique de « maîtrise des risques » s'est très largement confondue avec le contrôle des allocataires, le *data mining* a été essentiellement conçu, au moins au départ, dans le but de détecter les indus et, parmi eux, les indus frauduleux.

L'appropriation institutionnelle d'un instrument statistique

La deuxième condition du succès d'une technique identifiée par Alain Desrosières, le circuit de sa diffusion, a en l'occurrence également orienté ses usages dans le sens de la détection des indus et des fraudes. Ces usages sont aussi ceux de Pôle emploi et de la branche Maladie de la Sécurité sociale, conférant à la technique *a priori* « neutre » du *data mining* le sens d'un outil de contrôle de l'État social. Une telle orientation s'est dessinée, pour une part, dans l'origine de cet usage. Avant d'être promu comme instrument national de la politique de contrôle, le *data mining* a d'abord été expérimenté localement, à l'initiative d'un agent comptable qui, considérant les limites des moyens disponibles pour lutter efficacement contre la fraude, et étant potentiellement en cause dans le cas de fraudes qui auraient pu être détectées, s'est investi dans la recherche de nouvelles méthodes. Il découvre alors celle du *data mining*, que la Caf de la Gironde expérimente en lien avec la politique nationale de « maîtrise des risques » à partir de 2004. La première étape consiste en la constitution d'une base nationale de données de référence de trois mille dossiers frauduleux, utilisée pour constituer les premiers modèles testés à la Caf de la Gironde. Le « profilage » est ensuite affiné en 2006-2007, grâce à l'extension de la base de données. À partir de 2009, la collecte de données sur la base d'un échantillonnage aléatoire de dossiers allocataires est opérée sur l'ensemble du territoire. Dix mille cinq cents dossiers allocataires sont sélectionnés annuellement par le système et font l'objet de vérifications systématiques par des agents de contrôle assermentés ; des contrôles sur pièces et sur place sont donc menés. Dans la continuité de ce qui avait été réalisé à l'échelon local, puis étendu à plusieurs Caf, cette opération spécifique dite « cible 021 » « *s'apparente à une enquête de veille sur les risques et leurs évolutions* » (Collinet, 2013, p. 130), qui permet l'élaboration puis l'adaptation des modèles dégagés à partir des corrélations entre variables les plus statistiquement prédictives d'un indu. Ces différents modèles, sur lesquels l'article revient plus précisément *infra*, permettent d'attribuer un « score de risque » aux dossiers des allocataires et de sélectionner sur cette base les cas à contrôler. Il s'agit donc d'un instrument de « précontrôle », qui permet de détecter les dossiers dont le niveau de risque fonde le déclenchement d'un contrôle effectif. C'est un outil dont la mise en œuvre peu coûteuse améliore l'efficacité des opérations de contrôle. En 2010, le dispositif est expérimenté pour le déclenchement des contrôles dans plusieurs Caf volontaires. Il est généralisé à l'ensemble du réseau en 2011. En complément, un échantillon plus restreint de sept mille

allocataires est destiné à calculer le taux de risques financiers non couverts par le système de contrôle : l'« indicateur de risque résiduel » (IRR).

L'usage fait par la branche Famille – mais également Pôle emploi et d'autres branches de la Sécurité sociale – du *data mining* renvoie ainsi à l'évolution des fonctions assignées à (et revendiquées par) les organismes de Sécurité sociale. Ces derniers ont désormais fortement intégré des logiques gestionnaires dans un contexte de raréfaction de leurs ressources. Données et instruments statistiques ne sont désormais plus seulement utilisés à des fins de socialisation des risques auxquels sont exposés les individus (maladie, vieillesse, chômage, accidents, événements familiaux), dans la logique de la « *société assurantielle* » analysée par François Ewald (1986). Il ne s'agit pas non plus d'identifier les « *populations à risques* » qui, par leur situation et leur comportement potentiellement dangereux pour elles-mêmes ou pour la société, suscitent la prévention d'institutions spécialisées, comme dans la « *gestion des risques* » analysée par Robert Castel (1981). Les risques dont traite le *data mining* comme instrument de « *maîtrise des risques* » sont ceux que des informations manquantes, incomplètes ou erronées concernant les allocataires et, pour une bonne part, fournies par eux, font courir à la bonne marche légale et financière de l'organisme payeur. Dans les deux cas, la statistique est bien un « *outil de gouvernement* » (Desrosières, 2014) ; il sert en l'occurrence avant tout une maîtrise des risques qui rime avec l'effort de maîtrise des dépenses.

Un dispositif de contrôle recentré autour du *data mining*

Légitimé par des résultats financiers probants, le *data mining* apparaît comme la nouvelle pierre angulaire du contrôle, à l'articulation d'une augmentation des échanges de données informatisées et d'un ciblage restreint des contrôles à domicile sur les allocataires « à risques », qui sont aussi souvent les plus précaires. Le recours au *data mining* favorise la remontée au niveau national de données localisées à des fins de pilotage à distance, s'inscrivant en cela dans le cadre plus général d'un « tournant électronique du travail social » en Europe occidentale depuis le début des années 2000 (Parton, 2008). Son rôle désormais central dans la politique de contrôle de la branche Famille marque un renforcement de l'encadrement par l'échelon central, en limitant non seulement la marge de manœuvre des directions des Caf, mais aussi les initiatives de leurs agents de terrain.

La nouvelle pierre angulaire du contrôle

Si le *data mining* apparaît désormais comme une pièce essentielle de la politique de contrôle et de lutte contre la fraude définie au niveau national, c'est que son expérimentation y est présentée comme un facteur explicatif majeur de la progression constante des indus – et dans une moindre mesure des rappels – détectés depuis la fin des années 2000. Les rapports d'activité de la Cnaf valorisent le rendement croissant des contrôles à domicile et la généralisation du *data mining* dans le ciblage des allocataires qui a permis ce progrès. La proportion des contrôles sur place qui conduisent à des régularisations financières a en effet considérablement augmenté, passant de 17 % d'indus et 14 % de rappels en 2009 à respectivement 44 % et 30 % en 2015. Cette évolution est également significative en termes de montants détectés, lesquels passent de 122 millions à 225 millions d'euros d'indus et de 40 millions à 59 millions d'euros de rappels au cours de cette période. Cette corrélation positive résulterait de la capacité des modèles statistiques établis à identifier des facteurs de risque d'irrégularités à l'impact financier important, quand bien même les modèles prédictifs du montant des indus ou des rappels restent à construire. Dès lors, le ciblage des situations présentant les risques financiers les plus importants par le biais de leur modélisation ne saurait que s'accroître.

Si le *data mining* apparaît comme la nouvelle pierre angulaire de la politique de contrôle, c'est aussi que son essor programmé et justifié, avec chiffres à l'appui, s'articule à deux autres évolutions marquantes de cette politique. En premier lieu, l'efficacité croissante des contrôles ne saurait être dissociée de l'augmentation et l'amélioration constante de la qualité des données entrantes. Celle-ci repose sur un ensemble d'innovations en matière de stockage, de consultation et d'échanges de données personnelles sur les allocataires entre organismes de Sécurité sociale, administrations fiscales et partenaires privés, notamment dans le secteur bancaire. Les échanges de données informatisées et, plus largement, les vérifications sur fichier, ont été favorisés par la mise en place d'une série d'identifiants, de fichiers et de croisements entre fichiers, qui permettent un contrôle à distance, à grande échelle et relativement peu coûteux en personnel. Autorisé légalement dès 1998, mais déployé seulement depuis 2008, un fichier national, référentiel national des bénéficiaires (RNB) identifie les allocataires à l'aide d'un numéro unique (Nir, numéro d'inscription au registre), regroupe l'ensemble des fichiers des Caf et permet d'identifier les affiliations multiples et de croiser à grande échelle les données d'autres fichiers. Le répertoire national

commun de la protection sociale (RNCPS) favorise l'accès aux informations personnelles détenues par les autres organismes de protection sociale. Institué en 2009, il est accessible par le biais d'un portail informatisé commun, l'espace des organismes partenaires de la protection sociale (EOPPS), le partage de ce type d'interface informatique contribuant à renforcer des liens institutionnels préexistants (Baudot, 2011).

Dans le même temps, le nombre de contrôles à domicile (voir *infra*) a diminué. Cependant, cette évolution marque moins un déclin de leur importance dans les dispositifs de contrôle que leur focalisation sur les dossiers les plus « risqués », et donc souvent les plus complexes, identifiés par les algorithmes du *data mining* qui constitue désormais la principale technique de déclenchement de ces contrôles. L'identification d'un « risque » potentiel dont la nature n'est pas précisée renforce, en effet, la nécessité des contrôles à domicile, mais ces derniers exigent un temps de travail plus important des agents assermentés, qu'il s'agisse des enquêtes menées en amont ou de la rédaction des rapports. La baisse significative et continue du nombre de contrôles sur place, qui passe de deux cents quatre-vingt mille en 2009 à cent soixante-six mille en 2015, suit donc logiquement l'augmentation de la part de leur déclenchement par le *data mining*. Ainsi, plusieurs tendances se conjuguent : la montée en charge du *data mining* dans le déclenchement des contrôles à domicile, la spécialisation sur les cas difficiles, l'allongement de la durée des enquêtes plus ciblées et approfondies, l'amélioration de leurs résultats en termes de détection d'irrégularités.

Un encadrement renforcé de l'échelon central sur le ciblage

Orienté vers l'amélioration de l'efficacité des contrôles et l'augmentation de la détection des indus et fraudes, le *data mining* a contribué à limiter l'autonomie des Caf et à estomper les différences qui marquaient auparavant leurs pratiques. La définition de requêtes ciblées sur certains allocataires, à partir d'une connaissance fondée sur l'administration locale des prestations, était jusqu'alors l'une des principales possibilités pour les organismes d'orienter leur propre politique de contrôle. Or, le *data mining* est devenu progressivement le principal outil de sélection des dossiers à contrôler, et ce d'abord dans des départements d'expérimentation précoce et massive, tels que la Seine-Saint-Denis, puis sur l'ensemble du territoire. Avant la généralisation de l'outil en 2011, moins d'un quart des contrôles sur place résultaient de cibles nationales. En 2016 (**tableau 1**), près des deux tiers des contrôles sur place ont été lancés à

partir d'un ciblage national issu du *data mining*. L'augmentation de la proportion des contrôles issus de l'outil, qui semble néanmoins atteindre un plafond depuis 2015, a pour principal corollaire la baisse drastique des contrôles initiés par le biais des gestionnaires conseils allocataires (seulement 20 % en 2016 contre 51 % en 2011) et, dans une moindre mesure, de ceux déclenchés à partir de cibles définies localement – le plus souvent liées à des prestations et en particulier au RSA – ou des contrôles activités-ressources (Rac).

Cette évolution suit les plans de contrôle nationaux, qui préconisent de donner systématiquement la priorité au *data mining* sur les autres modalités de déclenchement des contrôles, notamment les demandes de contrôle des gestionnaires conseils allocataires dont le nombre est désormais encadré. L'introduction du *data mining* réduit non seulement la part des cibles définies par les directions des Caf, mais aussi la place accordée aux initiatives de leurs agents de terrain. Ces derniers n'en restent pas moins mobilisés dans la lutte contre la fraude, mais leur vigilance est davantage sollicitée pour l'identification résiduelle des cas non détectés par l'outil statistique. La technique ne remplace donc pas le savoir-faire des agents, mais encadre sa mobilisation et l'oriente vers les cas les plus complexes. La montée en puissance d'un système automatisé d'identification des risques affecte le sens du contrôle sur place. Le *data mining* et les scores de risque constituent pour les contrôleurs une « boîte noire » entourée de mystère, dont certains effets sont indésirables, tels que les contrôles répétés effectués d'une année sur l'autre sur une même personne, le profil de celle-ci étant considérée « à risques » selon l'outil, et souvent sans raison. La finalité du contrôle sur place a fait l'objet d'un recadrage, la fonction de conseil autrefois incarnée par les agents s'effaçant au profit d'une seule fonction de contrôle. Par le biais des formations auxquelles ils assistent, mais aussi, moins directement, en raison de la généralisation des « objectifs fraude » – eux-mêmes déterminés par recours au *data mining* – qu'on leur assigne, les contrôleurs sont donc incités à axer leur rencontre avec l'allocataire sur le contrôle de la conformité de ses droits au regard de sa situation, et plus encore à la détection d'une anomalie volontaire, donc d'une fraude. Dans la mesure où les dossiers à contrôler sont les plus fortement scorés, et que les cibles du *data mining* sont elles-mêmes conçues pour détecter les cas de fraude, les contrôleurs ont tendance à considérer la sélection des dossiers à contrôler comme une présomption de fraude, et à guider leur enquête en conséquence.

La Cnaf renforce d'autant mieux l'encadrement des pratiques locales en matière de ciblage qu'elle assure le pilotage du système d'information intégrée sur lequel repose le *scoring* des allocataires. Le ciblage des dossiers et des situations à contrôler est, en effet, actualisé mensuellement, à travers une prise en compte plus efficace des changements de situation des allocataires, grâce à l'essor des échanges de données informatisées avec les institutions partenaires évoquées *supra*. Les gestionnaires du réseau national exploitent également les modélisations statistiques pour fixer aux Caf des objectifs locaux de réalisation de contrôles ou « objectifs fraude », à des fins d'harmonisation de la couverture des risques sur le territoire. La marge de manœuvre des directions des Caf se limite alors, en principe, à l'ajustement mensuel du nombre de contrôles à réaliser en fonction d'objectifs annuels fixés par l'échelon national, la Cnaf. Les objectifs promus par cette dernière demeurent néanmoins difficiles à atteindre, en particulier dans les départements qui présentent les plus hauts niveaux de risque. Il s'agit plus largement d'aligner les pratiques des Caf locales sur celles souhaitées par la Cnaf. À ce titre, la COG 2013-2017 accorde une place centrale au *data mining*, mis en place les années précédentes. Sur le plan de l'organisation, l'accent est mis sur la mise en cohérence des pratiques au plan national, visant à limiter les disparités locales, comme en témoignent la constitution d'indicateurs de « suivi de la réduction des écarts entre Caf » en matière de rendement des contrôles, d'avertissements et de pénalités financières. La « politique des indicateurs » déjà engagée précédemment est encore renforcée, notamment par le « *Suivi du nombre de fraudes détectées qui distingue dans le total le nombre de récidives* » et, surtout, l'« indicateur de risque résiduel » qui, si l'on peut dire, « boucle » la logique de maîtrise des risques.

Le *data mining* s'est imposé comme instrument de la politique de contrôle de la branche Famille par le biais de la reconfiguration des rôles au sein de son organisme central, la Cnaf. Le champ de compétences des agents comptables s'est étendu, tandis que l'outil *data mining* a permis d'orienter le contrôle des données entrantes dans le sens de l'identification de la fraude. L'usage du *data mining* n'est donc pas sans incidence sur l'économie de la politique de contrôle.

Vers un contrôle renforcé des plus précaires

L'application du *data mining* repose, en théorie, sur l'identification de « facteurs de risques » dans les dossiers, et non de populations « à risques ». Cependant, en pratique, les modèles de prédiction des

risques élaborés se focalisent sur les caractéristiques de la situation des allocataires plutôt que sur les prestations perçues ou les modalités de gestion de leurs dossiers. De surcroît, la tendance à la convergence de certains facteurs de risques chez des ménages en situation de précarité et d'isolement conduit à les désigner de fait comme des cibles privilégiées. C'est ainsi que la généralisation progressive du *data mining* sur le terrain aboutit à une surexposition croissante au contrôle des populations les plus défavorisées. Ce surcontrôle a lui-même augmenté ces dernières années.

Des risques associés aux caractéristiques des allocataires

Le processus d'identification et de mesure des risques d'indus étudié répond à un protocole bien précis. Il importe maintenant d'offrir un aperçu des techniques statistiques et des matériaux empiriques mobilisés par le service statistique de la Cnaf. Il s'agit ainsi de mieux comprendre, sans dévoiler certaines données sous le sceau de la confidentialité, les différentes modélisations prédictives des risques sur lesquelles se fonde l'attribution de scores aux allocataires en vue de renforcer l'efficacité du ciblage des contrôles. On verra ainsi comment les premiers modèles élaborés et utilisés entre 2010 et 2014 privilégient la prise en compte de caractéristiques propres à la situation des ménages des allocataires.

Le premier défi auquel font face les statisticiens consiste à identifier des caractéristiques ou des comportements particulièrement « à risques » parmi la profusion d'informations disponibles sur les allocataires et leur ménage. Deux techniques de modélisation statistique peuvent être utilisées : les régressions logistiques et les arbres de décision (2). L'usage, privilégié dans les faits, de la seconde méthode apparaît comme relativement courant en matière de prédiction des risques (Peretti-Watel, 2005). Il s'inscrit, en outre, dans une certaine continuité de l'expérimentation initiale menée par la Caf de la Gironde, tout en mettant à disposition des statisticiens de la Cnaf un matériel empirique bien plus important (Collinet, 2013). Il convient de souligner que toutes les informations sur les allocataires sondés sont connues sur une même date de référence – avril 2009 – ce qui facilite le travail de constitution d'une base de données homogène, mais revient à prendre en compte des attributs et des comportements temporellement situés, dont on ne saurait oublier qu'ils connaissent des évolutions. Son exploitation doit néanmoins permettre de mesurer – jusqu'au renouvellement de cette base de données – les risques d'indus versés sur une période de plus de six mois parmi plus de mille cinq cents modalités de variables caractéristiques possibles, qui se révèlent donc d'une très grande diversité. Des variables elles-mêmes

démultipliables à l'infini grâce aux recodages opérés. Par exemple, la variable « contact » est fondée sur la soustraction du nombre d'appels et de visites des allocataires dans les locaux des Caf.

Différents modèles de régression logistique visent ensuite à identifier les risques d'indus. Il s'agit de déclinaisons du risque associé aux allocataires et à leurs ménages qui renvoient à leur situation sociale et familiale, plutôt qu'aux prestations qu'ils perçoivent, même si ces dernières entrent en compte *a posteriori*, dans le processus de ciblage des contrôles. Ces modèles de régression sont au nombre de cinq. Les quatre premiers modèles correspondent à autant de « familles de risques », liés aux ressources des ménages des allocataires ; à leur composition selon l'État civil ; à leur situation résidentielle et à la situation professionnelle des adultes. Le cinquième modèle, dit « global », apparaît comme un florilège des caractéristiques les plus significativement associées aux indus. Ces différents modèles présentent chacun leurs spécificités, à commencer par la proportion plus ou moins importante de l'échantillon disponible qu'ils permettent de caractériser. Ainsi, dans le modèle relatif à la situation professionnelle, l'absence d'activité du conjoint est tout particulièrement associée au risque d'indu. Concernant le modèle qui prend en compte les diverses situations familiales recensées, le risque apparaît presque doublé lorsque l'allocataire est isolé ou seul avec un enfant, par rapport aux situations où le conjoint est salarié. Le risque est aussi fortement aggravé dans le cas où la charge d'un enfant âgé de plus de 18 ans est déclarée. Pour ce qui est de la situation résidentielle, les rares allocataires dont le conjoint est étudiant ou retraité multiplient par 2,6 le risque de versement d'indus (en particulier de prestations liées au logement), par rapport aux allocataires ne déclarant pas de conjoint. Quant à la modélisation des risques relatifs aux ressources, ceux-ci apparaissent renforcés lorsque les revenus des allocataires sont faibles, en particulier lorsque le taux d'effort de leur ménage pour se loger est supérieur à 35 %. Enfin, le modèle dit « global » pointe de nouveau l'absence d'activité professionnelle, la charge d'enfant(s) de plus de 18 ans ou la faiblesse des revenus. À ces facteurs caractéristiques des ménages des allocataires s'ajoutent des modalités de variables de gestion du dossier (exemple, l'absence de certification du Nir) ou spécifiques au mode de contact des allocataires avec leur Caf, en particulier le mode de paiement des prestations autre que par virement bancaire, ou l'écart entre le nombre d'appels et de déplacements au guichet. Si le modèle global porte sur l'essentiel de l'échantillon, les modélisations par famille de risques s'appliquent à des populations plus restreintes. Ceci implique une dégradation plus rapide de la significativité des

corrélations révélées lorsque l'on ajoute des facteurs explicatifs des indus. Ces facteurs sont donc moins nombreux que pour le modèle global, ce qui tend à renforcer la spécificité des résultats observés. Il convient, par ailleurs, de préciser succinctement les usages de ces modèles élaborés sur un échantillon d'allocataires, afin de permettre l'attribution d'un score de risque d'indu à l'ensemble des allocataires de la population de référence. De fait, c'est la création d'une base d'apprentissage des modèles de *data mining* qui a permis de modéliser un risque d'indu déclenché par une modification d'une information relative au logement, à la situation familiale, aux ressources ou encore à la situation professionnelle. Le modèle dit « global » est également utilisé. Les résultats de trois vagues successives d'expérimentations locales ont amené le service statistique à privilégier une méthode particulière de combinaison des modèles de régression logistique précédents pour attribuer un score à l'ensemble des allocataires au niveau national. La variable de score est issue de cinq scores calculés par les modélisations par type de risque : elle est égale au maximum atteint par les scores de chaque sous-famille de risque ou par le score relatif au modèle dit « global ». Ainsi, un allocataire avec un score de 0,3 aurait ainsi 30 % de chances de présenter un indu, ce qui justifie son placement sur la liste des priorités des contrôleurs. Des modèles théoriques élaborés au niveau national aux pratiques de ciblage des allocataires sur le terrain, de l'attribution généralisée d'un score à la sélection localisée des allocataires contrôlés, il existe un ensemble d'étapes précédemment identifiées. Mais, en dépit des difficultés d'appropriation et des résistances localisées que suscite la généralisation du *data mining*, celle-ci accompagne concrètement un ciblage croissant des populations les plus défavorisées.

La surexposition au contrôle des plus défavorisés

Après l'identification d'erreurs internes ou de comportements frauduleux associés à certaines prestations fondée sur l'expérience des agents de terrain, l'étape suivante est la généralisation de modèles de prédiction des risques associés à des variables caractéristiques de la situation des allocataires. Ainsi, l'hypothèse d'un glissement possible d'une politique de « maîtrise des risques » vers une forme de gestion des « populations à risques » est donc formulée. Sans pour autant qu'il s'agisse d'une focalisation volontaire sur les populations les plus défavorisées, qui sont aussi les plus dépendantes des prestations sociales, il apparaît que les transformations des logiques de déclenchement des contrôles conduisent, *de facto*, au surcontrôle des allocataires précaires, c'est à dire à leur plus grande probabilité d'être contrôlés

que l'ensemble de la population des allocataires. En l'occurrence, un tel écart croît en proportion de l'intensité des difficultés économiques et sociales de la fréquence des changements de situation (Chantel, 2017), intensité et fréquence caractéristique des populations précaires.

Suivant la perspective diachronique qui structure la réflexion ici, il est possible de montrer que cette tendance au ciblage des populations les plus défavorisées est allée croissant au cours de la période étudiée, en particulier en matière de contrôle à domicile des allocataires. Les données collectées permettent, en effet, de comparer les caractéristiques de la population faisant l'objet de contrôles pour les années 2006, 2010, 2013 et 2014 (**tableau 2**). Si la tendance au ciblage croissant des allocataires les plus précaires peut être constatée dans le cadre d'échanges de données informatisées ou de contrôles sur pièces, elle se révèle la plus manifeste en matière de contrôle à domicile. Ce qui apparaît conforme aux objectifs des plans de contrôle nationaux, dès lors que des gains de productivité relatifs à la montée en charge du *data mining* sont les plus attendus pour ce type de contrôle (voir *supra*). Certes, les allocataires les plus fragiles étaient déjà l'objet d'une attention particulièrement marquée du personnel affecté au contrôle et à la lutte contre la fraude durant la première moitié des années 2000 (Dubois, 2003). Néanmoins, les données collectées permettent d'établir précisément en quoi cette tendance s'est renforcée, et en direction de quels allocataires en particulier. Ces derniers se distinguent, en effet, du point de vue de la composition de leur ménage, de leur âge et de celui de leurs enfants, de leur activité professionnelle, ainsi que de leurs revenus.

Concernant la composition familiale des ménages des allocataires, il apparaît tout d'abord clairement que les familles monoparentales – essentiellement des mères isolées – sont de plus en plus surcontrôlées : on constate 29,2 points de pourcentage d'écart entre la proportion de familles monoparentales contrôlées sur place durant l'année en 2014 et celle qu'elle représente parmi l'ensemble des allocataires, contre seulement 16,4 pour l'année 2006. L'âge des allocataires et de leurs enfants est, en outre, déterminant, mais c'est parmi les ménages incluant des enfants âgés de 18 ans à 25 ans – autrement dit en âge de travailler et donc de ne pas déclarer d'éventuelles ressources – que l'augmentation des contrôles sur place est la plus frappante (seulement 2,2 points de pourcentage d'écart avec l'ensemble des allocataires contre 24,4 en 2013 et 23,7 en 2014). Des évolutions significatives des écarts en matière de contrôle sont également observées selon l'activité professionnelle. Tandis que les allocataires au chômage ou sans

activité professionnelle sont de plus en plus souvent contrôlés, la tendance s'inverse pour ceux qui exercent un emploi. On constate enfin une évolution concomitante des écarts en matière de contrôle selon les revenus des allocataires. Un clivage majeur apparaît entre les allocataires aux revenus mensuels inférieurs à 500 euros par unité de consommation du ménage – ce qui les situe aux environs de la moitié du seuil de pauvreté à 60 %**(3)** – et les autres. Les plus pauvres faisaient déjà l'objet d'un surcontrôle au milieu des années 2000, mais cette tendance est allée *crescendo*, notamment en ce qui concerne les visites à domicile, l'écart avec l'ensemble des allocataires évoluant de 25,1 points de pourcentage en 2006 à 33,6 en 2014.

Certes, n'ont pas disparu les contrôles faisant suite à un doute sur le comportement au guichet ou des suspicions de la part d'un gestionnaire conseil allocataires, sans parler des signalements par des institutions partenaires ou des dénonciations. On peut d'ailleurs penser qu'ils sont également plus fréquents parmi les populations les plus défavorisées, davantage marquées que d'autres par l'instabilité, et donc sujets à des erreurs ou des retards dans les informations fournies. Néanmoins, c'est la généralisation du *data mining* qui tend le plus systématiquement et le plus fortement à orienter le ciblage des contrôles vers les allocataires aux difficultés sociales et économiques les plus accentuées.

L'application d'outils statistiques de ciblage en fonction des caractéristiques sociodémographiques conduit d'ailleurs logiquement à ce que les allocataires les plus régulièrement contrôlés soient ceux qui bénéficient de prestations sous conditions de ressources et, parmi eux, plus encore ceux qui perçoivent des aides au titre de la solidarité. Et cela non par décision *a priori* de les contrôler, comme dans le cadre de cibrages localisés sur les titulaires du RSA – l'avènement du *data mining* rend d'ailleurs d'autant plus obsolètes de telles opérations qu'elles sont difficiles à assumer publiquement –, mais parce que des modèles statistiques aux apparences de neutralité et aux usages désormais éprouvés sur le terrain identifient davantage de « facteurs de risques » chez les populations concernées.

Conclusion

Le recours au *data mining* et la politique de maîtrise des risques dont il constitue un élément central s'inscrivent dans un processus de rationalisation qui associe rationalité « en finalité » et rationalité « en valeur », pour reprendre la distinction wébérienne. Il s'agit, en effet, à la fois de mobiliser de nouveaux moyens ayant pour finalité de limiter les dépenses sociales, et de lutter contre une fraude sociale dont

l'immoralité, régulièrement dénoncée dans l'espace public, est constituée en repoussoir de la « valeur travail » qu'il s'agit de promouvoir. Ce processus de rationalisation est ainsi éminemment politique, de par les modalités de son déclenchement et de son accélération, avec la remise en cause croissante, au plan gouvernemental et dans le débat politique, des formes traditionnelles d'administration de l'État social accusées de « laxisme » et la dénonciation corrélative des « assistés » qui en profiteraient abusivement. Éminemment politique, ce processus a cependant toutes les apparences de la technique. Il combine la raison juridique, qui implique une stricte application des règles en vue d'assurer le « *paiement à bon droit* » des allocataires (Buchet, 2013), la raison gestionnaire, qui implique la maîtrise de la bonne affectation des ressources, et la raison statistique, au service des deux premières, qui intègre la pensée probabiliste dans l'administration du social (Desrosières, 1993).

Conduite au nom de l'exactitude et de l'efficacité, cette rationalisation, dont le *data mining* est à la fois le produit et le vecteur, n'est pas pour autant socialement neutre. Elle conduit, sans que cela n'ait été souhaité, mais pas sans raisons, à faire des populations précaires les cibles privilégiées d'un contrôle qui, s'il ne se confond pas entièrement avec la lutte contre la fraude, leur accorde cependant une place prépondérante. En témoignent l'importante différence entre le montant des rappels et celui des indus détectés, ainsi que la multiplication des sanctions à l'encontre des allocataires. La rationalisation à l'œuvre va ainsi de pair avec une orientation coercitive à l'égard des plus précaires. La montée en charge du *data mining* s'accompagne ainsi d'une judiciarisation croissante de la fraude. Une telle évolution s'inscrit dans un mouvement plus large de judiciarisation des politiques sociales (Bernheim et Commaille, 2012 ; Weill, 2017). L'analyse des effets sociaux de cette judiciarisation de la fraude aux prestations sociales apparaît d'autant plus nécessaire que leur public cible apparaît de plus en plus majoritairement composé de populations précaires, et que l'on sait depuis longtemps à quel point les inégalités face à la justice sont profondément marquées en leur défaveur (Galanter, 1974). À l'instar de la différenciation sociale du contrôle fiscal (Spire, 2012), on aboutit finalement à un renforcement des sanctions à l'encontre des ménages des fractions précarisées des classes populaires, pour la surveillance et la détection des comportements frauduleux desquels on mobilise des moyens plus importants que pour les autres catégories.

Ainsi, un outil statistique tel que le *data mining* ne détermine pas en lui-même ses usages, qui sont toujours définis dans le processus social et politique de sa mobilisation (Bruno *et al.*, 2014). Si les conditions de l'importation du *data mining* dans l'administration sociale française en ont fait une technologie au service d'une « maîtrise des risques » définis avant tout en référence à un impératif de vérification et de limitation des dépenses sociales, rien n'empêche de l'utiliser aussi pour faciliter l'accès aux droits sociaux et lutter contre le non-recours, en ciblant les populations à partir des facteurs prédictifs de la non-demande (Warin, 2016). Un tel usage existe en Grande-Bretagne ou au Pays-Bas (Zuurmond, 2008). Il fait l'objet de préconisations en France (Biémouret et Costes, 2016), où il impliquerait, cependant, de surmonter des conditions sociales, politiques et financières qui ne lui sont guère favorables.

Notes

(1) Vidéo en ligne : <https://www.youtube.com/watch?v=aXPhEPRTyBs>

(2) En 2015, le seuil de pauvreté se situe à 60 % du niveau de vie médian, soit 987 euros par mois. Une part croissante des ménages s'écarte cependant de ce seuil, leur revenu par unité de consommation étant inférieur à 30 % du niveau de vie médian (Fontaine et Sisic, 2015).

(3) La régression logistique est une technique statistique prédictive. Elle vise à construire un modèle permettant de prédire et d'expliquer les valeurs prises par une variable cible qualitative à partir d'un ensemble de variables explicatives. Le recours à des arbres de décisions permet de sélectionner ces variables et/ou de les discrétiser.

Tableau 1 – Répartition en % des contrôles sur place clos l'année N en fonction de la cible du contrôle (en %)

	2011	2012	2013	2014	2015	2016
<i>Data mining</i>	23	40	53	60	63	63
Cibles locales	8	5	3	5	6	6
Gestionnaires conseils allocataires	51	42	30	24	20	20
Signalements externes	3	3	3	3	4	4
Contrôles ressources activités	4	3	3	1	1	1
Autres	11	7	8	7	6	6
Total	100	100	100	100	100	100

Source : Cnaf - fichiers mensuels ALLNAT.

Tableau 2 – Écart entre la proportion d'allocataires contrôlés sur place et celle observée parmi l'ensemble des allocataires en 2006, 2010, 2013 et 2014 (points de pourcentage)

	2006	2010	2013	2014

Situation familiale				
Isolés sans enfant	9,0	5,4	- 9,0	- 11,6
Couples sans enfant	1,2	0,4	- 0,6	- 0,9
Couples avec enfants	- 28,9	- 27,1	- 17	- 17,9
Familles monoparentales	16,4	18,1	25,2	29,2
Âge du responsable du dossier				
Moins de 30 ans	5,6	3,9	- 6,6	- 3,7
30 ans à 39 ans	- 4,2	- 2,5	- 3,0	- 2,6
40 ans à 59 ans	0,6	0,0	8,0	7,0
60 ans et plus	- 4,9	- 4,8	- 5,1	- 6,6
Âge des enfants				
0 à 13 ans	- 13,2	- 10,6	0,7	3,7
14 à 17 ans	- 5,1	- 3,7	7,3	7,4
18 à 25 ans	2,2	1,6	24,4	23,7
Situation professionnelle du responsable du dossier				
Activité	- 21,9	- 19,4	-21	-22,7
Travailleurs indépendants	- 0,6	0,2	1,7	1,0
Congé longue durée	- 0,9	- 0,1	1,2	1,2
Retraités	- 4,7	- 4,8	- 5,0	- 6,5
Étudiants	- 3,3	- 3,8	- 5,5	- 5,8
Chômage	11,1	7,3	9,0	8,2
Inactivité	18,6	18,7	20,7	26,1
Revenus mensuels par unité de consommation				
0 à 500 euros	25,1	21,6	31,1	33,6
500 euros et plus	- 29,3	- 22,6	- 30,8	- 32,2

Source : Cnaf - fichiers mensuels ALLNAT. Champ : ensemble de la population des allocataires en 2006, 2010, 2013, 2014. Lecture : la proportion d'allocataires isolés sans enfant parmi la population contrôlée est supérieure de 9,9 points de pourcentage à celle que l'on retrouve parmi l'ensemble des allocataires.

Références bibliographiques

- Bernheim E., Commaille J., 2012, Quand la justice fait système avec la remise en question de l'État social, *Droit et société*, vol. 81, n° 2, p. 281-298.
- Baudot P.-Y., 2011, L'incertitude des instruments. L'informatique administrative et le changement dans l'action publique (1966-1975), *Revue française de science politique*, vol. 61, n° 1 p. 79-103.
- Biémouret G., Costes J.-L., 2016, *Rapport d'information sur l'évaluation des politiques publiques en faveur de l'accès aux droits sociaux*, n° 4158, Assemblée nationale, Paris.
- Bruno I., Didier E., Prévieux J., 2014, *Statactivisme comment lutter avec des nombres*, Paris, Zones.
- Buchet D., 2013, Le paiement à « bon droit » des prestations sociales des Caf, *Informations sociales*, n° 178, p. 97-103.
- Castel R., 1981, *La Gestion des risques. De l'antipsychiatrie à l'après-psychoanalyse*, Paris, Éditions de Minit.
- Chantel C., 2017, Les allocataires contrôlés par les caisses d'allocations familiales en 2015, *l'e-ssentiel*, n° 169.
- Collinet P., 2013, Focus – Le *data mining* dans les Caf : une réalité, des perspectives, *Informations sociales*, n° 178, p. 129-132.
- Desrosières A., 2014, *Prouver et gouverner. Une analyse politique des statistiques publiques*, Paris, La Découverte.
- Desrosières A., 1993, *La politique des grands nombres*, Paris, La Découverte.
- Dubois V., 2003, Les conditions sociopolitiques de la rigueur juridique. Politique de contrôle et lutte contre la fraude aux prestations sociales, Caisse nationale des Allocations familiales, *Dossier d'études*, n° 48.
- Ewald F., 1986, *L'État-providence*, Paris, Grasset.
- Fontaine M., Sicsic M., 2015, Des indicateurs précoces de pauvreté et d'inégalités - Résultats expérimentaux pour 2014, *Insee Analyses*, n° 23.
- Galanter M., 1974, Why the haves come out ahead?, *Law and Society Review*, vol. 9, n° 1, p. 95-160.
- Harcourt B. E., 2014, Digital security in the expository society: Spectacle, surveillance and exhibition in the neoliberal age of Big Data, *Columbia Public Law Research Paper*, n° 14-404.
- Lazarus J. 2012, Prévoir la défaillance de crédit : l'ambition du scoring, *Raisons politiques*, n° 48, p. 103-118.
- Lipsky M., 1980, *Street-level bureaucracy: Dilemmas of the individual in public services*, New York, Russell Sage Foundation.

- Ollion E., 2015, L'abondance et ses revers. Big data, open data et recherches sur les questions sociales, *Informations sociales*, n° 191, p. 70-79.
- Parton N., 2008, Changes in the form of knowledges in social work: From the social to the informational?, *British Journal of Social Work*, vol. 38, n° 2, p. 253-269.
- Peretti-Watel P., 2005, La culture du risque, ses marqueurs sociaux et ses paradoxes. Une exploration empirique, *Revue économique*, vol. 56, n° 2, p. 371-392.
- Spire A., 2012, *Faibles et puissants face à l'impôt*, Paris, Raisons d'agir.
- Supiot A., 2015, La gouvernance par les nombres. Cours au Collège de France 2012-2014, Paris, Fayard.
- Warin P., 2016, *Le non-recours aux politiques sociales Libres cours politique*, Grenoble, Presses universitaires de Grenoble.
- Weill P.-É., 2017, *Sans toit ni loi ? Genèse et conditions de mise en œuvre de la loi Dalo*, Rennes, Presses universitaires de Rennes.
- Zuurmond A., 2008, Bureaucratic bias and access to public services: The fight against non-take-up, *in* Jong J. (de), Rizvi L. G. (dir.), *The state of access Success and failure of democracies to create equal opportunities*, Washington, DC : Brookings Institution Press, p. 148-166.