

Modélisation statistique du surendettement des particuliers

Amandine Tran¹

Février 2021, WP #807

RÉSUMÉ

La procédure de surendettement est née vers la fin des années 1980, dans un contexte de forte expansion des crédits à la consommation. En 2010, un meilleur encadrement des pratiques commerciales de la part des pouvoirs publics a permis de limiter l'ampleur des crédits à la consommation. Le surendettement se concentre désormais sur des populations financièrement fragiles à faibles niveaux d'endettement. L'objectif de cette étude est de mettre en place un modèle explicatif du surendettement. Un premier modèle permet de mettre en évidence l'impact significatif du niveau de vie sur le fait d'être surendetté. Des disparités demeurent toutefois entre les structures familiales. La limite principale de ce premier modèle est le fait qu'il se fonde essentiellement sur des données agrégées. Un second modèle s'intéressant uniquement à la population des ménages surendettés et exploitant une base de données individuelles riche permet cette fois-ci de rendre compte de différents facteurs jouant un rôle dans l'orientation vers le rétablissement personnel, qui est l'orientation proposée aux ménages surendettés les plus en difficulté.

Mots-clés : Surendettement, Particuliers, Niveau de Vie, Structure Familiale, Logit, Enquête des Revenus Fiscaux et Sociaux, Rétablissement Personnel

JEL classification : C15, C51, C52

¹ ENSAE, amandine.tran@ensae.fr

Je remercie Karine Jean et Dominique Nivat de m'avoir supervisée et conseillée tout au long de la mise en œuvre de cette étude. Je tiens également à remercier Pierre Sicsic pour tous ses conseils et ses remarques concernant les méthodes et modèles statistiques utilisés. Enfin, je remercie très vivement toute l'équipe de la DPAR pour leur accueil.

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Ils sont disponibles sur publications.banque-france.fr

NON-TECHNICAL SUMMARY

Ce papier vise à présenter les principaux facteurs jouant un rôle dans le fait d'être un ménage surendetté en France. La procédure de surendettement est née vers la fin des années 1980, dans un contexte de forte expansion des crédits à la consommation et d'absence d'encadrement de ces derniers. Cette procédure a pour objectif d'apporter des solutions aux ménages en difficulté face à leurs dettes non professionnelles. Depuis 2010, un meilleur contrôle des pratiques commerciales par les pouvoirs publics a permis de considérablement limiter l'ampleur des crédits à la consommation. Aujourd'hui, les ménages affectés par le surendettement sont donc principalement des ménages vulnérables à faibles revenus. En 2019, une étude menée à la Banque de France a par exemple permis de mettre en évidence la surreprésentation des femmes seules avec enfant(s) parmi les ménages surendettés. Il s'agira donc de présenter une modélisation statistique du surendettement en France.

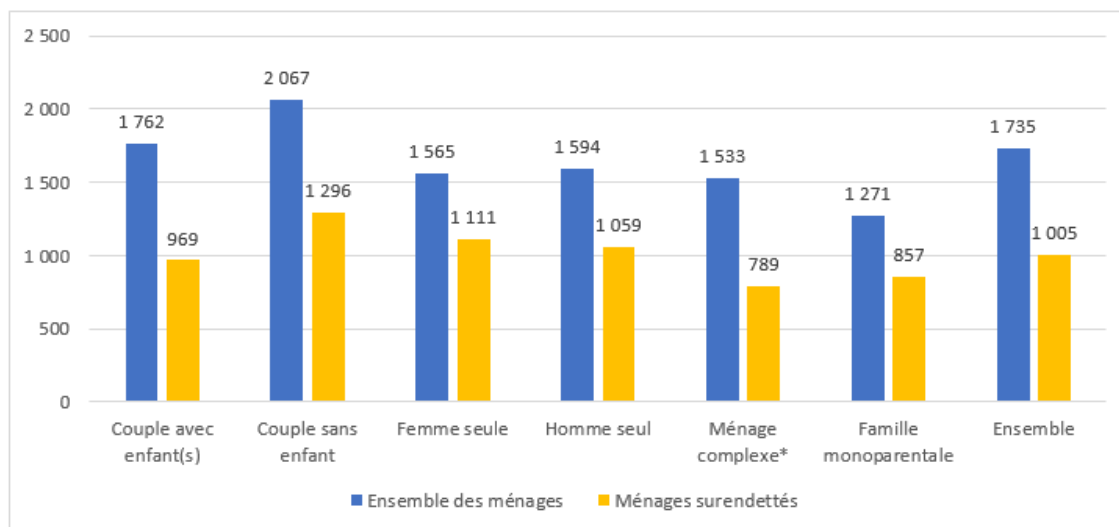
Pour analyser l'impact de certains facteurs le fait d'être surendetté, il est nécessaire d'analyser les caractéristiques de l'ensemble des ménages français. Or, les données individuelles concernant les ménages ne sont pas disponibles. Une alternative à cela consiste à fabriquer une base de données de ménages simulés grâce aux informations agrégées disponibles sur la structure familiale et le revenu des ménages (partie 1). Cette base de données est utilisée afin d'implémenter un modèle statistique permettant de quantifier les impacts respectifs des deux facteurs exploités (partie 2). Enfin, dans une dernière partie, il s'agit de se pencher sur la population même des ménages surendettés et de caractériser les facteurs déterminants de l'orientation vers le rétablissement personnel qui est proposée aux ménages surendettés aux situations financières les plus dégradées (partie 3).

De ces analyses, il ressort que le niveau de vie demeure le facteur essentiel du surendettement (cf. figure 1). Son impact est d'autant plus important que les conditions financières initiales sont fragiles : tandis qu'un complément de revenu aura un impact négligeable pour ménage aisé, cela permettrait de diminuer significativement la probabilité qu'un ménage aux revenus modestes soit surendetté. Cependant, la structure familiale joue également un rôle significatif, à niveaux de vie égaux : les cheffes de famille monoparentale sont par exemple les plus susceptibles d'être affectées par le surendettement.

Le rétablissement personnel est la solution proposée par la commission aux ménages dont les conditions financières sont telles qu'aucun plan de remboursement ne peut être envisagé. Cela implique l'effacement total des dettes dues. Parmi les ménages surendettés, ceux qui orientés par la commission vers cette solution sont également ceux dont les conditions de vie sont les plus dégradées : les individus sans emploi, à faibles revenus, les familles monoparentales ou les hommes seuls sont les plus vulnérables face à cette situation. Ils présentent également très souvent des montants de dettes faibles par rapport aux autres ménages surendettés. Il en résulte finalement qu'en 2019, seule 26,6% de la dette

globale des situations closes cette même année a été effacée alors que presque la moitié des ménages surendettés ont été orientés vers le rétablissement personnel.

Figure 1. Niveaux de vie mensuels médians des ménages et des ménages surendettés en 2017 (en €)



* Selon l'Insee, « un ménage complexe se définit par rapport aux autres types de ménages. Il s'agit d'un ménage qui n'est pas composé soit d'une seule personne, soit d'une seule famille (un couple sans enfant, un couple avec enfants ou une famille monoparentale) ».

Champ : Ménages ordinaires résidant en France métropolitaine en 2017 dont la personne de référence n'est pas un étudiant et ménages ayant déposé un dossier de surendettement recevable en 2017.

Source : Insee et Banque de France.

Statistical Modelization of Overindebtedness

ABSTRACT

The procedure for overindebtedness was born in the late 1980s, in a context of strong expansion of consumer credit. In 2010, better supervision of business practices by the public authorities helped to limit the scale of consumer credit. Overindebtedness now concerns financially fragile populations with low levels of debt. The main goal of this paper is to set up an explanatory model of overindebtedness. A first model highlights the significant impact of the standard of living on being overindebted. However, disparities remain between family structures. The main limitation of this first model is the fact that it relies essentially on aggregated data. Therefore, a second model, which is focused on the population of overindebted households only and uses a rich database, allows accounting for different factors playing a role in the orientation towards personal recovery, which is the orientation that is offered to the most financially vulnerable overindebted households.

Keywords : overindebtedness, standard of living, family structure, logit, Enquête des Revenus Fiscaux et Sociaux, personal recovery

Working Papers reflect the opinions of the authors and do not necessarily express the views of the Banque de France. This document is available on publications.banque-france.fr/en

1 Introduction

Vers la fin des années 1980, la levée de l'encadrement du crédit a entraîné une forte expansion des crédits à la consommation, ce qui a donné naissance à la procédure de surendettement. Cette procédure a pour objectifs d'apporter des solutions aux ménages surendettés tout en permettant aux créanciers de recouvrer la totalité ou une partie des sommes dues. Le surendettement est défini en France « par l'impossibilité manifeste de faire face à l'ensemble de ses dettes non professionnelles exigibles et à échoir » [1].

Depuis 2010, un meilleur contrôle des pratiques commerciales par les pouvoirs publics a permis de limiter considérablement l'ampleur des crédits à la consommation. On assiste alors à une diminution du nombre de situations de surendettement et à une concentration de ces situations sur des populations financièrement fragiles. Le tableau 1 ci-dessous représente le positionnement des ménages surendettés en termes de niveau de vie. Ainsi, près de 95% des ménages surendettés présentent un niveau de vie inférieur au niveau de vie médian. Deux tiers des ménages surendettés se situent en-dessous du second décile, ce qui illustre le fait que les ménages surendettés font essentiellement partie des ménages les plus modestes en France.

TABLE 1 – Niveau des ressources mensuelles nettes par unité de consommation (UC)

	Ménages surendettés 2017 ^{a)}	Population française 2017 ^{b)}
Ressources (R) < D1	42,5	10
$D1 \leq R < D2$	23,8	10
$D2 \leq R < D3$	13,4	10
$D3 \leq R < D4$	8,2	10
$D4 \leq R < D5$	4,9	10
$D5 \leq R < D6$	3,1	10
$D6 < D7$	1,9	10
$D7 \leq R < D8$	1,2	10
$D8 \leq R < D9$	0,7	10
$R \geq D9$	0,2	10

a) Situations de surendettement recevables. Ressources mensuelles après déduction de l'impôt sur le revenu et de la taxe d'habitation.

b) France métropolitaine. Revenu disponible des ménages par unité de consommation dont le revenu est positif ou nul et dont la personne de référence n'est pas étudiante – en

Note : les ressources nettes d'un ménage par UC correspondent à son niveau de vie. Une définition du niveau de vie est fournie en annexe 5.

Sources : Banque de France; Insee, DGFIP, Cnaf, Cnav et CCMSA, Enquête Revenus fiscaux et sociaux 2017.

Concernant la situation vis-à-vis du logement, la part de propriétaires parmi les ménages surendettés est très faible : en 2009, on comptabilisait seulement 8,3% de propriétaires (dont 5,3% de propriétaires accédants) au sein des ménages surendettés, contre 58,3% dans l'ensemble des

ménages français [3]. Ce constat a peu évolué par la suite. En effet, en 2019, parmi les ménages surendettés, on compte par exemple 74,5% de locataires et 11,5% de propriétaires contre 39% de locataires et 36,8% de propriétaires au sein de l'ensemble des ménages français. Les revenus des ménages apparaissent alors être un déterminant primordial du surendettement, contrairement au patrimoine puisque, à l'exception des propriétaires accédants de leur logement principal, très peu de surendettés détiennent un patrimoine [5].

La structure familiale du ménage semble également avoir une influence sur le fait d'être un ménage surendetté. Une étude réalisée en 2019 a notamment mis en évidence l'influence de la structure familiale du ménage sur le fait d'être surendetté, en particulier la vulnérabilité des femmes seules avec enfant(s) face à cette situation [2]. Cette catégorie de ménages figure cependant aussi parmi celles qui ont les revenus par unité de consommation les plus faibles. Ainsi, cela amène à s'interroger sur les impacts respectifs des ressources et de la structure familiale, mais également d'autres facteurs dans l'exposition au surendettement. L'objectif de la présente étude est donc de mettre en place un modèle statistique explicatif du surendettement.

Il s'agira d'abord de quantifier et d'analyser l'impact de la structure familiale et des ressources des ménages sur la probabilité d'être surendetté, en utilisant des données agrégées portant sur les ménages de France métropolitaine en 2017 et les données individuelles concernant les ménages surendettés. L'absence de données individuelles concernant les ménages non surendettés invite à reconstruire une population fictive de ménages afin de pouvoir mettre en place un modèle logistique à partir de la base de données ainsi créée. Dans une première partie, les caractéristiques des ménages en 2017 en termes de niveau de vie et de structure familiale seront donc exposées et la méthodologie déployée pour la création d'un jeu de données simulées à partir de ces deux variables sera également décrite. Il en ressort ainsi que les ménages surendettés ont un niveau de vie médian plus faible que l'ensemble des ménages et ce, quelle que soit la structure familiale. Parmi les ménages surendettés, les familles monoparentales et les hommes seuls sont en surreprésentation tandis que les couples avec ou sans enfant sont sous-représentés. On souhaite mettre en place un modèle logistique utilisant les variables de niveau de vie et de structure familiale. Pour cela, une population fictive de ménages non surendettés est construite à partir des données agrégées issues d'une enquête annuelle de l'Insee. En concaténant ces données avec les données disponibles concernant les ménages non surendettés, nous obtenons ainsi une pseudo-population de ménages sur laquelle le modèle logistique peut être implémenté.

Dans une seconde partie, les résultats issus de la modélisation de la probabilité d'être surendetté en fonction de ces deux facteurs seront détaillés. La principale conclusion qui se dégage de ces résultats est le fait que le niveau de vie est un facteur explicatif essentiel du surendettement. En effet, les ménages à revenus faibles ont plus de risques d'être surendettés que les ménages plus aisés. Accorder un complément de revenu à un ménage dont la situation financière est modeste permet aussi de diminuer plus fortement sa probabilité d'être surendetté par rapport à un ménage à

hauts revenus. La structure familiale demeure néanmoins un facteur significatif : les femmes seules avec enfant(s) sont, à niveaux de vie égaux, le type de ménage ayant la probabilité la plus forte d'être surendetté. Cette modélisation a pour limites principales l'absence de données individuelles concernant les ménages non surendettés et l'exploitation de seulement deux variables.

Dans une troisième partie, il s'agira de cibler l'analyse sur la population des surendettés uniquement. Le but sera d'étudier les caractéristiques des ménages les plus en difficulté parmi les ménages surendettés, c'est-à-dire ceux orientés par la commission vers le rétablissement personnel. Le rétablissement personnel, impliquant l'effacement total des dettes, est proposé lorsque la situation financière du ménage est telle qu'aucun plan de remboursement ne peut être envisagé. Une modélisation statistique de l'orientation vers cette solution sera ainsi développée. Ce second modèle a l'avantage principal d'utiliser une base de données individuelles disponibles comportant un nombre important de variables. De cette analyse, il en ressort que la probabilité d'être orienté vers le rétablissement personnel est plus élevée pour les ménages à faibles niveaux de vie, confrontés au chômage et pour les familles monoparentales ou les ménages composés d'un individu seul. À l'inverse, les caractéristiques des dettes n'ont pas d'impact significatif sur cette orientation.

2 Construction d'une population fictive de ménages utilisant le niveau de vie et la structure familiale

Afin d'étudier l'impact du niveau de vie et de la structure familiale sur le fait d'être surendetté, il est nécessaire d'analyser les caractéristiques de l'ensemble des ménages français. Or, les données individuelles concernant les ménages ne sont pas disponibles dans le cadre de cette étude. Une alternative à l'exploitation de données réelles est la simulation d'une pseudo-population de ménages ayant les mêmes caractéristiques que les ménages français. Pour cela, des données agrégées concernant la structure familiale et les ressources des ménages, issues des enquêtes les plus récentes de l'Insee, sont exploitées. La base de données ainsi fabriquée via R et SAS servira de base de travail pour la mise en œuvre d'un modèle de régression binaire. Dans un premier temps, les caractéristiques des ménages français en 2017 seront décrites et comparées à celles des ménages surendettés cette même année. La méthodologie mise en place pour la création du jeu de données sera détaillée dans une seconde section.

Dans le cadre de ce premier modèle (modélisation de la probabilité qu'un ménage soit surendetté en fonction du niveau de vie et de la structure familiale), deux sources de données sont donc utilisées :

- Les données concernant les dossiers de surendettement disponibles au sein de la Banque de France : les variables « structure familiale » et « niveau de vie » sont construites à partir des variables présentes dans les tables de données (statut matrimonial, nombre de personnes à

charge, nombre d'enfants à charge, montant détaillé des ressources) ;

- Des données agrégées issues de l'Enquête Revenus Fiscaux et Sociaux (ERFS) : cette enquête annuelle menée par l'Insee permet l'analyse des revenus en France selon des critères sociodémographiques usuels (CSP, âge des individus, composition du ménage, etc.). Les données utilisées provenant de cette enquête sont les moyennes et médianes de niveaux de vie détaillées par décile pour chaque structure familiale.

Les données les plus récentes concernant l'ERFS sont celles concernant les ménages français en 2017. Le premier modèle utilise donc uniquement les données de cette même année.

2.1 Caractéristiques des ménages français en 2017

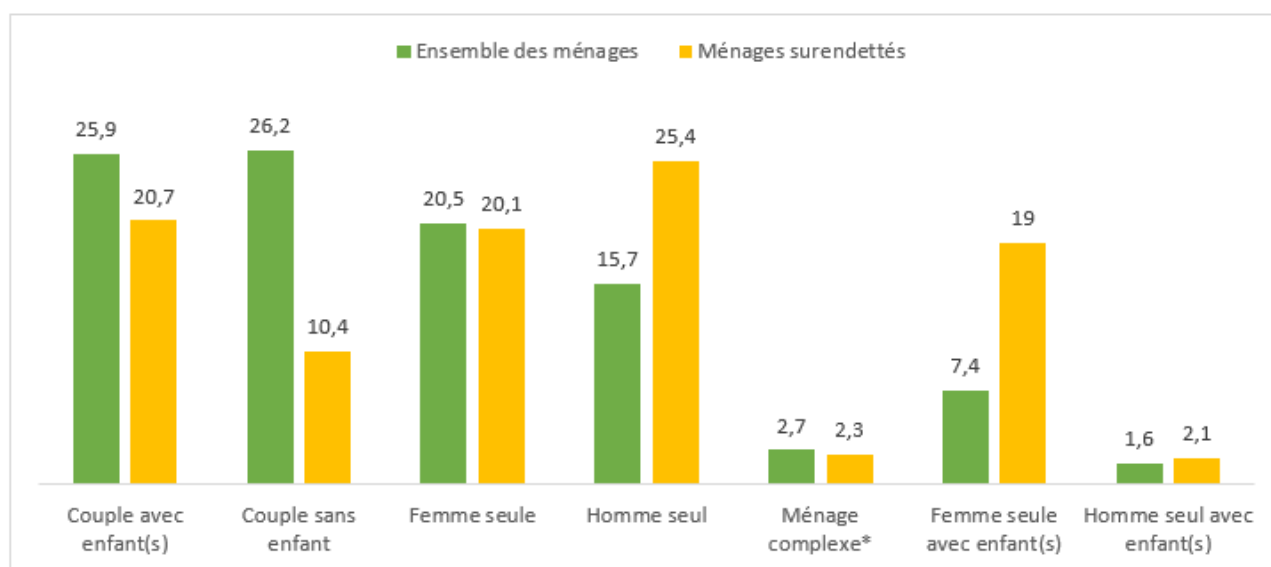
En 2017, on comptait en France près de 28,5 millions de ménages, dont 166 839 ont vu leur dossier de surendettement déclaré recevable¹, ce qui représente 0,6% des ménages. Par rapport à l'ensemble de la population, les ménages surendettés présentent des caractéristiques particulières, en termes de niveaux de vie et de structures familiales.

2.1.1 Structure familiale des ménages français

Tout d'abord, les hommes seuls et les femmes seules avec enfant(s) sont surreprésentés parmi les ménages surendettés (cf. figure 1 ci-dessous). On compte par exemple 25,4% d'hommes seuls parmi les ménages surendettés, contre 15,7% dans l'ensemble des ménages français. À l'opposé, les couples avec ou sans enfant(s) sont sous-représentés au sein de des ménages surendettés. Les couples sans enfant représentent ainsi 10,4% des ménages surendettés en 2017 contre 26,2% des ménages français cette même année, soit une différence de 15,8 points de pourcentage. Les proportions de ménages complexes, de femmes seules et d'hommes seuls avec enfant(s) parmi les ménages surendettés sont quant à elles équivalentes aux proportions observées au sein de la population française. Parmi les ménages surendettés, les hommes seuls sont la catégorie la plus nombreuse. À l'inverse, les hommes seuls avec enfant(s) représentent la catégorie la moins fréquente, que ce soit parmi les ménages surendettés ou dans l'ensemble des ménages.

1. « Une situation est considérée recevable par la commission de surendettement lorsque le ou les déposants sont dans l'impossibilité manifeste de faire face à l'ensemble de leurs dettes non professionnelles exigibles ou à échoir et qu'ils remplissent les conditions d'éligibilité à la procédure. La décision de recevabilité entraîne l'arrêt des poursuites, la suspension des pénalités financières et le gel des intérêts de retard. », (*Le surendettement des ménages. Enquête typologique 2019*, p.9, note en bas de page n°2, Banque de France)

FIGURE 1 – Structures familiales des ménages et des ménages surendettés en 2017 (en %)



* Selon l’Insee, « un ménage complexe se définit par rapport aux autres types de ménages. Il s’agit d’un ménage qui n’est pas composé soit d’une seule personne, soit d’une seule famille (un couple sans enfant, un couple avec enfants ou une famille monoparentale) ».

Champ : Ménages ordinaires résidant en France métropolitaine en 2017 dont la personne de référence n’est pas un étudiant et ménages ayant déposé un dossier de surendettement recevable en 2017.

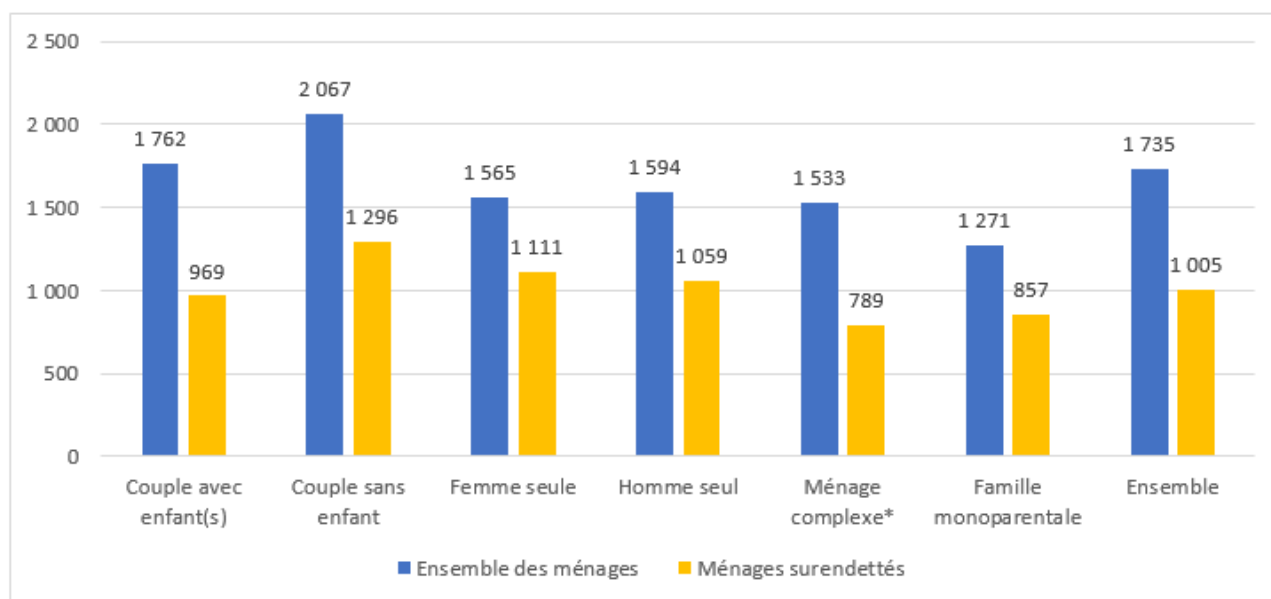
Sources : Insee et Banque de France.

Lecture : 19% des ménages surendettés en 2017 sont des femmes seules avec enfant(s).

2.1.2 Niveaux de vie des ménages français

Les ménages surendettés disposent généralement de ressources beaucoup plus modestes que le reste de la population. Les niveaux de vie médians constatés en 2017 par type de ménage ont été représentés sur la figure 2. Le niveau de vie mensuel médian des ménages surendettés est de 730 € inférieur au niveau de vie mensuel médian de l’ensemble des ménages. Les familles monoparentales et les ménages complexes sont les catégories de ménages présentant le plus faible niveau de vie médian parmi les surendettés et dans l’ensemble de la population. À l’inverse, les couples sans enfant sont la catégorie de ménage ayant le niveau de vie médian le plus élevé, à la fois chez les surendettés et parmi l’ensemble des ménages.

FIGURE 2 – Niveaux de vie mensuels médians des ménages et des ménages surendettés en 2017 (en €)²



* Selon l’Insee, « un ménage complexe se définit par rapport aux autres types de ménages. Il s’agit d’un ménage qui n’est pas composé soit d’une seule personne, soit d’une seule famille (un couple sans enfant, un couple avec enfants ou une famille monoparentale) ».

Champ : Ménages ordinaires résidant en France métropolitaine en 2017 dont la personne de référence n’est pas un étudiant et ménages ayant déposé un dossier de surendettement recevable en 2017.

Lecture : Le niveau de vie médian des hommes seuls surendettés s’établit à 1 059 € par mois en 2017, celui de l’ensemble des hommes seuls en 2017 s’établit à 1 594 € par mois.

Sources : Insee et Banque de France.

Parmi les ménages surendettés, les couples avec enfant(s), les familles monoparentales et les ménages complexes sont les catégories les plus pauvres : ils présentent des niveaux de vie médians inférieurs à celui de l’ensemble des surendettés. Cependant, dans l’ensemble des ménages, les couples avec enfant(s) ont un niveau de vie médian légèrement supérieur au niveau de vie médian général.

Au sein des ménages surendettés, les couples sans enfant, les hommes seuls et les femmes seules présentent des niveaux de vie médians plus élevés que celui de l’ensemble des ménages surendettés. Dans l’ensemble des ménages, au contraire, les hommes seuls et les femmes seules ont un niveau de vie médian inférieur au niveau de vie médian général en 2017.

Les niveaux de vie de l’ensemble des ménages et des ménages surendettés en 2017 peuvent être analysés de façon plus détaillée grâce aux déciles. Les tableaux 2 et 3 présentent les déciles de niveaux de vie par structure familiale, respectivement pour l’ensemble des ménages en 2017 et pour les ménages dont le dossier de surendettement a été déclaré recevable cette même année.

2. L’information supplémentaire concernant les familles monoparentales surendettées est disponible : le niveau de vie médian des femmes seules avec enfant(s) surendettées est de 891 € par mois ; celui des hommes seuls avec enfant(s) surendettés est de 939 € par mois.

<i>Déciles</i>	Ensemble	Famille monoparentale	Couple avec enfant	Couple sans enfant	Femme seule	Homme seul	Ménage complexe
p10*	933	757	966	1 206	906	773	824
p20	1 172	893	1 204	1 470	1 110	1 078	1 034
p30	1 371	1 017	1 406	1 675	1 275	1 263	1 188
p40	1 551	1 133	1 585	1 867	1 418	1 441	1 369
p50	1 735	1 271	1 762	2 067	1 565	1 594	1 533
p60	1 936	1 414	1 951	2 288	1 739	1 773	1 726
p70	2 178	1 577	2 176	2 566	1 961	1 990	1 924
p80	2 523	1 837	2 503	2 979	2 261	2 352	2 205
p90	3 184	2 269	3 140	3 723	2 816	3 027	2 745

TABLE 2 – Niveaux de vie mensuels des ménages en 2017 (en €)

* « p10 » est le seuil du 1^{er} décile de niveau de vie. Dans ce tableau, les 9 seuils des déciles sont présentés.

Champ : Ménages ordinaires résidant en France métropolitaine en 2017 dont la personne de référence n'est pas un étudiant.

Lecture : Le seuil du premier décile de niveau de vie mensuel de l'ensemble des ménages s'établit à 933 € en 2017.

Source : ERFS, Insee, 2017.

<i>Déciles</i>	Ensemble	Famille monoparentale	Femme seule avec enfant	Homme seul avec enfant	Couple avec enfant	Couple sans enfant	Femme seule	Homme seul	Ménage complexe
p10*	607	605	606	595	637	702	663	490	528
p20	730	700	701	697	742	888	766	722	611
p30	823	773	772	788	846	1 032	912	808	665
p40	935	829	826	858	933	1 164	1 021	969	730
p50	1 036	894	891	939	1 013	1 287	1 111	1 080	805
p60	1 140	972	965	1,035	1 103	1 420	1 228	1 205	885
p70	1 269	1 064	592	1,129	1 210	1 580	1 359	1 365	970
p80	1 439	1 174	1 057	1,249	1 355	1 782	1 522	1 544	1 082
p90	1 711	1 332	1 165	1 422	1 590	2 095	1 778	1 826	1 268

TABLE 3 – Niveaux de vie mensuels des ménages surendettés dont le dossier a été déclaré recevables en 2017 (en €)

* « p10 » est le seuil du 1^{er} décile de niveau de vie. Dans ce tableau, les 9 seuils des déciles sont présentés.

Champ : Ménages dont le dossier de surendettement a été considéré recevable en 2017.

Lecture : Le seuil du 1^{er} décile de niveau de vie mensuel de l'ensemble des ménages surendettés s'établit à 607 € en 2017.

Source : Banque de France

Encadré 1 : *Le niveau de vie*

Parmi les indicateurs permettant d'évaluer et de comparer les ressources des ménages ou des personnes, le niveau de vie semble le plus pertinent dans la mesure où il ramène des ménages de tailles différentes à une base standardisée.

Le niveau de vie d'un ménage correspond à son revenu disponible divisé par le nombre d'unités de consommation (uc). Le premier adulte du ménage compte pour 1 uc, les autres personnes du ménage de 14 ans comptent pour 0,3 uc chacun.

Par l'analyse des déciles de niveau de vie, on s'aperçoit que la distribution du niveau de vie des ménages surendettés est systématiquement décalée vers la gauche par rapport à la distribution du niveau de vie de l'ensemble des ménages, quelle que soit la structure familiale considérée. De plus, le seuil du 9^e décile du niveau de vie des surendettés (valant 1 711 € mensuels) est inférieur au niveau de vie médian de l'ensemble des ménages en 2017 (soit 1 735 € mensuels). On remarque également que le niveau de vie mensuel médian des ménages surendettés (qui vaut 1 036 €) est tout juste inférieur au seuil de pauvreté qui s'établit en 2017 à 1 041 € par mois³. Le surendettement concerne ainsi en grande majorité des ménages en situation de pauvreté monétaire.

Encadré 2 : *La procédure de surendettement en 2020* [1,5]

Selon les articles L. 711-1 et L. 712-2 du Code de la consommation, « la situation de surendettement est caractérisée par l'impossibilité manifeste de faire face à l'ensemble de ses dettes non professionnelles exigibles et à échoir ». Face à ces dettes, la procédure de surendettement, gratuite, permet d'apporter des solutions pour le(s) créancier(s) et le(s) débiteur(s).

Après le dépôt d'un dossier de surendettement par le débiteur (et le codébiteur éventuel), la commission de surendettement dispose d'un délai de 3 mois pour déclarer la recevabilité ou l'irrecevabilité du dossier, et son orientation le cas échéant. Le dossier de surendettement contient des informations précises concernant les caractéristiques des dettes, l'identité du débiteur et de son éventuel codébiteur, leur situation familiale et professionnelle, les ressources perçues, etc. Si le dossier est recevable, différentes orientations possibles peuvent être proposées par la commission en fonction de la situation constatée du ménage endetté :

- Le **rétablissement personnel** avec ou sans liquidation judiciaire : il s'agit de l'effacement total des dettes. Cette orientation est proposée lorsque la capacité de remboursement est quasi-inexistante et lorsque les difficultés financières sont si importantes qu'aucun plan de redressement ne peut être envisagé. Si le débiteur possède un patrimoine immobilier (financier ou autre), une procédure de liquidation judiciaire est

3. Il s'agit du seuil de pauvreté monétaire défini par l'Insee et correspondant à 60% du niveau de vie médian.
Source : *Tableaux de l'économie française*, Insee, 2017.
<https://www.insee.fr/fr/statistiques/2569358?sommaire=2587886#>

appliquée, avec l'accord du débiteur.

- La mise en place d'un **plan** de remboursement : si un bien immobilier est présent, un plan de remboursement des dettes est établi après une phase de conciliation avec les créanciers.
- Une **mesure imposée** avec ou sans effacement de dettes : si aucun bien immobilier n'est présent, une mesure de remboursement qui peut comprendre reports, rééchelonnement et/ou effacement partiel des dettes est établie.

Dans tous les cas, le déposant d'un dossier recevable est inscrit au FICP (Fichier national des Incidents de remboursement des Crédits aux Particuliers) pour une durée de 5 ans. Celui-ci peut être consulté par les établissements bancaires et financiers.

2.2 Création du jeu de données

On dispose d'une population de ménages surendettés dont on connaît de nombreuses données individuelles (situation familiale, revenus, âge, sexe, etc.). Pour étudier l'effet de différents facteurs tels que le niveau de vie et la structure familiale sur la probabilité d'être surendetté par un modèle de régression binaire, il faut également disposer de données individuelles sur la population de référence. Or, pour des raisons de secret statistique, nous n'avons pas accès à ces données.

En leur absence, nous reconstruisons une population française fictive de 28,5 millions de ménages dont les caractéristiques correspondent à celles des données agrégées publiées par l'Insee. Nous avons à notre disposition les moyennes et les déciles de niveaux de vie selon la structure familiale. L'utilisation d'autres variables potentiellement pertinentes telles que l'âge ou la situation professionnelles nécessiterait des informations précises auxquelles nous n'avons pas accès. Par conséquent, seules les informations concernant le niveau de vie et la structure familiale des ménages en 2017 seront utilisées pour la simulation de données. L'estimation de la probabilité qu'un ménage soit surendetté en fonction des ressources et de la structure familiale s'effectuera donc sur cette base de données simulées.

2.2.1 Simulation d'une population de ménages

L'objectif est de simuler une pseudo-population de ménages ayant une structure similaire aux ménages de France métropolitaine de 2017, en termes de moyennes et de déciles de niveaux de vie et de type de ménage. Pour cela, les données agrégées issues du tableau « Niveau de vie des individus et revenu des ménages correspondants par décile de niveau de vie selon le type de ménage » de l'*Enquête Revenus Fiscaux et Sociaux* (2017) sont utilisées. Ce tableau contient des informations concernant le niveau de vie annuel selon la structure familiale des ménages ordinaires résidant en France métropolitaine de 2017 dont la personne de référence n'est pas un étudiant [4].

Afin de simuler une population de ménages de référence, les éléments suivant sont exploités :

- La **distribution des ménages selon leur type** (couples avec enfant, couples sans enfant, femmes seules, hommes seuls, familles monoparentales, ménages complexes) ;
- Les **vingtiles de niveau de vie** pour chaque type de ménage ;
- Les **moyennes de niveau de vie** pour chaque décile.

Chaque type de ménage est traité séparément. La méthodologie suivante est appliquée pour chaque type de ménage :

1. Pour le premier décile :

Une grande partie des ménages surendettés présente des niveaux de vie inférieurs au seuil du premier décile de niveau de vie de l'ensemble des ménages en 2017 (environ 40% d'entre eux, d'après les tableaux 2 et 3). Par conséquent, une attention particulière doit être apportée à cette catégorie.

Le tableau « Répartition des niveaux de vie en 2015 » du dossier *Les revenus et le patrimoine des ménages* [7] de l'Insee suggère que la distribution du premier décile de niveau de vie est croissante et convexe. Ainsi, après divers essais sur plusieurs lois de probabilité (loi beta, loi gamma, densité de forme polynomiale), il a finalement été décidé de générer des observations issues d'une loi dont la densité est de la forme $ax^2 + b$ [8]. Il s'agissait en effet de la densité permettant d'obtenir les résultats en termes de moyennes et de médianes les plus proches des données réelles.

Les coefficients a et b peuvent être choisis de sorte que le niveau de vie moyen ou le niveau de vie médian du premier décile simulé soit similaire à celui réellement observé. Le niveau de vie moyen a finalement été choisi comme indicateur de référence, puisqu'il permettait d'obtenir les résultats agrégés les plus proches des données réelles en termes de moyennes et de médianes. Les données ont été simulées via une méthode de Monte-Carlo, en utilisant l'inverse de la fonction de répartition associée à la loi choisie (voir annexe A1).

2. Au-dessus du premier décile, jusqu'au quantile à 95% :

Des observations issues d'une loi uniforme simulée entre chaque quantile de niveau de vie sont générées.

3. Au-dessus du quantile à 95% :

Il s'agit des plus hauts niveaux de vie. Les hauts niveaux de vie sont généralement simulés en utilisant une loi de Pareto [8]. C'est le choix qui a été effectué ici (voir annexe A2 pour plus de détails).

Enfin, les données de chaque type de ménage sont concaténées afin d'obtenir la population finale. On s'assure que la distribution des ménages selon leur structure familiale est bien similaire à la population de départ. On compare ensuite les résultats obtenus sur ces données simulées aux résultats sur les ménages français de 2017. Les résultats principaux de cette comparaison sont récapitulés dans le tableau 4.

<i>Catégories de ménage</i>	<i>Déciles</i>	Déciles réels	Différences (en %)	Médianes réelles	Différences (en %)	Moyennes réelles	Différences (en %)
Ensemble	p10	11 190	1,3	9 140	3,9	8 460	3,3
	p20	14 060	0,4	12 800	0,7	12 790	1,1
	p30	16 450	0,2	15 350	0,9	15 340	0,8
	p40	18 610	0,3	17 540	0,7	17 540	0,6
	p50	20 820	0,5	19 660	0,7	19 670	0,6
	p60	23 230	0,7	21 920	0,5	21 950	0,5
	p70	26 140	0,6	24 600	0,3	24 610	0,4
	p80	30 270	0,1	27 940	-0,1	28 050	0,2
	p90	38 210	-0,1	33 540	-0,1	33 760	-0,1
	p95	47 650	0,0	42 050	-0,3	42 350	0,0
	>p95	,	,	59 960	-0,1	72 250	-0,2
	Ensemble	,	,	20 820	0,5	23 920	0,3
Homme seul	p10	9 280	0,0	7 330	8,8	6 120	0,0
	p50	19 130	0,0	18 220	0,0	18 200	-0,1
	>p95	,	,	57 990	-2,8	73 680	0,3
	Ensemble	,	,	19 130	0,0	22 300	-0,1
Femme seule	p10	10 870	0,0	8 720	5,9	7 570	0,0
	p50	18 780	0,0	17 840	0,0	17 870	0,0
	>p95	,	,	52 030	1,1	60 440	0,1
	Ensemble	,	,	18 780	0,0	21 400	-0,1
Famille monoparentale	p10	9 080	0,0	7 930	4,8	7 280	0,0
	p50	15 250	0,0	15 000	0,0	15 010	2,0
	>p95	,	,	40 770	-0,9	48 940	0,1
	Ensemble	,	,	15 250	0,0	17 330	-1,7
Couple sans enfant	p10	14 470	0,0	12 070	4,1	10 990	0,0
	p50	24 800	0,0	23 530	0,0	23 560	0,0
	>p95	,	,	68 370	-2,9	84 300	0,2
	Ensemble	,	,	24 800	0,0	28 550	-0,1
Couple avec enfant(s)	p10	11 590	0,0	9 750	0,3	9 390	0,0
	p50	21 140	0,0	20 610	0,0	20 600	1,2
	>p95	,	,	61 400	2,2	72 960	0,0
	Ensemble	,	,	21 140	0,0	24 080	-1,4
Complexe	p10	9 890	0,0	7 070	12,4	5 760	0,0
	p50	18 400	0,0	17 900	0,0	17 870	1,2
	>p95	,	,	NA	NA	NA	NA
	Ensemble	,	,	18 400	0,0	20 440	-1,4

TABLE 4 – Comparaison entre la simulation et les données réelles (montants en € annuels)

NB : Les écarts en rouge dans ce tableau sont ceux qui sont supérieurs à 3%. Les résultats concernant les autres déciles ne sont pas présents dans ce tableau, car peu de différences sont constatées avec les données réelles.

Champ : ménages ordinaires résidant en France métropolitaine en 2017 dont la personne de référence n'est pas un étudiant.

Lecture : en 2017, le seuil de 1er décile du niveau de vie des hommes seuls s'établit à 9 280 €. Au sein de ce décile, le niveau de vie médian est de 7 330 €. La médiane simulée pour cette catégorie est inférieure à la médiane réelle de 8,8%.

Sources des données réelles : ERFS, Insee, 2017

Le tableau 4 ci-dessus montre peu d'écarts entre les moyennes simulées et les moyennes réelles par décile et par structure familiale. Les écarts sont en effet tous inférieurs en valeur absolue à 2%. Lorsque les données sont agrégées, on constate néanmoins un écart supérieur à 3% entre la moyenne réelle et simulée du niveau de vie annuel au sein du premier décile. Étant donné que le premier décile du niveau de vie de chaque type de ménage a été distribué en utilisant la moyenne, on constate donc peu d'écarts aux moyennes, mais des écarts plus importants aux médianes du premier décile, sauf pour les couples avec enfant(s). Après concaténation des données simulées par structure familiale, la distribution du niveau de vie simulé demeure relativement proche de la distribution réelle.

Les distributions des niveaux de vie des femmes seules avec enfant(s) et des hommes seuls avec enfant(s) ne sont pas disponibles : seule celle des familles monoparentales peut être exploitée. Afin de pouvoir étudier tout de même la catégorie des cheffes de famille monoparentale, on suppose que les femmes et les hommes seuls avec enfant(s) présentent des distributions identiques aux familles monoparentales termes de niveau de vie. Cette hypothèse biaise faiblement les données concernant les cheffes de famille monoparentale, puisque ces dernières représentent plus de 80% des familles monoparentales. En revanche, les résultats concernant les hommes chefs de famille monoparentales seront à interpréter avec précaution.

Après l'obtention de la population de ménages simulés, les ménages fictifs aux caractéristiques les plus proches des ménages surendettés ont été remplacés par les ménages surendettés dont le dossier a été déclaré recevable en 2017, afin d'obtenir une pseudo-population de référence composée à la fois de ménages surendettés et de ménages non surendettés.

2.2.2 Description de la base de données finale exploitée

Afin de mettre en évidence les effets du niveau de vie et de la structure familiale sur le fait d'être surendetté, des régressions logistiques ont été réalisées via R. Pour des raisons techniques (problèmes de mémoire vive), ces modèles ont été appliqués sur un échantillon représentant 20% de la base initiale, contenant près de 5,7 millions d'observations. Des essais ont également été effectués sur une base composée de 50% des observations, ainsi que sur une base où seules les observations ayant un niveau de vie inférieur à 3 000€ par mois (soit plus de 99% des surendettés) ont été conservées. Les résultats se sont avérés tous quasi-identiques. Par conséquent, les résultats présentés ici sont ceux du modèle implémenté sur l'échantillon de 20% représentatif de la base initiale. Comme les ménages surendettés représentent une très faible proportion des ménages français, on examine au préalable le nombre d'observations par type de ménage dans l'échantillon des surendettés via le tableau 5 ci-contre :

TABLE 5 – Structures familiales des ménages surendettés de l'échantillon de travail

Structure familiale	Nombre d'observations	% des ménages surendettés
Complexe	772	2,3
Couple avec enfant	6 915	20,7
Couple sans enfant	3 467	10,3
Femme seule avec enfant(s)	6 356	19,1
Femme seule	6 726	20,1
Homme seul avec enfant(s)	706	2,1
Homme seul	8 476	25,4

Champ : Ménages surendettés issus de l'échantillon représentatif de 20% des données initiales, soit 33 368 observations

Lecture : 2,3% des observations de l'échantillon des ménages surendettés sont des ménages complexes, ce qui représente 772 unités.

L'échantillon des surendettés (qui est la catégorie minoritaire) présente a priori un nombre suffisamment important d'observations par modalité de structure familiale pour effectuer une régression logistique : on observe plus de 700 observations par modalité. L'échantillon final utilisé pour la modélisation contient ainsi près de 5,7 millions dont 33 368 ménages surendettés, soit 0,6% des observations de l'échantillon (voir annexe A3 pour la structure de la base de données finale).

3 Revenu disponible, structure familiale et surendettement

En 2017, on comptait près de 0,6% de ménages surendettés en France. Le surendettement est donc une situation qui concerne une partie spécifique de la population française. Malgré le faible nombre de personnes dans l'impossibilité de faire face à l'ensemble de leurs dettes, on peut relever des facteurs influençant la probabilité d'être surendetté, en particulier le niveau de vie et la structure familiale.

Afin de mettre en évidence ces effets, des régressions logistiques ont été réalisées via R sur un échantillon représentatif de 20% de la base initiale, ce qui représente 5,7 millions de lignes. Pour rappel, l'unité considérée est le ménage dans le cadre de ce modèle.

3.1 Le niveau de vie est un facteur explicatif déterminant

Tout d'abord, l'effet de la seule structure familiale sur le surendettement sera analysé via un modèle logistique utilisant uniquement la variable de structure familiale. Ensuite, le niveau de vie sera ajouté au modèle pour permettre d'évaluer l'impact des ressources par rapport à la structure familiale sur la probabilité d'être exposé au surendettement. Si le niveau de vie est un facteur explicatif déterminant du surendettement, la structure familiale joue tout de même un rôle

significatif additionnel.

Par la suite, la modalité « Homme seul » sera utilisée en tant que modalité de référence pour la variable de structure familiale dans les régressions effectuées. En effet, il s'agit de la catégorie de ménage la plus fréquente parmi les surendettés. On pourra donc se référer à cette dernière en tant que ménage surendetté dit « typique », ce qui facilitera les comparaisons et les interprétations.

3.1.1 Régression sur la structure familiale

Dans un premier temps, une régression logistique du fait d'être un ménage surendetté sur la structure familiale du ménage est réalisée dans le but de voir l'impact de cette dernière lorsqu'elle est utilisée en tant que seule variable explicative.

Soit la variable Y telle que $Y = 1$ si le ménage est surendetté, et $Y = 0$ si le ménage n'est pas surendetté. Soit $F = (\mathbf{1}_{\text{complexe}}, \mathbf{1}_{\text{couple avec enfant(s)}}, \mathbf{1}_{\text{couple sans enfant}}, \mathbf{1}_{\text{femme seule}}, \mathbf{1}_{\text{femme seule avec enfant(s)}}, \mathbf{1}_{\text{homme seul}}, \mathbf{1}_{\text{homme seul avec enfant(s)}})$ un vecteur d'indicatrices du type de ménage. Le modèle logistique exploité s'écrit de la façon suivante :

$$\text{logit}(P(Y = 1|F)) = \beta_0 + \sum_{m \in \text{TypesMenages}} \beta_m \mathbf{1}_m$$

Les informations concernant les coefficients du modèle estimé sont résumées dans le tableau 6 ci-dessous :

TABLE 6 – Coefficients estimés du modèle utilisant la structure familiale

Variable	Coefficient	Écart-type	P valeur
Constante	-4,65	0,011	<0,01
Complexe	-0,64	0,038	<0,01
Couple avec enfant	-0,71	0,016	<0,01
Couple sans enfant	-1,41	0,020	<0,01
Femme seule avec enfant(s)	0,48	0,017	<0,01
Femme seule	-0,50	0,016	<0,01
Homme seul avec enfant(s)	-0,23	0,039	<0,01

Les coefficients associés aux différentes modalités de la structure familiale sont tous négatifs, à l'exception de celui associé à « femme seule avec enfant(s) ». Ils sont aussi tous statistiquement significatifs au seuil de 1%. Ainsi, toutes choses égales par ailleurs, par rapport à un homme seul (qui correspond au type de ménage le plus fréquent parmi les surendettés), une femme seule avec enfant(s) présente une probabilité plus élevée d'être exposée au surendettement. À l'inverse, les ménages appartenant aux autres catégories ont une probabilité plus faible d'être surendettés que les hommes seuls, toutes choses égales par ailleurs. Le pseudo-R2 permettant de juger de la qualité

du modèle est cependant faible : il vaut 0,03 pour ce modèle. Ce modèle est donc très faiblement ajusté selon ce critère.

La performance du modèle peut être étudiée de façon plus approfondie via une matrice de confusion. Afin de construire cette matrice, on choisit un seuil d'affectation à la classe positive de 0,6%, ce qui correspond à la proportion de ménages surendettés dans l'ensemble des ménages. Cela signifie que si un ménage a une probabilité d'être surendetté supérieure à 0,6% selon le modèle, on considère que le modèle prédit que le ménage est surendetté ; lorsque la probabilité qu'un ménage soit surendetté est inférieure à 0,6% selon le modèle, on considère que le modèle prédit l'absence de surendettement. Le choix de ce seuil est motivé par le fait qu'il correspond à la probabilité « simple » d'être surendetté, c'est-à-dire en ne tenant pas compte de facteurs explicatifs tels que le niveau de vie ou la structure familiale. Ainsi, avoir une probabilité d'être surendetté supérieure à ce seuil signifie avoir une probabilité plus élevée que la probabilité moyenne.

TABLE 7 – Matrice de confusion du modèle utilisant la structure familiale seulement

	Réels non surendettés	Réels surendettés
Prédits non surendettés	75,4 %	53,5 %
Prédits surendettés	24,6 %	46,5 %
	100 %	100 %

Lecture : 46,5% des ménages surendettés présents dans la base de travail ont été détectés comme « surendettés » par le modèle.

La sensibilité (% de bien prédits parmi les ménages surendettés) est à 45,6% et la spécificité (% de bien prédits parmi les ménages non surendettés) vaut 75,4%.

Ce modèle présente ainsi une spécificité élevée mais une sensibilité très faible : il prédit mal la catégorie des surendettés. On remarque qu'il tend à classer la quasi-totalité des observations dans la catégorie des « non surendettés », ce qui expliquerait la faible sensibilité et le fort taux de non surendettés correctement prédits. Ainsi, ce modèle n'incluant que la variable « structure familiale » semble avoir une performance faible. À présent, tentons de l'améliorer en ajoutant la variable « niveau de vie ».

3.1.2 Régression sur la structure familiale et le niveau de vie

Dans un second temps, une régression logistique est effectuée en utilisant les deux variables explicatives disponibles : la structure familiale et le niveau de vie. Soient N la variable « niveau de vie mensuel » (en milliers d'euros) et F un vecteur d'indicateurs du type de ménage défini comme précédemment. On a dans cette partie le vecteur de variables explicatives $X = (N, F)$ et le modèle

logistique exploité s'écrit de la façon suivante :

$$\text{logit}(P(Y = 1|X = (N, F))) = \beta_0 + \beta_n N + \sum_{m \in \text{Types Menages}} \beta_m \mathbf{1}_m$$

Le tableau 8 ci-contre résume les informations concernant les coefficients estimés :

TABLE 8 – Coefficients estimés du modèle utilisant la structure familiale et le niveau de vie

Variable	Coefficient	Écart-type	P valeur
Constante	-2,52	0,015	<0,01
Ménage complexe	-0,71	0,038	<0,01
Couple avec enfant(s)	-0,37	0,017	<0,01
Couple sans enfant	-0,74	0,021	<0,01
Femme seule avec enfant(s)	0,27	0,017	<0,01
Femme seule	-0,43	0,017	<0,01
Homme seul avec enfant(s)	-0,44	0,039	<0,01
Niveau de vie	-1,61	0,010	<0,01

NB : Le coefficient associé à la variable « niveau de vie » est exprimé en milliers d'euros.

Les coefficients de ce modèle sont tous statistiquement significatifs au seuil de 1%. Ceux qui correspondent aux modalités de la structure familiale ont le même signe que dans le modèle précédent. Ainsi, l'interprétation des effets de la structure familiale sur la probabilité d'être surendetté demeure inchangée concernant le sens de l'effet. Le coefficient associé au niveau de vie est négatif, ce qui traduit le fait que pour un ménage, plus le niveau de vie est élevé, plus faible est sa probabilité d'être exposé au surendettement. Cela est cohérent avec le constat effectué depuis une dizaine d'années : le surendettement touche surtout des personnes et des ménages dont la situation financière est particulièrement difficile.

Les coefficients associés à la variable « structure familiale » sont plus faibles en valeur absolue dans ce second modèle que dans le modèle précédent. Par exemple, le coefficient associé à la modalité « couple avec enfant(s) » vaut -0,71 dans le premier modèle, contre -0,37 dans le second modèle. Il est donc presque deux fois plus faible en valeur absolue dans ce second modèle que dans le premier modèle. Les odds ratios, définis comme le rapport des cotes d'un événement entre groupes d'individus (voir annexe B1 pour plus de détails), peuvent être obtenus en prenant les exponentielles des coefficients correspondants. Par exemple :

- L'odd ratio pour la modalité « couple avec enfant(s) » vaut $e^{-0,71} = 0,49$ dans le premier modèle : les risques d'être surendetté plutôt que de ne pas être surendetté pour un couple avec enfant(s) valent 0,49 fois ceux d'un homme seul.
- L'odd ratio pour la modalité « couple avec enfant(s) » vaut $e^{-0,37} = 0,69$ dans le second modèle : les risques d'être surendetté plutôt que de ne pas être surendetté pour un couple

avec enfant(s) valent 0,69 fois ceux d'un homme seul.

Les disparités entre les structures familiales sont moins marquées dans ce second modèle. Sans la variable « niveau de vie », l'effet de la structure familiale sur le fait d'être surendetté était donc surestimé : si les femmes seules avec enfant(s) apparaissent comme la catégorie de ménage la plus vulnérable face au surendettement, cela s'explique en grande partie par leur niveau de vie plus faible en moyenne que celui des autres structures familiales.

Le pouvoir explicatif de ce modèle étudié via le pseudo-R2 apparaît bien meilleur que celui du premier modèle qui n'utilisait que la structure familiale. Il faut cependant prendre en compte le fait que le pseudo-R2 augmente mécaniquement avec le nombre de variables ajoutées dans un modèle. De plus, le pseudo-R2 demeure ici très faible. Afin de rendre compte de la performance du modèle et de le comparer au modèle précédent, une matrice de confusion peut à nouveau être construite, en utilisant un seuil d'affectation identique à celui choisi pour la matrice précédente (cf tableau 7). Les résultats correspondants sont représentés dans le tableau 9 :

TABLE 9 – Matrice de confusion du modèle utilisant la structure familiale et le niveau de vie

	Réels non surendettés	Réels surendettés
Prédits non surendettés	72,5 %	25,5 %
Prédits surendettés	27,5 %	74,5 %
	100 %	100 %

Lecture : 74,5% des ménages surendettés présents dans la base de travail ont été détectés comme « surendettés » par le modèle.

Les critères de sensibilité et spécificité suggèrent que le modèle est plutôt performant avec le seuil choisi. En effet, plus de 70% des situations de surendettement sont correctement prédites, le modèle détectant également plus de 70% des ménages non surendettés. De plus, la sensibilité de ce modèle est bien meilleure que celle du modèle n'exploitant que la variable de structure familiale, puisqu'il ne classe pas systématiquement les observations dans la classe majoritaire (qui est dans notre cas, la classe des non surendettés comprenant 99,4% des observations). La courbe ROC associée à ce modèle, représentée par la figure B3.1 en annexe B3, permet également d'évaluer la performance du modèle. L'AUC qui correspond à l'aire sous la courbe ROC vaut environ 0,80. L'allure de la courbe et l'AUC élevé suggèrent que le modèle présente une bonne performance.

On en conclut que le niveau de vie a un impact significatif et important sur le fait d'être surendetté, mais que la structure familiale demeure un facteur explicatif non négligeable.

3.2 Probabilités estimées et interprétations

Le modèle que nous avons construit montre que tant le niveau de vie que la structure familiale jouent un rôle dans le fait d'être surendetté. En effet, plus un ménage dispose de ressources importantes,

moins il a de probabilité d'être surendetté. De plus, les femmes seules avec enfant(s) apparaissent comme le type de ménages le plus susceptible d'être surendetté, tandis que les couples sans enfant sont ceux qui semblent avoir le moins de risques d'être exposés au surendettement. Néanmoins, ces effets ne sont pas linéaires : si augmenter le niveau de vie d'un certain montant permet d'améliorer significativement la situation d'un ménage modeste, cela aura moins d'impact sur un ménage aisé qui présente déjà une probabilité très faible d'être surendetté.

3.2.1 Les ménages aux situations précaires et les cheffes de famille monoparentales plus vulnérables face au surendettement

Les probabilités d'être surendetté estimées par le modèle pour plusieurs ménages hypothétiques peuvent être analysées. Le tableau 10 ci-dessous présente quelques exemples de probabilités estimées pour les différentes structures familiales, et en fonction de la situation sur l'échelle des niveaux de vie.

TABLE 10 – Exemples de probabilités de surendettement estimées (en %)

NB : les écarts-types des estimations sont entre parenthèses.

		Femme seule avec enfant(s)	Homme seul avec enfant(s)	Femme seule	Homme seul	Couple avec enfant(s)	Couple sans enfant	Famille complexe
D1 surendettés*	607 €	3,82 (0,0005)	1,90 (0,0007)	1,95 (0,0003)	2,95 (0,0003)	2,04 (0,0003)	1,37 (0,0003)	1,54 (0,0005)
D5 surendettés	1 036 €	1,95 (0,0002)	0,96 (0,0004)	0,98 (0,0001)	1,50 (0,0002)	1,03 (0,0001)	0,69 (0,0001)	0,77 (0,0005)
D5 tous ménages	1 735 €	0,64 (0,0001)	0,31 (0,0001)	0,32 (0,0000)	0,49 (0,0001)	0,34 (0,0000)	0,23 (0,0000)	0,24 (0,0001)
D8 tous ménages	2 523 €	0,18 (0,0000)	0,09 (0,0000)	0,09 (0,0000)	0,14 (0,0000)	0,09 (0,0000)	0,06 (0,0000)	0,07 (0,0000)

* « D1 surendettés » correspond au seuil du 1^{er} décile de niveau de vie des ménages surendettés et « D5 surendettés », à la médiane du niveau de vie des ménages surendettés. « D5 tous ménages » et « D8 tous ménages » sont respectivement la médiane et le seuil du 8^e décile du niveau de vie des ménages français.

Lecture : Une femme seule avec enfant(s) ayant un niveau de vie mensuel de 607 € a une probabilité estimée de 3,82% d'être surendettée.

Pour un niveau de vie donné, les femmes seules avec enfant(s), ont systématiquement une probabilité plus élevée d'être surendettées que tous les autres types de ménages, ce qui correspond au coefficient positif constaté dans le tableau 8. Ainsi, pour un niveau de vie mensuel égal à 1 036€ (soit le niveau de vie médian des ménages surendettés), une femme cheffe de famille monoparentale aura une probabilité d'être surendettée égale à 1,95%, contre 1,50% pour un homme seul ou encore 0,69% pour un couple sans enfant. La probabilité qu'une femme seule avec enfant(s) soit surendettée est aussi presque deux fois plus élevée que celle d'une femme seule sans enfant,

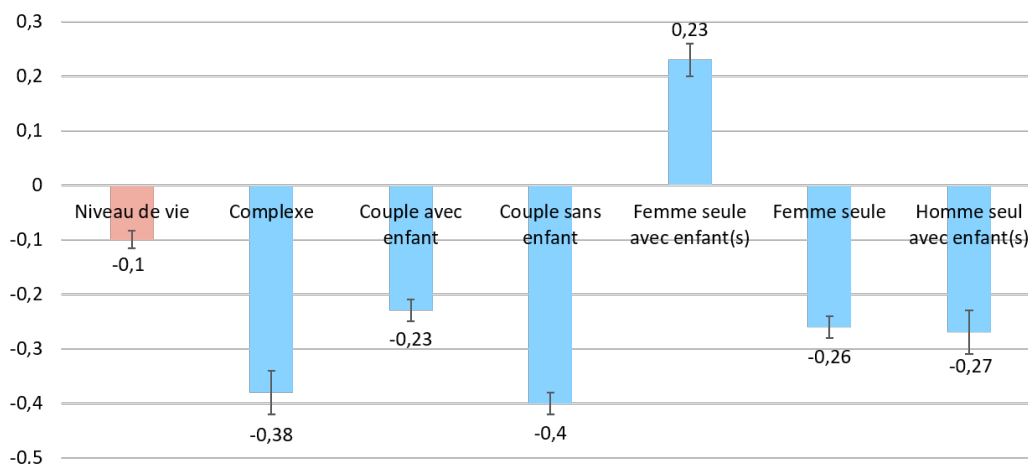
à niveaux de vie égaux. À l’opposé, à niveau de vie comparable, les couples sans enfant(s) apparaissent comme le type de ménage ayant les plus faibles probabilités d’être surendetté. Ainsi, pour un niveau de vie égal au niveau de vie médian des ménages en 2017, ils ont une probabilité d’être surendettés égale à 0,23%, contre 0,49% pour un homme seul ou 0,34% pour un couple avec enfant(s). Concernant les ménages complexes et les hommes seuls avec enfant, les résultats sont à interpréter avec précaution puisqu’ils représentent une infime proportion des ménages en France. Enfin, il reste que le niveau de vie joue un rôle primordial dans la probabilité d’être surendetté : celle-ci décroît en effet à mesure que le niveau de vie s’élève, quel que soit le type de ménages concerné. La probabilité qu’une femme seule soit surendettée diminue par exemple d’environ 67% lorsque son niveau de vie mensuel passe de 1 036 € à 1 735 €.

En utilisant les probabilités de surendettement du tableau 10, il est possible de calculer les écarts de probabilité moyens causés par une augmentation du niveau de vie mensuel d’une certaine quantité. Soit $p_{i,j}$ la probabilité de surendettement (en %) constatée à la ligne i et la colonne j du tableau 10. Soient $(r_1, r_2, r_3, r_4) = (607, 1036, 1735, 2523)$ les paliers de niveaux de vie du même tableau et f_j la proportion de ménages de structure familiale j au sein de l’échantillon (voir figure 1). L’effet marginal moyen du niveau de vie sur la probabilité d’être surendetté peut s’obtenir de la façon suivante :

$$EMM_{niveau\ de\ vie} = \sum_{j=1}^7 \left(\frac{f_j}{3} \sum_{i=1}^3 \frac{1}{r_{i+1} - r_i} (p_{i+1,j} - p_{i,j}) \right)$$

Une augmentation de 100 euros mensuels entraîne alors une diminution d’environ 0,1 point de pourcentage (pp.) de probabilité d’être surendetté, qui correspond à l’effet moyen calculé sur l’ensemble de l’échantillon (voir figure 3 ci-dessous).

FIGURE 3 – Effets marginaux moyens (en pp.)



NB : Sur ce graphique, pour le niveau de vie, est représenté l'effet marginal moyen d'une augmentation de 100 € mensuels. Les intervalles de confiance représentés sont ceux à 95%.

Lecture : En moyenne, une augmentation du niveau de vie de 100€ par mois conduit à une diminution de 0,1 point de pourcentage de la probabilité d'être surendetté.

En moyenne, être une femme seule avec enfant(s) plutôt qu'un homme seul augmente la probabilité d'être surendetté de 0,23 point de pourcentage.

Les effets décrits par le graphique ci-dessus traduisent uniquement des effets « moyens » et ne prennent donc pas en compte les effets de composition des ménages, ce qui rend difficile leur interprétation. On peut imaginer que les liens entre niveau de vie et probabilité d'être surendetté soient différents entre les types de ménages et selon la situation des ménages sur l'échelle des niveaux de vie. Par exemple, entre deux ménages à revenus modestes, une faible différence de niveau de vie peut avoir un impact important en termes de probabilité de surendettement, alors que la même différence entre deux ménages aisés, ayant une faible probabilité d'être surendetté, n'aura que peu d'effet. Afin de rendre compte de ces effets différenciés, nous pouvons analyser quelques exemples.

3.2.2 Des effets différenciés du niveau de vie selon la structure familiale

Afin de rendre compte des effets différenciés d'une augmentation du niveau de vie ou d'un changement de structure familiale, nous pouvons analyser plusieurs exemples. Le tableau 11 ci-dessous présente l'effet marginal d'une augmentation du niveau de vie de 100 € mensuels pour différents ménages hypothétiques :

TABLE 11 – Effets d’une augmentation de 100 € du niveau de vie mensuel sur la probabilité estimée par le modèle d’être surendetté (en pp.)

		Femme seule avec enfant(s)	Homme seul avec enfant(s)	Femme seule	Homme seul	Couple avec enfant(s)	Couple sans enfant	Famille complexe
D1 surendettés*	607 €	-0.55	-0.28	-0.28	-0.43	-0.3	-0.21	-0.22
D5 surendettés	1 036€	-0.28	-0.14	-0.14	-0.22	-0.15	-0.11	-0.11
D5 tous ménages	1 735€	-0.09	-0.05	-0.05	-0.07	-0.05	-0.03	-0.04
D8 tous ménages	2 523 €	-0.03	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01

* « D1 surendettés » correspond au seuil du 1^{er} décile de niveau de vie des ménages surendettés et « D5 surendettés », à la médiane du niveau de vie des ménages surendettés. « D5 tous ménages » et « D8 tous ménages » sont respectivement la médiane et le seuil du 8^e décile du niveau de vie des ménages français.

Lecture : Pour une femme seule avec enfant(s) ayant un niveau de vie mensuel initial de 607 €, une augmentation du niveau de vie de 100 € mensuels entraîne une diminution de 0,55 point de pourcentage de la probabilité d’être surendettée.

1. *Comment évolue la probabilité d’un ménage d’être surendetté selon sa situation initiale, lorsque son niveau de vie mensuel augmente de 100 €?*

Pour un ménage ayant un niveau de vie très faible (inférieur au niveau de vie médian des ménages surendettés), un complément monétaire de 100 € par mois permet de diminuer significativement la probabilité d’être exposé au surendettement, quelle que soit sa structure familiale. Prenons l’exemple d’un niveau de vie mensuel initial de 607 €. Pour un ménage surendetté « typique » (homme seul), la probabilité d’être surendetté vaut alors 2,95% et diminue de 0,43 pp. lorsque son niveau de vie augmente de 100 € mensuels. Pour une femme seule avec enfant, la probabilité d’être surendettée vaut initialement 3,82% et diminue de 0,55 pp. lorsque son niveau de vie mensuel diminue de 100 €. À l’inverse, pour un ménage ayant des revenus par uc plus élevés, augmenter le niveau de vie semble avoir beaucoup moins d’impact (ex : une diminution de 0,03 pp. pour une femme seule avec enfants ayant un niveau de vie mensuel de 2523 €).

On constate néanmoins des disparités entre les catégories de ménages. Ainsi, à niveau de vie initial équivalent, augmenter le niveau de vie mensuel de 100 € permet de diminuer plus fortement la probabilité d’être surendetté chez les femmes seules avec enfant(s) que pour les autres catégories de ménages. Les effets de l’augmentation du niveau de vie sont les plus faibles pour les couples sans enfant. Les écarts de ces effets entre les structures familiales deviennent néanmoins plus faibles à mesure que le niveau de vie initial augmente.

Tentons à présent de quantifier les effets d’une augmentation fixe du niveau de vie sur la probabilité d’être surendetté.

2. *De combien faut-il augmenter le niveau de vie mensuel d'un ménage composé d'une femme seule avec enfant(s) pour obtenir une probabilité d'être surendettée égale à celle d'un homme seul, à niveau de vie initial identique (exemple : pour qu'une femme seule avec enfant(s) ait la même probabilité d'être surendetté qu'un homme seul à niveau de vie fixé à 1 735 €) ?*

Il faudrait l'augmenter de 165 € par mois environ (voir annexe B4 pour le calcul) :

- Pour un niveau de vie initial égal au niveau de vie médian des ménages surendettés, cela permet une diminution de 0,45 point de pourcentage de probabilité d'être surendetté ;
 - Pour un niveau de vie initial égal au niveau de vie médian de l'ensemble des ménages, cela permet une diminution de 0,15 point de pourcentage de probabilité d'être surendetté
- Augmenter de 165 € par mois le niveau de vie d'une femme seule avec enfant(s) permet une plus forte diminution de sa probabilité d'être exposée au surendettement lorsqu'elle présente un niveau de vie initial faible.

3. *De combien faut-il augmenter le niveau de vie mensuel d'une femme seule avec enfant(s) ou d'un homme seul pour obtenir une probabilité d'être surendetté égale à celle d'un couple sans enfant, à niveaux de vie initiaux égaux ?*

- Pour une femme seule avec enfant(s), il faudrait l'augmenter d'environ 625 € :
 - Pour un niveau de vie initial égal au niveau de vie médian des ménages surendettés, cela permet une diminution de 1,26 point de pourcentage de probabilité d'être surendetté. On peut aisément observer cela dans le tableau 10 : pour une femme seule avec enfant(s), passer d'un niveau de vie mensuel 1 036 € à 1 735 € (soit une augmentation de 700 € environ) permet de passer d'une probabilité d'être surendettée de 1,95% à 0,64% (soit une diminution de 1,31 point de pourcentage), donc de s'approcher de probabilité qu'un couple sans enfant ayant un niveau de vie égal à 1 735 € mensuels soit surendetté (qui vaut 0,69%) ;
 - Pour un niveau de vie initial égal au niveau de vie médian de l'ensemble des ménages, cela permet une diminution de 0,41 point de pourcentage de la probabilité d'être surendetté.
- Pour un homme seul, il faudrait l'augmenter d'environ 460 € :
 - Pour un niveau de vie initial égal au niveau de vie médian des ménages surendettés, cela permet une diminution de 0,81 point de pourcentage de la probabilité d'être surendetté ;
 - Pour un niveau de vie initial égal au niveau de vie médian de l'ensemble des ménages, cela permet une diminution de 0,26 point de pourcentage de la probabilité d'être surendetté.

Ainsi, le niveau de vie apparaît comme un facteur explicatif déterminant du surendettement : les ménages présentant des situations financières difficiles ont plus de risques d'être surendettés que

les ménages plus aisés. La structure familiale demeure cependant un facteur significatif : à niveau de vie égal, les femmes cheffes de famille monoparentale ont une probabilité d'être surendettées plus élevée que la moyenne, tandis que les couples sans enfant présentent à l'inverse moins de risque de l'être. Pour un niveau de vie initial proche du seuil de pauvreté, une femme seule avec enfant(s) aura besoin d'environ 625 € de niveau de vie mensuel supplémentaire pour atteindre la probabilité d'endettement d'un couple sans enfant.

Un tel modèle pourrait être amélioré par l'ajout d'autres facteurs tels que l'âge, le patrimoine, la catégorie socio-professionnelle ou encore l'aspect géographique. On suppose en effet que ces variables jouent un rôle significatif dans le fait d'être exposé au surendettement. Cela nécessiterait cependant d'avoir des informations bien plus précises que celles auxquelles nous avons accès. Une deuxième perspective d'amélioration serait de prendre en compte le nombre d'enfants par ménage. On s'attend à ce que le nombre d'enfants ait un impact positif et significatif sur la probabilité d'être surendetté. La distribution du niveau de vie selon la structure familiale et le nombre d'enfants n'étant pas accessible, cette modélisation n'a pas pu être mise en œuvre. Il serait également possible d'effectuer un autre modèle de régression dont la variable à expliquer serait cette fois-ci le montant de l'endettement et voir si les résultats varient selon l'ampleur de l'endettement du ménage. Bien entendu, la précision des estimations et la qualité des résultats seraient meilleures s'il était possible d'exploiter des données individuelles empiriques au lieu de données agrégées.

Par ailleurs, une autre difficulté liée à l'analyse de la procédure de surendettement est le fait que le surendettement soit défini au sein d'un cadre juridique précis : dans notre modèle, un ménage est considéré surendetté lorsque ce dernier a déposé un dossier de surendettement auprès d'une commission. De plus, seules les dettes non professionnelles sont considérées. Or, il peut également exister d'autres ménages surendettés, n'ayant pas déposé de dossier, voire ne connaissant pas la procédure de surendettement. Une problématique intéressante à creuser serait donc de voir quelles différences sont susceptibles d'exister entre les ménages surendettés tels que définis par la Banque de France, et les ménages modestes (dont le niveau de vie est par exemple inférieur au seuil du premier décile) ainsi que les ménages n'ayant pas déposé de dossier mais présentant des dettes non professionnelles. Il aurait néanmoins fallu un travail considérable pour évaluer numériquement la population des ménages surendettés et, plus encore, pour connaître ses caractéristiques sociales, familiales, financières, etc. Cela explique le fait de s'être limité ici à la population des ménages surendettés, définie par les personnes qui ont déposé un dossier et pour laquelle nous disposons des données individuelles.

Ce premier modèle, utilisant les données concernant la structure familiale, pourrait par ailleurs être approfondi en étudiant de façon plus précise les caractéristiques de l'endettement en fonction du type de ménage : les dettes des couples sont-elles différentes de celles des individus seuls ? Les dettes d'un individu divorcé surendetté sont-elles des dettes accumulées lorsque ce dernier était encore en couple, ou sont-elles des dettes contractées après son divorce ? Ainsi, d'autres recherches

complémentaires pourraient être réalisés afin de compléter la présente étude.

Dans cette partie, la probabilité qu'un ménage français soit surendetté a été modélisée en fonction du niveau de vie et de la structure familiale. À présent, penchons-nous sur les caractéristiques des ménages les plus vulnérables au sein des ménages surendettés, soit ceux dont le dossier de surendettement a été orienté vers le rétablissement personnel (voir l'encadré 2 pour la définition des différents types d'orientations possibles). On souhaite ainsi modéliser la probabilité qu'un dossier de surendettement soit orienté vers cette solution par la commission. L'avantage majeur de cette modélisation est la disponibilité des données individuelles concernant les ménages surendettés et la présence de nombreuses variables potentiellement pertinentes.

4 Modélisation de l'orientation d'un dossier vers le rétablissement personnel

Pour un dossier de surendettement, plusieurs orientations possibles sont susceptibles d'être proposées par la commission (voir encadré 2). L'orientation proposée au ménage surendetté lorsque sa situation financière ménage est particulièrement dégradée est le rétablissement personnel. Cela consiste à effacer l'intégralité des dettes dues. Cette orientation est systématiquement proposée lorsque la capacité de remboursement du débiteur et de son éventuel codébiteur est quasi nulle, voire négative et concerne ainsi surtout des ménages en très grande difficulté financière⁴.

Ainsi, après avoir analysé l'impact du niveau de vie et de la structure familiale sur la probabilité d'être surendetté dans la partie précédente, nous nous pencherons ici sur l'orientation vers le rétablissement personnel selon différentes variables socio-démographiques et de surendettement. Seule l'orientation vers le rétablissement personnel sans liquidation judiciaire (RP sans LJ) sera modélisée, puisque le RP avec LJ concerne une infime proportion des cas de rétablissement personnel (environ 2% des cas en 2019).

Dans le cadre de ce second modèle (modélisation de la probabilité qu'un ménage surendetté soit orienté vers le RP sans LJ), une seule source de données est donc utilisée. Il s'agit des données concernant les dossiers de surendettement disponibles au sein de la Banque de France. La base de données complète la plus récente concerne les ménages ayant déposé un dossier recevable en 2019. Par conséquent, alors que pour le premier modèle, nous étions contraints à utiliser des données concernant les ménages en 2017 (dû à l'utilisation de l'Enquête Revenus Fiscaux et Sociaux), des

4. « La capacité de remboursement d'un ménage surendetté, calculée sur une base mensuelle, est la différence entre ses ressources nettes et son « budget de vie courante ». Elle est déterminée selon les modalités décrites dans la circulaire ministérielle du 22 juillet 2014 relative à la procédure de traitement des situations de surendettement des particuliers. », *Le surendettement des ménages. Enquête typologique 2019*, p.164, Banque de France

données plus récentes peuvent être exploitées dans le cadre du second modèle.

4.1 Champ de la modélisation et variables utilisées

Afin de mettre en place une modélisation de l'orientation vers le RP sans LJ, les données concernant les dossiers déposés en France métropolitaine déclarés recevables en 2019 sont utilisées, ce qui représente 135 014 observations. La variable cible est la décision ou non d'orientation vers RP sans LJ suite à commission⁵. Parmi les dossiers recevables en 2019, 60 106 ont été orientés vers le RP sans LJ suite à commission, ce qui représente 44,5% des observations.

4.1.1 Variables explicatives disponibles

On dispose de variables socio-démographiques (par exemple l'âge ou la situation professionnelle du codébiteur et de son éventuel codébiteur) et d'informations précises concernant l'endettement du ménage (montant des dettes, types de dettes, etc.) grâce aux données recueillies via les dossiers de surendettement. Comme la procédure de surendettement s'effectue par l'analyse du dossier de surendettement dans sa globalité, l'unité d'observation considérée est donc le dossier (qui équivaut à un ménage), et non le débiteur. Cela pose cependant un problème au niveau de l'exploitation de plusieurs variables socio-démographiques individuelles telles que l'âge, en particulier lorsque le dossier est déposé par un débiteur accompagné d'un codébiteur. Afin de considérer les caractéristiques du ménage plutôt que de l'individu, la variable « structure familiale » a été construite comme alternative aux variables « sexe » et « situation matrimoniale ». Concernant d'autres facteurs comme l'âge, seule l'information concernant la personne de référence du ménage⁶ est conservée. La liste des variables explicatives potentiellement exploitables dans le cadre de ce modèle est disponible en annexe C1. L'unité considérée reste néanmoins le ménage dans le cadre de ce modèle, même si certaines caractéristiques individuelles (concernant la personne de référence du ménage uniquement) sont prises en compte.

Ci-dessous, en figures 4 et 5, sont représentées la distribution du niveau de vie et de la structure familiale selon l'orientation du dossier (orientation vers le RP sans LJ ou non). Des représentations graphiques concernant d'autres variables explicatives (ressource principale du ménage, premier dépôt ou redépôt, montant de la dette globale et montant des dettes autres) sont disponibles en annexe D.

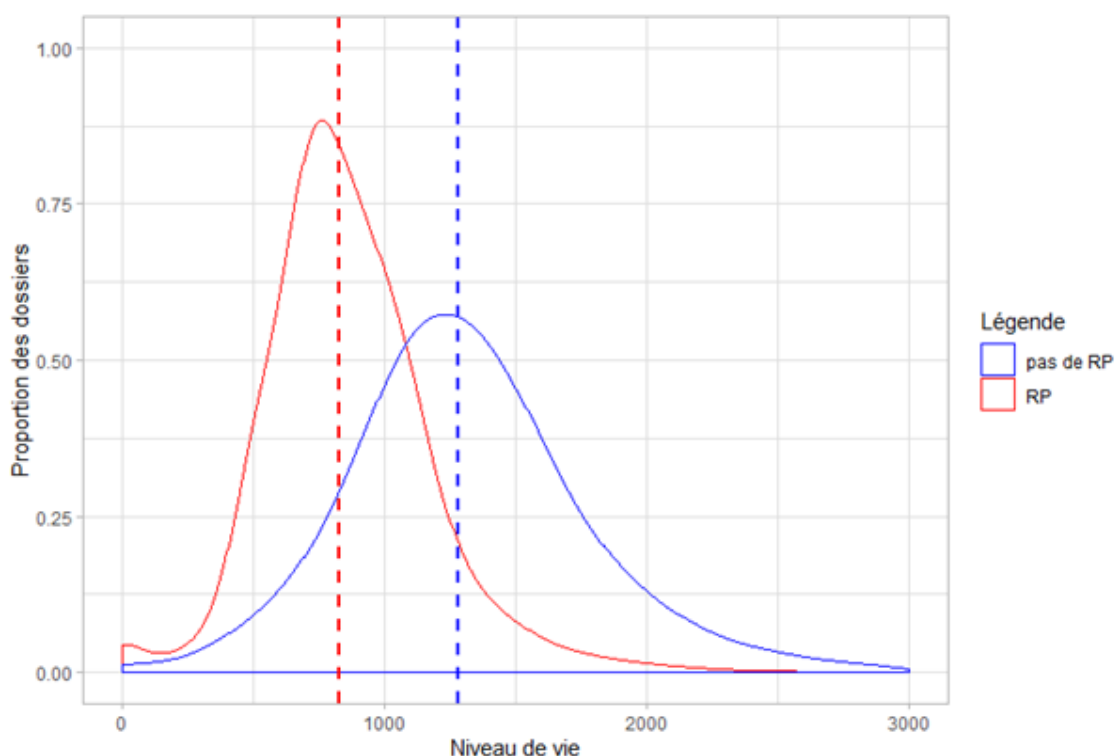
5. La proposition de la commission peut être différente de l'orientation finale du dossier, par exemple après une contestation de la proposition la part du déposant ou d'un créancier.

6. À l'exploitation principale du recensement de la population, « la personne de référence d'un ménage [...] est :
— La personne active la plus âgée ayant un conjoint ;
— À défaut de personne active ayant un conjoint, la personne la plus âgée ayant un conjoint ;
— À défaut de personne ayant un conjoint, la personne active la plus âgée ;
— À défaut de personne active, la personne la plus âgée. »

Source : Insee, <https://www.insee.fr/fr/metadonnees/definition/c1944>

Le niveau de vie semble jouer un rôle essentiel dans le fait d’être orienté vers le RP sans LJ d’après la figure 4. En effet, le niveau de vie médian des ménages orientés vers le RP sans LJ est d’environ 800 € mensuels, soit près de 500 € plus faible que le niveau de vie mensuel des ménages non orientés vers le RP sans LJ. Ainsi, tout comme les ménages surendettés ont une situation financière plus précaires que les ménages non surendettés, les ménages orientés vers le RP sans LJ sont parmi les ménages surendettés les plus pauvres.

FIGURE 4 – Distribution du niveau de vie selon l’orientation du dossier



NB : Les médianes sont représentées en pointillés.

Champ : Dossiers de surendettement déclarés recevables en 2019.

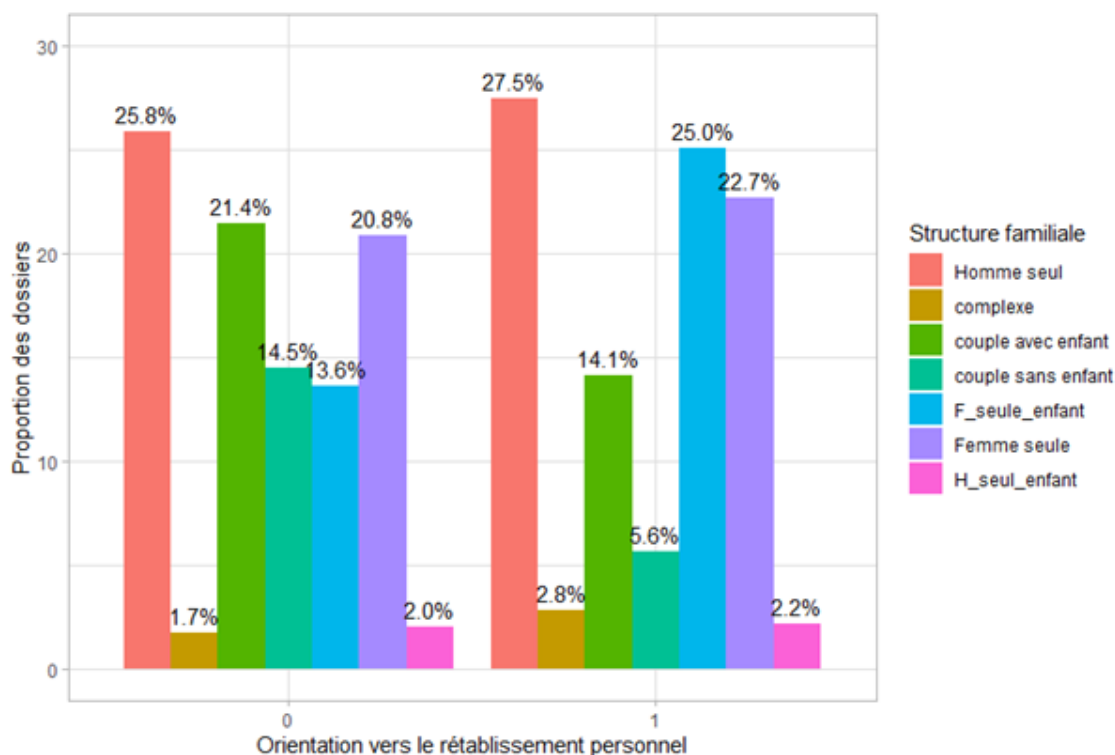
Lecture : Le niveau de vie médian des ménages dont le dossier a été orienté vers le RP sans LJ est d’environ 800 € mensuels.

L’orientation d’un dossier diffère également beaucoup selon la structure familiale du ménage déposant d’après la figure 5. Alors que les cheffes de famille monoparentale sont surreprésentées parmi les ménages surendettés par rapport aux reste des ménages, elles représentent également un quart des ménages orientés vers le RP sans LJ, contre moins d’un septième des ménages orientés vers d’autres voies. Elles sont ainsi le deuxième type de ménage majoritaire chez les ménages orientés vers le RP sans LJ, après les hommes seuls.

Les couples avec enfant(s) sont proportionnellement moins nombreux chez les ménages orientés vers le RP sans LJ (14,1%) que chez les ménages ayant une autre orientation (21,4%). Les couples sans enfant(s) sont très peu représentés parmi les ménages orientés vers le RP sans LJ (5,6%) alors

qu'ils représentent une proportion conséquente des autres ménages surendettés (14,5%). Enfin, les hommes seuls et les femmes seules sont présentés dans des proportions similaires chez les ménages orientés vers le RP sans LJ et les ménages orientés vers d'autres voies.

FIGURE 5 – Structure familiale selon l'orientation du dossier



Champ : Dossiers de surendettement déclarés recevables en 2019.

Lecture : 5,6% des ménages dont le dossier a été orienté en RP sans LJ sont des couples sans enfant

4.1.2 Première sélection de variables

Une première sélection de variables peut être mise en œuvre par l'analyse des colinéarités et des redondances entre les variables explicatives. En effet, la présence de variables colinéaires entraîne un biais dans l'estimation des coefficients de la régression. Ainsi, parmi les variables qualitatives disponibles (voir annexe C1), on supprime :

- « Montant des dettes immobilières » et « situation vis-à-vis du logement » : ces variables sont en effet très liées au montant du patrimoine ;
- « Cause licenciement/chômage » car la variable situation professionnelle permet déjà de savoir si la personne de référence est au chômage ou non ;
- « CSP » car elle est redondante avec la situation professionnelle : les deux variables présentent par exemple toutes les deux une catégorie « sans activité professionnelle ». De plus, la CSP comporte plusieurs modalités rares, dont une modalité peu interprétable (la catégorie « Agriculteurs exploitants, artisan, commerçant, chef d'entreprise »). La variable

- « situation professionnelle » semble plus interprétable que la variable « CSP » ;
- « Nombre de personnes à charge » et « présence d'un codébiteur » car ces variables sont redondantes avec la structure familiale.

Concernant les variables quantitatives, une matrice des corrélations peut être construite (cf. tableau D1.1 en annexe D1). Celle-ci permet de se rendre compte d'une forte corrélation positive entre la capacité de remboursement et le niveau de vie (0,71), et entre le montant global de la dette et le montant des autres dettes (0,66). On décide de garder le niveau de vie car il permet de comparer les ressources de ménages de différentes tailles. De plus, la capacité de remboursement est endogène à la procédure de surendettement : une capacité de remboursement négative entraîne systématiquement l'orientation du dossier vers le RP sans LJ. Par conséquent, on retire cette variable. On conserve la variable « dette globale » plutôt que le montant des dettes autres car les dettes autres concernent des montants très faibles et semble être une variable moins discriminante que la dette globale (cf. figures D0.3 et D0.4 en annexe D). 15 variables sont ainsi conservées à l'issue de cette première étape.

4.2 Choix du modèle

Nous avons toujours à notre disposition un nombre important de variables : cela permet un meilleur pouvoir explicatif, mais réduit la précision des coefficients à cause de l'estimation de nombreux paramètres. Afin d'améliorer la précision des estimateurs, nous pouvons procéder à une sélection de modèles selon le critère *AIC* ou le critère *BIC*. Cela permet de sélectionner le modèle qui s'approche le plus du « vrai modèle ».

4.2.1 Utilisation de critères d'information pour la sélection du modèle final

Le tableau 12 ci-dessous présente l'*AIC* et le *BIC* calculés pour le modèle minimisant l'*AIC* (modèle AIC) et pour celui minimisant le *BIC* (modèle BIC). Aucun des deux modèles ne permet à lui seul de minimiser les deux critères d'information. Cependant, le *BIC* permettant un meilleur choix de modèle lorsque le nombre d'observations est élevé, on choisit de garder le modèle minimisant le *BIC*.

TABLE 12 – *AIC* et *BIC* des modèles AIC et BIC

	AIC	BIC
Modèle AIC	110 032	110 277
Modèle BIC	110 039	110 265

Les 12 variables finalement conservées dans le modèle BIC sont :

- Premier dépôt ;

- Accompagnement par un travailleur social ;
- Poids des dettes de charges courantes ;
- Montant des dettes financières non immobilières ;
- Patrimoine global ;
- Situation professionnelle de la personne de référence ;
- Âge de la personne de référence ;
- Structure familiale ;
- Niveau de vie ;
- Ressource principale ;
- Cause de surendettement : séparation/divorce ;
- Cause de surendettement : montant élevé des dettes et charges.

Les variables retirées par la procédure de sélection de modèle selon le critère *BIC* sont :

- Dette globale ;
- Nombre de dettes ;
- Cause de surendettement : maladie/accident.

Ces trois dernières variables sont également celles qui n'apparaissent pas ou peu significatives à 1% dans le modèle implémenté sur les 15 variables initialement présélectionnées. Ainsi, la sélection du modèle par le critère *BIC* permet d'améliorer la précision des estimations, tout en conservant un modèle pertinent.

4.2.2 Performance du modèle sélectionné

La performance du modèle final est étudiée via une matrice de confusion. La matrice de confusion est construite en utilisant un seuil d'affectation égal à 0,5 :

- Si la probabilité estimée par le modèle qu'un dossier soit orienté en RP est supérieure à 0,5 on considère que son orientation vers le RP sans LJ est prédite par le modèle ;
- Sinon, on considère que le modèle prédit que le dossier ne sera pas orienté vers le RP sans LJ.

TABLE 13 – Matrice de confusion du modèle minimisant le BIC

	Réels non surendettés	Réels surendettés
Prédits non surendettés	82,1 %	17,9 %
Prédits surendettés	17,9 %	82,1 %
	100 %	100 %

La spécificité et la sensibilité du modèle final valent chacune 82,1%. Aussi, dans plus de 80% des cas, les décisions d'orientation sont correctement prédites par le modèle. Notre modèle présente ainsi de bonnes performances, ayant des taux de détection élevés à la fois parmi les dossiers orientés vers le RP sans LJ, et parmi les dossiers qui ne sont pas orientés vers le RP sans LJ.

4.3 Résultats principaux du modèle final

Le modèle final présente ainsi des performances satisfaisantes. Il s'agit à présent d'analyser les principaux résultats de ce modèle, à travers l'analyse des coefficients des variables et de leurs effets marginaux.

4.3.1 Analyse des coefficients du modèle final

Les informations concernant les coefficients du modèle final sont récapitulés dans le tableau D3.1 en annexe D3. Toutes les variables quantitatives du modèle sont significatives. Parmi les variables qualitatives, la variable « structure familiale » est la seule qui présente une modalité non significative au seuil de 1%. En effet, la modalité « complexe » de cette variable présente une p-valeur très élevée. Cette modalité concerne un effectif très faible de la base de données, ce qui explique cette non significativité.

4.3.2 Effets marginaux moyens

Les effets marginaux moyens peuvent permettre de quantifier l'impact de chaque variable explicative sur la probabilité d'être orienté vers le rétablissement personnel sans liquidation judiciaire. Le tableau 14 ci-contre présente les effets marginaux moyens de chaque prédicteur du modèle.

Prédicteurs	Effets marginaux moyens (en pp.)	IC 95% inf.	IC 95% sup.
<i>Premier dépôt</i>			
Non	Réf.	Réf.	Réf.
Oui	-4,02	-4,42	-3,63
<i>Aide d'un travailleur social</i>			
Non	Réf.	Réf.	Réf.
Oui	0,85	0,45	1,25
<i>Poids des dettes de charges courantes</i>			
	4,19	3,53	4,85
<i>Montant dettes financières non immobilières</i>			
	-0,0001	-0,0001	-0,0001
<i>Patrimoine global</i>			
<30 000 euros	Réf.	Réf.	Réf.
≥ 30 000 euros	-49,91	-50,17	-49,65
<i>Situation professionnelle</i>			
CDI/libéral/artisan/commerçant	Réf.	Réf.	Réf.
Autre	6,95	6,09	7,81
CDD/intérim	4,91	4,08	5,74

Chômage	11,67	11,00	12,35
Retraité	3,19	2,14	4,23
Autre sans activité professionnelle	16,53	15,48	17,57
<i>Âge</i>	0,13	0,11	0,15
<i>Structure familiale</i>			
Homme seul	Réf.	Réf.	Réf.
Complexe	0,39	-0,97	1,75
Couple avec enfant(s)	-3,43	-4,09	-2,78
Couple sans enfant	-6,86	-7,66	-6,05
Femme seule avec enfant(s)	2,17	1,54	2,80
Femme seule	1,29	0,72	1,86
Homme seul avec enfant(s)	-2,23	-3,65	-0,81
<i>Niveau de vie</i>	-0,04	-0,04	-0,04
<i>Ressource principale</i>			
Revenus d'activité	Réf.	Réf.	Réf.
Autre	8,23	6,88	9,58
Revenus de transfert	7,26	6,67	7,84
<i>Cause : séparation/divorce</i>			
Non	Réf.	Réf.	Réf.
Oui	3,62	3,08	4,15
<i>Cause : poids élevé des dettes et charges</i>			
Non	Réf.	Réf.	Réf.
Oui	-6,53	-7,14	-5,93

TABLE 14 – Effets marginaux moyens du modèle BIC

Lecture : En moyenne, le fait que la personne de référence du ménage soit au chômage entraîne une augmentation de 11,67 pp. de la probabilité d'être orienté en RP sans LJ par rapport à un ménage dont la personne de référence est en CDI/libéral/artisan/commerçant.

Concernant les variables qualitatives, on observe les effets marginaux les plus élevés pour les variables « situation professionnelle de la personne de référence » et « ressource principale ». Ainsi, en moyenne, le fait que la personne de référence soit au chômage entraîne une augmentation de 12 points de pourcentage (pp) de la probabilité d'être orienté en RP sans LJ par rapport à un ménage dont la personne de référence est en CDI ou exerce une profession libérale. En moyenne également, le fait que les revenus de transfert soient la ressource principale du ménage entraîne une augmentation de 7 pp de la probabilité d'être orienté en RP sans LJ par rapport à un ménage dont les revenus d'activité représentent la ressource principale. Le patrimoine global influe également de façon très conséquente sur l'orientation vers le RP sans LJ, avec un effet marginal proche de

50 pp de probabilité, mais cela est dû en grande partie au fait que le rétablissement personnel sans liquidation judiciaire concerne les dossiers sans bien immobilier, donc présentant la plupart du temps un faible patrimoine. À l’opposé, le fait d’être accompagné par un travailleur social par exemple a un impact assez faible sur la probabilité d’être orienté en RP sans LJ, d’après le modèle : l’effet marginal moyen est en effet inférieur à 1 pp.

Parmi les variables quantitatives présentes dans le modèle, le niveau de vie présente l’effet marginal le plus conséquent : une augmentation relativement faible du niveau de vie (100 €) entraîne une diminution plutôt importante de la probabilité d’être orienté en RP sans LJ par la commission (4 pp). À l’inverse, le montant des dettes financières non immobilières ou l’âge par exemple ont des effets assez faibles sur l’orientation vers le RP sans LJ. En effet, en moyenne, une augmentation de 10 ans de l’âge de la personne de référence du ménage entraîne une augmentation de 1,3 pp de la probabilité d’être orienté en RP sans LJ, et l’effet marginal moyen correspondant au montant des dettes financières non immobilières d’environ 1 pp pour 10 000 €.

L’orientation vers le rétablissement personnel sans liquidation judiciaire concerne ainsi surtout des ménages aux revenus très modestes. Il s’agit surtout de personnes au chômage, dont les ressources financières sont aussi essentiellement composées de revenus de transfert. La probabilité d’être orienté vers le RP sans LJ est également croissante avec l’âge et est plus élevée pour les familles monoparentales et les individus seuls. Aussi, si ces ménages sont caractérisés par des ressources financières très limitées, ils présentent également des montants de dettes très peu élevés. Ainsi, le rétablissement personnel ou l’effacement total des dettes concerne des ménages dont l’endettement est également le plus faible parmi les ménages surendettés : seule 26,6% de la dette globale des situations closes au cours de l’année 2019 a été effacée [5], alors que le RP sans LJ concernait 44,5% des cas de surendettement.

D’autres variables intéressantes pourraient être ajoutées à ce modèle. Par exemple, plusieurs éléments de la procédure de surendettement sont aujourd’hui automatisés, notamment le calcul de quantités telles que le reste à vivre du ménage (montant minimum permettant au débiteur de faire face aux charges courantes essentielles). L’ajout de ces quantités précalculées à notre modèle permettrait de complexifier ce dernier. Un autre aspect absent de notre modèle est l’analyse géographique : une problématique intéressante à étudier serait en effet les différences potentiellement existantes entre les différentes zones géographiques. Les écarts constatés entre les commissions, concernant l’orientation des dossiers de surendettement, sont-ils uniquement dus à des différences individuelles, ou existe-t-il des causes structurelles ? Une possibilité serait de considérer un modèle multiniveaux permettant d’analyser et quantifier la part de la variance totale liée à des différences inter-groupes (ici, inter-commissions) et celle liée à de la variance entre individus.

Ce modèle pourrait également être enrichi par la mise en place d’une étude longitudinale afin d’ana-

lyser les éventuelles évolutions au fil des années. Par ailleurs, un modèle multinomial permettrait d'analyser de façon plus générale l'orientation du dossier de surendettement. En effet, ce second modèle se concentre uniquement sur l'orientation vers le rétablissement personnel sans considérer les autres orientations possibles (mesures imposées, plan de redressement, et non recevabilité notamment).

5 Conclusion

La première modélisation a permis de mettre en évidence le niveau de vie en tant que facteur explicatif déterminant du surendettement : les ménages présentant des situations financières difficiles ont plus de risques d'être surendettés que les ménages plus aisés. La structure familiale demeure cependant un facteur significatif : à niveau de vie égal, les femmes cheffes de famille monoparentale ont une probabilité plus élevée que la moyenne d'être surendettées, tandis que les couples sans enfant présentent à l'inverse moins de risque de l'être. Pour un niveau de vie initial proche du seuil de pauvreté, une femme seule avec enfant(s) aura besoin d'environ 625 € de niveau de vie mensuel supplémentaire pour atteindre la probabilité d'endettement d'un couple sans enfant. Une faible augmentation des ressources semble avoir un impact négligeable sur la probabilité qu'un ménage aisé soit surendetté. En revanche, pour un ménage financièrement fragile, cela semble permettre une diminution significative de la probabilité d'être exposé à cette situation.

Plusieurs perspectives d'amélioration peuvent être envisagée. Le modèle utilisé dans le cadre de cette étude pourrait être amélioré par l'ajout d'autres facteurs tels que l'âge, la catégorie socio-professionnelle ou encore l'aspect géographique. On suppose en effet que ces variables jouent un rôle significatif dans le fait d'être exposé au surendettement. Cela nécessiterait cependant d'avoir des informations bien plus précises que celles auxquelles nous avons accès. Une deuxième perspective d'amélioration serait de prendre en compte le nombre d'enfants par ménage. On s'attend à ce que le nombre d'enfants ait un impact positif et significatif sur la probabilité d'être surendetté. La distribution du niveau de vie selon la structure familiale et le nombre d'enfants n'étant pas accessible, cette modélisation n'a pas pu être mise en œuvre. Enfin, des analyses complémentaires plus détaillées, notamment concernant les trajectoires des ménages par structure familiale pourraient permettre de remonter aux origines du surendettement pour différents types de ménages.

Ce premier modèle était surtout limité par l'absence de données individuelles. Le second modèle, se penchant sur uniquement sur la population des ménages surendettés, a permis d'utiliser l'ensemble des données individuelles disponibles concernant ces derniers. Ce modèle a mis en évidence des facteurs essentiels jouant un rôle dans le fait d'être orienté vers le rétablissement personnel, orientation proposée aux ménages surendettés les plus vulnérables financièrement. Ainsi, la situation professionnelle, le niveau de vie et la structure familiale sont par exemple des variables

qui jouent significativement sur l'orientation vers le rétablissement personnel. En fin de compte, les ménages dont les dettes sont effacées sont également ceux dont les niveaux d'endettement font partie des plus faibles.

Cependant, ce second modèle a été implémenté en utilisant les données concernant les ménages en 2019. Il serait alors intéressant de mettre en oeuvre une étude longitudinale, en prenant en compte plusieurs années antérieures afin de mettre en évidence l'évolution des caractéristiques des ménages surendettés orientés vers le rétablissement personnel. Enfin, l'ajout de l'aspect géographique serait également une perspective d'amélioration permettant d'enrichir l'étude.

Références

- [1] Banque de France, « Le surendettement en bref ».
<https://particuliers.banque-france.fr/surendettement/comprendre-le-surendettement/le-surendettement-en-bref>.
- [2] Berardi N. et al., « Le surendettement et les femmes », *Bulletin de la Banque de France*, Juillet-Août 2019.
https://publications.banque-france.fr/sites/default/files/medias/documents/bdf224-3_surendettement_vf.pdf.
- [3] Fraisse H., Muller A., « Les commissions de surendettement des ménages : de l'objectif de négociation à la prévention de la rechute », *Économie et statistique*, n°443, 2011.
https://www.persee.fr/doc/estat_0336-1454_2011_num_443_1_9622.
- [4] Insee, « Revenu, niveau de vie et pauvreté en 2017 », *Enquête Revenus Fiscaux et Sociaux (ERFS) - Insee Résultats*, 2017.
<https://www.insee.fr/fr/statistiques/4261132>.
- [5] Banque de France, *Le surendettement des ménages. Enquête typologique 2019*, 2020.
https://particuliers.banque-france.fr/sites/default/files/media/2020/02/06/suren2019_web_0.pdf.
- [6] Banque de France, *Le surendettement des ménages. Enquête typologique 2017*, 2018.
https://particuliers.banque-france.fr/sites/default/files/media/2020/02/06/suren2019_web_0.pdf.
- [7] « Les revenus et le patrimoine des ménages », *Insee Références*, Édition 2018, p.115.
<https://www.insee.fr/fr/statistiques/fichier/3549502/REVPME18.pdf>.
- [8] Fournier J., *Generalized Pareto curves : Theory and application using income and inheritance tabulation for France 1901-2012*, Septembre 2015.
<http://piketty.pse.ens.fr/files/Fournier2015.pdf>.

Annexes

A Annexes de la partie 1

A1 Simulation de données par Monte-Carlo

Afin de simuler des données dont la distribution est non usuelle, une méthode possible consiste à utiliser l'inverse de la fonction de répartition de la loi de probabilité en question. Il s'agit d'une des méthodes de simulation Monte-Carlo.

Soit f définie sur l'intervalle $[a, b]$ la fonction de densité de la loi selon laquelle on souhaite générer un jeu de n données. Soit F inversible sa fonction de répartition. On commence par générer un ensemble U de n observations issues d'une loi uniforme sur $[0, 1]$. Pour tout $u \in U$, il existe un unique $x \in [a, b]$ tel que :

$$\begin{aligned} u &= F(x) \\ \iff x &= F^{-1}(u) \end{aligned}$$

Ainsi, en appliquant cette fonction aux n observations de l'échantillon U , on génère un jeu de données X de taille n qui suit une loi dont la fonction de densité est f .

Dans le cadre de cette étude, la fonction de densité utilisée est :

$$f(x) = \begin{cases} ax^2 + b & \forall x \in \left[\sqrt{-\frac{b}{a}}, q_{10} \right] \\ 0 & \forall x \in \left[0, \sqrt{-\frac{b}{a}} \right] \end{cases} \quad \text{si } b < 0 \text{ avec } a > 0$$
$$f(x) = ax^2 + b \quad \forall x \in [0, q_{10}] \quad \text{si } b \geq 0 \text{ avec } a > 0.$$

avec q_{10} le seuil du premier décile de niveau de vie de la catégorie de ménage considérée.

A2 La loi de Pareto

La loi de Pareto est souvent utilisée afin d'étudier la structure des très hauts revenus. Soit X une variable aléatoire suivant une loi de Pareto de paramètres (α, k) avec $k > 0$ et $\alpha > 0$. Sa fonction de répartition s'écrit :

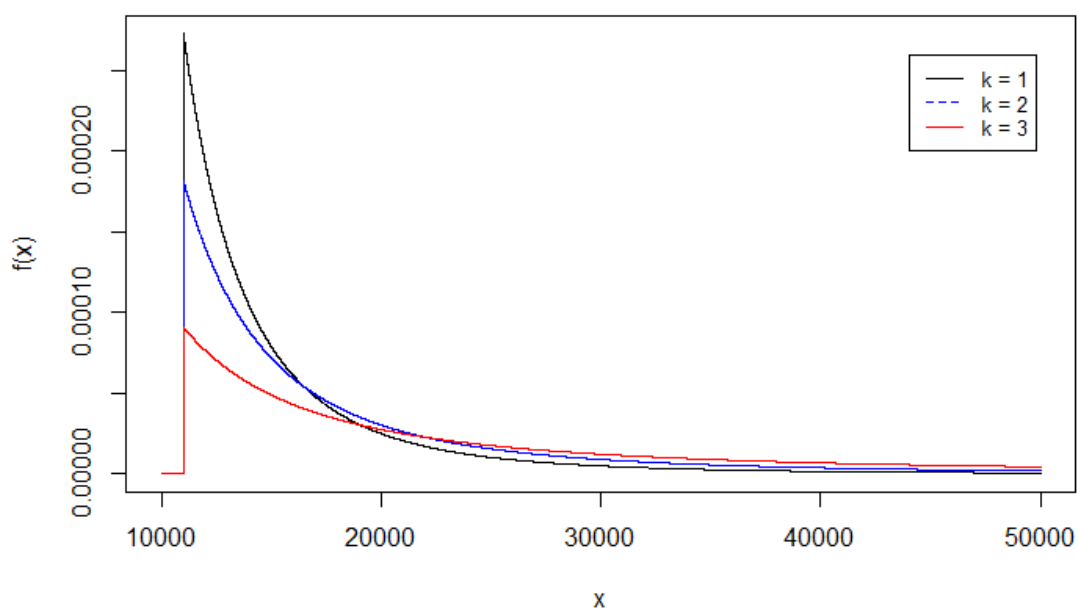
$$P(X \leq x) = 1 - \left(\frac{\alpha^k}{x} \right) \text{ avec } x \geq \alpha$$

Sa densité quant à elle s'écrit, pour tout $x \geq \alpha$: $f(x) = \frac{k\alpha^k}{x^{k+1}}$.

L'espérance associée à cette loi est : $E(X) = \frac{k\alpha}{k-1}$.

Afin de visualiser l'allure de cette distribution dans le cadre d'une simulation des hauts niveaux de vie, représente ci-dessous en figure A2.1 la fonction de densité de la loi de Pareto, pour $\alpha = 11\ 000$ et pour plusieurs valeurs possibles de k :

FIGURE A2.1 – Fonction de densité de la loi de Pareto pour plusieurs combinaisons de paramètres



A3 Structure de la base de données finale

TABLE A3.1 – Structure de l'échantillon de travail

	Population totale		Non surendettés		Surendettés	
	<i>Structure familiale</i>					
	Effectifs	% de l'effectif total	Effectifs	% de l'effectif total	Effectifs	% de l'effectif total
Femme seule avec enfant(s)	418 922	7,4	412 566	7,3	6 356	19,0
Homme seul avec enfant(s)	93 858	1,6	93 152	1,7	706	2,1
Femme seule	1 165 757	20,5	1 159 031	20,6	6 726	20,1
Homme seul	894 347	15,7	855 871	15,2	8 476	25,4
Couple avec enfant	1 476 566	25,9	1 469 651	26,1	6 915	20,7
Couple sans enfant	1 492 376	26,2	1 488 909	26,4	3 467	10,4
Famille complexe	153 367	2,7	152 595	2,7	772	2,3
	<i>Niveau de vie mensuel (en €)</i>					
Q1		1 267		1 272		779
Médiane		1 727		1 731		1 040
Q3		2 332		2 336		1 348
	<i>Surendetté/Non surendetté</i>					
	Effectifs	% de l'effectif total	Effectifs	% de l'effectif total	Effectifs	% de l'effectif total
Non surendetté	5 695 193	99,4	5 695 193	100	0	0
Surendetté	33 368	0,6	0	0	33 368	100

A4 Coefficients associés au modèle implémenté sur différents types d'échantillons

Les tableaux A4.1 et A4.2 ci-dessous présentent les coefficients associés au modèle logistique de la probabilité d'être surendetté selon le niveau de vie et la structure familiale, respectivement :

- en utilisant un échantillon représentatif de 50% de la base de données initiale;
- en utilisant un échantillon comprenant uniquement les ménages dont le niveau de vie (simulé si le ménage n'est pas surendetté, réel sinon) est inférieur à 3 000 € par mois. Cela permet de retirer 3,5 millions d'observations sur les 28,5 millions de la base initiale, dont moins d'1% des observations surendettées.

On remarque que ces coefficients sont très proches de ceux obtenus sur un échantillon représentatif de 20% de la population initiale.

TABLE A4.1 – Coefficients estimés sur un échantillon représentatif de 50% de la population initiale

Variable	Coefficient	Écart-type	P valeur
Constante	-2,524	0,010	<0,01
Complexe	-0,712	0,024	<0,01
Couple avec enfant	-0,361	0,010	<0,01
Couple sans enfant	-0,756	0,013	<0,01
Femme seule avec enfant(s)	0,272	0,011	<0,01
Femme seule	-0,430	0,010	<0,01
Homme seul avec enfant(s)	-0,449	0,025	<0,01
Niveau de vie	-0,002	0,00001	<0,01

TABLE A4.2 – Coefficients estimés sur un échantillon stratifié

Variable	Coefficient	Écart-type	P valeur
Constante	-2,499	0,007	<0,01
Ménage complexe	-0,676	0,017	<0,01
Couple avec enfant(s)	-0,347	0,007	<0,01
Couple sans enfant	-0,742	0,009	<0,01
Femme seule avec enfant(s)	0,284	0,008	<0,01
Femme seule	-0,421	0,007	<0,01
Homme seul avec enfant(s)	-0,442	0,018	<0,01
Niveau de vie	-0,002	0,00001	<0,01

B Annexes de la partie 2

B1 Odds ratios

L'odd ratio est défini comme le rapport des cotes d'un événement entre groupes d'individus. Soient p_1 et p_2 les probabilités respectives que l'événement A arrive à un individu du groupe 1 et à un individu du groupe 2. Soient $1 - p_1$ et $1 - p_2$ les probabilités respectives que l'événement complémentaire A^C arrive à un individu du groupe 1 et à un individu du groupe 2. L'odd ratio de l'événement A sur l'événement A^C s'écrit alors :

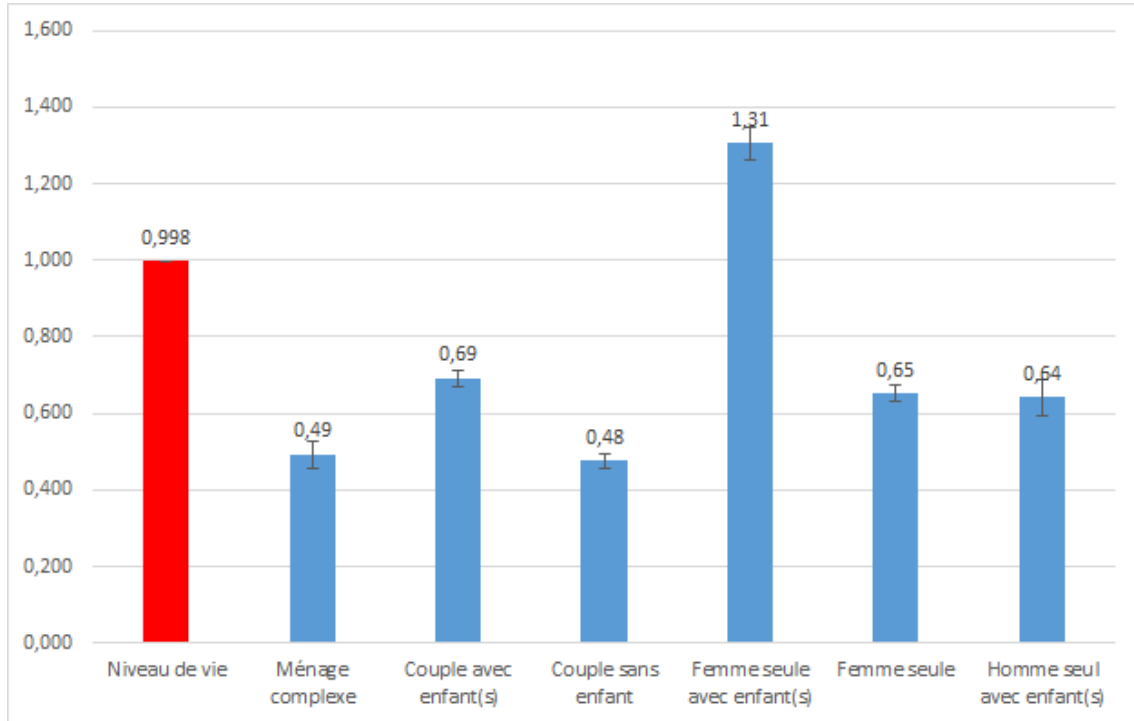
$$OR_{1/2} = \frac{p_1(1 - p_1)}{p_2(1 - p_2)}$$

L'odd ratio peut ainsi traduire l'effet d'une variable sur la probabilité qu'un événement A adienne :

- si $OR_{1/2} < 1$, alors l'événement A est moins fréquent dans le groupe 1 que dans le groupe 2;
- si $OR_{1/2} = 1$, alors l'événement A est indépendant du groupe ;

— si $OR_{1/2} > 1$, alors l'événement A est plus fréquent dans le groupe 1 que dans le groupe 2.

FIGURE B1.1 – Odds ratios associés au modèle utilisant la structure familiale et le niveau de vie



NB : Les intervalles de confiance représentés sont ceux à 95%.

B2 Le test de Wald

Le test de Wald est, avec le test du score et le test du maximum de vraisemblance, l'un des trois tests liés au maximum de vraisemblance. Il permet de tester une hypothèse du type :

$$H_0 : R\beta_0 = 0 \text{ contre } H_1 : R\beta_0 \neq 0$$

où R est une matrice $p \times k$, $p \leq k$, avec p le nombre de prédicteurs du modèle.

La statistique de test correspondante est :

$$\xi_n^W = n\hat{\beta}'R' \left[R\widehat{I_1}(\beta_0)^{-1}R' \right]^{-1} R\hat{\beta}$$

où $\widehat{I_1}(\beta_0)$ est l'estimateur convergent de l'information de Fisher calculée en β_0 .

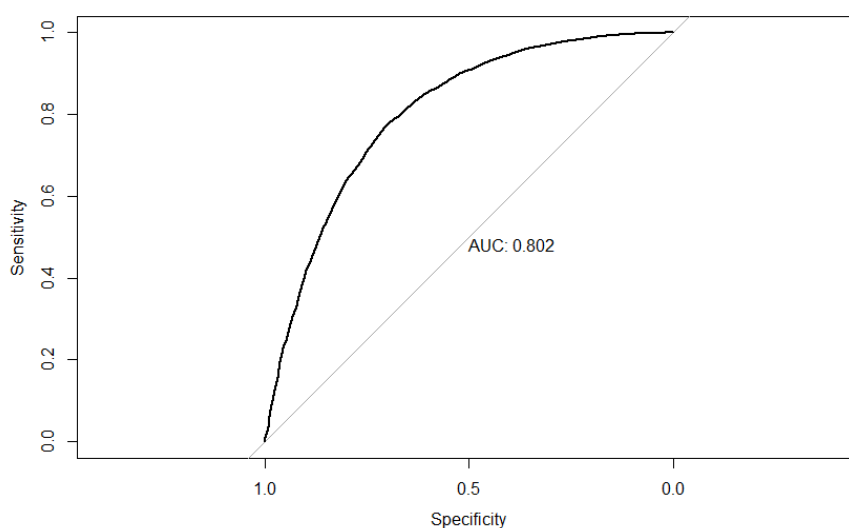
On souhaite tester sur le modèle utilisant la structure familiale et le niveau de vie l'hypothèse de nullité jointe des coefficients via le test de Wald. Les résultats du test sont résumés dans le tableau B2.1 ci-dessous :

Statistique de test	Degrés de liberté	P Valeur
30 449,2	7	<0,01

TABLE B2.1 – Résultat du test de Wald du modèle utilisant la structure familiale et le niveau de vie

B3 La courbe ROC

FIGURE B3.1 – Courbe ROC associé au modèle utilisant le niveau de vie et la structure familiale



B4 Compensation monétaire et probabilité d'être surendetté

Soient le vecteur de variables explicatives $X = (m, N)$, où m est le type de ménage, et N le niveau de vie mensuel. On souhaite calculer, pour le vecteur de variables explicatives $X = (m, N)$ donné, la compensation de niveau de vie mensuel en euros N_c nécessaire à un ménage m pour atteindre une certaine probabilité p d'être surendetté. Or, dans le cadre d'un modèle logistique, on a l'équation suivante :

$$P(Y = Surendette|X = (m, N)) = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 N + \beta_m)\}}$$

On souhaite avoir l'égalité suivante :

$$\frac{1}{1 + \exp\{-(\beta_0 + \beta_1(N + N_c) + \beta_m)\}} = p$$

Par conséquent, on en déduit :

$$N_c = -\beta_1^{-1} \frac{\log(p^{-1} - 1)}{\exp\{-(\beta_0 + \beta_1 N + \beta_m)\}}$$

C Annexes de la partie 3

C1 Liste des variables présélectionnées

TABLE C1.1 – Liste des variables potentiellement exploitables pour la modélisation de l’orientation vers le RP sans LJ

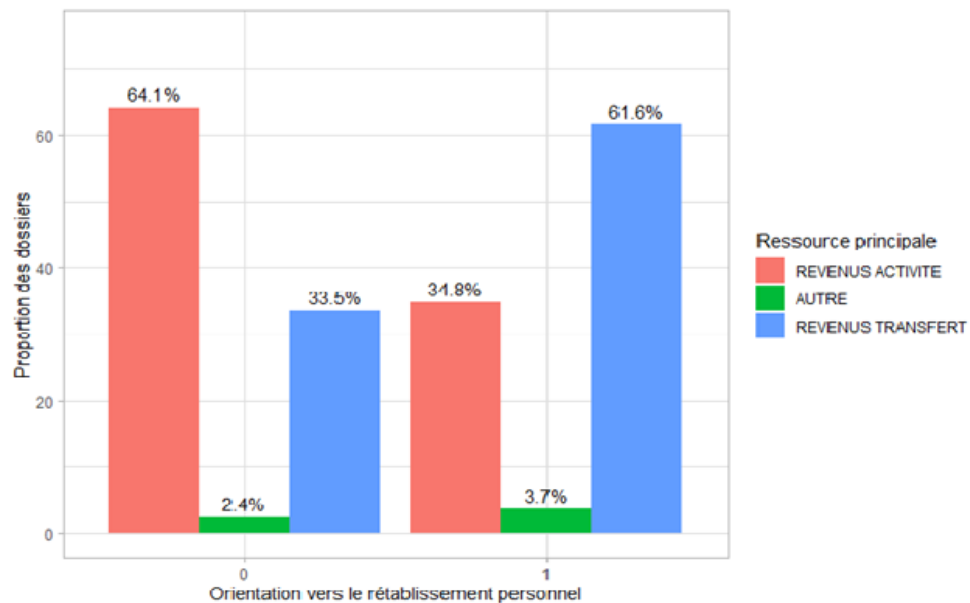
Variabiles	Type
Variabiles de surendettement	
montant global des dettes	numérique
nombre de dettes	catégorielle
s’agit-il d’un premier dépôt ?	catégorielle
montant dettes financières non immobilières	numérique
montant des dettes immobilières	catégorielle
poids des dettes de charges courantes	numérique
montant des dettes autres	numérique
accompagnement par un travailleur social	catégorielle
présence d’un codebiteur	catégorielle
niveau élevé des dettes et charges**	catégorielle
cause : séparation/divorce**	catégorielle
cause : maladie/accident/invalidité**	catégorielle
cause : licenciement/chômage**	catégorielle
capacité de remboursement	catégorielle
Variabiles socio-démographiques	
âge*	numérique
nombre de personnes à charge	catégorielle
niveau de vie	numérique
montant du patrimoine du dossier	catégorielle
ressource principale du dossier	catégorielle
situation vis-à-vis du logement	catégorielle
situation professionnelle*	catégorielle
catégorie socio-professionnelle*	catégorielle
structure familiale	catégorielle

* De la personne de référence du ménage

** Il s’agit de la « cause » de surendettement renseignée au moment du dépôt du dossier. C’est information est donc déclarative. Plusieurs causes peuvent être citées pour un même dossier.

D Représentations graphiques de quelques variables explicatives

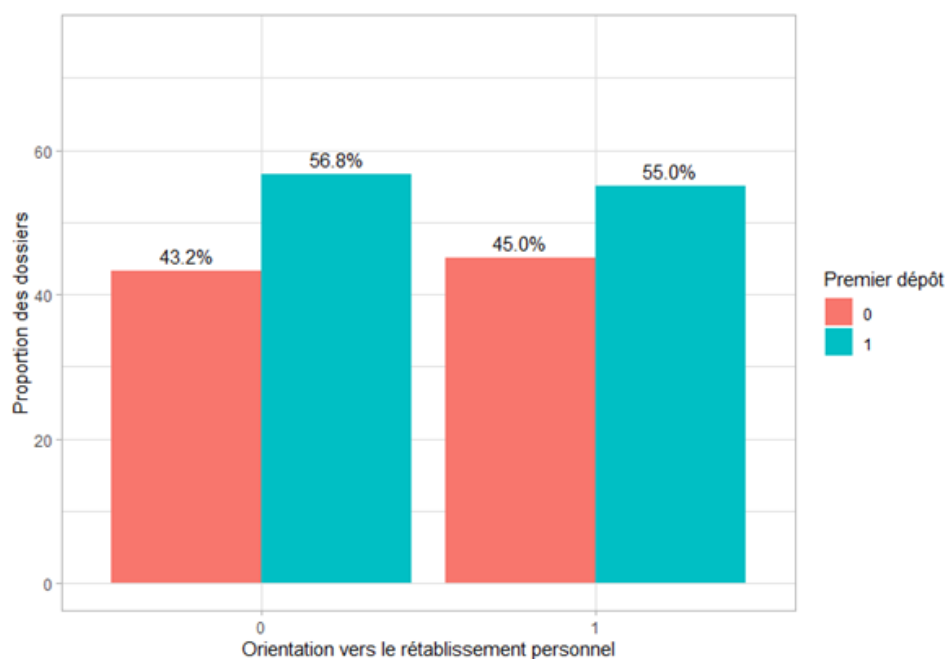
FIGURE D0.1 – Ressource principale du ménage selon l'orientation du dossier



Champ : Dossiers de surendettement déclarés recevables en 2019.

Lecture : Les revenus d'activités représentent la ressource principale de 64,1% des dossiers qui n'ont pas été orientés vers le RP sans LJ.

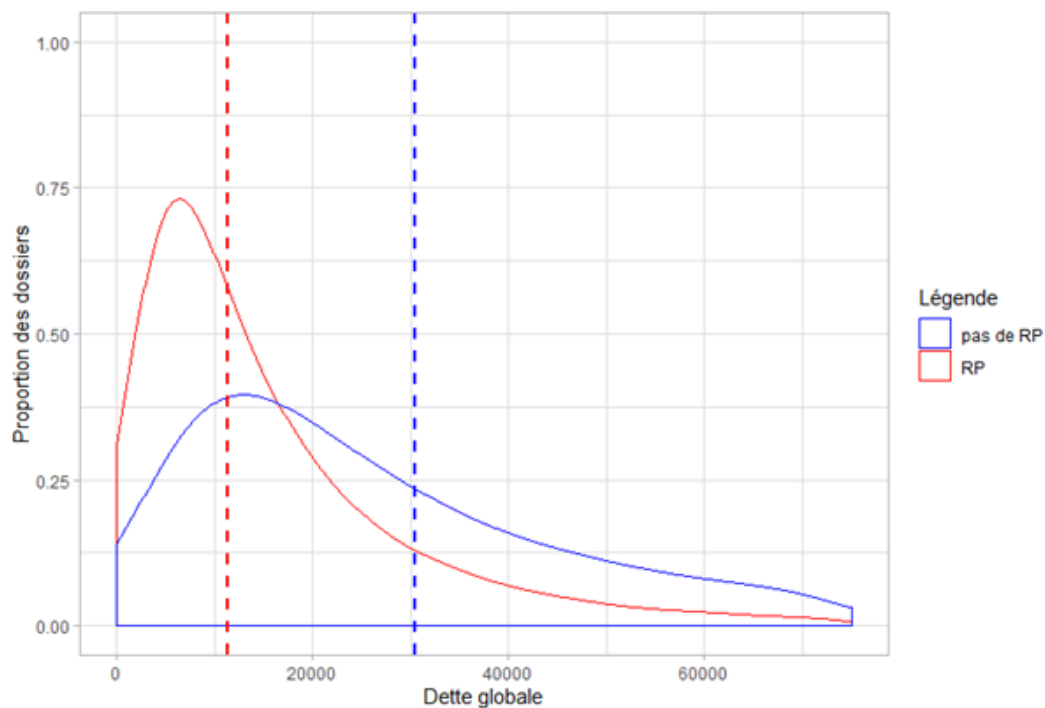
FIGURE D0.2 – Premier dépôt selon l'orientation du dossier



Champ : Dossiers de surendettement déclarés recevables en 2019.

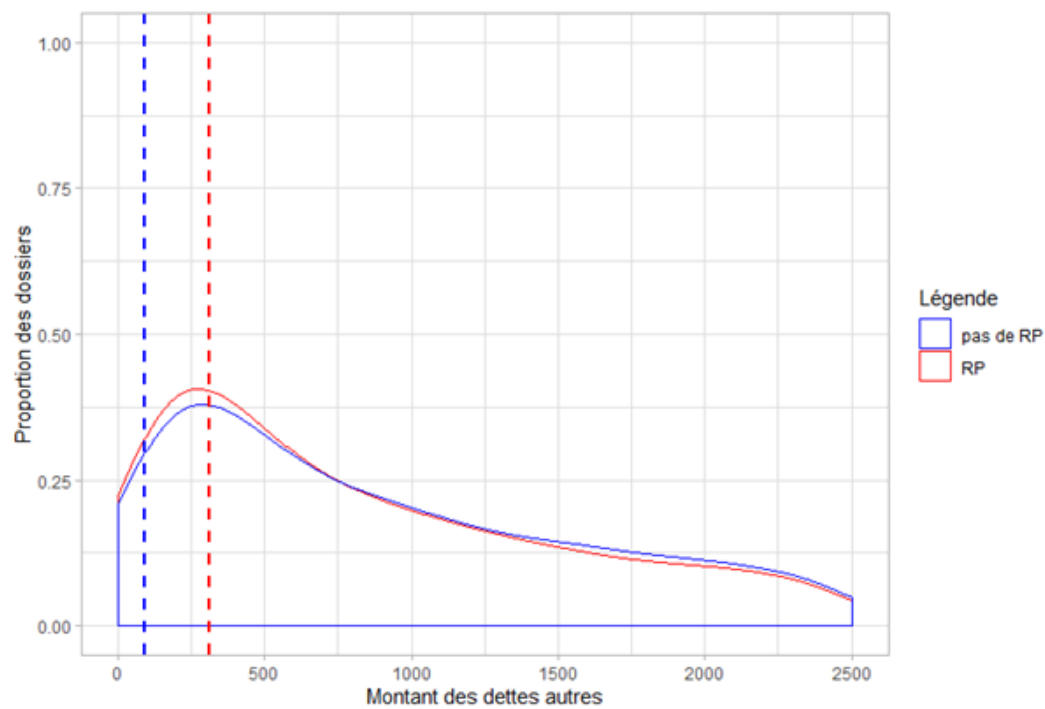
Lecture : 55% des dossiers orientés vers le RP sans LJ sont des premiers dépôts.

FIGURE D0.3 – Dette globale selon l'orientation du dossier



Champ : Dossiers de surendettement déclarés recevables en 2019.

FIGURE D0.4 – Montant des dettes autres selon l'orientation du dossier



Champ : Dossiers de surendettement déclarés recevables en 2019.

D1 Matrice des corrélations entre les variables explicatives quantitatives

TABLE D1.1 – Matrice des corrélations

	CAR	POIDS_ DETTE_CC	NB_DETTEES	DETTE_ GLOBALE	MONT_ DETTE_ FIN_ NOIMMO	MONT_ DETTE_ AUTRE	AGE	NIV_VIE
CAR	1,00	-0,33	0,23	0,30	0,45	0,04	0,14	0,71
POIDS_ DETTE_CC	-0,33	1,00	-0,15	-0,22	-0,39	-0,07	-0,07	-0,26
NB_DETTEES	0,23	-0,15	1,00	0,25	0,45	0,05	-0,08	0,16
DETTE_ GLOBALE	0,30	-0,22	0,25	1,00	0,39	0,66	0,05	0,19
MONT_ DETTE_ FIN_ NOIMMO	0,45	-0,39	0,45	0,39	1,00	0,06	0,16	0,34
MONT_ DETTE_ AUTRE	0,04	-0,07	0,05	0,66	0,06	1,00	0,02	0,04
AGE	0,14	-0,07	-0,08	0,05	0,16	0,02	1,00	0,16
NIV_VIE	0,71	-0,26	0,16	0,19	0,34	0,04	0,16	1,00

D2 AIC et BIC

L'*AIC* et le *BIC* sont deux critères d'information généralement utilisés pour la sélection de modèles. Soient J modèles paramétriques possibles :

$$\{M_{\beta^{(1)}}, \dots, M_{\beta^{(j)}}\}$$

Soit p_j le nombre de paramètres du modèle j et n le nombre d'observations. Le critère d'information d'Akaike (*AIC*) associé au modèle j est défini par :

$$AIC(j) = -2\ell_n(Y|X; \widehat{\beta}^{(j)}) + 2p_j$$

Où $-2\ell_n(Y|X; \widehat{\beta}^{(j)})$ est le maximum de la log-vraisemblance du modèle j . Ce critère ne permet pas de pénaliser suffisamment le nombre de paramètres lorsque n est très grand. Le critère d'information bayésien (*BIC*), défini par Schwarz en 1978, permet de corriger cela :

$$BIC(j) = -2\ell_n(Y|X; \widehat{\beta}^{(j)}) + p_j \ln(n)$$

Le choix d'un modèle s'effectue alors en minimisant l'un de ces deux critères d'information.

Exemple fictif de choix de modèle par sélection backward utilisant l'AIC :

Soit un modèle $M_{(P1,P2,P3,P4)}$ à 4 paramètres implémenté sur un échantillon de n observations. $P1$, $P2$, $P3$ et $P4$ sont les 4 paramètres du modèle initial complet. On souhaite sélectionner le modèle final en minimisant le critère d'information AIC. On commence donc par calculer l'AIC du modèle $M_{(P1,P2,P3,P4)}$. On calcule également les AIC de chaque modèle dont un des paramètres a été retiré du modèle initial. Le tableau D2.1 suivant présente les résultats :

TABLE D2.1 – Exemple 1

	AIC
$M_{(P1,P2,P3,P4)}$	10 000
$M_{(P2,P3,P4)}$	11 000
$M_{(P1,P3,P4)}$	9 000
$M_{(P1,P2,P4)}$	11 100
$M_{(P1,P2,P3)}$	11 100

Parmi les 5 modèles présents dans le tableau ci-dessus, le modèle $M_{(P1,P3,P4)}$ présente l'AIC le plus faible. Il s'agit du modèle où le paramètre $P2$ a été retiré du modèle initial complet. À présent, réitérons la procédure précédente à partir du nouveau modèle $M_{(P1,P3,P4)}$:

TABLE D2.2 – Exemple 2

	AIC
$M_{(P1,P3,P4)}$	9 000
$M_{(P1,P4)}$	10 000
$M_{(P1,P3)}$	10 500
$M_{(P3,P4)}$	10 300

D'après le tableau D2.2, la suppression de paramètre ne permet plus de diminuer l'AIC. On conserve donc le modèle $M_{(P1,P3,P4)}$ en tant que modèle final.

D3 Coefficients du modèle BIC

Prédicteurs	Coefficients	Écart-type	P-valeurs
<i>Constante</i>	2,163	0,056	<0,001
<i>Premier dépôt</i>			
Non	Réf.	Réf.	Réf.
Oui	-0,303	0,015	<0,001
<i>Aide d'un travailleur social</i>			

Non	Réf.	Réf.	Réf.
Oui	0,064	0,015	<0,001
<i>Poids des dettes de charges courantes</i>			
	0,314	0,025	<0,001
<i>Montant des dettes financières non immobilières</i>			
	-0,00001	0,0000005	<0,001
<i>Patrimoine global</i>			
<30 000 euros	Réf.	Réf.	Réf.
≥ 30 000 euros	-6,522	0,117	<0,001
<i>Situation professionnelle</i>			
CDI/libéral/artisan/commerçant	Réf.	Réf.	Réf.
Autre	0,472	0,029	<0,001
CDD/intérim	0,332	0,029	<0,001
Chômage	0,800	0,023	<0,001
Retraité	0,215	0,036	<0,001
Autre sans activité professionnelle	1,154	0,037	<0,001
<i>Âge</i>			
	0,009	0,0008	<0,001
<i>Structure familiale</i>			
Homme seul	Réf.	Réf.	Réf.
Complexe	0,029	0,051	0,577
Couple avec enfant(s)	-0,252	0,025	<0,001
Couple sans enfant	-0,502	0,030	<0,001
Femme seule avec enfant(s)	0,161	0,024	<0,001
Femme seule	0,096	0,022	<0,001
Homme seul avec enfant(s)	-0,164	0,053	0,002
<i>Niveau de vie</i>			
	-0,003	0,00003	<0,001
<i>Ressource principale</i>			
Revenus d'activité	Réf.	Réf.	Réf.
Autre	0,594	0,051	<0,001
Revenus de transfert	0,523	0,021	<0,001
<i>Cause de surendettement : séparation/divorce</i>			
Non	Réf.	Réf.	Réf.
Oui	0,273	0,021	<0,001
<i>Cause de surendettement : poids élevé des dettes et charges</i>			
Non	Réf.	Réf.	Réf.
Oui	-0,481	0,022	<0,001

TABLE D3.1 – Coefficients du modèle BIC