



HAL
open science

Note méthodologique sur un travail d'appariement en vue d'analyser la position sociale des agriculteurs dans la société française

Antonin Favaro

► To cite this version:

Antonin Favaro. Note méthodologique sur un travail d'appariement en vue d'analyser la position sociale des agriculteurs dans la société française. 2024. hal-04525659

HAL Id: hal-04525659

<https://hal.inrae.fr/hal-04525659v1>

Preprint submitted on 28 Mar 2024

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

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

Note méthodologique sur un travail d'appariement en vue d'analyser la position sociale des agriculteurs dans la société française

Antonin Favaro

Contextualisation

Ce document présente un travail d'appariement préalable à une analyse de la position sociale des agriculteurs. L'objectif était d'actualiser la position sociale des agriculteurs à partir de statistiques en population générale contenant en priorité des informations sur les revenus et le patrimoine, les études sociologiques récentes (Bessière 2011, Laferté 2018) montrant l'importance et l'interdépendance de ces variables dans la distinction des conditions de vie des ménages agricoles.

Pour déterminer la position sociale des agriculteurs, l'approche configurationnelle (Cayouette-Remblière et Ichou, 2019) avait été choisie. Pour cela, il nous fallait un moyen de caractériser les ménages, et pas seulement les individus, à partir de variables sociologiques sur leurs membres et le logement. De plus, il s'agissait de tester l'hypothèse d'une division en trois groupes de la population des ménages agricoles (Laferté, 2021) : des ménages à dominante économique situés en bas à droite de l'espace social (Beaumont, Challier et Lejeune, 2018), un groupe d'agriculteurs alternatifs proche de la petite-bourgeoisie culturelle (Samak, 2014) et un groupe de ménages agricoles au sein de la bourgeoisie économique (Laferté, 2018).

Cela nécessitait des variables qui permettraient de distinguer les ménages agricoles au sein de la population générale mais également au sein de la population agricole. De ce fait, sachant l'importance du patrimoine professionnel dans le capital économique des agriculteurs (Bessière, De Paoli, Roger, Gouraud, 2012), il était important d'avoir des variables de revenus et de patrimoine sur l'ensemble de la population mais aussi des variables sur les exploitations agricoles afin d'évaluer les corrélations entre positions sociales et types d'agriculteurs.

Aussi, nous avons sélectionné 4 bases de données qui permettaient de rassembler ces informations : le recensement de la population (RP) de 2018, la base FIDELI de 2018, le recensement agricole (RA) de 2020 et l'enquête Histoire de Vie et Patrimoine (HVP) de 2017. La base FIDELI étant la base de sondage du RP, les ménages du RP enquêtés en 2018 étaient censés se retrouver dans la base de FIDELI, ce qui rendait l'appariement exact possible. Cet appariement permettait d'associer les revenus fiscaux aux ménages dans le RP. De plus, le RA étant exhaustif sur la population des exploitations agricoles, les agriculteurs présents dans le RP étaient censés se retrouver dans le RA (à ceci près que les dates de recensement pour le RP et le RA diffèrent ici, ce qui implique qu'une partie des agriculteurs du RP puisse ne pas apparaître dans le RA).

Aussi, nous présenterons dans la suite de ce document le travail d'appariement entre les bases du RP et de FIDELI, puis entre la base obtenue (nommée base de référence) et le RA. Le travail d'estimation du patrimoine à partir de la base HVP n'ayant pas abouti, nous ne le présenterons pas ici.

L'ensemble de ce travail a été réalisé à partir d'un accès sécurisé via CASD suite à la validation par le comité du secret statistique du projet, dans le cadre d'un contrat doctoral financé par l'Université de

Bourgogne et grâce au soutien financier de la Région Bourgogne-Rhône-Alpes. Le travail doctoral était encadré par Gilles Laferté (Inrae) et Joanie Cayouette-Remblière (Ined)

Périmètre

Le périmètre de la population est celui des ménages ordinaires de France métropolitaine. On identifie le ménage à la personne de référence et son conjoint. La personne de référence et le conjoint doivent avoir plus de 20 ans et avoir déjà travaillé (on enlève les individus des PCS 7 et 8, soit les retraités et les inactifs).

Appariements

Méthode générale

On résume ici le principe de la méthode qui sera détaillée ensuite dans chaque cas.

Toute la difficulté ici repose sur le fait qu'il n'y ait pas d'identifiant direct entre les bases, ce qui exclut l'appariement direct². Il faut donc une façon d'identifier, de façon indirecte, les individus dans les différentes bases.

Le principe de l'appariement indirect n'est pas de chercher à tout appairier mais de chercher à effectuer le plus d'appariements possibles tout en minimisant le nombre de faux, c'est-à-dire soit de faux positif (deux individus seraient appariés alors qu'ils sont différents) ou de faux négatif (deux individus identiques ne sont pas appariés).

Il y a donc un arbitrage à faire entre maximiser le nombre d'appariement et minimiser le nombre d'erreurs possibles. Pour cela, nous avons divisé l'appariement en plusieurs étapes : un premier appariement sur des variables assurées, pour lesquelles on ne doute pas de la qualité, c'est-à-dire, du faible nombre d'erreurs possibles, et qui sont discriminantes. Ensuite, nous distinguons les concordances uniques, les concordances multiples et les absences de concordances : les individus ne trouvant qu'une seule concordance sont considérés comme appariés, les individus ayant plusieurs concordances sont utilisés dans l'étape suivante et les individus sans concordance sont mis de côté. On effectue ensuite un nouvel appariement sur les individus ayant des concordances multiples en utilisant des variables de moins bonne qualité. On répète éventuellement l'opération jusqu'à ce qu'il ne reste plus beaucoup d'individus à appairier. Enfin, suivant les cas, on reprend éventuellement les individus non-appariés pour refaire un appariement avec d'autres variables. Finalement, la dernière opération consiste en l'évaluation de l'appariement.

Appariement RP-FIDELI : base de référence (BR)

Description des bases

La définition des unités de base n'est pas la même dans les deux cas : le ménage dans le RP et le ménage fiscal dans FIDELI. Or, les ménages sont identifiés, pour l'appariement et pour la classification, par les variables sur les personnes du couple de référence, ce qui suppose de définir une personne 1 et une personne 2⁵, une personne de référence et un conjoint. Si la méthode d'attribution est explicite pour ce qui est du recensement de la population⁶, elle ne l'est pas vraiment pour ce qui est de FIDELI : les logements sont divisés en foyers fiscaux, tels que définis par l'administration fiscale, c'est-à-dire selon les personnes inscrites sur la déclaration des revenus du foyer, ce qui fait qu'il peut exister plusieurs personnes de référence (fiscale) dans un même logement d'abord. Ensuite, les concepteurs de FIDELI ont construit la variable (LIEN_FAMILIAL) qui attribue à la personne, « de façon conventionnelle », le statut de « personne de référence » du logement suivant l'estimation qui est faite de la probabilité d'être la personne s'étant acquittée de la taxe d'habitation. Dans le recensement, il existe une procédure d'attribution : la personne la plus âgée en couple et étant en emploi. Ces trois critères sont hiérarchisés mais rien ne permet d'affirmer que la personne de référence de FIDELI concorde avec celle du recensement.

Ainsi, nous avons deux définitions différentes de la personne de référence. La solution trouvée est d'effectuer l'appariement sur les individus et non pas le couple de référence.

Nous avons ensuite évalué la qualité des différentes variables pour l'appariement :

- L'année et la commune de naissance ainsi que le sexe ne posent pas de problèmes : elles ont la même définition dans les deux bases et ne sont pas susceptibles de prendre des valeurs différentes d'une base à l'autre pour un même individu ;
- Le lieu de résidence : entre 2008 (référence du recensement 2018) et 2013 (référence de FIDELI 2018), il y a eu un changement d'IRIS. En comparant les bases IRIS des deux années, on voit que les différences sont minimales (environ 300 communes modifiées pour 50 mille, toutes concernant des villes de petite taille). Je suis donc parti du principe que les erreurs dans l'appariement seraient minimales. J'ai sélectionné cette variable dans l'appariement ;
- Le statut conjugal : il ne pose pas de problème puisqu'il est constitué de presque toutes les modalités du recensement dans FIDELI. Cependant la distribution peut différer entre les deux bases :

	RP	FIDELI
Célibataire	39%	32%
Marié	46%	47%
Pacsé	9%	9%
Divorcé	6%	11%

Veuf	1%	1%
------	----	----

On constate un certain écart entre la part de célibataires et entre la part des divorcés, ce qui explique qu'on ait plutôt pris la variable dans un troisième temps, après avoir opéré déjà un appariement. Les distributions restent tout de même proches ;

- Type de logement : étant donné le périmètre, il n'y a que des maisons et des appartements et les distributions sont proches :

	RP	FIDELI
Appartement	43%	40%
Maison	57%	60%

- Statut d'occupation : les distributions sont proches :

	RP	FIDELI
Propriétaire	55%	58%
Locataire	43%	40%
Logé à titre gratuit	2%	2%

- Logement HLM : l'indicateur a été recodé à partir des variables HLM dans le recensement et LOGEMENT dans FIDELI qui distingue les logements suivant le type de personne morale propriétaire. Nous avons simplement recodé HLM et non-HLM dans la base FIDELI. En effet, la variable nous a semblé relativement discriminante et assurée. Les distributions concordent relativement bien entre les deux bases de données :

	RP	FIDELI
HLM	15%	13%
Non HLM	85%	87%

- La date de fin de la construction : elle pose des difficultés. Dans FIDELI, il est noté dans la documentation qu'elle est construite comme variable de synthèse entre les variables DATACHEV (date d'achèvement de la construction) et DATCONST (date d'entrée dans le

fichier des bâtis) : la date la plus ancienne est sélectionnée pour définir la variable de synthèse. Les différences entre RP et FIDELI sont importantes en termes de répartition pour les années après 2016, ce qui est dû aux difficultés inhérentes à la construction de FIDELI (le problème est souligné dans la documentation sans plus de précision quant aux raisons) : le nombre de logements dans FIDELI sur la période est beaucoup moins élevé que dans le recensement et que pour les autres années de FIDELI.

- Le nombre de pièces : il ne semble pas correspondre correctement : il semble y avoir une différence de distribution entre le recensement et FIDELI, d'environ 3 pièces. Nous avons préféré ne pas prendre la variable, le nombre de pièces de décalage nous paraissant relativement élevé et les résultats d'appariement suffisants avec les variables sélectionnées précédemment;
- Les pièces et l'équipement du logement : la présence d'une douche, le nombre de salles de bains et de WC. Les modalités nous semblaient peu discriminantes au vue des variables déjà sélectionnées pour l'appariement et nous ne les avons pas maintenues.

Au vu des limites posées par chaque variable, nous nous sommes restreint à un ensemble de variables dont assurées, qui nous semblaient suffisamment identifiantes et qui nous semblaient nécessaires pour différencier les individus : l'année de naissance, le sexe et le lieu de résidence. Nous avons ensuite ajouté ensuite le statut conjugal, le type de logement et le logement HLM parce qu'ils étaient suffisamment assurés et discriminants.

Appariement

Nous nous sommes restreint à 2018, l'année la plus récente dont nous disposons pour la base FIDELI. La population dans les deux bases de données est la suivante :

Base de données	Nombre de ménages	Dont ménages agricoles	Nombre d'individus	Nombre d'agriculteurs
Recensement de la population (RP non pondéré)	1 156 492	12 184	1 683 163	13 929
FIDELI	18 991 690		28 439 113	

1ère étape

Nous avons apparié les individus des deux bases selon la procédure suivante :

- Concordance unique (il n'y a qu'un individu dans FIDELI qui concorde avec l'individu du RP) : nous considérons qu'il y a appariement entre les deux ménages ;
- Concordance multivoque (aucun appariement unique dans le ménage et l'individu concorde avec plusieurs individus dans FIDELI) : il est utilisé pour l'étape suivante ;
- Aucune concordance : il est repris à l'étape 4 ;

Cela permet de diminuer le nombre d'individus pour lesquels il existe plusieurs candidats à l'appariement dans FIDELI et diminue les erreurs d'appariement qui peuvent exister sur les variables en plus des erreurs de déclaration éventuelles : pour les couples, il suffit qu'un seul individu s'apparie de façon univoque pour pouvoir appairer le couple entier.

Après cette première étape, nous obtenons les résultats suivants :

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairer	Nombre d'individus écartés en raison de non-appariements
RP	1 710 163	906 154	507 843	269 166	1 059 035	354 962	269 166
Dont agriculteurs	13 929	7 729	4 303	1 897	9 569	2 463	1 897

Il ne reste plus que 2 463 agriculteurs à appairer, du fait que 9 569 soient regroupés dans un ménage. Il reste 354 962 individus avec plusieurs concordances mais pas dans un ménage avec une concordance unique. Nous les reprendrons en étape 5. Les individus sans concordance seront repris en étape 4.

2ème étape

Pour les individus avec plusieurs concordances, on ajoute deux variables supplémentaires dans l'appariement qui sont relativement fiables. Les variables HLM et de statut d'occupation concordent bien sur les ménages appariés en première vague. Nous obtenons les résultats suivants :

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairer	Nombre d'individus écartés en raison de non-appariements
RP	354 962	135 543	101 112	118 307	153 751	82 904	118 307
Dont agriculteurs	2463	957	663	843	1 107	513	843

3ème étape

Nous ajoutons le statut conjugal dans cette étape. Nous obtenons les résultats suivants :

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairer	Nombre d'individus écartés en raison de non-appariements
RP	82 904	29 293	43 492	10 119	33 217	39 568	10 119
Dont agriculteurs	513	187	254	72	209	232	72

4ème étape : appariement des individus sans concordance à la première étape

Nous reprenons ici les individus ayant été écartés à la première étape. Nous utilisons le département au lieu de l'IRIS de résidence et faisons les mêmes appariements ensuite avec l'ensemble des autres variables : statut conjugal, HLM, surface du logement, type de logement.

Il y a 269 166 individus concernés au départ. Les résultats sont très proches entre l'appariement avec le sexe, l'année de naissance, la commune de naissance et le département de résidence et l'appariement avec toutes les variables. Pour améliorer la précision, nous ajoutons toutes les variables :

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairer	Nombre d'individus écartés en raison de non-appariements
RP	269 166	86 571	112 153	70 442	92 900	105 824	70 442
Dont agriculteurs	1 897	418	1 011	468	484	945	468

On voit que cette méthode nous permet de récupérer un nombre d'appariements relativement assurés étant donné 92 900 individus appartiennent à un ménage ayant un individu avec une concordance unique alors même que l'on avait choisi le département comme variable de localisation.

On abandonne les 468 agriculteurs et 70 442 individus non-appariés à la fin de la cette étape de manière définitive du fait de la trop grande incertitude.

5ème étape : attribution d'un revenu

On reprend ici les individus restant à appairer aux étapes 3 et 4 soit 145 392 individus.

À défaut de pouvoir récupérer l'ensemble des variables de FIDELI, nous avons cherché à attribuer un revenu crédible aux ménages restants, afin de les conserver pour les analyses futures.

Il s'agissait, dans ce cas, d'attribuer un revenu crédible aux individus afin de limiter les probabilités d'erreur d'attribution. Pour cela, nous avons considéré qu'il valait mieux prendre comme critère le salaire par rapport à la PCS, c'est-à-dire que pour une PCS donnée, nous avons défini un intervalle de revenu probable au regard des distributions par PCS publiées par l'INSEE et avons choisi de ne garder que les concordances qui seraient dans cet intervalle. Puis, dans le cas où il y aurait encore plusieurs concordances possibles, nous attribuons comme concordance effective l'individu

dont les revenus se rapprochent le plus de la médiane de la distribution pour la PCS donnée. Cela permet, même si l'individu attribué n'était pas le bon, de s'assurer que les revenus attribués ne soient pas aberrants au regard de la population concernée. Pour l'année 2018, nous nous référons à cette distribution de revenus :

	1 ^{er} décile	Médiane	9 ^{ème} décile
Agriculteurs exploitants	0	868	3 117
Artisans, commerçants et chefs d'entreprise	56	1 268	4 148
Cadres et professions intellectuelles supérieures	1294	2 794	5 703
Professions intermédiaires	767	1 783	2 843
Employés	339	1 226	1 920
Ouvriers	311	1 355	2 073
Ensemble des actifs ayant travaillé	437	1 536	3 218

Lecture : en 2018, en France, les 10% d'agriculteurs exploitants, femmes et hommes, aux revenus les plus élevés gagnent au moins 3 117 euros par mois.

Champ : France métropolitaine, actifs ayant déjà travaillé vivant en logement ordinaire.

Source : Insee, enquête Revenus fiscaux et sociaux 2018 (<https://www.nomenclature-pcs.fr/decire/revenus-patrimoine>, consulté le 16/09/2022).

Parmi la population appariée jusque-là, nous avons comme distribution :

	1 ^{er} décile	Médiane	9 ^{ème} décile
Agriculteurs exploitants	334	1 671	3 359
Artisans, commerçants et chefs d'entreprise	522	1 795	4 320
Cadres et professions intellectuelles	946	2 521	5 450

supérieures			
Professions intermédiaires	782	1 906	3 345
Employés	514	1 541	2 708
Ouvriers	623	1 682	2 976
Actifs sans emplois	231	1 264	2 710

Pour ce qui est des agriculteurs, l'écart est conséquent mais pour les autres PCS, la distribution est semblable.

Nous avons finalement estimé le revenu de 144 306 individus sur 145 392.

Résultats finaux

Nous avons finalement apparié tous les individus excepté 70 442 individus, ceux sans concordance à la 5^{ème} étape. Il y a 1 012 800 ménages, soit 1 400 072 individus appariés à la suite des 4 premières étapes, donc, 1 544 378 à la fin de l'appariement.

Parmi les individus appariés, 21 535 déclarent un revenu agricole non nul.

Base de données	Nombre d'individus appariés	Nombre de ménages appariés	Nombre d'individus déclarant un revenu agricole non nul	Nombre d'agriculteurs déclarés
Recensement de la population (RP non pondéré)	1 424 492	967 046	21 535	13 929

Ensuite, l'échantillon n'est peu, voire pas déformé par l'appariement suivant les PCS :

PCS	Part dans la population de référence	Part dans l'échantillon final	Différence (pourcentage de points de pourcentage)

Agriculteurs	0,83	0,85	+2%
Commerçants et artisans et chefs d'entreprise	5,15	5,17	+0,4%
CPIS ¹²	16,84	18,46	+9,6%
Professions intermédiaires	22,95	24,21	+5,4%
Employés	24,84	24,72	-0,5%
Ouvriers	17,02	16,63	-2,2%
Inactifs	12,38	9,96	-19,5%

Comme nous l'avons souligné, la part des agriculteurs est relativement stable dans l'échantillon et correspond à la part dans la distribution de référence¹³.

Appariement BR-RA : population agricole de référence

Périmètre et limites

Nous nous sommes restreint, du côté de la BR, à la population des individus déclarés « agriculteur exploitant » ou qui déclarent un bénéfice agricole non nul et non absent, du côté du recensement agricole.

En croisant les PCS avec le bénéfice déclaré, nous obtenons la distribution suivante :

CS	Bénéfices positifs	Bénéfices négatifs	Bénéfices nuls	Valeurs manquantes	Total
1	4 326	748	5 186	3 669	13 929
2	545	123	0	0	668
3	792	266	0	0	1 058
4	1 532	400	0	0	1 962
5	1 576	326	0	0	1 902
6	1 272	282	0	0	1 554
8	373	89	0	0	462
Total	10 446	2 234	5 186	3 669	21 535

Environ un tiers des agriculteurs déclarent un bénéfice agricole nul, un tiers un bénéfice agricole positif et un tiers un bénéfice agricole négatif ou absent. On peut être assuré que le bénéfice agricole fiscal nul pour les agriculteurs est effectivement nul. Il peut cependant y avoir eu des investissements effectués, qui augmenteraient le capital productif et donc, in fine, le capital. Il ne peut cependant y avoir eu de transfert direct, qui sont toujours prélevé après déduction des impôts.

Il existe un premier critère pour évaluer le nombre d'agriculteurs dans la base FIDELI : le bénéfice agricole. Cependant, on ne peut distinguer entre les valeurs nulles et les valeurs absentes : par exemple, si un individu déclare des revenus provenant de salaires mais n'a pas déclaré de bénéfice agricole, le bénéfice agricole sera codé « 0 » et non pas « . », les « . » étant réservés aux individus pour lesquels il n'existe aucune déclaration fiscale (l'ensemble de la ligne des revenus fiscaux est codée « . »). On ne peut donc pas distinguer les individus ayant un bénéfice agricole nul et ceux n'ayant réellement pas de bénéfice agricole. Nous ne prendrons donc en compte, pour la population des agriculteurs, que ceux ayant un bénéfice agricole non nul et non absent.

D'après le recensement agricole 2020, nous avons cette distribution en déciles pour le nombre d'exploitations par commune :

Décile	9	8	7	6	5	4	3	2	1	Max
Valeur	2	3	4	6	7	9	12	16	25	518

Cela veut donc dire qu'il y a plus de 90% des communes qui comportent moins de 25 exploitations. Théoriquement, il peut y avoir plus d'un exploitant sur une même exploitation. De plus, d'après les résultats agrégés du recensement agricole, il y a 514 694 exploitations pour 629 039 exploitants et coexploitants¹⁴, donc 20% d'agriculteurs en plus que d'exploitations. Si nous n'avons pas la distribution du nombre d'exploitant et coexploitant par commune, nous pouvons avoir une certaine idée de la distribution avec celle des exploitations. De plus, un même exploitant peut-être exploitant sur plusieurs exploitations.

Une difficulté de cet appariement est que les variables de localisation ne sont pas identiques a priori: l'une renvoie à la commune de résidence de l'individu, l'autre, à la commune du siège de l'exploitation. Nous partons du principe que la plupart des exploitants habite la même commune que le bâtiment principal de l'exploitation et que, à défaut, nous prendrons le département comme localisation de référence. Une fois l'appariement effectué avec la commune, les individus non appariés seront isolés du reste et appariés avec le département de résidence. Aller plus loin, au niveau de la région, serait pour nous trop approximatif. Nous finirons, pour les individus sans appariement possible au niveau du département.

Base de données	Nombre d'individus ayant un bénéfice agricole non nul	Nombre d'agriculteurs déclarés	Nombre de ménages

	et non absent		
Recensement de la population (RP non pondéré)	21 535	13 929	16 938
RA (exploitants et coexploitants)	496 392	496 392	

Appariement

1ère étape : appariement des individus en utilisant la commune

La première étape consiste à appairer les individus avec les variables d'année de naissance, de sexe et la commune de localisation. A la fin de la première étape, nous obtenons les résultats suivants :

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairer	Nombre d'individus écartés en raison de non-appariements
RP	21 535	9 243	4 153	8 139	9 737	3 659	8 139

2ème étape : appariement des équivoques avec le diplôme

Le diplôme est donné dans les deux bases de données : le niveau de diplôme le plus élevé dans le RP et le niveau de diplôme général d'un côté et agricole de l'autre dans le RA. Il suffit de recoder le diplôme du RA dans le RP pour retrouver les mêmes niveaux de diplôme. Du diplôme général ou agricole, on prend celui qui est le plus élevé pour faire l'appariement.

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre	Nombre	Nombre de	Nombre	Nombre	Nombre

		d'individus appariés de façon univoque	d'individus avec concordance multiples	non-appariements	d'individus appartenant à un ménage avec une concordance unique	d'individus restant à appairier	d'individus écartés en raison de non-appariements
R P	3 586 ¹	1 350	1 068	1 168	1 369	1 041	1 168

Nous avons donc notre ensemble d'individus appariés. Il nous faut désormais traiter les non-appariés de la première étape.

3ème étape : appariement des non-appariés de la première étape avec le département de résidence

Ici, nous refaisons un appariement avec le département de résidence et le département de localisation du siège de l'exploitation. C'est un moyen pour récupérer les cas où le siège d'exploitation ne se situe pas dans la même commune que le lieu de résidence.

Avec le département, on obtient 7 882 individus avec concordance et seulement 257 qui n'en ont pas sur 8 139. 3 544 ont une concordance unique et 3 764 individus appartiennent à un ménage qui possède une concordance unique.

	Nombre total d'individus	Résultat de l'appariement par individus			Résultat de l'appariement des ménages avec au moins un individu apparié de façon univoque		
		Nombre d'individus appariés de façon univoque	Nombre d'individus avec concordance multiples	Nombre de non-appariements	Nombre d'individus appartenant à un ménage avec une concordance unique	Nombre d'individus restant à appairier	Nombre d'individus écartés en raison de non-appariements
R P	8 139	3 544	4 338	257	3 764	4 118	257

Résultats finaux

Finalement, nous obtenons ces résultats pour l'appariement entre BR et RA :

¹ La différence qui apparaît entre le nombre d'individus au départ de la seconde étape et le nombre d'individus restant à appairier à la première étape est due à une différence de résultats entre plusieurs essais d'appariement, les deux chiffres étant censés concorder. L'accès au CASD nous étant désormais fermé, nous n'avons pas pu corriger cette erreur. La différence relativement faible entre ces deux chiffres nous incite tout de même à présenter les résultats finaux

	Nombre de concordances uniques	Nombre d'individus ayant des concordances multiples	Nombre d'individus appartenant à un ménage avec une concordance unique	Aucune concordance	Nombre total d'individus
RP	14 137	5 159	14 870	1 425	21 454 ²

Bibliographie

BEAUMONT, Amélie, CHALLIER, Raphaël, LEJEUNE, Guillaume. « En bas à droite. Travail, visions du monde et prise de position politiques dans le quart en bas à droite de l'espace social. » *Politix*, 2018/2, n°122, pp 9-31

BESSIERE, Céline, DE PAOLI Caroline, ROGER, Murielle et GOURAUD, Bénédicte, « Les agriculteurs et leur patrimoine: des indépendants comme les autres ? » *Economie et statistique*, 2011, 444-445, pp. 55-74.

CAYOUEITE-REMBLIERE, Joanie, ICHOU, Mathieu, « Saisir la position sociale des ménages : une approche par configurations. » *Revue française de sociologie*, 2019/3, vol. 60, pp. 385-

JEANNEAUX, Philippe, et VELAY, Nathalie. « Capitalisation du revenu agricole et formation du patrimoine professionnel des exploitations agricoles », *Économie rurale*, vol. 378, no. 4, 2021, pp. 97-117.

LAFERTE, Gilles. *L'embourgeoisement : une enquête chez les céréaliers*. Paris : Raisons d'agir, 1^{ère} éd., 2018.

LAFERTE, Gilles. « Des revenus à la position sociale : reclasser les agriculteurs. » *Économie rurale*, 2021/4, n° 378, pp. 159-174.

PARANTHOEN, Jean-Baptiste. « Déplacement social et entrées an agriculture. Carrières croisées de deux jeunes urbains devenus maraîchers. » *Sociétés contemporaines*, 2014/4, n°96, pp. 51-76.

SAMAK, Madlyne. « La politisation variable des alternatives agricoles », *Savoir/Agir*, 2016/4 (N° 38)

² Nous ne pouvons expliquer en l'état la différence qui apparaît entre le nombre d'individus pris au départ, soit 21 535, et le nombre d'individus calculés à partir des résultats, soit 21 454. L'accès au CASD nous étant désormais fermé, nous n'avons pas pu corriger cette erreur. La différence relativement faible entre ces deux chiffres nous incite tout de même à présenter les résultats finaux.