



HAL
open science

L'université française, lieu de brassage ou de ségrégation sociale? Mesure de la polarisation du système universitaire français (2007-2015)

Romain Avouac, Hugo Harari-Kermadec

► To cite this version:

Romain Avouac, Hugo Harari-Kermadec. L'université française, lieu de brassage ou de ségrégation sociale? Mesure de la polarisation du système universitaire français (2007-2015). *Economie et Statistique* / *Economics and Statistics*, INSEE, 2022, pp.63-83. 10.24187/ecostat.2021.528d.2058 . hal-03547191

HAL Id: hal-03547191

<https://hal-univ-orleans.archives-ouvertes.fr/hal-03547191>

Submitted on 28 Jan 2022

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.

L'université française, lieu de brassage ou de ségrégation sociale ? Mesure de la polarisation du système universitaire français (2007-2015).

Romain Avouac* et Hugo Harari-Kermadec**

Résumé

Les transformations récentes de l'enseignement supérieur (politiques de regroupement, influence des classements universitaires, etc.) ont pu alimenter la crainte d'assister à une polarisation entre, d'un côté, un nombre restreint de "world-class universities" et, de l'autre, un vaste ensemble d'universités de second rang. Néanmoins, peu de travaux quantitatifs permettent d'éclairer cette question. À partir de la base des inscriptions universitaires en France (bases « Système d'information sur le suivi de l'étudiant »), nous proposons une visualisation exhaustive de l'espace universitaire selon les capitaux détenus par les différents publics étudiants. Nous mobilisons ensuite des mesures de ségrégation et de polarisation pour caractériser la dynamique de cette hétérogénéité, qui s'accroît entre 2007 et 2015. Nous mettons enfin en relation cette polarisation avec les dispositifs nationaux (initiatives d'excellence) et internationaux (classements universitaires) qui structurent une globalisation universitaire réaffirmée depuis le milieu des années 2000.

Codes JEL : I24, I23, I28, C38

Mots-clés : enseignement supérieur, polarisation, inégalités scolaires, classements universitaires, origine sociale, France

* ENS Paris-Saclay et ENSAE (romain.avouac@ensae.fr) ; ** IDHES, ENS Paris-Saclay (hugo.harari@ens-paris-saclay.fr)

Nous remercions Frédéric Lebaron et Thibaut de Saint Pol, et les participants du séminaire quantitativisme réflexif à l'IDHES pour leurs remarques sur une version préliminaire de ce travail.

La dernière décennie a été marquée par une dynamique de globalisation de l'enseignement supérieur, symbolisée par l'émergence du classement de Shanghai dans le débat public français. Si l'explosion de la mobilité étudiante est bien réelle (+20% entre 2010 et 2020¹), le défi posé par la massification de l'accès au supérieur ne devrait pas être sous-estimé. Pourtant, les politiques publiques ont mis au premier plan la recherche de visibilité internationale de quelques établissements à travers des procédures concurrentielles d'attribution de fonds importants, à commencer par les « Initiatives d'excellence » (Idex). De très nombreux établissements ont alors participé à des projets de regroupement visant à atteindre la "taille critique" nécessaire pour intégrer les classements internationaux ou s'y élever. La question de la performativité (Espeland & Sauder, 2007, Paradeise & Thoenig, 2015) de ces dispositifs est alors posée : les classements internationaux ne participent-ils pas à produire un effet sur la réalité qu'ils sont censés décrire ?

Cependant, il n'est pas certain que l'excellence de quelques champions nationaux dans l'espace concurrentiel global entraîne mécaniquement celle d'un système d'enseignement supérieur massifié. Amsler et Bolsmann (2012) soutiennent ainsi que, dans le cas du Royaume-Uni, les classements internationaux mettent en avant une vision plus étroite de l'enseignement, centrant le débat public sur les établissements d'élite et reléguant au second plan les enjeux de démocratisation. Les politiques comme les financements publics tendraient alors à se concentrer sur quelques établissements, avec le risque d'assister à une généralisation de systèmes éducatifs à deux vitesses : d'un côté, un nombre restreint d'universités de rang mondial captant à elles seules une large part des ressources et, de l'autre, un vaste ensemble d'universités de second rang se partageant les ressources résiduelles (Van Parijs, 2009). Merton (1968) parle d'effet Matthieu pour décrire l'effet cumulatif de la renommée scientifique dans le monde de la recherche.

L'université est le lieu principal de la massification de l'accès au supérieur en France au cours des dernières décennies : entre 1984 et 2009, la part des enfants d'ouvriers diplômés de l'université est passée de 6 % à 24 % (Peugny, 2013). Elle voit donc désormais se côtoyer des étudiants d'origines et de profils variés (Brinbaum *et al.*, 2018). Bien que l'espace universitaire soit lui-même stratifié (Frouillou 2017, Convert 2006), on pourrait espérer qu'il permette ainsi aux étudiants d'origine populaire de poursuivre l'accumulation d'un capital scolaire mais aussi de compenser partiellement le capital social déficitaire de la socialisation initiale, grâce aux pairs côtoyés à l'université (Truong, 2015). À côté des grandes écoles déjà très fermées aux classes populaires (Ichou & Vallet, 2013), un système universitaire lui-même polarisé entraînerait au contraire une entrée des jeunes dans la vie adulte fortement ségréguée, dans

¹ Repères et références statistiques (2011) et Note d'information n°5 (2020) du « Systèmes d'Information et Études statistiques » du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.

la continuité de l'expérience scolaire, renforçant le déterminisme des trajectoires scolaires puis professionnelles.

Mais observe-t-on en pratique une polarisation du système universitaire français en matière de composition sociale ? Afin d'apporter des éléments de réponse à cette question, nous étudions l'évolution de la composition sociale des populations étudiantes à partir des données de la base exhaustive des inscriptions dans les universités françaises (base SISE inscriptions du SIES), disponibles entre 2007 et 2015.

Dans une première partie, nous proposons une image statistique exhaustive du paysage universitaire français à travers les caractéristiques sociales et scolaires des publics étudiants de différents établissements. Pour ce faire, nous appliquons une analyse en composantes principales (ACP) afin de synthétiser l'information contenue dans les données initiales. Une classification ascendante hiérarchique (CAH) nous permet ensuite de construire une typologie des universités, à partir de laquelle nous analysons le niveau d'hétérogénéité des établissements en coupe instantanée. Nous mettons en évidence un paysage universitaire socialement hétérogène et hiérarchisé, corroborant les résultats que Brusadelli et Lebaron (2012) obtiennent à partir d'une méthodologie similaire.

Dans un second temps, nous développons une analyse visant à quantifier l'évolution de l'hétérogénéité des compositions sociales des établissements d'enseignement supérieur entre 2007 et 2015. Nous introduisons d'abord une mesure de polarisation issue de la littérature économétrique (Esteban & Ray, 1994), qui met en évidence une hausse de la polarisation sociale sur la période étudiée. Nous comparons ces résultats à ceux obtenus à partir d'un indice d'entropie normalisé, classiquement utilisé dans la littérature sur la ségrégation scolaire.

Enfin, nous mettons en relation l'hétérogénéité sociale qui caractérise l'enseignement supérieur français avec les dispositifs nationaux (initiatives d'excellences, regroupements universitaires) et internationaux (classements universitaires) qui structurent la globalisation universitaire. Ces dispositifs se révèlent distinguer les établissements aux populations étudiantes les plus favorisées. C'est donc vers les étudiants les plus dotés en capitaux de toutes sortes que se concentrent les financements supplémentaires (publics ou privés) orientés par les dispositifs d'excellence. Ce constat pose la question de la redistribution des ressources opérée par les politiques de l'enseignement supérieur.

1. Un paysage universitaire socialement hétérogène

Si la massification de l'enseignement supérieur est indubitable, la démocratisation de l'accès à l'ensemble de ses établissements pose question. Pour traiter de l'écart entre ces deux dynamiques, nous proposons d'introduire une notion de polarisation sociale. Pour cela, nous devons tout d'abord caractériser la composition sociale des universités sur le plan statique.

Les bases exhaustives des inscriptions dans les universités françaises (bases SISE inscriptions du SIES), disponibles entre 2007 et 2015, contiennent pour chaque étudiant inscrit dans une université française son établissement d'inscription, des informations sociodémographiques ainsi que de données sur sa scolarité présente et passée (encadré 1). Elles permettent de caractériser la composition de la population étudiante de chaque établissement. Par exemple, pour chaque université, on calcule la part d'étudiants dont le parent référent est cadre supérieur. Les tables contenant des variables qualitatives décrivant les étudiants conduisent ainsi à des tables agrégées contenant des variables quantitatives décrivant les universités.

1.1. Une analyse en composantes principales comme outil de réduction de la dimension

Afin de visualiser la distribution des caractéristiques sociales et scolaires des populations étudiantes, nous appliquons aux données transformées une analyse en composantes principales (ACP). Cette méthode permet de réduire la dimension du problème – dimension initialement égale au nombre de variables incluses pour caractériser la composition sociale des publics étudiants des universités. Cette approche a été particulièrement utilisée par des auteurs cherchant à produire des indices de statut socioéconomique (SES) à partir de multiples variables caractérisant imparfaitement le milieu social (Vyas & Kumaranayake, 2006). Plus récemment, Rocher (2016) note que "les classifications des PCS présentent des limites dès qu'il s'agit de décrire la tonalité sociale de groupe d'élèves (classes, établissements, académies)" (Rocher, 2016, p. 16) et recourt donc également à des méthodes factorielles pour construire un "indice de position sociale des élèves".

Plusieurs variables déclinées au niveau des étudiants sont à même de caractériser la composition des universités une fois agrégées. Les catégories socio-professionnelles (CS) des parents d'un étudiant sont un marqueur majeur de son milieu social d'origine et des différents capitaux qu'ils peuvent lui transmettre, bien que ces variables ne permettent pas de quantifier totalement l'influence du milieu d'origine sur les trajectoires scolaires individuelles (Boutchénik *et al.*, 2015). Si une pratique courante consiste à résumer l'origine sociale des individus par la CS du père, le rôle prépondérant joué par le niveau d'éducation des mères dans le niveau d'étude atteint par les enfants (Place & Vincent, 2009) plaide pourtant pour l'inclusion des deux CS. Dans notre cas, nous sommes limités par un souci d'homogénéité : la CS du second parent référent de l'enfant n'est présente dans SISE qu'à partir de 2009. Nous caractérisons donc l'origine sociale des étudiants par la CS de leur premier parent référent (voir annexe 1 pour le choix de la nomenclature).

L'inclusion de variables relatives à la scolarité des étudiants peut permettre de décrire plus finement la composition des universités. En particulier, la série du bac constitue un marqueur social important qui joue un rôle majeur dans les trajectoires dans l'enseignement supérieur (Duru-Bellat & Kieffer, 2008). Le statut de boursier (du supérieur) de l'étudiant ainsi que son échelon paraissent également pertinents.

Cependant, ces variables n'étant disponibles que de 2007 à 2011, nous ne les intégrons pas afin de conserver une approche homogène sur la période. Des analyses complémentaires (non reproduites ici) effectuées sur cette période réduite montrent que leur inclusion ne modifie pas les résultats obtenus.

Enfin, nous incluons le sexe des étudiants ainsi que le cycle d'inscription (licence/master/doctorat) comme variables supplémentaires dans l'ACP : elles ne participent pas à la construction des axes. Si les trajectoires scolaires dans l'enseignement supérieur des étudiants et des étudiantes présentent des différences marquées (Rosenwald, 2006), la part des différents sexes ne nous semble cependant pas constitutive en soi de la composition des universités. De même, le niveau d'études atteint dans le supérieur dépend de l'origine sociale (Brinbaum *et al.*, 2018), mais l'inclure comme variable active poserait un problème d'endogénéité étant donné que l'on s'intéresse à l'effet de l'origine sociale sur les trajectoires dans l'enseignement supérieur.

1.2. La première composante principale comme indice synthétique de composition sociale

Au final, l'ACP est construite à partir de 13 variables actives. Nous présentons seulement les résultats issus de l'ACP réalisée sur les données de 2007. Ce sont en effet les axes de celle-ci que nous utilisons comme référence pour construire l'indice de composition sociale et assurer la comparabilité de ce dernier dans le temps (encadré 2).

L'"éboulis" des valeurs propres permet d'évaluer la qualité de la réduction du nuage initial opérée par l'ACP (figure A1-I). L'application du critère du coude² amène à retenir les deux premières composantes principales (CP) pour mener l'analyse. Le critère de Keiser³ conduit quant à lui à retenir également la troisième CP. Cependant, contrairement aux deux premiers, le troisième axe apparaît difficilement interprétable, et n'apporte que très peu d'information supplémentaire par rapport à la moyenne des axes. Aussi, nous restreignons l'analyse au premier plan factoriel, i.e. le plan formé des deux premières CP. Ce dernier résume 64 % de l'inertie totale ; les représentations obtenues sur les deux premières dimensions synthétisent une grande partie de l'information contenue dans le nuage initial et se prêtent bien à l'interprétation.

Le graphique des variables (figure I) permet d'analyser la structure des corrélations entre les variables actives et de donner une interprétation aux axes de l'ACP. Les variables contribuant le plus fortement au premier axe (horizontal) sont la part d'enfants de cadres supérieurs à gauche de l'axe et, à l'opposé,

² L'analyse en composante principale ordonne les dimensions par ordre décroissant d'inertie. L'inertie a tendance à décroître rapidement sur les premières dimensions, puis à décroître plus lentement et régulièrement pour les dimensions suivantes, moins informatives. Le critère du coude consiste à retenir les dimensions avant la rupture de pente dans l'inertie.

³ Le critère de Keiser consiste à choisir comme seuil l'inertie totale divisée par le nombre de dimensions. On retient alors les dimensions dont l'inertie est supérieure à la moyenne et on omet les autres.

les pourcentages d'étudiants titulaires d'un baccalauréat professionnel ou technologique, ainsi que les parts d'enfants d'ouvriers non qualifiés, d'employés et, dans une moindre mesure, d'ouvriers qualifiés. Cette proximité entre les variables à droite de l'axe confirme le constat selon lequel "les élèves de milieu populaire [...] [sont] surreprésentés dans ces filières technologiques et professionnelles" (Duru-Bellat & Kieffer, 2008, p. 126). Ainsi, l'axe 1 semble opposer tout particulièrement les enfants des salariés les plus favorisés aux plus défavorisés, ce qui confirme la persistance d'écarts substantiels entre ces origines sociales quant à leurs trajectoires dans l'enseignement supérieur (Albouy & Tavan, 2007). Les enfants d'indépendants sont assez mal représentés sur le plan factoriel – leurs corrélations avec les deux axes sont modérées – même si leurs positions sur l'axe 1 paraissent sociologiquement pertinentes. Seule la position de la variable indiquant la part d'étudiants dont le parent référent exerce une profession intermédiaire apparaît surprenante : elle présente un degré de corrélation avec l'axe 1 légèrement inférieur à celui des ouvriers qualifiés ; cela peut résulter de l'hétérogénéité importante de cette catégorie. Au final, cette opposition très nette entre les milieux sociaux et leurs caractéristiques sur l'axe 1 nous permet d'interpréter celui-ci comme ordonnant les universités selon la composition sociale de leurs publics respectifs : d'origine favorisée à gauche, d'origine populaire à droite. Notons enfin que cette opposition apparaît très discriminante dans le paysage universitaire français : l'axe 1 résume à lui seul 44 % de l'information du nuage initial. Cet axe apparaît donc à de multiples égards comme une mesure pertinente et robuste de la composition des universités.

Figure I

1.3. Des inégalités importantes entre établissements sur le plan statique

L'interprétation du deuxième axe (vertical) est moins immédiate, mais apporte des éléments utiles à l'analyse. Les variables contribuant le plus fortement à cet axe sont, en bas de l'axe, la part d'étudiants titulaires d'un baccalauréat scientifique (S). En haut de l'axe, ce sont les parts d'étudiants titulaires respectivement d'un baccalauréat économique/social (ES) et littéraire (L) qui s'avèrent déterminantes. Le deuxième axe semble donc opposer, à composition sociale donnée, les établissements dont le domaine disciplinaire est à dominante scientifique de ceux à dominante littéraire et/ou économique et social. Cette interprétation est confirmée par le graphique des individus (figure II) : les universités à dominante littéraire apparaissent en haut du graphique là où celles à dominante scientifique, notamment les écoles d'ingénieur rattachées à des universités, apparaissent en bas.

Les positions des variables supplémentaires sur le plan factoriel apportent des informations additionnelles sur la structure des corrélations. Comme ces dernières n'ont pas participé à la construction des axes de l'ACP, leur position et l'interprétation que l'on peut en faire apparaissent d'autant plus pertinentes. Tout d'abord, la proximité importante entre respectivement la part de filles et celle de bacheliers ES/L est notable, de même qu'entre la part de garçons et celle de bacheliers S. Cette proximité

confirme le constat fait, entre autres, par Rosenwald (2006) selon lequel bien que les filles tendent à connaître plus de succès dans leurs études quel que soit le niveau et la filière, elles restent surreprésentées dans les filières à dominante littéraire. Par ailleurs, l'opposition entre les variables indiquant les taux d'inscrits en licence (à droite) et celles indiquant les taux d'inscrits en masters (à gauche) – la qualité de reproduction de la variable associée aux taux de doctorants est faible ce qui ne permet pas son interprétation – est également révélatrice. Toutes les universités ne présentent pas le même rapport entre effectifs de licence et de master, et ce rapport est socialement marqué. Bien que les variables portent ici sur les établissements et non directement sur les étudiants, ce constat semble corroborer l'idée que non seulement les enfants d'origine populaire ont moins de chance d'accéder à l'enseignement supérieur de manière générale, mais que ceux qui y accèdent ont moins de chance d'accéder au grade de master (Peugny, 2013, Selz & Vallet, 2006).

L'ordre observé sur l'axe de composition sociale apparaît sociologiquement pertinent. Ainsi, les établissements qui présentent les compositions étudiantes les plus favorisées sont en général ceux qui peuvent procéder à une sélection et parfois prélever des frais d'inscription importants : les grands établissements (Dauphine, l'IEP de Paris, l'Observatoire de Paris) ainsi que les écoles d'ingénieurs rattachées à des universités (INP Toulouse, Polytech Grenoble, etc.). Toujours à gauche de l'axe, la surreprésentation des universités parisiennes est patente. À l'opposé, à droite du graphique, apparaissent d'abord les universités des DOM, puis des villes moyennes de province, généralement du Nord (Valenciennes, Artois, Haute-Alsace) et du Sud-Est (Perpignan, Avignon, Nîmes, Toulon) de la France. L'ordre des universités sur l'axe de composition sociale recoupe donc largement la distribution des niveaux de vie au sein des différents territoires (telle que décrite par exemple par Auzet *et al.*, 2007), et confirme ainsi des travaux montrant qu'une part substantielle des inégalités entre universités peut être directement reliée à l'inégale distribution sociale au sein des territoires français (Nicourd *et al.*, 2011).

Enfin, l'étalement des positions des établissements tout au long de l'axe 1 de l'ACP amène à faire le constat d'une hétérogénéité sociale importante des établissements (figure II). L'échelle de l'ACP n'ayant pas d'interprétation directe, un retour aux variables initiales permet d'illustrer cette hétérogénéité. Si l'on s'intéresse aux extrémités de l'axe de composition sociale, on note par exemple que la part d'enfants de cadres supérieurs atteint 66 % à Dauphine et à l'IEP de Paris, ce qui traduit une très forte surreprésentation par rapport à la structure socioprofessionnelle de la génération des parents (Marchand, 2010). À l'opposé de l'axe, cette part n'est que de l'ordre de 20 % dans les universités des DOM⁴, et de 25 % dans les universités du Nord et du Sud-Est situées à droite du plan factoriel. Réciproquement, la part d'enfants d'ouvriers non qualifiés atteint 7 % dans les universités d'Avignon et de Nîmes quand ces

⁴ Nous avons vérifié que la restriction du champ aux seuls établissements métropolitains ne modifiait pas les principaux résultats.

derniers sont pratiquement absents de plusieurs établissements parisiens. De même, la part d'enfants d'ouvriers qualifiés dépasse 20 % dans les universités du Nord contre 2 % dans les grands établissements parisiens (Dauphine, IEP Paris, Observatoire de Paris) et 3 à 5 % dans les universités de Paris intramuros (Panthéon-Assas, Paris 7, Paris 5, Paris 4, Panthéon-Sorbonne, etc.).

Figure II

1.4. Une typologie des établissements d'enseignement supérieur français

L'importante dispersion des établissements sur le premier plan factoriel suggère qu'une typologie peut être construite à partir des résultats de l'ACP. Pour ce faire, nous procédons à une classification ascendante hiérarchique (CAH) réalisée à partir des coordonnées des établissements sur les composantes principales de l'ACP⁵.

Nous choisissons de retenir les deux premières dimensions de l'ACP pour construire la typologie. Le premier axe apparaît comme une mesure pertinente de la composition sociale des universités, mais il ne résume cependant pas complètement l'hétérogénéité sociale qui s'observe sur le plan factoriel. La variable indiquant les parts de bacheliers S, par exemple, apparaît significativement corrélée à la fois à l'axe 1 et à l'axe 2, ce qui souligne le rôle de marqueur social du baccalauréat scientifique. Ainsi, si l'axe vertical mesure d'abord la discipline dominante d'une université, il contient également une partie de l'hétérogénéité sociale des établissements qui n'est pas captée par l'axe horizontal. En construisant la typologie à partir des deux premières dimensions, les logiques de différenciation sociale à l'œuvre dans l'enseignement supérieur sont donc plus complètement décrites. Par ailleurs, nous montrons *a posteriori* que les classifications obtenues sont sociologiquement pertinentes (cf. encadré 3).

Au-delà de leur pouvoir descriptif, les typologies que nous présentons vont nous permettre d'appliquer par la suite la mesure de polarisation d'Esteban et Ray (1994) dans le cadre d'une analyse dynamique. Cette mesure repose en effet sur l'existence de groupes bien définis dans la population étudiée. Pour que l'analyse de l'évolution de la polarisation soit justifiée, il faut que les typologies produites conservent leur interprétation sur la période étudiée d'une part, et qu'elles permettent de mettre en lumière une déformation des groupes d'autre part, i.e. les éventuels changements de groupe de certains établissements sur la période étudiée qui se manifestent par leur déplacement sur le plan factoriel. Pour que ces deux conditions soient remplies, nous réalisons la CAH sur les données de toutes les années en même temps. Ainsi, les groupes produits gardent nécessairement la même interprétation au fil des ans,

⁵ Les méthodes factorielles et de classification sont en effet complémentaires : le fait de réaliser la classification directement sur les composantes principales et non à partir des données initiales permet de réduire l'influence du bruit statistique présent dans ces dernières (Husson *et al.*, 2010).

et l'on observe effectivement les changements de groupe de certains établissements sur la période étudiée.

De nombreux critères existent dans la littérature pour guider le choix du nombre de groupes. En pratique, ce choix résulte d'un arbitrage entre parcimonie et homogénéité des groupes, "le critère principal étant que la typologie finalement sélectionnée soit cohérente et porteuse d'enseignements du point de vue de la recherche qui est menée" (Robette, 2011, p. 19). Suivant cette logique, nous choisissons de concentrer l'analyse sur la typologie à quatre groupes issue de la CAH (figure III). De manière schématique, la gauche du plan regroupe des établissements à composition favorisée, majoritairement parisiens, ainsi que des universités de grandes villes de province. La droite du plan est occupée par des établissements dont les publics sont d'origine sociale intermédiaire à populaire, majoritairement situés dans des villes de taille moyenne ainsi que dans les DOM. Par ailleurs, la classification distingue parmi les établissements à composition favorisée ceux dont le domaine disciplinaire principal est littéraire ou économique et social (en haut du plan) de ceux à caractère principalement scientifique (en bas). L'hétérogénéité suivant l'axe 2 est en effet beaucoup plus forte à gauche du plan, entre établissements à composition favorisée, alors que les établissements à population étudiante intermédiaire ou populaire, à droite du plan, présentent une importante homogénéité suivant l'axe 2. Ainsi, on peut distinguer au sein des établissements à compositions favorisées des établissements à dominante scientifique et d'autres à dominante littéraire, alors que les établissements à population étudiante intermédiaire comme populaire ne semblent pas marqués disciplinairement. En cela, il apparaît pertinent de séparer ces différents groupes car ils peuvent être caractérisés par des logiques de polarisation distinctes. Afin de limiter la sensibilité des résultats à la typologie retenue, toutes les analyses qui suivent ont été reproduites à partir des typologies en deux (figure A1-II) et trois (figure A1-III) groupes. Les résultats obtenus sont analogues dans tous les cas.

Figure III

La pertinence de la typologie obtenue est confirmée par la description statistique des différents groupes (voir annexe 2). Sur le plan de la composition sociale, le gradient social est net pour chacune des variables déterminantes de l'axe 1 entre les trois ensembles d'établissements (publics favorisés, publics intermédiaires, publics populaires). Ainsi, la part d'enfants de cadres supérieurs atteint 52-54 % dans les deux groupes des publics favorisés, contre 38 % dans celui des publics intermédiaires et 28 % pour les publics d'origine populaire. À l'inverse, la part d'enfants d'ouvriers qualifiés ne dépasse pas 6 % dans les établissements à composition favorisée contre 13 % en moyenne pour ceux à publics populaires. Ce gradient social s'observe également très clairement lorsque l'on s'intéresse à la répartition des différents niveaux d'étude (L/M/D) au sein des groupes.

L'analyse de la répartition des différents types de baccalauréat – à l'exception des baccalauréats professionnels et technologiques, qui sont principalement des marqueurs sociaux et se comportent donc comme les variables décrites précédemment – montre la pertinence de séparer les groupes favorisés par domaine disciplinaire. La part de titulaires d'un baccalauréat scientifique varie de 48 points entre les deux groupes à composition favorisée. Notons cependant que la part de titulaires d'un baccalauréat scientifique dans le groupe des établissements à composition sociale populaire est aussi faible que dans celui des établissements à composition favorisée et dominante littéraire. Ce constat confirme que le baccalauréat scientifique joue également un rôle de marqueur social important, ce qui justifie notamment le choix de réaliser la typologie sur le plan factoriel – et non sur le seul axe 1 – afin de capter l'ensemble des processus de différenciation sociale. À l'inverse, les parts de titulaires d'un baccalauréat ES ou L sont fortement sous-représentées dans le groupe des compositions sociales favorisées à dominante scientifique et inversement surreprésentées dans les établissements à dominante littéraire. Néanmoins, l'hétérogénéité des groupes relativement à ces types de baccalauréat apparaît moins marquée ; le rôle social de ces types de baccalauréat semble donc moins marqué, corroborant la position des variables associées sur le plan factoriel. Enfin, les variables indiquant la part de femmes et d'hommes sont respectivement très proches de celles indiquant les parts de titulaires d'un bac S et d'un bac ES/L, ce qui est conforme à une différenciation importante des disciplines selon le sexe déjà observée.

Encadré 1 – La base « Système d'Information sur le Suivi de l'Étudiant » (SISE)-----

Présentation de la base

Réalisée par la sous-direction des systèmes d'information et des études statistiques (SIES) du ministère de l'enseignement supérieur, de la recherche et de l'innovation, la base SISE recense depuis 1995 les étudiants inscrits dans les principaux établissements universitaires. Un intérêt majeur de cette base est son caractère exhaustif : SISE recense la *totalité* des inscriptions (principales) prises dans les établissements recensés. En 2007, 1,4 millions d'étudiants sont présents dans SISE sur les 2,2 millions d'étudiants inscrits dans l'enseignement supérieur, soit un taux de couverture d'environ 63%. Dans le cas des universités, toutes les composantes sont prises en compte : IUT, écoles d'ingénieurs rattachées, enseignement à distance, etc. Pour chaque étudiant recensé, on dispose d'informations sur sa scolarité antérieure (type de baccalauréat, année de première inscription dans le supérieur, ...) et actuelle (description de l'établissement d'inscription, du diplôme préparé, du niveau atteint dans le supérieur, ...), ainsi que des informations sociodémographiques (sexe, CS du parent référent, nationalité, ...) et géographiques (pays et commune de résidence de l'étudiant, de ses parents, ...).

Une fois le codage de la CS des parents référents effectué et harmonisé (voir annexe 1) les tables des universités à partir desquelles nous travaillons sur toute la période contiennent 81 établissements et 18 variables construites à partir des tables individus initiales portant sur les étudiants.

-----fin encadré

Encadré 2 – Comparabilité dans le temps de l'indice de composition sociale des universités-----

Le premier axe de l'ACP permet d'ordonner les établissements selon leur composition sociale : des publics d'origine favorisée (à gauche du plan factoriel) aux publics d'origine populaire (à droite). Par

ailleurs, cet axe résume une large part de l'information contenue dans le nuage initial, ce qui renforce sa pertinence. Nous retenons donc la distribution des coordonnées des établissements sur cet axe comme indice synthétique de composition sociale des universités.

Dans la suite de l'article, nous analysons l'évolution de la polarisation de la distribution des compositions sociales entre 2007 et 2015. Pour que cette analyse soit pertinente, il est nécessaire de s'assurer de la comparabilité de cet indice au cours du temps. Bien que la structure des corrélations issue de l'ACP présente une forte stabilité au fil des années – le graphique des corrélations de l'ACP réalisée en 2007 est très semblable à ceux obtenus lorsque nous réalisons une ACP équivalente sur les données des années ultérieures – les positions des variables par rapport aux deux premières composantes principales varient quelque peu, de même que la qualité de la représentation du premier axe. L'évolution des coordonnées des universités dans le temps n'est donc pas immédiatement comparable. Pour pallier cette limite, les axes de l'ACP en 2007 sont fixés – l'ACP n'est réalisée qu'une seule fois, en 2007 – puis les données des années ultérieures sont projetées comme individus supplémentaires sur le plan factoriel de 2007. Les coordonnées des établissements sur le premier axe correspondent à la valeur de leur indice de composition sociale dans le temps. Cette procédure garantit que les axes de l'ACP mesurent la même chose chaque année, et que les valeurs successives de l'indice dans le temps sont comparables.

Une limite de cette analyse est qu'elle donne à l'année de référence – en l'occurrence, 2007 – un rôle singulier : les données des années ultérieures sont analysées par rapport à la structure intrinsèque des données de l'année 2007. Afin de s'assurer que les résultats obtenus ne dépendent pas de manière prépondérante de ce choix, nous avons reproduit l'analyse développée dans la suite de l'article en prenant chaque année de la période étudiée comme année de référence. La tendance observée à la hausse de la polarisation est robuste entre 2009 et 2015 quelle que soit l'année de référence, avec une variabilité plus marquée concernant les deux premières années de la période.

-----fin encadré

2. Une polarisation sociale croissante des établissements d'enseignement supérieur français

Le premier plan factoriel issu de l'ACP met en lumière une hétérogénéité sociale importante des établissements sur le plan statique. La littérature qui traite des évolutions récentes des politiques de l'enseignement supérieur, notamment du recours accru aux indicateurs quantitatifs de performance, suggère qu'il est possible d'observer un mouvement de polarisation du système d'enseignement supérieur (Van Parijs, 2009). Par ailleurs, il est possible que la démocratisation de l'accès à l'université s'accompagne d'une dimension ségrégative, à la manière de celle qui a pu être observée dans l'accès au baccalauréat (Merle, 2002). Par rapport à ces hypothèses, nous proposons une analyse dynamique basée sur deux indices différents : un indice de polarisation issu de la littérature économétrique, et un indice de ségrégation multi-groupes classiquement utilisé dans la littérature sur la ségrégation scolaire.

Nous examinons d'abord l'évolution de l'indice de composition sociale issu de l'ACP entre 2007 et 2015 (Figure IV). Cette analyse a l'avantage de ne dépendre d'aucune typologie. L'écart-type de la distribution – qui constitue de fait une mesure du degré d'inégalité de la distribution des compositions sociales (McKenzie, 2005) – s'accroît de 15 % entre 2007 et 2015. L'écart interquartile augmente quant à lui de 11 %. Enfin, l'écart maximal, i.e. l'écart entre les établissements les plus éloignés sur l'axe de

composition sociale – il s’agit tout au long de la période l’université Paris-Dauphine et l’université de la Polynésie Française – augmente de 18 % entre 2007 et 2015.

Figure IV

Notre première analyse met en évidence un accroissement des inégalités de composition sociale sur la période étudiée. Cependant, elle ne permet pas de déterminer à quels endroits de la distribution s’opèrent les changements. Une manière de contourner cette limite consiste à analyser l’évolution des positions des différents établissements, en fonction de leur groupe, sur le plan factoriel. Pour ce faire, nous représentons l’évolution des barycentres de chaque groupe – i.e. les points moyens de chaque groupe sur les deux dimensions de l’ACP, pondérés par le nombre d’étudiants inscrits dans les différents établissements (figure V). Parmi les établissements à composition sociale favorisée, le groupe à dominante scientifique paraît se fermer, là où celui à dominante littéraire présente une légère ouverture sociale. De même, le groupe des compositions sociales intermédiaires s’ouvre légèrement. Enfin, le groupe d’établissements à composition populaire présente un décrochement continu vers la droite. Au final, il semble donc que l’aggravation des inégalités observée s’explique essentiellement par la dynamique des extrêmes : fermeture du groupe des établissements scientifiques à population étudiante favorisée et ouverture des établissements à population étudiante populaire.

Figure V

Afin d’approfondir sur ce point, nous mobilisons à présent deux indices économétriques permettant de mesurer l’hétérogénéité d’une population divisée en groupe : la ségrégation et la polarisation.

2.1. Une réduction de la ségrégation...

Nous employons tout d’abord l’indice d’entropie normalisé, classiquement utilisé dans la littérature traitant de la ségrégation scolaire et de son évolution. Les indices de ségrégation ont pour point commun de reposer sur la division d’une population au sein de plusieurs catégories, puis répartie dans différentes unités. La distribution observée des catégories au sein des différentes unités est ensuite comparée à la distribution globale de ces catégories dans la population, afin de quantifier l’écart à ce qui serait une distribution homogène des catégories. De nombreux indices ont été développés et appliqués dans la littérature, traduisant la multiplicité des situations auxquelles ces indices peuvent être confrontés – distribution binaire ou multi-catégories, variable utilisée pour caractériser les catégories, choix des unités, etc. Parmi les indices disponibles, nous retenons l’indice d’entropie normalisé H (voir une présentation formalisée en Annexe en ligne), qui est adapté à l’étude de la ségrégation multi-catégories ; cet indice possède un ensemble de propriétés souhaitables pour un tel indice, en particulier, des propriétés statistiques qui permettent la comparaison dans le temps, ainsi que la propriété de décomposabilité additive à la fois selon les catégories et les unités (voir Reardon & Firebaugh, 2002

pour une présentation détaillée de ces propriétés, ainsi qu'une comparaison formalisée avec plusieurs autres indices multi-catégories usuels). Du fait de ces avantages, cet indice a été largement utilisé dans la littérature traitant de la ségrégation multi-catégories en milieu scolaire (voir par exemple Fack, Grenet & Benhenda, 2014 ; Givord *et al.*, 2016 pour des applications à l'enseignement secondaire en France).

Ici, les étudiants sont catégorisés selon leur origine sociale mesurée par la CS du parent référent, et répartis dans les établissements d'enseignement supérieur. Nous mesurons alors comment les CS du parent référent des étudiantes et étudiants se répartissent dans les quatre groupes issus de la CAH. L'indice de ségrégation affiche une baisse relativement lisse à partir de 2010, de l'ordre de 10 % au total (figure VI).

Figure VI

2.2. Alors que la polarisation croît

Nous testons maintenant l'hypothèse d'une polarisation des établissements d'enseignement supérieur français relativement à leur composition sociale. Traditionnellement, la littérature économique aborde ce type de questionnement à l'aide de mesures d'inégalité telles que le coefficient de Gini, afin de déterminer si la concentration d'une distribution (en général, celle du revenu) s'accroît dans le temps. Or, comme le notent Esteban et Ray (1994), dans le cas d'un passage d'une distribution unimodale à une distribution bimodale au cours du temps – i.e. d'une polarisation de deux groupes autour de moyennes locales – les mesures d'inégalité classiques fondées sur le principe de transfert de Pigou-Dalton concluraient à une baisse des inégalités. Pourtant, une telle situation pourrait effectivement caractériser un effet polarisant (les établissements initialement mieux positionnés voient leur avance d'accroître). Le recours aux mesures de polarisation permet d'aller au-delà de cette limite des mesures d'inégalité.

Nous cherchons donc à appréhender une situation où la répartition de la composition sociale et des capitaux scolaires des populations étudiantes dans les universités françaises tend à évoluer d'une situation uni-modale vers une situation bi-modal ou multi-modal, i.e. où la répartition autour de chaque mode tend à se concentrer et/ou les modes s'éloignent les uns des autres. Le cadre économétrique de la polarisation que nous mobilisons a été proposé par Esteban et Ray (1994), présenté de façon plus détaillée en Annexe en ligne. Ici, les groupes sont identifiés a priori grâce à la typologie issue de la CAH, nous utilisons donc la mesure développée dans l'article initial (voir Échevin et Parent, 2002 pour une application de cette mesure à des données de revenu en France).

Plusieurs éléments justifient l'application du cadre de la polarisation à notre objet d'étude. D'abord, les conditions nécessaires pour que les axiomes de la polarisation soient vérifiés apparaissent remplies. En effet, chacune des typologies issues de la CAH produit un petit nombre de groupes (de deux à quatre), homogènes en leur sein mais fortement hétérogènes entre eux comme le montrent les statistiques

descriptives (cf. annexe 2, tableau A2 pour la configuration à quatre groupes). Une condition est que chaque groupe soit "de taille significative". Bien que cela soit assez vague, cela paraît vérifié ici, puisque les tailles relatives des groupes sont proches dans chacune des configurations.

Par ailleurs, la constitution de groupes de pression regroupant des présidents d'université à la fin de la période étudiée nous amènent à penser que les aspects subjectifs de la polarisation pensée par Esteban et Ray (1994) sont également présents. En 2015, 21 présidents d'universités françaises ont signé une tribune collective dénonçant "l'obsession des classements internationaux [qui prévaut] dans les choix de soutien budgétaire, la notion d'excellence scientifique et de masse critique [qui] deviennent centrales au détriment des autres missions de l'université dont la réussite pour tous ou la formation tout au long de la vie"⁶. Cette tribune, qui conduira à l'Alliance des Universités de Recherche Et de Formation (AUREF), regroupe des signataires issus des groupes intermédiaire et populaire de notre typologie. A l'opposé du plan principal, on retrouve les établissements présidés par des signataires de la tribune de 2017 distinguant « quelques universités de recherche qui ont vocation à être compétitives au niveau mondial »⁷, et les membres de la coordination des universités de recherche intensive françaises (CURIF, créée en 2008 et depuis reconfigurée en Udice). Ces exemples illustrent autant le sentiment d'identification qui peut exister entre des établissements présentant des compositions sociales similaires, que l'aliénation qu'ils peuvent ressentir vis-à-vis des établissements qui se situent loin d'eux dans la distribution, pour reprendre les termes d'Esteban et Ray (1994).

L'approche dans le cadre économétrique proposé par Esteban et Ray (1994) pour étudier l'évolution du degré de polarisation sociale entre les établissements (voir Annexe en ligne) conduit à un résultat inverse de celui qui ressortait de l'approche par la ségrégation. En effet, la polarisation sociale des établissements d'enseignement supérieur s'accroît entre 2005 et 2017 (figure VII). L'évolution du degré de polarisation est de l'ordre de 20 % pour notre typologie principale à quatre groupes. Ce résultat est robuste si l'on retient une typologie à 2 ou 3 groupes et il l'est également au choix du paramètre α de sensibilité à la polarisation, i.e. à l'hypothèse que l'on retient sur le degré d'homogénéité des groupes sur le plan statique (voir Annexe en ligne).

Figure VII

⁶ <https://blogs.mediapart.fr/edition/les-invites-de-mediapart/article/290515/quel-avenir-pour-l-enseignement-superieur-et-la-recherche-francais>

⁷ https://lemonde.fr/idees/article/2017/05/30/enseignement-superieur-rapprocher-les-meilleurs-organismes-de-recherche-des-meilleurs-etablisements_5135898_3232.html

Ségrégation et polarisation présentent donc des dynamiques opposées. L'absence d'interprétabilité directe de ces deux indices rend difficile l'explication de ces évolutions divergentes. Un retour aux données brutes sur lesquelles sont basés les calculs de ces indices permet cependant d'avancer deux explications.

Une première possibilité est que cette divergence découle en partie de la différence de concepts que ces deux indices entendent mesurer. La décomposition de l'indice d'entropie selon les groupes issus de la CAH en une composante inter- et une composante intra-groupe montre que la quasi-totalité de la diminution de l'entropie totale sur la période s'explique par la diminution de la composante intra-groupe, là où la composante inter-groupes reste stable (figure VII). Ce constat signifie que l'essentiel de la baisse du niveau de ségrégation observée sur la période s'explique par une homogénéisation des différents groupes d'établissements en termes de composition sociale, et non par un rapprochement des différents groupes entre eux. Les différents établissements d'un même groupe ont donc des populations étudiantes de plus en plus similaires, en termes de CP du parent référent, mais l'écart entre un groupe et un autre reste stable. Une telle dynamique fait précisément partie de celles qu'entend mesurer l'indice de polarisation et qui le distingue des indices d'inégalité et de ségrégation classiques : une telle situation, où les centres des groupes étudiés ne s'écartent pas, mais où, au sein de chaque groupe, les établissements se rapprochent du centre, tend à produire une distribution multimodale plus marquée, ce qui correspond à l'axiome 1 de la polarisation (Esteban & Ray, 1994). Ségrégation et polarisation évoluent conjointement avec l'hétérogénéité entre les centres des groupes, mais de façon opposée avec l'hétérogénéité intra-groupe : lorsque l'hétérogénéité intra-groupe décroît, la ségrégation décroît mais la polarisation augmente.

Une seconde explication, également corroborée par les données, est que ces deux analyses ne reposent pas sur une même définition de l'origine sociale des étudiants. L'indice d'entropie n'est calculé qu'à partir de la CS du parent référent. L'indice synthétique de composition sociale produit par l'ACP – à partir duquel est calculé l'indice de polarisation – repose sur une définition plus large de la composition sociale des établissements, qui inclut les capitaux scolaires. En particulier, le graphique des variables de l'ACP (figure I) montre que le baccalauréat professionnel constitue un marqueur social équivalent au fait d'être enfant d'employé ou d'ouvrier non-qualifié. Or l'analyse précise de l'évolution du profil socio-scolaire des étudiants sur la période montre une augmentation considérable du nombre de titulaires d'un baccalauréat professionnel dans l'enseignement supérieur : leur nombre est multiplié par 2.6 entre 2007 et 2015, là où les évolutions des autres types de baccalauréat restent stables ou d'ampleur limitée (10 %). En déclinant cette évolution selon la typologie issue de la CAH, on observe que l'essentiel de cette augmentation est capté par les établissements appartenant au groupe des compositions sociales populaires, qui initialement concentraient déjà un taux significativement plus élevé de baccalauréats professionnels au sein de leurs effectifs (figure VIII). Ainsi, il semble que l'évolution particulièrement

marquée du barycentre du groupe des établissements à composition sociale populaire (figure V) s'explique en large partie par un accès sur la période 2007-2015 des titulaires d'un baccalauréat professionnel au système d'enseignement supérieur à la fois très conséquent et très inégalement réparti entre les établissements. Cette dynamique, captée par l'analyse de polarisation grâce à l'indice synthétique des capitaux socio-scolaires produits par l'ACP, ne pouvait être mise en évidence via une analyse de ségrégation reposant sur la seule CS du parent référent. C'est le caractère endogène de l'indice de polarisation qui permet de saisir les différentes dynamiques, là où la ségrégation nécessite de choisir a priori les catégories pertinentes (ici le taux de bacheliers professionnels).

Figure VIII

La conclusion de l'analyse dynamique dépend donc directement de l'opérationnalisation du concept utilisé pour mesurer la transformation de la distribution des compositions sociales. L'analyse d'un indice de polarisation conclut à son augmentation, dans la mesure où cette distribution apparaît de plus en plus marquée par l'existence de groupes d'établissements clairement identifiés par les capitaux socio-scolaires de leurs publics. En revanche, l'analyse d'un indice de ségrégation conclut à sa diminution, bien que celle-ci soit avant tout le produit de l'homogénéisation des compositions sociales au sein des groupes issus de la typologie. De manière générale, ces différentes analyses amènent à conclure que l'espace universitaire français, socialement fortement hétérogène sur le plan statique, n'affiche pas d'homogénéisation globale qui refléterait la convergence de la composition des publics étudiants dans les établissements sur la période étudiée.

Annexe en ligne 1 – L'indice d'entropie normalisé

Formellement, l'indice d'entropie normalisé est défini de la manière suivante :

$$H = \sum_{m=1}^M \sum_{j=1}^J \frac{t_j}{TE} \pi_{jm} \ln \frac{\pi_{jm}}{\pi_m}$$

où t_j correspond au nombre d'étudiants inscrits dans l'établissement j , T au nombre total d'étudiants, π_{jm} à la proportion d'étudiants inscrits dans l'établissement j et dont le parent référent a pour CS m , et π_m la proportion d'étudiants dont le parent référent a pour CS m dans la population totale. L'indice d'entropie correspond donc à une moyenne pondérée (par la taille de chaque établissement) des écarts entre la distribution des compositions sociales dans les établissements et celle de la population étudiante complète, normalisée par la taille de cette dernière.

Afin de faire le lien entre les résultats obtenus avec l'indice de polarisation et l'indice de ségrégation, nous décomposons ce dernier selon les quatre groupes d'établissements issus de la CAH. Cette opération est rendue possible par la propriété de décomposabilité de l'indice de ségrégation selon les unités. Si les J unités peuvent être regroupées dans K groupes, l'indice se réécrit :

$$H = H_{INTER} + \sum_{k=1}^K \frac{t_k E_k}{TE} H_k$$

où H_{INTER} est l'indice de ségrégation inter-groupes et les $\{H_k, k = 1, \dots, K\}$ correspondent aux indices de ségrégation intra-groupes.

Comme pour l'indice de polarisation ci-dessous, la valeur de l'indice de ségrégation n'a pas de sens en elle-même : s'il est normalisé entre 0 et 1, ses propriétés mathématiques font qu'il est en pratique très rare d'observer des valeurs élevées de l'indice. C'est donc dans la mesure de l'évolution de la ségrégation qu'il s'avère particulièrement pertinent. Pour une présentation très détaillée et pédagogique des propriétés de l'indice d'entropie normalisé, nous invitons le lecteur à se référer à Givord *et al.* (2016).

Annexe en ligne 2 – La mesure de polarisation d'Esteban et Ray

Un outil qui se distingue des mesures d'inégalités traditionnelles

Présentation du cadre de la polarisation

La polarisation est bien plus proche de la ségrégation que des mesures d'inégalité de type Gini. Mais la première s'intéresse à l'émergence de modes au sein d'une distribution unique, alors que la seconde traite de catégories prédéfinies distinctes (typiquement des catégories raciales) dont les membres sont répartis dans des dispositifs prédéfinis (typiquement des écoles) et non autour de valeurs distinctes dans la distribution d'une variable continue commune. On pourrait ainsi parler de ségrégation entre PCS populaires et PCS favorisées (ou entre bacs généraux et bacs technologiques et professionnels) dans des groupes d'universités différentes.

Le cadre économétrique de la polarisation que nous mobilisons a été proposé par Esteban et Ray (1994). Depuis, plusieurs auteurs ont développé des variantes de celle-ci – notamment dans les cas de groupes non identifiés et/ou définis de manière continue – et l'ont appliqué à différentes populations (Wang & Tsui, 2000 ; Duclos *et al.*, 2004 ; Anderson, 2011).

La mesure d'Esteban et Ray est basée sur une axiomatisation de la polarisation que nous avons reprise dans la définition ci-dessus. Ces axiomes imposent trois conditions sur la distribution étudiée : un fort degré d'homogénéité au sein des groupes ; un fort degré d'hétérogénéité entre les différents groupes ; un faible nombre de groupes chacun de taille significative. Dans leur article, les auteurs considèrent une distribution de revenus. Cependant, ils notent que "plus généralement, même si l'espace des attributs est multidimensionnel, l'analyse n'est pas perturbée tant que les différences entre les attributs peuvent être représentés par une métrique" (Esteban et Ray, 1994, p. 823 ; notre traduction de l'anglais). Cette remarque caractérise bien notre cas : l'attribut multidimensionnel correspond à la composition sociale des universités, que nous résumons par un indice synthétique construit à partir de l'ACP.

Nous présentons à présent le cadre formel tel que développé par Esteban et Ray pour définir la mesure de polarisation. Nous justifions ensuite la pertinence de ce cadre relativement à notre objet d'étude. Soit $(\pi, y) \equiv (\pi_1, \dots, \pi_n, y_1, \dots, y_n)$ une distribution, où $\pi_i > 0$ est le poids attribué au groupe i – dans notre cas, la part d'étudiants inscrits dans les établissements du groupe i – et $y_i \in \mathbb{R}$ la moyenne de l'attribut dans le groupe i – dans notre cas, l'indice de composition sociale. Soit $I(p)$ une fonction d'"identification", qui traduit le fait qu'un individu ressent un sentiment d'identification vis-à-vis des individus qui ont un indice proche du sien. Soit $a(\delta(y_i, y_j))$ une fonction d'"aliénation", qui représente l'aliénation que peut ressentir un individu par rapport à ceux qui sont "loin" de lui dans la distribution de l'indice. Enfin, soit $T(I, a)$ une fonction d'"antagonisme effectif", qui traduit le fait que le sentiment d'identification d'un individu peut influencer l'expression de son aliénation. La polarisation totale est alors définie comme la somme de tous les antagonismes effectifs :

$$P(\pi, y) = \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j T[I(\pi_i), a(\delta(y_i, y_j))]$$

Les auteurs montrent que pour que cette mesure respecte les axiomes mentionnés précédemment, elle doit être de la forme :

$$P^*(\pi, y) = K \sum_{i=1}^n \sum_{j=1}^n \pi_i^{1+\alpha} \pi_j |y_i - y_j|$$

Avec $K > 0$ une constante et $\alpha \in (0; 1.6]$.

L'expression P^* correspond à l'indice de polarisation d'Esteban et Ray. Cette mesure est directement calculable en choisissant les valeurs de K et α . La constante K peut être utilisée pour normaliser la population, mais n'a aucune influence sur l'ordre lors de la comparaison des degrés de polarisation de différentes distributions. Nous la fixons à l'unité. Le paramètre α peut être interprété comme un degré de "sensibilité à la polarisation" de la population, qui dépend du degré d'homogénéité des groupes. Si $\alpha = 0$, les groupes sont supposés parfaitement homogènes et la sensibilité à la polarisation est donc maximale. De fait, la sensibilité à la polarisation de la population est difficile à évaluer *a priori*. Une pratique commune dans la littérature consiste à calculer la mesure pour différentes valeurs de α afin de tester la robustesse des résultats à des variations de ce paramètre.

Enfin, notons que pour $\alpha = 0$, la mesure de polarisation P^* correspond de fait à un indice de Gini calculé entre les groupes, à une constante multiplicative près. Comme le remarquent les auteurs, "c'est précisément le fait que les poids des groupes soient élevés à une puissance supérieure à un qui est à l'origine du comportement différent de la mesure de polarisation" par rapport aux mesures d'inégalités classiques. Plus la valeur de α est élevée, plus cette différence de comportement est marquée.

3. Les « politiques d'excellence » de l'enseignement supérieur : une redistribution inversée ?

Nos résultats indiquent que les établissements d'enseignement supérieur se caractérisent par une forte hétérogénéité sociale sur le plan statique. Par ailleurs, cette hétérogénéité paraît s'accroître entre 2007 et 2015. Il apparaît cependant difficile d'identifier les causes d'un tel processus, qui est vraisemblablement multifactoriel.

Par exemple, nous avons montré que, sur le plan statique, l'hétérogénéité des compositions sociales reflétait en partie l'inégale distribution de celles-ci au sein des différents territoires. Il est donc possible qu'une augmentation des inégalités sociales entre les territoires se répercute mécaniquement sur la composition sociale des universités associées. Cette explication suppose néanmoins une faible mobilité des étudiants par rapport à leur domicile familial. Or dans les faits, la propension à la mobilité des étudiants n'est pas négligeable, et dépend elle-même de l'origine sociale. En 2007, dans les groupes d'établissements dont les compositions sociales sont intermédiaires ou populaires, environ un étudiant sur quatre est inscrit dans un établissement donc l'académie ne correspond pas au lieu de résidence de ses parents. Dans les établissements accueillant des publics favorisés, ce même taux monte à 40 % dans le groupe à dominante scientifique, et 50 % dans le groupe à dominante littéraire. L'existence de cette mobilité rend difficile de conclure à une transmission mécanique de l'évolution des inégalités territoriales, Par ailleurs, la distribution des compositions sociales au sein des territoires apparaît

relativement stable sur la période étudiée. Pour arriver à ce constat, nous reprenons l'indice d'entropie normalisé – qui permet utilement de quantifier l'évolution du degré de séparation entre plusieurs groupes sur un territoire – et l'appliquons aux distributions des catégories socioprofessionnelles par académie, calculées à partir des données des recensements récents. On observe ainsi une augmentation de 2 % de l'indice entre 1999 et 2010, suivi d'une diminution de 3,8 % entre 2010 et 2015, des évolutions qui ne semblent pas suffisantes pour expliquer les dynamiques observées du recrutement social dans l'enseignement supérieur.

Une polarisation accrue est une des prédictions théoriques émergeant de la littérature qui analyse la portée performative des dispositifs qui orientent en partie les politiques récentes de l'enseignement supérieur : initiatives d'excellence (I dex), utilisation des palmarès universitaires pour justifier les regroupements en Communautés d'universités et établissements (COMUEs), etc. (Espeland & Sauder, 2007 ; Van Parijs, 2009 ; Halffman & Leydesdorff, 2010 ; Brusadelli & Lebaron, 2012 ; Paradeise & Thoenig, 2015). Cependant, la causalité entre ces deux phénomènes apparaît difficile à établir empiriquement, d'autant qu'une causalité inverse est possible : par exemple, une composition sociale favorisée pourrait « attirer » les initiatives d'excellence dans la mesure où le niveau moyen des étudiants dans ces établissements serait plus élevé. Dans cette partie, nous nous intéressons au lien entre la composition sociale des établissements et le fait d'être distingué par de tels dispositifs. Nous nous concentrons particulièrement sur le cas des I dex, qui apparaissent comme un dispositif majeur – du fait des avantages financiers en jeu – des politiques de l'enseignement supérieur au cours de la période étudiée. Compte tenu de l'hétérogénéité sociale mise en évidence dans les précédentes analyses, cette relation pose de manière forte la question de l'équité et de la redistribution des ressources au sein de l'enseignement supérieur.

Engagés par la loi de finances rectificative du 9 mars 2010, les "investissements d'avenir" (PIA) ont pour objectif de soutenir des projets innovants dans un certain nombre de secteurs jugés "priorités nationales". La recherche y occupe une place majeure : la dotation accordée aux I dex représente 7,7 milliards d'euros, ce qui en fait de loin le poste le plus important⁸. En 2018, à l'issue des deux premières vagues des investissements d'avenir (PIA1 et PIA2), 10 établissements sont porteurs du label I dex – dont 3 en période probatoire : les projets de Saclay, Paris Sciences et Lettres et Université de Paris. Par ailleurs, les PIA2 ont également distingué parmi les I dex des "Initiatives Science - Innovation - Territoires - Economie" (I-SITE), dont la capacité d'innovation est reconnue mais qui n'ont pas vocation à devenir

⁸Voir le *Rapport relatif à la mise en œuvre et au suivi des investissements d'avenir* de 2015, disponible à l'adresse : https://www.performance-publique.budget.gouv.fr/sites/performance_publique/files/farandole/ressources/2015/pap/pdf/jaunes/jaune2015_investissements_avenir.pdf

des universités de rang mondial, et qui par conséquent reçoivent des dotations substantiellement plus faibles que les IDEX.

Nous représentons sur le plan factoriel les universités du champ couvert qui appartiennent à un regroupement – généralement, une COMUE – détenant le label IDEX (figure IX). Pour ce faire, nous avons dû construire *ex ante* l'Université Clermont-Auvergne (issue de la fusion en 2017 des universités Clermont I et Clermont II) à partir des données 2015 de la base SISE, afin de rendre notre périmètre comparable aux universités bénéficiant du label IDEX.

Figure IX

Plusieurs observations peuvent être formulées à partir de cette représentation. D'abord, les universités membres d'un établissement IDEX se situent pour la plupart dans le cadran de gauche du plan factoriel, et correspondent donc à des universités dont les populations étudiantes sont socialement favorisées. En réalité, ce fait est même certainement sous-estimé dans notre représentation : en pratique, ces universités sont généralement regroupées au sein de COMUEs avec des grandes écoles et des établissements de recherche. Si l'on disposait des données nécessaires pour reconstituer complètement les composantes d'enseignement de ces regroupements, il est très vraisemblable qu'ils se déplaceraient encore vers la gauche du plan. C'est notamment le cas pour l'ensemble des universités qui se situent sur le cadran de droite : l'université d'Évry-Val-d'Essonne (UEVE, située à droite du plan) fait partie de l'Université Paris-Saclay aux côtés de l'Université de Versailles-Saint-Quentin-en-Yvelines, de Paris 11 et de grandes écoles (ENS Paris-Saclay, AgroParisTech, CentraleSupélec, ...). Lyon 2 et Lyon 3 font partie de la COMUE Université de Lyon aux côtés de Lyon 1 et de grandes écoles (ENS Lyon, IEP de Lyon, EM Lyon, ...), etc.

On remarque par ailleurs que la hiérarchie des IDEX s'ordonne assez naturellement sur l'axe de composition sociale de l'ACP. Si l'on reprend les classifications issues de la CAH, les IDEX appartiennent très majoritairement aux groupes d'établissements dont les publics sont d'origine sociale favorisée, les I-SITE aux groupes à composition sociale intermédiaire, mais aucun établissement appartenant au groupe à composition sociale populaire ne détient le label IDEX – à l'exception de l'université d'Évry-Val-d'Essonne (UEVE), mais ce cas relève plus de l'artefact statistique pour les différentes raisons évoquées précédemment.

Ainsi, les labels d'excellence et les financements exceptionnels correspondants sont attribués à des établissements qui concentrent les populations étudiantes les mieux dotées, reproduisant au sein même du système universitaire une anti-redistributivité des moyens publics déjà bien connue entre ce système universitaire et le champ classes préparatoires-grandes écoles (Flacher *et al.*, 2018). D'autre part, la mise en concurrence des établissements pour l'obtention de ces labels se révèle sélectionner suivant la

composition sociale des populations étudiantes, alors même qu'elle ne se base que sur une évaluation des projets institutionnels et scientifiques des établissements.

Un constat analogue peut être fait à propos des principaux palmarès universitaires, dont l'utilisation croissante dans la justification des politiques de l'enseignement supérieur est notée dès 2007 (Hazelkorn, 2007). Ces derniers sont notamment largement utilisés dans le cadre des appels à projets liés aux initiatives d'excellence : parmi les vingt lauréats des initiatives Idex et I-SITE, quatorze justifient leur projet par l'opportunité d'améliorer leur position dans le classement de Shanghai (Charpin *et al.*, 2017). Sans surprise, la projection des établissements apparaissant au classement de Shanghai en 2015 sur le plan factoriel recoupe largement celle des initiatives d'excellence, en particulier les établissements titulaires d'un Idex (figure X). Ainsi, ces palmarès, qui ne mesurent que très imparfaitement l'excellence universitaire du fait d'importantes faiblesses méthodologiques (Billaut *et al.*, 2010), s'avèrent également distinguer principalement des établissements dont la composition sociale est favorisée. On peut alors se demander dans quelle mesure ces différents dispositifs permettent de stimuler une réelle concurrence et d'attribuer davantage de moyens aux établissements les plus méritants, ou bien au contraire s'ils ne viennent pas conforter des établissements qui ne prennent en charge que marginalement la massification de l'enseignement supérieur.

Figure X

Conclusion

À partir des bases exhaustives des inscriptions étudiantes, nous montrons que le système universitaire français se caractérise par une forte hétérogénéité sociale. Si l'enseignement supérieur poursuit le processus d'ouverture aux jeunes issues des classes populaires observé depuis plusieurs décennies, cette ouverture ne permet pas d'accroître la mixité sociale dans les universités. Les indicateurs que nous mobilisons suggèrent que l'hétérogénéité sociale observée n'a pas diminué significativement entre 2007 et 2015. Nous montrons aussi que les initiatives d'excellence (Idex), dispositif majeur des politiques récentes de l'enseignement supérieur, distinguent prioritairement des établissements dont la composition sociale est favorisée. Dans le contexte d'une forte massification scolaire qui n'est pas uniformément supportée par les établissements, ce constat justifie de repenser en profondeur la répartition des financements dans l'enseignement supérieur.

D'autres évolutions contemporaines laissent penser que cette polarisation pourrait se poursuivre dans les années qui viennent. Après des hésitations sur les statuts des regroupements universitaires (regroupements, pôles, communautés, etc.), la rédaction de statuts d'exception pour les Idex au second semestre de l'année 2018 pourrait traduire dans le code de l'éducation l'existence de deux types distincts d'universités en France. En lien avec cette évolution légale, la possibilité pour les universités disposant

du statut de grand établissement de mettre en place des frais d'inscription bien plus élevés que les autres peut renforcer la polarisation sociale, à la fois parce qu'elle joue le rôle de barrière à l'entrée dans les établissements à population étudiante socialement favorisée et parce qu'elle accroît les différences de moyens entre les établissements des deux groupes.

Plusieurs approfondissements peuvent-être envisagés à partir de cette étude. En remontant dans le temps, la présence d'une période d'ouverture sociale du système universitaire sans polarisation, ou moins forte, pourrait être testée. L'extension progressive du périmètre de la base SISE universités risque néanmoins de réduire la faisabilité d'une telle étude. Nous envisageons une extension vers les établissements sous la tutelle des autres ministères, qui pourrait être plus fructueuse et permettrait de tester l'hypothèse d'une convergence des universités vers les grandes écoles. La lecture que nous proposons des IDEX s'en verrait renforcée. Enfin, une analyse multi-échelle est également envisageable : au sein même des établissements, peut-on observer une polarisation sociale interne entre Licence et Master par exemple, entre disciplines ou entre différents campus d'une même université ? Ou au cours du cycle Licence, à l'occasion de l'arrivée en L3 des étudiants issus des CPGE ou dans les doubles licences et autres dispositifs dérogatoires. On pourrait alors décomposer les rôles respectifs de l'établissement, du campus, de la discipline, du cycle, etc. dans la mixité ou la polarisation sociales vécues par la jeunesse étudiante.

Bibliographie

Albouy, V. & Tavan, C. (2007). Accès à l'enseignement supérieur en France: une démocratisation réelle mais de faible ampleur. *Économie et Statistique*, 410, 3–22. <https://doi.org/10.3406/estat.2007.7053>

Albouy, V. & Wanecq, T. (2003). Les inégalités sociales d'accès aux grandes écoles. *Économie et Statistique*, 361, 27–52. <https://doi.org/10.3406/estat.2003.7351>

Amsler, S.-S., & Bolsmann, C.-H. (2012). University ranking as social exclusion. *British Journal of Sociology of Education*, 33, 283–301. <https://doi.org/10.1080/01425692.2011.649835>

Anderson, G. (2011). Polarization measurement and inference in many dimensions when subgroups can not be identified. *Economics: The Open-Access, Open-Assessment E-Journal*, 5, 1–19. <https://dx.doi.org/10.5018/economics-ejournal.ja.2011-11>

Auzet, L. & Février, M. (2007). Niveaux de vie et pauvreté en France : les départements du Nord et du Sud sont les plus touchés par la pauvreté et les inégalités. *Insee Première* N°1162. <http://www.epsilon.insee.fr/jspui/bitstream/1/91/1/ip1162.pdf>

Billaut, J.C., Bouyssou, D. & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84, 237–263. <https://akjournals.com/view/journals/11192/84/1/article-p237.xml>

Boutchénik, B., Coron, C., Grobon, S., Goffette, C. & Vallet, L.-A. (2015). Quantifier l'influence totale de la famille d'origine sur le devenir scolaire et professionnel des individus. *Économie et statistique*, 5–23. <https://doi.org/10.3406/estat.2015.10532>

- Brinbaum, Y., Hugrée, C. & Poullaouec, T. (2018).** 50% à la licence... mais comment? Les jeunes de familles populaires à l'université en France. *Économie et Statistique* 499, 79–105. <https://doi.org/10.24187/ecostat.2018.499s.1941>
- Brusadelli, N. & Lebaron, F. (2012).** Les indicateurs de "performance" universitaire. *Savoir/Agir*, 4, 97–104. <https://doi-org.ezproxy.universite-paris-saclay.fr/10.3917/sava.022.0097>
- Charpin, J.-M., Filliatreau, G. & Aimé, P. & Dulbecco, P. & Foucault, M. (2017).** La prise en compte des classements internationaux dans les politiques de site. Rapport de l'IGF et de l'IGAENR. <https://www.enseignementsup-recherche.gouv.fr/cid119176/la-prise-en-compte-des-classements-internationaux-dans-les-politiques-de-site.html>
- Convert, B. (2006).** *Les impasses de la démocratisation scolaire*. Paris: Raisons d'Agir.
- Duclos, J.-Y., Esteban, J.-M. & Ray, D. (2004).** Polarization: concepts, measurement, estimation. *Econometrica*, 72, 1737–1772. <https://www.econstor.eu/bitstream/10419/49713/1/66798416X.pdf>
- Duru-Bellat, M. & Kieffer, A. (2008).** Du baccalauréat à l'enseignement supérieur en France: déplacement et recomposition des inégalités. *Population*, 63, 123–157. <https://doi.org/10.3917/popu.801.0123>
- Échevin, D. & Parent, A. (2002).** Les indicateurs de polarisation et leur application à la France. *Économie & prévision*, 155, 13–30. <https://doi.org/10.3917/ecop.155.0013>
- Espeland, W.-N. & Sauder, M. (2007).** Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113, 1–40. <https://doi.org/10.1086/517897>
- Esteban, J.-M. & Ray, D. (1994).** On the measurement of polarizarion. *Econometrica*, 62, 819–851. <https://doi.org/10.2307/2951734>
- Fack, G., Grenet, J. & Benhenda, A. (2014).** L'impact des procédures de sectorisation et d'affectation sur la mixité sociale et scolaire dans les lycées d'Île-de-France. Institut des Politiques Publiques, *Rapport IPP N°3*. <https://hal.archives-ouvertes.fr/hal-02464102/>
- Falcon, J., & Bataille, P. (2018).** « Equalization or Reproduction? Long-Term Trends in the Intergenerational Transmission of Advantages in Higher Education in France ». *European Sociological Review* 34(4), 335-47. <https://doi.org/10.1093/esr/jcy015>
- Flacher, D., Harari-Kermadec, H. & Moulin, L. (2018).** Régime par répartition dans l'enseignement supérieur : fondements théoriques et estimations empiriques. *Économie et institutions* [En ligne], 27. <https://doi.org/10.4000/ei.6233>
- Frouillou, L. (2017).** Des publics hétérogènes ? Tableau des étudiants en Géographie des universités françaises. *Carnets de géographes*, 10, 1–17. <https://doi.org/10.4000/cdg.1129>
- Givord, P. & Guillerm, M. & Monso, O. & Murat, F. (2016).** La ségrégation sociale entre les collèges : Quelles différences entre public et privé, aux niveaux national, académique et local ? *Éducation et Formations*, Massification scolaire et mixité sociale, 53-76. <https://hal.archives-ouvertes.fr/halshs-01447198/>
- Hazelkorn, E. (2007).** L'impact du classement des établissements sur la prise de décision dans l'enseignement supérieur. *Politiques et gestion de l'enseignement supérieur*, 19, 95–122. <https://doi.org/10.1787/hemp-v19-art12-fr>
- Halfman, W. & Leydesdorff, L. (2010).** Is Inequality Among Universities Increasing? Gini Coefficients and the Elusive Rise of Elite Universities. *Minerva*, 48, 55–72. <https://arxiv.org/abs/1001.2921>
- Husson, F., Josse, J. & Pagès, J. (2010).** Principal component methods – hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? Agrocampus, *Rapport technique*. http://www.sthda.com/english/upload/hcpc_husson_josse.pdf

- Ichou, M. & Vallet, L.-A. (2013).** Academic achievement, tracking decisions, and their relative contribution to educational inequalities : Change over four decades in France. *In* M. Jackson (Ed.), *Determined to succeed? Performance versus choice in educational attainment*, 116-148. Stanford University.
- Jauneau, Y. (2009).** Les employés et ouvriers non qualifiés. *Insee Première* N° 1250. <http://www.epsilon.insee.fr/jspui/bitstream/1/672/1/ip1250.pdf>
- Marchand, O. (2010).** 50 ans de mutations de l'emploi. *Insee Première* N° 1312. <http://www.epsilon.insee.fr/jspui/bitstream/1/583/1/ip1312.pdf>
- McKenzie, D.-J. (2005).** Measuring Inequality with Asset Indicators. *Journal of Population Economics*, 18, 229–260. <https://doi.org/10.1007/s00148-005-0224-7>
- Merton, R.-K. (1968).** The Matthew effect in science. *Science*, 159, 56–63. <https://doi.org/10.1126/science.159.3810.56>
- Nicourd, S., Samuel, O. & Vilter, S. (2011).** Les inégalités territoriales à l'université: effets sur les parcours des étudiants d'origine populaire. *Revue française de pédagogie*, 176, 27–40. <https://doi.org/10.4000/rfp.3153>
- Paradeise, C. & Thoenig, J. C. (2015).** *In Search of Academic Quality*. London: Palgrave MacMillan.
- Peugny, C. (2013).** *Le destin au berceau. Inégalités et reproduction sociale*. Paris: Seuil.
- Place, D. & Vincent, B. (2009).** L'influence des caractéristiques sociodémographiques sur les diplômes et les compétences. *Économie et statistique*, 424–425, 125–147. <https://doi.org/10.3406/estat.2009.8034>
- Reardon, S.-F. & Firebaugh, G. (2002).** Measures of Multigroup Segregation. *Sociological Methodology*, 32, 33-67. <https://doi.org/10.1111/1467-9531.00110>
- Robette, N. (2011).** Explorer et décrire les parcours de vie : les typologies de trajectoires. Paris : CEPED.
- Rocher, T. (2016).** Construction d'un indice de position sociale des élèves. *Éducation et formations*, 5–27. <https://hal.archives-ouvertes.fr/hal-01350095/>
- Rosenwald, F. (2006).** Filles et garçons dans le système éducatif depuis vingt ans. Insee, *Données sociales : La société française*, 87-94. <https://www.insee.fr/fr/statistiques/fichier/1371989/donsoc06i.pdf>
- Selz, M. & Vallet, L.-A. (2006).** La démocratisation de l'enseignement et son paradoxe apparent. Insee, *Données sociales: La société française*, 101-107. <https://www.insee.fr/fr/statistiques/fichier/1371955/donsoc06k.pdf>
- Truong, F. (2015).** Comment les "nouveaux étudiants" s'ajustent aux études supérieures par le bas. *Diversité*, 181, 68–73.
- Van Parijs, P. (2009).** European higher education under the spell of university rankings. *Ethical perspectives*, 16, 189–206. <https://doi.org/10.2143/EP.16.2.2041651>
- Vyas, S. & Kumaranayake, L. (2006).** Constructing Socio-Economics Status Indices: How to Use Principal Components Analysis. *Health Policy and Planning*, 21, 459-468. <https://doi.org/10.1093/heapol/czl029>
- Wang, Y.-Q. & Tsui, K.-Y. (2000).** Polarization Orderings and New Classes of Polarization Indices. *Journal of Public Economic Theory*, 2, 349–363. <https://doi.org/10.1111/1097-3923.00042>

Annexes

1. Codage de la CS des parents référents et harmonisation

Choix de la nomenclature de codage de la catégorie socio-professionnelle du parent référent

L'information sur la CS du parent de référence de l'étudiant est codée dans SISE dans une nomenclature proche de celles des PCS à deux chiffres de l'Insee. Compte tenu de la taille de notre échantillon – l'analyse portant non sur les étudiants mais sur les 81 établissements d'enseignement supérieur – il est impossible d'inclure directement l'ensemble de ces modalités comme variables dans l'ACP. La nomenclature des PCS à un chiffre n'apparaît cependant pas non plus entièrement satisfaisante : certaines catégories présentent une hétérogénéité importante, en particulier en termes de capitaux sociaux et culturels, qui pourrait masquer des différences marquées de trajectoires dans l'enseignement supérieur. Aussi, nous employons une version affinée de cette dernière, en procédant à des découpages fréquemment employés par les sociologues de la stratification sociale. Nous séparons notamment les chefs d'entreprise de dix salariés ou plus (CS 23) des artisans-commerçants (CS 21 et 22). Nous distinguons également les ouvriers qualifiés (CS 61 dans SISE) des non qualifiés (CS 66 et 69 dans SISE). Au même titre, il aurait sans doute été pertinent de distinguer le niveau de qualification des employés, mais cela n'est possible qu'à partir de nomenclatures plus détaillées (Jauneau, 2009).

Modifications apportées au champ de la base

Le codage des CS du parent référent dans SISE nous contraint à exclure des individus de notre analyse. D'abord, les catégories de retraités (CS 71 à 76 dans SISE) sont agrégées : par exemple, la CS 76 correspond aux retraités employés et ouvriers. Il est donc impossible de ventiler ces catégories dans la nomenclature des CS que nous retenons. Aussi, nous excluons du champ les étudiants dont le parent de référence est retraité. Par ailleurs, nous excluons également les individus dont le parent de référence est codé comme "chômeur n'ayant jamais travaillé" (trop peu d'observations pour créer une variable associée à cette modalité), "autre personne sans activité de référence" (catégorie imprécise et potentiellement hétérogène), ou dont la CS n'est pas renseignée.

Afin d'assurer la comparabilité de nos analyses dans le temps, nous devons procéder à des restrictions du champ couvert par la base. D'abord, certaines composantes universitaires – telles que les IUFM de Martinique, Guadeloupe et Guyane, le centre universitaire de Mayotte ainsi que le PRES de Paris-Est – apparaissent ou se modifient de manière importante au cours de la période étudiée, nous les supprimons donc afin de travailler à champ constant. De même, entre 2007 et 2008, certaines écoles d'ingénieur rattachées à un établissement passent dans le champ des écoles d'ingénieurs non universitaires, et ne sont donc plus présentes dans les tables SISE universités.

Plusieurs universités fusionnent en une entité unique sur la période. Pour assurer la comparabilité de nos résultats dans le temps, nous choisissons de reconstruire ces fusions *ex-ante*. Par exemple, l'Université de Bordeaux est créée en 2014 suite à la fusion des universités Bordeaux I, Bordeaux II et Bordeaux IV. Nous reconstruisons donc cet établissement agrégé dès 2007 en regroupant chaque année entre 2007 et 2013 les effectifs de ses trois composantes. Une telle opération n'est pas neutre sur le plan statistique : sur le plan factoriel de l'ACP, les fusions ont en général un effet "moyennisant" – l'établissement qui en résulte tend à se trouver au centre du plan factoriel – puisqu'elles agrègent de fait des établissements qui auparavant pouvaient avoir des positions très différentes sur ce plan. Cependant, du fait de leur nombre et de leur importance, supprimer les établissements issus de fusions produirait un biais de représentativité important.

2. Description des groupes issus de la CAH en 2007 (typologie à quatre groupes)

Tableau A2

Pourcentages

Groupes	Publics favorisés	Publics favorisés	Publics intermédiaires	Publics populaires	Ensemble
---------	-------------------	-------------------	------------------------	--------------------	----------

	(dominante scientifique)	(dominante littéraire)			
Homme	50	37	43	41	43
Femme	50	63	57	59	57
Bac S	77	29	48	31	48
Bac ES	10	34	24	26	23
Bac L	5	28	15	22	16
Bac Pro	0	1	1	3	1
Bac Techno	8	8	12	19	12
Agriculteur	2	1	3	3	2
Art.-Comm.	6	6	6	7	6
Chef d'ent.	3	4	2	2	2
Cadre sup.	52	54	38	28	41
Prof. inter.	17	13	19	20	18
Employé	13	14	17	22	17
Ouvrier qual.	6	6	11	13	10
Ouvrier non qual.	2	2	3	5	3
Licence	53	61	67	79	66
Master	42	34	30	20	31
Doctorat	5	5	3	2	3
Nb. Etudiants	136656	122464	581775	115510	956405

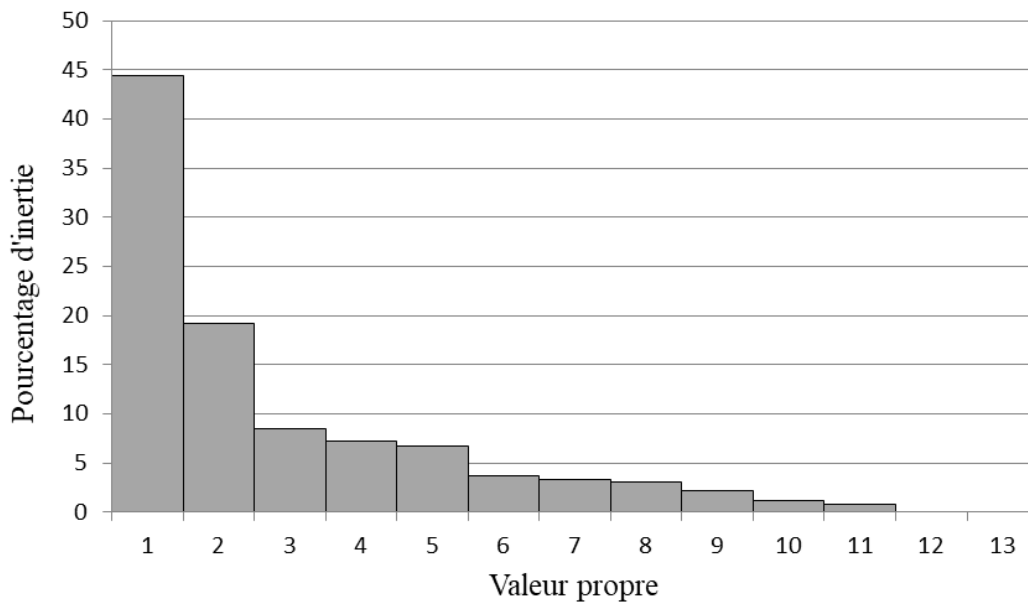
Lecture : en 2007, les établissements à dominante scientifique dont les publics sont d'origine sociale favorisée comptent 50 % d'hommes, 77 % de titulaires d'un baccalauréat scientifique, 2 % d'étudiants dont le parent référent est agriculteur et 53 % d'étudiants inscrits en licence.

Champ : universités françaises et grands établissements placés sous la tutelle du ministre chargé de l'enseignement supérieur

Source : *MESRI-SIES, SISE*

3. L'ACP et la typologie

Figure A3-I – Variance expliquée par les différentes dimensions de l'ACP en 2007

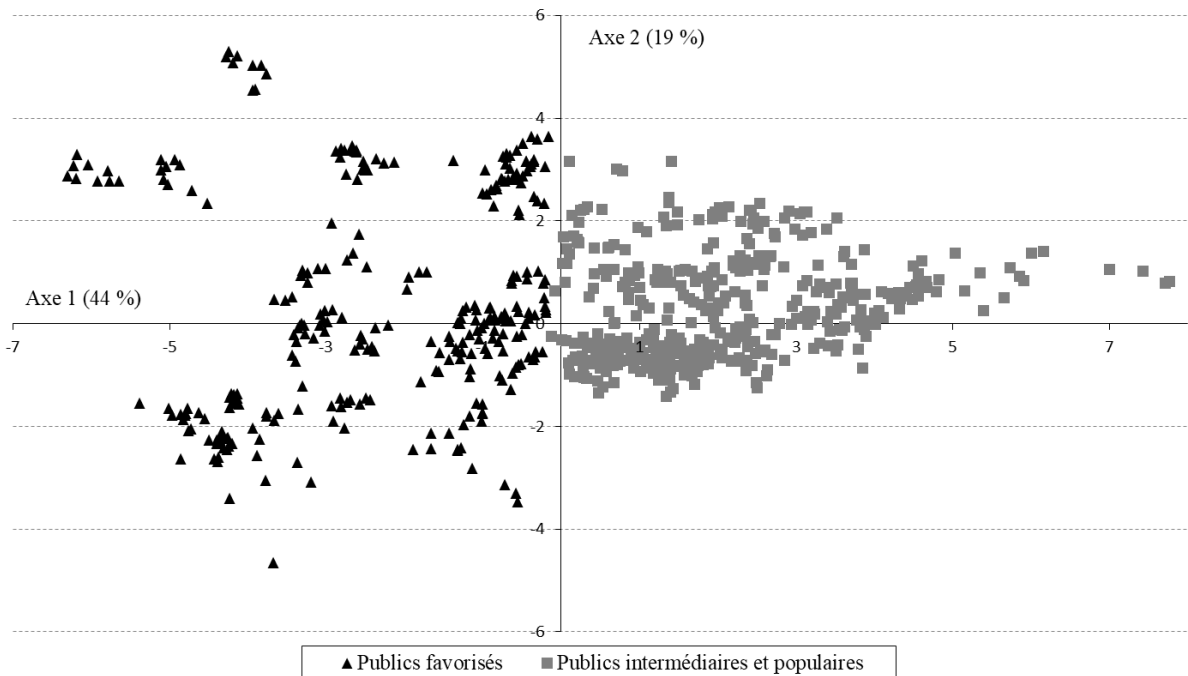


Lecture : la 1^{ère} dimension de l'ACP reproduit 44 % de l'inertie du nuage initial

Champ : universités françaises et grands établissements placés sous la tutelle du ministre chargé de l'enseignement supérieur

Source : MESRI-SIES, SISE

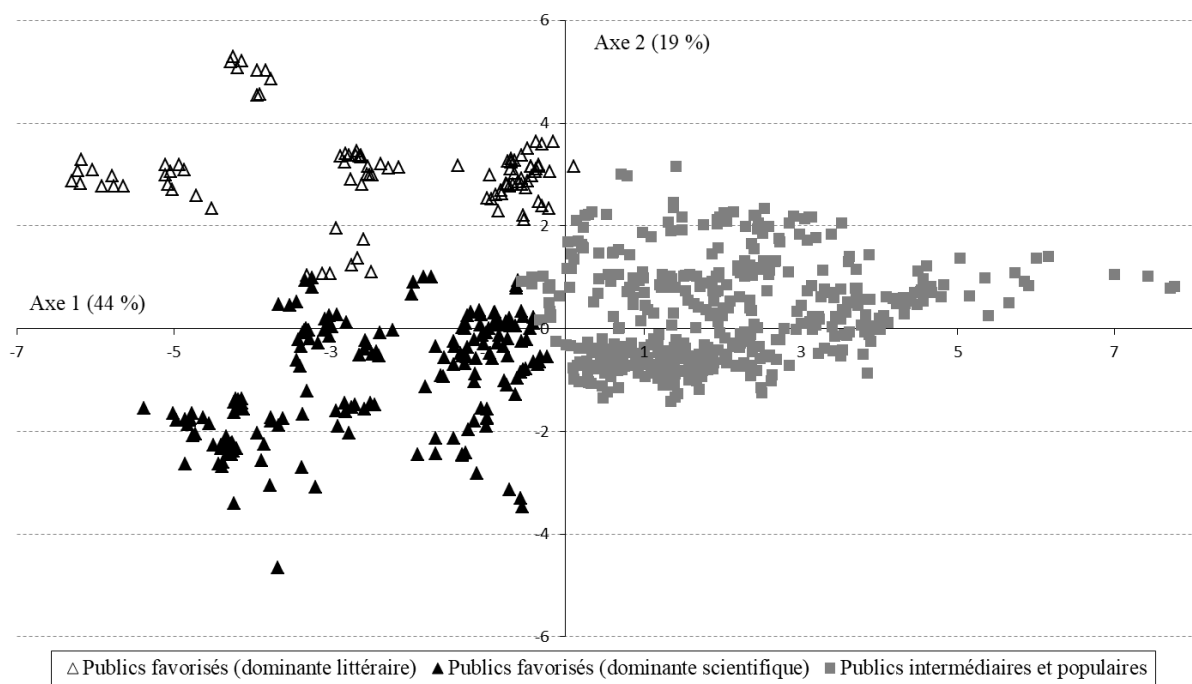
Figure A3-II – Représentation de la typologie à deux groupes issue de la CAH



Champ : universités françaises et grands établissements placés sous la tutelle du ministre chargé de l'enseignement supérieur

Source : MESRI-SIES, SISE

Figure A3-III – Représentation de la typologie à trois groupes issue de la CAH



Champ : universités françaises et grands établissements placés sous la tutelle du ministre chargé de l'enseignement supérieur

Source : MESRI-SIES, *SISE*