



HAL
open science

Mixité sociale à l'Université : une analyse sur la période 2006-2016 en France

Johan Seux

► **To cite this version:**

Johan Seux. Mixité sociale à l'Université : une analyse sur la période 2006-2016 en France. Economies et finances. Université Panthéon-Sorbonne - Paris I, 2023. Français. NNT : 2023PA01E051 . tel-04714982

HAL Id: tel-04714982

<https://theses.hal.science/tel-04714982v1>

Submitted on 30 Sep 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.

UNIVERSITE PARIS I PANTHÉON SORBONNE

ED d'ECONOMIE

Laboratoire de rattachement: Centre d'Economie de la Sorbonne

THÈSE

Pour l'obtention du titre de Docteur en Sciences Economiques

Présentée et soutenue publiquement

le 28 NOVEMBRE 2023 par

Johan SEUX

Titre de la thèse

Mixité sociale à l'Université:

une analyse sur la période 2006-2016 en France

Sous la direction de M. Pierre Courtioux, Professeur, et de M. Jérôme Gauté, Professeur

Membres du Jury

M. Arnaud Chéron, Professeur, Le Mans Université

M. Pierre Courtioux, Professeur, *Paris School of Business*

M. Jérôme Gauté, Professeur, Université Paris I

M. Jean-François Giret, Professeur, Université de Bourgogne

Mme Elise Huillery, Professeure, Université Paris-Dauphine

Mme Muriel Roger, Professeure, Université Paris I

UNIVERSITE PARIS I PANTHÉON SORBONNE
ED d'ECONOMIE

Laboratoire de rattachement: Centre d'Economie de la Sorbonne

THÈSE

Pour l'obtention du titre de Docteur en Sciences Economiques

Présentée et soutenue publiquement

le 28 NOVEMBRE 2023 par

Johan SEUX

Titre de la thèse

Mixité sociale à l'Université:

une analyse sur la période 2006-2016 en France

Sous la direction de M. Pierre Courtioux, Professeur, et de M. Jérôme Gauté, Professeur

Membres du Jury

M. Arnaud Chéron, Professeur, Le Mans Université

M. Pierre Courtioux, Professeur, *Paris School of Business*

M. Jérôme Gauté, Professeur, Université Paris I

M. Jean-François Giret, Professeur, Université de Bourgogne

Mme Elise Huillery, Professeure, Université Paris-Dauphine

Mme Muriel Roger, Professeure, Université Paris I

Résumé

Notre travail fournit, d'abord, un cadre méthodologique pour la mesure de la mixité sociale à l'Université. Celui-ci permet de définir les concepts d'origine sociale, de mixité sociale et de sélectionner une unité d'observation pertinente pour mesurer la mixité sociale au sein de l'Université française. Nous analysons, ensuite, de manière empirique la ségrégation sociale à l'Université en France sur la période 2006-2016. Sur la base de l'indice d'exposition normalisé, nous montrons qu'au niveau national, la ségrégation se fixe à des niveaux plus faibles que ceux mis en évidence par d'autres travaux pour l'enseignement secondaire. Nous montrons également que le niveau de ségrégation à l'Université a plutôt tendance à baisser avec le niveau de diplôme au niveau national. Par ailleurs, nous montrons une très grande variabilité de la ségrégation au cours du cursus par académie, avec des académies peu ségréguées pour certains niveaux d'études et fortement pour d'autres.

Afin de compléter notre analyse empirique, nous décomposons nos résultats sur le degré de mixité sociale afin de connaître les parts contributives de chaque facteur au niveau total de la ségrégation. Au niveau national, environ un tiers de la ségrégation correspond à une composante inter-disciplinaire alors que les deux tiers restants résultent d'un facteur inter-universitaire interne à chaque discipline. Nos résultats décrivent également la dimension fortement locale et inter-disciplinaire de la ségrégation dans le système universitaire français.

La dernière étape de notre travail analyse les liens entre mixité sociale et réussite à l'Université. Afin de mesurer les effets de la mixité sociale, nous avons choisi d'analyser les liens potentiels entre mixité sociale des formations suivies et réussite universitaire mesurée par le passage en année supérieure. Nous parvenons à mettre au jour un effet significatif de la part des étudiants très favorisés et défavorisés dans les formations sur la probabilité de passage en année supérieure pour le groupe des étudiants défavorisés uniquement. Au regard de la taille limitée de notre échantillon et de certains problèmes de représentativité identifiés, ce résultat ne nous permet pas, toutefois, de connaître précisément le type d'effet de pairs en jeu.

Summary

Our work provides, first, a methodological framework for measuring social segregation in French Universities. We define the concepts of social origin, social diversity (social segregation) and we select a relevant observation unit to measure social segregation within the French Universities. Over the period 2006-2016, we calculate the Normalized Exposure index of very advantaged and disadvantaged students for each French académies and levels of

education (one-year degree, two-year degree, Bachelor and Master). At the national level, values of Normalized Exposure index reveal the existence of social segregation in French universities, although at lower levels than those highlighted by articles on secondary education. We also study the levels of segregation by education level and we observe a downward trend in segregation as students progress through their studies. However, this overall trend does not seem to be carried by all académies, but only by a limited number. In addition, there is an important variability of segregation across education levels as far as one académie can be characterized by low level of segregation for certain education levels and important ones for others. Then, we decompose a segregation index and demonstrate that social segregation at university is rather a local phenomenon, and argue that the main driver of social segregation is the choice of academic discipline at the local level.

The last part of our work analyzes the links between social segregation and academic success in French universities. In order to measure the effects of social segregation, we analyze the potential links between the level of social segregation and the probability of success at the end of the first academic year. We observe a significant effect of the part of highly advantaged and disadvantaged students on this probability for the group of disadvantaged students only. Due to the limited size of our sample and identified representativeness problems, this result does not allow us, however, to know precisely what type of peer effect could be at play.

Mots-clés

Origine sociale – Mixité sociale – Université

Diversité régionale – Indice de ségrégation – Décomposition des indices

Facteurs de réussite à l'Université – Probabilité d'échec – Effet de pairs

Keywords

Social Origin – Social diversity – University

Regional diversity – Segregation Index – Index decomposition

Academic success determinants – Probability of dropout – Peer effect

L'Université Paris I Panthéon-Sorbonne n'entend donner aucune approbation, ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Remerciements

Je tiens à remercier, suivant l'ordre chronologique de déroulement de mon projet de thèse :

- D'abord Philippe Bich qui m'a reçu dans son bureau à la MSE en mai 2018 pour évoquer la possibilité de réaliser une thèse à Paris I, m'a exposé les différentes options possibles et m'a orienté vers les personnes susceptibles d'encadrer mon projet. Ensuite, Christine Erhel pour avoir évoqué avec moi pour la première fois la possibilité d'un sujet concret. Même si ce n'est pas finalement le sujet retenu, cela m'a permis de mûrir mes motivations et mon projet.
- Pierre Courtioux pour les nombreuses discussions sur les projets de thèse envisageables, les modalités pratiques d'organisation de celle-ci et surtout pour être finalement devenu mon directeur de thèse. Sans l'encadrement de Pierre et sa capacité à fixer des exigences croissantes au fur et à mesure du déroulement de la thèse, ce projet n'aurait pas pu aboutir.
- Jérôme Gautié pour être devenu bien sûr mon directeur de thèse ainsi que pour ses retours sur le contenu de mon travail. Je suis particulièrement reconnaissant envers Jérôme pour ses conseils méthodologiques et d'orientation qui ont fortement contribué au bon déroulement de ma thèse.
- Muriel Roger pour ses retours sur le contenu de mes chapitres de thèse, leur solidité méthodologique et leur adéquation aux problématiques concrètes de l'enseignement supérieur.
- Les organisateurs du séminaire STEP, Corinne Perraudin, Malo Mafhakami puis Eric Lefèvre, qui m'ont offert la possibilité à deux reprises de présenter mon travail et d'obtenir ainsi des retours riches et stimulants sur celui-ci.
- Tristan-Pierre Maury pour les discussions à la fois fournies et structurées que nous avons eu sur les questions liées à l'économie géographique dans le cadre

du projet de recherche dont les résultats sont très largement utilisés dans le chapitre II¹.

- les doctorants passés par le bureau 215 lors de ces quatre années d'études à l'Université Paris I, en particulier Thomas Delcey, Sophie Dessenin, Niels Boissonnet, Malo Mafhakami, Marion Bruna et Zeynep Kan pour les nombreuses informations échangées sur le déroulement du doctorat ainsi que pour les moments de convivialité.
- Claudia pour son soutien et sa patience ainsi que pour avoir pris soin de Diana et David durant les longues journées consacrées à mon travail de recherche.

¹ Les résultats du chapitre II ont fait l'objet d'une publication dans la Revue d'Economie Régionale et Urbaine : Pierre Courtioux, Tristan-Pierre Maury, Johan Seux, « Ségrégation sociale à l'Université: Des disparités académiques sur la période 2006-2016 », *REVUE D'ÉCONOMIE RÉGIONALE ET URBAINE* (3/2022), pp. 385-417.

*« Le premier cycle est ouvert à tous les titulaires
du baccalauréat et à ceux qui ont obtenu l'équivalence
ou la dispense de ce grade en justifiant d'une qualification
ou d'une expérience jugées suffisantes ».*
Article L612-3 du Code de l'Education

INTRODUCTION GENERALE

La mixité sociale dans l'enseignement fait l'objet de nombreux travaux académiques depuis plusieurs années. Cet intérêt se justifie au regard des conséquences potentielles de celle-ci. Les travaux de l'OCDE consacrés à l'Enquête PISA, que nous décrivons plus en détail ci-dessous, analysent ainsi dans quelle mesure la ségrégation sociale peut réduire l'efficacité du système éducatif dans son ensemble. Cette situation peut justifier l'allocation de moyens supplémentaires aux établissements concernés par une concentration d'élèves issus d'un milieu défavorisé, comme c'est le cas en France avec les politiques d'éducation prioritaire.

Si la lutte contre la ségrégation scolaire a été érigée depuis plusieurs années au rang d'objectif central des politiques éducatives, comme le soulignent Fack et Grenet (2013), il existe également des débats relatifs à cette question pour l'enseignement supérieur. Ainsi, Fack et Huillery (2021) constatent qu'il existe non seulement des inégalités d'accès, mais aussi de parcours dans l'enseignement supérieur, qui se traduisent par un caractère régressif, c'est-à-dire anti-redistributif de la dépense d'enseignement supérieur. En effet, les jeunes issus d'un milieu défavorisé ont, en moyenne, des taux d'accès plus faibles à l'enseignement supérieur et ont tendance à suivre des formations plus courtes. L'ambition de ce travail de recherche est ainsi de contribuer aux travaux sur la question de la mixité sociale au sein de l'Université française, notre démarche nous semblant justifiée au regard du nombre encore relativement limité de travaux portant sur cette question.

Afin d'aborder la question de la mixité sociale au sein de l'Université française, il nous semble nécessaire d'introduire préalablement les concepts nécessaires à l'analyse et de présenter les principaux travaux disponibles sur ce sujet.

1- Définir la mixité sociale à partir d'indices synthétiques

- Les concepts de mixité sociale et de ségrégation sociale

La mixité sociale peut être définie de manière préliminaire et intuitive comme la cohabitation de personnes ayant des origines sociales variées au sein d'un espace commun, d'un lieu partagé. A l'inverse, il existe une situation de ségrégation sociale lorsqu'un espace commun, un lieu partagé comprend principalement des personnes provenant du même milieu social. Dans ce cas, les différents groupes de la population sont en partie séparés géographiquement (par exemple les personnes d'un même milieu social peuvent être particulièrement présentes dans une même ville, etc). Un faible degré de mixité sociale peut résulter soit d'une volonté de réduire les contacts avec les personnes issues d'autres milieux sociaux, c'est-à-dire de côtoyer principalement des individus du même milieu social, soit d'une contrainte subie lorsqu'un espace de vie commun est peu diversifié en raison des stratégies d'évitement mises en œuvre par des individus d'autres groupes sociaux².

Cette première définition de la mixité sociale et de son contraire, la ségrégation sociale, nous permet ainsi d'apporter plusieurs précisions concernant notre travail. L'objet de celui-ci est la mixité sociale : il n'aborde donc ni la mixité scolaire, ni la « mixité ethnique », pas plus que les questions liées au genre. Afin de justifier cette distinction, en l'absence, à notre connaissance, de traitement explicite de cette question dans les analyses économiques, nous pouvons nous appuyer sur les analyses sociologiques de Pierre Merle. Comme le montre Merle (2012), la ségrégation dans l'enseignement peut reposer sur quatre éléments : le genre, l'origine ethnique, l'origine sociale ainsi que la réussite académique. Si les trois premiers types de ségrégation peuvent apparaître au premier abord relativement intuitif, la ségrégation suivant le niveau académique est moins intuitive. Celle-ci peut notamment passer par différents biais :

² Pierre Merle (2011) analyse les conséquences de l'assouplissement de la carte scolaire qui autorise à compter de 2007 des dérogations antérieurement prosrites. Ce travail montre que si cet assouplissement devait initialement profiter davantage aux boursiers, ce sont finalement surtout les catégories sociales moyennes et supérieures qui les utilisent afin de contourner, d'éviter certains établissements au profit d'autres qui sont davantage conformes à leurs attentes.

institutionnalisation dans certains pays avec des filières créées explicitement sur la base des résultats des élèves, établissements dont le niveau moyen des élèves est varié, constitution de classes de différents niveaux avec le système des options et dichotomie entre secteur public et secteur privé... Les quatre facteurs de ségrégation identifiés peuvent bien entendu s'accumuler ou entretenir des liens entre eux. C'est le cas, par exemple, de l'origine sociale qui peut être corrélée aux performances académiques qui deviennent elles-mêmes un facteur de ségrégation potentiel.

- Les objectifs des indices synthétiques de mesure de la mixité sociale

Les indices synthétiques de mesure de la mixité sociale reposent sur le principe d'une comparaison entre la part occupée par des groupes sociaux au sein d'une population de référence et celle qu'ils occupent au sein des différentes unités dans lesquelles cette population a été répartie. L'avantage d'un indicateur synthétique est de résumer en un seul chiffre une série de statistiques. Faute d'indice, il n'y a qu'une série de données statistiques relatives à une population de référence et aux unités qui la composent. Ces indices font aujourd'hui l'objet d'une littérature conséquente et détaillée. Les articles de référence de Reardon et Firebaugh (2002) ainsi que celui de Frankel et Volij (2011) identifient ainsi les principaux indices disponibles et discutent leurs propriétés mathématiques.

2- L'Université française comme champ d'analyse

L'enseignement supérieur français, qui comprend l'ensemble des formations postérieures à l'obtention du baccalauréat, se caractérise par une grande diversité. Afin de définir le périmètre de notre travail et d'identifier plus clairement les types d'établissements qui sont compris dans celui-ci, nous avons introduit la double distinction suivante : enseignement public et privé, d'une part, et enseignement universitaire et non universitaire, d'autre part.

- Première distinction : enseignement supérieur public et privé

Une première source de diversité dans l'enseignement supérieur est la coexistence d'établissements publics et privés³. Les établissements publics d'enseignement supérieurs, qui ont un statut d'établissement public à caractère scientifique, culturel et professionnel (EPCSP)⁴ sont décrits plus en détail dans les paragraphes suivants. Les établissements privés peuvent obtenir la qualification d'établissement d'enseignement supérieur privé d'intérêt général (EESPIG) qui leur permet d'offrir des formations conduisant à des diplômes conférant un grade universitaire ou visé par l'État. Les établissements privés comprennent des établissements privés d'enseignement supérieur libres et des instituts privés d'enseignement supérieur techniques. A la rentrée 2016, la première catégorie comprenait 13 établissements dont 5 instituts catholiques. La seconde catégorie inclut des écoles d'ingénieurs, des écoles de commerce⁵ ainsi que des Lycées privés. Ces derniers peuvent comprendre, en effet, des sections de techniciens supérieurs (STS) et des classes préparatoires aux Grandes Ecoles (CPGE)⁶. A la rentrée 2016, l'enseignement supérieur privé accueillait environ 20% du total des inscrits.

- Seconde distinction : enseignement supérieur universitaire et non universitaire

Une seconde source de diversité au sein de l'enseignement supérieur français résulte de la coexistence d'établissements universitaires et non universitaires. Il convient de noter que les distinctions d'enseignement public et privé d'une part et d'enseignement universitaire et non universitaire d'autre part ne se recoupent pas. Un établissement public peut ainsi être ou non universitaire. L'article L711-2 du Code de l'Education

³ Conformément à la loi du 12 juillet 1875, l'enseignement supérieur est libre.

⁴ Une liste des acronymes utilisés dans le cadre de ce travail ainsi que leur description est disponible dans l'annexe 11.

⁵ Il existe actuellement une cinquantaine d'écoles d'ingénieurs et une soixantaine d'écoles de commerce privées qui ont obtenu une reconnaissance par l'Etat et dont le diplôme est accrédité ou visé par le MESRI.

⁶ Ils sont soumis dans ce cadre aux règles qui régissent les établissements de l'enseignement supérieur.

distingue plus précisément quatre types d'établissements publics à caractère scientifique, culturel et professionnel : a) les Universités auxquelles sont assimilés les instituts nationaux polytechniques b) les écoles et instituts extérieurs aux Universités c) Les Ecoles Normales Supérieures, les écoles françaises à l'étranger et les grands établissements d) les communautés d'Universités et établissements.

Le système universitaire accueille la majeure partie des étudiants. A la rentrée universitaire 2016-2017, suivant les données issues de la publication du Ministère de l'Enseignement Supérieur et de la Recherche (MESRI) « L'état de l'enseignement supérieur et de la recherche, édition 2018 », sur un total d'environ 2 610 000 étudiants inscrits dans l'enseignement supérieur en France métropolitaine et dans les départements d'outre-mer, environ 62% sont inscrits à l'Université. Parmi ceux-ci, environ 117 000 étaient inscrits en formation dans un institut universitaire technologique (IUT)⁷ permettant d'obtenir un diplôme universitaire technologique (DUT). Au sein des Universités, des instituts ou écoles peuvent être créés, mais ceux-ci conservent leur statut universitaire : cela concerne notamment les IUT, les Instituts Nationaux Polytechniques (INP), les Instituts de Préparation à l'Administration Générale (IPAG) ou encore les Instituts d'Administration des Entreprises (IAE). Même si le principe général pour l'enseignement universitaire est que le premier cycle est ouvert à tous les titulaires du baccalauréat et à ceux qui ont obtenu une équivalence, comme le précise l'article L612-3 du Code de l'Education, des procédures d'admission sélective peuvent, toutefois, exister, comme dans le cas des INP ou des IUT.

Le périmètre des établissements universitaires français a évolué significativement depuis les années 1990 en raison des orientations des politiques d'enseignement supérieur prises au niveau national. Les établissements universitaires existants ont commencé, d'abord, à créer des antennes universitaires à compter de 1992⁸. Ces créations ont eu lieu dans le contexte du plan « Université 2000 » de modernisation des Universités dont l'un des objectifs était de développer des antennes délocalisées d'Universités dans les villes moyennes. Parallèlement à ces créations d'antennes

⁷ Un IUT est une composante d'une université publique qui prépare en deux années (à partir de la rentrée 2021, la formation se déroule sur trois ans) aux fonctions de certains secteurs de la production, de la recherche appliquée et des services. Il existait à la rentrée 2016 108 IUT.

⁸ La première antenne créée concernait un DEUG de Droit, ouvert en partenariat avec l'Université Toulouse 1.

délocalisées, un mouvement de fusion des établissements universitaires est à l'œuvre depuis la fin des années 2000. Entre 2009 et 2017, plusieurs établissements ont fusionné en raison de l'objectif défini sur le plan national de défragmenter l'organisation de l'enseignement supérieur. La loi relative aux libertés et responsabilités des Universités (LRU) de 2007, qui renforce l'autonomie des Universités, en particulier sur le plan budgétaire, prévoit, en effet, pour celles-ci, la possibilité de se regrouper après un vote à la majorité absolue des membres en exercice du Conseil d'administration et après approbation également du MESRI. Si les fusions se réalisent à l'initiative des établissements eux-mêmes, les différentes fusions qui ont eu lieu traduisent les objectifs nationaux de créer, d'une part, des synergies entre les établissements et, d'autre part, d'atteindre pour ces établissements une certaine taille leur permettant potentiellement de figurer en bonne position dans les classements internationaux.

Le système que nous appelons par défaut « non universitaire » comprend différents types d'établissements qui accueillent à la rentrée 2016-2017 environ 38% du total des étudiants. Celui-ci comprend une grande diversité d'établissements (écoles de commerce, Ecoles Normales Supérieures, « Grands établissements », etc.). Suivant « L'état de l'enseignement supérieur et de la recherche, édition 2018 », à la rentrée 2016, l'enseignement supérieur non universitaire comprenait notamment environ 87 000 étudiants inscrits en classe préparatoire aux Grandes Ecoles (CPGE)⁹. Il convient de noter que l'enseignement non universitaire ne peut, toutefois, être réduit au périmètre des « Grandes Ecoles »¹⁰. A la rentrée 2016, l'enseignement non universitaire comprenait ainsi environ 257 000 étudiants inscrits dans une section de technicien supérieur (STS) permettant de préparer le brevet de technicien supérieur (BTS) dans un Lycée ou un centre d'apprentissage.

Les COMUE sont des établissements publics à caractère scientifique, culturel et professionnel (EPSCP) qui regroupent des établissements d'enseignement supérieur et de recherche en vue de leur fusion potentielle. Ce statut a été créé par la loi du 22

⁹ Dans le cadre d'une mise en œuvre progressive de la double inscription obligatoire, ceux-ci sont de manière croissante inscrits à l'Université en parallèle à leur inscription en CPGE. Le nombre d'élèves inscrits en CPGE est en baisse à 85 000 à la rentrée 2020 selon la dernière version disponible de l'état de l'enseignement supérieur.

¹⁰ La conférence des « Grandes Ecoles », qui est une association d'établissements d'enseignement supérieur, comptabilise au mois de Juin 2017, 223 « Grandes Ecoles » celles-ci proposant des formations ainsi que des statuts très variés : écoles d'accès à la haute fonction publique civile ou militaire, écoles de commerce et de management...

juillet 2013 relative à l'enseignement supérieur et à la recherche, pour succéder aux pôles de recherche et d'enseignement supérieur (PRES) créés par la loi de programme pour la recherche du 18 avril 2006, pour faciliter la coopération inter-universitaire et mettre en place des structures de coordination des financements de la recherche entre institutions, mais sans les inciter à fusionner. Les COMUE sont susceptibles de regrouper des établissements à la fois universitaires et non universitaires, comme nous le décrivons plus en détail dans le paragraphe 2.2 du chapitre II.

Au regard des éléments susmentionnés, l'intérêt d'une analyse de la mixité sociale dans le secteur universitaire résulte selon nous des facteurs suivants :

- S'il existe un principe général de non sélection à l'entrée de l'Université, cela n'implique pas nécessairement que les phénomènes de ségrégation y soient absents. Le principe de recrutement des étudiants dans les établissements non universitaires est la sélection : ceux-ci peuvent utiliser de manière systématique des concours de recrutement dès le premier cycle, ce qui est susceptible de favoriser des phénomènes de ségrégation sociale et scolaire. Au niveau du premier cycle de l'enseignement non universitaire, en effet, le principe général de non sélection des étudiants ne s'applique pas. Un travail centré sur l'Université peut permettre ainsi de vérifier que la mixité sociale est bien présente à l'Université et, si ce n'est pas le cas, d'analyser pour quelles raisons.
- L'organisation de l'enseignement non universitaire est distincte de celle de l'Université, avec par exemple des cursus pluridisciplinaire ou non spécialisés au niveau de la Licence alors que la spécialisation est le principe à l'Université¹¹. Alors qu'il est envisageable de mesurer à l'Université des phénomènes de ségrégation non seulement entre les établissements, mais aussi entre les disciplines, cela semble plus compliqué pour les établissements non universitaires, voire impossible pour certaines d'entre eux.

¹¹ C'est cas par exemple le cas dans une certaine mesure pour le Collège Universitaire de Sciences Po Paris.

- Comme nous le détaillerons davantage dans les paragraphes suivants, il existe encore peu de travaux à ce jour qui analysent la question de la mixité sociale sur l'ensemble de l'Université française.

Pour ces différentes raisons, nous avons choisi de recentrer notre analyse de la mixité sociale sur le champ suivant : l'ensemble des étudiants inscrits dans les Universités de France métropolitaine¹². Notre travail aborde ainsi uniquement la question de la répartition des étudiants au sein du système universitaire et non celle de l'accès différencié à l'Université, la population de l'Université étant elle-même considérée ici comme une donnée. Comme le décrit le Tableau 1 ci-dessous, notre champ d'analyse inclut uniquement les établissements universitaires de l'enseignement public.

Tableau 1 : les principaux types d'enseignement supérieur et le champ de notre étude

Type d'établissement	Enseignement Universitaire	Enseignement Non-Universitaire
Public	X	
Privé ¹³ .		

Source : Auteur

Il convient de noter que les analyses portant sur la mixité sociale dans l'enseignement supérieur pourraient tout à fait être étendues à l'avenir à l'enseignement supérieur non universitaire¹⁴. En effet, comme expliqué dans l'encadré 1 ci-dessous, il existe aujourd'hui des bases de données SISE sur l'enseignement supérieur non

¹² Suite à cette première définition générale et conceptuelle, le périmètre de notre analyse sera précisé de manière plus détaillée dans la section 2 du chapitre II et restreint en particulier aux étudiants de Licence et Master.

¹³ D'un point de vue juridique, les établissements d'enseignement privés n'ont pas le statut d'université, même si le nom de ceux-ci peut faire référence à l'Université (Université catholique de Lille...).

¹⁴ Comme expliqué dans le chapitre II, la base SISE Universités contient des informations relatives à un nombre limité d'établissements non universitaires, mais ceux-ci ne sont pas représentatifs de l'ensemble de celui-ci.

universitaire sur lesquelles pourraient s'appuyer de futurs travaux de recherche le jour où celles-ci seront diffusées par le MESRI. Il nous semble, toutefois, que l'exploitation de celles-ci correspondrait à un travail de recherche complémentaire mais distinct du nôtre en raison des spécificités des établissements concernés et de la méthodologie probablement différente qui devrait être employée pour mesurer les niveaux de mixité sociale entre ceux-ci ou en leur sein. Par ailleurs, au moment de définir le périmètre de notre sujet en 2019, les bases de données relatives à l'enseignement supérieur non universitaire n'étaient pas accessibles via l'ADISP-PROGEDO¹⁵.

Encadré 1 : un premier aperçu des bases de données SISE

Les bases SISE (Système d'Information sur le Suivi de l'Etudiant) du Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation (MESRI) sont des bases exhaustives concernant l'ensemble des étudiants du système d'enseignement supérieur français.

Il existe aujourd'hui de multiples bases SISE correspondant chacune à un type ou une voie d'accès à l'enseignement supérieur. Ainsi, il y en avait 7 à la rentrée 2020 : SISE Université, SISE Ingénieurs, SISE Management, SISE ENS, SISE PRIV, SISE ENQ26bis et SISE CULTURE.

Dans le cadre de notre travail, nous avons utilisé la base SISE Université à laquelle nous avons pu accéder via l'ADISP-PROGEDO. Celle-ci, qui fera l'objet d'une description plus détaillée dans le chapitre II, se scinde en deux bases de données distinctes SISE-Inscrits et SISE-Résultats, cette dernière incluant des informations relatives aux résultats des examens.

¹⁵Nous n'avons pas pu accéder à ces bases de données dans le cadre de nos travaux, le MESRI nous ayant indiqué en 2019 qu'il n'y avait pas de date de diffusion prévue.

3- Si de nombreux travaux abordent la question de la mixité sociale dans l'enseignement, ceux qui analysent spécifiquement ce sujet au sein de l'Université sont récents et en nombre limité

- Les travaux portant sur la question de la mixité sociale dans l'enseignement primaire et secondaire

La ségrégation académique et sociale à l'école fait aujourd'hui l'objet de nombreux travaux comparatifs menés au niveau international. Celle-ci constitue par exemple l'un des champs d'investigation de l'OCDE, via l'enquête PISA que nous décrivons dans l'Encadré 2 ci-dessous. Afin d'analyser cette question, les travaux de l'OCDE, qui portent essentiellement sur l'enseignement secondaire et abordent ce sujet dans une perspective internationale, recourent depuis plusieurs années aux indices synthétiques de mesure de la mixité sociale. La publication OECD (2019) *Balancing School Choice and Equity* aboutit par exemple à des conclusions portant sur la majeure partie des pays ainsi qu'à des conclusions spécifiques à chaque pays. Pour les premières, l'OCDE indique ainsi que la ségrégation académique est plus fréquente que la ségrégation sociale, qu'il existe une tendance marquée à la concentration des élèves aux performances les plus basses dans certains établissements et qu'en revanche, il y a davantage d'écoles avec une forte concentration des élèves favorisés que d'écoles avec une forte concentration d'élèves défavorisés.

Encadré 2 : L'enquête internationale PISA de l'OCDE

Le PISA désigne le programme international pour le suivi des acquis des élèves. Il s'agit de la plus grande étude internationale effectuée directement auprès des élèves dans le domaine de l'éducation. Menée par l'OCDE, l'enquête PISA permet de comparer les résultats obtenus par les élèves dans différents pays. Il s'agit d'une enquête périodique (elle a lieu tous les trois ans) effectuée auprès de jeunes âgés de 15 ans. En 2022, l'enquête couvrait ainsi 85 pays, environ 8 000 élèves français ayant répondu à celle-ci. L'enquête PISA aborde la question de la mixité sociale et scolaire au sein de chaque système éducatif, celle-ci étant susceptible d'être à la fois un facteur d'iniquité et d'inefficacité.

En France, en l'absence de données ethniques, la littérature sur la ségrégation éducative s'est centrée sur l'origine sociale. Pour un certain nombre de travaux, la dimension spatiale est essentielle en raison du rôle particulier de la ségrégation résidentielle liée au principe de sectorisation¹⁶. Ly et al. (2014) analysent les très fortes inégalités sociales et scolaires qui existent entre les territoires en Île-de-France. Ces inégalités semblent venir en grande partie de facteurs extérieurs au système scolaire, tels les inégalités spatiales, qui structurent la société et limitent donc les leviers d'actions des établissements pour réduire la ségrégation. Toutefois, ce travail qui analyse de manière quantitative la ségrégation intra-établissements générée par la constitution des classes montre que les établissements contribuent de manière importante à la ségrégation locale vécue par les collégiens et les Lycéens. Courtioux et Maury (2018) analysent quant à eux sur la période 2004-2014 le degré de mixité sociale des Collèges dans le secteur privé, le secteur public hors ZEP et dans les ZEP. A partir de la Base Centrale de Scolarité (BCS) du Ministère de l'Education Nationale, ils parviennent à la conclusion que pour un même niveau mesuré de mixité sociale, le secteur public accueille davantage d'élèves défavorisés. Courtioux et Maury (2020) soulignent également le caractère essentiel de la dimension géographique de la ségrégation dans l'enseignement secondaire. Ce travail explique comment la contribution des divergences public/privé augmente et est plus élevée dans les aires urbaines moyennes et les communes centrales. De plus, cette divergence public /privé est en moyenne plus élevée au niveau local, par exemple au sein d'une agglomération, qu'entre régions géographiques distinctes. Givord et al. (2016) distinguent les territoires suivant le degré plus ou moins fort de ségrégation entre les Collèges, celle-ci étant plus forte dans les académies et dans les départements urbains. Grenet et Souidi (2021) analysent quant à eux la nouvelle méthode d'affectation des élèves dans les Collèges publics qui existe dans certains arrondissements de Paris depuis la rentrée 2017 : les secteurs multi-Collèges, qui permettent de définir des secteurs communs à plusieurs Collèges géographiquement proches de manière à rééquilibrer leur recrutement sur le plan social. Les auteurs constatent que sur la période 2017-2019, deux des trois secteurs concernés ont atteint leur objectif de mixité sociale et

¹⁶ La répartition des élèves entre Collèges publics repose en France sur la carte scolaire. Les élèves sont normalement affectés à un unique Collège public qui dépend de leur lieu de résidence. Toutefois, il est possible de demander une dérogation pour un autre Collège public, en raison, par exemple, des options demandées ou bien de s'orienter vers un Collège privé, dont le recrutement ne dépend pas de la carte scolaire.

que les stratégies d'évitement vers le privé ont diminué. Dans le troisième secteur où les résultats sont plus mitigés, l'évitement vers le privé a reculé également.

- Pour l'Université, des sujets proches mais distincts tels que celui de l'équité de l'accès à l'Université sont abordés dans la littérature

La question de l'accès différencié à l'Université suivant les origines sociales est distincte de celle de la mixité sociale au sein de l'Université au sens où nous l'entendons comme évaluation du degré de ségrégation entre les formations. L'accès à l'Université fait l'objet de nombreux travaux, une part significative des élèves interrompant leurs études à la fin du secondaire/n'intégrant pas l'enseignement supérieur. Dans cette ligne de travaux, O'Connell et al. (2006) évaluent la question de savoir si la hausse du nombre d'étudiants entrant chaque année à l'Université au cours des années 1990 et 2000 en Irlande a pu atténuer les inégalités sociales d'accès à l'Université. Ils aboutissent à la conclusion selon laquelle les étudiants issus des milieux les plus favorisés ne progressent plus en termes relatifs en raison d'un phénomène de saturation alors que les taux de participation des enfants des travailleurs manuels augmentent. Kottmann et al. (2019) abordent la question de l'accès différencié à l'enseignement supérieur dans les 28 pays membres de l'Union européenne. Les travaux de Kaye (2020) incluent à la fois la question de l'accès différencié à l'Université et celle de la mixité sociale au sein des établissements d'enseignement supérieur via la notion d'« élargissement de la participation ». Les politiques « d'élargissement de la participation » ont pour objectif de réduire les inégalités sociales non seulement à l'entrée de l'Université, mais aussi au sein de celle-ci. La place des boursiers dans l'« élargissement de la participation » à l'Université est analysée de manière spécifique.

D'autres travaux analysent également les liens entre origines sociales, accès à l'Université et type de diplôme dans le cas de l'Université française. Selon Albouy et al. (2007), l'accès à l'Université est différencié suivant les origines sociales. De plus, les inégalités concernent aujourd'hui davantage la nature des études suivies : à partir des données des Enquêtes Emploi de 1990 à 2002, les auteurs constatent à un niveau fin un lien entre origine sociale et type de diplôme. Selz et Vallet (2006) analysent quant à eux le paradoxe suivant : évaluées sur l'ensemble de la population, d'une part,

les inégalités d'accès aux diplômes en fonction de l'origine sociale ont diminué entre les générations sur cinquante ans. D'autre part, si l'on prend en compte les seuls diplômés du baccalauréat, les inégalités d'obtention d'un diplôme de l'enseignement supérieur, en fonction de l'origine sociale ont augmenté entre les mêmes générations. Ce paradoxe apparent résulterait d'un phénomène de sur-sélection au sein des générations anciennes : en raison de l'élimination différentielle très forte selon l'origine sociale au niveau du bac, seuls restaient les élèves d'origine populaire notoirement brillants et ceux-ci ne rencontraient en général que peu de difficultés dans la suite de leurs études supérieures.

Fack et Huillery (2021) constatent qu'il existe non seulement des inégalités d'accès à l'enseignement supérieur, un jeune issu d'une famille aisée ayant trois fois plus de chance d'accéder à l'enseignement supérieur qu'un jeune issu d'une famille aux revenus modestes, mais aussi de parcours au sein de celui-ci. Cela se traduit par un effet anti-redistributif : « l'investissement public pour la formation supérieure des jeunes dont les parents sont les plus aisés est beaucoup plus important que pour les jeunes dont les parents sont plus modestes ». A partir de cette analyse, les auteurs questionnent la stratégie d'investissement actuelle dans l'enseignement supérieur et proposent plusieurs pistes pour renforcer l'efficacité, l'efficience et l'équité de celui-ci.

Afin uniquement de préciser le périmètre de notre travail, par contraste avec les travaux mentionnés dans les trois paragraphes antérieurs, il convient de noter que si ceux-ci abordent des thématiques proches de celle de la mixité sociale au sein du système universitaire et utilisent fréquemment des bases de données identiques, leur perspective nous semble différente. Dans ces travaux, en effet, la question de la possibilité de l'ascension sociale permise par les différents diplômes ainsi que celle de la reproduction sociale sont centrales. L'analyse de la question de la cohabitation de différents groupes sociaux au sein d'un espace commun, en revanche, ne constitue pas leur objectif premier.

- Les travaux consacrés à l'enseignement supérieur non universitaire

La littérature académique s'est depuis longtemps consacrée à la question de la mixité sociale au sein des Grandes Ecoles françaises en raison probablement du présupposé

implicite que c'est au sein de celles-ci pour l'essentiel qu'ont lieu les phénomènes de ségrégation, l'Université par contraste étant naturellement considérée comme un lieu d'enseignement ouvert. Depuis les années 1970 et 1980, à la suite des travaux de Pierre Bourdieu (1989), les Grandes Ecoles et en particulier les plus prestigieuses d'entre elles ont fait l'objet de nombreuses analyses revenant sur l'origine sociale très majoritairement privilégiée de leurs étudiants : pour une analyse relativement récente, on peut par exemple se reporter à l'article de Yves-Marie Abraham (2007) sur les étudiants de HEC Paris. Bonneau et al. (2022), qui analysent les inégalités sociales d'accès aux Grandes Ecoles, soulignent que certains établissements ont mis en place des dispositifs destinés à favoriser la mixité sociale et à diversifier le profil de leurs étudiants. Toutefois, ce travail montre qu'au cours des quinze dernières années, les élèves issus de milieux sociaux défavorisés y sont encore très nettement sous-représentés et que ces établissements peinent à diversifier leur base de recrutement. D'un point de vue statistique, selon les données issues du MESRI, et publiées par la Direction de l'Évaluation de la Prospective et de la Performance¹⁷, à la rentrée 2016-2017, les Ecoles Normales Supérieures accueillaient ainsi 59,7% d'enfants de « cadres et de professions intellectuelles supérieures » et seulement 2,8% d'enfants d'ouvriers. Geiss et al. (2006) montrent que la probabilité d'intégrer une Grande Ecole pour un élève de milieu modeste est environ vingt fois inférieure à celle d'un élève très favorisé. Le travail de Luc Rouban (2015) sur l'École Nationale d'Administration démontre qu'aux antipodes des objectifs affichés de l'ENA, le profil moyen des énarques est un concentré de ressources sociales qui peut se mesurer via un indice multicritères qui repose sur les Professions et Catégories Socio-professionnelles (PCS), le passage par d'autres Grandes Ecoles, l'exercice de fonctions politiques d'élus ou de militants, l'existence de relations familiales dans le monde de la politique et des affaires, etc. Ce phénomène est croissant dans le temps depuis la création de l'ENA en 1945. Il convient, toutefois, de noter que ces travaux, dans leur ensemble, concernent, comme indiqué ci-dessus, des problématiques distinctes et potentiellement éloignées de celles qui concernent l'Université.

¹⁷ Section « Origine sociale des étudiants français dans les principales filières de l'enseignement » « Repères et Références Statistiques », Édition 2017, MESRI.

- Des travaux récents abordent la question de la mixité sociale au sein de l'Université française

Dans le champ de l'enseignement supérieur, les données disponibles ont eu tendance à s'accroître au cours des dernières années. Le Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation (MESRI), par exemple, a ainsi mis à disposition des chercheurs la base de données SISE Université qui comprend de nombreuses données administratives relatives à l'inscription des étudiants ou encore le panel bacheliers 2014 qui est une enquête comprenant des données relatives à leurs parcours ainsi qu'à leurs conditions de vie. Sur cette base, un certain nombre d'analyses portant sur la mixité sociale à l'Université se sont développées au cours de la période récente.

Le travail de Frouillou (2016) est précurseur au sens où il est le premier à aborder la question de la mixité sociale des formations au sein de l'Université française sur une base quantitative. Celui-ci porte sur les inégalités de répartition sociale des étudiants entre les formations des 16 Universités d'Île-de-France. Il repose sur l'exploitation de la base de données SISE (décrite de manière sommaire dans l'encadré 1 ci-dessus et de façon plus détaillée dans le chapitre II), de données communales portant sur l'Île-de-France ainsi que sur des entretiens avec des acteurs universitaires et des étudiants. Ces différentes sources permettent de montrer qu'il existe en Île-de-France des phénomènes de ségrégation entre les formations des différentes Universités. Au moyen d'une méthodologie quantitative reposant sur des entretiens, Frouillou (2016) souligne l'importance prépondérante des effets de pairs pour les étudiants qui ont un capital scolaire limité et sont issus principalement du groupe social le plus défavorisé : ils ont davantage tendance que les étudiants plus favorisés à se laisser influencer par les choix collectifs pour choisir une Université ainsi qu'une discipline d'inscription. L'une des principales limites de ce travail est la restriction de son champ géographique à l'Île-de-France. Par ailleurs, celui-ci n'utilise pas d'indices synthétiques de ségrégation, ce qui le rend plus difficilement comparable avec les travaux évoqués précédemment sur la ségrégation dans le secondaire.

L'article d'Avouac et Harari-Kermadec (2021), qui repose également sur l'exploitation de la base de données SISE, est centré sur les différences qui existent entre les Universités sur le plan de l'origine sociale des étudiants. L'article établit ainsi pour la

période 2007-2015 trois principaux constats. Il existe, d'abord, de manière statique une forte hétérogénéité sociale entre établissements. De manière dynamique, ensuite, il existe une réduction de la ségrégation et simultanément « une polarisation sociale croissante des établissements », la conclusion du travail indiquant que « l'hétérogénéité sociale observée n'a pas diminué significativement entre 2007 et 2015 ». Les niveaux de mixité sociale observés sont, enfin, articulés aux politiques récentes de l'enseignement supérieur telles que les initiatives d'excellence (IDEX), qui sont décrites dans l'encadré 3 ci-dessous et bénéficient davantage aux établissements dont la composition sociale est favorisée. Il nous semble, toutefois, que ce travail n'aborde pas toutes les dimensions envisageables de la ségrégation et notamment celles qui sont liées aux effets de pairs ainsi qu'aux choix individuels et familiaux, abordés par Frouillou (2016). Dans les chapitres I et II, nous discuterons ce constat ainsi que le choix de l'unité d'observation, à savoir l'Université.

Encadré 3 : Les Initiatives d'excellence (IDEX)

Les initiatives d'excellence correspondent à une partie des « investissements d'avenir », engagés par la loi de finances rectificative du 9 mars 2010. L'objectif de ces programmes d'investissement est de créer en France des ensembles pluridisciplinaires d'enseignement supérieur et de recherche qui soient de rang mondial. Entre 2010 et 2022, le Parlement a voté trois programmes successifs d'investissements d'avenir pour un total de 57 milliards d'euros dont plus de la moitié est réservé à l'enseignement supérieur, à la recherche et au transfert d'innovation. A la fin 2022, 18 sites avaient reçu un label d'excellence, dont 9 au titre de l'Initiative d'Excellence-IDEX (début effectif de labellisation en 2012) et 9 au titre des I-SITE (Initiative-Science-Innovation-Territoires-Economie), qui sont apparus dans le cadre de la seconde vague des plans d'investissement d'avenir. Avec un début de labellisation effectif en 2016, les I-Site correspondent à des Universités dans lesquelles le spectre des disciplines ayant vocation à l'excellence sur le plan mondial est plus limité.

Le travail de Bechichi et al. (2021) repose sur un ensemble de données administratives issues des plateformes APB et Parcoursup¹⁸, récemment mises à disposition par le MESRI et qui offrent la possibilité d'étudier de manière très précise la ségrégation qui s'opère à l'entrée de l'enseignement supérieur. Celui-ci analyse l'impact de l'introduction de la plateforme Parcoursup en 2018 sur la ségrégation à l'entrée des études supérieures en France en fonction du niveau scolaire, de l'origine sociale et du genre des étudiants. Selon les résultats présentés, environ 50% de la ségrégation scolaire et de la ségrégation sociale résulterait de la répartition des étudiants entre grands types de formations (Licence, IUT...). Par ailleurs, une partie importante de la ségrégation à l'entrée dans l'enseignement supérieur serait déjà contenue dans les vœux exprimés par les candidats, les classements de ceux-ci sur la plateforme n'y contribuant que faiblement. Une des principales limites de ce travail est la limitation de son périmètre à la première année d'étude à l'Université, qui résulte de l'utilisation de données administratives de pré-inscription dans le supérieur de 2013 à 2019.

Finalement, si de nombreux travaux abordent la question de la mixité sociale dans l'enseignement supérieur, ceux qui analysent spécifiquement ce sujet au sein de l'Université sont récents et en nombre limité. Un certain nombre de travaux portant sur le secondaire, principalement au niveau Collège, utilisent des indices synthétiques de mesure de la mixité sociale suivant une méthodologie bien établie. Toutefois, ceux-ci ne sont pas systématiquement appliqués à une analyse de la mixité au sein de l'Université alors qu'il existe une base de donnée exhaustive portant sur ce sujet. Il s'agit de la base de données SISE qui contient des informations concernant l'origine sociale de chaque étudiant et la formation dans laquelle il est inscrit. Les outils de mesure utilisés pour le Collège pourraient ainsi potentiellement être appliqués à la base SISE. Toutefois, avant d'appliquer directement ces outils de mesure utilisés pour le Collège à la base SISE, il nous semble nécessaire de clarifier certains éléments méthodologiques de manière préalable.

¹⁸ Le « système d'admission Post-Bac » (APB) était un service accessible en ligne du MESRI entre 2009 et 2017 permettant d'affecter les Lycéens de terminale en première année des formations de l'enseignement supérieur en France. Ce système qui succédait à « RAVEL » (Recensement automatisé des vœux des élèves), réservé à l'Île-de-France, a été remplacé à partir de 2018 par la nouvelle plateforme « Parcoursup » sur laquelle les candidats peuvent formuler leurs vœux d'inscription. La procédure APB a été abandonnée notamment en raison des problèmes liés au tirage au sort dans certaines filières en tension.

4- Trois questions méthodologiques préalables

Au vu des résultats précédents, nous comprenons que les analyses portant sur la mixité sociale dans l'enseignement supérieur ne constituent pas une simple suite de celles abordant cette question dans l'enseignement primaire ou secondaire. Avant de calculer des indices de mixité sociale au sein de l'Université, il nous semble nécessaire de définir un cadre méthodologique adapté à la question de la mixité sociale à l'Université, ce qui implique selon nous de répondre aux trois questions suivantes.

- *Comment définir et mesurer l'origine sociale ?*

S'il existe plusieurs définitions du milieu socio-économique et de l'origine sociale, il nous semble nécessaire d'aboutir à une définition adaptée aux questions d'éducation. Afin de rassembler les individus dans des groupes relativement homogènes d'un point de vue social, par exemple, il nous semble indispensable de clarifier préalablement les concepts de catégorie socioprofessionnelle, de statut socio-économique, de hiérarchie sociale, de « distance à l'école » ou encore de proposer un regroupement par groupes sociaux homogènes d'une manière détaillée¹⁹.

- *Comment définir la mixité sociale à l'Université, mesurer celle-ci au moyen d'indices synthétiques et analyser ses différentes composantes ?*

Il est nécessaire de définir le concept de mixité sociale et de répertorier l'ensemble des facteurs pouvant contribuer à celle-ci dans le cas particulier de l'enseignement supérieur. En l'absence d'une telle définition, il nous semble qu'Avouac et Harari-Kermadec (2021) prennent en compte les facteurs géographiques et institutionnels de la ségrégation alors que les choix individuels et des familles mis en avant par Frouillou (2016) ainsi que les effets de pairs sont davantage mis au second plan.

Il convient de définir un cadre d'analyse adapté pour traiter la question spécifique de la mixité sociale à l'Université. En effet, la répartition des élèves dans l'enseignement

¹⁹ A cette fin, de nombreux travaux peuvent être mobilisés comme par exemple de ceux de Brinbaum et al. (2018), de Ly et Riegert (2014) ou de Rocher (2016).

primaire et secondaire reflète en partie des phénomènes de ségrégation résidentielle, comme l'ont analysé Boutchenik et al. (2018). Toutefois, à l'entrée de l'enseignement supérieur, la situation est distincte, dans la mesure où la majorité des étudiants s'orientent vers une discipline particulière ou un domaine d'études spécialisé et où une proportion non négligeable d'entre eux quittent le domicile parental.

Afin de mesurer le niveau de mixité sociale au sein de l'Université, il est possible de s'appuyer sur des statistiques descriptives, comme le réalise Frouillou (2016) qui souligne la dimension géographique des phénomènes de ségrégation. Les analyses économiques utilisent, toutefois, de manière fréquente des indices synthétiques afin d'évaluer les phénomènes de ségrégation. Ceux-ci font aujourd'hui l'objet d'une littérature largement développée. Celle-ci définit les différents indices disponibles et décrit leurs principales propriétés. Il convient, toutefois, d'adapter ce cadre d'analyse au sujet spécifique de la mixité sociale à l'Université, notamment pour sélectionner les indices pertinents, ceux-ci n'étant pas nécessairement les mêmes suivant que l'objectif est de mesurer la mixité sociale dans une perspective pluriannuelle ou de décomposer les indices calculés.

- *Comment choisir une unité d'observation pertinente pour mesurer la mixité sociale à l'Université ?*

La question de l'unité d'observation, c'est-à-dire de l'unité au niveau de laquelle on mesure la mixité sociale, est problématique au niveau du système universitaire alors que celle-ci ne soulève pas de difficultés particulières au niveau du Collège ou du Lycée. Dans le secondaire, l'unité de mesure de la ségrégation est l'établissement ou la classe, la ségrégation étant alors mesurée entre les établissements ou au sein des établissements eux-mêmes, comme l'expliquent par exemple Ly et al. (2014). A l'inverse, dans l'enseignement supérieur, au moment de définir le champ d'études, différentes questions doivent être examinées. Au vue des effectifs parfois très élevés de certaines Universités, faut-il mesurer les différences entre établissements ou les différences entre filières qui tiennent compte des choix de disciplines ? Est-il nécessaire de tenir compte des niveaux d'études puisqu'une part très significative d'étudiants n'accède pas au niveau du Master ?

5- Les trois objectifs de la thèse

Répondre aux trois questions mentionnées ci-dessus est le premier objectif de notre travail. La réalisation de celui-ci nous permettra de définir un cadre méthodologique cohérent grâce auquel nous pourrons aborder notre deuxième objectif, à savoir mesurer les niveaux de ségrégation existants. Afin de traiter la question de la mixité sociale au sein de l'Université de la manière dont nous l'avons définie, nous nous demanderons dans quelle mesure il existe au sein du système universitaire des phénomènes de ségrégation sociale par établissement, filière ou niveau de formation. Quel était le niveau de mixité sociale existant à l'Université avant la mise en œuvre de la loi « Orientation et réussite des étudiants » (ORE), promulguée en mars 2018 ? La mixité sociale se réduit-elle au fur et à mesure de la progression dans le cursus LMD ? Comment décomposer les indices synthétiques et quelle est la part relative des différents facteurs (tels les différences inter-disciplinaires et les différences géographiques) pour rendre compte du niveau total de ségrégation ?

La réalisation du deuxième objectif nous permettra d'aborder le troisième, à savoir l'analyse des conséquences possibles de la mixité sociale sur les apprentissages. L'un des présupposés implicites mais inhérent au concept de mixité sociale est, en effet, que les individus issus de différents groupes sociaux ne font pas que coexister au sein d'un espace commun, mais interagissent, ce qui est susceptible d'engendrer des effets variés. En ce sens, notre dernier objectif n'est pas accessoire ou subsidiaire, mais, au contraire, indissociable des deux premiers.

Afin de répondre aux trois objectifs identifiés ci-dessus, notre travail est structuré en quatre chapitres de la manière suivante :

- Le chapitre I discute du cadre méthodologique pour la mesure de la mixité sociale à l'Université. Celui-ci définit, d'abord, l'origine sociale dans le contexte de l'économie de l'éducation, répertorie les facteurs permettant de mesurer celle-ci et envisage une méthode pour constituer des groupes sociaux homogènes. Ce chapitre fournit, ensuite, une définition de la mixité sociale à l'Université et sélectionne les indices de mesure les plus adaptés pour répondre à nos objectifs. Celui-ci, enfin, présente les critères permettant de choisir une

unité d'observation adaptée à la mesure de la mixité sociale à l'Université. Ces trois éléments fournissent ainsi un cadre d'analyse qui pourra être utilisé de manière opérationnelle dans les chapitres suivants.

- Le chapitre II a pour objectif d'utiliser le cadre défini dans le chapitre I pour produire des résultats sur le niveau de mixité sociale au sein de l'Université française. Les indices de mesure de la mixité sociale retenus sont d'abord calculés au niveau national. Ensuite, afin d'obtenir une compréhension plus fine des résultats d'ensemble obtenus, ceux-ci sont décomposés par zone académique, des indices synthétiques étant calculés pour chacune d'entre elles.
- Le chapitre III a pour objectif de décomposer suivant différentes dimensions (inter-académiques, inter-universitaires...) les indices calculés dans le chapitre II afin de connaître les parts contributives de chaque facteur au niveau total de la ségrégation observée. Cette décomposition des résultats est réalisée aussi bien au niveau national que pour des sous-populations comme l'Ile-de-France. Le chapitre III constitue ainsi une seconde application plus fine du cadre d'analyse que nous avons défini dans le chapitre I.
- Le chapitre IV propose une analyse du lien entre ségrégation sociale à l'Université et réussite scolaire : si des phénomènes de ségrégation sociale sont identifiés, quelles peuvent être les conséquences potentielles de cette situation sur les apprentissages ? Afin de mesurer les effets de la mixité sociale, nous avons ainsi choisi d'analyser les liens potentiels entre mixité sociale des formations suivies et réussite universitaire (mesurée par le passage en année supérieure). Nous développons ainsi un modèle au sein duquel la probabilité de passage en année supérieure peut dépendre du degré de mixité sociale des formations.

CHAPITRE I : MESURER LA MIXITE SOCIALE A L'UNIVERSITE

L'introduction générale nous a permis de montrer qu'aborder la question de la mixité sociale au sein de l'Université (c'est-à-dire la partie « non sélective » de l'enseignement supérieur français) était une question importante qui n'a été que très peu traitée jusqu'à présent. Pour analyser cette question, une première étape consiste à vérifier que les travaux disponibles²⁰ utilisent un cadre d'analyse commun et à en préciser les contours/les hypothèses.

L'objectif du premier chapitre est ainsi de clarifier certains concepts en répondant aux trois questions suivantes. *Comment définir et mesurer l'origine sociale dans le cadre de l'économie de l'éducation ? Comment définir la mixité sociale à l'Université, mesurer celle-ci au moyen d'indices synthétiques et analyser ses différentes composantes ? Comment choisir une unité d'observation pertinente pour mesurer la mixité sociale à l'Université ?* Les réponses à ces trois questions fourniront ainsi un cadre d'analyse qui nous servira pour développer un travail empirique dans les chapitres II et III.

Notre premier chapitre est structuré de la manière suivante. La première section discute du concept d'*origine sociale*²¹ dans le cadre de l'économie de l'éducation. La deuxième section aborde la mesure de la mixité sociale. La troisième section aborde la question essentielle du choix de l'unité d'observation permettant de mesurer la mixité sociale. La quatrième section fournit un premier cadrage statistique et la cinquième une conclusion.

²⁰ Ceci concerne notamment les travaux suivants : Avouac et Harari-Kermadec (2021), Bechichi et al. (2021), Frouillou (2016), Fack et Huillery (2021).

²¹ Dans le cadre de ce premier chapitre où nous cherchons à définir celui-ci, le concept d'*origine sociale* est écrit en italique. Dans la suite de la thèse, le terme d'*origine sociale* sans italique renvoie à la définition établie à la fin de cette section.

1 L'origine sociale dans le champ de l'éducation

Analyser la question de la mixité sociale à l'Université nécessite de définir préalablement l'*origine sociale* des étudiants. Comme expliqué dans l'introduction générale, cette notion est fréquemment employée dans les travaux empiriques sans être définie de manière univoque. Afin d'obtenir une définition intégrant les principales caractéristiques du concept d'*origine sociale*, il nous a semblé nécessaire d'effectuer une revue de travaux récents portant soit directement sur la question de la mixité sociale au sein des établissements (à tous les niveaux d'enseignement, y compris le primaire et le secondaire), soit sur des sujets connexes comme par exemple celle de l'accès différencié à ceux-ci. Cette section, toutefois, n'a pas vocation, à réaliser une revue de la littérature exhaustive mais a pour objectif de cartographier les concepts à partir d'exemples de travaux académiques. Cela peut impliquer également de notre point de vue de ne pas nous fonder uniquement sur l'économie de l'éducation et de solliciter d'autres disciplines comme les sciences de l'éducation, la sociologie ou même des œuvres littéraires classiques traitant de ce sujet.

La section 1 est organisée ainsi de la manière suivante : le paragraphe 1.1 introduit la définition de l'*origine sociale*, le paragraphe 1.2 présente les types de classification de celle-ci, le paragraphe 1.3 décrit les méthodes permettant de former des groupes sociaux homogènes et le paragraphe 1.4 conclut et fournit une définition opérationnelle de l'*origine sociale* pour notre cadre d'analyse.

1.1 Définir l'origine sociale

Le concept d'*origine sociale* n'est pas facile à appréhender de manière univoque. Afin de le définir plus précisément, le présent paragraphe aborde dans un premier temps le concept d'*origine sociale* dans le contexte très général des sciences sociales (sous-paragraphe 1.1.1) puis, ensuite, dans le contexte plus spécifique de l'économie de l'éducation (sous-paragraphe 1.1.2).

1.1.1 Principes généraux

L'*origine sociale* désigne, dans son acception la plus courante, le milieu socio-économique dont un individu est issu. Pour le chercheur en sciences sociales, ce concept renvoie plus précisément au statut socio-économique de la famille dans

laquelle l'individu a grandi. Le statut socio-économique d'un individu fait référence à la position de celui-ci ou de sa famille dans une structure sociale hiérarchique qui reflète la capacité à mobiliser des ressources telles que les revenus financiers, le prestige, le capital culturel ou encore les relations sociales (Mueller et Parcel, 1981).

Dans ce cadre, afin de caractériser l'*origine sociale* d'un individu, c'est-à-dire les caractéristiques du milieu social dont il provient, une approche multicritères tenant compte d'un faisceau de caractéristiques est idéalement requise. Comme le rappelle à de multiples reprises, nous semble-t-il, le narrateur de l'ouvrage de Marcel Proust *A la Recherche du Temps Perdu*, le revenu ou le patrimoine financier ne permettent pas à eux seuls de situer la position sociale d'un individu. Le personnage de Madame Verdurin qui cherche par différents biais à améliorer sa position sociale est évoquée par celui-ci de la manière suivante : « Mme Verdurin fût elle-même vertueuse et d'une respectable famille bourgeoise excessivement riche et entièrement obscure ». Par opposition à ce personnage, la Duchesse Oriane de Guermantes cumule un ensemble de caractéristiques permettant de la positionner au sommet de la hiérarchie sociale du point de vue du narrateur (grande fortune, culture générale apparente, reconnaissance sociale, nombreuses relations dans le monde de la politique, de la diplomatie et des arts, simplicité des manières et grande aisance relationnelle, origines familiales aristocratiques). L'*origine sociale* repose ainsi sur différents critères et présuppose implicitement l'existence d'une « hiérarchie sociale ». De très nombreux critères sont ainsi susceptibles d'être utilisés pour caractériser l'*origine sociale* d'un individu.

1.1.2 Les différents types de critères utilisés

Dans le cadre de nos travaux, il convient toutefois de noter que nous ne devons pas définir l'*origine sociale* en général, mais dans le contexte spécifique de l'éducation, ce qui implique d'adapter notre définition et de sélectionner uniquement les dimensions pertinentes. En effet, les économistes de l'éducation utilisent souvent le terme d'*origine sociale* pour mesurer implicitement la « proximité » des individus à l'école : il s'agit de définir l'*origine sociale* en fonction du caractère plus ou moins propice de celle-ci à l'acquisition des apprentissages et à la réussite scolaire ou universitaire, des capitaux culturels et scolaires différents étant transmis suivant les milieux sociaux.

Nous avons répertorié dans la littérature récente sur la question neuf critères principaux permettant de caractériser l'*origine sociale* d'un individu dans le cadre de l'économie de l'éducation²² :

1. Les professions et catégories socioprofessionnelles (PCS) dans l'environnement familial.
2. Les niveaux de diplôme dans l'environnement familial.
3. Le revenu ainsi que le patrimoine financier et immobilier de la famille. Le statut de boursier de l'étudiant qui est utilisé dans certains travaux, est directement lié à ce critère.
4. Les ressources éducatives disponibles à la maison.
5. La présence d'une culture classique à la maison.
6. Les relations familiales, avec l'idée que des contacts avec des personnes provenant d'un milieu socialement privilégié facilitent la réussite scolaire et professionnelle de l'individu.
7. Les origines géographiques de la famille, celles-ci pouvant contribuer à caractériser l'*origine sociale* (milieu rural isolé, pays en développement, centre historique de certaines grandes métropoles, etc.).
8. La participation à des clubs ou des activités politiques.
9. La scolarité antérieure de l'individu. Ce critère repose sur l'idée que les établissements dans lequel l'individu a étudié antérieurement peuvent refléter ou non une situation privilégiée.

Il convient de noter que l'indice de statut économique, social et culturel défini par l'OCDE (*International Socio-Economic Index of Occupational Status*, ISEI), utilisé dans le cadre des travaux sur le PISA²³ ou mentionné par exemple dans l'ouvrage OECD (2019) *Balancing School Choice and Equity*, repose uniquement sur les cinq premiers critères. Toutefois, ceux-ci ne sont pas complètement indépendants les uns des autres. Ainsi le critère de PCS est généralement lié au niveau de diplôme et de revenu ainsi qu'éventuellement aux ressources éducatives disponibles à la maison. Celui-ci reflète une dimension à la fois sociale, économique et culturelle, comme nous le

²² Un tableau synthétique permettant d'illustrer chacune de ces critères par quelques travaux est proposé ci-dessous dans le sous-paragraphe 1.2.3.

²³ Les objectifs du Programme International pour le Suivi des Acquis (PISA) sont présentés dans notre introduction générale.

décrivons plus en détail dans le sous-paragraphe 1.2.2 ci-dessous. En ligne avec ce travail, il nous semble que les critères 6 à 9, utilisés dans certains travaux, doivent être considérés comme subsidiaires ou complémentaires. Cette distinction se justifie pour deux raisons : ces quatre derniers critères ne sont généralement pas disponibles dans les bases de données usuelles et, par ailleurs, ils entretiennent généralement des liens étroits avec les cinq premiers et en particulier avec le critère de PCS.

Dans le cadre de notre réflexion sur la mixité sociale à l'Université il peut être envisageable d'utiliser une variable liée au statut de boursier ou non des étudiants. Les bourses d'enseignement supérieur sur critères sociaux (BCS) peuvent, en effet, être attribuées en fonction des études, des diplômes et des ressources. Le critère de boursier permettrait ainsi d'identifier de manière directe les étudiants issus d'un milieu modeste. La population des étudiants boursiers fait aujourd'hui l'objet de travaux de recherche concernant notamment leur réussite aux examens (Fack et Grenet, 2013) ou encore la persistance de barrières pour ces étudiants dans l'enseignement supérieur (Pierrel, 2015).

Un autre critère susceptible d'être mobilisé dans le cadre de notre travail pour caractériser l'*origine sociale* est le niveau de diplôme des parents de l'étudiant (qui peut refléter dans une certaine mesure, le niveau de revenu ainsi que les ressources culturelles et éducatives). Dans ce contexte, les étudiants dont les parents n'ont pas accédé à l'Université font l'objet de travaux spécifiques depuis plusieurs années, les premiers de ce type portant sur les Universités américaines. L'administration des Universités américaines a, en effet, créé la notion d'« étudiant de première génération » qui désigne un étudiant dont aucun parent n'a obtenu au moins un diplôme de niveau Licence. Le lien classique entre niveau d'éducation des parents et trajectoire d'étude des enfants est ainsi mobilisé, mais avec un effet de seuil. Lima et al., (2019) utilisent ce concept dans le cas de la France en utilisant les données de l'enquête « Conditions de vie 2016 » de l'Observatoire Nationale de la Vie Etudiante. Ils constatent notamment que ces étudiants sont surreprésentés dans les filières courtes de l'enseignement supérieur et qu'ils se caractérisent en outre par des taux d'échec et/ou d'abandon plus élevé que les autres étudiants.

Un autre apport de la littérature récente est de prendre en compte pour définir l'*origine sociale* de l'individu non seulement les caractéristiques de ses ascendants mais également de la fratrie (Boutchenik et al., 2015). Ce travail repose sur le constat que

le milieu familial est fréquemment caractérisé dans les analyses à partir du diplôme et de la profession des deux parents de l'individu uniquement. Or, selon les auteurs, la part de variance expliquée par ces variables fournit une estimation biaisée de l'influence du milieu social car elle ignore d'autres aspects familiaux et, en particulier, le rôle majeur joué par les fratries sur le parcours scolaire et professionnel des individus. Cette analyse repose sur des modèles de fratrie, initialement développés par Hauser et Wong (1989) pour mesurer l'influence réciproque entre frères et sœurs dans le cadre de l'éducation en fonction du genre et des âges dans la fratrie.

Un critère complémentaire mais non négligeable pour caractériser l'*origine sociale* d'un individu est l'origine géographique de la famille. Brinbaum et al. (2018) aborde ainsi la question de l'accès différencié à l'Université et de l'obtention différenciée de la Licence en fonction de la PCS des parents, des résultats scolaires antérieurs, des choix de filières de bac, mais aussi des origines géographiques de la famille²⁴ (avec l'usage explicite des trois catégories suivantes : « Français d'origine », « Europe du sud », « Origine maghrébine »). L'utilisation de ce dernier critère par les auteurs repose sur le présupposé que ces origines sont susceptibles d'avoir une influence sur les apprentissages de l'individu via de multiples canaux (maîtrise de la langue, relations sociales, attitude face au travail, etc).

1.2 Les types de classification de l'*origine sociale*

Les critères retenus dans les travaux empiriques pour caractériser l'*origine sociale* dépendent à la fois des objectifs retenus et de la disponibilité des données. Le présent paragraphe décrit ainsi l'appréhension de l'*origine sociale* au moyen d'indices quantitatifs basés sur des variables continues (sous-paragraphe 1.2.1) et des PCS (sous-paragraphe 1.2.2) avant d'introduire une discussion des usages de classification (sous-paragraphe 1.2.3).

²⁴ Bien que les statistiques à caractère ethnique *stricto sensu* soient illégales en France, il est possible d'évoquer la question de la langue maternelle et du pays de naissance des ascendants. Des précisions sont fournies sur ce sujet directement sur le site de l'INSEE (INSEE, 2020).

1.2.1 Les approches basées sur un indice quantitatif continu

L'*origine sociale* peut être appréhendée à partir d'indices quantitatifs basés sur des variables continues. C'est notamment le cas lorsqu'il est possible d'apparier plusieurs bases de données ou de constituer des bases de données *ad hoc*. Kuzmanic et al. (2021) utilisent, par exemple, un double critère de revenu et de niveau de diplôme des parents pour mesurer l'*origine sociale*. La question de la mixité sociale dans l'enseignement supérieur sur la période 2009-2017 au Chili est analysée à partir de deux sources de données : le *Sistema de Información en Educación Superior* (SIES) qui fournit des informations relatives à l'inscription administrative des étudiants et le *Sistema Nacional de Evaluación de Resultados de Aprendizaje*, SIMCE, qui fournit des données sur le niveau de diplôme des parents ainsi que sur le revenu mensuel du ménage. A partir de ces deux informations les auteurs calculent un indice de statut socio-économique. Une pondération plus importante est accordée au niveau d'éducation des parents qu'au revenu, avec une pondération de 0,43 pour le niveau de diplôme du père, de 0,37 pour le niveau de diplôme de la mère et de 0,2 pour le revenu. Cet indice permet de définir une variable d'*origine sociale* de type numérique et de donner à celle-ci une valeur pour chaque étudiant.

Cette approche tenant compte à la fois des revenus et du niveau de diplôme en lien avec le champ de l'éducation doit, toutefois, être distinguée de celles qui reposent uniquement sur les revenus ou le patrimoine financier et permettent par exemple d'identifier les classes moyennes ou aisées de manière générale (hors champ de l'éducation). Ainsi Courtioux et Maury (2020) privilégient une approche en termes de revenus pour identifier les classes moyennes. Trois sous-catégories sont ainsi distinguées dans cette analyse : la « classe moyenne aisée », la « classe moyenne centrale » et la « classe moyenne moins aisée » en fonction de la distance au revenu médian. Fack et Huillery (2021) analysent les inégalités d'accès et de parcours dans l'enseignement supérieur en utilisant non seulement la PCS des parents des étudiants, mais aussi leur décile de revenu. De même Vaughan-Whitehead (2016) utilise une approche fondée sur les revenus pour identifier la classe moyenne, même si d'autres possibilités sont introduites dans ce travail (en particulier, le patrimoine, la PCS et la perception d'appartenir à une classe de revenu).

Dans une perspective différente, Rouban (2015) crée une base de données *ad hoc* pour appréhender les origines sociales des anciens élèves de l'Ecole Nationale

d'Administration. Afin de caractériser une *origine sociale* très favorisée un indice de ressources sociales est calculé de la manière suivante : père ayant une PCS considérée comme supérieure ; passage par une autre Grande Ecole au moins en dehors des I.E.P. ; fonctions politiques d'élu, de militant ou participation fréquente à des activités politiques ; relations familiales dans le monde de la haute fonction publique, de la politique ou des affaires.

L'indice multicritère le plus largement diffusé actuellement est certainement l'indice ISEI (*International Socio-Economic Index of occupational status*) de l'OCDE qui repose sur les cinq critères suivants analysés au niveau de l'environnement familial : les PCS et les niveaux de diplôme, le revenu ainsi que le patrimoine financier et immobilier, les ressources éducatives disponibles et la présence d'une culture classique à la maison. Pour définir les PCS, l'OCDE utilise la classification internationale à quatre chiffres ISCO (*International Standard Classification of Occupations*) définie en 2012 par le Bureau International du Travail. Afin de construire cet indice, l'OCDE diffuse un questionnaire *ad hoc* et a constitué une base de données spécifique.

Si d'un point de vue théorique, l'*origine sociale* renvoie à une série de critères, d'un point de vue pratique, les bases de données ne contiennent fréquemment qu'une information partielle sur l'*origine sociale* des individus. Rouban (2015) parvient à caractériser l'*origine sociale* des anciens élèves de l'ENA suivant quatre critères en raison uniquement des spécificités de l'échantillon analysé. En effet, celui-ci est limité, avec 1 299 individus, et concerne principalement des acteurs publics dont les données biographiques sont accessibles de manière relativement aisée, comme l'explique l'auteur. En revanche, lorsque le champ d'études inclut des dizaines de milliers, voire des centaines de milliers d'observations, une caractérisation de l'*origine sociale* sur la base de quatre critères est difficilement envisageable. Dans le cas des étudiants inscrits à l'Université et dont les données personnelles ne sont généralement pas accessibles en raison des règles du secret statistique, une telle méthode ne saurait être retenue. Par conséquent, il est le plus souvent nécessaire d'employer un nombre de critères plus restreint ou d'utiliser d'autres méthodes afin de caractériser l'*origine sociale* d'un individu.

1.2.2 L'utilisation des professions et catégories socioprofessionnelles

Au vu des difficultés fréquentes dans la pratique pour caractériser l'*origine sociale* d'un individu via une approche multicritères, dans de nombreux travaux, l'*origine sociale* est définie uniquement à partir des PCS des parents (Rocher, 2016). Le choix de ce critère se justifie dans la mesure où le critère de PCS reflète simultanément une dimension sociale, économique et culturelle. Par exemple, Avouac et Harari-Kermadec (2021) justifient l'utilisation des PCS pour mesurer l'*origine sociale* au regard du fait que ceux-ci sont pour un étudiant un « marqueur majeur de son milieu social d'origine » ainsi que des capitaux sociaux, scolaires et culturels transmis. De même Bechichi et al. (2021) expliquent que sur le plan de la réussite scolaire et universitaire, la PCS du parent de référence peut être utilisée car celle-ci reflète une « dotation différentielle en capital socio-économique et scolaire ».

L'INSEE codifie les professions au moyen de la nomenclature professions et catégories socioprofessionnelles (PCS) 2003 ou 2020, les deux faisant l'objet d'une table de correspondance²⁵. La nomenclature PCS 2003 (à laquelle recourt la base SISE), est disponible suivant quatre niveaux d'agrégation. Au niveau le plus fin, un poste de la nomenclature correspond à une profession, codifiée par trois chiffres et une lettre. Au niveau le plus agrégé, il existe huit groupes socioprofessionnels (dont 6 pour les actifs). Les niveaux d'agrégation intermédiaires comportent 24 ou 42 postes. Il convient de noter que la nomenclature PCS 2003 de l'INSEE est fondée non seulement sur le libellé de la profession déclarée, mais aussi sur d'autres dimensions comme la nature de l'emploi (salarié ou indépendant) et de l'employeur (public ou privé) ou encore le niveau de qualification ou la taille de l'entreprise.

La nomenclature des PCS disponible dans la base SISE comprend quant à elle uniquement 32 postes (catégories) dont l'annexe 1 fournit une description détaillée. En effet, la Direction de l'évaluation, de la prospective et de la performance (DEPP) a adapté la nomenclature PCS 2003 de l'INSEE en regroupant certaines des 42 PCS de la nomenclature de l'INSEE. L'annexe 2 du présent travail discute de manière plus détaillée de la représentativité de cette nomenclature des PCS en 32 postes dans la

²⁵ La nomenclature a fait l'objet d'une refonte en 1982, qui a ajouté un niveau détaillé, celui des professions, aux niveaux agrégés des catégories. En 2003, une rénovation plus limitée a été conduite afin de prendre en compte l'évolution des emplois et celle-ci a été à nouveau modifiée en 2020.

mesure où certains travaux de recherche portant sur la mixité sociale en proposent une version modifiée.

L'utilisation des PCS dans le cadre de travaux portant sur la mixité sociale peut se justifier dans la mesure où l'un des objectifs de la nomenclature des PCS est de refléter la réalité des professions mais aussi la structure de la société et, en particulier, les hiérarchies sociales et de diplôme existantes. En effet, dans la version la plus agrégée à huit groupes de la nomenclature des PCS, la catégorie 3 comprend, par exemple, les salariés les plus diplômés (« cadres et professions intellectuelle supérieures »), qui remplissent généralement une fonction hiérarchique, par exemple d'encadrement des salariés classés dans la catégorie 4 (« professions intermédiaires »). Malgré ses limites (la PCS « Policiers et militaires » dans la nomenclature à 42 postes regroupent ainsi l'ensemble des personnels de ces deux institutions dont les statuts et le niveau de diplôme peuvent être très différents), l'utilisation de la nomenclature des PCS se justifie au regard des multiples dimensions que celle-ci reflète de manière directe ou indirecte : le niveau hiérarchique, de revenu, de diplôme, mais aussi, potentiellement, les ressources culturelles et sociales.

Dans une approche économique cherchant à mesurer la productivité ou la performance, un indice quantitatif peut apparaître comme un prolongement naturel de cette approche basée sur les PCS. Rocher (2016) a ainsi proposé un indice de position sociale (IPS). Celui-ci constitue une mesure de la situation sociale des élèves face aux apprentissages, une valeur élevée de l'indice indiquant un contexte familial favorable à ceux-ci. Chaque PCS se voit ainsi attribuer une valeur numérique, celle-ci reposant sur l'hypothèse que dans le domaine de l'éducation, face à l'école et aux apprentissages, il existe une hiérarchie implicite des PCS.

1.2.3 Discussion des usages de classification

Afin d'illustrer la pertinence des neuf critères susmentionnés pour caractériser l'*origine sociale*, nous présentons de manière synthétique dans le Tableau 2 ci-dessous l'usage qui en est fait dans une sélection de travaux récents. Celui-ci comprend ainsi 15 travaux portant sur la mixité sociale dans l'éducation, seuls les travaux de l'OCDE et ceux de Kuzmanic et al. (2021) ne concernant pas la France.

Tableau 2 : critères de définition de l'*origine sociale* dans une sélection de travaux récents abordant la question de la mixité sociale dans l'enseignement secondaire ou supérieur

Travaux	PCS	Niveau de diplôme dans la famille	Revenus ou patrimoine financier et immobilier	Ressources éducatives disponibles à la maison	Présence d'une culture classique à la maison	Relations familiales	Origines géographiques	Participation à des clubs ou des activités politiques	Scolarité antérieure
Avouac et Harari-Kermadec (2021)	X								
Bechichi et al. (2021)	X								
Bonneau et al. (2021)	X						X		
Brinbaum et al. (2018)	X						X		
Courtioux, Maury (2018, 2020)	X								
Fack et Huillery (2021)	X		X						
Frouillou (2016)	X								
Guyon et Huillery (2016)	X						X		
Lima et al. (2019)		X							
Indice <i>ISEI</i> de l'OCDE	X	X	X	X	X				
Kuzmanic et al. (2021)		X	X						
Ly et al. (2014)	X								
Merle (2011)	X								
Rocher (2016)	X								
Rouban (2015)	X					X		X	X

Source: synthèse de l'auteur sur la base des travaux mentionnés.

Note : les références complètes sont précisées en bibliographie. L'ensemble des travaux mentionnés concernent la France, sauf ceux de l'OCDE et de Kuzmanic et al. (2021)

Le Tableau 2 ci-dessus permet d'établir les constats suivants :

- l'indice *ISEI* de l'OCDE est celui qui utilise le plus de critères.
- L'ensemble des travaux à l'exception de Kuzmanic et al. (2021) et de Lima et al. (2019), qui se fondent notamment sur le niveau de diplôme des parents, utilisent la PCS des parents pour caractériser l'*origine sociale*.
- De nombreux travaux utilisent le seul critère de PCS pour mesurer l'*origine sociale* notamment car ce critère est susceptible de refléter différentes dimensions comme le niveau *de revenu et de diplôme*.

1.3 Définir des groupes sociaux homogènes

Les individus d'une population peuvent être répartis en groupes ou catégories relativement homogènes en fonction de leur *origine sociale*. Ces regroupements permettent de mesurer le degré de mixité sociale d'un espace partagé et constituent un prérequis pour le calcul des indices de ségrégation présentés dans le paragraphe 2.3. Dans ce paragraphe, nous présentons les méthodes possibles pour constituer des groupes sociaux homogènes au moyen d'indices quantitatifs continus (sous-paragraphe 1.3.1) et, ensuite, grâce au regroupement des PCS (sous-paragraphe 1.3.2).

1.3.1 Les indices quantitatif continus

Les indices quantitatifs permettent de donner à chaque observation une valeur numérique correspondant à son statut socioéconomique. La définition de seuils permet alors de définir des groupes ou catégories présentant un certain degré d'homogénéité sur le plan socioéconomique. Courtioux et Maury (2020), dans un travail centré sur la place occupée par les classes moyennes dans la société, distinguent ainsi trois sous-catégories de classe moyennes en fonction de seuils de revenus. Kuzmanic et al. (2021), sur la base de l'indice numérique calculé antérieurement, regroupent ainsi les étudiants en dix catégories de statut socioéconomique. De même, les travaux de l'OCDE définissent des groupes ou catégories homogènes à partir de l'indice ISEI.

Ceux-ci se concentrent plus particulièrement sur deux groupes : les étudiants issus d'un milieu socio-économique défavorisé et favorisé. Les premiers sont définis comme le quart des étudiants pour lesquels la valeur de l'indice est la plus basse et les seconds comme le quart des étudiants pour lesquels la valeur de l'indice est la plus élevée. Trois groupes sont ainsi implicitement définis sur la base de l'indice : les étudiants issus d'un milieu défavorisé, ceux issus d'un milieu favorisé et les autres. Il convient de noter que ce type de regroupement présente certaines limites inhérentes à l'usage d'indices quantitatifs, les seuils choisis étant susceptibles de modifier les résultats de la classification. Une solution possible est alors d'effectuer des tests de robustesse destinés à estimer la sensibilité des résultats aux hypothèses choisies.

1.3.2 Le regroupement des PCS

Lorsque l'*origine sociale* est mesurée à l'aune des PCS uniquement, il est possible de regrouper celles-ci pour définir des groupes ou catégories homogènes. La Direction de l'évaluation, de la prospective et de la performance (DEPP) du MESRI regroupe ainsi les PCS suivant quatre groupes sociaux. Selon Rocher (2016), ce regroupement en quatre catégories reflète une hiérarchie implicite suivant leur distance à l'école et une dotation que l'on suppose différenciée en capital socio-économique et scolaire. Toutefois, les critères de ce regroupement n'ont pas été explicités lors de sa création dans les années 1990. Les quatre groupes sociaux identifiés par la DEPP regroupent les PCS suivantes :

- groupe « très favorisé » : « chefs d'entreprise de dix salariés ou plus », « profession libérale », « cadre de la fonction publique », « professeur et assimilé », « profession information, arts, spectacles », « cadre administratif et commercial d'entreprise » et « ingénieur-cadre technique d'entreprise ».
- groupe « favorisé » : « instituteur et assimilé », « profession intermédiaire santé - travail social », « clergé, religieux », « profession intermédiaire administratif fonction publique », « profession intermédiaire administrative commerce entreprise ».

- groupe « intermédiaire » : « agriculteur exploitant », « artisan », « commerçant et assimilé », « technicien contremaître », « agent de maîtrise », « emploi civil - agent service fonction publique », « policier et militaire », « employé administratif d'entreprise », « employé de commerce », « personnel de service direct aux particuliers », « retraité agriculteur exploitant », « retraité artisan, commerçant, chef d'entreprise », « retraité cadre, profession intermédiaire ».
- groupe « défavorisé » : ouvrier qualifié, ouvrier non qualifié, ouvrier agricole, retraité employé et ouvrier, chômeur n'ayant jamais travaillé, autre personne sans activité professionnelle, non renseignée (inconnue ou sans objet)²⁶.

Dans notre cadre d'analyse, l'utilisation de cette méthode de regroupement en quatre groupes se justifie par son utilisation dans de nombreux travaux abordant la question de la mixité sociale dans l'enseignement, ce qui permet d'effectuer des comparaisons des résultats obtenus.

Il convient par ailleurs de noter que dans le cas de notre sujet, si cette méthode de regroupement est utilisée, les parts moyennes occupées par les groupes d'étudiants changent d'une année à l'autre, même de façon marginale, ce qui peut impliquer pour le calcul des indices des problèmes de variance de la composition, comme expliqué dans la section suivante. En effet, les parts des groupes changent au cours du temps et les indices ne sont pas strictement comparables d'une année sur l'autre.

Il est possible de constituer des groupes sociaux à partir de la PCS des deux parents lorsque celles-ci sont toutes les deux disponibles. Dans le domaine de l'économie de l'éducation, une pratique courante, selon Avouac et Harari-Kermadec (2021) consiste à résumer l'*origine sociale* des individus au moyen de la seule PCS du père. Toutefois, lorsque la PCS des deux parents est disponible, il ne nous semble pas qu'il y ait de justification à utiliser la seule PCS du père. En effet, la PCS des mères joue un rôle non négligeable dans le parcours scolaire des enfants, comme l'ont montré Place et Vincent (2009). Ainsi, Courtioux (2011) propose un regroupement des groupes sociaux en six catégories reposant sur la PCS des deux parents. La définition de catégories ou groupes sociaux lorsque la PCS des deux parents est disponible peut s'avérer

²⁶ L'annexe 1 fournit des détails supplémentaires sur la méthode de regroupement retenue par la DEPP.

cependant plus complexe que lorsqu'une seule PCS est prise en compte. Il convient alors d'envisager une typologie des situations possibles en fonction des groupes d'appartenance des deux parents, comme nous le proposons dans l'encadré 4 ci-dessous.

Encadré 4 : proposition de regroupements lorsque la PCS des deux parents est prise en compte et que la population de référence est composée de trois groupes sociaux

Si la population de référence est composée de trois groupes sociaux, par exemple favorisé (A), intermédiaire (B) et défavorisée (C) :

- L'origine familiale est considérée comme favorisée (A) dans cinq cas : si les deux parents appartiennent au groupe A, si le parent 1 appartient au groupe A et le parent 2 au groupe B, si le parent 1 appartient au groupe A et le parent 2 au groupe C, si le parent 2 appartient au groupe A et le parent 1 au groupe B, si le parent 2 appartient au groupe A et le parent 1 au groupe C.

- L'origine familiale est considérée comme intermédiaire (B) dans trois cas : si les deux parents appartiennent au groupe intermédiaire (B), si le parent 1 appartient au groupe B et le parent 2 au groupe C, si le parent 2 appartient au groupe B et le parent 1 au groupe C.

- L'origine familiale est considérée comme défavorisée dans un seul cas : si les deux parents appartiennent au groupe défavorisé (C)²⁷.

Source : auteur

1.4 Conclusion de la section 1

Différentes méthodes sont utilisées pour constituer des catégories ou groupes sociaux homogènes à partir de leur *origine sociale*. Le Tableau 3 ci-dessous présente ainsi de manière synthétique les choix effectués sur ce point dans la sélection de travaux récents déjà mentionnés.

²⁷ Comme dans le cas de la méthode de la DEPP retenue pour définir l'*origine sociale*, nous faisons l'hypothèse qu'une absence de réponse renvoie à une origine sociale défavorisée.

Tableau 3 : regroupement d'individus d'origines sociales différentes en groupes homogènes dans une sélection de travaux récents

Critères	Constitution de groupes sociaux	Nombre de groupes sociaux identifiés	Critères utilisés pour le regroupement	Utilisation de la méthode de regroupement de la DEPP	Prise en compte de la PCS des deux parents
Avouac et Harari-Kermadec (2021)	Oui		PCS	Non	Non
Bechichi et al. (2021)	Oui	4	PCS	Oui	Non
Bonneau et al. (2021)	Oui	4	PCS	Oui	Non
Brinbaum et al. (2018)	Non	---	---	---	Non
Courtioux et Maury (2018)	Oui	4	PCS	Oui	Non
Courtioux (2011)	Oui	6	PCS	Non	Oui
Frouillou (2016)	Non	---	---	---	Non
Guyon et Huillery (2016)	Oui	2	PCS	Non	Non
Indice <i>ISEI</i> de l'OCDE	Oui	3	Indice multicritères	Non	Non
Kuzmanic et al. (2021)	Oui	10	Indice multicritères	Non	Non
Ly et al. (2014)	Oui	4	PCS	Oui	Non
Merle (2011)	Oui	4	PCS	Oui	Non
Rocher (2016)	Non	---	---	---	Non
Rouban (2015)	Non	---	---	---	Non

Source : synthèse de l'auteur sur la base des travaux mentionnés. Note : la DEPP désigne la Direction de l'évaluation, de la prospective et de la performance.

Le Tableau 3 ci-dessus permet d'établir les constats suivants concernant cette sélection de travaux récents :

- Une majorité de travaux reposent sur la définition de groupes sociaux homogènes. Les analyses qui n'utilisent pas de méthode de regroupement se fondent directement sur l'observation de certaines PCS comme Brinbaum et al. (2018) ou sur l'utilisation d'une valeur quantitative de l'*origine sociale* comme Rocher (2016) ou Rouban (2015).
- Parmi les travaux qui définissent des groupes sociaux homogènes, le nombre de catégories répertoriées va de 3 à 10. Les catégories se situant aux extrêmes

de la distribution (groupe « défavorisé » et « favorisé » ou « très favorisé ») font généralement l'objet d'une attention spécifique.

Dans le contexte spécifique de l'économie de l'éducation, l'*origine sociale* peut être appréhendée à partir de neuf types de critères qui permettent ensemble de mesurer la « distance à l'école » de l'individu : la PCS, le niveau de diplôme, le revenu ainsi que le patrimoine financier et immobilier de la famille, les ressources éducatives disponibles à la maison, la présence d'une culture classique à la maison, les relations familiales, les origines géographiques de la famille, la participation à des clubs ou des activités politiques et la scolarité antérieure de l'individu. De notre point de vue, les cinq premiers critères peuvent être considérés comme fondamentaux alors que les quatre derniers sont davantage complémentaires ou subsidiaires pour les raisons mentionnées ci-dessus au sous-paragraphe 1.1.2.

Si d'un point de vue théorique, une solution idéale consiste à prendre en compte simultanément l'ensemble de ces critères, dans la pratique, les bases de données ne contiennent fréquemment que des informations limitées. La base de données SISE, telle qu'elle est diffusée via l'ADISP-PROGEDO n'inclut ainsi d'information ni sur le revenu des parents ni sur le statut de boursier. Celle-ci ne renseigne pas non plus le niveau de diplôme des parents ou le statut de « première génération » des étudiants. En l'absence d'information disponible sur les revenus, le patrimoine ainsi que le niveau de diplôme des parents des étudiants, une solution fréquemment utilisée dans la littérature est de caractériser l'*origine sociale* des individus à l'aune du seul critère de la PCS. Ce choix se justifie dans la mesure où ce critère comporte une dimension sociale, économique et culturelle, comme nous l'avons expliqué antérieurement dans le sous-paragraphe 1.2.2. Il n'est toutefois envisageable que si la nomenclature des PCS est suffisamment détaillée pour distinguer et caractériser différents groupes sociaux. Il facilite, par ailleurs, les comparaisons avec de nombreux autres travaux utilisant également ce critère.

2 Mesurer la mixité sociale

Afin d'analyser la question de la mixité sociale à l'Université de manière empirique, il nous semble préférable de définir préalablement ce que l'on entend par mixité sociale et son contraire, la ségrégation sociale. Comme expliqué dans l'introduction générale, cette notion, tout comme celle d'*origine sociale*, est fréquemment utilisée dans les travaux empiriques dans plusieurs sens, la présente section pouvant contribuer à distinguer ces différentes acceptions. Afin de parvenir à une définition que nous pourrions utiliser pour notre cadre d'analyse, il nous a semblé nécessaire d'effectuer une revue de travaux récents portant sur cette question. Ceci peut impliquer de mobiliser des sources variées telles que des travaux à caractère géographique ou sociologique. La section 2 est structurée ainsi de la manière suivante : le paragraphe 2.1 définit les éléments nécessaires à la mesure la mixité sociale, le paragraphe 2.2 présente les propriétés attendues des indices de mixité, le paragraphe 2.3 présente les principaux indices de mesure de la mixité sociale et le paragraphe 2.4 conclut et propose une définition opérationnelle de la mixité sociale que nous utiliserons dans notre cadre d'analyse.

2.1 Les éléments nécessaires à la mesure

Le présent paragraphe introduit d'abord le concept de population de référence dans le contexte de la mixité sociale (sous-paragraphe 2.1.1), expose, ensuite, les unités d'observation pertinentes pour mesurer la mixité sociale dans le champ de l'éducation (sous-paragraphe 2.1.2) et introduit finalement le principe des indices synthétiques de mesure de la mixité sociale (sous-paragraphe 2.1.3).

2.1.1 Mesure et population de référence

La « mixité sociale » a été définie dans l'introduction générale de ce travail de façon intuitive comme la cohabitation de personnes ayant des origines sociales variées au sein d'un espace commun, d'un lieu partagé. Toutefois, cette première définition intuitive peut être précisée à l'aide du concept de « population de référence » : lorsqu'une population de référence est divisée ou répartie en différentes sous-

populations localisées géographiquement, on peut considérer qu'il existe une situation de « mixité sociale » lorsque la composition sociale de celles-ci est semblable ou proche de celle de la population de référence. Si dans une population de référence (par exemple une ville) le groupe A représente par exemple 80% de la population totale, le groupe B 10% et le groupe C 10%, il existe une situation de mixité sociale lorsque les parts occupées par chacun des groupes au sein des sous-populations (par exemples les quartiers d'une ville) sont semblables ou proches de celle qu'ils occupent dans la population de référence. Cette définition classique de la mixité sociale est celle sur laquelle repose implicitement les indices de mixité sociale (de ségrégation) en général. En effet, comme le rappelle par exemple Bechichi et al. (2021), leur objectif est de « mesurer à quel point les caractéristiques des individus regroupés au sein d'unités éducatives (ici, des formations d'enseignement supérieur) s'éloignent des caractéristiques moyennes de la population considérée dans son ensemble ».

Par opposition à la mixité sociale, lorsqu'une population de référence est divisée ou répartie en sous-populations, il existe une situation de « ségrégation sociale » si la composition sociale de celles-ci est éloignée de celle de la population de référence.

Mesurer la mixité sociale (la ségrégation) revient ainsi à comparer les parts occupées par les groupes sociaux au sein de la population de référence identifiée et des sous-populations qui la composent. Si la répartition des individus d'un groupe en plusieurs sous-groupes ou unité, que nous pouvons noter « k », ne se fait pas de manière complètement aléatoire, mais résulte par exemple des préférences individuelles ou des appartenances collectives, alors la composition des unités n'est plus identique à celle de la population de référence et il existe un certain degré de ségrégation au sens où nous l'avons défini.

2.1.2 Rappel des principaux facteurs susceptibles d'influencer les niveaux de ségrégation sociale dans le champ de l'éducation

Les raisons pour lesquelles il peut y avoir un degré limité de mixité sociale et donc une composition sociale des unités « k » éloignée de celle de la population de référence sont multiples. Dans la littérature, il nous semble possible de distinguer cinq principaux facteurs susceptibles d'influencer les niveaux de ségrégation sociale dans le champ de l'éducation :

1) *la ségrégation résidentielle (spatiale)*, qui est particulièrement importante dans le cas de l'enseignement primaire et secondaire, la majorité des élèves étant scolarisés dans l'établissement de leur secteur en raison de la carte scolaire, qui permet d'affecter les élèves dans une école, un Collège ou un Lycée en fonction de la localisation géographique de leur domicile. Dans le cas de la France, il semble que la répartition des élèves entre les établissements reproduit dans une grande mesure la répartition entre quartiers de résidence (le lieu de résidence peut, par ailleurs, bien sûr être choisi ou modifié en fonction des établissements qui correspondent à la carte scolaire)²⁸. Le rapport de l'OCDE *Balancing School Choice and Equity* (2019) constate également l'importance de la ségrégation résidentielle, même si ce phénomène a eu tendance à se réduire depuis plusieurs années dans la moyenne des pays de l'OCDE, la répartition géographique des élèves dans les Collèges dépendant moins du lieu de résidence en 2015 qu'en 2000. Les analyses de Boutchenik et al. (2018) sur le Collège confirment le rôle central joué par la carte scolaire, qui reflète la ségrégation résidentielle, sur les niveaux de ségrégation mesurés au Collège, même si les « stratégies d'évitement » que nous décrivons dans le paragraphe suivant sont également essentielles.

Ce facteur de ségrégation résidentielle persiste également au niveau de l'Université même si celui-ci y joue un rôle plus faible²⁹. Brodaty et Jacotin (2016) rappellent ainsi qu'avec le système « RAVEL », qui était encore utilisé avant 2009 en Ile-de-France, les bacheliers étaient affectés dans les Universités de cette région en fonction de la localisation de leur Lycée et de la discipline souhaitée. L'affectation des bacheliers reposait ainsi sur une base essentiellement géographique, indépendamment du niveau scolaire des futurs étudiants. Toutefois, ce

²⁸ Fack et Grenet (2009) relèvent ainsi que la carte scolaire pourrait renforcer cette ségrégation résidentielle, en orientant les choix de domiciliation des familles aisées vers les quartiers bénéficiant des meilleurs établissements.

²⁹ La section 1 du chapitre II consacrée à la dimension géographique de la ségrégation sociale décrit de façon plus précise la place du facteur résidentiel.

facteur de ségrégation est moins important au niveau de l'Université en raison de l'absence de principe de carte scolaire et du nombre élevé d'étudiants qui quittent le domicile parental au moment de l'entrée dans l'enseignement supérieur. Ainsi, en 2006, au niveau national, 41,6% des étudiants avaient obtenu leur bac dans une autre académie que celle où ils étaient inscrits. Ce taux atteignait un maximum de 77,1% dans l'académie de Paris et un minimum de 28,1% dans l'académie d'Amiens.

2) *les stratégies d'évitement liées à des choix individuels et familiaux* (au-delà du choix concernant le lieu de résidence). Dans le cas de l'enseignement primaire et secondaire, ce facteur correspond aux stratégies d'évitement destinées à contourner la carte scolaire. Ces stratégies peuvent être mises en œuvre par une famille via le choix des options disponibles dans les établissements ou encore le recours aux établissements privés. Merle (2011) décrit ainsi la façon dont l'assouplissement de la carte scolaire depuis 2007 autorise des dérogations auparavant proscrites, cette plus grande liberté profitant essentiellement aux catégories sociales moyennes et supérieures qui y recourent fortement afin de contourner les établissements dont la réputation ou le niveau d'exigence ne leur paraît pas suffisant³⁰. Boutchenik et al. (2018) constatent également que dans une majorité de pays, ce sont d'abord les familles des milieux sociaux les plus favorisés qui utilisent la possibilité de choisir un établissement scolaire. Dans le cas de la France, les auteurs montrent que si la ségrégation sociale observée entre Collèges reflète en partie la ségrégation résidentielle, les choix des familles amplifient nettement celle-ci à l'entrée en 6^{ème}. A Paris, Bordeaux et Clermont-Ferrand celle-ci serait entre un tiers et moitié moins élevée si les élèves étaient tous scolarisés dans le Collège public de leur quartier. Ce surcroît de ségrégation s'explique ainsi avant tout par le choix de certaines familles en faveur de l'enseignement privé.

³⁰ Dans le cas des Etats-Unis, Reardon et al. (2000) constatent ainsi que la scolarisation dans les écoles privées (*private schools* ou *charter schools*) des élèves relativement favorisés contribue à accentuer la concentration d'élèves issus des minorités dans les écoles publiques, renforçant la ségrégation globale.

Dans le cas de l'Université, ces stratégies peuvent également opérer afin d'éviter par exemple une formation au sein d'un établissement dans lequel le niveau moyen des élèves ou les exigences sont considérés comme insuffisantes par l'étudiant ou sa famille. Frouillou (2016) documente ainsi pour l'Ile-de-France les différentes stratégies d'« évitement » et de « placement » universitaire que les étudiants sont susceptibles de déployer en fonction de leurs ressources individuelles et familiales. Comme nous le décrivons de manière détaillée dans le chapitre II, ce choix peut se faire au profit d'un établissement non universitaire, ce qui renvoie alors à la question de l'accès différencié au système d'enseignement supérieur (comme expliqué dans l'introduction générale), mais il peut également se faire au profit d'un autre établissement universitaire, ce qui contribue à favoriser des logiques de ségrégation au sein de l'Université.

- 3) *les effets de pairs*³¹, qui désignent le processus par lequel des individus ayant un environnement commun s'influencent mutuellement par l'intermédiaire de leurs caractéristiques et de leur comportement (Fougère et al., 2018). Ces effets de pairs peuvent jouer un rôle essentiel, en particulier au moment de l'entrée dans l'enseignement supérieur. Frouillou (2016) relève ainsi que les futurs étudiants qui ont peu de capital scolaire ont davantage tendance que les étudiants plus favorisés à se laisser influencer par les choix des pairs, c'est-à-dire de leurs camarades de Lycée, au moment de sélectionner une Université. Il convient de noter que si les effets de pairs peuvent potentiellement se recouper et/ou interagir avec les deux premiers facteurs (la ségrégation résidentielle et les choix individuels et familiaux)³², d'un point de vue conceptuel, il s'agit, toutefois, de phénomènes distincts. Les choix individuels et familiaux peuvent, par exemple, résulter d'objectifs

³¹ Les effets de pairs sont mentionnés de manière synthétique dans le présent paragraphe, ceux-ci faisant l'objet d'une description spécifique et détaillée dans le cadre de notre chapitre IV.

³² Les phénomènes ne sont pas indépendants : le choix du Lycée et donc des pairs peut être influencé voire conditionné par la localisation géographique et les choix de la famille.

personnels qui ne sont pas nécessairement influencés par les pairs identifiables dans l'environnement immédiat.

- 4) *les politiques des établissements ainsi que leurs stratégies de communication.* Ce facteur concerne particulièrement les établissements qui cherchent à se donner une image élitiste et sont susceptibles d'attirer davantage les étudiants les plus favorisés. Dans le cas de la France, Frouillou (2016) montre ainsi que pour ces établissements le nombre de doubles Licences est, par exemple, essentiel. De même, ce travail montre que la politique de communication de l'Université Paris 1 cherche à mettre au premier plan son passé, ses locaux prestigieux ainsi que sa localisation en plein cœur du Quartier latin alors que l'Université Paris VIII, par exemple, met davantage en valeur l'effectif plus réduit de ses étudiants dans les amphis et les cours ainsi que son encadrement soutenu.

- 5) *le cadre légal et les politiques publiques mises en œuvre au plan national.* Ce facteur est lui-même susceptible de comprendre plusieurs dimensions. Sur le plan juridique *stricto sensu*, un exemple provient des Etats-Unis avec les divers arrêts de la Cour Suprême depuis les années 1970 qui ont favorisé la déségrégation et encouragé les efforts des districts scolaires à progresser en ce sens. En 1971, ainsi, l'arrêt Swann contre Charlotte Mecklenburg instaure le principe d'acceptabilité de quotas raciaux lorsqu'une situation de ségrégation est avérée. Dans le cas de l'Université française, le travail d'Avouac et Harari-Kermadec (2021) fait l'hypothèse que les niveaux de ségrégation constatés entre Universités résultent dans une certaine mesure des politiques d'enseignement supérieur décidées au niveau national et notamment des politiques liées aux investissements d'avenir, votés dans le cadre de la loi de finances rectificative du 9 mars 2010. En effet, l'objectif de cette loi est de soutenir des projets innovants dans un certain nombre de secteurs jugés prioritaires, les Universités bénéficiant d'un label d'excellence attirant alors davantage les étudiants favorisés ou très favorisés.

Il nous semble essentiel de ne pas perdre de vue le rôle potentiel de ces cinq facteurs dans nos analyses afin de conserver une vision globale du problème de la ségrégation au sein de l'Université française. Nous comprenons ainsi que l'analyse d'Avouac et Harari-Kermadec (2021) qui adopte comme unité d'observation l'Université prend en compte uniquement les deux derniers facteurs de ségrégation, une des hypothèses sous-jacentes à leur travail étant que la ségrégation dépendrait essentiellement des choix institutionnels et des politiques éducatives. Il nous semble, en revanche, que Frouillou (2016) tient compte de ces cinq facteurs différents pour appréhender la question de la mixité sociale dans l'enseignement supérieur. Il convient, par ailleurs, de noter que ces facteurs, conceptuellement distincts, sont susceptibles d'interagir entre eux, en particulier pour les trois premiers.

2.1.3 Le principe des indices de mixité sociale

Les indices de mixité sociale (de ségrégation), de manière générale, reposent sur le principe d'une comparaison entre la part occupée par chacun des groupes sociaux dans une population et celle qu'ils occupent au sein des différentes unités dans lesquelles cette population a été répartie. Selon Monso et al. (2015), la composition de la population de référence sert ainsi nécessairement de point de comparaison.

Le développement de ces indices et le recours intensif à ceux-ci dans les travaux récents traitant de la question de la ségrégation dans l'enseignement est justifié au regard des limites intrinsèques aux statistiques descriptives. Celles-ci ne fournissent pas une mesure directe du degré de ségrégation et ne permettent pas d'effectuer des comparaisons d'ensemble entre plusieurs situations en raison de la multiplicité des informations fournies. Faute de recours aux indices synthétiques, seuls les contrastes les plus marqués, les comparaisons les plus emblématiques sont mises en avant. Ainsi, dans le travail de Frouillou (2016), les formations susceptibles de présenter des niveaux de contraste élevés (le Droit à l'Université Paris I, l'AES à l'Université Paris VIII...) sont mises en avant. Le recours aux indices synthétiques de mixité sociale constitue ainsi une étape plus générale et complémentaire d'analyse.

Par rapport aux statistiques descriptives, l'avantage d'un indicateur synthétique, comme précisé par Frankel et Volij (2011), est de résumer en un seul chiffre une série de statistiques, ce qui présente les avantages suivants :

- résumer à l'aide d'un seul chiffre le degré de mixité d'une population de référence répartie en plusieurs unités. Faute d'indice, il n'y a qu'une série de données statistiques relatives à une population de référence et aux unités qui la composent.
- prendre en compte le poids de tous les groupes sociaux qui composent la population alors qu'à la seule lecture des statistiques descriptives, le risque est de se focaliser sur certains d'entre eux et donc de limiter l'analyse, à la comparaison de deux catégories situées par exemple aux extrêmes de la distribution.
- effectuer des comparaisons immédiates entre différentes populations ou différentes situations.

2.2 Les propriétés attendues des indices de mixité

Un grand nombre d'indices synthétiques de mesure de la mixité sociale sont documentés dans la littérature ainsi que leurs avantages et inconvénients respectifs. Les travaux de Bourguignon (1979), Hutchens (2001) Mora et al. (2003), ainsi que ceux de Reardon et Firebaugh (2002) et Frankel et Volij (2011) font partie des travaux de référence sur cette question. L'article de référence de Reardon et Firebaugh (2002) répertorie ainsi sept critères. Les analyses ainsi que la terminologie employée dans le présent paragraphe reposent principalement sur l'article plus récent de Frankel et Volij (2011), qui définit 10 critères. Dans la mesure où notre objectif ici est uniquement d'identifier des propriétés permettant de choisir les indices les plus adaptés pour notre travail, nous avons retenu uniquement trois propriétés issues de ce travail et proposé une quatrième. Le sous-paragraphe 2.2.1 présente ainsi le premier critère : un degré de mixité supérieur au sein de la population de référence qu'au sein des sous-populations. Le sous-paragraphe 2.2.2 introduit le second, l'invariance à la composition et le sous-paragraphe 2.2.3 décrit le troisième, à savoir la décomposabilité forte. Le sous-paragraphe 2.2.4 traite du caractère facilement interprétable des indices. Le sous-paragraphe 2.2.5 présente finalement les autres critères de sélection qui nous paraissent moins pertinents pour notre sujet.

2.2.1 Un degré de mixité de la population de référence au moins égal à celui des unités

Frankel et Volij (2011) démontrent que lorsqu'une population est divisée en plusieurs sous-unités, la valeur de l'indice ne doit pas diminuer (« *school division property* »). En effet, si une ou plusieurs sous-unités étaient plus mixtes socialement que les unités dont elles sont issues, cela impliquerait un niveau de ségrégation plus élevé dans d'autres sous-unités, ce qui contrebalancerait le premier effet. Cette propriété peut être illustrée au moyen de l'exemple suivant. Si l'ensemble des étudiants inscrits dans le système universitaire sont répartis par établissement universitaire (par Université), l'unité d'observation est alors l'Université. Si chacune de ces unités, est elle-même subdivisée, dans un second temps, par matière, l'unité d'observation devient « Université x matière ». Le degré de ségrégation observé, dans ce cas, est susceptible d'augmenter, éventuellement de demeurer identique, mais non de diminuer. En effet, dans le cas le plus favorable, les étudiants se répartissent de manière aléatoire au sein de chaque Université et, alors, le degré de mixité sociale au sein de la population de référence demeure identique, malgré la subdivision des unités d'observation et la valeur mesurée des indices de ségrégation doit demeurer également identique. Cela se comprend dans la mesure où les sous-unités ne peuvent être en moyenne plus mixtes socialement que les unités dont elles sont issues. Il est donc nécessaire que les indices mesurant la ségrégation reflètent cette réalité.

2.2.2 L'invariance à la composition

Un deuxième critère que doivent idéalement respecter les indices de mesure est celui d'invariance à la composition. Selon celui-ci, la valeur de l'indice ne doit pas varier en fonction de la part occupée par chacun des groupes au sein de la population de référence. Ainsi, si la part occupée par un groupe double au sein de la population de référence, mais aussi au sein de chaque unité, alors la valeur mesurée de l'indice doit demeurer identique. Cette propriété établit que si le nombre d'individus d'une catégorie donnée venait à augmenter, mais que ces nouveaux individus étaient répartis dans les formations selon la répartition initiale, alors l'indice de ségrégation ne changerait pas. L'invariance à la composition permet donc d'effectuer des comparaisons géographiques ou temporelles : les niveaux de ségrégation de deux académies différentes (ou pour deux années académiques différentes) peuvent être comparés

même si leur composition, c'est-à-dire les parts occupées par les différents groupes sociaux, ne sont pas les mêmes. Dans ce cas, on mesure, en effet, l'écart à la population de référence pour une année donnée.

2.2.3 La « décomposabilité forte » des unités

Un troisième critère de sélection est celui de la « décomposabilité forte » des unités, mentionné par Frankel et Volij (2011). Si ce critère est respecté, la valeur finale de l'indice calculé est égale à la somme pondérée des indices mesurant la mixité des composants des unités (*strong school decomposability*).

L'utilisation de ce critère se justifie au regard des objectifs de notre analyse. Dans le cadre de notre travail, les phénomènes de ségrégation peuvent, en effet, résulter de différents éléments, comme nous l'avons décrit dans le sous-paragraphe 2.1.2. Si l'unité d'observation choisie est, par exemple, la formation, soit « Université x matière », les phénomènes de ségrégation peuvent aussi bien résulter de différences entre les Universités qu'au sein de chaque Université de phénomènes de ségrégation inter-disciplinaires. Afin d'identifier ces éléments et d'estimer leur importance relative, il est nécessaire de pouvoir décomposer les valeurs calculées des indices en différents éléments.

A quoi correspond dès lors l'exigence de « décomposabilité forte »? Un indice est considéré comme fortement ou hautement décomposable si et uniquement si la valeur de l'indice est égale à la somme pondérée de ses composantes, cette pondération correspondant à l'effectif de chacune. Dans le cadre de notre travail, si les étudiants sont regroupés « par Universités et par discipline », l'indice peut être considéré comme « fortement décomposable » dès lors que la valeur de celui-ci est égale à la somme pondérée (en fonction de l'effectif de chaque Université) de deux éléments : l'indice mesurant le degré de ségrégation inter-universitaire et, pour chaque Université, l'indice mesurant le degré de ségrégation inter-disciplinaire. Le respect du principe de « forte décomposabilité » permet ainsi dans ce cas d'indiquer quelle part de la ségrégation est liée à la répartition entre Universités et quelle part résulte de la répartition entre disciplines académiques au sein des Universités. Un indice « faiblement » décomposable, comme l'expliquent Frankel et Volij (2011), ne permet pas d'identifier l'importance relative de chaque composante dans le niveau total de ségrégation, notamment parce que la pondération utilisée est complexe et ne reflète pas l'effectif

de chaque Université. Seule la « décomposabilité forte », permet, dans notre cas, de savoir, par exemple, de combien serait réduite la valeur de l'indice s'il n'existait plus d'écart de composition entre les Universités. Givord et al. (2015) précisent ainsi que les décompositions obtenues au moyen de l'Indice d'entropie normalisé sont de type additif et ne sont pas parfaites. Les termes de pondération employés ne permettent pas de quantifier exactement de combien diminuerait la valeur totale de l'indice si l'une de ses composantes disparaissait.

2.2.4 La facilité d'interprétation

Si la définition de ce critère n'appelle pas de commentaire particulier, il nous semble en revanche nécessaire de justifier son utilisation comme critère de choix d'indice à part entière des indices synthétiques de mesure de la mixité sociale. En effet, dans les travaux empiriques, ce critère est fréquemment utilisé pour justifier du choix d'un indice, en sus des propriétés mathématiques présentées par exemple dans Frankel et Volij (2011). Ainsi, Bechichi et al. (2021) retiennent deux critères pour choisir un indice synthétique de mesure de la mixité sociale comme base de leur travail : la simplicité d'interprétation et sa propriété de décomposabilité. A l'aune de ces deux critères, ils sélectionnent l'indice d'exposition normalisé. Ly et Riegert (2015), justifient également le recours à l'indice d'exposition normalisé sur la base de ces deux critères. Il nous semble donc important de retenir la facilité d'interprétation comme un critère de sélection explicite et à part entière des indices. Ainsi, les indices normalisés peuvent être considérés comme relativement faciles à interprétés car ils sont rapportés à 100.

2.2.5 Les autres propriétés

D'autres propriétés des indices sont également répertoriées dans la littérature. Toutefois, celles-ci nous semblent moins prioritaires dans le cadre de notre sujet car elles sont soit redondantes avec certaines propriétés retenues ou correspondent à des situations qui ne se présentent pas, comme indiqué dans le Tableau 4 ci-dessous.

Tableau 4 : propriétés des critères non retenus et les raisons de leur rejet

N	Propriété	Définition	Raison pour laquelle nous ne retenons pas ce critère
1	Continuité	Un indice de ségrégation est continu s'il est une fonction continue du nombre de membres de chaque groupe dans chaque unité.	Respectés par les huit indices
2	Symétrie	Un indice est symétrique si la valeur de celui-ci n'est pas impactée lorsque le nom des groupes est interchangé.	La situation ne se présente pas.
3	Indépendance	Si des individus sont réaffectés au sein d'une unité, alors le degré de ségrégation de la population totale augmente seulement si celui-ci augmente au sein de cette unité.	La situation ne se présente pas.
4	Divisibilité des groupes (<i>Group Division Property</i>)	Diviser un groupe en deux sous-groupes n'a pas d'impact sur le niveau mesuré de l'indice si la distribution de ce groupe est identique dans les différentes unités.	La situation ne se présente pas.
5	Agrégation (<i>Aggregation, AGG</i>)	Un indice est agrégatif si la valeur mesurée de l'indice peut être calculée comme une fonction continue de ses composantes.	Critère redondant avec celui de décomposabilité forte (<i>strong school decomposability</i>).
6	Décomposabilité additive (<i>Additive decomposability, AD</i>)	Un indice est décomposable de manière additive si la valeur mesurée de l'indice peut être calculée comme somme pondérée de ses composantes.	Critère redondant avec celui de décomposabilité forte (<i>strong school decomposability</i>).
7	Forte décomposabilité des groupes (<i>Strong group decomposability, SGD</i>)	Pendant de la <i>Strong School Decomposability</i> abordée par Frankel et Volij : si, dans le cas d'une subdivision d'un groupe en plusieurs sous-groupes, la valeur mesurée de l'indice est égale à une somme des indices de chaque sous-groupe pondérée en fonction de leur importance numérique.	La situation ne se présente pas au sens où les groupes sociaux identifiés ne sont pas subdivisés en sous-groupes (par exemple suivant le genre).

Source : Frankel et Volij (2011) pour les deux premières colonnes. Troisième colonne : application des principes définis à notre sujet.

Suivant le Tableau 4 ci-dessus, certains critères ne sont pas retenus pour les raisons suivantes :

- le critère est respecté par l'ensemble des indices (ou au moins par tous ceux que nous présenterons dans le cadre de notre travail et qui seront introduits ci-dessous) et n'est donc pas discriminant. C'est le cas du critère n°1.
- la situation ne se présente pas dans le cadre d'une analyse de la mixité sociale à l'Université française. En effet, si la situation décrite est envisageable d'un point de vue théorique, de notre point de vue, soit celle-ci ne peut être décrite avec les bases de données existantes, soit celle-ci n'est *a priori* pas pertinente dans le cadre d'une analyse de la mixité sociale à l'Université. C'est le cas des

critères n°2, 3, 4 et 7. Ainsi, dans le cas du critère n°7, si une subdivision des groupes suivant le genre est envisageable au moyen des bases de données usuelles, il nous semble que celle-ci serait requise dans un champ d'investigation distinct.

- les critères n°5 et 6 correspondent à des conditions nécessaires mais non suffisantes du critère de décomposabilité forte déjà retenu. Seule la décomposabilité forte nous semble être un critère pertinent dans le cadre de notre travail, les autres types de décomposabilité ne nous permettant pas de quantifier le rôle joué par chacune des composantes dans le niveau de ségrégation total. Toutefois, le critère de décomposabilité forte implique le critère 6 qui implique lui-même le critère 5. Autrement dit, si les critères 5 et 6 ne sont pas retenus comme critères de sélection dans notre travail, ceux-ci sont nécessairement respectés par un indice conforme au principe de décomposabilité forte que nous avons retenu.

2.3 Les huit principaux indices de mesure de la mixité sociale

Sur la base des résultats du paragraphe précédent, cette section introduit ce qui nous semble être les huit principaux indices disponibles dans la littérature actuelle. Dans le cadre de travaux portant sur la mixité sociale à l'Université, la population de référence est fréquemment subdivisée en trois ou quatre groupes sociaux. Nous faisons, aussi, le choix de centrer généralement notre description sur une version des indices compatible avec une analyse incluant quatre groupes, des versions adaptées à davantage de groupes ou seulement à trois groupes pouvant être facilement dérivées. Chacun des sous-paragraphes suivants présente brièvement un indice de mesure de la mixité sociale : l'indice d'entropie normalisé ou indice de Theil (sous-paragraphe 2.3.1), l'indice d'information mutuelle (sous-paragraphe 2.3.2), l'indice d'Atkinson (sous-paragraphe 2.3.3), l'indice d'exposition normalisé (sous-paragraphe 2.3.4), l'indice de dissimilarité (sous-paragraphe 2.3.5), l'indice de Gini Multigroupes (sous-paragraphe 2.3.6), l'indice de diversité relative (sous-paragraphe 2.3.7) et l'indice du coefficient de variation au carré (sous-paragraphe 2.3.8).

2.3.1 L'indice d'entropie normalisé ou indice de Theil

a) Le concept d'entropie : principe général et calcul

Les indices d'entropie ont initialement été introduits par Theil (1967). Ceux-ci sont utilisés de manière fréquente dans les travaux sur la mixité sociale. Le calcul d'un indice d'entropie présuppose l'identification d'une population de référence généralement notée R . L'entropie de la population de référence notée h_R permet de résumer en un seul chiffre le degré de représentation de chacun des groupes sociaux au sein d'une population donnée.

L'indice d'entropie peut être calculé pour un nombre de groupes sociaux variables. Nous faisons ici le choix de présenter une version de cet indice sur la base de quatre groupes notés a , b , c et d , les PCS étant fréquemment regroupées en quatre groupes dans le champ de l'éducation, comme nous l'avons montré dans la section 1 du présent chapitre. Dans le cadre d'un travail sur la mixité sociale, cela pourrait, par exemple correspondre aux groupes sociaux suivants : Très favorisé, Favorisé, Intermédiaire et Défavorisé. Le calcul de l'entropie de la population repose alors sur la proportion occupée par chaque groupe dans la population, que l'on peut noter q_a , q_b , q_c et q_d .

L'entropie h de la population de référence R se calcule ainsi de la manière suivante :

$$h_R = q_a \times \ln\left(\frac{1}{q_a}\right) + q_b \times \ln\left(\frac{1}{q_b}\right) + q_c \times \ln\left(\frac{1}{q_c}\right) + q_d \times \ln\left(\frac{1}{q_d}\right) \quad (1)$$

La condition pour réaliser ce calcul est qu'il y ait au moins un individu par groupe : $q_a > 0$, $q_b > 0$, $q_c > 0$ et $q_d > 0$.

Dans l'hypothèse où la presque totalité des individus appartiennent à un seul groupe (s'il y a par exemple un seul individu dans chacun des trois groupes restants), la valeur de l'entropie est minimale et tend vers 0. De manière symétrique, la valeur maximale est atteinte si les quatre groupes sociaux sont équi-pondérés, celle-ci étant d'environ 1,38.

Il convient de distinguer l'entropie de la population de référence notée h_r et l'indice d'entropie noté H . En effet, la première peut s'interpréter comme la simple description statistique de la répartition d'une population en plusieurs groupes sociaux, alors que le second présuppose à un moment donné la répartition des individus de cette

population en différentes unités (groupes de travail, Universités...) et, donc, des phénomènes ségrégatifs potentiels. Le concept de ségrégation est absent lors de la phase de calcul de l'entropie décrite ci-dessus puisque la population de référence est subdivisée en catégories, mais pas en unités à ce stade.

b) Le calcul de l'indice d'entropie normalisé

Il est possible d'évaluer le degré de mixité sociale (de ségrégation) d'une population dès lors que celle-ci est subdivisée. Pour se faire, il convient de mesurer l'indice d'entropie des unités k qui composent la population de référence, le tout en reprenant la méthodologie utilisée précédemment. Nous obtenons ainsi le calcul d'entropie suivant pour les unités k :

$$h_k = q_a \times \ln\left(\frac{1}{q_a^k}\right) + q_b \times \ln\left(\frac{1}{q_b^k}\right) + q_c \times \ln\left(\frac{1}{q_c^k}\right) + q_d \times \ln\left(\frac{1}{q_d^k}\right) \quad (2)$$

La mesure indicielle des niveaux de ségrégation repose de manière sous-jacente sur une comparaison entre l'indice d'entropie de la population totale et celui des différentes unités k qui la compose. L'indice d'entropie normalisé est ainsi égal à la moyenne pondérée des différences entre le niveau d'entropie de la population de référence et celui de chacune des sous-unités k qui la composent, cette soustraction étant normalisée au moyen d'une division par h_r qui garantit que la valeur finale de l'indice H soit comprise entre 0 et 1.

$$H = \sum_{k=1}^K \pi_k \frac{h_R - h_k}{h_R} \quad (3)$$

Le calcul de l'indice H est pondéré au moyen de π_k , qui indique le poids que représente chaque unité k dans la population.

Le calcul de l'indice H peut se réécrire de la manière suivante :

$$H = \frac{h_R - \sum_{k=1}^K \pi_k h_k}{h_R} \quad (4)$$

Suivant cette réécriture, l'indice H est simplement égal à la différence normalisée entre la valeur de l'entropie de la population de référence R et la moyenne pondérée de celle

de toutes les unités k . La valeur de l'indice H est ainsi de 0 si la composition de chacune des unités k (c'est-à-dire les parts occupées par les différents groupes au sein de chaque unité k) est identique à celle de la population de référence, ce qui serait le cas si la mixité sociale était parfaite. Une valeur de 0 ne renvoie pas, ainsi, à une situation d'équi-pondération des différents groupes sociaux, mais à l'absence de phénomènes ségrégatifs dans la subdivision de la population : la composition sociale des unités est identique à celle de la population. Dans ce cas extrême, la distribution des individus de la population R est semblable à une distribution aléatoire entre les différentes unités k .

Toutefois, ce cas extrême n'écarte aucunement la possibilité d'une ségrégation en amont ou en aval : s'il n'y a par exemple pas de phénomènes ségrégatifs dans les modalités suivant lesquelles les étudiants d'une Université se répartissent entre différentes disciplines, des phénomènes ségrégatifs peuvent tout à fait avoir lieu préalablement, par exemple dans la répartition des étudiants au sein d'une académie. Une valeur de 0 de l'indice d'entropie normalisée est ainsi compatible avec, par exemple, la prépondérance d'un ou de plusieurs groupes sociaux dans la population de référence.

De manière symétrique, si la presque totalité des individus appartiennent à un seul groupe, alors, la valeur de l'entropie des différentes unités h_k , est minimale et tend vers 0. Par suite, la valeur de l'indice d'entropie normalisé, qui est alors maximale, tend vers 1 et ce, quel que soit le degré initial de mixité sociale au sein de la population de référence. Dans ce second cas extrême, les individus se répartissent entre les différentes unités k uniquement en fonction de leurs appartenances sociales et non de façon aléatoire.

c) Avantages et inconvénients de l'indice d'entropie normalisé

Le principal avantage de l'indice d'entropie normalisé est qu'en raison de sa valeur comprise entre 0 et 1, des comparaisons entre différentes observations sont pratiques et immédiatement réalisables. Le fonctionnement de celui-ci, basé sur une comparaison normalisée entre la composition en groupes sociaux d'une population de départ et celle des différentes unités qui la composent, de plus, est aisément

compréhensible. Cet indice est ainsi choisi par Givord et al. (2015), en raison notamment de sa comparabilité.

Une même valeur de l'indice H peut, toutefois, renvoyer à des configurations distinctes voire opposées puisque l'indice est indifférent à la nature qualitative des groupes représentés, seule leur pondération quantitative étant prise en compte. Une unité k qui comprend près de 50% d'individus très favorisés et 50% d'individus défavorisés aura ainsi un niveau d'entropie similaire à une autre unité k composée dans les mêmes proportions d'individus issus du groupe intermédiaire et défavorisé.

Il est, par ailleurs, possible de démontrer que l'indice d'entropie normalisé n'est pas fortement décomposable, comme l'expliquent Frankel et Volij (2011). Dans le cadre de ce travail, cela empêcherait, de comparer la ségrégation entre différents niveaux (académie, Université au sein des académies, disciplines au sein des Universités, etc.) de manière robuste (c'est-à-dire avec des indices cohérents et décomposables pour les différents niveaux) et de savoir, par exemple, dans quelle proportion celle-ci résulte de la répartition par Universités ou par disciplines.

En résumé, l'indice d'entropie normalisé est utilisé car il fournit une appréhension synthétique de la mixité et permet de réaliser des comparaisons de manière directe. Toutefois, celui-ci ne peut pas être utilisé pour décomposer les résultats. Si l'objectif de l'analyse est de décomposer la ségrégation entre différents niveaux, il conviendra d'utiliser une version non normalisée de l'indice d'entropie.

2.3.2 L'indice d'entropie non normalisé ou indice d'information mutuelle

a) Définition et calcul de l'indice d'information mutuelle

L'indice d'entropie peut être utilisé dans sa version non normalisée afin, notamment, de satisfaire la propriété de forte décomposabilité. L'indice d'entropie non normalisé ou indice d'information mutuelle, se réécrit en modifiant la formule de l'équation 4 décrite ci-dessus de la manière suivante :

$$M = h_R - \sum_{k=1}^K \pi_k h_K \quad (5)$$

Il convient de noter qu'il s'agit de la même formule qu'en (4), la seule différence étant qu'ici il n'y a plus de normalisation au moyen d'une division par h_r , ce qui garantissait que la valeur finale de l'indice H soit comprise entre 0 et 1.

Si la valeur calculée de l'indice n'est plus comprise entre 0 et 1 et que de ce fait, celui-ci est moins facilement interprétable, celui-ci est « fortement décomposable ». Cette propriété peut être illustrée à travers l'exemple suivant. Si la population de référence d'étudiants R est subdivisée en différentes unités k au sein desquelles les étudiants sont regroupés par Universités et par disciplines, la valeur de l'indice d'information mutuelle est égale à la somme de l'indice calculé pour ses composantes suivant la formule (6) ou (7) :

$$M = M_{\text{INTERDISCIPLI}} + \pi_{\text{DISCIPLI}_1} \times M_{\text{INTRADISCIPLI}_1} + \pi_{\text{DISCIPLI}_2} \times M_{\text{INTRADISCIPLI}_2} \dots \quad (6)$$

$$M = M_{\text{INTERUNIV}} + \pi_{\text{UNIV}_1} \times M_{\text{INTRAUNIV}_1} + \pi_{\text{UNIV}_2} \times M_{\text{INTRAUNIV}_2} \dots \quad (7)$$

Suivant la formule (6), l'indice d'information mutuelle M est égal à la somme de l'indice « inter-disciplines » $M_{\text{INTERDISCIPLI}}$ et de la somme pondérée des indices « intra-disciplines » $M_{\text{INTRADISCIPLI}}$ calculé pour chaque discipline individuelle. Suivant la formule (7), l'indice M est égal à la somme des indices « inter-Universités » $M_{\text{INTERUNIV}}$, c'est-à-dire de l'indice de mixité sociale mesuré entre Universités et de la somme pondérée des indices « intra-Universités » $M_{\text{INTRAUNIV}}$, c'est-à-dire de l'indice de mixité sociale mesuré au sein de chaque Université entre disciplines. π_{UNIV_1} indique la pondération de l'Université 1 dans la totalité des Universités. Sur la base de cette formule, il est possible de quantifier les parts respectives de la ségrégation qui résultent de la répartition des étudiants par disciplines et par Universités.

b) Avantages et inconvénients de l'indice d'information mutuelle

Les principaux avantages de cet indice sont son caractère intuitif et sa décomposabilité forte. Toutefois, son absence de normalisation limite la possibilité de réaliser des comparaisons, ce qui a conduit par exemple Givord et al. (2015) à lui préférer l'indice d'entropie normalisé. De plus, comme démontré par Frankel et Volij (2011), l'indice

d'information mutuelle n'est pas invariant à la composition. Il convient de noter que cela peut poser problème, comme nous l'avons expliqué ci-dessus au paragraphe 1.3, dans la mesure où les parts des groupes changent au cours du temps et les indices ne sont pas strictement comparables d'une année sur l'autre.

Par ailleurs, comme dans le cas de l'indice d'entropie normalisé, une même valeur de l'indice d'information mutuelle peut renvoyer à des configurations distinctes puisque celui-ci est indifférent à la nature qualitative des groupes représentés, seule leur pondération quantitative étant prise en compte.

2.3.3 L'indice d'Atkinson

a) Définition et calcul de l'indice d'Atkinson

L'indice dit d'«Atkinson » a été introduit par Anthony B. Atkinson en 1970 dans le cadre de ses travaux portant sur les inégalités. Celui-ci existe dans une version dite symétrique ainsi que dans une version dite asymétrique qui se différencie par le choix de pondération des groupes sociaux.

En reprenant la notation utilisée antérieurement, l'indice d'Atkinson dans sa version symétrique peut s'écrire de la manière suivante :

$$A = 1 - \sum_{k=1}^K (q_a^k q_b^k q_c^k q_d^k)^{1/4} \quad (8)$$

q_a^k indique ici la part du groupe a dans l'unité k. Il convient de noter que la version symétrique de l'indice d'Atkinson ne prend pas en compte le poids occupé par chacun des groupes sociaux au sein de la population de référence, en se focalisant uniquement sur celui qu'ils occupent au sein de chaque unité. Cette absence de prise en compte n'est pas problématique en cas d'équi-pondération des groupes dans la population de référence. Toutefois, si comme dans le cas de notre étude, les groupes sociaux ne sont pas équi-pondérés au sein de la population de référence, il existe alors un risque de biais. Comme dans l'exemple mentionné au sous-paragraphe 2.2.2 ci-dessus, si un groupe social est surreprésenté au sein de la population de référence, alors la simple surreprésentation de ce groupe au sein des unités k ne saurait être

considérée comme un facteur de ségrégation sociale. Cependant, si l'indice d'Atkinson est utilisé dans sa version symétrique, cette surreprésentation diminuera la valeur de l'indice, et donc le niveau mesuré de mixité sociale, ce qui créera un biais.

Dans sa version asymétrique, l'indice d'Atkinson prend en compte le poids occupé par chacun des groupes sociaux au sein de la population de référence. La pondération est différenciée afin de tenir compte des poids relatifs des différents groupes sociaux au sein de la population de référence. Dans ce cas, l'indice se base sur la répartition des groupes sociaux au sein de la population totale pour une année de référence des données. L'indice d'Atkinson dans sa version asymétrique peut alors s'écrire de la manière suivante :

$$A = 1 - \sum_{k=1}^K q_{ak}^{q_{at}} q_{bk}^{q_{bt}} q_{ck}^{q_{ct}} q_{dk}^{q_{dt}} \quad (9)$$

$q_{ak}^{q_{at}}$ indique ici la part du groupe a dans l'unité k élevée à la puissance (pondérée) q_{at} qui désigne la part du groupe a dans la population totale l'année de référence t.

b) Avantages et inconvénients de l'indice d'Atkinson

Si la version symétrique de l'indice d'Atkinson implique un risque de biais dans le cadre de notre travail, la version asymétrique de celui-ci présente deux avantages principaux : il est invariant à la composition et respecte le principe du degré de mixité de la population de référence supérieur ou égal à celui des unités (*school division property*). De ce point de vue, il s'agit du seul indice permettant d'effectuer des comparaisons dans le temps rigoureuses mathématiquement.

Celui-ci, toutefois, a comme principal désavantage de ne pas être fortement décomposable. De plus, son calcul est sensible à la présence de « zéro », c'est-à-dire à des situations où la part d'un groupe dans une unité k donnée est nulle. Frankel et Volij (2011) recommandent ainsi l'usage de cet indice avec des données suffisamment agrégées où les « zéros » sont rares.

2.3.4 L'indice d'exposition normalisé

a) Définition et calcul de l'indice d'exposition normalisé

L'indice d'exposition normalisé, que l'on peut noter P , repose sur l'analyse de la part occupée par une catégorie sociale de référence, par exemple le groupe « très favorisé » au sein des différentes unités k . L'indice d'exposition normalisé, qui mesure la façon dont un groupe de référence se voit « exposé » au reste de la population dans les différentes unités k , présuppose d'un point de vue méthodologique le choix d'un groupe au regard des objectifs de l'étude, ce groupe étant distingué du reste de la population.

En reprenant la notation utilisée antérieurement, l'indice d'exposition normalisé E se calcule au moyen de la formule suivante :

$$E = \frac{1}{q_a(1 - q_a)} \sum_{k=1}^K \pi_k (q_a^k - q_a)^2 \quad (10)$$

La part q_a^k occupée par le groupe de référence a au sein de chaque unité k est comparée à la part q_a de ce groupe au sein de la population totale. La valeur de l'indice est de 0 lorsque la part occupée par le groupe de référence au sein de chaque unité k est identique à celle qu'il occupe dans la population totale. Dans ce cas, s'il est impossible de soutenir de manière générale que la mixité sociale est parfaite ou même élevée, puisqu'aucune information relative aux autres groupes et à leur répartition n'est disponible, on constate que le groupe a est « exposé » au reste de la population de manière identique dans toutes les unités k . De manière inverse, la valeur de l'indice est de 1,66 lorsque le groupe a n'est jamais « exposé » au reste de la population puisque les individus qui le composent évoluent en cercle fermé et évitent les unités où des individus appartenant au reste de la population sont présents. Chacune des unités k comprend soit exclusivement des individus du groupe a soit uniquement des individus appartenant au reste de la population, le degré de mixité sociale du groupe a étant nul.

Il est possible de calculer la valeur de cet indice pour différents groupes sociaux afin de vérifier si des valeurs significatives apparaissent pour certains groupes.

Il convient de noter qu'il existe également une version multigroupes de cet indice. Dans le cas où il y a quatre groupes sociaux, celui-ci se calcule simplement comme la moyenne pondérée des indices calculés pour chaque groupe social a , b , c et d via la formule n°10.

b) Avantages et inconvénients de l'indice d'exposition normalisé

La valeur de l'indice d'exposition normalisé unigroupe peut aisément être interprétée et comparée. Il respecte, de plus, le principe du degré de mixité de la population de référence supérieur ou égal à celui des unités (*school division property*), suivant la démonstration effectuée par Frankel et Volij (2011).

Quant à l'indice d'exposition normalisé multigroupe, s'il respecte le principe du degré de mixité de la population de référence supérieur ou égal à celui des unités (*school division property*), celui-ci présente également le défaut d'être ni fortement décomposable ni invariant à la composition. Toutefois, dans le cas particulier traité par Reardon et O'Sullivan (2004)³³, il varie très peu selon la composition. Si son interprétation est plus complexe et moins intuitive que dans le cas de l'indice unigroupe, il fournit une mesure synthétique et plus globale de la mixité sociale que les indices unigroupes, en ligne avec les autres indices présentés.

2.3.5 L'Indice de dissimilarité

L'indice de dissimilarité a été proposé en 1955 par Duncan et Duncan. Dans sa version binaire la plus simple pour un groupe a et un groupe b , celui-ci peut s'écrire de la manière suivante :

$$D = \frac{1}{2} \sum_{k=1}^K \pi_k \left| \frac{q_a^k}{q_a} - \frac{q_b^k}{q_b} \right| \quad (11)$$

π_k indique le poids que représente chaque unité k dans la population, q_a^k la part occupée par le groupe a au sein de chaque unité k et q_a la part de ce groupe au sein de la population totale.

³³ Les travaux de Reardon et O'Sullivan (2004) concernent la ségrégation résidentielle aux Etats-Unis.

La version multigroupe de l'indice de dissimilarité, calculée pour quatre groupes a , b , c et d peut s'écrire au moyen de la formule suivante :

$$D = \frac{1}{2I_{SIMS}} \sum_{k=1}^K \pi_k [|q_a^k - q_a| + |q_b^k - q_b| + |q_c^k - q_c| + |q_d^k - q_d|] \quad (12)$$

L'indice de Simpson I_{SIMS} compris dans la formule (11) peut s'écrire de la manière suivante :

$$I_{SIMS} = q_a(1 - q_a) + q_b(1 - q_b) + q_c(1 - q_c) + q_d(1 - q_d) \quad (13)$$

L'indice de Simpson compris dans la formule de calcul de l'indice de dissimilarité correspond à la probabilité que deux élèves tirés au hasard proviennent de milieux sociaux distincts.

L'indice de dissimilarité peut s'interpréter comme l'écart absolu moyen entre l'exposition d'un élève au groupe de référence et la part de ce groupe dans la population, divisée par la valeur maximale que peut prendre cet écart moyen, atteinte en situation de ségrégation complète. Cette normalisation permet de faire en sorte que D évolue entre zéro et un. Une interprétation alternative est de voir D comme la part minimale de la population qu'il faudrait déplacer d'une unité à l'autre pour atteindre l'équilibre parfait (une répartition uniforme du groupe d'intérêt à travers l'ensemble des unités), là encore divisée par la part maximale qu'il faudrait déplacer, en cas de ségrégation complète.

Si l'indice de dissimilarité respecte la propriété de *school division property*, celui-ci n'est invariant à la composition que s'il est utilisé sur la base de deux groupes et en aucun cas il n'est fortement décomposable.

2.3.6 L'indice de Gini multigroupes

a) L'indice de Gini calculé pour deux groupes

L'indice de Gini a été défini par Jahn et al. (1947) pour deux groupes a et b . Dans un schéma représentant la distribution de revenus d'une population, l'indice de Gini mesure la superficie comprise entre la courbe de Lorenz³⁴ et une ligne à 45°.

³⁴ La courbe de Lorenz est introduite dans les cours de statistiques de niveau intermédiaire. Se reporter par exemple à Statistique pour économistes et gestionnaires, Brigitte Tribout, 2^{ème} édition, Pearson, 2013.

Il se calcul de la manière suivante :

$$G = \frac{1}{q_a(1-q_a)} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |q_{ai}^k - q_{aj}^k| \quad (14)$$

N désigne la taille de la population ; les individus sont notés i et $i \in \{1, \dots, N\}$

q_a indique la part occupée par le groupe a au sein de la population totale.

q_{ai}^k désigne la part occupée par le groupe a au sein de l'unité k dans laquelle se trouve l'individu i

La valeur de l'indice de Gini peut s'interpréter comme l'écart moyen d'exposition au groupe a pour deux individus i et j choisis au hasard. L'indice repose sur le calcul de la moyenne de tous les écarts en valeur absolue pour tous les couples de la variable statistique étudiée (cette différence moyenne mesure l'écart espéré de l'exposition au groupe a de deux individus pris au hasard avec remise dans la population étudiée). L'égalité parfaite (répartition du groupe a dans les unités k égale à la part occupée par ce groupe dans la population totale) correspond à un indice égal à zéro alors que l'inégalité totale correspond à une valeur égale à 1.

b) L'indice de Gini multigroupes

L'indice de Gini multigroupes est une généralisation proposée par Reardon et Sean (1998) de la version de l'indice de Gini pour deux groupes. Il se calcule de la manière suivante pour quatre groupes a, b, c et d :

$$D = \sum_{i=1}^N \sum_{j=1}^N \frac{N_j N_i}{2N^2} [|q_{ai}^k - q_{aj}^k| + |q_{bi}^k - q_{bj}^k| + |q_{ci}^k - q_{cj}^k| + |q_{di}^k - q_{dj}^k|] \quad (15)$$

N désigne la taille de la population ; les individus sont notés i et $i \in \{1, \dots, N\}$

q_a indique la part occupée par ce groupe au sein de la population totale.

q_{ai}^k désigne la part occupée par le groupe a au sein de l'unité k dans laquelle se trouve l'individu i

N_j désigne la taille de l'unité k dans laquelle se trouve l'individu j .

Ce calcul peut être interprété comme la somme, pour les groupes répertoriés, de la différence absolue moyenne pondérée entre toutes les paires possibles d'unités de la

part occupée par le groupe d'intérêt, divisée par la valeur maximum possible de cette somme, obtenue si le système était dans un état de ségrégation.

Si l'indice de Gini respecte la propriété de *school division property*, celui-ci n'est invariant à la composition que s'il est utilisé sur la base de deux groupes et n'est en aucun cas fortement décomposable.

2.3.7 L'indice de diversité relative

L'indice de diversité relative a été proposé par Goodman et al. (1954) puis Carlson (1992). Celui-ci peut se calculer de la manière suivante pour quatre groupes a , b , c et d :

$$DR = \sum_{j=1}^N \frac{N_j}{NI_{SIMS}} [|q_{ai}^k - q_a| + |q_{bi}^k - q_b| + |q_{ci}^k - q_c| + |q_{di}^k - q_d|]^2 \quad (16)$$

N désigne la taille de la population ; les individus sont notés i et $i \in \{1, \dots, N\}$

q_a indique la part occupée par le groupe a au sein de la population totale.

q_{ai}^k désigne la part occupée par le groupe a au sein de l'unité k dans laquelle se trouve l'individu i

N_j désigne la taille de l'unité k dans laquelle se trouve l'individu j .

L'indice de Simpson I_{SIMS} a été défini en (12).

L'indice de diversité relative peut être interprété comme le pourcentage d'individus qu'il faudrait transférer entre les unités pour que les proportions occupées par les groupes soient identiques entre les unités et la population totale, divisé par le pourcentage d'individus qu'il serait nécessaire de transférer en cas de ségrégation complète. L'égalité parfaite (répartition du groupe a dans les unités k égale à la part occupée par ce groupe dans la population totale) correspond à un indice égal à zéro alors que l'inégalité totale correspond à une valeur égale à 1, ce cas n'étant possible que si chaque unité a un degré de diversité nul (un seul groupe social est présent).

Si l'indice de diversité relative respecte la propriété de *school division property*, celui-ci n'est ni invariant à la composition ni fortement décomposable.

2.3.8 L'indice du coefficient de variation au carré

Le coefficient de variation au carré a été proposé par Reardon et O'Sullivan (2004). Celui-ci peut se calculer de la manière suivante pour quatre groupes a , b , c et d :

$$C = \sum_{j=1}^N \frac{N_j}{N} \frac{[|q_{ai}^k - q_a| + |q_{bi}^k - q_b| + |q_{ci}^k - q_c| + |q_{di}^k - q_d|]^2}{(4-1)q_a X(4-1)q_b X(4-1)q_c X(4-1)q_d} \quad (17)$$

N désigne la taille de la population ; les individus sont notés i et $i \in \{1, \dots, N\}$

q_a indique la part occupée par le groupe a au sein de la population totale.

q_{ai}^k désigne la part occupée par le groupe a au sein de l'unité k dans laquelle se trouve l'individu i

N_j désigne la taille de l'unité k dans laquelle se trouve l'individu j .

Si l'indice du coefficient de variation au carré respecte la propriété de *school division property*, celui-ci n'est ni invariant à la composition ni fortement décomposable.

2.4 De la nécessité de recourir à une stratégie multi-indicielle

Ce paragraphe synthétise les résultats exposés ci-dessus afin d'identifier les indices qui paraissent les plus importants pour analyser la mixité sociale à l'Université. Ce paragraphe rappelle ainsi la situation des huit indices présentés au regard des quatre propriétés souhaitables et propose l'utilisation d'une stratégie multi-indicielle.

Afin de choisir parmi les huit indices présentés antérieurement celui ou ceux qui seraient le plus adaptés à un travail portant sur la mixité sociale à l'Université, nous avons utilisé quatre critères tirés notamment de l'article de référence de Frankel et Volij (2011) : un degré de mixité plus élevé au sein de la population de référence que dans la moyenne des sous-populations, l'invariance à la composition, la décomposabilité forte et la facilité d'interprétation. Les caractéristiques des huit indices répertoriés au regard des quatre critères de sélection retenus sont résumés dans le Tableau 5 ci-dessous.

Tableau 5 : Les critères de sélection des indices de mixité sociale

Critère	<i>H</i>	<i>M</i>	<i>A</i>	<i>E</i>	<i>D</i>	<i>G</i>	<i>DR</i>	<i>C</i>
<i>School Division Property</i>	Oui	Oui						
Invariance à la composition	Non	Non	Oui	Non	2	2	Non	Non
Décomposabilité forte	Non	Oui	Non	Non	Non	Non	Non	Non
Interprétation intuitive	Non	Non	Non	Oui	Non	Non	Non	Non

Source : Frankel et Volij (2011) sauf pour le dernier critère relatif au caractère intuitif de l'interprétation (nous proposons d'ajouter cet élément comme critère supplémentaire de sélection)

Les huit indices de mixité sociale retenus sont désignés par les sigles suivants (première ligne du tableau, à partir de la gauche):

H : indice du Theil ou indice d'Entropie Normalisé

M : Indice d'Information mutuelle ou indice d'entropie non normalisé

A : indice d'Atkinson (version symétrique et non symétrique)

E : Indice d'Exposition Normalisé

D : Indice de dissimilarité

G : Indice de Gini multigroupes

DR : Indice de diversité relative

C : Indice du coefficient de variation au carré

Le chiffre « 2 » permet de préciser que le critère n'est respecté par l'indice synthétique que s'il y a deux groupes sociaux définis.

Le Tableau 5 ci-dessus nous permet d'effectuer les constats suivants :

- les huit indices retenus respectent le critère de *school division property*.
- aucun des huit indices retenus ne respecte simultanément les critères d'invariance à la composition et de décomposabilité forte. *A fortiori*, aucun de ces huit indices ne respecte les quatre critères de sélection.
- dans le contexte d'utilisation multigroupe (avec en général au moins trop groupes sociaux), trois indices semblent particulièrement intéressants : l'indice d'Information mutuelle ou indice d'entropie non normalisé, qui est fortement décomposable, mais non invariant à la composition, l'indice d'Atkinson, qui est invariant à la composition mais pas fortement décomposable et l'Indice d'Exposition Normalisé, qui est facilement interprétable et varie très peu selon la composition suivant les estimations réalisées par Reardon et O'Sullivan (2004).

Comme aucun des huit indices synthétiques répertoriés ne respectant les quatre critères de sélection retenus, une solution envisageable nous semble être d'utiliser un ou deux indices à titre principal suivant le contexte d'utilisation et de calculer les six ou sept indices restants afin d'obtenir des résultats complémentaires. Cela peut permettre de tester la robustesse des résultats et de vérifier que les différents indices calculés ne conduisent pas à des conclusions divergentes, en ligne avec la stratégie de recherche basée sur une approche multi-indicielle employée dans d'autres travaux portant sur la question. Ainsi Ly et Riegert (2015), utilisent, par exemple, pour analyser la ségrégation scolaire l'indice d'exposition normalisé, mais en calculent également trois autres à des fins de comparaison : l'indice de dissimilarité, l'indice de Gini, l'indice d'entropie ou Indice du Theil. Givord et al. (2016) afin d'étudier la ségrégation sociale entre les Collèges français privilégient l'indice d'entropie mais calculent également l'indice de dissimilarité, l'indice d'Atkinson et l'indice d'information mutuelle afin de vérifier la robustesse des résultats.

Une analyse de la mixité sociale à l'Université, pourrait reposer ainsi sur l'exploitation des données SISE présentées de manière synthétique en introduction générale et la distinction de deux contextes d'utilisation :

- Dans le premier cas, le degré de mixité sociale est mesuré de manière dynamique sur plusieurs années. L'utilisation de données ayant une profondeur temporelle suffisante est justifiée pour identifier des tendances de moyen terme avant la mise en place de la loi O.R.E. évoquée dans l'introduction générale. Il convient, toutefois de noter qu'en raison de l'utilisation de données pluriannuelle, la composition de la population n'est pas rigoureusement identique d'une année sur l'autre. L'indice d'Atkinson étant le seul parmi les huit indices retenus à être *stricto sensu* invariant à la composition, il nous semble pleinement légitime de recourir à celui-ci, d'autant plus qu'à notre connaissance, il n'existe pas de travaux portant sur la mixité sociale au sein de l'Université française reposant sur l'utilisation de cet indice. Une solution alternative, mais moins rigoureuse d'un point de vue mathématique peut être également d'utiliser à titre principal l'Indice d'Exposition Normalisé (IEN) dans la mesure où suivant les estimations réalisées par Reardon et O'Sullivan (2004) il varie très peu à la composition et se caractérise par son interprétation intuitive.

Les sept indices restants sont calculés afin d'obtenir des résultats complémentaires. Sur ce point, Avouac et Harari-Kermadec (2021) montrent qu'il existe une baisse relativement lisse de la ségrégation à compter de 2010 en utilisant un indice d'entropie normalisée. Notre revue montre, toutefois, que lorsqu'on traite des évolutions il est préférable d'utiliser un indice d'Atkinson. Pour notre analyse il apparaît aussi important de vérifier que les conclusions des auteurs tiennent quand on utilise un indice d'Atkinson afin de fournir un diagnostic plus complet de l'évolution de la mixité sociale à l'Université.

- Dans le second cas, l'objectif est de décomposer les valeurs mesurées des indices en différents éléments et de quantifier leur importance respective. Comme expliqué ci-dessus dans le sous-paragraphe 2.1.2, la ségrégation sociale dépend selon notre analyse de cinq facteurs principaux. Aussi, il nous semble essentiel de décomposer les niveaux de ségrégation mesurés en différents éléments pour comprendre à quel niveau celle-ci se forme et potentiellement distinguer des points permettant de renforcer les niveaux de mixité sociale à l'Université. Or, comme rappelé dans le Tableau 5 ci-dessus, seul l'Indice d'information mutuelle (ou indice d'entropie non normalisé) respecte le critère de décomposabilité forte. Celui-ci pourra donc dans ce cas être utilisé à titre principal alors que les sept indices restants peuvent être calculés afin d'obtenir des résultats complémentaires.

3 L'unité d'observation de la mixité sociale

Mesurer la mixité sociale à l'Université nécessite de choisir une unité d'observation. Dans le cadre de notre travail portant sur la mixité sociale à l'Université, il existe, en effet, une série de possibilités pour choisir les unités d'observations k , telles que nous les avons définies dans la section 2.3 et destinées à mesurer le degré de mixité sociale : les zones académiques, les Universités, les facultés (au sens d'une unité pédagogique donnée au sein d'une Université donnée), les groupes de travaux dirigés, etc. Or, le choix de celle-ci nous paraît crucial pour bien appréhender ce phénomène dans toutes ses dimensions. En effet, le calcul des indices synthétiques de mesure dépend de celui-ci. De notre point de vue, deux critères doivent être retenus pour décider de l'unité d'observation : l'objectif exact de la description et le problème statistique des petites unités. La section 3 est structurée de la manière suivante : le paragraphe 3.1 introduit la question de la granularité des unités d'observation, le paragraphe 3.2 évoque le problème statistique lié à l'existence des petites unités et le paragraphe 3.3 décrit une méthode de correction envisageable du biais liés aux petites unités .

3.1 La granularité des unités d'observation

Le choix de l'unité d'observation doit reposer, d'abord, sur les objectifs de la description et les facteurs de ségrégation décrit ci-dessus au sous-paragraphe 2.1.2 que l'on cherche à prendre en compte. En effet, ce choix ne sera pas identique suivant que l'objectif est de décrire les différents niveaux de ségrégation de la manière la plus précise possible ou seulement, par exemple la ségrégation résidentielle.

Si, dans un premier cas, l'objectif retenu est de décrire la ségrégation de la manière la plus précise et complète possible, l'unité d'observation retenue doit permettre de rendre compte des cinq facteurs mentionnés au sous-paragraphe 2.1.2. Afin de prendre en compte ceux-ci, il est nécessaire de s'assurer que l'unité d'observation choisie soit suffisamment désagrégée pour refléter si possible l'environnement de l'étudiant et, plus précisément son groupe de pairs.

Dans le cas de notre travail portant sur la mixité sociale au sein de l'Université française, avec la mise à disposition pour la recherche de données administratives de

type SISE, l'unité désagrégée la plus fine disponible dans les données est ce que l'on peut appeler l'*unité de formation* par année universitaire et que l'on peut définir comme : « Etablissement universitaire x Discipline d'inscription x Année du cursus ». Les étudiants d'une faculté au sein d'un établissement universitaire sont alors regroupés suivant le niveau atteint dans le cursus LMD ou même suivant l'année d'études. Un exemple d'une telle unité d'observation pourrait être les étudiants de L3 de la faculté de droit de l'Université d'Aix-Marseille. Un tel choix permet de mesurer de manière précise les niveaux de ségrégation et de se rapprocher au maximum de la répartition des étudiants, telle qu'elle est susceptible d'exister dans la pratique. Il convient de noter, en effet, que, contrairement au cas de l'enseignement secondaire, les groupes de pairs des étudiants peuvent être difficiles à identifier en raison des spécificités des études universitaires (groupes de travaux dirigés, de travail en bibliothèque, etc.). Cela permet de rendre compte du fait qu'une formation peut par exemple être fortement ségrégée au niveau du Master et non de la Licence à cause de la sélection opérée au fur et à mesure de l'avancement dans les études. Un tel niveau de désagrégation permet par exemple de refléter également les différences entre les disciplines au sein d'une même Université, ce phénomène ayant été analysé par Frouillou (2016). Ce niveau de désagrégation permet également de prendre en compte les aspects géographiques de la ségrégation, avec notamment les différences susceptibles d'exister entre des Universités de différentes régions, mais aussi au niveau local, par exemple au sein d'une même ville. Par exemple, Courtioux et Maury (2020) ont ainsi souligné le caractère essentiel de la dimension géographique de la ségrégation dans l'enseignement secondaire, avec des différences entre enseignement privé et public particulièrement élevée au niveau local dans les centres urbains.

Toutefois, des unités d'observation plus agrégées peuvent également être intéressantes. Dans ce cas, l'unité d'observation retenue doit permettre de rendre compte du ou des facteurs de ségrégation sélectionnés. Ainsi le travail de Courtioux et Maury (2020) est centré sur les aspects spatiaux, géographiques de la ségrégation. Dans le travail d'Avouac et Harari-Kermadec (2021), l'unité d'observation choisie pour mesurer la ségrégation est l'Université. Ce choix permet de notre point de vue de refléter les facteurs de ségrégation n°1, 4 et 5 mentionnés au sous-paragraphe 2.1.2, à savoir la ségrégation résidentielle, celle liée aux décisions institutionnelles des

établissements et aux politiques nationales en matière d'enseignement supérieur. En revanche, ce choix ne permet pas, de notre point de vue, d'intégrer de manière directe à la mesure de la ségrégation les microdécisions des étudiants et de leur famille (facteur de ségrégation n°3) ou les effets de pairs (facteur de ségrégation n°4). Ainsi, les indices calculés ne pourront refléter ni les différences inter-disciplinaires ni celles liées aux niveaux d'études qui sont par ailleurs documentées par Frouillou (2016).

3.2 Le problème statistique des petites unités

Le choix de l'unité d'observation, quand cette dernière est petite, est susceptible d'engendrer un risque de biais dans la mesure de la ségrégation, comme le rappellent par exemple Ly et Riegert (2015) ou Bechichi et al. (2021) : il est statistiquement impossible d'obtenir une valeur nulle des indices de mixité sociale choisis si les unités d'observation ne sont pas d'une taille suffisamment grande. Une valeur nulle d'un indice de ségrégation impliquerait, en effet, que la composition de chacune des unités k (c'est-à-dire la part occupée par chacun des groupes sociaux) soit rigoureusement identique à celle de la population de référence. Toutefois, cela est impossible, même en cas d'allocation aléatoire, si les unités d'observation sont petites.

Le problème suivant lequel les niveaux de ségrégation mesurés par les indices sont supérieurs à zéro même en cas d'allocation aléatoire de la population est d'autant plus important que les unités d'observation k sont petites et qu'un des groupes sociaux identifiés représente une faible part de la population totale. Si par exemple un groupe social représente 5% de la population de référence, soit un individu sur 20, alors, un indice de mixité sociale nul implique que chaque unité comprenne un individu de ce groupe, ce qui n'est possible que si chacune des unités qui compose la population est elle-même un multiple de 20 (si un groupe représente 1% de la population, il est nécessaire que chacune des unités soit un multiple de 100). Ainsi dans une unité k de taille 19, il y a en moyenne 0,95 individus. Du fait de la discontinuité de la distribution, toutes les classes sont ainsi éloignées de la moyenne de la population de référence : dans ce cas, la ségrégation sociale observée résulte du biais statistique lié aux unités de petite taille.

3.3 La correction des biais liés aux petites unités

Afin de répondre au problème lié à l'existence des petites unités d'observation, une solution envisageable est de ne pas choisir des unités d'observation trop petites et, de recourir, en plus, à des méthodes statistiques correctives. Le présent paragraphe est structuré de la manière suivante : le sous-paragraphe 3.3.1 explique comment le choix de l'unité d'observation doit reposer sur des critères potentiellement contradictoires et le sous-paragraphe 3.3.2 décrit une solution statistique envisageable basée sur les méthodes de simulation pour corriger les risques de biais résiduels.

3.3.1 Le choix de l'unité d'observation dans la pratique

Les deux paragraphes antérieurs 3.1 et 3.2 ont mis au jour deux critères pour sélectionner une unité d'observation adéquate : les objectifs exacts de la description et le problème statistique des petites unités. Si l'objectif de l'analyse n'est pas de décrire un phénomène de ségrégation de la manière la plus fine possible, alors il n'est pas nécessaire de recourir au niveau de désagrégation maximal, comme c'est le cas pour Harari et Kermadec (2021), dont le travail est centré sur les différences inter-universitaires, et pour lequel le problème des petites unités est limité. Si en revanche, l'objectif de l'analyse est de décrire un phénomène de ségrégation de la manière la plus fine possible, dans ses différentes composantes, le choix de l'unité d'observation repose sur deux critères potentiellement contradictoires, à savoir l'utilisation d'une unité d'observation suffisamment désagrégée, voire la plus désagrégée possible et, simultanément, la nécessité de ne pas choisir une unité d'observation trop petite afin d'éviter un biais statistique trop important lié à la présence de petites unités. Le choix de l'unité d'observation résulte alors d'un arbitrage entre ces deux objectifs contradictoires et doit correspondre à un point d'équilibre entre ces deux objectifs. Par ailleurs, l'information la plus désagrégée théoriquement envisageable, par exemple au niveau des groupes de travaux dirigés, n'est pas disponible dans la base de données SISE, ce qui constitue une limitation propre aux données.

Dans le cadre d'un travail portant sur l'Université française, il y a des raisons de penser que le choix d'un niveau de décomposition maximal décrit dans le paragraphe 3.1, s'il est justifié au regard de l'objectif de description fine des phénomènes de ségrégation, pourrait engendrer d'un point de vue statistique un risque important d'obtenir des

unités de taille insuffisante, en particulier pour les plus petits établissements de Province. Dès lors, une solution pertinente de recours (*second best*) pourrait être de décrire les phénomènes de ségrégation sociale à l'Université au niveau de désagrégation supérieur, soit l'*unité de formation* (ou faculté) considérée sur l'ensemble du cycle universitaire que nous définissons comme le croisement de l'Université et de la discipline. Par exemple, l'ensemble des étudiants de droit de l'Université Lyon II sont regroupés dans une même *unité de formation*, sans que soit pris en compte le niveau atteint dans le « cursus LMD ». Ces unités de formation, qui sont définies ici de manière formelle et théorique, font l'objet d'une description détaillée dans la section 2 du chapitre II.

3.3.2 Corriger les biais

Si le choix de l'unité d'observation utilisée tient compte de la double exigence exposée dans le paragraphe antérieur, le problème statistique des petites unités est réduit mais non éliminé. Dès lors, afin de tenir compte de ce risque de biais résiduel, une solution possible est de recourir à une méthode corrective issue de la statistique inférentielle, comme le proposent Ly et Riegert (2015) ainsi que Bechichi et al. (2021). Les méthodes de simulation permettent, en effet, d'estimer, d'abord, le « niveau de ségrégation » en cas de répartition aléatoire des individus sous contrainte de prise en compte de la taille des unités existantes et, ensuite, de déduire celui-ci des résultats obtenus avec les données réelles :

- Dans un premier temps, une méthode de simulation est appliquée. Les individus de la population de référence sont répartis entre chacune des unités k (dont la taille demeure identique) grâce à une méthode de tirage aléatoire sans remise, aucun phénomène de ségrégation n'étant alors à l'œuvre. Un certain niveau de ségrégation est, toutefois, mesuré par les indices de mixité sociale en raison du problème statistique des petites unités évoqué ci-dessus dans le paragraphe 3.2. Afin d'obtenir une valeur moyenne des indices robuste, il est généralement nécessaire de répliquer cette démarche un certain nombre de fois, par exemple 1 000 fois, ce qui permet d'obtenir une valeur moyenne, un écart-type ainsi qu'un intervalle de confiance.

- Dans un second temps, le niveau de ségrégation mesuré via l'application des méthodes aléatoires est comparé au niveau de ségrégation mesuré sur les unités réelles. Si l'intervalle de confiance autour de la valeur moyenne de l'indice calculé sur la base des méthodes aléatoires comprend la valeur de l'indice calculé à partir des données réelles, alors il est possible d'affirmer qu'avec un certain degré de probabilité, il n'y a pas de ségrégation sociale. A l'inverse, si la valeur de l'indice calculé à partir des données réelles est supérieure à la borne haute de l'intervalle de confiance, alors il est légitime d'affirmer qu'avec un certain degré de probabilité, il existe un certain degré de ségrégation sociale.

4 Un premier cadrage statistique sur l'évolution de la mixité sociale à l'Université

En ligne avec les éléments mentionnés dans les sections précédentes, il nous semble nécessaire de réaliser un premier cadrage statistique relatif à la mixité sociale au sein de l'Université française. En effet, il existe très peu de travaux sur le sujet et ceux-ci ne permettent pas selon nous d'obtenir une vision d'ensemble de la question de la mixité sociale à l'Université et de son évolution. Or, de ce point de vue, il nous semble essentiel d'évaluer la situation sur ce plan avant la mise en œuvre de « Parcoursup ». Nous noterons que les travaux de Frouillou (2016) ne concernent que l'Ile-de-France et ne reposent pas sur le calcul d'indices synthétiques de mesure de la mixité sociale et ceux de Bechichi et al. (2021) ont une profondeur temporelle trop limitée au regard de cette question. Par ailleurs, il nous semble que la robustesse du constat effectué par Avouac et Harari-Kermadec (2021) sur la base de l'indice d'entropie normalisé devrait être vérifiée avec un indice d'Atkinson sur une période pluriannuelle suffisamment longue. Nous avons vu dans la section 2 que l'indice d'entropie normalisé n'était pas invariant à la composition. Or la part des élèves de différentes PCS varie d'une année sur l'autre. Dès lors, seuls les indices d'Atkinson permettent d'effectuer des comparaisons pluriannuelles. Dans la mesure où nous disposons des données issues des bases SISE Université sur la période 2006-2016, nous avons mesuré les niveaux de mixité sociale au moyen de cet indice sur l'ensemble de cette période. Afin de décrire l'estimation réalisée, la présente section est structurée de la manière suivante : le paragraphe 4.1 décrit la méthodologie employée, le paragraphe 4.2 présente les principaux résultats obtenus et le paragraphe 4.3 fournit des éléments d'interprétation de ces résultats.

4.1 Méthode

Afin d'estimer les niveaux de ségrégation sur l'ensemble de la période 2006-2016, nous avons utilisé un indice d'Atkinson, qui est le seul à être invariant à la composition lorsqu'il y a plus de deux groupes, comme expliqué plus en détail dans le paragraphe 2.3 ci-dessus. Le choix d'utiliser un indice d'Atkinson se justifie d'autant plus qu'il n'existe pas à notre connaissance à ce jour de travaux portant sur la mixité sociale au

sein de l'Université française reposant sur son utilisation : le travail de Frouillou (2016) repose ainsi uniquement sur des statistiques descriptives, celui d'Avouac et Harari-Kermadec (2021) utilise un indice d'information mutuelle et celui de Bechichi et al. (2021) un indice d'exposition normalisé.

Nous avons vu dans la section 2 de ce chapitre qu'il était possible de choisir entre plusieurs versions de ces indices qui respectent toutes cette propriété d'invariance à la composition. Pour que cette propriété soit vérifiée, il convient de choisir un indice dont la structure correspond au poids de chaque groupe dans la population. Dans le cas où il y a quatre groupes, l'indice symétrique prend comme référence de mixité une structure (composition) de la population où chaque groupe occupe une part identique, soit 25%. Toutefois, dans la mesure où les quatre groupes sociaux identifiés ne sont pas équipondérés dans la population de référence analysée, il est plus cohérent de ne pas utiliser l'indice d'Atkinson dans sa version symétrique. La structure retenue correspond alors au poids relatif de la population des différentes catégories une année de référence donnée : l'indice d'Atkinson asymétrique utilisé est ainsi « en structure 2006 », c'est-à-dire que celui-ci est calculé pour chaque année en référence à la part occupée par les groupes sociaux identifiés en 2006 (il est bien entendu possible de construire un indice basé sur une autre année de référence, invariante dans le temps). Il convient de noter qu'afin de pouvoir réaliser des comparaisons pluriannuelles, la pondération des groupes doit être constante dans le temps. Par ailleurs, si son calcul est sensible à la présence de « zéro », c'est-à-dire à des situations où la part d'un groupe dans une unité k donnée est nulle, cette situation ne se présente quasiment pas dans nos données. Une solution alternative, mais mathématiquement moins rigoureuse peut être d'utiliser l'Indice d'Exposition Normalisé (IEN). En effet, suivant les estimations réalisées par Reardon et O'Sullivan (2004), celui-ci varie très peu à la composition et se caractérise par ailleurs par son interprétation intuitive.

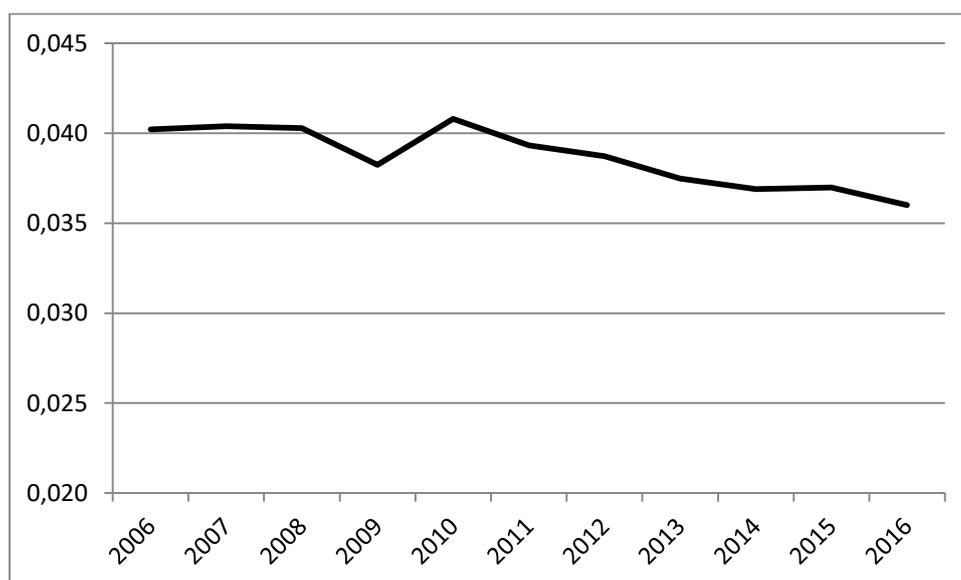
4.2 Résultats de l'analyse

L'indice d'Atkinson asymétrique en structure 2006 a été calculé en utilisant comme unité d'observation la « formation » telle que nous l'avons définie dans ce chapitre, pour l'ensemble de l'Université (sous-paragraphe 4.2.1) ainsi que pour chaque niveau de diplôme (sous-paragraphe 4.2.2).

4.2.1 Les résultats de l'analyse pour l'ensemble de l'Université

Pour calculer nos indices, nous avons utilisé des données provenant de la base SISE sur la période 2006-2016 pour l'ensemble de l'Université hors doctorat. Les résultats obtenus sont présentés ci-dessous dans la Figure 1.

Figure 1: Indice d'Atkinson non corrigé du biais des petites unités pour l'ensemble de l'Université hors doctorat (2006-2016)



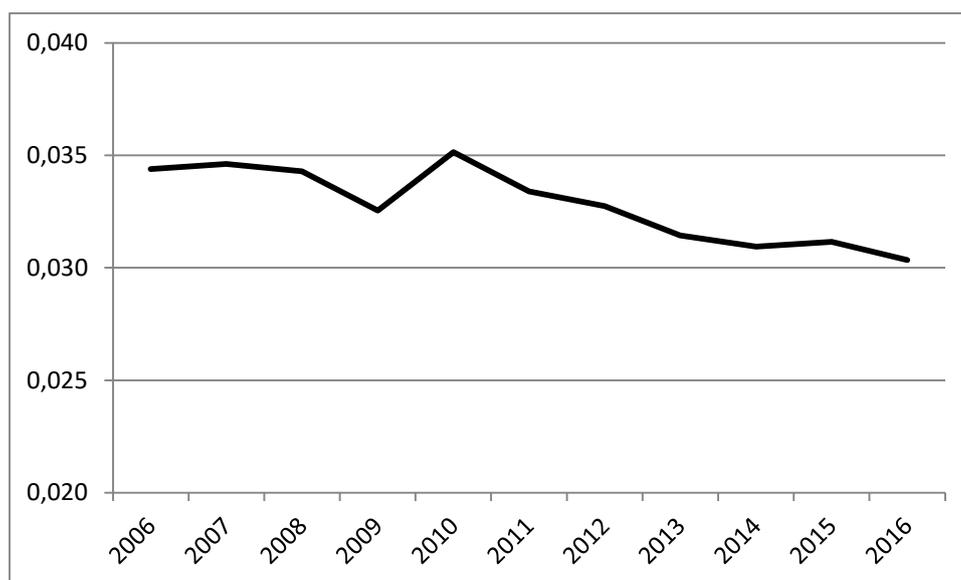
Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.
Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.
Note de lecture : la valeur de l'indice d'Atkinson en 2006 est de 0,04.

Comme indiqué ci-dessus sur la Figure 1, l'estimation réalisée indique qu'il existe un niveau de ségrégation sociale significatif sur la période 2006-2016 parmi les étudiants inscrits dans les Universités de France métropolitaine (hors doctorat). Il convient, toutefois de noter que les résultats présentés ci-dessus sont susceptibles d'être affectés par le problème des petites unités décrits ci-dessus au paragraphe 3.2. A cette fin, comme expliqué précédemment, une solution envisageable est de corriger les résultats obtenus en calculant au moyen d'une méthode de simulation de type *Bootstrap* le niveau de ségrégation qui résulterait en cas de distribution aléatoire des étudiants entre les différentes unités existantes (les résultats des estimations réalisées au moyen des méthodes de simulation sont présentés dans l'annexe 8).

En comparant les valeurs des indices calculés sur la population distribuée de manière aléatoire et les données observées, nous parvenons à la conclusion que les niveaux

de ségrégation calculés sont significativement distincts de 0 avec une probabilité supérieure à 99%. Nous avons aussi, obtenu des valeurs plus robustes statistiquement en déduisant des indices calculés sur les données observées des valeurs moyennes issues d'un jeu de simulation. Les résultats obtenus sont présentés dans la Figure 2 ci-dessous.

Figure 2: Indice d'Atkinson corrigé du biais des petites unités pour l'ensemble de l'Université hors doctorat (2006-2016)³⁵



Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.
 Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.
 Note de lecture : la valeur de l'indice d'Atkinson en 2006 est de 0,034.

Comme indiqué ci-dessus sur la Figure 2, il existe un niveau de ségrégation sociale significatif sur la période 2006-2016 parmi les étudiants inscrits dans les Universités de France métropolitaine (hors doctorat), même si le calcul des indices est corrigé du biais lié aux petites unités. De manière dynamique, les niveaux de ségrégation constatés ont plutôt tendance à diminué faiblement, surtout après 2010. Pour l'ensemble de l'Université, la valeur de l'indice est de 0,034 en 2006 et de 0,030 en 2016, soit une baisse d'environ 12%³⁶. Ceci confirme plutôt le résultat obtenu par Avouac et Harari-Kermadec (2021) qui montrent qu'il existe une baisse relativement lisse de la ségrégation à compter de 2010 en utilisant un indice d'entropie normalisé.

³⁵ L'ensemble des indices suivants présentés dans la suite de ce travail sont corrigés du biais liés aux petites unités.

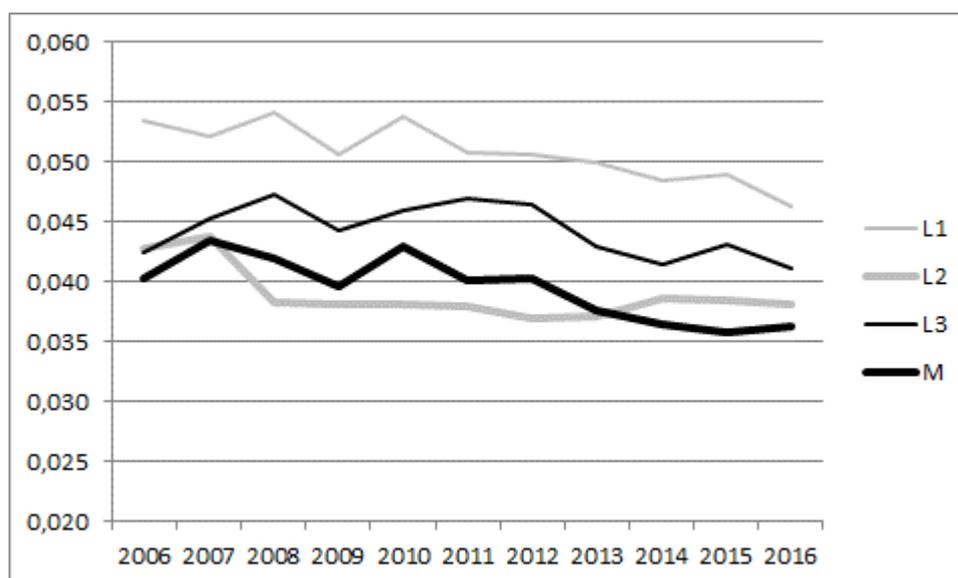
³⁶ Le calcul des indices est disponible de manière détaillée dans l'annexe 8 du présent travail.

Afin de mieux comprendre ce premier résultat, une solution envisageable est de préciser ces tendances par niveau de diplôme, c'est-à-dire de calculer un indice d'Atkinson pour chaque niveau de diplôme³⁷.

4.2.2 Le résultat des estimations par niveau de diplôme

De la même manière que nous avons calculé un indice d'Atkinson pour l'ensemble de l'Université (hors doctorat), il est également possible de calculer des indices pour chaque niveau de diplôme : dans ce cas l'unité d'observation est la formation subdivisée par niveau d'études (soit « Université x discipline x niveau d'études »). Nous avons ainsi calculé cet indice pour les trois années de Licence et le Master. Les résultats obtenus sont présentés ci-dessous dans la Figure 3.

Figure 3 : Indice d'Atkinson corrigé du problème des petites unités pour les trois années de Licence et le Master (2006-2016)



Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : l'indice d'Atkinson en 2006 de 0,053 pour la Licence 1.

Pour la période précédant la mise en place de « Parcoursup », il est possible d'établir les constats suivants à partir de la Figure 3 ci-dessus :

³⁷ Les étudiants sont alors regroupés par niveau de diplôme (L1,L2...). Un indice d'Atkinson est calculé pour chaque niveau.

- Dans cette analyse complémentaire, l'unité d'observation n'est plus la formation au sens où nous l'avons défini antérieurement comme le croisement d'une discipline et d'un établissement universitaire, mais le croisement d'une discipline, d'un établissement universitaire et d'un niveau de diplôme. C'est l'une des raisons pour laquelle les résultats obtenus sont en moyenne supérieurs à ceux obtenus dans le cadre de l'analyse précédente avec une unité d'observation plus agrégée. En effet, comme expliqué ci-dessus au paragraphe 2.2, lorsqu'une population est divisée en plusieurs sous-unités, la valeur de l'indice ne peut diminuer et celle-ci augmente dès lors que l'affectation entre ces sous-unités n'est pas complètement aléatoire.
- en moyenne les niveaux de ségrégation corrigés ont plutôt tendance à diminuer avec le niveau d'études, sauf pour la L3 dont le niveau moyen est supérieur à la L2.
- La baisse du niveau de ségrégation est particulièrement marquée entre la L1 et la L2 ; celle-ci est en moyenne d'environ 23% sur la période 2006-2016.
- Les niveaux de ségrégation ont tendance à diminuer sur la période analysée, surtout à compter de l'année 2010, sauf pour la L2.

Le calcul des indices d'Atkinson sur la période 2006-2016 permet d'estimer les niveaux de ségrégation existants ainsi que leur évolution. Toutefois, une même valeur de l'indice d'Atkinson, comme indiqué précédemment dans la section 2 de ce chapitre, peut renvoyer à des situations variées, les phénomènes de ségrégation pouvant, par exemple, concerner les « très favorisés » ou, inversement, les « défavorisés ». Cette approche pourrait ainsi être complétée, dans un second temps, par l'utilisation d'autres indices permettant d'identifier plus directement les groupes concernés par les phénomènes de ségrégation.

5 Conclusion

L'objectif principal de ce premier chapitre est de définir un cadre méthodologique permettant d'aborder la question de la mixité sociale à l'Université qui nous a permis d'éclaircir les trois points décrits ci-dessous.

Comment définir et mesurer l'origine sociale dans le cadre de l'économie de l'éducation ? Afin de caractériser l'*origine sociale* d'un individu, une solution idéale consiste à prendre en compte simultanément l'ensemble des 9 critères identifiés ou au moins plusieurs d'entre eux, ce qui permet de conférer à chaque individu une valeur sur l'indice de position sociale et de constituer des groupes sociaux homogènes. Toutefois, en raison de la limitation des informations généralement disponibles dans les bases de données usuelles, une autre solution pertinente est de caractériser l'*origine sociale* des individus à l'aune du seul critère de la PCS. Ce choix se justifie dans la mesure où le critère de PCS comporte une dimension sociale, économique et culturelle. L'utilisation de ce critère peut permettre, par ailleurs, d'effectuer des comparaisons avec de nombreux autres travaux utilisant une méthodologie identique ou proche.

Comment définir la mixité sociale à l'Université, mesurer celle-ci au moyen d'indices synthétiques et analyser ses différentes composantes ? A partir de la littérature récente, nous avons répertorié cinq facteurs susceptibles d'influencer les niveaux de ségrégation sociale dans le domaine de l'économie de l'éducation : la ségrégation résidentielle (spatiale), les choix individuels et familiaux, les effets de pairs, les politiques des établissements ainsi que leurs stratégies de communication et, enfin, le cadre légal et les politiques publiques mises en œuvre au plan national. Afin de mesurer le degré de mixité sociale sur plusieurs années, l'indice d'Atkinson peut être retenu car c'est le seul parmi les huit indices retenus à être *stricto sensu* invariant à la composition. Une solution alternative peut être également d'utiliser à titre principal l'Indice d'Exposition Normalisé (IEN) qui varie très peu à la composition et se caractérise par son interprétation intuitive. Afin de décomposer les valeurs mesurées des indices en différents éléments et de quantifier leur importance respective, l'Indice d'information mutuelle semble être plus adapté car il respecte le critère de décomposabilité forte.

Comment choisir une unité d'observation pertinente pour mesurer la mixité sociale à l'Université ? La solution qui nous semble la plus pertinente compte tenu des données disponibles est de décrire les phénomènes de ségrégation sociale à l'Université au niveau de la formation (ou faculté) que nous définissons comme le croisement de l'Université et de la discipline, les autres unités d'observation pouvant être utilisées comme résultat complémentaire. Afin de corriger le problème résiduel lié à l'existence des petites unités, une méthode de correction statistique reposant sur le calcul du niveau de ségrégation mesuré en cas de répartition totalement aléatoire des individus entre les unités existantes semble préférable.

Au-delà du cadre méthodologique défini, nous avons également établi un premier cadrage statistique robuste de l'évolution de la mixité sociale avant la mise en place de Parcoursup. Celui-ci nous a permis d'effectuer les constats suivants : les phénomènes de ségrégation sociale sont statistiquement significatifs au sein de l'Université française, même s'ils ont tendance à diminuer légèrement sur la période 2006-2016. Enfin, ceux-ci ont tendance à diminuer au fur et à mesure de l'avancement dans les études. En l'absence de statistiques ou d'indices complémentaires, nous ne sommes, toutefois, pas en mesure de décrire les ressorts de cette hausse et de proposer une interprétation intuitive sur la base de l'indice d'Atkinson. Afin de parvenir à une analyse à la fois plus fine et plus intuitive, une possibilité envisageable est de procéder à un basculement méthodologique en recourant à d'autres indices. Le chapitre II repose ainsi sur l'utilisation d'un nouvel indice et de statistiques complémentaires pour affiner cette première analyse.

CHAPITRE II LA MIXITE SOCIALE A L'UNIVERSITE : UN PREMIER CADRAGE GEOGRAPHIQUE³⁸

Le chapitre I nous a permis de préciser les principaux enjeux de mesure de la mixité sociale à l'Université et de constater à la fois l'existence d'un niveau de ségrégation sociale significatif au sein de celle-ci sur la période 2006-2016 ainsi qu'une tendance à la baisse, surtout à partir de l'année 2010. Afin d'analyser les niveaux de ségrégation au sein de l'Université française de manière plus fine et de pouvoir interpréter les résultats obtenus de façon intuitive, nous utilisons dans le cadre du chapitre II l'indice d'exposition normalisé. Celui-ci, en effet, peut être calculé pour un groupe en particulier et il permet d'interpréter les résultats obtenus de manière directe.

Ce chapitre propose une application sur données françaises cherchant à produire à la fois un panorama général pour l'ensemble du territoire français, mais également des résultats à un niveau géographique plus fin. Le choix de mesurer les niveaux de ségrégation non seulement au niveau national, mais également au niveau local permet de tester l'hypothèse présentée dans la section 2 du chapitre I suivant laquelle la ségrégation résidentielle et spatiale serait un des principaux facteurs susceptibles d'influencer les niveaux finaux de ségrégation dans le domaine de l'éducation.

Le présent chapitre est structuré de la manière suivante. La première section discute de l'importance de développer une analyse géographique de la ségrégation à l'Université en s'appuyant sur la littérature disponible dans le champ de l'éducation. La deuxième section présente de façon plus détaillée que dans le chapitre I la base de données SISE et apporte des précisions pratiques destinées à mettre en œuvre le calcul des indices de ségrégation/mixité sociale. La troisième présente des statistiques descriptives au niveau national et des éléments de cadrage géographique. La quatrième section expose les résultats obtenus au niveau national et la cinquième les résultats concernant les académies. La sixième section s'interroge sur le rôle de la taille des académies et de la mobilité géographique pour comprendre les niveaux de mixité sociale et la septième conclue.

³⁸ Ce travail réalisé en commun avec Pierre Courtioux et Tristan-Pierre Maury a fait l'objet d'une publication dans la RERU (cf. Courtioux et al. 2022).

1 Les enjeux géographiques de la mixité sociale dans l'enseignement

En complément de la littérature sur la mixité sociale dans l'éducation présentée dans le chapitre I, la littérature présentée dans cette section revient de manière plus spécifique sur les enjeux géographiques de la ségrégation.

La dimension géographique de la ségrégation est essentielle dans les travaux qui portent sur l'enseignement primaire et secondaire. Ly et Riegert (2015), par exemple, après avoir analysé les niveaux de ségrégation aussi bien au Collège qu'au Lycée sur la base de l'indice d'information mutuelle, comme nous l'avons décrit dans le chapitre I, détaillent leurs résultats par académie et montrent que les niveaux de ségrégation sont beaucoup plus élevés en Ile-de-France, notamment dans l'académie de Paris, qu'en Province. Ils constatent également qu'hors Ile-de-France, les académies où la ségrégation est la plus forte sont souvent celles liées à une grande agglomération (Lille, Aix-Marseille et, dans une moindre mesure, Lyon). Comme indiqué dans notre introduction générale, Courtioux et Maury (2020) analysent les phénomènes de ségrégation dans l'enseignement secondaire en utilisant une approche principalement géographique. Ils confirment notamment l'hypothèse que l'enseignement privé peut être utilisé pour contourner la carte scolaire et que celui-ci crée des différences additionnelles au niveau local. Boutechnik et al. (2021) analysent dans quelle mesure la ségrégation urbaine est susceptible de contribuer à la ségrégation scolaire. A partir de données géo-localisées et des secteurs de recrutement de trois villes françaises, les auteurs analysent comment la ségrégation résidentielle et les choix des parents renforcent la ségrégation scolaire entre Collèges.

Il existe également une analyse géographique de la ségrégation dans l'enseignement supérieur, même si celle-ci nous semble moins développée qu'au niveau de l'enseignement primaire et secondaire. Les travaux disponibles sur ce sujet peuvent être regroupés autour de trois pôles principaux :

- l'analyse des conséquences des choix de localisation résidentielle des étudiants qui peuvent conduire à une forme particulière de *gentrification* de certains quartiers parfois appelée *studentification* (Smith et Holt, 2007, Hubbard, 2008).

- l'analyse de la persistance des phénomènes de ségrégation résidentielle au niveau de l'Université. Ainsi dans le cas de l'Ile-de-France, Frouillou (2016) explique que « RAVEL », entre 1987 et 2008, en raison de son système de sectorisation, puis dans une moindre mesure « APB », entre 2009 et 2017, ont contribué à renforcer la ségrégation résidentielle. Ce travail décrit aussi dans quelle mesure les inégalités de fréquentation des formations universitaires en fonction des origines sociales sont liées à l'accessibilité, à l'éloignement géographique et au temps de transport. Comme indiqué dans le chapitre I, il nous semble que les principales limites de ce travail sont la limitation géographique de celui-ci à l'Ile-de-France et l'absence d'utilisation des indices synthétiques de ségrégation.

- Plusieurs travaux analysent le phénomène de la mobilité étudiante, en particulier à la sortie de l'enseignement secondaire. Perret (2007) décrit les différents facteurs explicatifs des niveaux de mobilité : discipline d'inscription de l'étudiant (cette mobilité étant par exemple plus élevée pour les étudiants des IUT), origine sociale (départ plus fréquent pour l'Ile-de-France pour les étudiants de Province provenant d'un milieu aisé), influence du parcours scolaire (avec un impact négatif des retards et de l'échec) et poids important du contexte régional. La mobilité géographique des jeunes à l'entrée dans le supérieur et notamment hors de l'académie d'origine, a par ailleurs été documentée par Bluntz (2018) ainsi que Fabre et al. (2019), qui mettent en avant l'influence des ressources sociales et économiques des familles et celle de l'implantation des sites d'enseignement supérieur.

Cette section nous permet de tirer les trois enseignements suivants en ce qui concerne l'analyse de la dimension géographique de la mixité sociale à l'Université :

1. Une double approche, incluant une dimension à la fois nationale, mais aussi géographique fine, reposant sur une comparaison des niveaux de ségrégation entre académies est nécessaire, dans la continuité des travaux réalisés par Ly et Riegert (2015) sur l'enseignement primaire et secondaire.

2. Afin de pouvoir effectuer des comparaisons avec les niveaux de ségrégation existants dans l'enseignement secondaire, il conviendrait d'utiliser l'indice d'information mutuelle, dont l'interprétation est facile et qui nous permettra à nouveau de nous situer dans le prolongement des travaux de Ly et Riegert (2015). Comme expliqué dans le chapitre I, celui-ci est par ailleurs peu sensible à la composition et son interprétation est intuitive.
3. La ségrégation résidentielle persiste au niveau de l'Université, mais il s'agit d'un facteur moins important qu'au niveau de l'enseignement primaire et secondaire en raison de la disparition de la carte scolaire et de la mobilité des étudiants.

2 Base de données utilisée et mesure de la mixité

A partir de la présente section, la stratégie de mesure de la mixité sociale présentée de manière conceptuelle dans le chapitre I est déclinée de manière opérationnelle et pratique. Si nous avons vu dans le chapitre I que pour produire un indice de mixité sociale, il était nécessaire d'avoir une base de données permettant notamment d'identifier des groupes sociaux, ainsi que des unités de base de mesure de la ségrégation, cette deuxième section montre que la base de données que nous utilisons, la Base SISE Université, convient tout à fait à ce type de calcul. Dans cette section, nous présentons la base de données utilisée (paragraphe 2.1), apportons des précisions pratiques concernant la population de référence identifiée (paragraphe 2.2), le contenu des disciplines académiques SISE (paragraphe 2.3) et l'unité de mesure de la ségrégation (paragraphe 2.4).

2.1 La base de données SISE

Nous utilisons la base SISE Université du MESRI. La période disponible s'étend des années universitaires 2006-2007 à 2016-2017³⁹. Celle-ci contient notamment des informations sur les différents établissements universitaires et sur les niveaux d'études, filières et origines sociales de l'ensemble des étudiants sur l'ensemble de la période 2006-2016. Le nombre de variables comprises dans la base SISE Université a augmenté au fil des années, passant de 42 variables pour l'année universitaire 2006-2007 à 74 pour l'année universitaire 2016-2017. Celles-ci concernent l'établissement fréquenté, des données sociodémographiques et géographiques relatives à l'étudiant ainsi qu'à sa scolarité antérieure et actuelle. Il convient de noter, toutefois, que seules 40 variables étaient diffusées par le MESRI via l'Adisp-Progedo pour l'année universitaire 2016-2017. L'Annexe 3 décrit, à titre d'exemple, pour l'année 2016-2017, la liste des variables diffusées par le MESRI.

³⁹ Il s'agit de la disponibilité des bases de données SISE au moment où ce travail de thèse a débuté, en 2019. De nouvelles années sont, toutefois, disponibles depuis cette date.

Cette base est exhaustive (elle contient des informations relatives à la totalité des étudiants inscrits à chaque rentrée dans le système universitaire). La base SISE Université que nous utilisons se scinde en SISE-Inscrits et SISE-Résultats, cette dernière complétant la première par des informations relatives aux résultats des examens. Dans la mesure où nous cherchons ici à mesurer la mixité sociale (et pas la mixité scolaire), nous n'utilisons ici que l'enquête SISE-Inscrits. Par ailleurs, s'il existe d'autres bases de données SISE, celles-ci ne sont pas exploitées dans le cadre du présent travail pour les raisons détaillées dans notre introduction générale.

2.2 Précisions concernant le champ des établissements retenus

Comme expliqué dans notre introduction générale, nous avons choisi de limiter notre champ d'analyse à l'ensemble des étudiants inscrits dans les Universités de France métropolitaine⁴⁰. Afin de demeurer au sein de ce périmètre, il convient de conserver dans le champ de notre analyse uniquement les établissements universitaires⁴¹. Des précisions sur ce point nous semblent d'autant plus nécessaires que le paysage de l'enseignement supérieur s'est complexifié depuis plusieurs années et que la base SISE Université inclut différents types d'établissements dont tous n'ont pas le statut universitaire. Afin d'analyser cette question de manière plus détaillée, le présent paragraphe est structuré de la manière suivante : le sous-paragraphe 2.2.1 décrit l'ensemble des établissements non universitaires, le sous-paragraphe 2.2.2 explique les raisons pour lesquelles les ESPE et les COMUE n'ont pas non plus été retenues et le sous-paragraphe 2.2.3 décrit les principaux types d'établissements conservés dans le périmètre de notre analyse.

2.2.1 Le cas des établissements non universitaires

Depuis 2013, 10 établissements non universitaires sont inclus dans la base SISE:

⁴⁰ A cette fin, seuls les établissements correspondants aux 25 académies de France métropolitaine en excluant la Corse ont été retenus.

⁴¹ Afin de retenir les établissements universitaires pertinents uniquement, nous avons utilisé la variable « ETABLI » qui indique le numéro d'identification de l'établissement dans lequel l'étudiant est inscrit, une variable « ETABLI_DIFFUSION » permettant de connaître également leur nom à compter de la rentrée 2013.

- 5 établissements spécifiques à l'agglomération parisienne : l'institut national des langues et civilisations orientales (INALCO), l'institut d'études politiques, l'observatoire de Paris, l'institut physique du globe de Paris et Paris Dauphine.
- 3 Universités de technologie (Compiègne, Troyes, Belfort-Montbéliard) qui sont des écoles d'ingénieurs habilitées à délivrer ce diplôme.
- l'école nationale du paysage de Blois⁴².
- L'institut national polytechnique de Grenoble (Grenoble INP), qui est devenu un grand établissement en mars 2007.

Ces 10 établissements sont exclus du périmètre de notre travail de recherche car ils n'ont pas le statut juridique d'Université. Ceux-ci présentent des caractéristiques académiques et de recrutement distinctes des établissements universitaires, avec notamment la possibilité de recourir à la sélection des étudiants dès le premier cycle, pour ceux d'entre eux qui en ont un. Comme expliqué dans notre introduction générale, le choix de ne pas inclure ces établissements dans le champ de notre analyse est d'autant plus justifié que les bases de données portant sur les établissements non universitaires ne sont pas diffusées par le MESRI, ce qui empêche d'avoir une vision globale de leur situation.

2.2.2 Le cas des ESPE et des COMUE

Les écoles supérieures du professorat et de l'éducation (ESPE) et les communautés d'Universités et établissements (COMUE) n'ont pas été davantage retenues dans le champ de notre analyse pour des raisons spécifiques à chacune de ces catégories d'établissements. Les ESPE, créées en 2013, et succédant ainsi aux instituts universitaires de formation des maîtres (IUFM), sont des écoles professionnelles destinées à apprendre progressivement et par l'alternance le métier d'enseignant. Les

⁴²L'ENS NATURE PAYSAGE BLOIS n'est présente dans la base SISE qu'en 2013 et 2014.

concours de recrutement sont organisés en fin de première année de Master et la formation en alternance en deuxième année de Master est rémunérée. En raison notamment du statut des élèves qui sont rémunérés après la réussite du concours en seconde année de Master, nous avons fait le choix de ne pas les conserver dans le champ de notre analyse. Les COMUE, qui ont été créés par la loi du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche, n'ont pas non plus été retenues pour les raisons suivantes :

- Conformément à leur objectif de renforcer la coopération entre différentes catégories d'établissements, celles-ci comprennent des établissements non universitaires.
- Les COMUE apparaissent pour la plupart dans la base de données SISE à partir de 2016 uniquement, soit la dernière année retenue dans notre travail.
- Ne pas prendre en compte les COMUE n'a pas d'impact significatif à l'échelle de la France métropolitaine, sauf pour l'année 2016 (exclusion de 16 416 étudiants).

2.2.3 Le cas des IUT et des INP

Après exclusion des établissements d'enseignement supérieur susmentionnés, il reste 88 établissements compris dans notre champ d'analyse, une liste exhaustive d'établissements non retenus et conservés étant disponibles respectivement dans les annexes 4 et 5. Les établissements conservés dans le champ de notre analyse comprennent ainsi les Universités de France métropolitaine ainsi que les instituts nationaux polytechniques qui sont « assimilés aux Universités » suivant l'article L711-2 du Code de l'Education, même si l'admission en premier cycle de ceux-ci est sélective. Il nous semble que deux précisions peuvent être apportées sur ce point. D'abord, les instituts universitaires de technologie (IUT) qui sont des instituts internes aux Universités ne peuvent être distingués dans la base SISE du reste de l'établissement auquel ils appartiennent et sont donc nécessairement conservés dans le champ de notre analyse. Concernant les INP, ensuite, l'institut national polytechnique de Lorraine (Lorraine INP), qui a fusionné avec les Universités de Nancy

et de Metz pour créer le grand établissement Université de Lorraine en janvier 2012 est conservé dans le champ de notre analyse. Il en est de même pour l'institut national polytechnique de Toulouse (Toulouse INP), qui a gardé ce statut jusqu'à sa transformation future en Toulouse Centrale Institut, annoncée en 2022. En revanche, l'INP de Grenoble n'est pas conservé : suite à un décret du 8 mars 2007, cet INP est devenu un grand établissement et il s'agit aujourd'hui d'une Ecole d'ingénieur habilitée à délivrer ce type de diplôme.

Comme rappelé dans l'introduction générale, entre 2009 et 2017, plusieurs Universités ont fusionné afin de défragmenter le paysage de l'enseignement supérieur et d'atteindre une taille critique permettant d'apparaître dans les classements internationaux. Comme décrit dans l'annexe 5, certains établissements disparaissent ainsi au cours de la période analysée (2006-2016) alors que d'autres établissements sont créés pour en remplacer plusieurs. Il convient par ailleurs de noter que la base SISE fournit des indications sur l'établissement d'inscription de l'étudiant, mais n'indique pas la localisation géographique de celui-ci et ne comprend pas d'informations relatives aux antennes universitaires. Afin d'analyser notamment les migrations des étudiants entre les différents pôles d'un même établissement dans le cadre d'une réflexion sur les politiques d'aménagement du territoire universitaire, le travail de Baron (2005) utilise ainsi des données antérieures à 1994 centralisées par la Direction de l'Évaluation et de la Prospective (DEP) dans le cadre de la norme « Effectifs Universitaires » (EFU) et non la base SISE⁴³.

2.3 L'appréhension des disciplines académiques avec SISE

Conformément aux résultats de notre chapitre I, nous avons choisi comme unité de mesure principale de la ségrégation la *formation* que nous définissons comme le croisement de l'établissement universitaire et de la discipline. Ayant décrit dans le paragraphe précédent le périmètre des établissements universitaires retenus dans le cadre de notre analyse, il convient également de préciser le contenu des disciplines SISE. La base de données SISE contient, en effet, une variable « DISCIPLI » qui indique la discipline académique correspondant au diplôme d'inscription de l'étudiant.

⁴³ La base de données SISE, telle qu'elle est diffusée par l'Adisp-Progedo, ne contient pas d'information sur ce plan.

L'unité de mesure principale de la ségrégation correspond ainsi au croisement des variables « ETABLI » et « DISCIPLI » de la base SISE.

Le tableau 6 ci-dessous liste l'ensemble des disciplines SISE répertoriées et rappelle les matières (ou « secteur disciplinaire ») que chacune d'elle comprend.

Tableau 6 : Les disciplines SISE présentes dans la base

Code SISE	Discipline SISE (variable « DISCIPLI »)	Secteur disciplinaire SISE et code secteur disciplinaire (variable « SECTDIS »)
1	DROIT - SCIENCES POLITIQUES	36 SCIENCES JURIDIQUES, 37SCIENCES POLITIQUES, 61 PLURI DROIT – SCIENCES POLITIQUES
2	SCIENCES ECONOMIQUES – GESTION (HORS AES)	38 SCIENCES ECONOMIQUES, 39 SCIENCES DE GESTION, 62 PLURI SCIENCES ECONOMIQUES - GESTION
3	ADMINISTRATION ECONOMIQUE ET SOCIALE (AES)	40 ADMINISTRATION ECONOMIQUE ET SOCIALE (AES)
4	LETTRES – SCIENCES DU LANGAGE – ARTS	17SCIENCES DU LANGAGE - LINGUISTIQUE, 18 LANGUES ET LITTERATURES ANCIENNES, 19 LANGUES ET LITTERATURES FRANCAISES, 20 LITTERATURE GENERALE ET COMPAREE, 21 ARTS, 64 PLURI LETTRES – SCIENCES DU LANGAGE – ARTS
5	LANGUES	22 FRANCAIS, LANGUE ETRANGERE, 23 LANGUES ET LITTERATURES ETRANGERES, 24 LANGUES ETRANGERES APPLIQUEES, 25 CULTURES ET LANGUES REGIONALES, 65 PLURI LANGUES
6	SCIENCES HUMAINES ET SOCIALES	26 PHILOSOPHIE, EPISTEMOLOGIE, 27 HISTOIRE, 28 GEOGRAPHIE, 29 AMENAGEMENT, 30 ARCHEOLOGIE, ETHNO., PREHISTOIRE, ANTHROPOLOGIE, 31 SCIENCES RELIGIEUSES, 32 PSYCHOLOGIE, SCIENCES COGNITIVES, 33 SOCIOLOGIE, DEMOGRAPHIE, 34 SCIENCES DE L'EDUCATION, 35 SCIENCES DE L'INFORMATION ET DE LA COMMUNICATION, 66 PLURI SCIENCES HUMAINES ET SOCIALES
9	SCIENCES DE LA VIE, DE LA SANTE, DE LA TERRE ET DE L'UNIVERS	05 SCIENCES DE L'UNIVERS, DE LA TERRE, DE L'ESPACE, 06 SCIENCES DE LA VIE, BIOLOGIE, SANTE, 69 PLURI SCIENCES DE LA VIE, DE LA SANTE, DE LA TERRE ET DE L'UNIVERS
10	STAPS	10 STAPS
11	MEDECINE	07 MEDECINE
12	ODONTOLOGIE	08 ODONTOLOGIE
13	PHARMACIE	09 PHARMACIE

Code SISE	Discipline SISE (variable « DISCIPLI »)	Secteur disciplinaire SISE et code secteur disciplinaire (variable « SECTDIS »)
15	SCIENCES FONDAMENTALES ET APPLICATIONS	01 MATHÉMATIQUES 02 PHYSIQUE 03 CHIMIE 43 PHYSIQUE ET CHIMIE 42 MATHÉMATIQUES ET INFORMATIQUE 04 MATHÉMATIQUES APPLIQUÉES AUX SCIENCES SOCIALES (MASS) 11 MÉCANIQUE, GENIE MÉCANIQUE, INGÉNIERIE MÉCANIQUE 15 SCIENCES FONDAMENTALES ET APPLICATIONS 12 GENIE CIVIL 13 GENIE DES PROCÉDES, MATÉRIAUX 14 INFORMATIQUE 15 ÉLECTRONIQUE , GENIE ÉLECTRIQUE, EEA 16 SCIENCES ET TECHNOLOGIE INDUSTRIELLES 41 FORMATION GÉNÉRALE AUX MÉTIERS DE L'INGÉNIEUR 68 PLURI SCIENCES FONDAMENTALES ET APPLICATIONS
31	PLURI DROIT-SCIENCES ECO- AES	63 PLURI DROIT-SCIENCES ECO-AES
32	PLURI LETTRES – LANGUES – SCIENCES HUMAINES	67 PLURI LETTRES – LANGUES – SCIENCES HUMAINES
33	PLURI SCIENCES	70 PLURI SCIENCES
34	PLURI SANTÉ ⁴⁴	71 PLURI SANTÉ

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo).

Le tableau 6 indique que le nombre de disciplines SISE varie au cours de la période analysée : il y en a 14 sur la période 2006-2008, 15 à compter de la rentrée 2009 avec l'introduction de la modalité « 31 - PLURI DROIT-SCIENCES ECO-AES », puis 16 en 2010 avec la modalité « 34 - PLURI SANTÉ »⁴⁵. Il convient, toutefois, de noter que si la base de données SISE inclut des effectifs dès la rentrée 2010 pour la modalité 34, celle-ci n'en comprend pour la modalité 31 qu'à compter de la rentrée 2013.

Le tableau 6 indique la correspondance entre la variable « DISCIPLI » que nous avons utilisé et la variable « SECTDIS » qui inclut 52 postes et décrit ainsi de manière fine la matière ou le secteur disciplinaire d'inscription de l'étudiant. Deux raisons sont

⁴⁴ Cette discipline correspond à la PACES, première année commune aux études de santé mise en place à la rentrée 2010.

⁴⁵ A compter de la rentrée 2012, la base SISE comprend également la discipline "Vétérinaire" et la discipline "Théologie". Toutefois, celles-ci sont utilisables respectivement par les écoles vétérinaires et les instituts catholiques uniquement. Le champ de notre analyse n'est donc pas concerné par ces nouvelles disciplines.

susceptibles de justifier l'utilisation de la variable « DISCIPLI » plutôt que la variable « SECTDIS » pour définir la formation telle que nous l'entendons. D'abord, en 2014, la mise en place du Cadre national des formations (CNF)⁴⁶ a modifié la répartition disciplinaire des diplômes, entraînant une rupture statistique alors que les évolutions par regroupement disciplinaire au niveau « DISCIPLI » ne pâtissent pas de cette rupture statistique. Ensuite, les unités d'observation retenues, soit le croisement d'un établissement et d'un secteur disciplinaire n'aurait pas été de taille suffisante en cas d'utilisation de la variable « SECTDIS » (au regard de la nécessité décrite dans le chapitre I de trouver un point d'équilibre entre l'exigence de finesse de la description et le problème statistique soulevé par les petites unités). Par ailleurs, la base de données SISE comprend une variable « GROUPE » qui est un regroupement de la variable « DISCIPLI » en 8 catégories. Il nous semble, toutefois, que le choix de cette variable aurait donné lieu, à l'inverse, à un niveau d'analyse insuffisamment granulaire.

2.4 La formation comme unité d'observation dans SISE

Comme indiqué dans le paragraphe précédent, l'unité d'observation choisie correspond au croisement des variables « ETABLI » et « DISCIPLI » de la base SISE. Ce choix permet de regrouper les étudiants par Université et par discipline.

Si les variables d'établissements universitaires (88 sur la période 2006-2016) et de disciplines (16) sont croisées, le nombre de formations pourrait au total être de 1 408. Toutefois, tous les établissements universitaires n'offrent pas la totalité des disciplines répertoriées dans SISE. Nous avons ainsi sélectionné dans la base SISE uniquement les formations existantes et comprenant au moins un étudiant et, ensuite, documenter la taille des unités d'observation que nous avons choisies et vérifier qu'elles sont compatibles avec les critères que nous avons énoncé dans le chapitre I sur les problèmes de taille des unités d'observation.

Pour cela, nous décrivons successivement la distribution de ces unités lorsque le niveau de décomposition maximal est utilisé et lorsque la formation est définie comme le croisement de l'établissement et de la discipline.

⁴⁶ L'arrêté du 22 janvier 2014 fixe le cadre national des formations conduisant à la délivrance des diplômes nationaux de Licence, de Licence professionnelle et de Master.

Le Tableau 7 ci-dessous décrit la distribution des formations si l'unité d'observation retenue correspondait au niveau de décomposition maximal, à savoir « Université x discipline x niveau d'études ».

Tableau 7 : distribution des tailles de formations (*) selon le niveau d'études (moyenne sur la période 2006-2016)

	L1	L2	L3	M1	M2	Doctorat
Percentile 1	9	10	9	6	4	1
Percentile 5	41	32	25	19	13	3
Percentile 10	89	54	46	32	25	6
1er quartile	235	121	111	77	65	19
Médiane	484	248	248	174	153	71
Nombre de formations	702	384	399	343	317	252

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note : (*) la formation comme unité d'observation est déterminée dans ce cas par le croisement Université/discipline/niveau d'étude.

La lecture du Tableau 7 ci-dessus permet de constater qu'au fur et à mesure de l'avancement dans les études, la taille des unités de formation diminue, en particulier à partir du Master. Ainsi, au niveau du doctorat, au niveau du premier quartile, la taille des formations est de 19 étudiants.

Afin de tenir compte du problème statistique soulevé par les petites unités, nous avons effectué les choix suivants :

- Comme indiqué dans le cadre du chapitre I, l'unité d'observation principale retenue dans le cadre de notre analyse est la formation définie comme le croisement d'un établissement universitaire et d'une discipline, le niveau de désagrégation maximal (croisement d'un établissement universitaire, d'une discipline et d'une année d'études) n'étant utilisé que de manière secondaire.
- Le champ de notre analyse ne retient pas les formations de doctorat dont la taille nous semble faible au regard du problème statistique soulevé par les petites unités décrit dans le cadre du chapitre I.

Le Tableau 8 ci-dessous décrit la distribution des formations lorsque l'unité d'observation retenue correspond au croisement d'une Université et d'une discipline.

Tableau 8 : distribution de la taille des formations (*) tous niveaux d'études confondus (2006-2016)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Percentile 1	9	7	7	8	9	5	8	9	9	4	9
Percentile 5	28	29	29	30	29	24	24	27	28	26	30
Percentile 10	89	88	84	79	70	61	70	84	93	82	89
1er quartile	411	383	352	347	391	397	385	419	420	421	432
Médiane	1045	985	976	994	997	1029	1064	1082	1095	1122	1105
Nombre de formations	721	721	740	750	769	746	750	745	734	746	744

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note : (*) la formation comme unité d'observation est déterminée dans ce cas par le croisement Université/discipline.

La lecture du Tableau 8 nous semble confirmer que le choix méthodologique effectué, à savoir l'utilisation de la formation définie comme le croisement d'une Université et d'une discipline hors doctorat réduit les risques qui résultent du problème statistique des petites unités. En effet, au niveau du 5^{ème} percentile, le nombre d'étudiants par unité d'observation (formation) est systématiquement supérieur à 20, ce qui réduit considérablement le problème statistique des petites unités.

Il convient de noter que la liste des formations telle que nous l'avons définie ne se recoupe pas nécessairement avec celle des Unités de Formation et de Recherche (UFR), des Ecoles doctorales ou des départements des différents établissements retenus. La composition et l'organisation de ces unités est, en effet, propre au fonctionnement de chaque établissement. A titre d'exemple, l'Université de Montpellier comprend huit facultés : droit et science politique, médecine, odontologie, pharmacie, STAPS, économie, éducation et Sciences. Les cinq premières facultés correspondent aux disciplines SISE 1, 11, 12, 13 et 10 mentionnées dans le tableau 6 ci-dessus, les deux suivantes peuvent être incluses dans les disciplines 2 et 6 (mais correspondent à une partie de celles-ci seulement) et la dernière se répartit entre les disciplines 9, 15 et 33. La formation telle que nous l'avons définie, permet, ainsi de regrouper les étudiants de l'ensemble du système universitaire dans les différentes disciplines identifiées de manière relativement fine sur la base des informations disponibles dans la base SISE.

Il convient par ailleurs de noter que la base de données SISE inclut à partir de la rentrée 2012 une variable indiquant si l'étudiant est inscrit en parallèle en CPGE. Toutefois, afin de conserver un certain niveau de cohérence de la population analysée

sur la période 2006-2016, nous n'avons pas utilisé celle-ci pour éliminer les étudiants concernés à partir de la rentrée 2012. En effet, les étudiants inscrits à la fois en CPGE et à l'Université sont déjà présents dans la base SISE sur la période 2006-2011, mais il n'existe pas pour ces années à notre connaissance de variable disponible permettant de les identifier.

3 Premiers éléments de cadrage statistique

Afin d'offrir des éléments de cadrage à la fois au niveau national et des zones académiques, avant de procéder au calcul des indices de ségrégation, cette section introduit des statistiques descriptives relatives aux étudiants inscrits à l'Université en France métropolitaine, via la description de la catégorie socioprofessionnelle du « parent référent ». Ainsi, le paragraphe 3.1 présente des statistiques descriptives pour l'ensemble de l'Université, le paragraphe 3.2 des statistiques descriptives par discipline alors que le paragraphe 3.3 porte sur les effectifs des académies et le paragraphe 3.4 sur la répartition des étudiants par origine sociale et par académies.

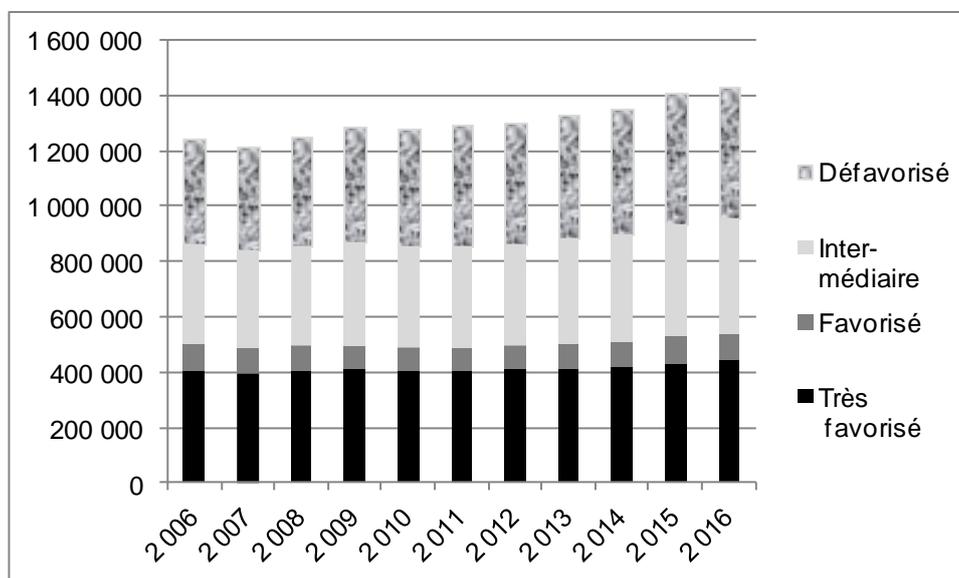
3.1 Origine sociale et formations à l'Université

Afin de décrire les liens entre origine sociale et formations à l'Université, on commence par présenter dans le sous-paragraphe 3.1.1 des statistiques descriptives pour l'ensemble du cycle universitaire. Le sous-paragraphe 3.1.2 introduit, ensuite, des statistiques descriptives par niveau de diplôme.

3.1.1 Statistiques descriptives concernant l'ensemble du cycle universitaire

Sur le plan national, des éléments de cadrage concernant les effectifs des étudiants ainsi que les parts respectives occupées par les quatre groupes sociaux identifiés sont indispensables. A cette fin, à partir de la base SISE, nous avons calculé les effectifs étudiants par année sur l'ensemble de la période disponible, à savoir 2006-2016. La Figure 4 ci-dessous indique ainsi l'origine sociale de l'ensemble des étudiants hors doctorants inscrits dans les Universités de France métropolitaine sur la période disponible. Il convient de rappeler que ces données excluent les établissements non universitaires compris dans la base SISE, pour les raisons exposées dans le paragraphe 2.2 du présent chapitre.

Figure 4 : effectifs de l'ensemble des Universités et parts respectives des quatre groupes sociaux (2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur

Sur la période 2006-2016, la Figure 4 permet ainsi d'effectuer les constats suivants⁴⁷ :

1. les effectifs totaux inscrits ont augmenté de manière ininterrompue, passant de 1 247 131 en 2006 à 1 435 216 en 2016, soit une hausse d'environ 15% sur la période.
2. en moyenne, sur l'ensemble de la période, les étudiants issus du groupe défavorisé sont légèrement majoritaires (32,7%), suivis de ceux issus du groupe très favorisé (31,4%), puis ceux du groupe intermédiaire (28,8%) et, finalement, ceux provenant du groupe favorisé (7,1%).
3. la part du groupe très favorisé est en moyenne de 31,4%, mais celle-ci a légèrement diminué, passant de 32,1% en 2006 à 30,9% en 2016.
4. la part du groupe favorisé a décru de manière assez régulière également, passant de 8,1% en 2006 à 6,8% en 2016.
5. la part du groupe intermédiaire a très faiblement progressé, passant de 28,9% en 2006 à 29% en 2016.

⁴⁷ Les chiffres correspondants sont disponibles en annexe 6.

6. la part du groupe défavorisé a augmenté de manière presque ininterrompue sur la période 2006-2016, passant de 30,9% en 2006 à 33,3% en 2016.

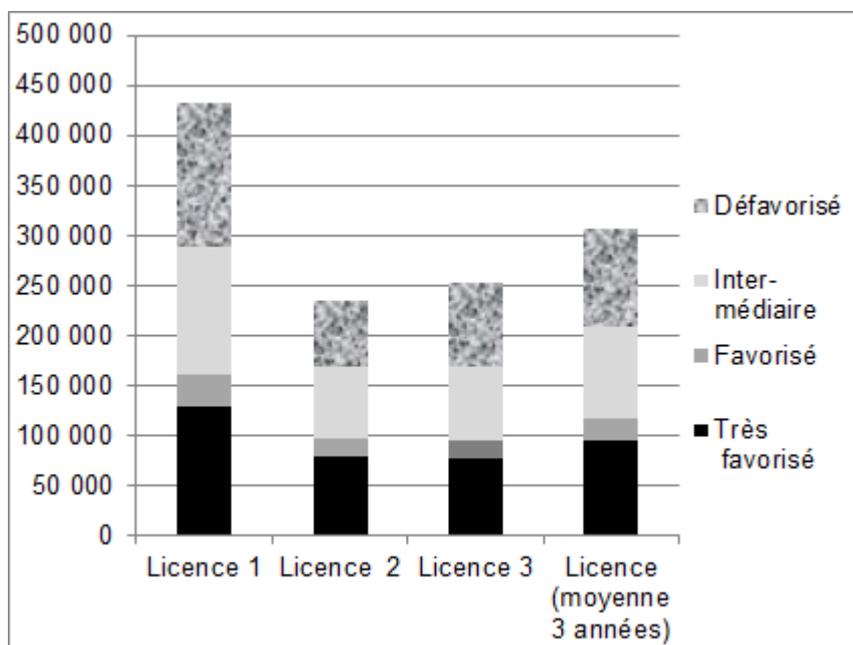
Au final, sur la période 2006-2016, la progression des effectifs totaux est surtout tirée par le groupe défavorisé (+23,9%), suivi du groupe intermédiaire (+15,4%) et du groupe très favorisé (+10,8%) alors que les effectifs du groupe favorisé ont diminué (- 3,3%).

3.1.2 Origine sociale et niveau d'études

Toujours au niveau national et pour la même population de référence, des éléments de cadrage supplémentaires concernant les effectifs des étudiants par niveau de diplôme nous semblent également nécessaires.

La Figure 5 ci-dessous présente ainsi l'origine sociale des étudiants du système universitaire français pour la Licence sur la période 2006-2016⁴⁸ :

Figure 5: effectifs pour la Licence et parts respectives des quatre groupes sociaux (moyenne sur la période 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur.

⁴⁸ Les chiffres correspondants sont disponibles en annexe 6.

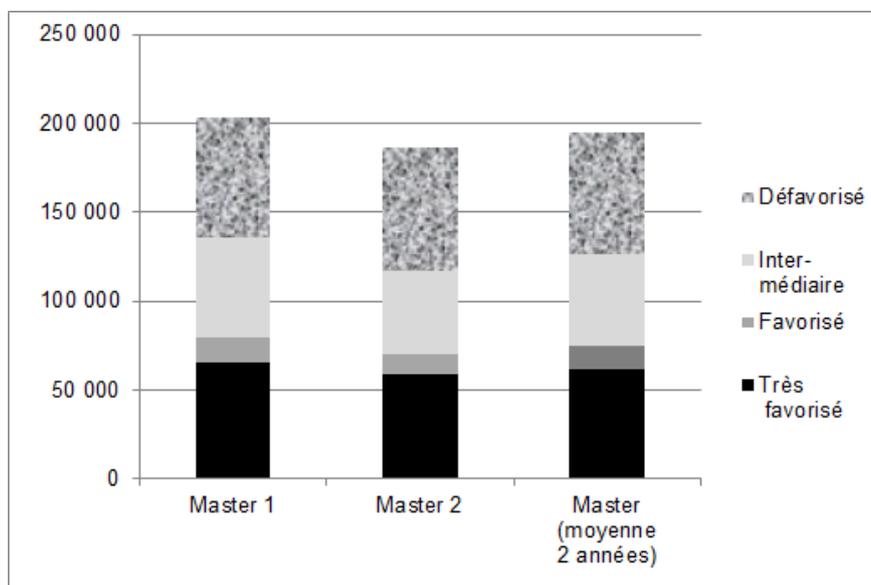
La Figure 5 permet d'effectuer les principaux constats suivants :

1. En Licence 1, les étudiants défavorisés sont nettement majoritaires, avec une part estimée à 33,2%, suivis des étudiants très favorisés (30%).
2. En Licence 2, la part des étudiants défavorisés diminue de façon très nette, celle-ci passant de 33,2% à 27,4% alors que celle des étudiants très favorisés passe de 30% à 34,1%.
3. L'effectif des étudiants augmente entre la L2 et la L3 de 7,9% en moyenne sur la période 2006-2016. Cette hausse peut être reliée au phénomène des redoublements à l'issue de la L3 (pour les étudiants qui n'obtiennent pas leur diplôme), mais aussi à l'arrivée d'étudiants provenant de formations non universitaires (CPGE et BTS en particulier). La progression des effectifs entre la L2 et la L3 est tirée par la progression en valeur absolue du nombre d'étudiants défavorisés (+30,5%) alors que les effectifs des étudiants intermédiaires est relativement stable (+2,2%) et que celle des favorisés et des très favorisés diminue (respectivement -0,6% et -3,0%). En Licence 3, les parts occupées par les étudiants très favorisés (30,6%) et défavorisés (33,2%) se situent ainsi à un niveau proche de la Licence 1.

La nette diminution de la part occupée par les étudiants défavorisés entre la L1 et la L2 peut s'interpréter à l'aide de certains éléments issus de la revue de la littérature. D'une part, comme l'a documenté Frouillou (2016), les effets de pairs à l'entrée de l'Université sont particulièrement importants pour les élèves issus d'un milieu défavorisé qui ont tendance à s'orienter vers les mêmes formations que leurs pairs. D'autre part, le taux d'échec en Licence des étudiants provenant d'un milieu défavorisé est plus élevé que celui de la moyenne des étudiants (Brinbaum et al., 2018) et ce taux est particulièrement élevé en première année de Licence. Ce taux d'échec semble ainsi se traduire directement par une baisse de la part occupée par les étudiants défavorisés à l'issue de la première année de Licence. En effet, d'après un panel d'élèves entrés au Collège en 1995, parmi les étudiants qui sont entrés dans les études supérieures par un premier cycle universitaire, 70% des enfants de cadres ou professions intermédiaires y obtiennent une Licence contre 52% des enfants d'ouvriers.

La Figure 6 ci-dessous présente les effectifs de Master et l'origine sociale des étudiants⁴⁹:

Figure 6: effectifs de Master pour l'ensemble des Universités et parts respectives des quatre groupes sociaux (moyenne sur la période 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
 Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur.

Sur la période 2006-2016, il est possible d'effectuer les constats suivants à partir des Figures 5 et 6 :

1. Les effectifs d'étudiants diminuent en moyenne entre la L3 et le M1 d'environ 50 000 étudiants, soit une baisse d'environ 20% des effectifs entre ces deux années.
2. Les effectifs diminuent à nouveau en moyenne entre le M1 et le M2, ceux-ci passant d'environ 203 000 à environ 186 000.
3. Entre le M1 et le M2, la part occupée par le groupe défavorisé passe en moyenne de 33,4% à 37,1% alors celle des autres groupes sociaux diminue.
4. La part des étudiants issus du groupe très favorisé est proche en moyenne en Master (31,9%) et en Licence (31,2%).
5. La part des étudiants issus du groupe favorisé est proche en moyenne en Master (7,3%) et en Licence (6,5%).

⁴⁹ Les chiffres correspondants sont disponibles en annexe 6.

6. La part des étudiants issus du groupe intermédiaire est plus faible en moyenne en Master (26,5%) qu'en Licence (29,8%).
7. La part des étudiants issus du groupe défavorisé est plus élevée en moyenne en Master (35,1%) qu'en Licence (31,7%).

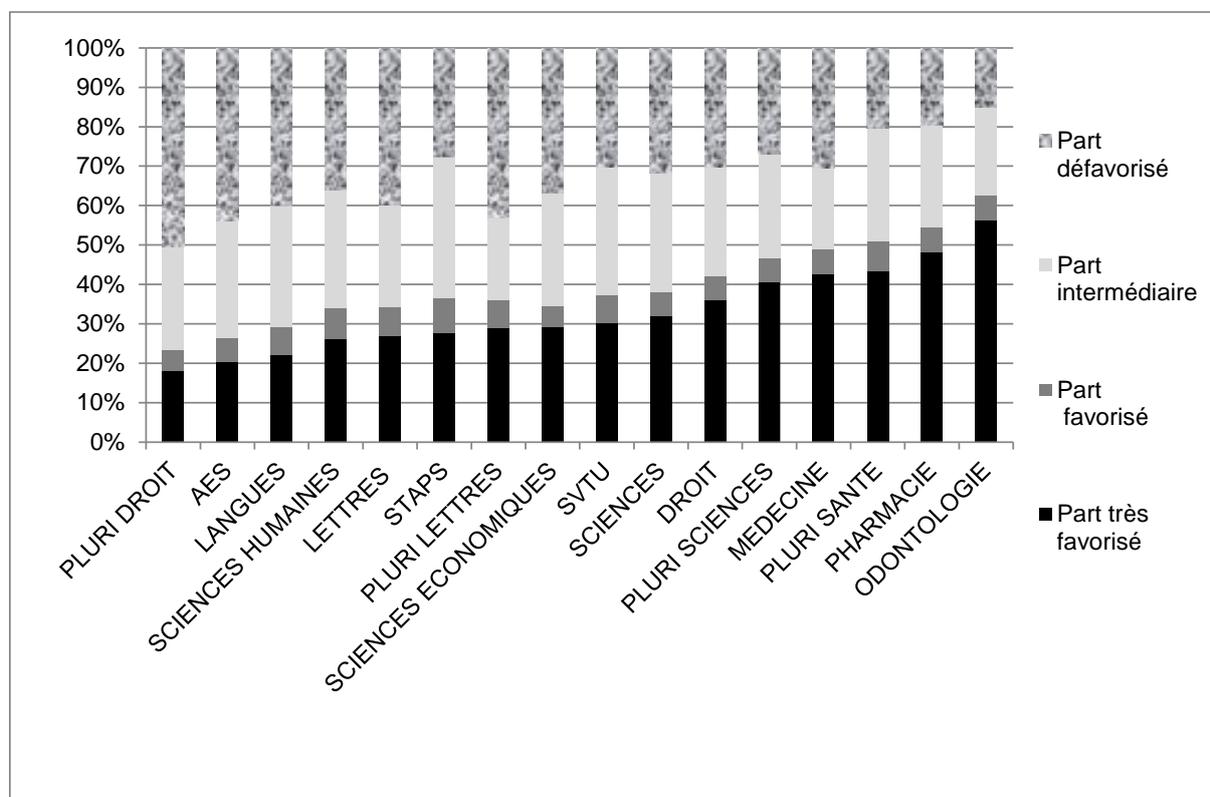
Les quatrième et septième points décrits ci-dessus pourraient sembler contradictoires avec l'autre constat mentionné dans le paragraphe précédent selon lequel le taux d'échec en Licence des étudiants provenant d'un milieu défavorisé est plus élevé que celui de la moyenne des étudiants (Brinbaum et al., 2018). En effet, cela pourrait inciter à penser qu'en Master la part des étudiants défavorisés diminue et que celle des étudiants très favorisés augmente. Or, la surreprésentation des étudiants défavorisés parmi ceux qui échouent en Licence et ne parviennent pas à obtenir le diplôme correspondant est compatible avec une hausse de la part occupée par ceux-ci entre la Licence et le Master dès lors qu'un grand nombre d'étudiants interrompent leurs études universitaires après la Licence. Avec une diminution moyenne de 20% des effectifs entre la Licence 3 et le Master 1 et, par ailleurs, l'arrivée d'étudiants en Master provenant de formation non universitaires, il est clair qu'une proportion significative d'étudiants interrompt ses études universitaires à la fin de la Licence, diplôme obtenu ou non, ce qui contribue à modifier en profondeur la répartition des groupes sociaux entre la Licence et le Master.

3.2 Origine sociale et discipline académique

Toujours au niveau national et pour la même population de référence, des éléments de cadrage supplémentaires concernant les disciplines d'inscription des étudiants par discipline nous semblent également nécessaires. Cela nous permet d'obtenir un premier éclairage général sur les différences de mixité sociale suivant les disciplines. La Figure 7 ci-dessous indique ainsi l'origine sociale des étudiants du système universitaire français en fonction des disciplines d'inscription sur la période 2013-2016⁵⁰ pour l'ensemble de l'Université.

⁵⁰ Les données de la période 2013-2016 ont été utilisées car les disciplines d'inscription sont identiques tout au long de cette période et l'ensemble des disciplines comprend des effectifs sur celle-ci, ce qui

Figure 7 : parts des quatre groupes sociaux pour l'ensemble du cycle universitaire (hors doctorat) suivant les disciplines (moyenne sur la période 2013-2016)



Champ : ensemble du système universitaire hors doctorat - France métropolitaine. Classement par parts de « très favorisés » en%.

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo), calculs de l'auteur.

Note de lecture : « Pluri-droit », AES, « Sciences Humaines », « Lettres », « STAPS », « Pluri-Lettres », « Sciences Economiques », « SVTU », « Sciences », « Droit », désignent respectivement les disciplines SISE suivantes « Pluri-droit-Sciences Eco-AES », « Administration Economiques et Sociales », « Sciences Humaines et sociales », « Lettres – Sciences du Langage - Arts », « Sciences et techniques des activités physiques et sportives », « Pluri-lettres – Langues – Sciences humaines », « Sciences Economiques – Gestion (hors AES) », « Sciences de la vie, de la Terre et de l'Univers », « Sciences Fondamentales et applications », « Droit – Sciences politiques ».

Pluri-Santé correspond uniquement à la première année commune aux études de santé (PACES). Il n'y a donc pas d'étudiants inscrits dans cette discipline SISE au-delà de la première année.

La Figure 7 ci-dessus permet de distinguer plusieurs catégories de disciplines académiques en fonction de la part occupée en moyenne par le groupe très favorisé

n'était pas le cas auparavant. La catégorisation en 16 disciplines sur cette période rendait ainsi possible le calcul d'une moyenne pluriannuelle. Comme indiqué dans le paragraphe 2.3, à compter de la rentrée 2010, les étudiants des filières médecine, odontologie et pharmacie sont d'abord inscrits dans la première année commune aux études de santé et seulement, à compter de la seconde année dans leur discipline respective pour ceux qui ont réussi l'examen de fin d'année.

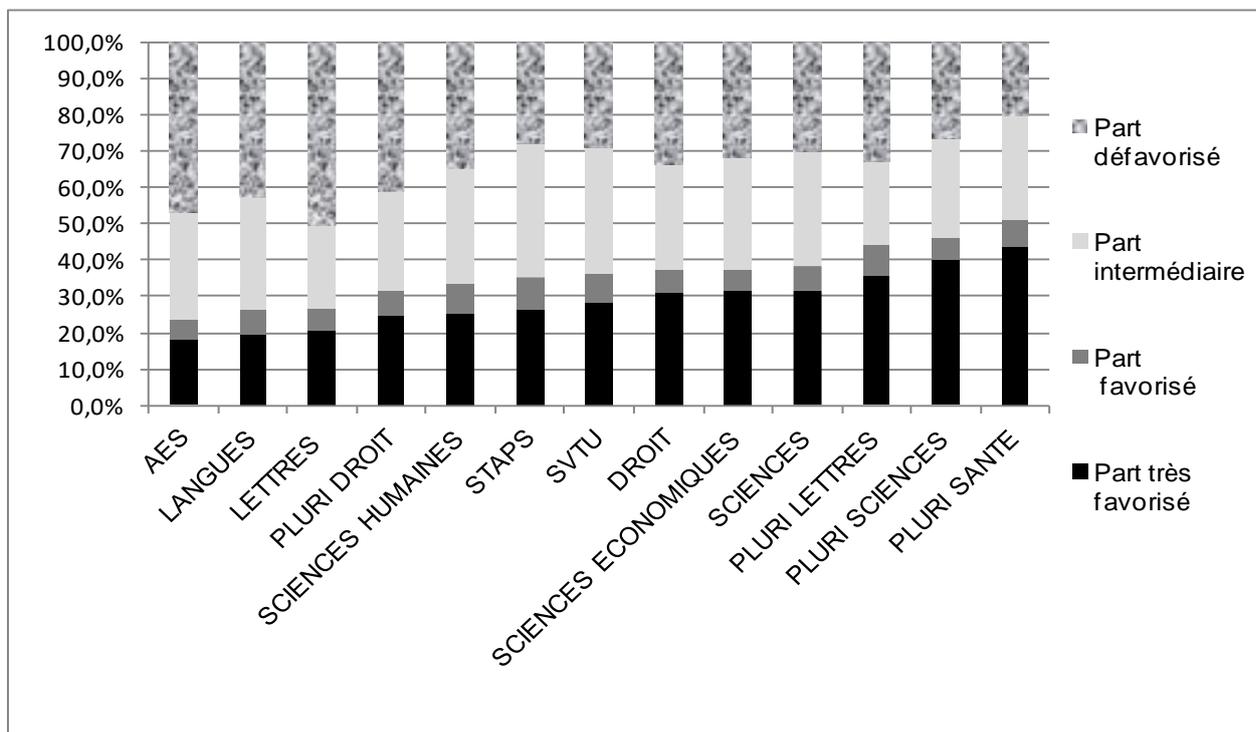
sur l'ensemble du cycle universitaire hors doctorat. Trois groupes peuvent ainsi être distingués sur l'ensemble de ce cycle⁵¹:

1. un groupe de disciplines où la part occupée par les très favorisés est supérieure ou égale à un tiers : odontologie, pharmacie, médecine, pluri-santé (qui correspond à la PACES), pluri-sciences, droit-sciences politiques. La part des étudiants appartenant au groupe défavorisé est inférieure à un tiers alors que ces étudiants sont en majorité à l'Université.
2. un groupe de disciplines intermédiaire où la part occupée par les très favorisés est comprise entre 20% et un tiers : « Sciences Fondamentales » et, « Sciences de la vie et de la Terre », « Sciences Economique-Gestion », « Lettres – Sciences du langage - Arts », « Sciences Humaines et Sociales », « Pluri Lettres – Langues - Sciences Humaines » et « Langues ».
3. un groupe de disciplines où la part occupée par les très favorisés est inférieure ou égale à 20% : « Administration Economique et Sociale » et « Pluri droit-Sciences Eco-AES ».

⁵¹ Les chiffres détaillés correspondant aux statistiques descriptives par discipline pour la première année de Licence et l'ensemble du cycle universitaire figurent en annexe 6. Comme indiqué précédemment dans le paragraphe 2.3, à compter de la rentrée 2010, les étudiants des filières médecine, odontologie et pharmacie sont d'abord inscrits dans la première année commune aux études de santé et seulement à compter de la seconde année dans leur discipline respective pour ceux qui ont réussi l'examen de fin d'année.

La Figure 8 ci-dessous indique l'origine sociale des étudiants du système universitaire français en fonction des disciplines d'inscription sur la période 2013-2016 pour la première année de Licence.

Figure 8 : parts des quatre groupes sociaux en première année de Licence suivant les disciplines (moyenne sur la période 2013-2016)⁵²



Champ : ensemble des étudiants de L1 du système universitaire - France métropolitaine. Les disciplines sont classées par parts de « très favorisés ».

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo), calculs de l'auteur.

Note de lecture: « Pluri-droit », AES, « Sciences Humaines », « Lettres », « STAPS », « Pluri-Lettres », « Sciences Economiques », « SVTU », « Sciences », « Droit », désignent respectivement les disciplines SISE suivantes « Pluri-droit-Sciences Eco-AES », « Administration Economiques et Sociales », « Sciences Humaines et sociales », « Lettres – Sciences du Langage - Arts », « Sciences et techniques des activités physiques et sportives », « Pluri-Lettres – Langues – Sciences Humaines », « Sciences Economiques – Gestion (hors AES) », « Sciences de la vie, de la Terre et de l'Univers », « Sciences Fondamentales et applications », « Droit – Sciences politiques ».

Pluri-Santé correspond uniquement à la première année commune aux études de santé. Il n'y a donc pas d'étudiants inscrits dans cette discipline SISE au-delà de la première année de Licence.

⁵² Suivant les informations communiquées par les gestionnaires de la base SISE en 2019, les étudiants qui apparaissent comme inscrits en Médecine, Odontologie et Pharmacie dès la première année sont inscrits dans des Licences professionnelles qui correspondent à des diplômes paramédicaux (herboristerie...). Il s'agit d'un nombre limité de cursus Licence qui ne durent qu'une année et pour lesquels les étudiants sont inscrits automatiquement en niveau 1. Dans la mesure où il s'agit de formations paramédicales qui sont en nombre limité et concernent environ 5 000 étudiants par an sur la période 2013-2016, nous ne les prenons pas en compte dans nos commentaires et sur la Figure 8 pour établir des comparaisons avec l'ensemble du cursus, même si la discipline SISE renseignée est identique. Toutefois, les statistiques relatives à ces étudiants sont disponibles en annexe 6.

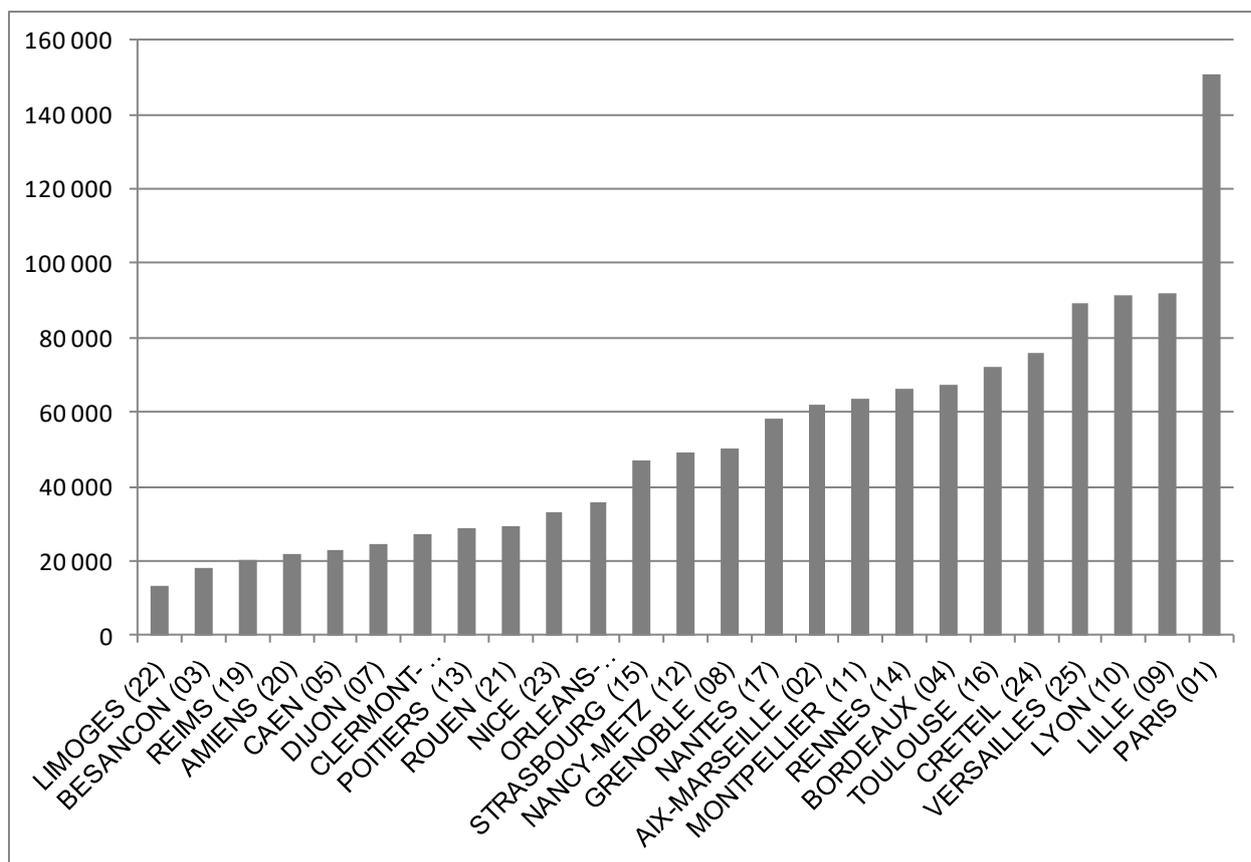
La Figure 8 ci-dessus montre que la part des étudiants très favorisés est élevée et supérieure à un tiers en première année de Licence dans les disciplines « Pluri-santé » (43,5%), « Pluri-Sciences » (39,9%) et « Pluri-Lettres » (35,9%) alors que celle-ci est inférieure à 20% en AES et en langues, les autres étant dans une situation intermédiaire. Toutefois, la Figure 8 met encore davantage en évidence la concentration des étudiants défavorisés dans certaines disciplines. Ainsi celle-ci est supérieure à un tiers dans les disciplines suivantes (hors études de santé) « Droit-Sciences Politiques » (33,6%), AES (46,9%), « Lettres – Sciences du Langage – Arts » (50,7%), Langues (42,9%), Sciences Humaines et Sociales (34,7%), « Pluri Droit-Sciences Eco – AES » (41,5%). Selon nous, cette concentration des étudiants défavorisés dans certaines disciplines en L1 peut être reliée à certains constats effectués dans les travaux de Frouillou (2016). En effet, comme nous l’avons expliqué dans la revue générale de la littérature du chapitre I, sur la base d’un nombre important d’entretiens qualitatifs, Frouillou (2016) insiste sur l’importance prépondérante des effets de pairs pour les étudiants qui ont un capital scolaire limité et sont issus principalement du groupe social défavorisé. Ceux-ci ont davantage tendance que les étudiants plus favorisés à se laisser influencer par leurs pairs au moment de sélectionner une Université ainsi qu’une discipline d’inscription. Comme nous l’avons expliqué ci-dessus dans le paragraphe 4.2 de ce chapitre, ce phénomène de concentration des étudiants défavorisés a tendance à diminuer après la L1. Leur part est ainsi plus faible sur l’ensemble du cursus universitaire qu’au niveau de la L1 dans 4 des 6 des disciplines mentionnées antérieurement : « Droit-Sciences Politiques » (30,2%), AES (44%), « Lettres – Sciences du Langage – Arts » (40%), Langues (40,1%). Ces disciplines accueillent ainsi un nombre important d’étudiants défavorisés en L1, mais ceux-ci en sont davantage écartés que les autres étudiants dans la suite du cursus⁵³.

⁵³ En revanche, la part des étudiants défavorisés remonte légèrement en Sciences Humaines et Sociales (36,1%) et de façon marquée dans la discipline « Pluri Droit-Sciences Eco – AES » (50,4%), mais cette dernière ne comprend que 0,1% du total des étudiants.

3.3 Les effectifs des académies sont très variables

Au-delà des moyennes observées au niveau national, un premier cadrage géographique semble nécessaire. A cette fin, nous avons choisi comme unité d'analyse géographique l'« académie », qui est la circonscription administrative de référence de l'Éducation nationale et du MESRI : il s'agit d'une unité d'organisation opérationnelle et de décision politique essentielle. Afin de connaître le poids relatif de chaque académie dans l'ensemble du système universitaire, la Figure 9 ci-dessous indique le nombre d'étudiants inscrits dans chacune d'entre elles en moyenne sur la période 2006-2016⁵⁴.

Figure 9 : effectifs par académies (moyenne annuelle sur la période 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
 Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur.

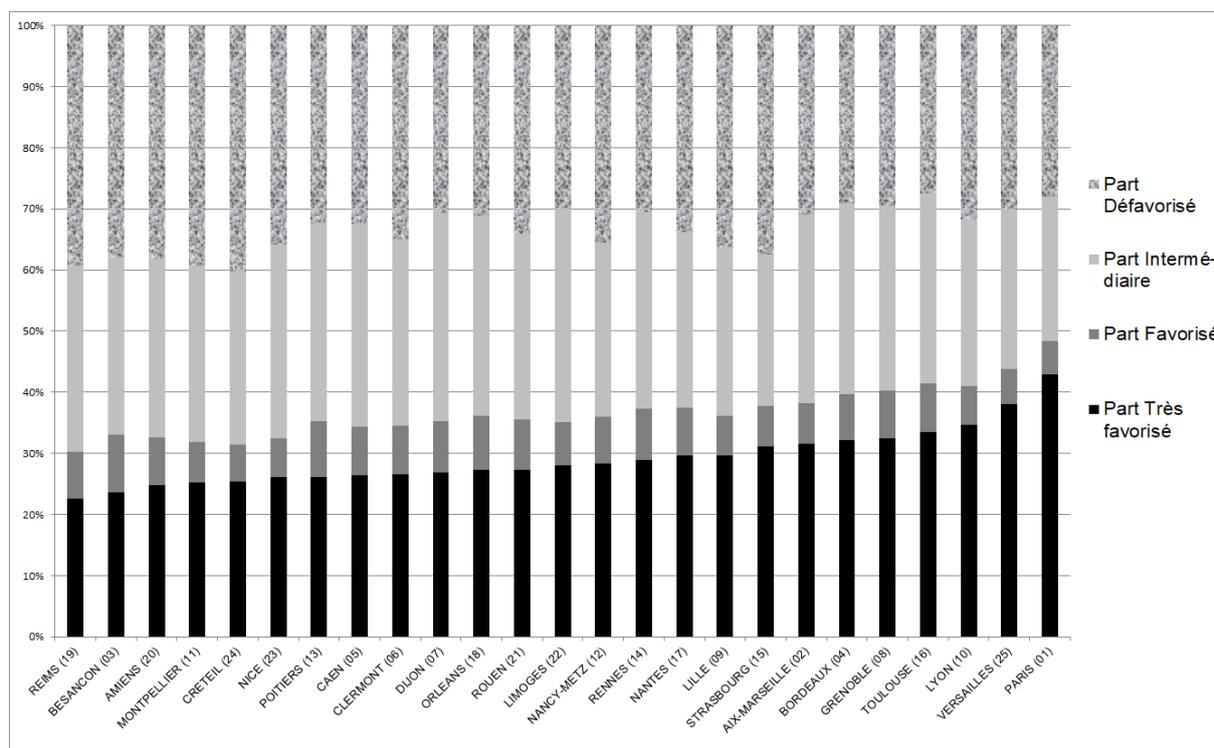
⁵⁴ Les tableaux disponibles en annexe 6 présentent les effectifs par académie pour chaque année sur la période 2006-2016.

La Figure 9 permet de constater la variabilité des effectifs inter-académies, avec par exemple, en 2016, 157 846 étudiants pour la plus grande (Paris) et 14 381 pour la plus petite (Limoges). Par ailleurs, il convient de noter que les trois académies d’Ile-de-France, à savoir les académies de Paris, Versailles et Créteil, représentent réunies 18,6% du total des étudiants de France métropolitaine en 2006 et 22,8% en 2016.

3.4 La répartition des étudiants par origine sociale et par académie

Au-delà des effectifs des académies, il nous semble également nécessaire de connaître les parts occupées par les différents groupes sociaux dans les différentes académies afin d’évaluer s’il existe des différences géographiques significatives de ce point de vue. La Figure 10 ci-dessous indique ainsi la répartition des étudiants par origine sociale suivant les académies sur la période 2006-2016⁵⁵ :

Figure 10 : parts respectives des quatre groupes sociaux suivant les académies (moyenne annuelle sur la période 2006-2016, en%)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
 Classement des académies de manière croissante suivant la part des très favorisés.
 Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l’auteur.

⁵⁵ Les chiffres correspondants sont disponibles dans l’annexe 6.

La Figure 10 ci-dessus indique qu'il existe des différences importantes entre académies suivant les parts respectives occupées par les étudiants issus des différents groupes sociaux avec, par exemple, sur la période 2006-2016, un maximum de 42,9% d'étudiants très favorisés dans l'académie de Paris et un minimum de 22,6% dans celle de Reims alors que pour les défavorisés le maximum est atteint dans l'académie de Créteil avec 40,3% et le minimum dans l'académie de Toulouse avec 27,4%. Il est possible de distinguer trois groupes d'académies suivant les effectifs et la répartition des étudiants par origine sociale :

- 1) Tout d'abord, les académies franciliennes (Paris, Versailles, Créteil) qui comprennent un nombre élevé d'étudiants. Il convient cependant de distinguer les académies de Paris et Versailles, où la part des étudiants très favorisés est élevée, de celle de Créteil qui, à l'inverse, a la part d'étudiants défavorisés la plus importante au niveau national (40,3%).
- 2) Ensuite, les grandes académies de Province, comme Lyon, Toulouse, Aix-Marseille, Grenoble, Rennes, Montpellier, Lille, Bordeaux et Nantes, dont l'effectif annuel moyen est supérieur à 50 000 étudiants. Pour celles-ci, la part occupée par les étudiants très favorisés est supérieure ou proche de 30%, sauf à Montpellier où elle est de 25,3%.
- 3) Enfin, les petites académies de Province dont l'effectif annuel moyen est inférieur à 50 000 étudiants (Strasbourg, Nancy-Metz, Limoges, Orléans-Tours, Rouen, Dijon, Clermont-Ferrand, Caen, Poitiers, Nice, Amiens, Besançon, Reims). Au sein de ces académies, la part occupée par les étudiants très favorisés est inférieure ou égale à 30%, sauf dans le cas de Strasbourg (31,1%).

Les écarts dans les parts occupées par chacun des groupes sociaux dans les différentes académies amènent deux types de commentaires distincts :

- Dans la mesure où le périmètre de notre analyse est constitué par l'ensemble de la population des étudiants de France métropolitaine, les différences inter-académiques dans les parts occupées par chacun des groupes sociaux

constituent déjà d'un point de vue statistique un facteur de ségrégation sociale. Cela reflète la part de ségrégation entre les formations qui résulte des différences inter-académiques. En effet, d'un point de vue statistique, un étudiant de Licence tiré au hasard aurait une probabilité plus élevée d'être dans une formation où les étudiants défavorisés sont en majorité à Créteil ou Montpellier qu'à Paris.

- Si l'objet de notre travail (la population de référence) était constitué, en revanche d'une ou plusieurs académies en particulier, alors il serait possible de considérer qu'il s'agit uniquement d'une description de la répartition des populations d'étudiants par origine sociale qui ne fournit pas d'information sur l'existence éventuelle de phénomènes de ségrégation. Une part élevée d'étudiants très favorisés au sein d'une académie est, par exemple, parfaitement compatible avec une répartition relativement aléatoire et homogène de ceux-ci entre les différentes formations proposées ou, au contraire, avec une concentration de ceux-ci au sein de formations ségréguées. Ces données statistiques ne fournissent pas d'indication, en effet, sur le degré de ségrégation au sein de chaque académie ou la propension des différents groupes à « pratiquer » la mixité sociale.

4 Au niveau national, des phénomènes de ségrégation réels mais plutôt stables et qui ont tendance à s'estomper suivant les années d'études

Comme expliqué dans chapitre I, les statistiques descriptives ne constituent qu'une première étape pour comprendre les phénomènes de mixité sociale, le recours au calcul d'indices synthétiques de mesure de la ségrégation étant indispensable pour appréhender ce phénomène de manière plus complète. A cette fin, dans la présente section, nous mesurons au niveau national les niveaux de ségrégation à l'Université avec l'indice d'exposition normalisé (IEN) entre 2006 et 2016. Comme expliqué dans le chapitre I, si l'indice d'Atkinson est le seul à permettre d'établir des comparaisons dans le temps rigoureuse d'un point de vue mathématique, le recours à un indice d'exposition normalisé peut, toutefois, se justifier pour plusieurs raisons : celui-ci est quasiment invariant à la composition suivant les estimations réalisées par Reardon et O'Sullivan (2004) et son interprétation est davantage intuitive que pour l'indice d'Atkinson⁵⁶. Cette section est structurée en deux parties, le paragraphe 4.1 décrivant l'évolution moyenne de la ségrégation au niveau national sur la période retenue et le paragraphe 4.2 expliquant comment celle-ci évolue en moyenne suivant les années d'études.

4.1 Au niveau national, une ségrégation plus faible que dans le secondaire et relativement stable dans le temps

a) Résultats d'ensemble

Afin d'apprécier les niveaux de ségrégation sociale au niveau national, nous avons calculé l'IEN uniquement pour les deux groupes qui se situent aux extrêmes de la distribution en quatre groupes définie par la DEPP : les très favorisés et les défavorisés. Dans le cadre de notre travail sur la mixité sociale, comme le montre les développements des paragraphes suivants, le rôle joué par ces deux groupes est, en effet, essentiel, mais distinct. Les étudiants très favorisés sont ainsi surreprésentés dans certaines filières sélectives du système universitaire : comme décrit dans la

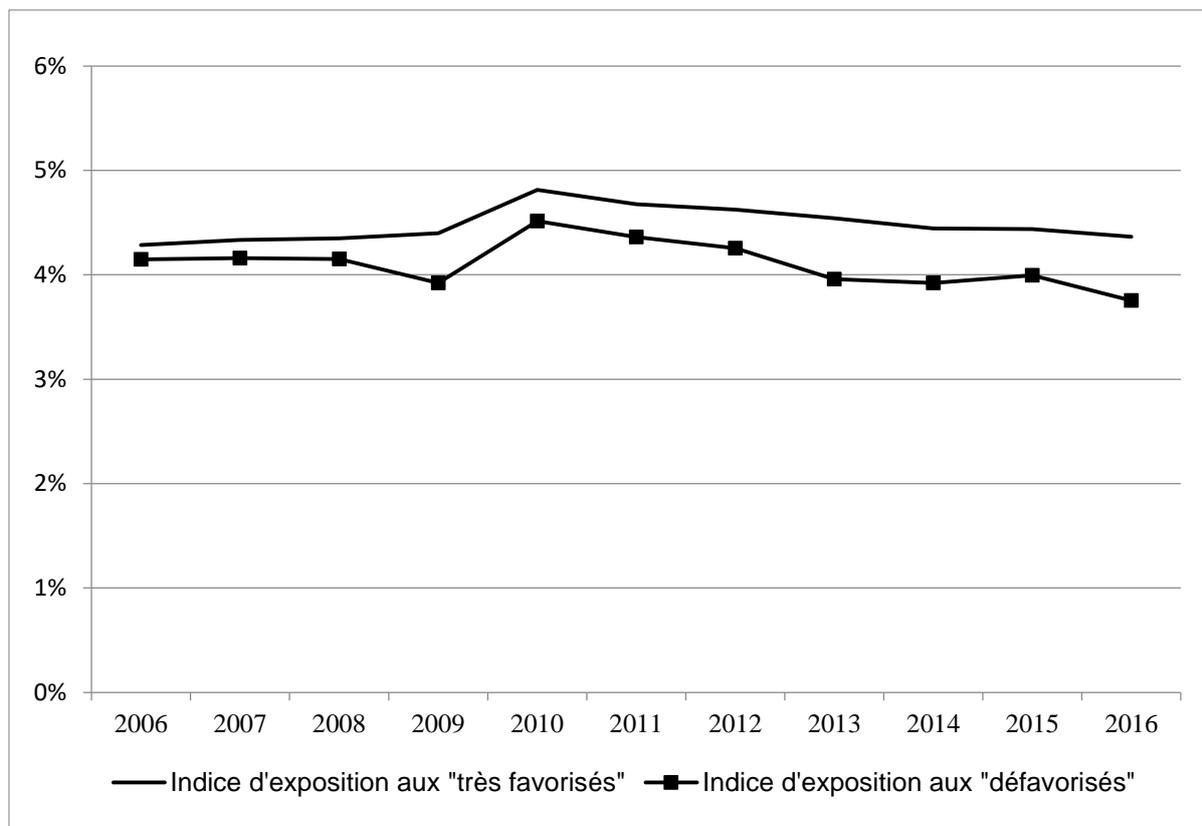
⁵⁶ L'ensemble des résultats du calcul des 8 indices présentés dans le chapitre I est disponible en annexe 8. Les valeurs de l'indice d'information mutuelle et d'entropie normalisé confirment en particulier les analyses menées sur la base de l'indice d'Atkinson et d'exposition normalisé dans le cadre des chapitres I et II.

section 3, dans la filière pluri-santé (qui correspond à la PACES), ils représentent 43,4% des étudiants sur la période 2013-2016, ce taux étant encore plus élevé pour les étudiants qui réussissent cette première année et accèdent aux différentes filières médicales. Une structure en deux groupes (qui distinguerait par exemple les étudiants défavorisés du reste de la population) ne permettrait pas, toutefois, d'analyser de manière spécifique et séparée la situation des étudiants très favorisés et défavorisés⁵⁷. Ainsi, nous conservons dans notre analyse une structure en quatre groupes, même si les deux groupes intermédiaires (« favorisé » et « intermédiaire ») ne sont pas au centre de notre analyse.

La Figure 11 montre que l'IEN des très favorisés et des défavorisés est positif : la ségrégation est bien présente, c'est-à-dire que le processus de répartition des étudiants entre formations est affecté par les origines sociales. Les IEN aux très favorisés et aux défavorisés sont, en moyenne, relativement stables : entre 2006 et 2016, ils sont compris entre 3,5% et 5%.

⁵⁷ Nous comprenons qu'au moins trois groupes sociaux doivent être définis afin de pouvoir calculer et utiliser de manière opérationnelle des IEN aux très favorisés et aux défavorisés.

Figure 11 : Evolution de l'indice d'exposition normalisé aux très favorisés et défavorisés en% (2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
 Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.
 Lecture : en 2006, l'IEN aux très favorisés était de 4,3%.

En 2016, l'IEN aux très favorisés est de 4,3%. Or, cette même année, il y a en moyenne 30,3% d'étudiants appartenant à ce groupe social pour l'ensemble de l'Université. Si l'on normalise la taille moyenne d'une formation à 100 (avec donc en moyenne 30 étudiants très favorisés), cela signifie qu'un étudiant lui-même très favorisé aura en moyenne dans sa formation environ 4,3 étudiants très favorisés de plus qu'un étudiant issu d'un autre groupe. Pour l'IEN aux défavorisés, le niveau de ségrégation est de 4,1%. Au final, ceci indique qu'en moyenne, les étudiants très favorisés comme les étudiants défavorisés sont regroupés dans des formations où ils côtoient relativement plus d'étudiants issus de leur propre milieu social.

L'existence de niveaux de ségrégation significatifs et plutôt stables (avec une légère tendance à la baisse à compter de 2010 que l'on peut observer sur la Figure 11) est cohérente avec les résultats obtenus dans la section 4 du chapitre I au moyen de

l'indice d'Atkinson⁵⁸. Il convient de rappeler que les valeurs de l'indice d'entropie normalisé calculées par Avouac et Harari-Kermadec (2021) diminuent sur la période 2007-2015. Toutefois, les données utilisées au sein de la base SISE ne sont pas identiques dans le travail mentionné : elles incluent, en particulier certains établissements non universitaires ainsi que des données relatives à des territoires ultramarins. De plus, l'unité retenue pour mesurer la ségrégation dans ce travail est l'Université et non la formation, contrairement à ce qui est effectué dans notre analyse.

b) Comparaisons avec le Collège

Les travaux portant sur le Collège aboutissent généralement à des résultats très marqués en termes de ségrégation. Ly et Riegert (2015) analysent la ségrégation sociale et scolaire dans les Collèges et Lycées sur l'ensemble du territoire français. L'IEN aux très favorisés passe d'une valeur légèrement supérieure à 15% en sixième à près de 20% en terminale. Ces niveaux de ségrégation sont donc nettement supérieurs à ceux que nous mettons en évidence à l'Université. Courtioux et Maury (2018) analysent, sur la période 2004-2014, le degré de mixité sociale des Collèges selon le secteur (privé, public hors ZEP et ZEP). L'entre soi des élèves très favorisés est particulièrement marqué dans les établissements privés et certains établissements publics hors ZEP. Les auteurs obtiennent un IEN aux très favorisés compris entre 15 et 17%, soit, là encore, des valeurs bien supérieures à celles obtenues pour les étudiants inscrits à l'Université.

Les niveaux de ségrégation à l'Université sont donc nettement inférieurs à ceux observés dans le secondaire. Cette baisse statistique des niveaux de ségrégation entre le secondaire et l'Université peut se comprendre à la lumière des trois facteurs suivants :

- 1) *Le nombre d'Universités est limité en comparaison du nombre d'établissements d'enseignement secondaire, ce qui réduit à la fois l'impact de la ségrégation résidentielle et la possibilité de mettre en œuvre des stratégies d'évitement au sein du système universitaire.* Ainsi, à titre

⁵⁸ L'ensemble des résultats du calcul des 8 indices présentés dans le chapitre I est disponible en annexe 8.

d'exemple, l'académie de Paris comptait, à la rentrée 2021, 177 Collèges et 166 Lycées, mais seulement 6 Universités et 2 IUT. Etant donné le nombre limité d'établissements universitaires, les étudiants résident en moyenne plus loin de leur établissement que dans le secondaire et ces derniers reflètent ainsi moins la ségrégation résidentielle. De plus, comme décrit dans le chapitre I, au niveau de l'enseignement secondaire, en raison du nombre élevé d'établissements, il est fréquent et relativement aisé, en particulier dans les grandes agglomérations, de mettre en place des « stratégies d'évitement », en contournant la carte scolaire. En revanche, pour les étudiants qui font le choix de s'inscrire à l'Université, ces possibilités sont plus limitées, étant donné le nombre restreint d'établissements universitaires. A titre d'exemple, en Ile-de-France, seules 7 Universités offraient à la rentrée 2017 la possibilité de suivre la PACES⁵⁹. Pour un étudiant francilien qui choisit de suivre ce cursus, les possibilités de choix d'un établissement, en particulier si l'on tient compte des contraintes de transport, sont plus limitées que dans l'enseignement secondaire. De notre point de vue, ce facteur contribue à renforcer le degré de mixité sociale.

- 2) *Les établissements universitaires reflètent moins la ségrégation résidentielle que dans l'enseignement secondaire en raison également de la mobilité inter-académique des élèves à la fin des études secondaires.* Ainsi, en 2006, au niveau national, 41,6% des étudiants avaient obtenu leur bac dans une autre académie que celle où ils sont inscrits. Ce taux atteignait un maximum de 77,1% dans l'académie de Paris et un minimum de 28,1% dans l'académie d'Amiens. Or, comme nous l'avons décrit dans la section 2 du chapitre I, les niveaux de ségrégation au Collège sont élevés dans une grande mesure en raison du principe de sectorisation scolaire (Monso et al., 2018 et Boutchenik et al., 2020, montrent que la sectorisation contribue de manière significative à la ségrégation sociale dans les Collèges français). De même, au niveau du Lycée, le recrutement des établissements se fait principalement sur la base de critères géographiques. La mobilité inter-académique des étudiants à la sortie de l'enseignement secondaire est ainsi

⁵⁹ Paris 5, Paris 6, Paris 7, Paris 11, Paris 12, Paris 13 et Paris Ouest-UVSQ.

un second facteur de mixité sociale et de réduction de la ségrégation résidentielle.

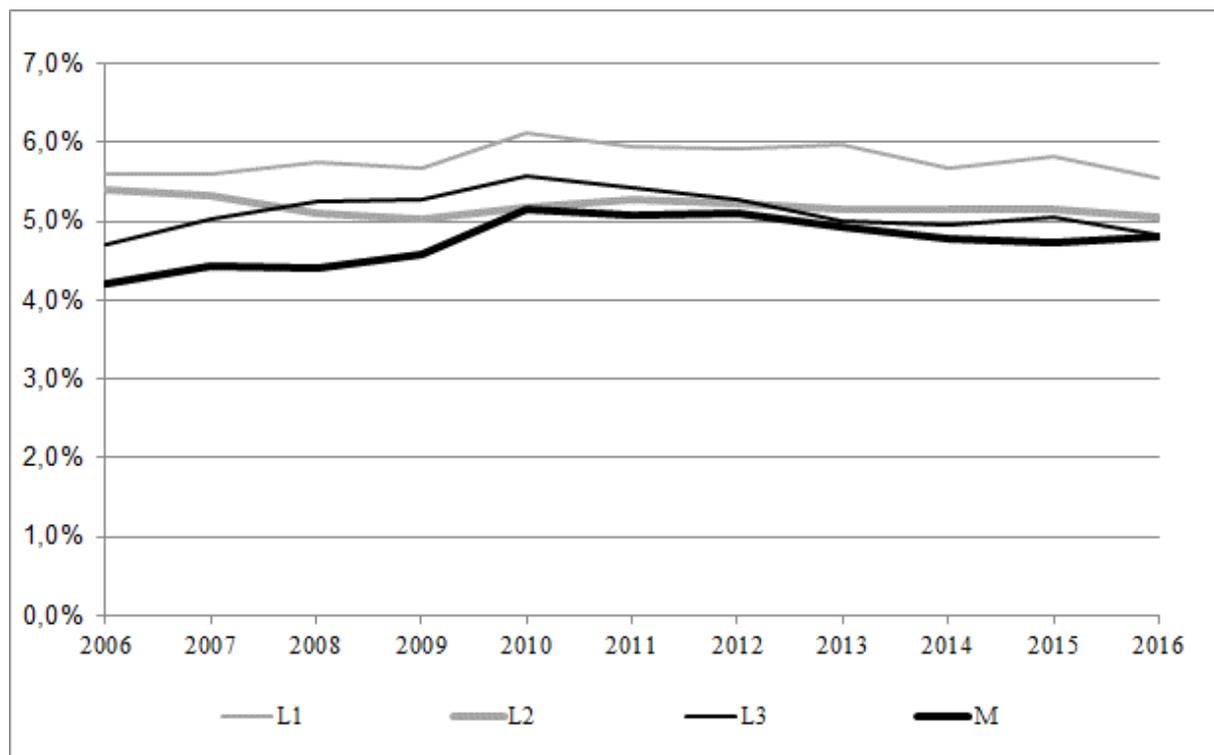
- 3) *Les stratégies d'évitement sont en partie déplacées à l'extérieur du système universitaire via l'enseignement supérieur non universitaire.* Comme indiqué dans notre introduction générale, une partie des étudiants peut chercher à mettre en place des stratégies d'évitement destinées à contourner l'Université, en particulier au niveau du premier cycle, via l'enseignement supérieur non universitaire. Dans la lignée de ce qui est constaté par Boutchenik et al. (2018) pour l'enseignement primaire et secondaire, ces stratégies d'évitement pourraient être utilisées plus particulièrement au sein des milieux sociaux les plus favorisés, en particulier au niveau de la Licence. Comme le soulignent Fack et Huillery (2021), qui reprennent les conclusions des travaux de Bonneau et Grobon (2021), l'accès aux filières sélectives de l'enseignement supérieur est particulièrement inégalitaire. A titre d'exemple, pour Sciences Po Paris et Paris Dauphine (qui ont tous les deux le statut de Grand Etablissement), sur la période 2010-2016, la part des étudiants très favorisés est de 63%, alors que celle des défavorisés est de 10,7% seulement, soit des niveaux bien plus élevés que ceux constatés dans l'académie de Paris et a fortiori sur l'ensemble du territoire métropolitain⁶⁰. Une partie des phénomènes de ségrégation a ainsi lieu entre le système universitaire et le système non universitaire : pour une partie des futurs étudiants, c'est l'Université elle-même qui est contournée, évitée. Dans la mesure où une partie des stratégies d'évitement sont mises en place non pas au sein de l'Université mais à l'extérieur de celle-ci, cela réduit les niveaux de ségrégation à l'intérieur du système universitaire.

⁶⁰Comme décrit dans l'introduction générale, nous ne disposons pas de statistiques descriptives sur les origines sociales de l'ensemble des étudiants du système non universitaire : des données sont disponibles dans la base SISE Université que pour un nombre très restreint d'établissements non universitaires.

4.2 Au niveau national, la ségrégation a plutôt tendance à s'estomper avec le niveau d'études

Les Figures 12 et 13 ci-dessous représentent l'évolution de l'IEN aux très favorisés et aux défavorisés, sur la période 2006-2016, suivant le niveau d'études⁶¹.

Figure 12: indice d'exposition normalisé aux très favorisés selon le niveau d'études en % (2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

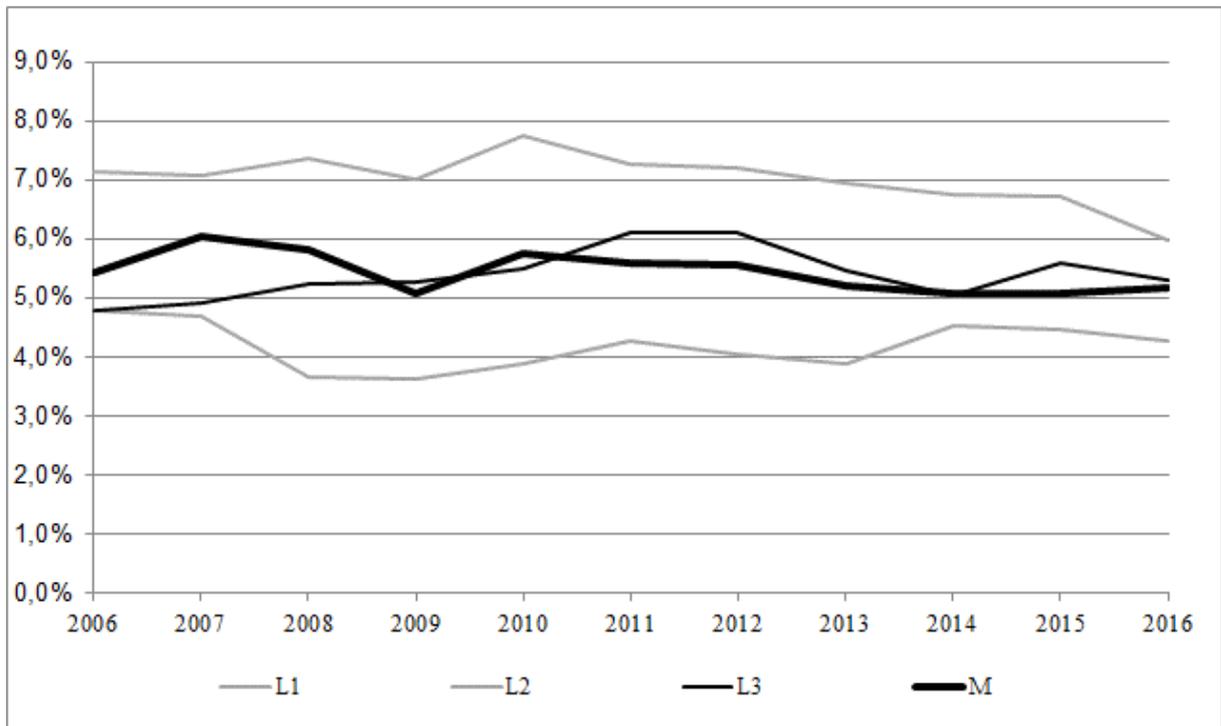
Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Note : L1 pour Licence 1, L2 pour Licence 2, L3 pour Licence 3, M pour Master 1 et 2.

Lecture : en 2006, l'IEN aux très favorisés en L1 était de près de 5,6%.

⁶¹ La population de référence permettant de calculer les indices est alors l'ensemble des étudiants inscrits dans une année d'étude particulière, par exemple les étudiants de L2. L'unité d'observation n'est alors plus la formation définie comme le croisement d'une discipline et d'un établissement universitaire, mais le croisement d'une discipline, d'un établissement universitaire et d'un niveau d'études (d'une année d'inscription).

Figure 13: indice d'exposition normalisé aux défavorisés selon le niveau d'études en % (2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Note : L1 pour Licence 1, L2 pour Licence 2, L3 pour Licence 3, M pour Master 1 et 2.

Lecture : en 2006, l'IEN aux très favorisés en L1 était de près de 7,1%.

a) Constats d'ensemble

Trois principaux constats peuvent être établis à partir des deux figures ci-dessus :

- La Figure 12 montre que globalement⁶², la ségrégation des très favorisés diminue avec le niveau d'étude. Elle diminue entre L1 et L2 et baisse encore en Master. Il convient de souligner, toutefois, que ces évolutions restent très modestes et que la ségrégation reste faible quel que soit le niveau d'études considéré.
- Le phénomène est un peu plus marqué pour la ségrégation des défavorisés. La Figure 13 montre que sur l'ensemble de la période, l'IEN diminue plus fortement entre la L1 et la L2 : il passe en moyenne de 7% à 4,2%. Toutefois, après le

⁶²A l'exception des années 2009, 2010 et 2011, où les niveaux de ségrégation observés en L3 sont supérieurs à ceux de L2.

passage de la L1 à la L2, la ségrégation des étudiants défavorisés repart légèrement à la hausse. En L3, la valeur moyenne de l'indice augmente de nouveau pour atteindre 5,4% et reste à cette valeur en moyenne sur la période en Master.

- Lorsque l'on compare l'évolution de la ségrégation des défavorisés et celle des très favorisés par niveau d'étude, la première année de Licence apparaît particulièrement contrastée. En 2016, en dépit d'une nette baisse des niveaux de ségrégation des défavorisés en L1 en l'espace de 10 ans, ceux-ci restent supérieurs aux niveaux observés pour les très favorisés.

Ces résultats sont à nouveau cohérents avec ceux qui ont été présentés dans la section 4 du chapitre I sur la base de l'indice d'Atkinson.

- b) Facteurs permettant d'expliquer la tendance à la baisse de la ségrégation avec le niveau d'études et notamment la baisse marquée de celle-ci à compter de la L2 pour les défavorisés

Afin de comprendre la tendance à la baisse des niveaux de ségrégation des étudiants au fur et à mesure de l'avancement dans les études, plusieurs facteurs peuvent être mobilisés :

- Comme l'a établi Frouillou (2016), les choix d'inscription en première année des étudiants défavorisés sont fortement influencés par les effets de pairs, ce qui favorise la concentration de ces étudiants dans certaines formations en L1 où ils sont surreprésentés, comme nous l'avons détaillé dans le paragraphe 3.2 ci-dessus. Plus généralement, l'origine sociale serait un facteur important à l'entrée de l'Université qui influence à la fois le choix de l'établissement, le type de filière suivie ainsi que la discipline d'inscription. Ces effets de pairs sont aussi susceptibles de jouer un rôle significatif pour la ségrégation des étudiants issus des autres groupes sociaux.

- La part occupée par les étudiants défavorisés passe en moyenne sur la période 2006-2016 de 33,2% à 27,4% entre la L1 et la L2. Cette baisse marquée peut être reliée au taux d'échec en Licence des étudiants provenant d'un milieu défavorisé qui est plus élevé que celui de la moyenne des étudiants, ce taux étant particulièrement élevé en première année de Licence (Brinbaum et al., 2018). A compter de la L2, la surreprésentation des étudiants défavorisés dans les filières où ils ont tendance à se concentrer en L1 diminue. Comme indiqué dans la section 3 ci-dessus, cela concerne notamment des disciplines comme le « Droit et les sciences politiques », l'AES ou encore les « Lettres-Sciences du Langage-Arts » où la part des étudiants défavorisés est respectivement de 33,6%, 46,9% et 50,7% en L1, mais de 30,2%, 44% et 40% pour l'ensemble de la Licence. La diminution de la part de ces étudiants après la L1 peut ainsi être reliée à des phénomènes d'abandon à l'issue de la première année de Licence, mais aussi potentiellement à leur réorientation vers d'autres filières moins ségréguées⁶³. Ces deux éléments permettent de comprendre la baisse de l'IEN des défavorisés entre la L1 et la L2.
- Entre la L2 et la L3, l'IEN des très favorisés est plutôt stable alors que celui des défavorisés remonte. Ce constat peut être relié pour les défavorisés à la hausse marquée de la part qu'ils occupent, celle-ci passant de 27,4% en L2 à 33,2% en L3, soit un niveau identique à la L1. Une hypothèse possible pour comprendre cette hausse est à nouveau le taux d'échec en Licence des étudiants provenant d'un milieu défavorisé. En raison des difficultés à obtenir le diplôme de Licence à l'issue de la troisième année et du redoublement qui les affecte plus particulièrement, la part des étudiants défavorisés revient au même niveau qu'en L1 et leur surreprésentation dans certaines formations repart à la hausse.
- Malgré des parts occupées identiques en L1 et en L3 pour les défavorisés, leur niveau de ségrégation est nettement inférieur en L3. Une hypothèse possible

⁶³ Dans le cadre de travaux ultérieurs, il nous semble envisageable d'approfondir cette analyse en estimant de manière fine et détaillée les taux de passage, d'abandon et de réorientation en fonction des disciplines et des origines sociales. A cette fin, il conviendrait de mobiliser d'autres bases de données telles que, par exemple, SISE-Résultats ; ces informations ne sont pas disponibles, en effet, dans la base SISE-Inscrits.

pour interpréter ce moindre niveau est le phénomène des réorientations entre formations après la L1 et de réduction du rôle joué par les effets de pairs qui permet de réduire la concentration des étudiants défavorisés dans certaines d'entre elles.

- Entre la L3 et le Master, la ségrégation des très favorisés diminue légèrement alors que celle des défavorisés est plutôt stable. Ceci indique que de manière globale, le rôle joué par l'origine sociale diminue légèrement pour les très favorisés à l'issue de la L3 alors que celui-ci reste plutôt constant pour les étudiants défavorisés. Il convient, par ailleurs, de noter que cette stabilité d'ensemble est bien entendu compatible avec des phénomènes localisés (par exemple au niveau des académies) de hausse ou de baisse de la ségrégation.

Afin d'affiner notre interprétation et de mieux comprendre ces résultats d'ensemble, une solution envisageable consiste à introduire une dimension géographique, en calculant des IEN par zone géographique pour l'ensemble du cycle universitaire, mais aussi par niveau d'études.

5 Des disparités géographiques importantes entre les académies et suivant les niveaux d'études

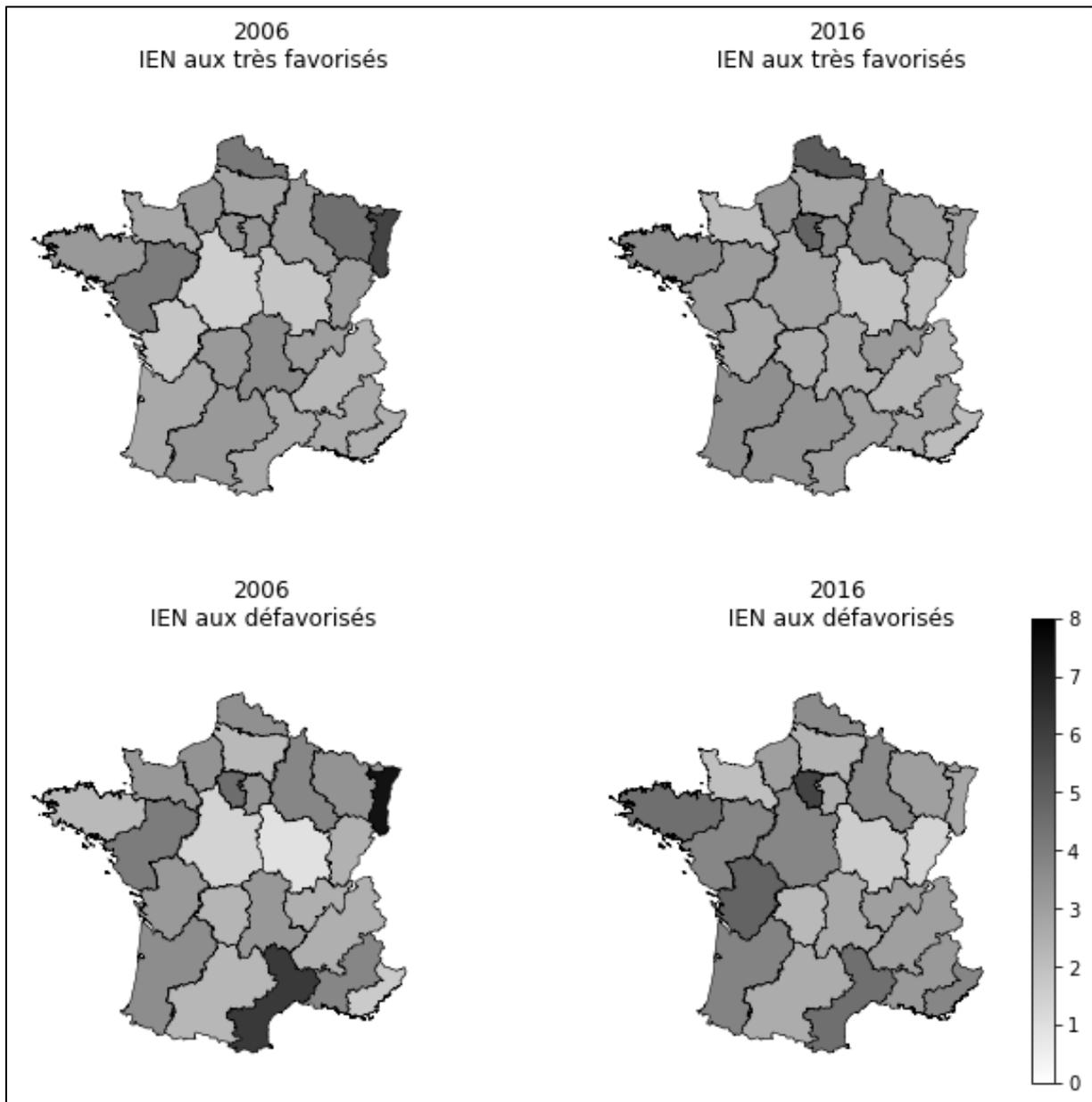
Afin de nuancer les résultats obtenus au niveau national, nous affinons géographiquement notre analyse et comparons les niveaux de ségrégation entre zones géographiques. En raison du rôle essentiel des académies comme circonscription administrative de référence pour le MESRI, nous mesurons les niveaux de ségrégation existants dans les différentes académies de France métropolitaine. La section 5 est ainsi structurée en deux parties, le paragraphe 5.1 faisant le constat d'une forte hétérogénéité suivant les académies alors que le paragraphe 5.2 analyse les niveaux de ségrégation suivant les académies, mais aussi suivant les années d'études.

5.1 Une forte hétérogénéité des niveaux de ségrégation selon les académies

La Figure 14 ci-dessous présente les niveaux de ségrégation par académie pour les années 2006 et 2016 pour les 2 indices (IEN aux très favorisés et aux défavorisés). Sur celle-ci, les valeurs de l'IEN sont représentées par l'intensité de gris⁶⁴.

⁶⁴ La population de référence permettant de calculer les indices est alors l'ensemble des étudiants inscrits au sein d'une académie particulière, par exemple l'ensemble des étudiants de l'académie de Créteil, l'unité d'observation demeurant la formation.

Figure 14: indices d'exposition normalisés (en%) selon les académies en 2006 et en 2016



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Note : IEN pour indice d'exposition normalisé ; la légende de droite indique la valeur de l'IEN en pourcentage en fonction de l'intensité de gris ; pour des questions d'échelle, l'académie de Paris n'est pas visible sur la carte.

Au premier abord, une forte hétérogénéité se profile sur la Figure 14 et il n'existe pas de schéma ou de tendance facilement identifiable. En particulier, il ne semble pas y avoir de corrélation spatiale évidente au niveau académique ni entre l'IEN aux très favorisés et l'IEN aux défavorisés, ni du point de vue de l'évolution de ces indices sur la période. En 2006, l'IEN pour les très favorisés va de 1,1% pour l'académie d'Orléans-Tours à 5,6% pour l'académie de Strasbourg, soit du simple au quadruple.

L'ampleur de l'hétérogénéité spatiale est presque comparable à celle observée dans le secondaire où, selon Ly et Riegert (2015), la ségrégation varie d'environ 6 ou 7% (pour Besançon) à plus de 25% à Paris. Ainsi, même si la ségrégation est en moyenne plus faible à l'Université que dans le secondaire, les écarts inter-académiques sont toujours très importants.

Par ailleurs, Ly et Riegert (2015) montrent que les niveaux de ségrégation des très favorisés dans le secondaire sont nettement plus élevés dans les académies franciliennes qu'en Province. Au contraire, nous montrons pour l'Université que les niveaux de ségrégation existant en Île-de-France en 2006 ne sont pas systématiquement supérieurs à ceux observés dans le reste du pays. Les niveaux de ségrégation à Paris (3,5%), Versailles (3%) et Créteil (2,9%) sont notamment inférieurs à ceux de Strasbourg, Nancy-Metz, Lille ou Nantes.

De même, comme montré dans la Figure 14 ci-dessus, la valeur de l'IEN des très favorisés varie entre un minimum de 1,8 % pour l'académie de Dijon et un maximum de 4,8% pour les académies de Versailles et de Lille en 2016. Là encore, nous n'observons pas de rupture aussi tranchée que dans le secondaire entre les académies franciliennes et celles de Province (même si les Universités d'Île-de-France semblent plus ségréguées que les autres en moyenne).

Notons qu'entre 2006 et 2016, les académies de Créteil et Versailles ont vu leurs niveaux de ségrégation augmenter et passer au-dessus de ceux de Paris. L'IEN des très favorisés de l'académie de Créteil est passé de 2,9 % à 3,2% et celui de l'académie de Paris de 3,6% à 2,9% et l'académie de Versailles est celle qui présente l'évolution la plus marquée : l'IEN des très favorisés a augmenté de manière régulière, passant de 3% en 2006 à 4,8% en 2016⁶⁵. Plusieurs facteurs peuvent expliquer cette évolution :

- 1) Des changements de l'*offre éducative*, notamment l'apparition de nouvelles disciplines où la part des étudiants très favorisés est nettement supérieure,

⁶⁵ Cette académie comprend en 2016 dans la base SISE cinq universités (Université de Versailles Saint-Quentin-en-Yvelines, Université d'Evry-Val-d'Essonne, Université Paris XI, Université Paris X et Université de Cergy-Pontoise) ainsi que la communauté d'universités et d'établissements (COMUE) de Paris-Saclay.

comme « pluri-santé » en 2016 à l'Université de Versailles-Saint-Quentin (avec une part de 65,5% d'étudiants très favorisés) et à l'Université Paris XI (avec une part de 56%).

2) Des changements de *public* au sein des formations déjà existantes dans l'académie, notamment une hausse de la part des étudiants très favorisés (par exemple les « STAPS » à l'Université Paris XI en 2016 ont 50,41% d'étudiants très favorisés en 2016 contre 35,21% en 2006). Ce type de hausse peut s'expliquer par les caractéristiques des unités de formation comme leur localisation géographique (nouveaux bâtiments, etc.) qui peut changer, les diplômes proposés et plus généralement leur attractivité. La faculté des Sciences du Sport de Paris XI, par exemple, a bénéficié de son rapprochement avec certaines disciplines scientifiques et propose désormais 8 Masters ainsi que des formations doctorales.

4) Des changements de *public* dans les académies proches géographiquement. Par exemple, pour l'Université d'Evry-Val d'Essonne qui propose 11 formations en 2016, la part des étudiants très favorisés se situe en-dessous de la moyenne de l'académie et a diminué par rapport à 2006. L'existence d'une mobilité étudiante entre cette Université et celle de Paris XI (Académie de Versailles), proche géographiquement, ainsi que la mise en place de stratégies d'évitement ne peut être exclue.

Parmi les académies de Province, ce ne sont pas nécessairement celles liées à une grande métropole qui connaissent les niveaux de ségrégation les plus importants, si l'on excepte l'académie de Lille. Ainsi, les niveaux de ségrégation à Aix-Marseille ou Lyon s'établissent à des niveaux inférieurs à ceux observés dans de plus petites académies comme Rouen ou Reims.

La Figure 14 permet également d'analyser la ségrégation des défavorisés. Les niveaux de ségrégation sont encore plus hétérogènes que pour les très favorisés. En 2006, l'IEN varie entre 0,7% pour l'académie de Dijon et 7,4% pour l'académie de Strasbourg. Dans les formations universitaires, la répartition géographique des

étudiants est donc encore moins homogène pour les défavorisés que pour les très favorisés.

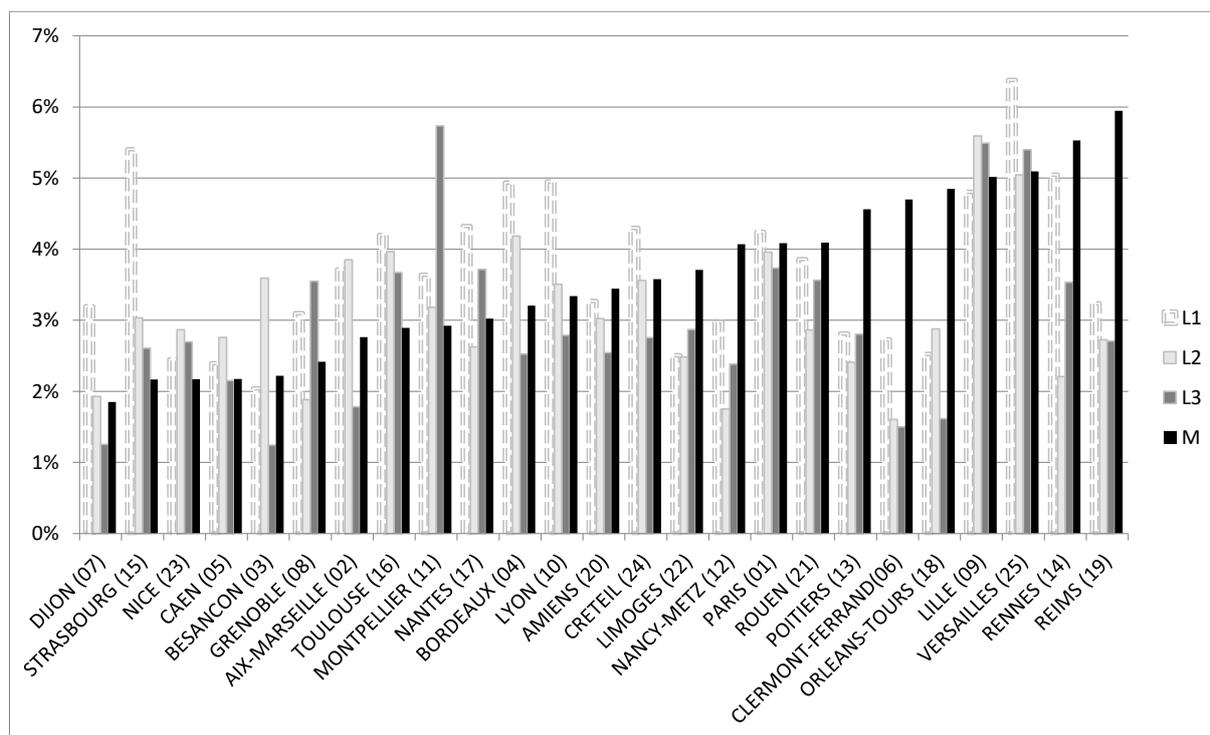
Quel que soit le groupe de référence choisi pour l'IEN (très favorisés ou défavorisés), certaines académies ont des niveaux de ségrégation très faibles, comme Dijon ou Besançon, alors que d'autres ont des niveaux de ségrégation élevés comme Strasbourg en 2006 ou Versailles en 2016. Cependant, le classement des niveaux de ségrégation par académie n'est pas exactement le même selon l'indice utilisé. A titre d'exemple, Montpellier et Poitiers sont des académies où, en 2016, la ségrégation des étudiants défavorisés est relativement élevée ce qui n'est pas le cas pour les très favorisés.

Enfin, comme avec l'IEN aux très favorisés, les trois académies parisiennes ne sont pas systématiquement les plus ségréguées. Si l'on excepte la hausse marquée de la ségrégation dans l'académie de Versailles entre 2006 et 2016, Paris et Créteil sont des académies avec des niveaux de ségrégation intermédiaires, voire relativement faibles en 2016. Ainsi, les niveaux de ségrégation dans les Universités parisiennes sont-ils légèrement inférieurs à ceux d'académies comme Reims, Nantes ou Orléans-Tours en 2016. La géographie de la ségrégation est donc largement redéfinie entre le Lycée et l'Université. Les Universités franciliennes ne sont pas des lieux aussi nettement ségrégués que peuvent l'être les Collèges ou Lycée franciliens par rapport à la Province. De même, au sein des académies de Province, les plus ségréguées ne sont pas mécaniquement celles comportant une grande agglomération (Lille, Aix-Marseille, Lyon), comme c'est le cas dans le secondaire.

5.2 Niveau de ségrégation et niveau d'étude : des disparités entre académies

L'analyse de la ségrégation par académie et par niveau d'études complète et affine notre approche géographique. La Figure 15 ci-dessous, qui indique les valeurs de l'indice d'exposition normalisé aux très favorisés par académies et selon le niveau d'études en 2016, révèle ainsi d'importantes différences inter-académiques⁶⁶.

Figure 15 : indice d'exposition normalisé aux très favorisés par académies et selon le niveau d'études en% (2016, classé par IEN croissant au niveau du Master)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Note : L1 pour Licence 1, L2 pour Licence 2, L3 pour Licence 3, M pour Master 1 et 2 ; les académies sont classées selon la valeur croissante de leur IEN, le numéro entre parenthèses sur l'axe des abscisses étant le numéro de l'académie.

⁶⁶ La population de référence permettant de calculer les indices est alors l'ensemble des étudiants inscrits au sein d'une académie particulière et d'une année d'études, par exemple l'ensemble des étudiants de l'académie de Créteil inscrits en L2, l'unité d'observation demeurant la formation.

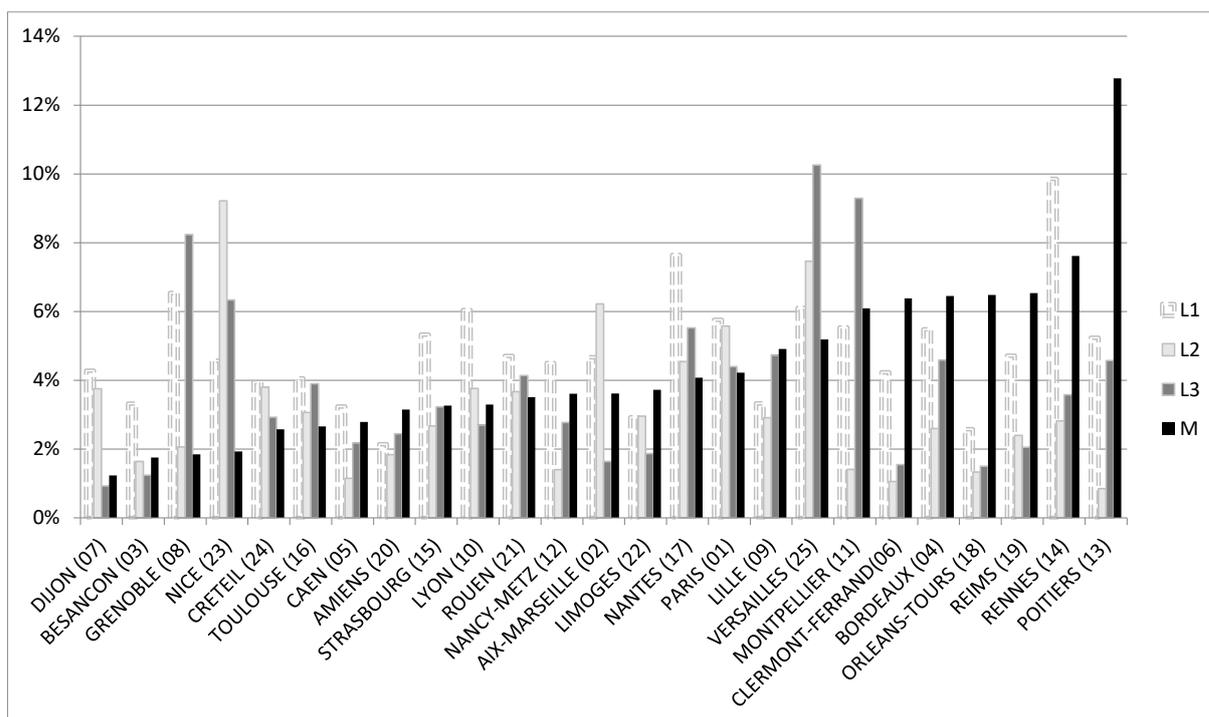
L'IEN est compris entre 1,2% (pour Besançon en L3) et 6,4% (pour Versailles en L1). En comparaison des moyennes obtenues par académies sur l'ensemble du cursus Licence et Master, nous pouvons extraire certaines tendances en distinguant les niveaux d'études :

- 1) Contrairement à la tendance relativement nette mise en évidence au niveau national, nous constatons un niveau d'entre soi légèrement supérieur en L2 qu'en L1 pour certaines académies (Besançon, Caen, Aix-Marseille, Orléans-Tours, Nice, Lille). La baisse de la ségrégation en L2 par rapport à L1 n'est donc pas portée par l'ensemble des académies, mais par certaines académies assez importantes démographiquement (comme Paris, Créteil, Versailles ou Lyon).
- 2) Nous observons un rebond assez marqué de l'entre soi des très favorisés en Master, après la L3, dans un nombre significatif d'académies de petite taille (par exemple Clermont-Ferrand, Poitiers, Orléans-Tours, Reims, Rennes, Limoges, Nancy-Metz), ainsi qu'à Paris et Créteil, en décalage avec ce qui est observé en moyenne au niveau national pour les très favorisés.

La Figure 16 ci-dessous, qui indique les valeurs de l'indice d'exposition normalisé aux défavorisés par académies et selon le niveau d'études en 2016, révèle également d'importantes différences inter-académiques⁶⁷.

⁶⁷ La population de référence permettant de calculer les indices est alors l'ensemble des étudiants inscrits au sein d'une académie particulière et d'une année d'études, par exemple l'ensemble des étudiants de l'académie de Créteil inscrits en L2, l'unité d'observation demeurant la formation.

Figure 16 : indice d'exposition normalisé aux défavorisés par académies et selon le niveau d'études en % (2016, classé par IEN croissant au niveau du Master)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Note : L1 pour Licence 1, L2 pour Licence 2, L3 pour Licence 3, M pour Master 1 et 2 ; les académies sont classées selon la valeur croissante de leur IEN, le numéro entre parenthèses sur l'axe des abscisses étant le numéro de l'académie.

La Figure 16 ci-dessus, où les académies sont classées selon leur IEN pour les défavorisés sur l'ensemble du cursus, montre que la relativement forte ségrégation des défavorisés pour certains niveaux d'étude, ne concerne en fait que certaines académies. Les niveaux d'exposition des défavorisés sont particulièrement hétérogènes (allant de 0,9% à Poitiers en L2 et à Poitiers également, mais en Master). Le principal enseignement de l'analyse géographique est que la ségrégation des défavorisés n'est pas un phénomène généralisé : par rapport aux tendances observées au niveau national nous constatons que les phénomènes de ségrégation semblent, au contraire, concentrés sur certaines académies et certains niveaux d'études. Nous constatons également des disparités de ségrégation entre niveaux d'études au sein de certaines académies. Ainsi, comme nous l'avons dit, Poitiers est l'académie la moins ségrégée en L2 et la plus ségrégée en Master. Ceci est valable pour d'autres académies : ainsi Nice est l'académie la plus ségrégée en L2 et, pourtant, l'une des moins ségrégées en Master. De même, Grenoble présente des niveaux de ségrégation assez élevés en L3, mais en-dessous de la moyenne en L2

ou en Master. De manière globale, il n'y a quasiment aucune académie présentant des niveaux de ségrégation durablement élevés tout au long du cursus (à l'exception notable de Versailles). De même, peu d'académies présentent des niveaux de ségrégation continument faibles depuis la L1 (comme Besançon, Caen, Amiens ou Limoges). De forts effets de recomposition du profil social des étudiants ont lieu d'un niveau d'études à l'autre au sein des académies et, sans doute, entre académies. Il y a peu d'académies intrinsèquement ségréguées ou mixtes et il est nécessaire d'effectuer un croisement entre académies et niveaux d'études pour obtenir une cartographie satisfaisante de la ségrégation sociale à l'Université en France.

6 La taille des académies et la mobilité géographique des étudiants : des facteurs explicatifs potentiels ?

Au-delà du constat effectué dans les sections 4 et 5 du présent chapitre, cette section s'interroge sur les facteurs géographiques susceptibles d'influencer les niveaux de ségrégation. Comme nous l'avons expliqué dans le paragraphe 2.1.2, cinq facteurs principaux sont susceptibles d'influencer les niveaux de ségrégation constatés dans l'enseignement supérieur, à savoir, la ségrégation résidentielle (spatiale), les choix individuels et familiaux, les effets de pairs, les politiques des établissements individuels ainsi que le cadre légal et les politiques d'enseignement mises en œuvre au niveau national. Dans la suite de ce chapitre, nous analysons de manière plus détaillée les liens entre disparités géographiques et ségrégation à l'Université. A cette fin, la présente section analyse le lien entre niveaux de ségrégation et, d'une part, la taille de l'académie (paragraphe 6.1) et, d'autre part, l'origine géographique des étudiants (paragraphe 6.2).

6.1 Niveau de ségrégation et taille des académies

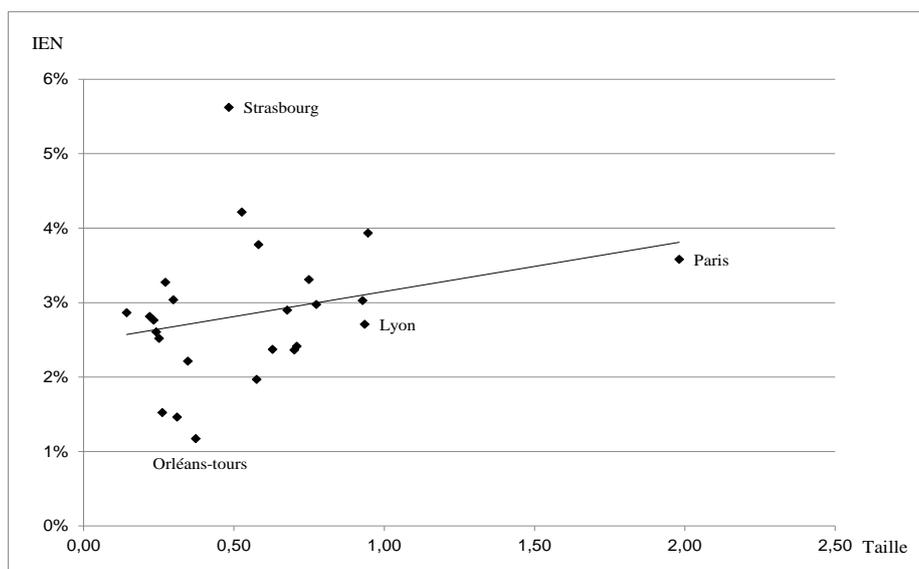
Comme le notent Ly et Riegert (2015), la ségrégation dans le secondaire a tendance à être plus élevée dans les grandes académies (qui sont souvent celles comprenant une ou plusieurs agglomérations importantes). La ségrégation *résidentielle*, fortement corrélée à la ségrégation éducative, est elle-même plus élevée dans les grandes agglomérations (notamment l'agglomération parisienne). Selon Ly et Riegert (2015), en effet, au niveau du secondaire, la ségrégation sociale est plus importante dans les départements urbains qui contiennent un nombre important de grandes villes. Cela se comprend dans la mesure où la multiplication des établissements dans les zones urbaines augmente les niveaux de ségrégation via deux mécanismes : d'abord, les phénomènes de ségrégation existants sont reflétés de manière fine dans le recrutement des élèves et, ensuite, il est plus aisé d'éviter certains établissements et d'en rechercher d'autres. A l'inverse, dans les départements ruraux, où la densité de population est faible, les établissements recrutent dans un périmètre large, ce qui favorise *de facto* la mixité.

Afin de mesurer l'impact potentiel du facteur taille sur le niveau de ségrégation observé, nous avons estimé une régression avec l'effectif de chaque académie comme variable explicative et l'IEN du groupe très favorisé comme variable expliquée. La Figure 17 ci-dessous illustre cette relation entre l'indice d'exposition normalisé des très favorisés et la taille de l'académie pour 2006. Le Tableau 9 ci-dessous nous indique, toutefois, que si l'estimateur de la pente a une faible valeur positive, cette relation n'est pas statistiquement significative, que cette variable explicative soit utilisée seule ou simultanément avec la variable d'origine géographique décrite dans le paragraphe suivant. Le lien non significatif entre taille de l'académie et ségrégation pour les très favorisés confirme ainsi que la répartition géographique de la ségrégation a été modifiée entre le secondaire et l'Université.

La Figure 18 illustre cette même relation pour les défavorisés. Le Tableau 10 indique que si l'estimateur de la pente de cette relation est significatif, c'est seulement au seuil de 10% (lorsque la taille de l'académie est utilisée en tant que seule variable explicative). Ce lien significatif aux seuils usuels entre taille de l'académie et ségrégation pour les défavorisés semble indiquer que la répartition géographique de la ségrégation n'a pas été complètement modifiée entre le secondaire et l'Université pour cette catégorie. Néanmoins ces résultats restent fragiles à ce stade. En effet, lorsque cette variable est estimée dans un modèle incluant une variable sur l'origine géographique des étudiants cette relation n'est plus significative (Tableau 10, dernière colonne).

Globalement, même si ces résultats restent à interpréter avec prudence, il semble que la répartition géographique de la ségrégation des favorisés et des défavorisés a été modifiée entre le secondaire et l'Université. Comme nous l'avons expliqué ci-dessus, il ne faut, toutefois, pas perdre de vue que seul un segment de l'offre éducative appartenant au supérieur est analysé dans le cadre de notre travail.

Figure 17: indice d'exposition normalisé aux très favorisés (en%) et taille de l'académie (en centaine de milliers) pour 2006

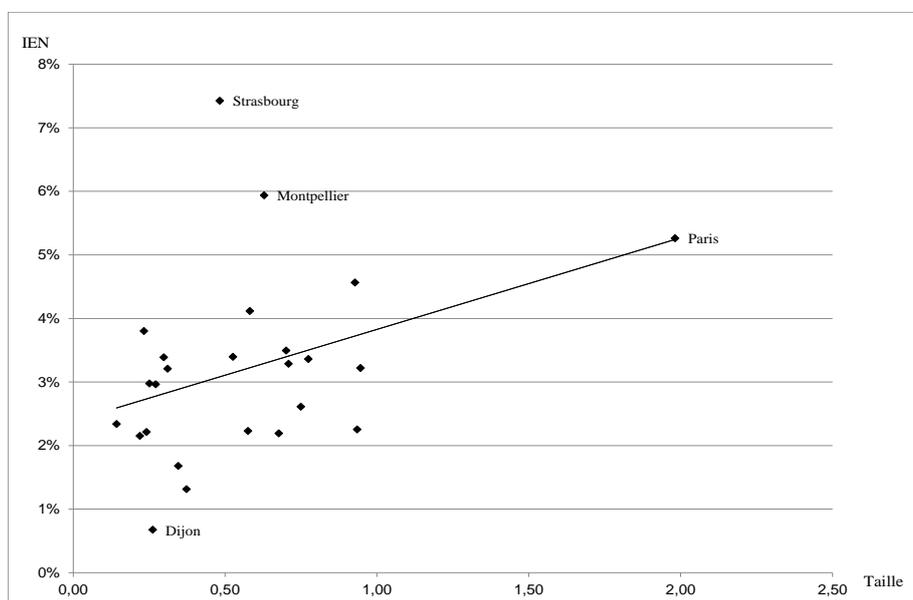


Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016 (Diffuseur : Adisp-Progedo). - calculs de l'auteur.

Note : La taille de l'académie est exprimée en centaines de milliers d'étudiants. IEN pour indice d'exposition normalisé.

Figure 18 : indice d'exposition normalisé aux défavorisés (en%) et taille de l'académie (en centaine de milliers) pour 2006



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE 2006-2016 (Diffuseur : Adisp-Progedo) - calculs de l'auteur.

Note : La taille de l'académie est exprimée en centaines de milliers d'étudiants. IEN pour indice d'exposition normalisé.

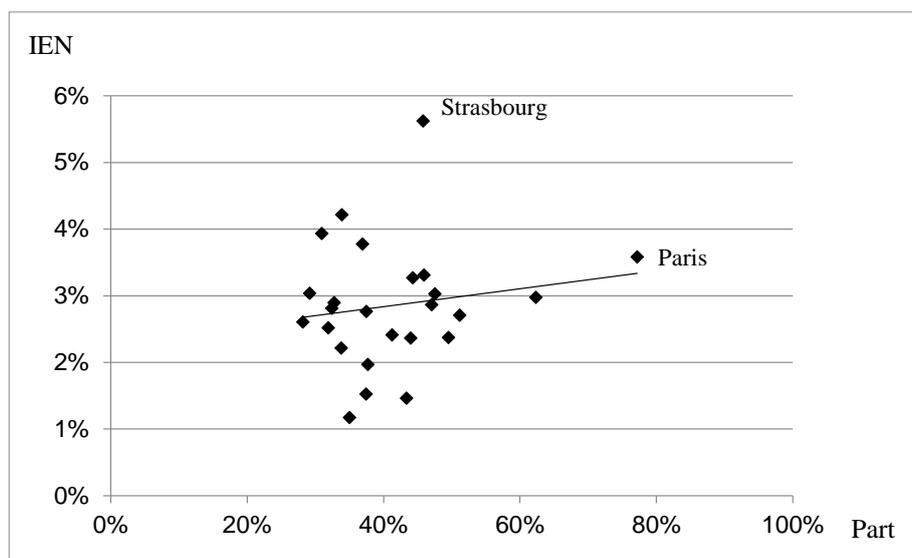
6.2 Niveaux de ségrégation et origine géographique des étudiants.

Avec les données dont nous disposons, nous pouvons également appréhender l'origine géographique des étudiants avec la proportion d'étudiants ayant obtenu leur Bac dans une autre académie⁶⁸. La Figure 19 et la Figure 20 ci-dessous illustrent ainsi la relation en 2006 entre l'origine géographique des étudiants et respectivement l'indice d'exposition normalisé des très favorisés et des défavorisés⁶⁹. Comme l'indique le Tableau 10 ci-dessous pour la ségrégation des défavorisés, si le modèle n'inclut pas d'autres variables, la relation semble statistiquement établie et l'estimateur de la pente est significatif au seuil de 5%. Ce résultat semble indiquer un lien positif entre la ségrégation des défavorisés et le fait que des étudiants originaires d'autres académies viennent y étudier. Néanmoins, cette relation est à interpréter avec prudence car la part d'élèves venant d'autres académies n'est pas indépendante de la taille de l'académie. Or, lorsque l'on contrôle de cette variable (Tableau 10, dernière colonne), cette relation n'est plus significative. Au final, le lien entre la taille et/ou l'attractivité d'une académie et son niveau de ségrégation ne semble pas pouvoir être établi.

⁶⁸ Afin d'évaluer le degré de la relation entre niveau de ségrégation et origine géographique des étudiants, nous utiliserons les variables « acarespa » et acabac » qui indiquent respectivement l'académie de résidence des parents et l'académie d'obtention du baccalauréat.

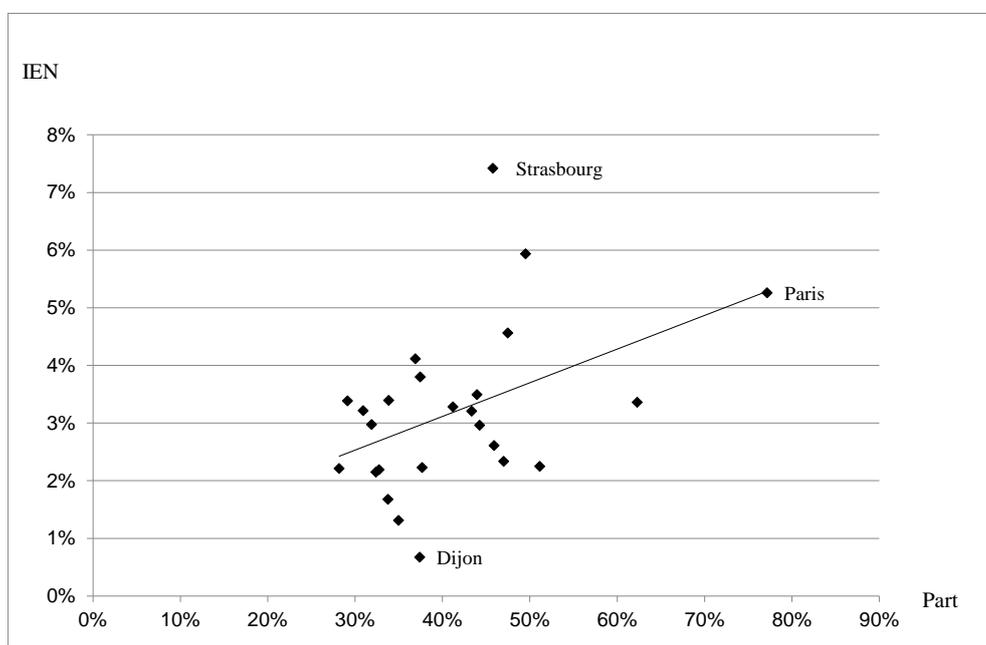
⁶⁹ Comme l'indique le Tableau 9 ci-dessous, pour la ségrégation des très favorisés, l'estimateur du paramètre de pente n'est pas statistiquement significatif, que cette variable explicative soit utilisée seule ou conjointement avec celle de la taille de l'académie.

Figure 19: indice d'exposition normalisé aux très favorisés (en%) et part d'étudiants ayant obtenu le Bac dans une autre académie (en%) en 2006



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat). Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.
 Note : IEN pour indice d'exposition normalisé.

Figure 20: indice d'exposition normalisé aux défavorisés (en%) et part d'étudiants ayant obtenu le bac dans une autre académie en 2006 (en%)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat). Source : SISE 2006-2016. Diffuseur : Adisp-Progedo. Calculs de l'auteur.
 Note : IEN pour indice d'exposition normalisé.

**Tableau 9: résultats des régressions
pour l'indice d'exposition normalisé aux très favorisés (2006)**

Variable	Valeur estimée	Erreur type	Significativité	Valeur estimée	Erreur type	Significativité	Valeur estimée	Erreur type	Significativité
Constante	0,025	(0,003)	***	0,023	(0,007)	***	0,027	(0,008)	***
Taille Académie	0,006	(0,005)	NS				0,008	(0,007)	NS
Part d'obtention du bac dans une autre académie				0,013	(0,017)	NS	- 0,007	(0,024)	NS
R ²	0,07			0,02			0,07		

**Tableau 10: résultats des régressions pour
l'indice d'exposition normalisé aux défavorisés (2006)**

Variable	Valeur estimée	Erreur type	Significativité	Valeur estimée	Erreur type	Significativité	Valeur estimée	Erreur type	Significativité
Constante	0,023	(0,005)	***	0,007	(0,01)	NS	0,01	(0,01)	NS
Taille Académie	0,014	(0,007)	*				0,005	(0,01)	NS
Part d'obtention du bac dans une autre académie				0,058	(0,02)	**	0,045	(0,03)	NS
R ²	0,14			0,19			0,20		

Champ : ensemble du système universitaire français (hors doctorat).

Source : SISE 2006. Diffuseur : Adisp-Progedo. Calculs de l'auteur

Note : *** significatif au seuil de 1%, ** significatif au seuil de 5%, * significatif au seuil de 10%, ns non significatif.

7 Conclusion

Nous avons analysé la ségrégation sociale des formations universitaires sur la période 2006-2016 en recourant à une approche géographique. Afin de comparer les niveaux de ségrégation à ceux mesurés dans le secondaire, nous avons utilisé l'indice d'exposition normalisé, ce qui nous a permis de mettre en évidence les faits suivants :

- Les IEN aux très favorisés et aux défavorisés sont, en moyenne, relativement stables : entre 2006 et 2016, ils sont compris entre 4% et 5,5%. Les phénomènes de ségrégation sociale sont ainsi bien présents à l'Université.
- La ségrégation des très favorisés a tendance à diminuer avec le niveau d'études. Elle diminue entre L1 et L2 et baisse encore en Master. Le phénomène de baisse est un peu plus marqué pour la ségrégation des défavorisés entre la L1 et la L2.
- Par ailleurs, les niveaux de ségrégation en Île-de-France ou dans les grandes académies de Province (Lille, Lyon, Aix-Marseille) ne sont pas systématiquement supérieurs à ceux observés dans de petites académies, contrairement à ce qui est constaté pour le secondaire.

A l'issue du chapitre II, l'importance relative des différents facteurs qui contribuent à la ségrégation (différences inter-universitaires, inter-disciplinaires et géographiques) n'a pas été quantifiée. Afin d'estimer celle-ci, un nouveau basculement méthodologique consistant à changer d'indice de mesure de la ségrégation est nécessaire. Les résultats de notre chapitre I suggèrent que pour cela il conviendrait de mener une analyse sur la base de l'indice d'information mutuelle. C'est ce que nous nous proposons de faire dans le chapitre III.

CHAPITRE III LA MIXITE SOCIALE A L'UNIVERSITE : LA DECOMPOSITION DES RESULTATS⁷⁰

Comme cela a été exposé dans le chapitre I, la ségrégation au niveau national se situe à un niveau inférieur à ce qui est observé dans l'enseignement secondaire. Afin de parvenir à un niveau d'analyse plus fin, une dimension géographique a été introduite dans le chapitre II. Pour cela, nous avons calculé des indices de ségrégation pour différentes académies, c'est-à-dire pour différentes sous-populations qui constituent notre champ d'analyse. Les résultats obtenus ont permis de montrer qu'il y avait une très forte diversité du niveau de ségrégation selon les académies.

Cependant, à ce stade, nous n'avons pas cherché à décomposer ce qui dans le niveau de ségrégation totale au niveau national relevait de différences de composition sociale des étudiants selon les académies de ce qui relevait d'autres dimensions de la ségrégation (nous avons notamment vu dans le chapitre II que les différentes disciplines n'étaient pas homogènes en termes de composition sociale des étudiants). Pour procéder de manière rigoureuse à une telle décomposition, nous avons vu dans le chapitre I, qu'il était nécessaire d'utiliser un indice de ségrégation fortement décomposable et que seul l'indice d'information mutuelle répondait à cette propriété. L'utilisation de ce type d'indice permet de répondre, par exemple, à la question suivante : dans quelle mesure les niveaux de ségrégation observés résultent-ils de différences inter-académiques, inter-disciplinaires et inter-universitaires ? A notre connaissance, le présent travail est le premier à décomposer un indice de ségrégation dans le domaine de l'enseignement supérieur et, ce faisant, à identifier la contribution des facteurs géographiques et disciplinaires au niveau de ségrégation des formations universitaires en France.

Ce chapitre est structuré de la manière suivante. La première section présente les enjeux de la décomposition des indices de ségrégation ainsi que des éléments de contexte portant sur la littérature disponible sur la question. La deuxième section présente la méthode utilisée et plus particulièrement la manière dont nous utilisons les indices d'information mutuelle pour contribuer à documenter la question. La troisième présente les résultats de la décomposition obtenus pour l'ensemble de la population

⁷⁰ Ce chapitre est issu d'un travail réalisé en collaboration avec Pierre Courtioux et Tristan-Pierre Maury.

au niveau national. La quatrième partie présente les résultats de la décomposition obtenus pour les étudiants de la région parisienne et pour chaque niveau d'études et la cinquième est une conclusion.

1 La décomposition des indices dans la littérature

Cette section décrit les méthodes de décomposition des indices employées dans la littérature. Les travaux utilisant ces méthodes portent à notre connaissance essentiellement sur l'enseignement primaire et secondaire.

Certains travaux mesurent les niveaux de ségrégation au sein du système scolaire d'autres pays. Ainsi Frankel et Volij (2011) mesurent l'évolution de la ségrégation « ethnique » aux Etats-Unis sur la période 1997/1998 à 2007/2008. En distinguant quatre groupes « ethniques »⁷¹, ils parviennent à la conclusion que le niveau de celle-ci demeure significatif, mais a diminué sur la période analysée. Grâce à l'utilisation d'un indice d'information mutuelle, ils décomposent le niveau de ségrégation ethnique en quatre éléments : 31,9% pour la ségrégation entre Etats, 15,9% pour la ségrégation entre zones municipales au sein des Etats (*Core Based Statistical Area*), 32,3% pour la ségrégation entre districts au sein de ces zones municipales (*CBSA*) et 19,8% pour la ségrégation entre les écoles au sein de ces districts⁷².

Boutechenik et al. (2018) mesurent la ségrégation sociale entre élèves à l'entrée en 6^{ème} dans trois grandes villes : Paris, Bordeaux et Clermont-Ferrand. Sur la base d'un indice d'entropie normalisé⁷³, le niveau global de ségrégation est décomposé en trois éléments : la ségrégation résidentielle liée à la carte scolaire, le contournement de la carte scolaire vers un autre établissement public que celui de secteur et le contournement de la carte scolaire vers un établissement privé. A Paris, à la rentrée 2015, le premier élément correspondait à environ 51% du niveau de ségrégation totale, le second à environ 44,5% et le dernier à 4,5%. La principale conclusion de ce travail est ainsi que la ségrégation totale serait entre un tiers et moitié moins élevée dans les trois villes comprises dans le champ si les élèves étaient scolarisés exclusivement dans le Collège public de leur quartier, sans que la carte scolaire ne soit contournée.

⁷¹ Comme rappelé dans la section 1 du chapitre I, les statistiques « ethniques » sont illégales en France.

⁷² La somme de ces quatre éléments permet ainsi d'arriver au total de la ségrégation ethnique, soit 100%.

⁷³ L'indice d'entropie normalisé est utilisé en raison de sa propriété de « décomposition additive ». Toutefois, Givord et al. (2015) précisent que les décompositions obtenues au moyen de l'Indice d'entropie normalisé ne sont pas parfaites. Les termes de pondération employés ne permettent pas de quantifier exactement de combien diminuerait la valeur totale de l'indice si l'une de ses composantes disparaissait. Les auteurs expliquent aussi que c'est la raison pour laquelle on peut préférer à cet indice, l'indice d'information mutuelle ou indice d'entropie non normalisé.

Givord et al. (2016), à partir de données portant sur le Collège sur la période 2003-2015 et en utilisant aussi un indice d'entropie normalisé, montrent que la ségrégation entre Collèges peut se décomposer en trois éléments. La première correspond aux différences de composition sociale entre secteurs public et privé, très variables d'une académie à l'autre. Les deux suivantes correspondent à la ségrégation au sein des Collèges de chaque secteur. Ainsi, en 2015, 18 % de la valeur de l'indice d'entropie, pour les Collèges, peut s'interpréter par l'écart de composition sociale moyen entre Collèges publics et privés, c'est-à-dire le caractère en moyenne plus favorisé des établissements privés. La ségrégation entre Collèges publics et entre Collèges privés permet d'expliquer respectivement 59 % et 23 % de l'entropie totale. Courtioux et Maury (2020), à l'aide d'un indice d'information mutuelle, montrent que la contribution des divergences public/privé au niveau total de la ségrégation entre Collèges est plus élevée dans les aires urbaines moyennes. Ces zones géographiques ne sont cependant pas celles où la ségrégation sociale est la plus forte. Ce résultat confirme l'hypothèse selon laquelle le secteur privé peut être utilisé comme un outil de contournement de la carte scolaire française et que celui-ci crée des différences sociales supplémentaires au niveau local.

Afin de pouvoir appliquer les méthodes de décomposition décrites ci-dessus au cadre de l'enseignement supérieur, il est toutefois nécessaire de tenir compte de plusieurs caractéristiques propres à ce type d'enseignement. A l'issue de l'enseignement secondaire, une partie significative des élèves n'accèdent pas à l'Université, soit parce qu'ils interrompent leurs études, soit parce qu'ils poursuivent celles-ci hors du système universitaire. Pour ceux qui accèdent au système universitaire, les étudiants peuvent, dans une certaine mesure, choisir à la fois leur Université et la discipline qu'il souhaite. Comme décrit dans la section 2 du chapitre I, le niveau de ségrégation sociale dépend principalement de cinq facteurs (susceptibles d'interagir entre eux): la ségrégation résidentielle, les stratégies d'évitement, les effets de pairs, les politiques des établissements et les orientations des politiques d'enseignement supérieur définies au niveau national. Afin de refléter l'importance relative de certains de ces facteurs dans le niveau global de ségrégation, il est par exemple possible de décomposer l'indice de mixité sociale utilisé en une dimension géographique (inter-académique), inter-disciplinaire et inter-universitaire.

2 La méthodologie utilisée : les principes de la décomposition

La section 2 décrit la méthodologie employée pour réaliser le travail de décomposition évoqué ci-dessus. A cette fin, le paragraphe 2.1 revient sur les principes de la décomposition et le paragraphe 2.2 distingue différents types de décomposition.

2.1 Les principes de la décomposition

Comme nous l'avons expliqué de manière détaillée dans le chapitre I, les phénomènes de ségrégation mesurés par les indices comprennent plusieurs composantes dont il convient idéalement d'estimer l'importance relative. Si l'unité d'observation choisie « k » est, par exemple « Université x matière » et que le degré de mixité sociale mesuré est faible, cela peut aussi bien résulter de différences entre les Universités qu'au sein de chaque Université de phénomènes de ségrégation inter-disciplinaires communs à l'ensemble du système universitaire mais lié à la place relative des disciplines au sein de chaque Université.

Comme expliqué dans la section 2 du chapitre I, seuls des indices « fortement décomposables » permettent d'attribuer à chaque élément son importance relative. Si un indice mesure le degré de ségrégation lorsque les étudiants sont regroupés « par Université et par discipline », alors l'indice est considéré comme « fortement décomposable » si la valeur de celui-ci, par exemple, est exactement égale à la somme du degré de ségrégation inter-universitaire et de la moyenne pondérée des indices calculés pour chaque Université des différences inter-disciplinaires⁷⁴. Comme expliqué par Givord et al. (2015), lorsqu'un indice n'est pas « fortement décomposable », les décompositions ne sont pas parfaites. Avec des indices « faiblement décomposables », les pondérations utilisées ne reflètent pas directement le poids de chaque unité. Dans ce cas, il n'est donc pas possible, par exemple, de quantifier de combien serait réduit l'indice s'il n'existait plus d'écart de composition sociale entre les disciplines. Afin de pouvoir quantifier cet effet, il convient d'utiliser un indice « fortement décomposable » comme l'indice d'information mutuelle.

⁷⁴ Comme expliqué dans la section 2 du chapitre I, un indice faiblement décomposable ne permet pas de savoir de combien serait réduit la ségrégation totale si, par exemple, les différences inter-disciplinaires au sein d'une Université disparaissaient.

Un indice fortement ou hautement décomposable permet d'estimer quelle part de la ségrégation totale est liée à la répartition entre Universités et quelle part résulte de la répartition par disciplines académiques. Afin de réaliser ce travail de décomposition, nous utilisons l'indice d'information mutuelle M décrit dans la section 2 du chapitre I. Suivant la formule (7) décrite dans cette section, l'indice d'information mutuelle peut se calculer comme la somme de deux composantes :

- Un indice inter-Universités. Dans ce cas, la population de référence est toujours constituée de l'ensemble des étudiants inscrits dans les Universités de France métropolitaine. L'unité d'observation, en revanche, n'est plus la formation, mais l'Université. La valeur de cet indice reflète ainsi les différences entre les Universités toutes disciplines confondues.
- La somme pondérée (en fonction du nombre d'étudiants inscrits dans chaque Université) des indices intra-Universités. Pour chaque Université, l'indice calculé reflète au sein de chaque établissement les différences entre disciplines (l'unité d'observation est alors implicitement la formation).

La décomposition de l'indice d'information mutuelle, suivant la formule décrite ci-dessus, permet de connaître l'importance relative d'une composante inter-universitaire et d'une composante inter-disciplinaire au sein de chaque Université.

2.2 Décomposition « simple » et « avancée »

Dans ce chapitre nous distinguons deux formes de décomposition que nous appelons « décomposition simple » et « décomposition avancée ». Nous considérons que la décomposition est simple si les indices sont décomposés uniquement suivant des critères inter-disciplinaires et inter-universitaires au sens où ces catégories sont déjà utilisées pour définir notre unité de base, la formation. Nous estimons, à l'inverse, que la décomposition est « avancée » dès lors que de nouvelles catégories sont introduites pour réaliser ce travail de décomposition, comme par exemple le niveau d'études ou la dimension régionale. La décomposition est « avancée » au sens où des critères de

décomposition additionnels sont introduits suivant les besoins de l'analyse. De nombreuses possibilités de décomposition sont alors envisageables.

Le tableau 11 ci-dessous liste les principaux critères de décomposition envisageables dans le cadre de notre travail et propose des abréviations destinées à faciliter la présentation des tableaux suivants.

Tableau 11: Critères de décomposition de l'indice d'information mutuelle envisagés dans le présent chapitre

Critères	Abréviation proposée	Finalité pour la décomposition
Formation	Inter-FOR	Mesurer la part de ségrégation résultant des différences entre formations. Ce critère peut être utilisé pour la décomposition par exemple en combinaison avec un indice inter-académique pour mesurer les différences entre formations au sein d'une même académie.
Université	Inter-UNIV	Mesurer la part de ségrégation résultant des différences inter-universitaires.
Discipline	Inter-DIS	Mesurer la part de ségrégation résultant des différences inter-disciplinaires.
Géographie Académie	Inter-AC	Mesurer la part de ségrégation résultant des différences entre zones académiques.
Niveaux d'études	Inter-NIV	Mesurer la part de ségrégation résultant des différences entre niveaux d'études.

Source : Auteur

Il convient de noter que ces critères doivent être sélectionnés en fonction de leur pertinence pour l'interprétation des données. Il est, par ailleurs, envisageable de les combiner suivant différentes possibilités.

3 La décomposition des indices de ségrégation portant sur la population de référence dans son ensemble

La section 3 décrit une première série de résultats portant sur la décomposition des indices de ségrégation obtenus pour la population de référence dans son ensemble, à savoir l'ensemble des étudiants inscrits dans les Universités françaises (hors doctorat). L'analyse de la décomposition portant elle-même sur des sous-groupes de la population de référence sera quant à elle introduite dans la section 4. Le paragraphe 3.1 fournit un exemple de décomposition simple suivant la discipline académique. Le paragraphe 3.2 fournit quant à lui deux exemples de décomposition avancée suivant les zones académiques et le niveau d'études.

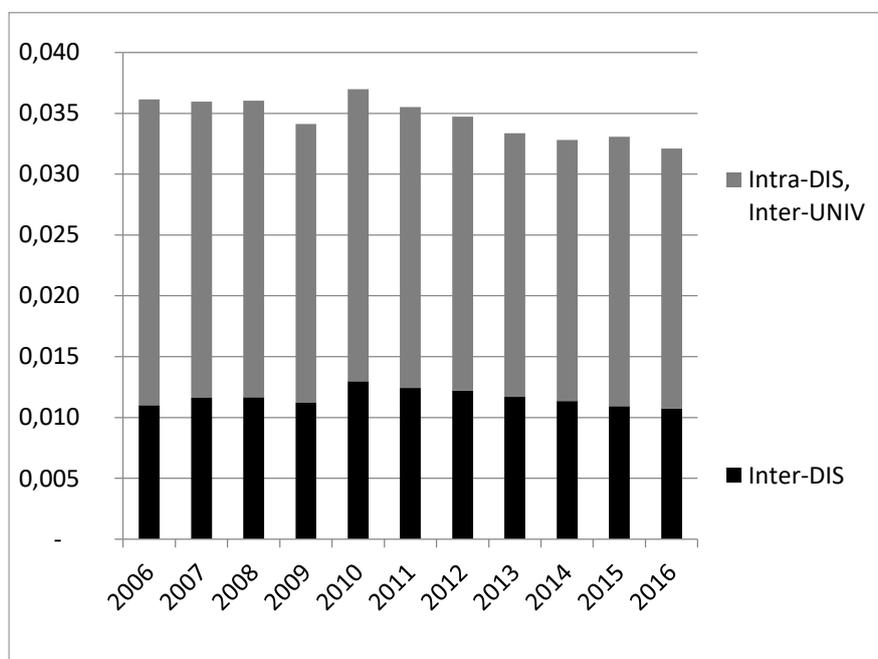
3.1 La décomposition simple suivant la discipline académique

Afin de savoir dans quelle mesure les niveaux de ségrégation observés résultent de différences entre disciplines ou de différence entre Universités, ce paragraphe décompose l'indice d'information mutuelle calculé pour l'ensemble du système universitaire. Une telle décomposition de l'indice d'information mutuelle peut être effectuée au choix suivant la discipline ou l'Université, c'est-à-dire en calculant de manière alternative soit un indice inter-disciplines puis la somme pondérée des indices inter-Universités de chaque discipline, soit un indice inter-Universités puis la somme pondérée des indices inter-disciplines de chaque Université. Nous avons choisi la première option afin de mesurer la contribution des différences entre disciplines au niveau total de ségrégation. Cela nous permet de mesurer les effets structurants au niveau national du choix de discipline de l'étudiant selon son origine sociale, décrits dans la section 2 du chapitre II.

La Figure 21 ci-dessous présente ainsi les résultats obtenus lorsque l'indice d'information mutuelle calculé au niveau de la formation est décomposé suivant les disciplines puis, pour chaque discipline, suivant les formations (les résultats détaillés des calculs de décomposition sont disponibles en annexe 9). Dans la mesure où il ne peut y avoir qu'une seule formation d'une discipline donnée par Université, par convention nous parlerons dans la suite de cette section de part liée à la ségrégation

intra-disciplinaire et inter-universitaire pour décrire cette part de la ségrégation que l'on aurait pu aussi qualifier de « intra-disciplinaire et inter-formations ».

Figure 21: Décomposition de l'indice d'information mutuelle suivant la discipline académique. 2006-2016



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Lecture: Les barres indiquent la valeur globale de l'indice chaque année. En 2006, la valeur de celui-ci est d'environ 0,036. Cette valeur se décompose en deux éléments : un facteur inter-disciplinaire (0,011) et un facteur intra-disciplinaire et inter-universitaire (0,025).

D'un point de vue dynamique, la légère tendance à la baisse de l'indice surtout à compter de 2010 observable sur la Figure 21 est cohérente avec les résultats présentés dans les chapitres I et II sur la base de l'indice d'Atkinson et d'exposition normalisé.

La Figure 21 ci-dessus nous indique que, depuis 2006, environ un tiers de la ségrégation totale observée à l'Université est due aux différences entre disciplines. L'indice d'information mutuelle inter-disciplinaire (« inter-DIS ») a, en effet, une valeur d'environ 0,011 en 2006 et en 2016, soit environ 30,5% de la valeur totale de cet indice en 2006 et 33,5% en 2016. Dans ce calcul, l'indice d'information mutuelle est calculé au niveau national entre les disciplines prises chacune comme un bloc qui correspond alors à l'unité d'observation. Ce résultat peut être relié au constat décrit dans le chapitre II suivant lequel les étudiants très favorisés sont surreprésentés dans les disciplines scientifiques (dentaire, pharmacie, santé, médecine et autres domaines

scientifiques), tandis que les étudiants défavorisés sont plus présents dans les sciences humaines et les langues (économie, sciences humaines et sociales, langues). Cette répartition peut être liée à plusieurs facteurs : l'attractivité de certaines disciplines qui peut varier selon les profils sociaux. Il est également possible qu'un phénomène d'auto-sélection conduise les élèves issus de milieux modestes à ne pas choisir des filières scientifiques jugées difficiles d'un point de vue académique. Des mécanismes de reproduction sociale peuvent également être en jeu, comme cela a été expliqué dans notre chapitre I. Enfin, la sélectivité de chaque discipline tout au long du cursus peut également expliquer ces niveaux élevés de ségrégation entre les disciplines. Bien qu'officiellement il n'y ait pas de sélection des étudiants à l'entrée de l'Université (c'est-à-dire en première année, L1), les taux d'abandon et d'échec des étudiants tout au long du cursus (de la première à la cinquième année) peuvent contribuer fortement à renforcer la ségrégation. Ainsi, il est possible qu'un taux d'échec différencié (par exemple plus élevé pour les élèves défavorisés dans les matières scientifiques) contribue à une ségrégation accrue dans les années supérieures. De manière dynamique, la Figure 21 indique, par ailleurs, que l'indice d'information mutuelle calculé au niveau inter-disciplinaire est plutôt stable sur la période 2006-2016. La décomposition des valeurs de l'indice d'information mutuelle permet ainsi d'aboutir au résultat suivant au niveau national : la ségrégation sociale serait diminuée en moyenne d'environ un tiers si les étudiants se répartissaient entre les disciplines académiques de manière aléatoire.

La Figure 21 indique également que, depuis 2006, environ deux tiers de la ségrégation totale observée à l'Université est due aux différences entre les Universités au sein d'une même discipline. L'indice d'information mutuelle intra-disciplinaire (« intra-DIS », « inter-UNIV ») a, en effet, une valeur d'environ 0,025 en 2006 et 0,021 en 2016, soit environ 69,50% de la valeur totale de cet indice en 2006 et 66,48% en 2016. Ces valeurs correspondent à la moyenne pondérée (en fonction du nombre d'étudiants) des indices calculés pour chaque discipline entre les Universités. Au sein de chaque discipline, les différences entre Universités sont évaluées et, ensuite, la moyenne pondérée des indices obtenus pour chacune des disciplines est calculé pour obtenir une valeur de l'indice pour la population totale. Ce constat accredit l'hypothèse que les différences entre les Universités au sein d'une même discipline sont très importantes. Toutefois, de manière dynamique, la valeur de l'indice calculé au sein de

chaque discipline entre les Universités a tendance à diminuer légèrement, en particulier depuis l'année 2010.

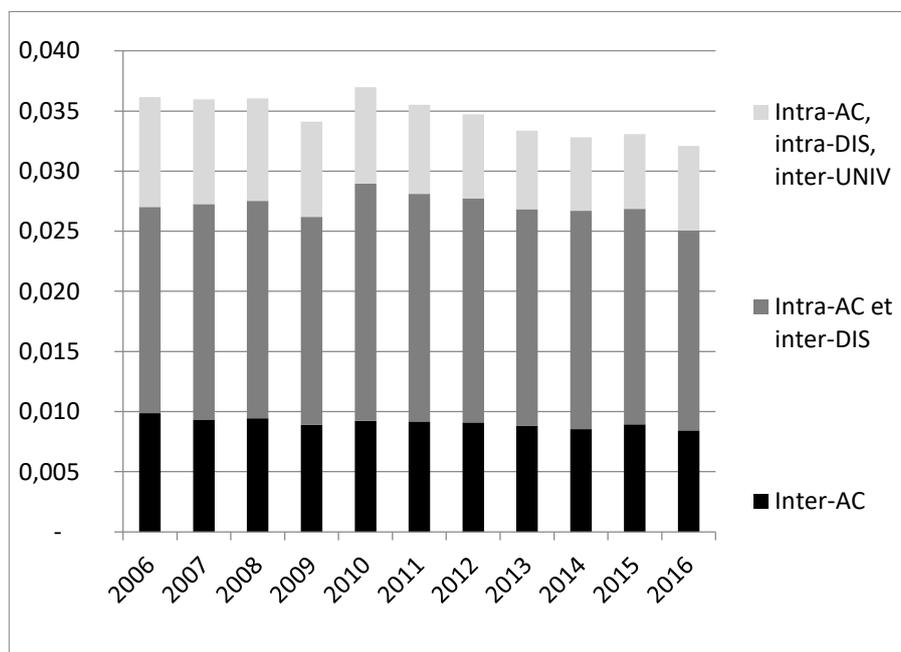
Au total, sur la période, les différences de composition sociale entre les disciplines sont plutôt stables au niveau national, mais les différences entre Universités pour une discipline donnée ont eu tendance à se réduire.

3.2 Décomposition avancée suivant les zones académiques

La décomposition des résultats obtenus suivant la discipline et les Universités ne constitue, toutefois, qu'un type d'interprétation possible des résultats. Ceux-ci peuvent être interprétés suivant d'autres dimensions. Ainsi, les niveaux de ségrégation constatés dans l'Université française pourraient être interprétés au moyen d'un facteur directement géographique. Dans quelle mesure les niveaux de ségrégation constatés dans l'enseignement supérieur sont-ils liés à un facteur géographique et aux différences inter-régionales de composition sociale ? Afin de vérifier cette hypothèse, le présent paragraphe décompose les résultats obtenus suivant les zones académiques.

Dans ce cadre, notre analyse porte d'abord sur les différences géographiques inter-académiques, puis au sein de celles-ci, sur les différences de composition sociale entre les disciplines (et non plus sur les différences de composition sociale entre les disciplines au niveau national, potentiellement influencées par la composition sociale de chaque aire géographique comme cela était proposé antérieurement). Par ailleurs, comme par construction une formation au sein d'une académie et d'une discipline donnée correspond à une Université unique, nous utilisons par convention le terme « intra-AC, intra-DIS, inter-UNIV », comme dans la section précédente, pour renvoyer aux différences entre formations au sein d'une académie et d'une discipline. La Figure 22 ci-dessous présente ainsi les résultats obtenus lorsque les indices calculés au niveau national sont décomposés suivant les zones académiques (les résultats détaillés des calculs de décomposition sont disponibles en annexe 9).

Figure 22 : décomposition de l'indice d'information mutuelle suivant les zones académiques. 2006-2016



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE (Diffuseur : Adisp-Progedo) – Calcul de l'auteur.

Lecture: Les barres indiquent la valeur globale de l'indice chaque année. En 2006, la valeur de celui-ci est d'environ 0,036. Cette valeur se décompose en trois éléments : un facteur inter-académique (0,01), un facteur intra-académique et inter-disciplinaire (0,017) et un facteur intra-académique, intra-disciplinaire et inter-universitaire (0,009).

La Figure 22 ci-dessus présente les résultats de la décomposition géographique suivant les zones académiques. Si nous choisissons d'introduire une dimension géographique à notre analyse afin de mesurer l'importance relative du facteur spatial dans le phénomène de ségrégation, la valeur de l'indice global résulte alors de la somme d'un indice portant sur les différences entre zones géographiques et d'un second indice portant sur les niveaux de ségrégation au sein des zones académiques. Afin de mener à bien cette analyse, nous avons choisi d'utiliser comme unité géographique les zones académiques et de subdiviser à nouveau l'analyse interne à chaque académie entre une dimension inter-universitaire et une dimension inter-disciplinaire. Ainsi, pour l'année 2006, la valeur de l'indice M de 0,036 est égale à la somme de trois composantes. La première composante correspond aux différences entre zones académiques ; il s'agit ainsi du niveau de ségrégation obtenu lorsque l'unité de base est la zone académique. La deuxième composante correspond à la moyenne pondérée (en fonction de la taille de chaque zone académique) des niveaux

de ségrégation obtenus entre disciplines dans chacune des zones académiques. La troisième et dernière composante correspond à la moyenne pondérée de la ségrégation entre les Universités au sein de chaque discipline pour chacune des zones académiques séparément (soit entre les formations au sein d'une académie et d'une discipline).

Nous montrons que la ségrégation entre zones académiques (première composante) a légèrement diminué sur la période sur la période d'étude. L'indice d'information mutuelle inter-académique (« inter-AC ») a, en effet, une valeur d'environ 0,01 en 2006 et d'environ 0,008 en 2016, soit environ 27,4% de la valeur totale de cet indice en 2006 et 26,2% en 2016. Dans ce calcul, l'indice d'information mutuelle est calculé au niveau national entre les zones académiques prises chacune comme un bloc qui correspond alors à l'unité d'observation. Cela signifie qu'environ un quart de la ségrégation totale est due aux différences de composition sociale entre zones académiques. Mécaniquement, cela signifie que près des trois quarts de la ségrégation en France est due à une répartition non homogène des profils d'étudiants au sein de chaque zone académique. La ségrégation est donc davantage un phénomène local qu'inter-académique dans le système universitaire français. Il y a plus de différences entre les compositions sociales de deux formations sélectionnées au hasard dans la même zone académique qu'entre la composition sociale moyenne de deux zones académiques sélectionnées au hasard. Ce résultat trouve un écho dans la littérature sur les différences régionales en France. Certaines références académiques soulignent l'importance de la redistribution interrégionale en France, qui limite considérablement les disparités (notamment de revenu) entre les régions. Dans une étude sur la ségrégation résidentielle (mesurée avec le revenu), Beaubrun-Diant et Maury (2021) montrent ainsi que la ségrégation est majoritairement locale (intra communale) et que les différences interrégionales contribuent peu à la ségrégation globale. De même, Courtioux et Maury (2020), mesurant la ségrégation au Collège, montrent qu'en 2016 seulement 17,63% de la ségrégation totale observée était due à sa composante inter-académiques. Notre étude, qui va donc dans ce sens, est la première à montrer le caractère davantage local qu'inter-académique de la ségrégation dans le système universitaire français.

La seconde composante reflète la part de la ségrégation totale due aux différences de discipline au sein de chaque zone académique, soit une mesure locale de la

ségrégation entre les disciplines. En 2006, l'indice de ségrégation inter-disciplinaire est en moyenne pondérée pour l'ensemble des zones académiques d'environ 0,017 alors qu'il est d'environ 0,01 lorsque celui-ci est calculé directement au niveau national entre les disciplines, comme indiqué sur la Figure 21 ci-dessus. Les filières les plus attractives ou sélectives ne sont pas les mêmes dans tout le pays et le choix local des filières semble générer un processus de ségrégation significatif à l'Université : les différences de composition sociale entre disciplines sont en moyenne plus marquées au niveau de l'académie qu'au niveau national. Il est possible, par exemple, que dans certaines académies, les filières scientifiques se caractérisent par une surreprésentation des étudiants les plus favorisés alors que d'autres filières seraient concernées par ce phénomène au sein d'autres académies. Cette seconde composante correspondant aux niveaux de ségrégation obtenus entre disciplines dans chacune des zones académiques représentait en 2006 47,5% du niveau total de ségrégation et 51,9% en 2016.

Nous avons vu ci-dessus que la part occupée par le facteur inter-disciplinaire dans le niveau total de ségrégation avait tendance à s'accroître au niveau national ; ce résultat confirme que cette accentuation est également visible au niveau local.

La troisième et dernière composante reflète la part de la ségrégation totale due aux différences entre Universités au sein de chaque discipline dans chaque zone académique. Celle-ci représente en 2006, 25 % de la ségrégation observée au niveau national, mais seulement 22% en 2016. Ce niveau peut se comprendre dans la mesure où chaque académie compte plusieurs Universités avec, dans la plupart des cas, une offre ample de disciplines et qu'il est très courant que deux Universités distinctes proposent le même cursus au sein d'une même académie. La valeur de ce troisième et dernier facteur passe de 0,09 en 2006 à 0,07 en 2016. Cette tendance à la baisse du facteur inter-universitaire est cohérente avec la baisse constatée ci-dessus d'une diminution de la valeur de l'indice calculé au niveau national au sein de chaque discipline entre les Universités.

Au final, mis en perspective avec les résultats antérieurs, cette analyse introduisant un niveau géographique confirme que la discipline apparaît comme un élément de plus en plus structurant de la ségrégation à l'Université même si le niveau de ségrégation a tendance par ailleurs à baisser sur la période.

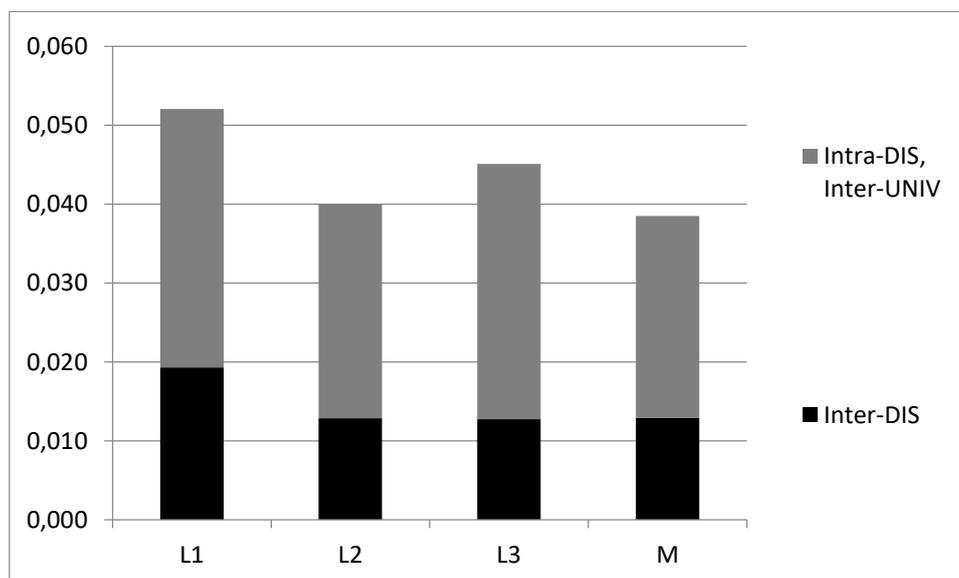
4 La décomposition des indices de ségrégation de sous-populations

La section 4 décrit une seconde série de résultats portant sur la décomposition des indices de ségrégation obtenus pour certaines sous-populations. Ce type d'analyse qui complète celle décrite dans la section antérieure permet de comprendre comment se construit la ségrégation au sein de certaines parties de la population de référence. Pour mener à bien celle-ci, trois étapes sont nécessaires : d'abord, la définition d'une sous-population dotée d'un intérêt particulier dans le cadre de notre travail, ensuite, le calcul d'indices de ségrégation pour celle-ci et, enfin, la décomposition de ces indices suivant la méthodologie utilisée antérieurement. Les paragraphes 4.1 et 4.2 ci-dessous décrivent ainsi la décomposition des indices de ségrégation calculés respectivement par niveau d'inscription et dans la région parisienne.

4.1 La décomposition des indices de ségrégation calculés pour chaque niveau d'inscription

Le présent paragraphe décrit la décomposition des indices de ségrégation par niveau d'inscription (L1, L2...). Nous avons ainsi choisi de subdiviser la population initiale de référence constituée par l'ensemble des étudiants inscrits à l'Université en quatre sous-populations : les étudiants de L1, de L2, de L3 et de Master. L'unité d'observation est définie alors par niveau d'études comme le croisement de l'Université, de la discipline et d'un niveau d'études. Ce choix se justifie par le besoin d'analyser les mécanismes de sélection à l'œuvre lorsque les étudiants passent ou non en année supérieure, se réorientent, abandonnent leurs études, etc. Si ce processus de sélection n'est pas socialement homogène, cela pourrait contribuer à renforcer les niveaux de ségrégation initiaux, avec, par exemple, des différences croissantes de composition sociale entre les disciplines académiques, à mesure que l'année d'inscription augmente. Les résultats obtenus pour ces quatre sous-populations lorsque l'on décompose suivant les disciplines sont résumés dans la Figure 23 ci-dessous (comme, par construction, une formation au sein d'une discipline donnée correspond à une Université unique, nous utilisons par convention le terme « intra-DIS, inter-UNIV » pour renvoyer au sein de chaque discipline aux différences entre formations).

Figure 23 : décomposition de l'indice d'information mutuelle calculé par niveau d'études suivant les disciplines (moyenne sur la période 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE (Diffuseur : Adisp-Progedo) – Calculs de l'auteur.

Lecture: Les barres indiquent la valeur globale de l'indice chaque année pour chaque niveau d'inscription. En moyenne sur la période 2006-2016, la valeur de celui-ci est ainsi de 0,052 pour la L1. Cette valeur globale se décompose en deux éléments : un facteur inter-disciplinaire (environ 0,019) et un facteur intra-disciplinaire et inter-universitaire (environ 0,033).

La Figure 23 qui synthétise les valeurs moyennes obtenues sur la période 2006-2016 pour l'indice d'information mutuelle peut être interprétée de deux manières distinctes, en comparant la valeur de l'indice calculée pour les différents niveaux, sans effectuer de travail de décomposition des résultats, mais aussi en décomposant ces valeurs pour chacun des niveaux de manière séparée. Nous pouvons, d'abord, comparer les valeurs de l'indice d'information mutuelle obtenues pour chaque niveau d'études⁷⁵. Nous constatons que la valeur de l'indice d'information mutuelle a tendance à diminuer avec le niveau universitaire, ce qui est cohérent avec les résultats obtenus antérieurement avec l'indice d'Atkinson dans le chapitre I et avec l'indice d'exposition normalisé dans le chapitre II.

Nous pouvons, ensuite, décomposer les indices obtenus pour chaque niveau d'études. Cette décomposition nous permet de constater que la valeur de l'indice inter-

⁷⁵ Il convient de noter que la valeur moyenne sur l'ensemble du cycle universitaire hors doctorat de l'indice d'information mutuelle calculé par niveau d'études est distincte de la valeur de cet indice calculé antérieurement, comme dans la figure 22. En effet, l'unité d'observation n'est pas la même dans les deux cas.

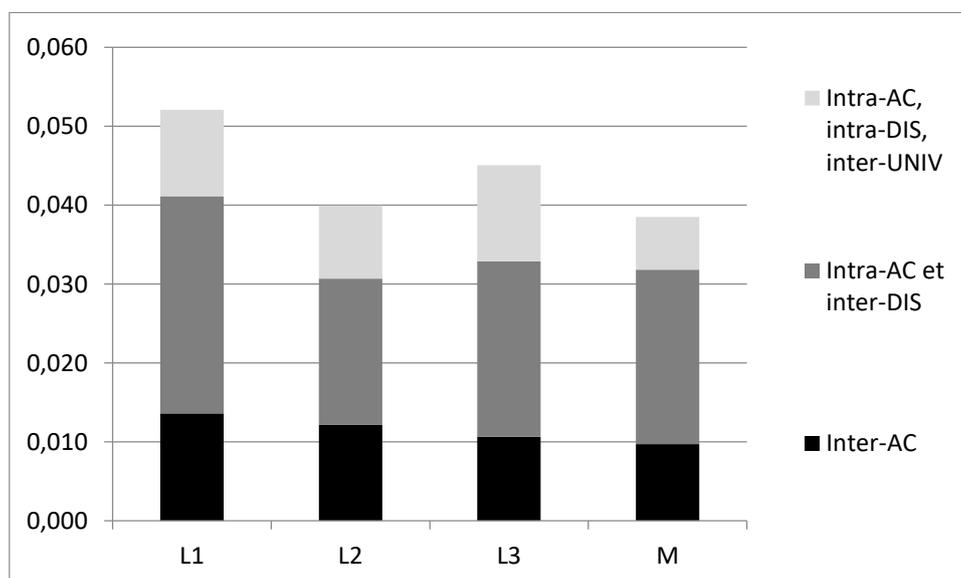
disciplinaire est en moyenne de 0,019 en L1, 0,013 en L2, 0,013 en L3 et 0,011 en Master⁷⁶. Ces résultats sont cohérents avec le résultat obtenu pour l'ensemble du cycle universitaire (hors doctorat) présenté dans le paragraphe 3.1 du présent chapitre suivant lequel le facteur inter-disciplinaire contribue à hauteur d'environ un tiers de la ségrégation totale. Ce résultat est, toutefois, à nuancer dans la mesure où il existe des différences non négligeables suivant le niveau d'études et où cette part a tendance à diminuer avec celui-ci (37% en L1, 32 % en L2, 28% en L3 et 30% en Master). Celui-ci confirme l'hypothèse présentée dans le paragraphe 4.2 du chapitre II suivant laquelle le facteur inter-disciplinaire est essentiel pour rendre compte des niveaux d'ensemble de ségrégation en L1, mais que l'importance de celui-ci diminue à compter de la L2 du fait des réorientations, mais aussi des taux d'échec des étudiants défavorisés.

La Figure 23 nous permet également de constater que la valeur de l'indice intra-disciplinaire et inter-universitaire est en moyenne de 0,033 en L1, 0,027 en L2, 0,032 en L3 et 0,026 en Master. A nouveau cela est cohérent avec le résultat obtenu antérieurement sur l'ensemble du cycle universitaire suivant lequel ce facteur contribue à hauteur d'environ deux tiers de la ségrégation d'ensemble.

La Figure 24 ci-dessous illustre les résultats obtenus lorsque les indices calculés par années d'études sont décomposés suivant les zones académiques (les résultats détaillés des calculs de décomposition sont disponibles en annexe 9).

⁷⁶ Les valeurs exactes sont disponibles en annexe 9.

Figure 24 : décomposition de l'indice d'information mutuelle calculé par niveau d'études suivant les zones académiques (moyenne 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE (Diffuseur : Adisp-Progedo) – Calculs de l'auteur.

Lecture: Les barres indiquent la valeur globale de l'indice chaque année pour chaque niveau d'inscription. En moyenne sur la période 2006-2016, la valeur de celui-ci est ainsi de 0,052 pour la L1. Cette valeur globale se décompose en trois éléments : un facteur inter-académique (0,014), un facteur intra-académique et inter-disciplinaire (0,028) et un facteur intra-académique, intra-disciplinaire et inter-universitaire (0,011).

La Figure 24 ci-dessus montre que l'importance du facteur inter-académique diminue au fur et à mesure de l'avancement dans les études, celui-ci passant notamment de 0,014 à 0,012 entre la L1 et la L2, puis encore à 0,011 en L3. Les résultats obtenus par année d'études sont, par ailleurs, plutôt cohérents avec le résultat obtenu antérieurement pour l'ensemble du cycle universitaire (hors doctorat), à savoir qu'environ un quart de la ségrégation totale est due aux différences entre zones académiques, près des trois quarts restants étant liés à une répartition non homogène des profils d'étudiants au sein de chaque zone académique. Le facteur inter-disciplinaire (facteur « intra-AC » et « inter-DIS ») diminue, enfin, entre la L1 et la L2, mais augmente de nouveau à compter de la L3.

4.2 La décomposition des indices de ségrégation calculés pour la région parisienne⁷⁷

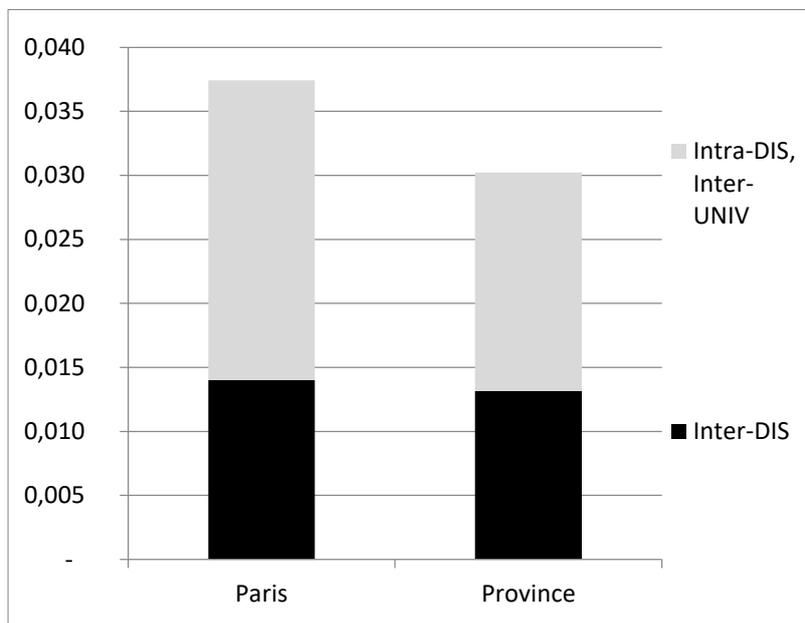
Le présent paragraphe décrit la décomposition des indices de ségrégation dans la région parisienne et dans le reste de la France. Nous avons ainsi choisi de subdiviser la population initiale de référence constituée par l'ensemble des étudiants inscrits à l'Université en deux sous-populations : celle des étudiants inscrits dans une Université de la région parisienne et celle des étudiants inscrits dans une Université située dans le reste de la France. La région parisienne telle que nous l'entendons comprend l'académie de Paris, de Créteil et de Versailles. Nous avons fait ce choix car, dans la littérature académique sur la ségrégation en France, le rôle spécifique de la région parisienne est souvent mis en avant, comme dans Oberti et Savina (2019), Bechichi et al. (2021) ou encore Ly et al. (2014). La région parisienne concentre une grande partie de la production de richesse du pays (plus du double de son poids démographique), mais est aussi le lieu de fortes inégalités de revenus. Les niveaux de ségrégation résidentielle et scolaire sont systématiquement plus élevés que dans le reste du pays (voir entre autres Ly et Riegert, 2015). Il est donc intéressant d'examiner les niveaux de ségrégation en région parisienne dans cette étude, même si la situation est assez différente à l'Université par rapport à la ségrégation au Collège ou au Lycée. En effet, la mobilité interrégionale des étudiants universitaires est un phénomène important. Près de 50% des étudiants en France ont obtenu leur baccalauréat (et donc terminé leurs études secondaires) dans une zone différente de celle où ils font leurs études supérieures. Dans l'académie de Paris, cette proportion atteint en 2006 77,1%⁷⁸ : la population des Universités n'y est pas du tout la même que celle des Lycées. Il est donc intéressant de voir dans quelle mesure cette forte mobilité a pu contribuer à modérer (ou renforcer) les niveaux de ségrégation en région parisienne et les écarts entre la région parisienne et le reste du pays.

Les résultats concernant les niveaux de ségrégation et leur décomposition en Île-de-France et dans le reste du pays suivant les différences inter-disciplinaires sont présentés dans la Figure 25 ci-dessous (les résultats détaillés du calcul des indices sont disponibles par année dans l'annexe 9).

⁷⁷ Dans ce paragraphe, le terme de « région parisienne » désigne l'Île-de-France.

⁷⁸ Estimation réalisée à partir de la base SISE Université (calculs de l'auteur).

Figure 25 : décomposition de l'indice d'information mutuelle calculé pour la région parisienne et la Province suivant les disciplines (moyenne 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Lecture: La première colonne indique la valeur de l'indice pour la région parisienne. Pour celle-ci, le facteur inter-disciplinaire a eu une valeur de 0,014.

La Figure 25 ci-dessus montre que les niveaux de ségrégation sont en moyenne supérieurs dans la région parisienne à ceux observés dans le reste du pays (ce qui n'implique pas que les académies de la région parisienne soient systématiquement parmi les plus ségréguées). La ségrégation dans celle-ci est en moyenne de 0,037 sur la période 2006-2016, contre 0,03 dans les autres zones académiques françaises. Par ailleurs, les résultats détaillés disponibles par année en annexe 9 montre que celle-ci est stable, voire en légère augmentation en région parisienne sur la période analysée (0,039 en 2016, soit +4% environ sur la période analysée) mais en nette baisse dans le reste du pays (environ -12% entre 2006-2016). Ces tendances sont conformes à celles mises en évidence dans d'autres travaux portant sur la ségrégation dans l'enseignement secondaire. Courtioux et Maury (2020) montrent ainsi que la ségrégation scolaire augmente plus vite en région parisienne que dans les autres zones académiques.

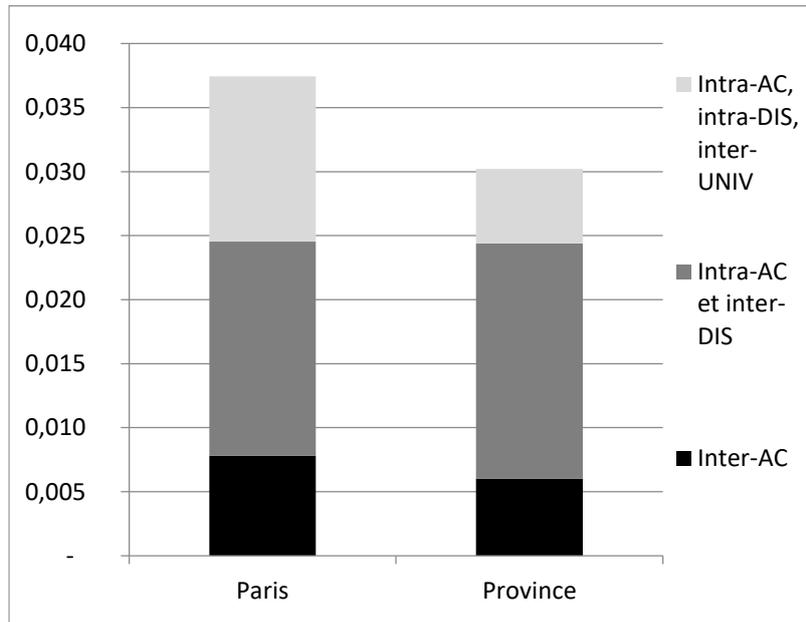
Nous montrons cependant que les différences entre la région parisienne et le reste du pays sont beaucoup plus faibles si l'on considère la ségrégation universitaire que les autres formes de ségrégation. Pour les Collèges, la ségrégation en région parisienne

est ainsi supérieure de 51,22% à la moyenne nationale en 2016 (Courtioux et Maury, 2020). Pour l'Université, la différence n'est que de 15,69%. Une telle diminution de l'écart Paris/Province entre le secondaire et l'Université peut s'expliquer par deux facteurs : 1) une redistribution géographique post-bac massive des étudiants au sein des zones académiques de la région parisienne et 2) l'importance des flux de mobilité étudiante en provenance et à destination de la région parisienne pour l'entrée à l'Université. Bien que les données dont nous disposons ne permettent pas d'identifier et de quantifier précisément ces effets (on ne connaît pas le Lycée d'origine du Lycéen), l'importance des flux des autres régions vers la région parisienne après le Lycée suggère que les nouveaux arrivants ont contribué à réduire la ségrégation dans la région parisienne (ou du moins à combler l'écart avec le reste du pays).

La Figure 25 montre également que les différences inter-disciplinaires contribuent environ à hauteur de 43% de la ségrégation totale en Province, contre environ 37% en Ile-de-France alors que la contribution des différences intra-disciplinaires et inter-universitaires dans ces deux zones géographiques est respectivement de 57% et 63 %. Ceci indique que les différences inter-universitaires contribuent davantage au niveau total de la ségrégation en région parisienne qu'en Province (où le facteur inter-disciplinaire est en proportion plus important). Ce résultat peut se comprendre au regard du nombre relativement élevé d'établissements universitaires présents en Ile-de-France comparativement à la Province. En effet, alors qu'en Province, le choix davantage limité d'établissements universitaires contribue à la mixité sociale, en région parisienne, le nombre d'établissements est élevé, ce qui implique des possibilités accrues de regroupement en fonction de l'origine sociale. Ces analyses recourent celles effectuées dans la section 4 du chapitre II concernant la comparaison des niveaux de ségrégation entre l'Université et le secondaire.

La Figure 26 ci-dessous présente la décomposition de l'indice d'information mutuelle calculé pour la région parisienne et la Province en introduisant le niveau zone académique.

Figure 26 : décomposition de l'indice d'information mutuelle calculé pour la région parisienne et la Province suivant les zones académiques (moyenne sur la période 2006-2016)



Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : SISE. Diffuseur : Adisp-Progedo. Calculs de l'auteur.

Lecture: La première colonne indique la valeur de l'indice pour la région parisienne (et non pour l'académie de Paris). Pour celle-ci la valeur du facteur inter-académies est de 0,07.

La Figure 26 ci-dessus montre que si le niveau de ségrégation est plus élevé en région parisienne qu'en Province, c'est en raison essentiellement des différences inter-universitaires qui existent au sein des académies et des disciplines académiques : la valeur de l'indice est pour cette composante de 0,013 pour la région parisienne contre 0,006 en Province en moyenne sur la période 2006-2016. Ce résultat conforte le résultat antérieur qui soulignait le rôle essentiel qu'occupent les différences inter-établissements au sein de la région parisienne.

5 Conclusion

Afin de compléter le cadrage géographique du chapitre II, nous avons utilisé les propriétés de décomposition de l'indice d'information mutuelle pour analyser plus finement les interactions entre disciplines, zones académiques, Universités et niveaux d'études. Une première série de résultats concerne la contribution des disciplines au niveau de ségrégation à l'Université. Le fait que les étudiants des différentes disciplines soient également très différents socialement contribue à un tiers de la ségrégation au niveau national. Nous montrons aussi que cette dimension ségrégative de la discipline est également présente au niveau local : les filières les plus attractives ou sélectives ne sont pas les mêmes dans tout le pays. De ce point de vue le choix local des filières semble générer un processus de ségrégation significatif à l'Université.

La seconde série de résultats obtenus concernent les niveaux d'études, la valeur de l'indice d'information mutuelle ayant tendance à diminuer avec l'avancement dans les études. De plus, nos résultats confirment l'hypothèse formulée dans le chapitre II suivant laquelle le facteur inter-disciplinaire est essentiel pour rendre compte des niveaux d'ensemble de ségrégation en L1, mais que l'importance de celui-ci diminue à compter de la L2 du fait des réorientations, mais aussi des taux d'échec des étudiants défavorisés.

Nos résultats confirment également la particularité de la région parisienne, avec des niveaux de ségrégation observés en moyenne supérieurs à ceux du reste du pays et en hausse sur la période 2006-2016 alors qu'une tendance inverse est observée dans le reste du pays. Les analyses de décomposition réalisées nous permettent de montrer que si le niveau de ségrégation est plus élevé en région parisienne qu'en Province, c'est principalement en raison des différences inter-universitaires qui existent au sein des académies et des disciplines académiques. Ce résultat, toutefois, n'implique pas que les académies de la région parisienne soient systématiquement parmi les plus ségréguées.

CHAPITRE IV : QUEL LIEN ENTRE MIXITE SOCIALE ET REUSSITE AUX EXAMENS ?⁷⁹

Les chapitres II et III nous ont permis de mesurer l'ampleur des phénomènes de ségrégation sociale existants au sein de l'Université française et de hiérarchiser l'importance de ses différentes dimensions, notamment géographiques et disciplinaires. Or, en abordant la question de la mixité sociale, nous faisons généralement et implicitement l'hypothèse que des individus issus de différents groupes sociaux ne font pas que coexister dans un même espace commun, mais interagissent au sein de celui-ci. En ce sens, d'un point de vue général, la mixité sociale n'est pas un processus neutre pour la société qui n'aurait par exemple aucun effet sur la cohésion de celle-ci. D'un point de vue économique, il peut sembler légitime de se demander si la mixité sociale n'a pas un effet sur l'efficacité et/ou la productivité. Analyser les conséquences potentielles de la mixité sociale constitue une étape complémentaire d'analyse qui suit la nécessaire quantification de ce phénomène effectuée dans les trois chapitres précédents, ce nouveau chapitre étant ainsi la suite naturelle de ceux-ci.

Notre chapitre IV est dérivé d'une question classique en économie de l'éducation qui est de s'interroger sur la productivité du système éducatif. A cette fin, nous avons fait le choix d'évaluer le lien entre mixité sociale des formations suivies à l'Université et réussite étudiante à l'issue de la première année de Licence à l'Université. Le choix de centrer notre analyse sur la réussite à l'issue de la première année de Licence se justifie au regard des niveaux de ségrégation particulièrement élevés en Licence 1, suivant le constat effectué dans les chapitres II et III. L'hypothèse sous-jacente de notre travail d'évaluation est que la mixité sociale peut jouer sur cette réussite via des phénomènes tels que les effets de pairs (diffusion des modèles de réussite, partage d'informations et de méthodes de travail...) ou encore la modification des pratiques pédagogiques des enseignants en fonction des groupes d'étudiants.

Notre analyse est susceptible de contribuer au débat sur la question de l'échec à l'Université qui est centrale en France en raison des faibles taux de réussite en Licence. Brinbaum et al. (2018) s'interrogent ainsi sur la faiblesse des taux de réussite

⁷⁹ Ce chapitre s'appuie sur un travail de recherche réalisé en commun avec Pierre Courtioux.

d'environ 50% en Licence et parviennent à la conclusion que les parcours scolaires antérieurs au bac jouent un rôle prépondérant. En effet, afin de comprendre la probabilité de passage en année supérieure, nous avons analysé les liens de cette variable avec deux catégories de facteurs bien documentés dans la littérature économique. Nous avons mobilisé, d'une part, une série de facteurs explicatifs individuels (genre, filière de bac...) liés aux caractéristiques personnelles de l'étudiant. Nous avons utilisé, d'autre part, une série de facteurs explicatifs de nature institutionnelle ou collective (Université, discipline choisie...) : dans ce cas, le passage en année supérieure peut être influencé par les caractéristiques de la formation suivie, des facteurs institutionnels tels que les politiques définies au niveau national ou les méthodes pédagogiques utilisées au sein de l'Université. Parmi cette seconde série de facteurs, nous introduisons de manière novatrice des indicateurs de mixité sociale des formations suivies.

Ce chapitre est structuré de la manière suivante. La première section discute des liens entre mixité sociale et réussite étudiante tels qu'ils apparaissent dans la littérature. La deuxième section présente la manière dont nous avons constitué une base de données originale permettant de décrire à la fois les caractéristiques socio-économique des étudiants de manière plus détaillée et celle des formations qu'ils suivent en termes de mixité sociale. La troisième s'appuie sur des statistiques descriptives pour présenter un premier cadrage de la réussite étudiante selon l'origine sociale. La quatrième section expose nos résultats sur les liens entre mixité sociale et réussite en Licence et la cinquième est une conclusion.

1 Les facteurs de réussite à l'Université et la question de la mixité sociale

La question de la réussite à l'Université est centrale en économie de l'éducation, mais aussi en sciences de l'éducation. Celle-ci fait l'objet d'une littérature abondante, notamment en France en raison précisément de l'ampleur du phénomène d'échec en Licence. Comme nous le verrons dans les paragraphes suivants, si la question des facteurs potentiels d'échec est fréquemment abordée, les analyses économiques des conséquences de celui-ci sont plus rares ; par exemple Chéron et Courtioux (2018) décrivent la question des bénéfices socio-économiques de la diplomation dans le supérieur, ce qui les conduit à évaluer le coût financier potentiel résultant du taux élevé d'abandon en Licence.

Il y a plusieurs manières d'analyser les facteurs de réussite à l'Université. Une possibilité est de les regrouper, comme le font notamment Monso et al. (2019), en trois catégories principales : les facteurs individuels, les facteurs de contexte qui ne dépendent pas des pairs tels les politiques éducatives et les facteurs de contexte qui dépendent des pairs, c'est-à-dire des autres étudiants qui suivent la même formation. Cette première section, aussi, est structurée à partir de ces trois catégories de facteurs identifiables. Le paragraphe 1.1 décrit ainsi les facteurs individuels de réussite à l'Université, le paragraphe 1.2 les effets de contexte non attribuables aux pairs et le paragraphe 1.3 les effets de contexte attribuables aux pairs. Le paragraphe 1.4 conclut sur les points qui nous paraissent les plus importants à retenir pour notre analyse.

1.1 Les facteurs individuels de réussite

Nous nous penchons dans un premier temps sur les facteurs liés aux caractéristiques individuelles et familiales de l'étudiant (1.1.1), avant d'aborder ceux liés aux conditions de vie matérielles (1.1.2) et la question de la personnalité et du temps de travail (1.1.3).

1.1.1 Les caractéristiques individuelles et familiales

La première catégorie de facteurs de réussite universitaire est constituée par l'ensemble des caractéristiques individuelles et familiales de l'étudiant. Le rôle de ces facteurs dans la réussite est bien établi dans la littérature, notamment parce que ceux-ci sont couramment disponibles dans les bases de données et ont été utilisés de manière fréquente dans les estimations économétriques. Ces caractéristiques sont liées à l'état civil de l'étudiant, à la PCS de ses parents ainsi qu'à son parcours scolaire antérieur. Au vu de la littérature, nous avons ainsi pu identifier les caractéristiques individuelles suivantes : l'origine sociale, le parcours scolaire antérieur, l'origine géographique et le genre.

1. L'origine sociale est susceptible de constituer un facteur de réussite universitaire. La proportion d'étudiants issus d'un milieu populaire en France ayant fortement augmenté depuis les années 1960, de nombreux travaux abordent la question de l'origine sociale comme facteur de réussite à l'Université. Comme expliqué dans les chapitres I et II, les travaux de Brinbaum et al. (2018), qui se basent sur un panel d'élèves entrés au Collège en 1995, montrent que d'un point de vue statistique l'origine sociale est un facteur de réussite en Licence si l'on ne tient pas compte du parcours scolaire antérieur. Toutefois, les auteurs nuancent eux-mêmes ce constat en intégrant dans un second temps les paramètres du parcours scolaire dans le secondaire et, en particulier, en distinguant les titulaires d'un bac général et ceux d'un bac technologique ou professionnel. Ainsi, 88% des jeunes inscrits en premier cycle universitaire après avoir obtenu un bac général avec mention obtiennent une Licence lorsque leurs parents sont tous les deux employés ou l'un d'entre eux employé et l'autre ouvrier ou inactif et cette proportion passe à 85% lorsque leurs deux parents font partie des cadres ou professions intermédiaires⁸⁰. Pour les jeunes ayant obtenu un bac général sans mention mais sans retard, ce même taux diminue à 71% pour ces deux catégories d'origine sociale susmentionnées. Brinbaum et al. (2018) montrent cependant que l'origine sociale joue un rôle majeur parmi les bacheliers technologiques ou professionnels. En

⁸⁰ Comme il est précisé dans le chapitre I et rappelé ci-dessous, certains résultats de ce travail reposent également sur les origines géographiques et migratoires des parents des étudiants.

effet, parmi ceux-ci, 21% obtiennent une Licence lorsque leurs parents sont tous les deux employés ou l'un d'entre eux employé et l'autre ouvrier ou inactif alors que cette proportion passe à 38% lorsque leurs deux parents font partie des cadres ou professions intermédiaires.

La PCS des parents n'a pas un impact significatif sur la probabilité d'obtenir une Licence universitaire en trois ans, dès lors que sont pris en compte les autres caractéristiques des étudiants et, en particulier, le parcours scolaire antérieur (Felouzis, 2000). Dans un travail de synthèse, Galand et al. (2005) montrent que les variables reflétant l'origine sociale ne rendent compte que de 1 à 5% de la variance dans les performances universitaires parmi quatre travaux analysés (Arulampalam et al., 2004 ; Degraeve et Martou, 1996 ; Galand et al., 2004 ; Romainville, 2000)⁸¹. Les analyses portant sur les origines sociales comprennent, toutefois, des effets de composition complexes, les étudiants défavorisés s'orientant par exemple davantage vers les bacs professionnels. Il nous semble, aussi, nécessaire de réaliser des estimations qui contrôlent ces effets de structure.

Comme expliqué de façon plus détaillée dans le chapitre I, il convient, par ailleurs, de noter qu'au-delà de la PCS, l'origine sociale peut aussi être caractérisée via d'autres variables comme le niveau de diplôme des parents ou encore le statut de boursier.

2. Le parcours scolaire antérieur. L'importance du parcours scolaire antérieur est également bien documentée dans la littérature. Celui-ci peut reposer sur différents éléments comme les notes obtenues antérieurement (en particulier dans l'enseignement secondaire), les filières suivies (typiquement la filière de bac) ainsi que le redoublement. Allen et Carter (2007) constatent ainsi que les notes obtenues dans les enseignements de base au début du cursus universitaire constituent un élément de prédiction essentiel de la réussite universitaire ultérieure des étudiants spécialisés sur le marché de l'immobilier⁸². Brinbaum et al. (2018) soulignent le rôle essentiel de la note obtenue au bac

⁸¹ Ces travaux concernent respectivement le Royaume-Uni, la Belgique et la France.

⁸² Cette analyse repose sur un panel d'étudiants de l'Université publique Florida Atlantic University dont le diplôme est spécialisé sur le marché immobilier.

ainsi que de la filière choisie, le taux d'obtention de la Licence étant significativement plus faible chez les titulaires d'un bac professionnel ou technologique. L'existence d'un redoublement n'est pas toujours disponible dans les bases de données, de nombreux travaux utilisent aussi l'âge ou la date de naissance des étudiants pour obtenir une indication indirecte sur le parcours scolaire antérieur.

3. Les origines ethniques et géographiques. Aux Etats-Unis, de très nombreux travaux font référence à la dimension ethnique au sein de l'éducation (voir par exemple, James et Taeuber, 1985 ; Frankel et Volij, 2011). Ce type de travaux est rendu possible par l'existence de statistiques ethniques, celles-ci ayant un fondement légal aux Etats-Unis. Dans le cas des Etats-Unis, il s'agit d'un facteur utilisé notamment pour prédire la probabilité de réussite aux examens universitaires. Dans le cas de la France, si les statistiques « ethniques » sont illégales⁸³, il est, toutefois, légalement possible d'approcher indirectement ces données via le concept d' « origine migratoire », les enfants nés en France de parents immigrés pouvant, en effet, être identifiés grâce au lieu et à la nationalité de naissance de leurs parents. Brinbaum et al. (2018) distinguent ainsi au sein de l'Université française trois groupes d'étudiants afin d'analyser l'échec en Licence (« Français d'origine », « origine maghrébine », « origine d'Europe du Sud »). Les auteurs indiquent que, pour un milieu social donné, les étudiants d'origine maghrébine obtiennent moins souvent la Licence suivant les estimations réalisées.
4. Le genre. D'après les travaux disponibles, le genre des étudiants aurait un impact significatif sur la probabilité de passage en année supérieure. Selon Lemaire (2000), le taux de passage en année supérieure des bacheliers généraux est significativement plus élevé pour les étudiantes que pour les étudiants. A caractéristiques comparables, en particulier scolaires, les garçons sont plus nombreux à redoubler, à se réorienter ou interrompre leurs études

⁸³ En France, le Conseil Constitutionnel, par sa décision du 15 novembre 2007, basée sur l'article 1 de la Constitution a jugé contraire à la Constitution le fait de faire apparaître directement ou indirectement les origines « raciales ou ethniques » ainsi que l'introduction de variables de « race » ou de religion dans les données administratives.

(Lemaire, 2000). Toutefois, suivant Michaut (2012), ces différences reposeraient principalement sur les pratiques de travail des étudiantes et le temps consacré au travail universitaire⁸⁴.

1.1.2 Les conditions de vie de l'étudiant

Au-delà des caractéristiques strictement individuelles liées à l'état civil de l'étudiant et à son parcours scolaire antérieur, les travaux analysant les facteurs de réussite universitaire ont progressivement intégré les questions liées aux conditions de vie des étudiants. Il s'agit de variables qui dépendent en partie des choix de l'étudiant et de son origine sociale, mais aussi des politiques publiques. Ces conditions incluent essentiellement trois dimensions : le logement, les revenus du travail et l'intégration sociale.

5. Le rôle du lieu d'habitation est analysé dans le travail de Grignon et Gruel (1999) effectué dans le cadre d'une enquête réalisée en 1997 par l'Observatoire de la Vie Étudiante. Suivant les statistiques descriptives disponibles, les étudiants ayant quitté le domicile parental auraient en moyenne de plus grandes chances de réussite pour les trois années de Licence. Le fait d'habiter en résidence universitaire, en foyer ou en internat serait un facteur positif de réussite, mais d'ampleur limité.
6. Au-delà des caractéristiques individuelles traditionnelles rappelées ci-dessus, la littérature s'est intéressée depuis quelques années aux sources de revenus des étudiants et notamment à la question du travail étudiant. Sur cette question, la littérature opère généralement des distinctions en particulier suivant la durée hebdomadaire de travail. Le temps de travail salarié n'aurait pas d'impact en dessous d'un certain seuil. Befy et al. (2009), sur la base de l'Enquête Emploi de l'INSEE de 1992 à 2002 montrent qu'un temps de travail salarié important augmente les risques d'échec pour les étudiants en cours d'études initiales à l'Université et préparant un diplôme universitaire de premier ou de second cycle, notamment si celui-ci excède 15 à 20 heures par semaine. Zilloniz (2017)

⁸⁴ Ces autres dimensions sont abordées dans les paragraphes suivants.

utilise comme critère central pour distinguer les différents types d'emploi étudiants l'existence d'un lien ou non avec le cursus universitaire. Sur la base des enquêtes Emploi 2013-2015 de l'INSEE, il est possible de constater que plus de 50% des étudiants qui travaillent exercent une activité en lien avec leur cursus d'inscription (stage, apprentissage, internat de médecine, etc.). Les activités rémunérées sans lien avec leurs études correspondent souvent à des emplois peu qualifiés et plus fréquemment à temps partiel que lorsqu'il existe un lien avec le cursus. Toutefois, cela reste plutôt un facteur négatif pour la réussite universitaire (Zilloniz, 2017).

7. L'intégration académique et sociale. L'absence d'isolement et le soutien familial et amical constituent des facteurs positifs de réussite (Bodson, 1999). Peu de travaux reflètent, toutefois, cette dimension qui repose généralement sur des questionnaires d'auto-évaluation.

1.1.3 Les facteurs liés à la personnalité de l'étudiant et à ses méthodes de travail

Ce que nous regroupons dans les facteurs liés à la personnalité de l'étudiant regroupent aussi bien ses capacités intellectuelles, ses traits psychologiques et ses motivations que ses méthodes de travail. Il s'agit de données individuelles sur lesquelles l'étudiant peut potentiellement agir mais qui ne sont pas l'objet directement des politiques publiques.

8. Les capacités intellectuelles mesurées via des tests standards. L'obtention de bons résultats à ces tests est associée à une probabilité de réussite accrue, mais également à un meilleur parcours scolaire (Robbins et al., 2004) ; Toutefois, Galand et al. (2005), à partir d'une revue de la littérature, estiment que les effets conjoints du parcours scolaire antérieur et des capacités intellectuelles rendent généralement compte de 5 à 15% de la variance des notes obtenues.
9. La motivation et la confiance en soi des étudiants sont d'autres facteurs essentiels pour la réussite étudiante, mais difficiles à mesurer et à intégrer dans des travaux quantitatifs. Peu de travaux empiriques prennent en compte ces facteurs en raison des difficultés existantes pour les appréhender. Lambert-Le

Mener (2012) s'est, toutefois, intéressé au rôle joué par la motivation sur la réussite des étudiants de première année universitaire. Le niveau de motivation est mesuré par l'étudiant lui-même qui se positionne sur une échelle par rapport à différents critères. La conclusion principale est que la motivation des étudiants joue un rôle positif et significatif sur la moyenne annuelle obtenue alors que le manque de motivation exerce un impact négatif. Toutefois, cet effet dépend lui-même du niveau des étudiants : via une analyse par quantiles de notes, on peut constater que la motivation n'a pas d'effet significatif sur les résultats obtenus par les étudiants les plus faibles. Algan et al. (2018) analysent le rôle des compétences socio-comportementales⁸⁵ qui influencent la capacité à apprendre et, ensuite, la réussite professionnelle. Les auteurs identifient un lien positif entre ces compétences et la réussite scolaire.

10. Les méthodes de travail et le temps de travail. Ce facteur fait plutôt l'objet de résultats contrastés. Toutefois, il semble qu'un travail régulier associé à une bonne gestion des temps de loisirs soit un facteur de réussite (Robbins et al., 2004).

1.2 Les effets de contexte non attribuables aux pairs

Les effets de contexte non attribuables aux pairs font référence au contexte, à l'environnement partagé par un groupe d'élèves ou d'étudiants qui est susceptible d'avoir une influence sur la progression des élèves. Cela peut correspondre par exemple au fait de partager un même enseignant qui peut avoir une influence sur la progression d'ensemble des élèves. Ces effets, qui doivent être distingués des interactions entre élèves (c'est-à-dire entre pairs), incluent les pratiques pédagogiques et les dispositifs institutionnels d'aide.

11. Les pratiques pédagogiques. Ce facteur, dans le cas de l'enseignement supérieur fait l'objet d'un nombre de travaux limité en France. Duguet (2014)

⁸⁵ Les compétences socio-comportementales regroupent dans ce travail des facteurs positifs tels que le sentiment d'efficacité personnelle, l'estime de soi, la persévérance, le sentiment d'avoir une certaine capacité à influencer le cours des choses ainsi que les compétences sociales.

analyse ainsi, par exemple, de manière empirique les conséquences des pratiques pédagogiques des enseignants sur la scolarité des étudiants inscrits en première année universitaire en termes de motivation et de réussite. Suivant ce travail, la perception qu'ont les étudiants des pratiques des enseignants, exercerait un impact modéré sur les notes obtenues et sur la probabilité de passer en seconde année.

12. Les dispositifs institutionnels d'aide à la réussite. A titre d'exemple, dans le cas de la France, certains dispositifs d'aide à la réussite inclus dans le Plan Réussite en Licence (PRL), mis en place à partir de 2007 ont fait l'objet d'évaluation. Parmi les nombreux travaux de l'IREDU portant sur ce type de dispositif, nous pouvons mentionner le résultat suivant : les dispositifs PRL ont conduit à une diminution des risques d'abandon des étudiants les plus faibles en 1^{ère} année (Perret et al., 2014).

1.3 Les effets de pairs et la mixité sociale

La troisième série de facteurs de réussite universitaire documentée dans la littérature est constituée par les effets de pairs. Il s'agit de facteurs de contexte qui permettent une interaction entre les pairs et, potentiellement, des phénomènes d'influence réciproque. Le degré de mixité sociale d'une formation est susceptible d'avoir un impact sur la réussite universitaire via ce mécanisme d'effets de pairs que nous décrivons dans le présent paragraphe. Celui-ci est structuré de la manière suivante : en 1.3.1, nous définissons les effets de pairs, en 1.3.2, nous introduisons les modèles non linéaires d'effets de pairs et, en 1.3.3, nous présentons les principaux résultats des travaux empiriques sur le sujet.

1.3.1 Les effets de pairs directs et indirects

Les effets de pairs renvoient aux conséquences qui résultent des interactions entre les individus. Dans le cadre de l'enseignement supérieur, cela fait référence au processus par lequel les caractéristiques et les comportements des personnes qui suivent une même formation et partagent un environnement commun s'influencent mutuellement.

Si les effets de pairs renvoient initialement aux interactions entre pairs uniquement, il est possible d'élargir ce concept en tenant compte de l'interaction indirecte des pairs via des tiers. Suivant Sacerdote (2011), selon les personnes par lesquels transitent les effets de pairs, il est possible d'en distinguer trois types principaux :

- des effets directs passant par les interactions directes entre élèves, par exemple avec le partage de connaissances ou d'expériences.
- des effets indirects qui passent par le biais des interactions avec l'enseignant, de bons élèves pouvant par exemple conduire à une hausse globale des exigences.
- des effets indirects qui passent par les familles, par exemple si des parents d'élèves ont une influence sur le fonctionnement de l'établissement.

Sur un autre plan, il est possible de distinguer chez un individu ce qui tient des caractéristiques strictement personnelles (sa personnalité individuelle) de ce qui tient davantage de son environnement social. Dans ce cadre, Manski (1993) définit ainsi deux types d'effets de pairs :

1. les effets de pairs « endogènes » qui correspondent aux interactions directes et aux influences réciproques entre les pairs. Cela peut comprendre par exemple des phénomènes d'entraide ou de compétition, le partage de connaissances ou d'expériences.
2. les effets de pairs « exogènes » ou contextuels qui désignent la façon dont un individu est influencé indirectement par l'environnement socioculturel de ses pairs. Dans le contexte de travaux portant sur la mixité sociale, ce dernier point est crucial puisque l'individu est influencé par les caractéristiques exogènes de ses pairs. Le capital social ou culturel provenant de l'environnement d'un élève exerce potentiellement une influence sur les autres élèves.

De notre point de vue, les typologies de Sacerdote (2011) et de Manski (1993) ne s'opposent pas nécessairement, la première étant davantage un approfondissement de la seconde. Dans la pratique, les effets de pairs se traduisent par des mécanismes d'influence réciproque. Nous pouvons ainsi distinguer trois canaux de transmission principaux :

1. un canal psychologique : un des groupes sociaux présent peut devenir le groupe de référence, ce qui a des conséquences sur l'image de soi, les motivations, les projets des étudiants.
2. un canal de l'information : c'est le processus par lequel les élèves échangent entre eux des informations (ressources culturelles, expériences de vie, de voyages ou projets d'orientation).
3. un canal des pratiques pédagogiques (Ireson et Hallam, 2001) : les pratiques pédagogiques adoptées par les enseignants sont susceptibles de s'adapter et de changer en fonction de la composition scolaire et sociale des groupes. Cela concerne par exemple la gestion de la discipline, le degré de couverture des programmes ou encore le niveau d'exigence des corrections.

Ces trois canaux de transmission ne constituent pas des modèles d'effets de pairs au sens où à ce stade ils ne reposent sur aucune hypothèse concernant les groupes sociaux présents et leurs modes d'interaction. Il s'agit d'une description des processus formels par lesquels les effets de pairs peuvent potentiellement se traduire.

1.3.2 Les effets de pairs hétérogènes

Pour mesurer les effets de pairs, dans les modèles d'effets de pairs « linéaires en moyenne », comme expliqué par Monso et al. (2019), les résultats d'un élève dépendent de façon linéaire des caractéristiques moyennes des autres élèves. Ce type de modèles constitue un point de départ essentiel qui ne permet pas, toutefois, de rendre compte des effets de pairs lorsque ceux-ci sont « non-linéaires » ou « hétérogènes » :

- les effets de pairs sont « non-linéaires » (en référence à leur source principale) lorsque ceux-ci proviennent davantage d'un groupe particulier comme par exemple les élèves provenant d'un milieu social très favorisé. Les effets de pairs deviennent non linéaires notamment si l'un des groupes sociaux présents parvient à s'imposer comme le groupe de référence (il ne s'agit pas nécessairement du groupe le plus nombreux).
- les effets de pairs sont « hétérogènes » (en référence à leurs conséquences, aux individus principalement affectés) lorsque ceux-ci ont des conséquences plus prononcées pour un groupe d'élèves, par exemple les élèves les plus faibles.

Dans la pratique, les effets de pairs sont complexes, multiples et peuvent être aussi bien positifs que négatifs sur le plan de la réussite. Afin d'illustrer ces situations de non-linéarité, nous pouvons reprendre la typologie de Hoxby et Weingarh (2005) qui décrivent huit situations-types qui peuvent s'ajouter aux effets directs et indirects décrits au 1.3.1 de ce chapitre dans un contexte où les élèves sont distingués suivant leur niveau scolaire⁸⁶ :

1. le modèle de la « mauvaise pomme » (*bad apple model*), où les élèves les plus faibles ou les moins disciplinés influencent négativement les autres élèves, soit en les distrayant soit en captant en grande partie l'attention de l'enseignant.
2. le modèle dit du « rayon de lumière » (*shining light*), opposé au modèle antérieur, où des élèves brillants servent de référence pour le restant du groupe et ont un effet d'entraînement positif.

⁸⁶ Les modèles d'effets de pairs dans le domaine de l'éducation reposent principalement sur des exemples liés à la mixité scolaire. Toutefois, de notre point de vue, il est tout à fait possible d'adapter les situations décrites à des cas de mixité sociale, même si, bien entendu, il n'existe pas de correspondance directe entre groupes sociaux et groupes de niveau.

3. le modèle de la « comparaison peu flatteuse » (*invidious comparison*), dans lequel la présence d'élèves brillants est un facteur de découragement pour la majorité du groupe alors que la présence d'élèves faibles accroît l'auto-estime des autres élèves.
4. le modèle d'« entraînement par le haut » (*single crossing*) où les meilleurs élèves ont un effet bénéfique sur les autres élèves à condition que les écarts de niveaux ne soient pas trop importants : l'apport des meilleurs élèves est d'autant plus marqué que l'élève a lui-même un bon niveau scolaire.
5. le modèle dit de la « classe homogène » (*focus model*) dans lequel l'homogénéité des classes favorise la réussite, par exemple si le travail de l'enseignant est plus facile en raison de l'absence de différences majeures de niveaux des étudiants.
6. le modèle dit de la « boutique » (*boutique model*), dans lequel l'homogénéité des capacités des élèves favorise leurs interactions. Suivant ce modèle, un élève a surtout intérêt à être dans une classe plutôt homogène, même s'il est lui-même différent des autres.
7. le modèle de l'« arc-en-ciel » (*rainbow model*), à l'opposé du modèle précédent fait l'hypothèse que dans une classe hétérogène, les élèves bénéficient de la diversité.
8. le modèle de la sous-culture (*subculture model*), dans lequel les plus mauvais élèves deviennent le groupe de référence pour une partie de la classe, ceux-ci diffusant des valeurs distinctes, voire opposées à celles de la réussite scolaire.

1.3.3 Les résultats empiriques des travaux disponibles

D'un point de vue empirique, nous pouvons distinguer les travaux portant sur le secondaire où un groupe de pairs est identifiable de manière relativement aisée, de ceux portant sur l'enseignement supérieur où l'identification d'un groupe des pairs est plus complexe.

a) Résultats pour le secondaire

La littérature empirique portant sur les effets de pairs dans le secondaire aboutit à un consensus sur différents points. Une large majorité de travaux aboutit à la conclusion que les origines sociales ainsi que le niveau scolaire des pairs exercent une influence réelle mais limitée sur les résultats scolaires des élèves. Les travaux de Burke et Sass (2013) constatent que ces effets sont plus forts à l'échelle de la classe que de l'établissement. Un certain nombre de travaux portant sur la France aboutissent à la conclusion que les origines sociales des pairs ont des effets modestes sur les résultats scolaires. Sur la base du panel d'élèves qui sont entrés en CP en 1997, Davezies (2005) constate que lorsqu'un élève est scolarisé dans une école parmi les plus favorisées socialement, plutôt que dans les écoles les plus défavorisées alors ses résultats aux évaluations nationales de début de sixième augmentent d'environ 5 points sur 100 en mathématiques, et de 2 à 3 points en français. Cet effet semble plutôt modeste si l'on tient compte du fait qu'il se matérialise après cinq années de scolarisation. Piketty et Valdenaire (2006) obtiennent des résultats similaires sur les scores obtenus aux évaluations nationales de CE2 et jugent cet effet faible, en particulier en comparaison de celui d'une réduction de la taille des classes. Fougère et al. (2017) parviennent à la conclusion que la composition des classes de Collège a des effets modestes et parfois non-significatifs, que l'on tienne compte de la part de camarades de classe étrangers, de la proportion d'élèves provenant d'un milieu social très favorisé, ou du nombre de redoublants.

De nombreux chercheurs reconnaissent que le modèle linéaire en moyenne rend compte en général de façon très imparfaite des effets de pairs observés dans la réalité. Hoxby et Weingarth (2005) est sur ce plan l'un des principaux travaux de référence. D'autres travaux testent des non-linéarités ainsi que des effets hétérogènes, comme par exemple Burke et Sass (2008) qui constatent que les élèves peu performants bénéficient davantage des capacités élevées de leurs pairs que les élèves très performants. Le résultat constatant des effets modestes moyens occulterait ainsi une diversité d'effets suivant le niveau des élèves, les élèves ayant les plus faibles résultats scolaires ou ayant le capital scolaire le plus limité étant les plus sensibles à la composition de leur établissement ou de leur classe. Ainsi, Davezies (2005) aboutit à la conclusion qu'être scolarisé dans une école parmi les plus favorisées socialement, plutôt que parmi les plus défavorisées, produit un impact deux à trois fois supérieur

pour des élèves qui sont parmi les plus faibles que pour les plus performants à l'entrée au CP (entendus respectivement comme les élèves appartenant aux deux premiers déciles et aux deux derniers déciles des scores d'entrée au CP). Boutchenik et Maillard (2018) constatent que cette asymétrie est encore plus prononcée dans le cas des classes de terminale, l'effet d'une hausse du niveau scolaire des pairs étant cinq fois supérieure pour les élèves initialement les plus faibles (entendus comme les élèves appartenant au premier quartile des notes aux épreuves écrites du Diplôme National du Brevet, les élèves les plus forts appartenant au dernier quartile). A l'inverse, lorsque l'on examine les effets de pairs pour les élèves bénéficiant d'un bon niveau scolaire initial, les résultats sont beaucoup moins convergents. Boutchenik et Maillard (2018), constatent quant à eux que la présence de bons élèves dans une classe ne bénéficie pas forcément aux autres bons élèves de la classe.

b) Résultats pour l'enseignement supérieur

Les travaux d'évaluation portant sur les effets de pairs au sein de l'enseignement supérieur sont rendus plus complexes par les difficultés d'identifier un groupe de pairs, celui-ci pouvant varier suivant les points de vue adoptés. Sacerdote (2011) identifie pour les performances académiques mesurées à partir du GPA (*Grade Point Average*) des effets de pairs liés à la répartition aléatoire des étudiants par chambre au sein de « *Dartmouth College* »⁸⁷. En moyenne, si un étudiant se voit assigné un *roommate* faisant partie du quart des meilleurs étudiants à l'examen d'entrée de l'Université, alors ses propres résultats progressent d'en moyenne 6%. Si Foster (2006), dans une démarche très proche, n'identifie pas d'effets de pairs lié à la répartition aléatoire des étudiants par chambre dans le cas de l'Université du Maryland, Zimmerman (2004) identifie un effet négatif pour les étudiants qui se voient attribués un *roommate* dont les performances à l'examen d'entrée de l'Université se situent dans les 15% les plus basses. Hoel et al. (2005) identifient d'importants effets de pairs liés aux *roommate*, mais pas au niveau des classes. Ce résultat permet de faire l'hypothèse que l'intensité des relations entretenues avec les pairs peut être une condition essentielle pour observer des effets sur les résultats obtenus.

⁸⁷ Il s'agit d'une Université privée de la côté est des Etats-Unis qui est membre de l'*Ivy League*.

1.4 Evaluer le lien entre mixité sociale et réussite à l'Université

A partir d'une revue de la littérature effectuée dans les paragraphes antérieurs, il nous semble possible de définir une stratégie d'évaluation opérationnelle. Nous avons identifié une série de facteurs de réussite à l'Université pouvant être regroupés en facteurs individuels, facteurs de contexte et des effets de pairs. Nous présentons de manière synthétique dans le Tableau 12 ci-dessous l'usage qui est fait de ces facteurs dans une sélection de dix-huit articles récents portant sur les facteurs de réussite universitaire.

Tableau 12 : les facteurs de réussite à l'Université dans une sélection de travaux empiriques récents

	Origine sociale	Parcours scolaire	Origines géographiques	Genre	Logement	Revenu salarié	Intégration sociale	Capacités intellectuelles	Personnalité	Méthodes et temps de travail	Pratiques pédagogiques	Dispositifs d'aides	Effets de pairs linéaires	Effets de pairs non linéaires
Allen et Carter (2007)		X	X	X										
Beffy et al. (2009)						X								
Brinbaum et al. (2018)	X	X	X											
Carrell et al. (2009)													X	
Felouzis (2000)	X	X												
Foster (2006)													X	
Galand et al. (2005)	X	X				X	X	X	X	X	X			
Gibbons et al. (2016)													X	
Giret et al. (2019)						X	X		X		X	X		
Grignon et Gruel (1999)					X									

	Origine sociale	Parcours scolaire	Origines géographiques	Genre	Logement	Revenu salarie	Intégration sociale	Capacités intellectuelles	Personnalité	Méthodes et temps de travail	Pratiques pédagogiques	Dispositifs d'aides	Effets de pairs linéaires	Effets de pairs non linéaires
Grignon (2000)						X								
Hoel et al. (2005)													X	
Hoxby et al. (2005)														X
Lemaire (2000)				X										
Michaut (2012)				X										
Robbins et al.(2004)								X						
Sacerdote (2011)													X	
Zilloniz (2017)						X								

Source : synthèse de l'auteur sur la base des travaux mentionnés

Le Tableau 12 ci-dessus nous permet de constater que la plupart des travaux analysent la question de la réussite à l'Université en utilisant de un à trois facteurs. Seuls Galand et al. (2005) et Giret et al. (2019) répertorient un nombre plus important de facteurs, mais il s'agit précisément de travaux de synthèse basés sur des revues de la littérature.

Selon la littérature empirique, même les modèles les plus complets ne parviennent à rendre compte que d'une part limitée du phénomène de réussite à l'Université. Robbins et al. (2004) montrent que les travaux les plus complets réalisés dans des pays anglo-saxons ne parviennent à rendre compte que d'un maximum de 25% environ de la variance dans les résultats universitaires.

Il convient de rappeler, enfin, que le modèle linéaire en moyenne repose sur deux hypothèses fortes concernant la forme des effets de pairs. D'une part, les caractéristiques des pairs interviennent uniquement par la moyenne des comportements et des caractéristiques des pairs. D'autre part, les effets de pairs modifient la moyenne de la distribution de la variable d'intérêt.

Les modèles non linéaires montrent, toutefois, qu'il existe, d'abord, des effets de pairs non-linéaires : l'existence et l'intensité des effets de pairs dépendent par exemple d'un groupe en particulier et de la part occupée par celui-ci ou encore de l'hétérogénéité de la composition d'une unité. Des effets hétérogènes peuvent également avoir lieu dès lors que les effets de pairs moyens sont plus prononcés pour certains groupes d'étudiants.

Au regard de notre objectif qui est en général d'analyser les facteurs de réussite à l'Université et, en particulier, d'évaluer le rôle potentiel de la mixité sociale sur celle-ci, nous pouvons définir une stratégie opérationnelle à partir de la revue de la littérature effectuée dans cette section. Les modèles définis devront idéalement satisfaire ainsi aux exigences suivantes :

1. Utiliser les facteurs de réussite répertoriés sous réserve de la disponibilité des données afin d'obtenir un modèle aussi complet que possible. L'inconvénient de cette stratégie est qu'elle implique de connaître de manière fine les conditions de vie des étudiants et de disposer des variables requises.
2. Estimer un effet moyen de la mixité sociale sur la réussite universitaire à partir d'un modèle linéaire, en sachant que la littérature conclut généralement à l'existence d'effets ambigus ou limités.
3. Estimer des effets de pairs non linéaires, en tenant compte du fait que l'intensité des effets de pairs peut dépendre d'un groupe en particulier et de la part

occupée par celui-ci. De notre point de vue, ces effets non linéaires répertoriés dans la littérature concernent les variables explicatives. Par conséquent, les méthodes suivantes pourraient être mises en œuvre pour définir et tester celles-ci :

- Utiliser des seuils, en définissant par exemple une variable explicative de ségrégation « active » à partir d'un certain seuil.
 - Accorder une attention particulière aux extrêmes de la distribution en élevant par exemple des variables explicatives au carré.
 - Centrer l'analyse sur un groupe en particulier, par exemple le groupe défavorisé. Dans le cas de la mixité sociale, il est probable que ce dernier ait une importance particulière⁸⁸. Il conviendrait alors d'utiliser une ou des variables explicatives liées spécifiquement à celui-ci.
4. Tenir compte des effets hétérogènes, au-delà des effets moyens, en cherchant à évaluer s'il existe des effets différenciés de la mixité sociale sur la réussite suivant les groupes sociaux. De notre point de vue, ces effets hétérogènes documentés dans la littérature concernent les variables expliquées. Par conséquent, pour aller au-delà des effets moyens, les méthodes suivantes pourraient être mises en œuvre pour définir et tester celles-ci :
- Evaluer de manière spécifique la situation des étudiants défavorisés.
 - Evaluer s'il existe des effets différenciés de la mixité sociale suivant les parts occupées par les différents groupes au sein des formations : dans ce cas, l'estimation de l'effet moyen sur l'ensemble du groupe dépend du groupe social dominant au sein de la formation. L'effet moyen au sein d'une formation très ségréguée avec surreprésentation des défavorisés est ainsi distingué de l'effet moyen au sein d'une formation très ségréguée avec surreprésentation des très favorisés.
5. Pour l'interprétation des résultats, conserver à l'esprit le rôle central des variables corrélées. Comme le montre les travaux mentionnés, le rôle de

⁸⁸ Dans le cadre de la mixité scolaire, la littérature tend à montrer des effets plus robustes pour les élèves les moins bons que pour les meilleurs élèves. Ces effets plus robustes, dans le cas de la mixité sociale, pourraient potentiellement concerner davantage les étudiants défavorisés qui, suivant les travaux de Brinbaum et al. (2018) sont davantage affectés par l'échec à l'Université.

l'origine sociale comme facteur de réussite est limité dès lors que sont prises en compte d'autres caractéristiques individuelles comme le parcours scolaire antérieur. L'interprétation des résultats devra tenir compte du rôle potentiel des variables omises dans le modèle si certaines sont indisponibles.

2 L'appariement de deux bases de données pour appréhender de manière fine les caractéristiques individuelles des étudiants

Dans le cadre de la stratégie d'évaluation décrite ci-dessus, il est nécessaire de disposer simultanément d'informations détaillées sur les caractéristiques individuelles des étudiants ainsi que sur le degré de mixité sociale des formations suivies. A cette fin, nous utilisons la base panel bacheliers 2014 produite par le SIES qui présente l'avantage d'inclure de nombreuses informations sur les caractéristiques individuelles des étudiants. L'inconvénient de celle-ci au regard de notre stratégie d'évaluation est, toutefois, qu'elle ne contient pas d'éléments concernant le degré de mixité sociale des formations suivies. Pour remédier à ce problème, nous avons procédé à un appariement de celle-ci avec la base de données SISE qui contient précisément des informations relatives aux origines sociales des étudiants inscrits dans les différentes formations. La section 4 est structurée de la manière suivante. Le paragraphe 2.1 décrit le panel bacheliers 2014, le paragraphe 2.2 présente l'utilisation la base de données SISE Université faite ici et le paragraphe 2.3 explique comment les deux bases sont appariées afin de créer une base de données originale.

2.1 Appréhender les caractéristiques individuelles des étudiants

La stratégie d'évaluation définie dans la section 1 de ce chapitre pour rendre compte du phénomène de réussite à l'Université repose notamment sur l'utilisation de variables individuelles. Ceci nécessite de disposer d'une information fine et détaillée concernant les étudiants, leurs conditions de vie et leur réussite aux examens. Toutefois, la base de données SISE-Inscrits ne comprend pas ce type d'information. Sur l'ensemble des variables mentionnées ci-dessus, seules 5 sont disponibles dans celle-ci : l'origine sociale individuelle, le parcours scolaire antérieur (avec l'âge et la filière de bac), la nationalité, l'origine géographique et le genre.

Afin d'obtenir davantage d'informations sur le profil des étudiants, le recours à une nouvelle base de données est donc requis. En l'absence de base de données administrative exhaustive comportant les informations nécessaires (la base SISE est la seule base contenant des informations sur l'ensemble des étudiants dont dispose le

MESRI), une solution possible est de recourir à une base comprenant un échantillon représentatif. Or, le MESRI (SIES) a réalisé de 2015 à 2018 des enquêtes de suivi du parcours dans l'enseignement supérieur des bacheliers 2014. Cette base panel bacheliers 2014, qui comprend 23 200 élèves, est constituée de deux sous-ensembles : un échantillon de 18 200 jeunes bacheliers issus du panel des élèves entrés en 6^{ème} en 2007 ainsi qu'un second échantillon de 5 000 jeunes tirés au sort dans les fichiers du baccalauréat sur la base des critères suivants : formation et âge (bacheliers ayant 19 ans et plus). Ce second sous-ensemble est compris dans le panel afin de tenir compte des bacheliers ayant accumulé une ou plusieurs années de retard au Collège et de garantir ainsi la représentativité de cette cohorte. Ces derniers sont entrés en 6^{ème} avant 2007, mais ont malgré tout obtenu leur bac à la session 2014.

La base panel bacheliers 2014 présente plusieurs avantages dans le cadre de notre travail. En effet, celle-ci permet d'observer le parcours des bacheliers dans l'enseignement supérieur et de mettre en évidence les différents facteurs susceptibles d'influer sur celui-ci. Un certain nombre de questions sont destinées à appréhender les conditions de vie des étudiants : il s'agit en particulier du logement occupé, des ressources financières, des difficultés rencontrées, des relations sociales (amis, famille...). Les données d'enquête sont par ailleurs complétées par des données administratives (APB, série et notes du baccalauréat, bourse sur critères sociaux). Le panel contient des informations sur le passage en année supérieure à l'issue de la première année universitaire. Cela permet de distinguer les trajectoires d'études « linéaires » caractérisées par un temps de diplomation « normal » et les trajectoires d'études « non linéaires » (affectées par des situations de redoublement de réorientation et/ou d'abandon).

La base de données panel bacheliers 2014 présente, toutefois, deux inconvénients au regard de notre travail. D'abord, le taux de réponse des étudiants diminue fortement après la première année. C'est une des raisons pour lesquelles nous avons fait le choix de faire reposer notre analyse sur la première année universitaire uniquement (une autre raison est l'existence de niveaux de ségrégation particulièrement élevés en Licence 1). Ensuite, celle-ci ne contient pas d'informations relatives au degré de mixité sociale des formations suivies.

2.2 Appréhender les caractéristiques des établissements et des formations

Recourir à la base panel bacheliers 2014 uniquement ne fournit qu'une réponse partielle à notre objectif qui est en général de rendre compte des facteurs de réussite à l'Université et, en particulier, d'estimer le rôle potentiel de la mixité sociale. En effet, cette enquête ne fournit par elle-même aucune information sur ce dernier aspect. Une solution possible est alors de réutiliser les données issues de la base SISE exploitée dans le chapitre II de notre travail. Deux types d'informations disponibles pour chacune des formations répertoriées nous semblent réutilisables : les parts occupées par les différents groupes sociaux ainsi que la contribution à la mixité sociale de chacune d'entre elles.

La contribution à la mixité sociale d'une formation peut être estimée à partir de l'indice d'exposition normalisé, qui a été défini dans le chapitre I et utilisé aussi bien en raison de sa facilité d'interprétation que de l'absence de nécessité de décomposer les résultats à cette étape d'analyse. Pour mémoire l'indice d'exposition normalisé au niveau national (E) peut se définir comme suit :

$$E = \frac{1}{q_a(1 - q_a)} \sum_{k=1}^K \pi_k (q_a^k - q_a)^2 \quad (18)$$

la part q_a^k occupée par le groupe a au sein de chaque formation k est comparée à la part q_a occupée par ce groupe au sein de la population totale ; π_k désigne le poids de la formation k , c'est-à-dire la fraction du nombre total d'étudiants suivant la formation k .

Par construction, cet indice repose sur la somme pondérée de la contribution individuelle de chacune des formations universitaires que nous notons CK :

$$CK = (q_a^k - q_a)^2 \quad (19)$$

Nous retenons cette formule pour construire notre variable de contribution individuelle à la ségrégation pour les formations universitaires. L'hypothèse sous-jacente est que si, au sein d'une formation donnée, la part prise par un groupe social s'éloigne significativement de la moyenne existante au niveau national (parce qu'un groupe est surreprésenté ou sous-représenté), alors cela peut avoir un effet positif ou négatif sur la réussite des étudiants.

2.3 L'appariement des deux bases

Dans la mesure où la base SISE permet de connaître le degré de mixité sociale des formations suivies et que la base panel bacheliers 2014 fournit des informations sur les parcours individuels, une solution possible est de procéder à un appariement de ces deux bases. L'objectif de celui-ci est de connaître le degré de mixité sociale des formations (au sens d'une discipline SISE suivie au sein d'une Université donnée) dans lesquels sont inscrits les étudiants du panel. Ce processus d'appariement implique les étapes suivantes. Après élimination des observations dont la formation ne rentre pas dans le champ des disciplines SISE (absence de réponse, étudiants de BTS, etc.)⁸⁹, 8 223 observations ayant une discipline SISE renseignée sont conservées. Après élimination des étudiants qui ne répondent pas au 1^{er} mars 2016 à la question concernant leur « situation 2015 » (poursuite des études, redoublement, abandon, réorientation...), 7 627 observations sont finalement conservées.

Nous obtenons à l'issue de cette procédure d'appariement (rendue possible grâce à la discipline SISE et à l'identifiant de l'établissement présent dans les deux bases) une base de données originale concernant 7 627 observations et comprenant désormais à la fois des variables de nature individuelle et des informations relatives au degré de mixité des formations. Cet appariement n'est, toutefois, effectué que dans un seul sens, c'est-à-dire que celui-ci permet uniquement de rajouter à la base panel bacheliers 2014 des informations relatives au degré de mixité sociale de la formation suivie sur la base d'une information commune : la formation SISE suivie.

Il convient par ailleurs de noter qu'au final sur les 23 200 étudiants du panel bacheliers 2014, 32,9% des observations sont finalement retenues. Ce pourcentage implique que notre échantillon pourrait être affecté par un biais de représentativité. La part des bacheliers 2014 issus du panel qui accède à l'Université devrait, en effet, être similaire ou proche de celle qui est constatée au niveau national. Toutefois, suivant MESRI (2017), 41,3% de l'ensemble des bacheliers 2014 se sont inscrits à l'Université (ce taux comprend une part s'élevant à 7,5% de ces bacheliers inscrits en IUT, ces formations étant également comprises dans notre base)⁹⁰. Cet écart de près de 10%

⁸⁹ Les étudiants des IUT sont conservés, comme pour les analyses des chapitre II et III.

⁹⁰ Parmi les bacheliers, le taux d'inscription atteint 74,5% pour l'ensemble de l'enseignement supérieur.

entre le panel et les données exhaustives provenant du MESRI concernant le nombre de bacheliers inscrits dans une formation à l'Université est liée à l'absence de réponse au 1^{er} mars 2016 de la part de certains bacheliers concernant leur situation au cours de l'année universitaire 2014-2015. Cette absence de réponse peut elle-même résulter de phénomènes variés (interruption des études, déménagement...). Par conséquent, les individus qui répondent sont (probablement) davantage impliqués dans leurs études que ceux qui ne répondent pas, ce qui est susceptible de biaiser les résultats.

3 Réussir sa première année de Licence : un premier cadrage statistique

La base de données utilisée nous permet d'obtenir un premier cadrage statistique portant sur des variables essentielles. La section 3 est structurée de la manière suivante. Le paragraphe 3.1 présente les taux de passage en année supérieure en fonction de l'origine sociale, le paragraphe 3.2 décrit ce taux en fonction des disciplines suivies et le paragraphe 3.3 présente des statistiques relatives au travail étudiant qui est susceptible d'être utilisé comme variable explicative.

3.1 Taux de passage en année supérieure suivant l'origine sociale et discussion relative à la représentativité de la base panel bacheliers 2014

Le paragraphe 3.1 est structuré de la manière suivante : en 3.1.1, nous décrivons les principaux résultats obtenus concernant les taux de passage suivant l'origine sociale et, en 3.1.2, sur la base de ces résultats, nous discutons de la représentativité de l'échantillon analysé.

3.1.1 Les taux de passage en L2 suivant l'origine sociale

Les observations sélectionnées dans le panel qui correspondent au sein de celui-ci aux étudiants inscrits dans les Universités de France métropolitaine nous permettent, d'abord, de connaître les taux de réussite en première année de Licence à l'Université suivant l'origine sociale. Le Tableau 13 situé ci-dessous synthétise ainsi la situation des étudiants du panel bacheliers 2014 à la fin de la première année de Licence suivant leur origine sociale.

Tableau 13 : Situation des étudiants à la fin de la première année selon leur origine sociale pour l'échantillon issu du panel bacheliers 2014, année universitaire 2014-2015

Groupe social	Nombre d'étudiants	Passage	Redoublement	Abandon	Part dans la population totale
Ensemble	7 627	55,7%	42,8%	1,5%	100%
Très favorisés	2 792	58,8%	39,9%	1,3%	36,6%
Intermédiaires	4 327	55,1%	43,4%	1,5%	56,7%
Défavorisés	508	43,9%	54,3%	1,8%	6,7%

Champ : ensemble des étudiants inscrits à l'Université en France métropolitaine pour la première fois en première année de Licence. Note : ces données brutes non pondérées ne comprennent pas les individus qui au 1^{er} mars 2016 ne répondent pas à la question concernant leur « situation 2015 ».

Source : Panel de bacheliers 2014 (SIES-MESRI) – Diffuseur : SIES. Calculs de l'auteur.

Le Tableau 13 situé ci-dessus nous indique ainsi qu'en moyenne, sur la base des observations brutes non pondérées, pour l'ensemble des étudiants du panel inscrits à l'Université en France métropolitaine, 55,7% passent en seconde année de Licence. Toutefois, ces taux de passage sont différenciés en fonction de l'origine sociale. Ainsi, ce taux est nettement plus faible (43,9%) pour les étudiants défavorisés, alors qu'il est sensiblement plus élevé (58,8%) pour les étudiants très favorisés. Les étudiants défavorisés se caractérisent également par un plus fort taux d'abandon (1,8% versus 1,5% pour l'ensemble des étudiants).

3.1.2 Le problème des taux de passage dans le panel bacheliers 2014

Les résultats présentés ci-dessus nous conduisent à discuter de la représentativité de l'échantillon obtenu à partir de la base panel bacheliers 2014. En effet, le taux de passage de 55,7% en seconde année de Licence mentionné dans le Tableau 13 ci-dessus sur la base des observations brutes non pondérées (si l'on utilise des observations pondérées, ce taux atteint alors 54,2%) ne correspond pas à celui qui est officiellement diffusé par le service statistique du MESRI sur une base exhaustive. Suivant la note du SIES N°18 (MESRI, 2017), le taux de passage entre la L1 et la L2 est pour les néo-bacheliers 2014 inscrits pour l'année universitaire 2014-2015 de 40,8%. Il convient également de noter que les taux d'abandon reportés dans le Tableau 13 ne correspondent pas à ceux qui sont diffusés par le MESRI, qui est de 27,7% pour les néo-bacheliers 2014 à l'issue de l'année universitaire 2014-2015 (l'annexe 10 fournit une description statistique détaillée sur ce point). Une hypothèse

possible pour expliquer cet écart significatif est que les étudiants ayant abandonné leurs études universitaires sont surreprésentés parmi les non répondants au questionnaire en mars 2016.

Il convient de noter que l'échantillon obtenu à partir de la base panel bacheliers 2014 semble également affecté par un biais de représentativité concernant les parts occupées par les trois groupes sociaux identifiés. En effet, suivant les données exhaustives de la base SISE, les 454 049 étudiants de première année de Licence en 2014 se répartissaient par groupes sociaux de la manière suivante : 29,2% pour le groupe très favorisés, 7% pour le groupe favorisé, 29,9% pour le groupe intermédiaire et 33,8% pour le groupe défavorisé⁹¹. Si les deux groupes intermédiaires sont regroupés, conformément à la méthodologie décrite dans l'annexe 7, nous obtenons la répartition suivante pour la base SISE 2014 : 29,2% pour le groupe très favorisés, 36,9% pour le groupe intermédiaire et 33,8% pour le groupe défavorisé. Dans l'échantillon obtenu à partir de la base panel bacheliers 2014, les étudiants très favorisés sont ainsi surreprésentés alors que les étudiants défavorisés sont sous-représentés. Il convient de noter que ce biais de représentativité de la base panel bacheliers 2014 diminue toutefois lorsque seule la profession du premier parent est utilisée, comme dans la base SISE⁹². Nous obtenons ainsi la répartition suivante en utilisant pour notre échantillon uniquement la profession du premier parent : 29,9% pour le groupe très favorisés, 41,7% pour le groupe intermédiaire et 28,3% pour le groupe défavorisé. En utilisant la pondération, ces taux atteignent respectivement 29,4%, 41,9% et 28,7%, soit des niveaux proches de la base SISE pour le groupe très favorisé, mais inférieurs d'environ 5% pour le groupe défavorisé.

3.2 Les taux de passage en année supérieure suivant les disciplines suivies

Ce panorama général sur les différences de réussite en Licence selon l'origine sociale peut être affiné en fonction des disciplines d'inscription. Dans la mesure où les taux de passage en deuxième année ne sont pas équivalents dans toutes les disciplines,

⁹¹ Ces valeurs correspondent aux données de la base SISE sur la base du champ que nous avons retenu et défini dans le cadre du chapitre II.

⁹² Dans la base SISE, telle qu'elle est diffusée par l'Adisp-Progedo, seule la profession du « parent 1 » est renseignée.

l'existence de taux de passage différents suivant les origines sociales peut renvoyer aux choix de formation de la part des étudiants.

Le Tableau 14 ci-dessous décrit ainsi les taux de passage en deuxième année de Licence suivant les disciplines d'inscription.

Tableau 14 : Situation des étudiants à la fin de la première année de Licence suivant la discipline pour l'échantillon issu du panel bacheliers 2014, année universitaire 2014-2015

Discipline SISE	Passage	Redoublement	Abandon
S.ECO-GEST	73,1%	25,8%	1,1%
PL LET LAN	71,0%	29,0%	0,0%
SC FOND AP	69,2%	29,9%	0,9%
PL SCIENCE	68,3%	31,7%	0,0%
SC.HUM-SOC	63,1%	34,1%	2,8%
LETTR-ARTS	62,7%	34,2%	3,1%
SC.NAT-VIE	59,0%	41,0%	0,0%
LANGUES	56,5%	39,7%	3,8%
DROIT-SC.P	56,4%	42,5%	1,1%
STAPS	53,4%	45,1%	1,5%
A.E.S.	47,5%	49,8%	2,8%
MEDECINE	20,0%	80,0%	0,0%
PL SANTE	17,4%	82,2%	0,4%
Total	55,7%	42,8%	1,5%

Champ : ensemble des étudiants inscrits à l'Université pour la première fois en première année de Licence. Source : Panel de bacheliers 2014 (SIES-MESRI) – Diffuseur : SIES. Calculs de l'auteur.

Note : Classement par taux de passage décroissant.

Le Tableau 14 ci-dessus montre que les taux de passage varient très fortement suivant les disciplines avec un taux minimum de 17,4% pour la discipline « Pluri-santé » et un taux maximum de 73,1% pour la discipline « Sciences Economique gestion (hors A.E.S) ». Les statistiques disponibles en annexe 10 qui portent sur les néo-bacheliers 2014 entrés en première année de Licence en 2014-2015 montrent, toutefois, qu'il existe sur ce point également un écart significatif avec les données diffusées par le MESRI. Suivant celles-ci, pour la discipline « Sciences Economique gestion (hors A.E.S) », le taux de passage est ainsi de 38,3% pour l'année universitaire 2014-2015 (soit une différence de +34,8% pour le panel) alors que le taux d'abandon à l'issue de la première année de Licence est de 29,9% (soit une différence de -28,8% pour le panel); pour les langues, par exemple, ces taux atteignent 39,9% et 31% (soit pour le panel des différences de respectivement +16,6,8% et -27,2%).

3.3 Statistiques concernant le travail étudiant

Comme indiqué dans la section 1 du présent chapitre, le travail salarié étudiant est susceptible de constituer un facteur de réussite ou d'échec dans les études. Le Tableau 15 ci-dessous décrit ainsi la répartition des étudiants du panel bacheliers 2014 en fonction de leur origine sociale et du nombre d'heures travaillées.

Tableau 15 : répartition des étudiants par origine sociale et suivant l'existence d'un revenu salarié en 2014 (%)

Revenus salariés	Très favorisé	Intermédiaire	Défavorisé	Total 3 groupes
Supérieurs ou égaux à 15H	2,4%	3,2%	3,9%	2,9%
Inférieurs à 15H	4,3%	5,2%	5,1%	4,9%
Nuls	93,3%	91,6%	91,1%	92,2%
Total	100%	100%	100%	100%

Champ : ensemble des étudiants inscrits à l'Université pour la première fois en première année de Licence ; observations brutes non pondérées. Seules 7348 observations figurent dans le tableau car 279 étudiants présents dans le panel ne fournissent pas de réponse à ces questions.

Source : Panel de bacheliers 2014 (SIES-MESRI). Diffuseur : SIES. Calculs de l'auteur.

Le Tableau 15 ci-dessus indique que sur un total de 7 348 individus provenant du panel, 2,9% exercent un travail régulier de 15 heures ou plus par semaine et 4,9% exercent un travail régulier de moins de 15 heures hebdomadaires, 92,2% déclarant ne pas exercer d'activité salariée régulière depuis le mois de juillet 2014. Le travail salarié étudiant concerne les trois catégories d'origine sociale qui nous intéressent ici, même si en moyenne les très favorisés sont moins fréquemment concernés par le travail salarié.

Ces données sont plutôt cohérentes avec celles présentées dans Beffy et al. (2009) et issues de l'Enquête Emploi de l'INSEE sur la période 1992-2002 : sur un échantillon de 1 603 étudiants, 5,4% exercent un travail régulier de moins de 16 heures par semaine. Toutefois, l'estimation du travail salarié supérieur à 16 heures hebdomadaire y est plus importante que dans notre base de données, puisqu'il concerne 7,0% des étudiants, 87,4% étant non-salariés. Il convient, toutefois de noter que ces écarts peuvent s'expliquer par les périodes analysées qui sont distinctes (de même que les

bases utilisées) et par le fait que dans Beffy et al. (2009), l'échantillon concerné n'est pas restreint aux étudiants de première année.

4 Le lien entre réussite et caractéristiques individuelles : une première appréhension

Grâce à l'appariement effectué et à la base de données nouvellement constituée, nous réalisons des estimations économétriques permettant une première appréhension du lien entre caractéristiques individuelles et passage en deuxième année de Licence. La section 4 est structurée de la manière suivante : le paragraphe 4.1 présente les variables utilisées dans le modèle, le paragraphe 4.2 présente les résultats des estimations pour l'ensemble de l'échantillon et le paragraphe 4.3 les résultats concernant plus spécifiquement le groupe défavorisé.

4.1 Les variables utilisées

Grâce à l'appariement effectué entre les bases de données, les variables utilisées dans notre modèle ont deux sources : la base SISE pour le degré de mixité sociale des formations et le panel bacheliers 2014 pour les autres variables.

Concernant la variable expliquée, deux cas sont distingués : d'une part les étudiants qui abandonnent leurs études (la variable « SITUATION2015 » a la modalité « 6 ») ou redoublent (la variable « SITUATION2015 » est codée en « 2 » ou en « 3 ») et, d'autre part, ceux qui passent en deuxième année de Licence (la variable « SITUATION2015 » est codée en « 1 », « 4 », « 5 » ou « 7 »). Afin de prendre en compte cette situation, nous avons utilisé un modèle de régression de type logit dichotomique, c'est-à-dire un modèle dans lequel la variable expliquée ne peut prendre que deux modalités.

Les variables explicatives utilisées sont décrites ci-dessous :

1. L'origine sociale de l'étudiant. Conformément à la possibilité expliquée dans le cadre méthodologique défini dans le chapitre I, l'origine sociale est définie au moyen de la PCS du ou des parents référents.
 - La nomenclature des PCS n'est, toutefois, pas exactement la même dans la base de données SISE que dans le panel bacheliers 2014. La base de données panel bacheliers 2014 répertorie 13 PCS via les variables

« PROF_PARENT_1_2014 » et « PROF_PARENT_2_2014 ». Nous avons fait le choix de regrouper ces 13 professions en trois catégories sociales uniquement : défavorisé, intermédiaire et très favorisé. L'annexe 7 décrit la méthode utilisée pour effectuer ces regroupements.

- De plus nous avons vu dans le chapitre 1 que lorsque la PCS des deux parents était disponible, il était préférable de les utiliser toutes les deux. La PCS du second parent étant disponible dans la base de données panel bacheliers 2014⁹³, nous avons choisi d'utiliser les PCS des deux parents pour constituer nos trois groupes sociaux, suivant la méthodologie décrite dans l'encadré 4 inséré dans la section 1 du chapitre I.
 - Nous avons utilisé de manière complémentaire une variable relative au statut boursier de l'étudiant : la variable « AGLAE-BCS » indique si l'étudiant a bénéficié ou non d'une bourse sur critères sociaux. Il convient de distinguer ce type de bourse qui repose sur la situation financière de l'étudiant et de sa famille de l'aide au mérite, qui est un complément versé aux étudiants bénéficiaires d'une bourse sur critères sociaux sur la base de leurs résultats scolaires, et n'est pas prise en compte ici.
 - Afin de prolonger notre réflexion sur les étudiants dits de « première génération », tels que nous les avons définis dans le cadre du chapitre I, nous avons utilisé une variable reposant sur le niveau de diplôme des parents de l'étudiant. Ainsi, si aucun des parents de l'étudiant n'est diplômé de l'enseignement supérieur (variable « DIP_PARENT »), celui-ci est considéré comme un étudiant de « première génération ».
2. Le parcours scolaire (notes, filières, âge). Trois variables provenant du panel ont été utilisées pour prendre en compte ce facteur : la filière de bac suivie, la note obtenue à celui-ci ainsi qu'une variable de redoublement.
 3. Les origines géographiques. Afin de rendre compte de cette dimension, nous avons utilisé une variable disponible dans les données du panel indiquant si l'étudiant a des origines non métropolitaines.

⁹³ Pour rappel, celle-ci n'est pas disponible dans la base SISE.

4. Genre. Nous avons utilisé une variable disponible dans les données du panel indiquant si l'étudiant est un homme.
5. Logement. Afin de rendre compte de cette dimension, nous avons utilisé une variable disponible dans les données du panel indiquant si l'étudiant réside principalement au domicile d'un ascendant.
6. Revenus. Afin de tenir compte de cette dimension, nous avons utilisé les variables qui indiquaient si l'étudiant disposait d'un travail salarié régulier depuis le mois de juillet 2014 comme source de revenu.
7. Intégration sociale. Il n'existe pas à notre connaissance dans le panel bacheliers 2014 de variable susceptible de rendre compte de cette dimension.
8. Capacités intellectuelles. Cette variable n'est pas disponible dans le panel au sens où celui-ci ne contient pas, par exemple, de données sur le quotient intellectuel des élèves.
9. Personnalité (Motivation, confiance en soi, stabilité). Afin de rendre compte de cette dimension, nous avons utilisé deux variables :
 - Une première variable indique si l'étudiant a choisi sa formation sans motivation particulière (« au hasard ») ou ne répond pas à la question concernant les raisons de son choix.
 - Une seconde variable indique si l'étudiant rencontre des « problème familiaux » au cours de l'année universitaire.
10. Méthodes et temps de travail. Afin de rendre compte de cette dimension essentielle, nous avons cherché à prendre en compte les facteurs suivants :
 - Une variable qui indique si l'étudiant est absent « souvent » à ses cours.
 - Une variable qui indique si l'étudiant a « de grandes difficultés » ou « de très grandes difficultés » pour s'organiser dans son travail universitaire.
 - Le nombre d'heures de travail personnel hebdomadaire pendant la semaine et le week-end n'a pas pu être pris en compte, toutefois, car cette variable n'est disponible que pour la seconde année de Licence.

11. Pratiques pédagogiques. Cette dimension nous semble constituer potentiellement un facteur important de réussite. Des variables comme la taille des groupes, le volume horaire de cours, la fréquence des examens ou encore les méthodes pédagogiques utilisées sont, en effet, susceptibles de jouer un rôle significatif pour la réussite des étudiants. Toutefois, à notre connaissance, le panel bacheliers 2014 ne contient pas de variable concernant ce domaine.
12. Dispositifs d'aides institutionnels à la réussite (accompagnement, suivi pédagogique individualisé des étudiants...). Cette variable n'est pas disponible en tant que telle dans le panel.
13. Effets de pairs.
 - Différents types d'effets de pairs pourraient théoriquement être pris en compte dans notre modèle. Au-delà de l'origine sociale des pairs de l'étudiant, nos estimations pourraient également chercher à évaluer les effets potentiels d'autres caractéristiques des pairs, comme leur niveau scolaire ou leur assiduité (absentéisme) en cours. Toutefois, afin de pouvoir estimer ces effets, il serait nécessaire de connaître les caractéristiques sur ces plans de chaque formation dans lesquelles sont inscrits les étudiants du panel. Or, la base SISE ne permet de caractériser les formations que sur le plan de la mixité sociale⁹⁴. Par conséquent, l'estimation que nous réalisons des effets de pairs potentiels repose uniquement sur le degré de mixité sociale des formations suivies au sens où nous les avons définies dans les chapitres II et III.
 - Afin de tenir compte plus spécifiquement de la dimension potentiellement « non linéaires » de ces effets, plusieurs possibilités sont envisageables, comme détaillé ci-dessus dans la conclusion de la section 1 : définir une variable explicative de ségrégation « active » à partir d'un certain seuil, accorder une attention particulière aux extrêmes de la distribution en élevant par exemple les variables explicatives de mixité sociale au carré ou centrer l'analyse sur un groupe en particulier, par exemple le groupe défavorisé.

⁹⁴ Les 7 267 étudiants du panel retenus dans notre échantillon se répartissent dans 598 formations, soit environ 12,7 étudiants par formation. Il ne serait pas possible, aussi, de caractériser chacune de ses formations sur le plan de l'absentéisme ou de la réussite scolaire uniquement sur la base du panel.

- Trois types de variables explicatives de mixité sociale sont ainsi utilisées dans notre modèle : la part occupée par les groupes très favorisés et défavorisés, la contribution individuelle des formations (calculée pour les très favorisés et les défavorisés) et une variable reposant sur la contribution individuelle des formations, mais fixée à un seuil minimal permettant de caractériser une formation très ségréguée. Celui-ci a été fixé à cinq fois la médiane des contributions individuelles des formations. Ce seuil, qui permet de tenir compte des effets potentiellement non linéaires de la mixité sociale, est calculé pour le groupe défavorisé qui, au regard des analyses antérieures est particulièrement susceptible d'être concerné par les effets non linéaires.

4.2 Les estimations réalisées pour l'ensemble des étudiants

Nous avons utilisé afin de réaliser nos estimations une régression logistique (un modèle logit) dichotomique. L'un des principaux avantages de ce type de modèle est qu'il permet de simuler directement la probabilité de passage d'un étudiant à partir des caractéristiques de son profil. Les résultats des estimations effectuées à partir de notre modèle de type logit dichotomique sont présentés dans le Tableau 16 ci-dessous.

Tableau 16 : Modélisation (logit dichotomique) du taux de passage en deuxième année des étudiants du panel à partir de leurs caractéristiques individuelles ainsi que celles des formations suivies

	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
Constante	-4,45 (0,36) ***	-4,49 (0,33) ***	-4,29 (0,38) ***	-4,47 (0,33) ***	-4,48 (0,33) ***
Caractéristiques individuelles					
Très favorisé	0,14 (0,11) NS	0,14 (0,11) NS	0,13 (0,11) NS	0,14 (0,11) NS	0,14 (0,11) NS
Autre	Réf.	Réf.	Réf.	Réf.	Réf.
Défavorisé	-0,28 (0,13) *	-0,27 (0,13) *	-0,27 (0,13) *	-0,27 (0,13) *	-0,28 (0,13) *
Genre ("Garçon")	0,01 (0,09) NS				
Note au Bac	0,39 (0,02) ***	0,39 (0,02) ***	0,39 (0,02) ***	0,39 (0,02) ***	0,39 (0,02) ***
Filière bac professionnel	-1,14 (0,10) ***	-1,15 (0,10) ***	-1,13 (0,10) ***	-1,14 (0,10) ***	-1,14 (0,10) ***
Revenus salariaux	-0,27 (0,15) **	-0,27 (0,15) **	-0,27 (0,15) **	-0,27 (0,15) **	-0,27 (0,15) **
Redoublement secondaire	-0,15 (0,12) NS				
Origine hors Métropole	-0,34 (0,33) NS	-0,34 (0,33) NS	-0,33 (0,33) NS	-0,34 (0,33) NS	-0,35 (0,33) NS
Domicile parental	-0,25 (0,08) ***	-0,24 (0,08) ***	-0,25 (0,08) ***	-0,25 (0,08) ***	-0,25 (0,08) ***
Faible motivation	-0,37 (0,14) ***	-0,37 (0,14) ***	-0,37 (0,14) ***	-0,37 (0,14) ***	-0,37 (0,14) ***
Absentéisme	-0,47 (0,17) ***	-0,46 (0,17) ***	-0,46 (0,17) ***	-0,47 (0,17) ***	-0,47 (0,17) ***
"Première Génération"	0,06 (0,09) NS	0,06 (0,09) NS	0,07 (0,09) NS	0,06 (0,09) NS	0,06 (0,09) NS
Boursier	0,25 (0,11) *	0,25 (0,11) *	0,24 (0,11) *	0,24 (0,11) *	0,24 (0,11) *
Difficultés d'organisation	-0,63 (0,10) ***	-0,63 (0,10) ***	-0,63 (0,10) ***	-0,63 (0,10) ***	-0,63 (0,10) ***
"Problèmes familiaux"	-0,12 (0,12) NS				
Caractéristiques des formations suivies					
Part d'étudiants très favorisés	-0,16 (0,54) NS				
Contribution individuelle très favorisés		-4,41 (3,39) NS			
Part d'étudiants défavorisés			-0,56 (0,49) NS		
Contribution individuelle défavorisés				-3,3 2,48 NS	
Formation très ségréguée défavorisés					-0,3 0,19 NS
Contrôle par discipline (a)	OUI	OUI	OUI	OUI	OUI
D de Somers	0,624	0,624	0,624	0,624	0,624
Pourcentage concordant	81,2	81,2	81,2	81,2	81,2
Poucentage discordant	18,8	18,8	18,8	18,8	18,8
Pourcentage lié	0,0	0,0	0,0	0,0	0,0

Champ : échantillon de 7 627 observations provenant du panel bacheliers 2014 (étudiants inscrits à l'Université pour la première fois en première année de Licence).

Note : ref. pour modalité de référence, (a) La discipline de référence est le « droit », *** significatif au seuil de 1%, ** significatif au seuil de 5%, * significatif au seuil de 10%, NS : non significatif.

Source : Panel de bacheliers 2014 (SIES-MESRI) – Diffuseur : SIES. Calculs de l'auteur.

A partir du résultat des estimations des différents modèles, il est possible de calculer pour chaque modèle la probabilité de passage d'un étudiant en fonction de ses caractéristiques individuelles et de celles de sa formation. Le Tableau 17 ci-dessous illustre le calcul de cette probabilité de passage à partir de trois cas-types pour le modèle n°1.

Tableau 17 : Probabilité calculée de passage sur la base du modèle n°1 suivant trois cas-types

	Modèle 1	Cas-Type n°1	Cas-Type n°2	Cas-Type n°3
Constante	-4,45 (0,36) ***	1,00 -(4,45)	1,00 -(4,45)	1,00 -(4,45)
Caractéristiques individuelles				
Très favorisé	0,14 (0,11) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Autre	Réf.			
Défavorisé	-0,28 (0,13) *	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Genre ("Garçon")	0,01 (0,09) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Note au Bac	0,39 (0,02) ***	12,00 (4,68)	12,00 (4,68)	14,00 (5,46)
Filière bac professionnel	-1,14 (0,10) ***	0,00 (0,00)	1,00 -(1,14)	0,00 (0,00)
Revenus salariaux	-0,27 (0,15) **	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Redoublement secondaire	-0,15 (0,12) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Origine hors Métropole	-0,34 (0,33) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Domicile Parental	-0,25 (0,08) ***	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Faible Motivation	-0,37 (0,14) ***	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Absentéisme	-0,47 (0,17) ***	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
"Première Génération"	0,06 (0,09) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Boursier	0,25 (0,11) *	0,00 (0,00)	0,00 (0,00)	1,00 (0,25)
Difficultés d'organisation	-0,63 (0,10) ***	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
"Problèmes familiaux"	-0,12 (0,12) NS	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Caractéristiques des formations suivies				
Part d'étudiants très favorisés	-0,16 (0,54) NS			
Probabilité de passage		55,6%	28,6%	77,8%

Champ : échantillon de 7 627 observations provenant du panel bacheliers 2014 (étudiants inscrits à l'Université pour la première fois en première année de Licence).

Note : *ref.* pour modalité de référence, *** significatif au seuil de 1%, ** significatif au seuil de 5%, * significatif au seuil de 1%, NS : non significatif.

Source : Panel de bacheliers 2014 (SIES-MESRI) – Diffuseur : SIES. Calculs de l'auteur

Le Tableau 17 ci-dessus montre ainsi que pour un cas de référence qui correspond à une étudiante inscrite en droit issue du groupe social intermédiaire ayant une note de 12 au bac, la probabilité de passage à l'issue de la première année de Licence est de 55,6% (cas-type n°1). Toutefois, si à partir de ce cas-type, on introduit le fait que cette même personne a suivi une filière de bac professionnel ou technologique, alors cette probabilité diminue à 28,6% (cas-type n°2). A l'inverse, si, à partir du cas-type n°1, la note obtenue au bac passe à 14 et que cette personne est bénéficiaire d'une bourse, la probabilité de passage augmente à 77,8% (cas-type n°3).

Comme expliqué antérieurement, les modèles estimés reposent sur l'utilisation de deux types de variables explicatives : celles relatives aux caractéristiques individuelles des étudiants et celles qui concernent le degré de mixité sociale des formations suivies. Les 7 modèles estimés sont identiques, la seule différence entre eux reposant sur l'utilisation d'une variable de mixité sociale distincte dans chacun d'entre eux.

a) Résultats liés aux caractéristiques individuelles des étudiants

Les modèles utilisés reposent sur l'utilisation de quinze variables explicatives liées aux caractéristiques individuelles des étudiants. Dans chacun de ces modèles, six variables explicatives liées aux caractéristiques individuelles des étudiants ne sont pas significatives. Il s'agit des variables suivantes :

- l'origine sociale très favorisée.
- le genre, ce qui confirme dans une certaine mesure les travaux de Michaut (2012) suivant lesquels cette variable n'est pas significative dès lors que des variables de contrôle liées aux pratiques de travail et au temps consacré aux études sont introduites.
- le redoublement dans le secondaire.
- le fait d'être un étudiant dit de « première génération ».
- l'origine géographique non métropolitaine.

- l'existence de « problèmes familiaux ».

L'absence de significativité de ces quatre dernières variables pourrait être liée au fait que nous contrôlons par le niveau scolaire (la note obtenue au bac) sur lesquelles ces variables influent probablement.

Les variables explicatives de type individuel suivantes, en revanche, sont significatives au seuil de 5% ou 10% :

- l'origine sociale défavorisée, pour laquelle le coefficient d'estimation est négatif et significatif au seuil de 10%. Suivant notre modèle logit dichotomique, par rapport à un cas-type de référence (une étudiante inscrite en droit issue du groupe social intermédiaire ayant une note de 12 au bac), le fait d'être issu d'un milieu social défavorisé réduit la probabilité de passage (-6,9% suivant le modèle n°1). Ce résultat indique que l'origine sociale défavorisée pourrait réduire de manière limitée la probabilité de passage en deuxième année de Licence.
- l'existence de revenus salariaux, pour laquelle le coefficient d'estimation est négatif et significatif au seuil de 5%. Par rapport au cas-type de référence, le fait d'avoir des revenus salariaux réduit la probabilité de passage (-6,7% suivant le modèle n°1). Ce résultat va dans le même sens que les travaux de Beffy et al. (2009) et de Zilloniz (2017), même si nous n'opérons pas de distinction fine des types d'emplois en fonction de leur durée hebdomadaire.
- le statut de boursier, pour laquelle le coefficient d'estimation est positif et significatif au seuil de 10%. Par rapport au cas-type de référence, le fait d'être boursier augmente la probabilité de passage (+5,9% suivant le modèle n°1). L'effet positif d'être boursier ne semble pas intuitif à première vue. Toutefois, notre modèle contrôle le fait d'être issu d'un milieu social défavorisé. Dans ce cas, cet effet positif pourrait relever d'une baisse de la contrainte financière qui pèse sur les étudiants et leur permettrait de mieux se consacrer à leurs études.

Les variables explicatives de type individuel suivantes sont significatives au seuil de 1% :

- la note au baccalauréat, pour laquelle le coefficient d'estimation est positif. Par rapport au cas-type de référence, le fait d'avoir eu 14 au lieu de 12 au bac augmente la probabilité de passage (+17,6% suivant le modèle n°1). Ce résultat est cohérent avec ceux présentés dans les travaux de référence dans la section 1 du présent chapitre.
- le fait d'avoir suivi une filière de baccalauréat professionnel ou technologique, pour laquelle le coefficient d'estimation est négatif. Par rapport au cas-type de référence, le fait d'avoir suivi l'une de ces filières de bac réduit la probabilité de passage (-27% suivant le modèle n°1). Ce résultat est à nouveau cohérent avec ceux présentés dans les travaux de référence dans la section 1 du présent chapitre.
- le fait de résider chez un ascendant, pour laquelle le coefficient d'estimation est négatif. Par rapport au cas-type de référence, le fait de résider chez un ascendant réduit la probabilité de passage (-6,3% suivant le modèle n°1). Ce résultat confirme ceux de Grignon et Gruel (2000) présentés dans la section 1.
- le fait d'être faiblement motivé, pour laquelle le coefficient d'estimation est négatif. Par rapport au cas-type de référence, le fait d'être faiblement motivé réduit la probabilité de passage (-9,2% suivant le modèle n°1). Ce résultat va dans le même sens que les travaux de Lambert-Lemener (2012) présentés dans la section 1 du présent chapitre.
- l'absentéisme, pour laquelle le coefficient estimé est négatif. Par rapport au cas-type de référence, le fait d'être souvent absent en cours réduit la probabilité de passage (-11,6% suivant le modèle n°1).
- l'existence de difficultés d'organisation, pour laquelle le coefficient d'estimation est négatif. Par rapport au cas-type de référence, le fait d'avoir des difficultés d'organisation réduit la probabilité de passage (-15,6% suivant le modèle n°1).

b) Résultats liés à la mixité sociale des formations suivies

Les coefficients de corrélation entre les variables de mixité sociale sont élevés et supérieurs aux seuils communément acceptés pour les variables explicatives. Ainsi ce coefficient est de -0,82 entre les parts d'étudiants très favorisés et défavorisés. Afin d'éviter des problèmes de multicolinéarité, les variables explicatives de mixité sociale ont, aussi, été employées de manière séquentielle.

Conformément aux conclusions issues de la section 1, nous avons recentré notre analyse sur les groupes situés aux extrêmes de la distribution avec les groupes très favorisés et défavorisés. Nous avons ainsi utilisé comme variable explicative les parts occupées par les étudiants issus de ces deux groupes dans les modèles 1 et 3. Toutefois, dans les deux cas, les coefficients estimés ne sont pas significatifs, même au seuil de 10%.

La contribution des formations à la ségrégation telle que définie dans le paragraphe 2.2 ci-dessus a été utilisée dans les modèles 2 et 4, mais aucun effet significatif n'a pu être identifié.

Pour tenir compte d'effets non linéaires potentiels, conformément aux recommandations formulées dans une partie de la littérature sur le sujet, nous avons cherché à estimer un effet de seuil potentiel. Ainsi, la variable « formation très ségréguée défavorisés » permet de sélectionner les formations pour lesquels ce niveau est supérieur à cinq fois la médiane. Aucun effet significatif n'a pu, toutefois, être identifié en utilisant cette méthode.

Différentes interprétations sont susceptibles d'éclairer la mise au jour d'effets non significatifs du degré de mixité sociale des formations suivies sur les taux de passage en année supérieure :

- Une première interprétation est d'ordre méthodologique et repose sur le nombre limité d'observations disponibles pour réaliser nos estimations et des problèmes de représentativité de la base panel bacheliers 2014 identifiés dans les sections 2 et 3 du présent chapitre. Davantage d'observations seraient nécessaires afin de mettre potentiellement en évidence des effets significatifs de la mixité sociale sur le passage en année supérieure.

- Une deuxième interprétation repose sur la comparabilité des formations identifiées. Nous avons tenu compte de l'existence de taux de passage hétérogènes entre disciplines en contrôlant nos régressions par un effet discipline. Il est toutefois, possible qu'au sein d'une même discipline, les degrés d'exigence et les probabilités de passage en année supérieure ne soient pas homogènes, indépendamment des effets liés à la mixité sociale. Ce facteur résulte de l'absence d'examen national homogène, les modalités de passage en année supérieure étant définie au niveau de chaque Université.

- Une troisième interprétation possible, selon nous, résulte de la difficulté à identifier des groupes de pairs à l'Université. En ligne avec les conclusions de la littérature, la mise au jour d'effets de pairs est plus complexe dans l'enseignement supérieur que dans le secondaire. L'identification d'un groupe de pairs qui pourrait influencer la réussite étudiante est moins immédiate et l'intensité des effets peut aussi être variable.

- Une quatrième interprétation possible est issue des conclusions présentées dans la section 1 du présent chapitre concernant l'importance des effets hétérogènes. Selon celle-ci, le degré de mixité sociale pourrait avoir des effets différenciés suivant les groupes sociaux. Afin de tester la validité de cette hypothèse, une solution possible est d'analyser la probabilité de passage en année supérieure non plus pour l'ensemble des étudiants, mais pour ceux issus d'un groupe social en particulier.

4.3 Estimations réalisées pour les étudiants défavorisés

Afin de tenir compte des effets potentiellement hétérogènes de la mixité sociale, nous avons choisi de recentrer notre analyse sur le groupe défavorisé car suivant la revue de la littérature effectuée, ce groupe serait le plus susceptible d'être concerné par les effets du degré de mixité sociale des formations suivies⁹⁵.

Les résultats des estimations effectuées pour les 5 modèles de type logit dichotomique pour les étudiants défavorisés sont présentés dans le Tableau 18 ci-dessous.

⁹⁵ Des estimations similaires ont été réalisées pour les groupes « très favorisés » et « intermédiaires », mais celles-ci n'ont pas permis de mettre au jour d'effets significatifs des variables concernant la mixité sociale des formations.

Tableau 18 : Modélisation du taux de passage en deuxième année des étudiants défavorisés (logit dichotomique) à partir de leurs caractéristiques individuelles ainsi que celles des formations suivies

	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
Constante	-8,19 (1,45) ***	-6,84 (1,30) ***	-5,71 (1,38) ***	-6,82 (1,30) ***	-6,82 (1,30) ***
Caractéristiques individuelles					
Genre ("Garçon")	-0,58 (0,30) *	-0,58 (0,30) *	-0,59 (0,30) *	-0,57 (0,30) *	-0,58 (0,30) *
Note au Bac	0,46 (0,08) ***	0,45 (0,08) ***	0,46 (0,08) ***	0,45 (0,08) ***	0,45 (0,08) ***
Filière bac professionnel	-1,17 (0,29) ***	-1,23 (0,28) ***	-1,17 (0,29) ***	-1,24 (0,28) ***	-1,22 (0,28) ***
Revenus salariaux	-0,43 (0,50) NS	-0,42 (0,51) NS	-0,41 (0,50) NS	-0,41 (0,50) NS	-0,42 (0,50) NS
Redoublement secondaire	0,32 (0,31) NS	0,26 (0,31) NS	0,32 (0,31) NS	0,26 (0,31) NS	0,26 (0,31) NS
Origine hors Métropole	0,67 (1,39) NS	0,96 (1,42) NS	0,87 (1,36) NS	0,97 (1,42) NS	0,94 (1,42) NS
Domicile parental	-0,67 (0,29) **	-0,67 (0,29) **	-0,60 (0,29) **	-0,65 (0,28) **	-0,65 (0,28) **
Faible motivation	0,32 (0,50) NS	0,26 (0,48) NS	0,33 (0,50) NS	0,27 (0,49) NS	0,30 (0,49) NS
Absentéisme	-1,04 (0,63) *	-1,02 (0,62) *	-1,06 (0,63) *	-1,02 (0,62) *	-1,02 (0,62) *
"Première Génération"	0,48 (0,49) NS	0,37 (0,49) NS	0,43 (0,49) NS	0,39 (0,49) NS	0,40 (0,49) NS
Boursier	1,61 (0,67) **	1,62 (0,65) **	1,62 (0,66) **	1,62 (0,65) **	1,65 (0,65) **
Difficultés d'organisation	-1,15 (0,33) ***	-1,10 (0,33) ***	-1,17 (0,33) ***	-1,10 (0,33) ***	-1,11 (0,33) ***
"Problèmes familiaux"	0,67 (0,41) NS	0,61 (0,40) NS	0,65 (0,41) NS	0,62 (0,40) NS	0,61 (0,40) NS
Caractéristiques des formations suivies					
Part d'étudiants très favorisés	4,40 (1,80) **				
Contribution individuelle très favorisés		8,28 (13,26) NS			
Part d'étudiants défavorisés			-3,40 (1,57) **		
Contribution individuelle défavorisés				0,49 (7,83) NS	
Formation très ségréguée défavorisés					-0,4 (0,56) NS
Contrôle par discipline (a)	OUI	OUI	OUI	OUI	OUI
D de Somers	0,71	0,703	0,712	0,702	0,704
Pourcentage concordant	85,5	85,2	85,6	85,1	85,2
Poucentage discordant	14,5	14,8	14,4	14,9	14,8
Pourcentage lié	0,0	0,0	0,0	0,0	0

Champ : échantillon de 508 observations provenant du panel bacheliers 2014 (étudiants défavorisés inscrits à l'Université pour la première fois en première année de Licence).

Note : ref. pour modalité de référence, (a) La discipline de référence est le « droit », *** significatif au seuil de 1%, ** significatif au seuil de 5%, * significatif au seuil de 10%, NS : non significatif.

Source : Panel de bacheliers 2014 (SIES-MESRI) – Diffuseur : SIES. Calculs de l'auteur.

a) Les résultats liés aux caractéristiques individuelles des étudiants

Concernant les variables liées aux caractéristiques individuelles des étudiants, comme le montre le Tableau 18 ci-dessus, il existe plusieurs différences par rapport aux résultats antérieurs portant sur l'ensemble des étudiants :

- La variable explicative de genre devient significative au seuil de 10% alors que celle-ci ne l'était pas auparavant, ce qui indique que ce facteur pourrait jouer un rôle dans le cas particulier des étudiants défavorisés, mais pas de manière générale.
- Deux variables explicatives qui étaient significatives pour l'ensemble des étudiants de l'échantillon ne le sont plus pour les étudiants défavorisés : les revenus salariaux et le niveau de motivation.
- Pour deux variables explicatives, le niveau de significativité a diminué : le fait d'habiter au domicile parental et l'absentéisme.
- Pour une variable explicative, le niveau de significativité a augmenté : le fait d'être boursier.

Si certaines de ces différences peuvent être difficiles à interpréter, il nous semble que cela met au jour l'existence de spécificités du groupe d'étudiants défavorisés par rapport à l'ensemble des étudiants sur le plan des variables explicatives individuelles.

b) Résultats liés aux caractéristiques des formations suivies sur le plan de la mixité sociale

En ce qui concerne les variables explicatives liées à la mixité sociale des étudiants favorisés, la variable liée à la part occupée par les étudiants issus du groupe très favorisé est devenue significative au seuil de 5% alors que ce n'était pas le cas dans les résultats qui concernaient l'ensemble des étudiants. Par rapport au cas d'une

étudiante boursière inscrite en droit (issue du groupe social défavorisé) ayant par exemple une note de 12 au bac, si la proportion d'étudiants très favorisés dans la formation dans laquelle l'étudiante est inscrite passe, par exemple, de 30% à 40% alors la probabilité de passage de l'étudiante passe de 55,3% à 65,7% (+10,4%).

Ce résultat peut être rapproché de ceux présentés plus haut en 1.3.3, qui concernent directement le niveau scolaire des élèves, suivant lequel les étudiants ayant le capital scolaire le plus limité ou le niveau le plus faible sont les plus sensibles à la composition de leur établissement ou de leur classe, ce qui peut être occulté dans les résultats d'ensemble (Burke et Sass, 2008 ; Davezies, 2005 ; Boutchenik et Maillard, 2018). Suivant nos estimations, la part d'étudiants très favorisés dans une formation ne serait pas significativement corrélée à la probabilité de passage des étudiants de manière générale, mais en revanche, celle-ci le serait dans le cas particulier des étudiants défavorisés.

Les éléments théoriques présentés en 1.3.1 et en 1.3.2 sont susceptibles d'être mobilisés pour interpréter ces résultats. Le modèle n°4 dit « d'entraînement par le haut » pourrait potentiellement être utilisé pour interpréter nos résultats si l'on faisait par exemple l'hypothèse que le niveau universitaire des étudiants très favorisés est en moyenne plus élevé que celui des étudiants défavorisés⁹⁶. Toutefois, il convient de noter que le résultat de notre estimation ne permet pas, en l'état actuel, de se prononcer en faveur d'un ou de plusieurs modèles d'interprétation théorique avec certitude. Le résultat de notre estimation concernant la part occupée par les étudiants favorisés est ainsi théoriquement compatible avec le modèle n°7 de l' « arc-en-ciel » si la hausse de la probabilité de passage des étudiants défavorisés résulte par exemple du fait qu'ils ont de moins bons résultats dans les formations les plus ségréguées où les étudiants défavorisés sont surreprésentés et les étudiants très favorisés également sous-représentés⁹⁷.

⁹⁶ Comme indiqué dans le Tableau 13 ci-dessus, la probabilité de passage en année supérieure des étudiants issus du groupe très favorisé, toutes choses égales par ailleurs, est en moyenne plus élevée que celle des étudiants issus du groupe intermédiaire ou défavorisé.

⁹⁷ Pour tester formellement l'effet du modèle arc-en ciel, il faudrait introduire un indicateur mesurant la mixité sociale compatible avec l'approche développée dans le chapitre II, c'est-à-dire qui ne se focalise pas sur les extrêmes de la distribution des origines sociales

En ce qui concerne la mixité sociale des étudiants défavorisés, la variable liée à la part occupée par les étudiants issus du groupe défavorisé dans les formations, celle-ci est également significative au seuil de 5%. Par rapport au cas d'une étudiante boursière inscrite en droit (issue du groupe social défavorisé) ayant une note de 12 au bac, si la proportion d'étudiants défavorisés dans la formation dans laquelle l'étudiante est inscrite passe, par exemple, de 30% à 40% alors la probabilité de passage diminue de 58,6% à 50,2% (-8,2%).

Les éléments théoriques présentés en 1.3.1 et en 1.3.2 sont à nouveau susceptibles d'être mobilisés pour interpréter ces résultats. Toutefois, sur la base des estimations obtenues à ce stade, il ne nous semble pas possible de choisir un ou plusieurs modèles théoriques pour interpréter ces résultats. D'un point de vue théorique, ceux-ci pourraient ainsi être à nouveau interprétés à la lumière des modèles n°1 (« *bad apple model* ») et n°8 (« *subculture model* »), mais aussi du modèle n°7 dit de « l'arc-en-ciel » dans la mesure où la situation des formations les plus ségréguées où se concentrent les étudiants défavorisés peuvent potentiellement expliquer ce résultat. Suivant ce dernier modèle, ce serait moins la part des étudiants défavorisés qui serait en général un facteur négatif pour la réussite que la concentration de ces étudiants dans certaines formations. Sur ce point, si la variable de seuil utilisée permettant d'identifier les formations les plus ségréguées n'est pas significative, il nous semble que ce résultat doit être interprété avec précaution, comme expliqué plus en détail ci-dessous.

4.4 Synthèse et axes de travail envisageables

Les estimations effectuées dans la section 4 nous ont permis d'aboutir aux résultats suivants concernant notre modèle logit dichotomique destiné à estimer la probabilité de passage en année supérieure à l'issue de la première année de Licence:

- Si 9 variables explicatives individuelles (l'origine sociale individuelle défavorisée, la note au Bac, la filière de bac, l'existence de revenus salariaux, le lieu de domicile, le niveau de motivation, l'absentéisme, le fait d'être boursier, l'existence de difficultés d'organisation) sont significatives pour les résultats qui

concernent l'ensemble des étudiants, les variables explicatives liées à la mixité sociale ne le sont pas.

- Les résultats sont, toutefois, différents pour les étudiants issus en particulier du groupe défavorisé. En ce qui concerne les variables explicatives de mixité sociale, les variables liées à la part occupée par les étudiants très favorisés et défavorisés sont ainsi toutes les deux significatives au seuil de 5% pour ce groupe en particulier. Ce résultat conforte ceux présentés dans la section 1 selon lesquels il est possible, au-delà des effets apparents modestes ou non significatifs pour l'ensemble des élèves, de parvenir à des résultats significatifs pour certains groupes d'élèves.

Il nous semble, toutefois, que les résultats obtenus à ce stade doivent être interprétés avec précaution au regard des facteurs suivants :

- les problèmes de représentativité identifiés dans la section 3
- sur les cinq variables significatives de mixité sociale identifiée, seules deux le sont au seuil de 5%. Ceci peut s'expliquer par la taille limitée de l'échantillon, en particulier pour les estimations réalisées pour le groupe défavorisé (508 observations) et la nécessité d'introduire de nombreuses variables individuelles de contrôle de la réussite en Licence.
- les variables explicatives liées à la contribution individuelle des formations (calculées pour les très favorisés et les défavorisés) ainsi que la variable de seuil indiquant si une formation est très ségréguée ne sont pas significatives (malgré les tests de sensibilité effectués pour d'autres valeurs de seuil).

Au regard de ces différents facteurs, il ne nous semble pas possible de nous prononcer en faveur d'un ou de plusieurs des modèles d'interprétation théorique présentés dans le paragraphe 1.3. De notre point de vue, cela nécessiterait par exemple de disposer d'un échantillon plus grand, dont le degré de représentativité soit mieux garanti, afin de tester de manière fine des effets de seuil.

Au regard des difficultés identifiées à ce stade, plusieurs pistes de travaux complémentaires nous semblent envisageables pour affiner et potentiellement confirmer les résultats obtenus à ce stade :

- exploiter la base de données SISE-Résultats qui est exhaustive et contient des informations relatives aux résultats des examens, en utilisant une méthodologie similaire. Une solution possible serait d'estimer par exemple la probabilité d'obtenir le diplôme de Licence à l'issue par exemple de la L3. Selon nous, cette base de données présente, toutefois, deux inconvénients majeurs pour réaliser un travail d'évaluation de ce type : dans la version de la base diffusée par l' « Adisp-Progedo », il n'est pas possible d'identifier les étudiants (dont les données d'identification individuelle sont cryptées) afin de le suivre leur trajectoire individuelle sur plusieurs années de Licence et le nombre de variables de contrôle relatives aux caractéristiques individuelles de l'étudiant est plus limité que dans la base panel bacheliers 2014.
- à partir de la base de données « SISE-Résultats », estimer, pour chaque formation identifiée, l'évolution pluriannuelle des taux de passage en fonction de leur degré de mixité sociale. Cette solution présenterait l'avantage de diminuer les risques liés au degré de comparabilité des formations, au sens où, au sein d'une même discipline, les Universités sont susceptibles d'avoir des degrés d'exigence hétérogènes et donc des taux de passage différents, indépendamment des variables de mixité sociale. Cette méthode présenterait, toutefois, deux inconvénients principaux : le nombre limité de variables de contrôle disponibles dans la base SISE et la possibilité que l'évolution pluriannuelle du degré de mixité sociale des formations prises individuellement soit trop limitée pour obtenir des résultats significatifs.

5 Conclusion

Ce chapitre a permis de mettre au jour deux résultats principaux concernant la probabilité de passage en année supérieure des étudiants en première année de Licence. Les premiers résultats confirment, d'abord, pour l'ensemble des étudiants, le lien (toutes choses égales par ailleurs) entre certaines caractéristiques individuelles des étudiants et leur réussite. Ceci concerne l'importance du niveau scolaire à l'entrée dans l'enseignement supérieur (note au bac, type de bac), mais également la possibilité de pouvoir s'engager dans ses études (un fort absentéisme et une faible motivation correspondent à une probabilité plus faible de passage) ainsi que l'organisation dans le travail universitaire. Il confirme également l'importance des conditions matérielles dans la réussite des études : le fait de bénéficier d'une bourse correspond à un plus fort taux de réussite tandis que travailler durant ses études correspond à un plus faible taux de réussite. Nous montrons également que lorsque ces variables individuelles sont introduites dans les estimations, la mixité sociale des formations suivies n'a pas d'effet significatif sur l'ensemble des étudiants.

Afin de tenir compte des effets potentiellement hétérogènes de la mixité sociale, nous avons choisi, ensuite, de recentrer notre analyse sur le groupe défavorisé. Nous montrons que pour celui-ci, une part d'élèves favorisés plus importante au sein des formations qu'ils suivent correspond à une probabilité de passage en deuxième année de Licence également plus importante. Si les variables de mixité sociale ne sont pas significatives pour l'ensemble des étudiants, celles-ci le sont pour les étudiants défavorisés en particulier. Même si ce résultat doit être interprété avec précaution, cela conforte les résultats présentés dans la section 1 selon lesquels il est possible, au-delà des effets apparents modestes ou non significatifs pour l'ensemble des élèves, de parvenir à des résultats significatifs pour certains groupes d'élèves. Si les résultats proposés dans ce chapitre ne permettent pas de trancher précisément sur le type d'effets de pairs à l'œuvre en Licence à l'Université et doivent être interprétés avec précaution, ils suggèrent néanmoins que ces effets existent pour les étudiants les plus défavorisés.

En termes de recommandations de politiques publiques relatives à la mixité sociale, il nous semble que nos résultats ne nous permettent pas à ce stade de choisir de manière claire entre un des modèles d'interprétation théorique présentés dans la

section 1, ce qui aurait pu permettre d'aboutir à des conclusions opérationnelles spécifiques. Toutefois les politiques publiques pourraient potentiellement intervenir sur la scolarité antérieure, le lieu de domicile, l'absentéisme et les bourses étudiantes afin de favoriser les conditions matérielles des études. Pour les étudiants défavorisés, nos résultats confirment également sur le plan de la réussite universitaire les risques liés à l'existence de formations où la part d'étudiants ayant une origine sociale également défavorisée est élevée.

CONCLUSION GENERALE

Notre analyse sur la mixité sociale au sein de l'Université française nous a permis d'aboutir à quatre conclusions principales. Notre travail fournit, premièrement, un cadre méthodologique pour la mesure de la mixité sociale à l'Université. Celui-ci définit, d'abord, l'origine sociale dans le contexte de l'économie de l'éducation. L'*origine sociale* peut être selon nous appréhendée à partir de neuf critères qui permettront ensemble de mesurer « la distance à l'école » de l'individu : la PCS, le niveau de diplôme, le revenu ainsi que le patrimoine financier et immobilier de la famille, les ressources éducatives disponibles à la maison, la présence d'une culture classique à la maison, les relations familiales, les origines géographiques de la famille, la participation à des clubs ou des activités politiques et la scolarité antérieure de l'individu. Sur ce point, nous parvenons, toutefois, à la conclusion qu'en raison de la limitation des informations généralement disponibles dans les bases de données usuelles, une solution pertinente est de caractériser l'*origine sociale* des individus à l'aune du seul critère de la PCS. Ce choix se justifie dans la mesure où le critère de PCS comporte une dimension sociale, économique et culturelle. Par ailleurs, la mixité sociale peut se définir de la manière suivante : lorsqu'une population de référence est divisée ou répartie en sous-populations (en différentes unités), il existe une situation de « mixité sociale » lorsque la composition sociale de celles-ci est semblable ou proche de celle de la population de référence. Un premier cadrage statistique est également calculé sur la base de l'indice d'Atkinson et celui-ci confirme l'existence de phénomènes de ségrégation sur la période 2006-2016, même s'il y a plutôt une tendance à la baisse et que le niveau de cette ségrégation a tendance à diminuer avec l'avancement dans les études.

Nous mesurons ensuite de manière opérationnelle le niveau de mixité sociale au sein de l'Université française sur la base du cadre méthodologique défini antérieurement. Les indices de mesure de la mixité sociale retenus sont ainsi calculés aussi bien au niveau national que pour les différentes zones académiques. Sur la période 2006-2016, nous parvenons ainsi à mettre en évidence différents faits. Les Indices d'Exposition Normalisés aux très favorisés et aux défavorisés sont, en moyenne, relativement stables : entre 2006 et 2016, ils sont compris entre 4% et 5,5%. L'analyse géographique révèle des disparités de ségrégation notables entre académies. La

géographie de la ségrégation observée dans le secondaire, en particulier, n'est plus de mise à l'Université. Les niveaux de ségrégation obtenus en Île-de-France ou dans les grandes académies de Province (Lille, Lyon, Aix-Marseille) ne sont pas systématiquement supérieurs à ceux observés dans de petites académies, contrairement à ce qui est constaté pour le secondaire. Les résultats montrent également que la ségrégation des très favorisés diminue avec le niveau d'étude. Elle diminue entre L1 et L2 et baisse encore en Master, le phénomène étant un peu plus marqué pour la ségrégation des défavorisés entre L1 et L2. Nous constatons également en tenant compte à la fois du niveau d'études et de l'académie que la ségrégation concerne plus spécifiquement certaines académies, pour certains niveaux d'étude précis. La mise en évidence d'un niveau de ségrégation en moyenne faible à l'échelle de la France est ainsi parfaitement compatible avec des niveaux localement élevés.

Afin de compléter notre analyse empirique, nous décomposons nos résultats sur le degré de mixité sociale afin de connaître les parts contributives de chaque facteur au niveau total de la ségrégation. Au niveau national, environ un tiers de la ségrégation correspond à une composante inter-disciplinaire alors que les deux tiers restant résultent d'un facteur inter-universitaire interne à chaque discipline. Nos résultats décrivent également la dimension fortement locale et inter-disciplinaire de la ségrégation dans le système universitaire français. Ceux-ci confirment également la particularité de la région parisienne, avec des niveaux de ségrégation observés supérieurs en moyenne à ceux du reste du pays et en hausse sur la période 2006-2016 alors qu'une tendance inverse est observée dans le reste du pays.

La dernière étape de notre travail analyse les liens entre mixité sociale et réussite à l'Université. Afin de mesurer les effets de la mixité sociale, nous avons ainsi choisi d'analyser les liens potentiels entre mixité sociale des formations suivies et réussite universitaire mesurée par le passage en année supérieure. En sus des facteurs de réussite à l'Université utilisés habituellement dans la littérature, nous évaluons dans quelle mesure le degré de mixité sociale des formations peut également jouer un rôle. A partir des données disponibles dans la base de données panel bacheliers 2014, nous parvenons à mettre au jour un effet significatif de la part des étudiants très favorisés et défavorisés dans les formations sur la probabilité de passage en année supérieure pour le groupe des étudiants défavorisés uniquement. Au regard de la taille

limité de notre échantillon et de certains problèmes de représentativité identifiés, il nous semble, toutefois, que ce résultat doit être interprété avec précaution. Ces estimations, en particulier, ne nous permettent pas d'identifier clairement quel type d'effet de pairs pourrait potentiellement jouer. A cette fin, il nous semble qu'il serait nécessaire de réaliser des estimations plus fines reposant sur le test de différents effets de seuil à partir d'un échantillon plus étendu.

Au-delà du travail réalisé dans le cadre de notre thèse, il nous semble que de nouveaux axes de recherche sont envisageables. Une première possibilité serait d'étendre nos travaux à l'enseignement supérieur non universitaire en exploitant si possible les bases de données contenant des informations relatives aux étudiants du système non universitaire (SISE ENS, SISE Ingénieurs...). Il conviendrait alors de réutiliser également les données de la base SISE Université relatives à des établissements non universitaires. Une autre possibilité envisageable serait selon nous d'approfondir nos analyses relatives aux effets potentiels de la mixité sociale sur la réussite étudiante en utilisant la base de données SISE Résultats qui est exhaustive et contient des informations relatives aux résultats des examens. L'inconvénient majeur de cette solution, toutefois, est que celle-ci contient des informations limitées sur le parcours individuel des étudiants.

Table des matières

Remerciements	9
INTRODUCTION GENERALE	13
CHAPITRE I : MESURER LA MIXITE SOCIALE A L'UNIVERSITE	30
1 L'origine sociale dans le champ de l'éducation.....	36
1.1 Définir l' <i>origine sociale</i>	36
1.1.1 Principes généraux.....	36
1.1.2 Les différents types de critères utilisés	37
1.2 Les types de classification de l'origine sociale.....	40
1.2.1 Les approches basées sur un indice quantitatif continu.....	41
1.2.2 L'utilisation des professions et catégories socioprofessionnelles.....	43
1.2.3 Discussion des usages de classification	44
1.3 Définir des groupes sociaux homogènes.....	46
1.3.1 Les indices quantitatif continus.....	46
1.3.2 Le regroupement des PCS	47
1.4 Conclusion de la section 1	49
2 Mesurer la mixité sociale	52
2.1 Les éléments nécessaires à la mesure.....	52
2.1.1 Mesure et population de référence	52
2.1.2 Rappel des principaux facteurs susceptibles d'influencer les niveaux de ségrégation sociale dans le champ de l'éducation	53
2.1.3 Le principe des indices de mixité sociale.....	58
2.2 Les propriétés attendues des indices de mixité.....	59
2.2.1 Un degré de mixité de la population de référence au moins égal à celui des unités	60
2.2.2 L'invariance à la composition.....	60
2.2.3 La « décomposabilité forte » des unités	61

2.2.4	La facilité d'interprétation.....	62
2.2.5	Les autres propriétés.....	62
2.3	Les huit principaux indices de mesure de la mixité sociale	64
2.3.1	L'indice d'entropie normalisé ou indice de Theil	65
2.3.2	L'indice d'entropie non normalisé ou indice d'information mutuelle.....	68
2.3.3	L'indice d'Atkinson.....	70
2.3.4	L'indice d'exposition normalisé	72
2.3.5	L'Indice de dissimilarité	73
2.3.6	L'indice de Gini multigroupes.....	74
2.3.7	L'indice de diversité relative	76
2.3.8	L'indice du coefficient de variation au carré	77
2.4	De la nécessité de recourir à une stratégie multi-indicielle.....	77
3	L'unité d'observation de la mixité sociale.....	81
3.1	La granularité des unités d'observation	81
3.2	Le problème statistique des petites unités.....	83
3.3	La correction des biais liés aux petites unités.....	84
3.3.1	Le choix de l'unité d'observation dans la pratique.....	84
3.3.2	Corriger les biais	85
4	Un premier cadrage statistique sur l'évolution de la mixité sociale à l'Université	87
4.1	Méthode	87
4.2	Résultats de l'analyse	88
4.2.1	Les résultats de l'analyse pour l'ensemble de l'Université	89
4.2.2	Le résultat des estimations par niveau de diplôme	91
5	Conclusion	93
CHAPITRE II LA MIXITE SOCIALE A L'UNIVERSITE : UN PREMIER CADRAGE GEOGRAPHIQUE		95

1	Les enjeux géographiques de la mixité sociale dans l'enseignement.....	96
2	Base de données utilisée et mesure de la mixité	99
2.1	La base de données SISE.....	99
2.2	Précisions concernant le champ des établissements retenus.....	100
2.2.1	Le cas des établissements non universitaires.....	100
2.2.2	Le cas des ESPE et des COMUE.....	101
2.2.3	Le cas des IUT et des INP.....	102
2.3	L'appréhension des disciplines académiques avec SISE	103
2.4	La formation comme unité d'observation dans SISE	107
3	Premiers éléments de cadrage statistique.....	111
3.1	Origine sociale et formations à l'Université	111
3.1.1	Statistiques descriptives concernant l'ensemble du cycle universitaire.....	111
3.1.2	Origine sociale et niveau d'études.....	113
3.2	Origine sociale et discipline académique.....	116
3.3	Les effectifs des académies sont très variables	121
3.4	La répartition des étudiants par origine sociale et par académie.....	122
4	Au niveau national, des phénomènes de ségrégation réels mais plutôt stables et qui ont tendance à s'estomper suivant les années d'études	125
4.1	Au niveau national, une ségrégation plus faible que dans le secondaire et relativement stable dans le temps	125
4.2	Au niveau national, la ségrégation a plutôt tendance à s'estomper avec le niveau d'études	131
5	Des disparités géographiques importantes entre les académies et suivant les niveaux d'études	136
5.1	Une forte hétérogénéité des niveaux de ségrégation selon les académies.....	136

5.2	Niveau de ségrégation et niveau d'étude : des disparités entre académies	141
6	La taille des académies et la mobilité géographique des étudiants : des facteurs explicatifs potentiels ?	145
6.1	Niveau de ségrégation et taille des académies	145
6.2	Niveaux de ségrégation et origine géographique des étudiants.	148
7	Conclusion	151

CHAPITRE III MIXITE SOCIALE A L'UNIVERSITE : LA DECOMPOSITION DES RESULTATS	153
1 La décomposition des indices dans la littérature.....	155
2 La méthodologie utilisée : les principes de la décomposition.....	157
2.1 Les principes de la décomposition.....	157
2.2 Décomposition « simple » et « avancée ».....	158
3 La décomposition des indices de ségrégation portant sur la population de référence dans son ensemble.....	160
3.1 La décomposition simple suivant la discipline académique.....	160
3.2 Décomposition avancée suivant les zones académiques.....	163
4 La décomposition des indices de ségrégation de sous-populations	167
4.1 La décomposition des indices de ségrégation calculés pour chaque niveau d'inscription	167
4.2 La décomposition des indices de ségrégation calculés pour la région parisienne.....	171
5 Conclusion	175

CHAPITRE IV : QUEL LIEN ENTRE MIXITE SOCIALE ET REUSSITE AUX EXAMENS ?	177
1 Les facteurs de réussite à l'Université et la question de la mixité sociale	179
1.1 Les facteurs individuels de réussite	179
1.1.1 Les caractéristiques individuelles et familiales	180
1.1.2 Les conditions de vie de l'étudiant	183
1.1.3 Les facteurs liés à la personnalité de l'étudiant et à ses méthodes de travail	184
1.2 Les effets de contexte non attribuables aux pairs	185
1.3 Les effets de pairs et la mixité sociale	186
1.3.1 Les effets de pairs directs et indirects	186
1.3.2 Les effets de pairs hétérogènes	188
1.3.3 Les résultats empiriques des travaux disponibles	190
1.4 Evaluer le lien entre mixité sociale et réussite à l'Université	193
2 L'appariement de deux bases de données pour appréhender de manière fine les caractéristiques individuelles des étudiants	199
2.1 Appréhender les caractéristiques individuelles des étudiants	199
2.2 Appréhender les caractéristiques des établissements et des formations	201
2.3 L'appariement des deux bases	202
3 Réussir sa première année de Licence : un premier cadrage statistique	204
3.1 Taux de passage en année supérieure suivant l'origine sociale et discussion relative à la représentativité de la base panel bacheliers 2014	204
3.1.1 Les taux de passage en L2 suivant l'origine sociale	204
3.1.2 Le problème des taux de passage dans le panel bacheliers 2014	205
3.2 Les taux de passage en année supérieure suivant les disciplines suivies	206
3.3 Statistiques concernant le travail étudiant	208
4 Le lien entre réussite et caractéristiques individuelles : une première appréhension	210
4.1 Les variables utilisées	210

4.2	Les estimations réalisées pour l'ensemble des étudiants	214
4.3	Les estimations réalisées pour les étudiants défavorisés	222
4.4	Synthèse et axes de travail envisageables	226
5	Conclusion	229

Annexe 1 : les 32 PCS de la base SISE et le regroupement en 4 groupes sociaux

La base SISE comprend une variable concernant la catégorie socioprofessionnelle des parents de l'étudiant. Cette variable est définie à partir de la nomenclature de l'INSEE PCS 2003 qui comporte 42 PCS. Toutefois, celle-ci ont été pré-regroupées dans la base SISE en 32 PCS dont le code ainsi que la désignation non codée apparaît dans le tableau ci-dessous. Il est possible, ensuite, de regrouper ces 32 PCS en 4 groupes sociaux (très favorisé, favorisé, intermédiaire et défavorisé) suivant la méthode retenue par la Direction de l'évaluation, de la prospective et de la performance (DEPP) du MESRI.

Groupe social	Code PCS	Dénomination de la PCS	Part 2014 (en%)
Intermédiaire	10	AGRICULTEUR EXPLOITANT	1,6%
Intermédiaire	21	ARTISAN	3,4%
Intermédiaire	22	COMMERCANT ET ASSIMILE	3,1%
Très favorisé	23	CHEF D'ENTREPRISE DE DIX SALARIES OU PLUS	2,1%
Très favorisé	31	PROFESSION LIBERALE	4,8%
Très favorisé	33	CADRE DE LA FONCTION PUBLIQUE	4,3%
Très favorisé	34	PROFESSEUR ET ASSIMILE	4,1%
Très favorisé	35	PROFESSION INFORMATION, ARTS, SPECTACLES	0,8%
Très favorisé	37	CADRE ADMINIS.ET COMMERC. D'ENTREPRISE	9,1%
Très favorisé	38	INGENIEUR - CADRE TECHNIQUE D'ENTREPRISE	6,3%
Favorisé	42	INSTITUTEUR ET ASSIMILE	1,4%
Favorisé	43	PROFESSION INTERM. SANTE - TRAVAIL SOCIAL	3,6%
Favorisé	44	CLERGE, RELIGIEUX	0,1%
Favorisé	45	PROFESSION INTERM. ADM. FONCT. PUBLIQUE	1,7%
Favorisé	46	PROFESSION INTERM. ADM. COMMERCE ENTR.	1,0%
Intermédiaire	47	TECHNICIEN	4,2%
Intermédiaire	48	CONTREMAITRE, AGENT DE MAITRISE	2,5%
Intermédiaire	52	EMPL. CIVIL - AGENT SERVICE FONCT. PUBL.	5,2%
Intermédiaire	53	POLICIER ET MILITAIRE	1,8%
Intermédiaire	54	EMPLOYE ADMINISTRATIF D'ENTREPRISE	4,3%
Intermédiaire	55	EMPLOYE DE COMMERCE	3,4%
Intermédiaire	56	PERSON. SERVICE DIRECT AUX PARTICULIERS	1,6%
Défavorisé	61	OUVRIER QUALIFIE	11,2%
Défavorisé	66	OUVRIER NON QUALIFIE	2,5%
Défavorisé	69	OUVRIER AGRICOLE	0,3%
Intermédiaire	71	RETRAITE AGRICULTEUR EXPLOITANT	0,1%
Intermédiaire	72	RETRAITE ARTISAN, COMMERC., CHEF ENTREP.	0,4%
Intermédiaire	73	RETRAITE CADRE, PROFES. INTERMEDIAIRE	1,2%
Défavorisé	76	RETRAITE EMPLOYE ET OUVRIER	2,2%
Défavorisé	81	CHOMEUR N'AYANT JAMAIS TRAVAILLE	0,3%
Défavorisé	82	AUTRE PERSONNE SANS ACTIVITE PROFESSIONNELLE	6,4%
Défavorisé	99	NON RENSEIGNEE (INCONNUE OU SANS OBJET)	5,2%

Source : base SISE-Inscrits 2014. Diffuseur : Adisp-Progedo. Méthodologie de regroupement de la DEPP – Calculs de l'auteur

Annexe 2 : discussion relative à l'utilisation de la nomenclature des PCS en 32 postes dans le cadre de travaux portant sur la mixité sociale

Il convient de noter que les travaux existants sur la mixité sociale et utilisant la base SISE du MESRI modifient parfois eux-mêmes la nomenclature à 32 postes susmentionnée. Ainsi Avouac et Harari-Kermadec (2021) indiquent que la nomenclature des PCS à deux chiffres ne leur paraît pas totalement pertinente, certaines catégories étant susceptibles de présenter une forte hétérogénéité en particulier en termes de capitaux sociaux et culturels. Aussi, les auteurs excluent du champ d'études les étudiants dont le parent de référence est retraité, les catégories de retraités étant de leur point de vue trop agrégées, le nombre d'observations concernées étant selon eux de surcroît limité (d'après nos calculs, cela représente 3,9% des observations dans la base SISE 2014). Par ailleurs, sont exclus dans ce travail les individus dont le parent de référence est codé comme « chômeur n'ayant jamais travaillé » en raison du faible nombre d'observations (0,3% des observations dans la base SISE 2014) ainsi que ceux dont le parent est codé comme « autre personne sans activité de référence » en raison des risques d'hétérogénéité ou dont la CS n'est pas renseignée. D'après nos calculs, ces deux derniers postes représentent respectivement 6,4% et 5,2% des observations dans la base SISE 2014⁹⁸. Il nous semble, toutefois, que l'utilisation de la nomenclature utilisée par la DEPP présente certains avantages :

- Cela permet de conserver un nombre très significatif d'observations (environ 15,8% pour l'année 2014) exclues dans Avouac et Harari-Kermadec (2021)
- La répartition de l'ensemble des retraités en quatre catégories uniquement, si elle n'est pas idéale d'un point de vue conceptuel, peut se justifier au regard du nombre restreint d'observations correspondant à cette catégorie (3,9% du total) ainsi qu'aux modifications des conditions de vie, des relations sociales et des revenus liées à la sortie de la vie active.

⁹⁸ Les raisons pour lesquelles nous avons fait le choix de conserver la nomenclature de la DEPP en l'état sont décrites ci-dessous.

- Les étudiants dont le parent de référence est codifié comme « autre personne sans activité de référence » ou pour lequel la PCS n'est pas renseigné ne sont pas exclus non plus car nous faisons l'hypothèse que celle-ci serait probablement renseignée dans le cas où le parent de référence exercerait une activité professionnelle mieux positionnée dans la hiérarchie des PCS (au sens de Bechichi et al. (2021) et de Rocher (2016)).

Annexe 3 : Liste des variables disponibles dans la base SISE Université diffusée via l'ADISP- PROGEDO pour l'année 2016-2017

ACABAC	Académie d'obtention du baccalauréat
ACAETA	Académie de l'établissement
ACARESPA	Académie de résidence des parents de l'étudiant
AGE	Âge révolu de l'étudiant au 31 décembre de l'année de la rentrée universitaire
ANBAC	Année d'obtention du baccalauréat
ANINSC	Année de première inscription dans le système universitaire français
ANSUP	Année d'entrée dans l'enseignement supérieur
BAC	Série du bac ou équivalence
BAC_RGRP	Série du baccalauréat ou équivalence (regroupée)
COMPOS	Unité de rattachement pédagogique et géographique de l'étudiant
CURPAR	Cursus suivi parallèlement dans un autre établissement d'enseignement supérieur
CURSUS_LM D	Cursus d'appartenance de l'étudiant
CYCLE	Cycle d'appartenance de l'étudiant
DEGETU	Niveau atteint par l'étudiant pour l'année en cours
DEPETA	Département de l'établissement
DEPRESPA	Département de résidence des parents de l'étudiant
DIPLOM	Diplôme de l'inscription
DISCIPLI	Discipline du diplôme d'inscription de l'étudiant
EFFECTIF	Inscription principale ou seconde
ETABLI	Numéro d'identification de l'établissement dans la BCE (Base Centrale des Etablissements)
ETABLI_DIFF USION	Nom de l'établissement
FLAG_MEEF	Masters MEEF 1ère année (ESPE)
FR_ETR_D	Regroupement géographique des établissements
GROUPE	Regroupement disciplinaire du diplôme d'inscription
IDET	Identifiant national de l'étudiant (crypté)
NATRG	Nationalité de l'étudiant (regroupée)
NBACH	Bachelier de l'année inscrit pour la première fois dans le système universitaire français en première année de premier cycle
NET	Nouvel entrant dans le système universitaire français
NIVEAU	Niveau dans le diplôme d'inscription atteint par l'étudiant pendant la présente année d'inscription
NUMED	Numéro de l'école doctorale
PCSPAR	Catégorie socio-professionnelle des parents de l'étudiant
REGIME	Régime d'inscription de l'étudiant
SECTDIS	Secteur disciplinaire du diplôme d'inscription de l'étudiant
SEXE	Sexe de l'étudiant
SPECIA	Spécialisation d'un diplôme national à un niveau plus fin que celui utilisé par les libellés du diplôme
SPECIB	Spécialisation d'un diplôme national à un niveau plus fin que celui utilisé par les libellés du diplôme
TYP_DIPL	Type du diplôme d'inscription de l'étudiant
TYPPREPA	Classe préparatoire fréquentée par les étudiants en formation d'ingénieur
UNIV	Type d'établissement (universitaire ou non)
VOIE	Voie du diplôme

Annexe 4 : 33 Etablissement de SISE hors champ de notre analyse

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Particularité	
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016		
0352756F	COMUE UNIV BRETAGNE LOIRE	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	COMUE
0755698L	COMUE UNIV PARIS LUMIERES	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	COMUE
0755700N	COMUE UNIV RECHE PARIS SC LET	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	COMUE
0062126D	COMUE UNIVERSITE COTE D'AZUR	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	COMUE
0383412C	COMUE UNIVERSITE DE GRENOBLE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	COMUE
0772710C	COMUE UNIVERSITE PARIS EST	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	COMUE/ EN DOUBLE
0912330N	COMUE UNIVERSITE PARIS-SACLAY	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	COMUE
	ESPE ACADEMIE DE BESANCON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	ESPE
	ESPE ACADEMIE DE LYON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	ESPE
	ESPE ACADEMIE DE PARIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	ESPE
	ESPE ACA	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	ESPE
038XXXXX	UNIVERSITE GRENOBLE ?	NON	NON	NON	OUI	NON	Non identifié							
077XXXXX	UNIVERSITE MARNE LA VALLE ?	NON	NON	NON	OUI	NON	Non identifié							
0410981U	ENS NATURE PAYSAGE BLOIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	NON	NON	Non universitaire
0753431X	INST ETUDES POLITIQUES PARIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0753488J	INST NAT LANG CIV ORIENT PARIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0753428U	INST PHYSIQUE DU GLOBE PARIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0381912X	INST POLYTECHNIQUE GRENOBLE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0753496T	OBSERVATOIRE PARIS	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0900424X	UNIV TECHN BELFORT MONTBELIARD	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0601223D	UNIV TECHNOLOGIE COMPIEGNE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Particularité	
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016		
0750736T	UNIVERSITE PARIS-DAUPHINE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
0101060Y	UNIVERSITE TECHNOLOGIE TROYES	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Non universitaire
	UNIVERSITE DE CORSE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	Autre zone géographique
9760358K	CUFR MAYOTTE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Outre-Mer
9710585J	UNIVERSITE ANTILLES	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Outre-Mer
9710939U	UNIVERSITE ANTILLES GUYANNE (GUADELLOUPE)';	NON	NON	OUI	NON	Outre-Mer								
9730224F	UNIVERSITE ANTILLES GUYANNE (GUYANNE)	NON	NON	OUI	NON	Outre-Mer								
9720706K	UNIVERSITE ANTILLES GUYANNE (MARTINIQUE)	NON	NON	OUI	NON	Outre-Mer								
9730429D	UNIVERSITE DE LA GUYANE	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	Outre-Mer
9830445S	UNIVERSITE DE NOUVELLE-CALEDONIE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Outre-Mer
9740478B	UNIVERSITE LA REUNION	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Outre-Mer
9840349G	UNIVERSITE POLYNESIE FRANCAISE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Outre-Mer

Source : base SISE-Inscrits. Diffuseur : Adisp-Progedo.

Annexe 5 : les 88 établissements inclus dans le champ de notre analyse

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Commentaire	
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016		
0811293R	CUFR NORD EST MIDI PYRENEES	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	Universitaire
0311381H	INP TOULOUSE	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	Universitaire
0541564G	NANCY INP	OUI	OUI	OUI	OUI	OUI	NON	Lorraine - Fusion en 2011						
0772502B	UNIV PARIS-EST MARNE-LA-VALLEE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0290346U	UNIVERSITE BREST	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0542493S	UNIVERSITE LORRAINE	NON	NON	NON	NON	NON	OUI	Lorraine - Fusion en 2011						
0131842G	UNIVERSITE AIX 1	OUI	OUI	OUI	OUI	OUI	NON	-						
0131843H	UNIVERSITE AIX 2	OUI	OUI	OUI	OUI	OUI	NON	-						
0132364Z	UNIVERSITE AIX 3	OUI	OUI	OUI	OUI	OUI	NON	-						
0134009M	UNIVERSITE AIX-MARSEILLE	NON	NON	NON	NON	NON	OUI	-						
0801344B	UNIVERSITE AMIENS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0490970N	UNIVERSITE ANGERS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0623957P	UNIVERSITE ARTOIS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0840685N	UNIVERSITE AVIGNON	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0251215K	UNIVERSITE BESANCON	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0333298F	UNIVERSITE BORDEAUX	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	OUI	-
0331764N	UNIVERSITE BORDEAUX 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	-
0331765P	UNIVERSITE BORDEAUX 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	-
0331766R	UNIVERSITE BORDEAUX 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0332929E	UNIVERSITE BORDEAUX 4	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	-
0561718N	UNIVERSITE BRETAGNE SUD	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0141408E	UNIVERSITE CAEN	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0951793H	UNIVERSITE CERGY PONTOISE	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	OUI	OUI	-
0730858L	UNIVERSITE CHAMBERY	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Commentaire	
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016		
0632035V	UNIVERSITE CLERMONT AUVERGNE	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	Clermont - Fusion en 2016
0631262E	UNIVERSITE CLERMONT FERRAND 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	Clermont - Fusion en 2016
0631525R	UNIVERSITE CLERMONT FERRAND 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	Clermont - Fusion en 2016
7200664J	UNIVERSITE CORSE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0342321N	UNIVERSITE DE MONTPELLIER	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	OUI	Montpellier - Fusion en 2014
0830766G	UNIVERSITE DE TOULON	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0211237F	UNIVERSITE DIJON	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0911975C	UNIVERSITE EVRY VAL D ESSONNE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0381838S	UNIVERSITE GRENOBLE 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	Grenoble - Fusion en 2015
0381839T	UNIVERSITE GRENOBLE 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	Grenoble - Fusion en 2015
0381840U	UNIVERSITE GRENOBLE 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	Grenoble - Fusion en 2015
0383493R	UNIVERSITE GRENOBLE ALPES	NON	NON	NON	NON	NON	NON	NON	NON	NON	NON	OUI	OUI	Grenoble - Fusion en 2015
0171463Y	UNIVERSITE LA ROCHELLE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0762762P	UNIVERSITE LE HAVRE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0720916E	UNIVERSITE LE MANS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0593559Y	UNIVERSITE LILLE 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0593560Z	UNIVERSITE LILLE 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0593561A	UNIVERSITE LILLE 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0870669E	UNIVERSITE LIMOGES	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0595964M	UNIVERSITE LITTORAL	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0691774D	UNIVERSITE LYON 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0691775E	UNIVERSITE LYON 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0692437Z	UNIVERSITE LYON 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Commentaire
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	
0572081C	UNIVERSITE METZ	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	NON	Lorraine - Fusion en 2011
0341087X	UNIVERSITE MONTPELLIER 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	Montpellier - Fusion en 2014
0341088Y	UNIVERSITE MONTPELLIER 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	Montpellier - Fusion en 2014
0341089Z	UNIVERSITE MONTPELLIER 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0681166Y	UNIVERSITE MULHOUSE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0541507V	UNIVERSITE NANCY 1	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	NON	Lorraine - Fusion en 2011
0541508W	UNIVERSITE NANCY 2	OUI	OUI	OUI	OUI	OUI	NON	NON	NON	NON	NON	NON	Lorraine - Fusion en 2011
0440984F	UNIVERSITE NANTES	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0060931E	UNIVERSITE NICE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0301687W	UNIVERSITE NIMES	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0450855K	UNIVERSITE ORLEANS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751717J	UNIVERSITE PARIS 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0921204J	UNIVERSITE PARIS 10	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0911101C	UNIVERSITE PARIS 11	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0941111X	UNIVERSITE PARIS 12	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0931238R	UNIVERSITE PARIS 13	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751718K	UNIVERSITE PARIS 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751719L	UNIVERSITE PARIS 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751720M	UNIVERSITE PARIS 4	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751721N	UNIVERSITE PARIS 5	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751722P	UNIVERSITE PARIS 6	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0751723R	UNIVERSITE PARIS 7	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0931827F	UNIVERSITE PARIS 8	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0640251A	UNIVERSITE PAU	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-

Numéro	Nom de l'établissement	Présent sur la période 2006-2016											Commentaire	
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016		
0660437S	UNIVERSITE PERPIGNAN	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0860856N	UNIVERSITE POITIERS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0511296G	UNIVERSITE REIMS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0350936C	UNIVERSITE RENNES 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0350937D	UNIVERSITE RENNES 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0761904G	UNIVERSITE ROUEN	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0421095M	UNIVERSITE SAINT ETIENNE	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0673021V	UNIVERSITE STRASBOURG	NON	NON	OUI	Strasbourg - Fusion en 2008									
0671712X	UNIVERSITE STRASBOURG 1	OUI	OUI	NON	Strasbourg - Fusion en 2008									
0671713Y	UNIVERSITE STRASBOURG 2	OUI	OUI	NON	Strasbourg - Fusion en 2008									
0671778U	UNIVERSITE STRASBOURG 3	OUI	OUI	NON	Strasbourg - Fusion en 2008									
0311382J	UNIVERSITE TOULOUSE 1	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0311383K	UNIVERSITE TOULOUSE 2	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0311384L	UNIVERSITE TOULOUSE 3	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0370800U	UNIVERSITE TOURS	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0593279U	UNIVERSITE VALENCIENNES	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-
0781944P	UNIVERSITE VERSAILLES ST QUENT	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	OUI	-

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).
Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo).

Annexe 6 : statistiques descriptives

Effectifs de l'ensemble des Universités et parts respectives des quatre groupes sociaux (2006-2016)

Année	Effectif	Très favorisé	Favorisé	Inter-médiaire	Défavorisé	Part très favorisé	Part favorisé	Part inter-médiaire	Part défavorisé
2006	1 247 131	399 983	100 412	360 827	385 909	32,1%	8,1%	28,9%	30,9%
2007	1 213 249	392 186	95 473	351 413	374 177	32,3%	7,9%	29,0%	30,8%
2008	1 250 868	399 172	96 085	360 240	395 371	31,9%	7,7%	28,8%	31,6%
2009	1 288 397	407 682	91 378	371 618	417 719	31,6%	7,1%	28,8%	32,4%
2010	1 280 898	404 688	85 929	368 585	421 696	31,6%	6,7%	28,8%	32,9%
2011	1 291 677	403 779	85 370	367 866	434 662	31,3%	6,6%	28,5%	33,7%
2012	1 301 190	406 739	86 748	372 192	435 511	31,3%	6,7%	28,6%	33,5%
2013	1 328 766	412 283	89 525	381 226	445 732	31,0%	6,7%	28,7%	33,5%
2014	1 353 512	417 148	90 893	390 037	455 434	30,8%	6,7%	28,8%	33,6%
2015	1 411 464	433 565	94 931	406 555	476 413	30,7%	6,7%	28,8%	33,8%
2016	1 435 216	443 352	97 069	416 487	478 308	30,9%	6,8%	29,0%	33,3%
Moyenne 2006-2016	1 309 306	410 962	92 165	377 004	429 176	31,4%	7,1%	28,8%	32,7%

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur

Effectifs et parts respectives des quatre groupes sociaux pour les différents niveaux d'études (moyenne 2006-2016)

Niveau	Effectif	Très favorisé	Favorisé	Inter-médiaire	Défavorisé	Part très favorisé	Part favorisé	Part inter-médiaire	Part défavorisé
Licence 1	432 116	129 291	31 276	127 816	143 733	30,0%	7,2%	29,6%	33,2%
Licence 2	234 485	80 000	17 852	72 224	64 409	34,1%	7,6%	30,8%	27,4%
Licence 3	253 192	77 563	17 740	73 844	84 045	30,6%	7,0%	29,2%	33,2%
Licence (total 3 années)	919 794	286 854	66 867	273 884	292 188	31,2%	7,3%	29,8%	31,7%
Master 1	203 139	65 499	13 843	55 999	67 798	32,2%	6,8%	27,6%	33,4%
Master 2	186 374	58 608	11 455	47 122	69 190	31,4%	6,1%	25,3%	37,1%
Master (total 2 années)	389 512	124 107	25 297	103 120	136 988	31,9%	6,5%	26,5%	35,1%
Licence + Master	1 309 306	410 962	92 165	377 004	429 176	31,4%	7,0%	28,8%	32,8%

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur

Effectifs par disciplines d'inscription et parts respectives des quatre groupes sociaux pour la L1 et l'ensemble de la Licence (moyenne 2013-2016)

Disciplines SISE	Ensemble de l'Université (hors doctorat)						Première année de Licence			
	Effectifs	Part du total des étudiants inscrits	Part très favorisés	Part favorisés	Part intermédiaires	Part défavorisés	Part très favorisés	Part favorisés	Part intermédiaires	Part défavorisés
DROIT - SCIENCES POLITIQUES	196 314	14,2%	36,0%	6,2%	27,6%	30,2%	30,8%	6,6%	28,9%	33,6%
SCIENCES ECONOMIQUES – GESTION (HORS AES)	192 676	13,9%	29,3%	5,3%	28,6%	36,8%	31,3%	6,1%	30,8%	31,8%
ADMINISTRATION ECONOMIQUE ET SOCIALE (AES)	35 719	2,6%	20,4%	6,1%	29,5%	44,0%	17,8%	6,1%	29,3%	46,9%
LETTRES – SCIENCES DU LANGAGE – ARTS	88 483	6,4%	26,9%	7,4%	25,6%	40,0%	20,7%	6,4%	22,3%	50,7%
LANGUES	105 806	7,7%	22,2%	6,9%	30,8%	40,1%	19,7%	6,8%	30,6%	42,9%
SCIENCES HUMAINES ET SOCIALES	249 946	18,1%	26,1%	8,0%	29,8%	36,1%	25,0%	8,6%	31,7%	34,7%
SCIENCES DE LA VIE, DE LA SANTE, DE LA TERRE ET DE L'UNIVERS	77 386	5,6%	30,3%	7,1%	32,2%	30,4%	28,2%	7,8%	34,7%	29,3%
STAPS*	49 466	3,6%	27,7%	8,8%	35,7%	27,8%	26,1%	9,0%	36,7%	28,3%
MEDECINE	61 627	4,5%	42,6%	6,3%	20,5%	30,5%	30,2%	6,3%	26,2%	37,3%
ODONTOLOGIE	5 638	0,4%	56,3%	6,3%	22,5%	15,0%	33,3%	1,9%	3,7%	61,1%
PHARMACIE	14 814	1,1%	48,2%	6,3%	25,9%	19,6%	3,3%	0,6%	6,3%	89,8%
SCIENCES FONDAMENTALES ET APPLICATIONS	209 572	15,2%	32,1%	6,1%	30,0%	31,8%	31,7%	6,7%	31,4%	30,3%
PLURI DROIT-SCIENCES ECO-AES	864	0,1%	17,9%	5,5%	26,1%	50,4%	24,9%	6,4%	27,2%	41,5%
PLURI LETTRES – LANGUES – SCIENCES HUMAINES	8 365	0,6%	28,8%	7,0%	21,0%	43,2%	35,9%	8,2%	22,8%	33,1%
PLURI SCIENCES	29 682	2,1%	40,6%	6,0%	26,5%	26,9%	39,9%	6,2%	27,1%	26,7%
PLURI SANTE**	55 883	4,0%	43,4%	7,7%	28,5%	20,4%	43,4%	7,7%	28,5%	20,4%
Total Université (hors Doctorat)	1 382 240	100,0 %	30,9%	6,7%	28,8%	33,6%	29,4%	7,1%	29,9%	33,6%

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo), calculs de l'auteur.

* : Sciences et techniques des activités physiques et sportives ** Pluri-Santé correspond uniquement à la première année commune aux études de santé.

Note de lecture : la colonne « part du total des étudiants inscrits » correspond à la part des étudiants inscrits dans une discipline parmi l'ensemble des étudiants. Ainsi, 14,2% du total des étudiants de notre champ d'analyse était inscrit en « Droit-Sciences Politiques » sur la période analysée.

Comme indiqué dans le paragraphe 2.3, à compter de la rentrée 2010, les étudiants des filières médecine, odontologie et pharmacie sont d'abord inscrits dans la première année commune aux études de santé et seulement, à compter de la seconde année dans leur discipline respective pour ceux qui ont réussi l'examen de fin d'année. Suivant les informations qui nous ont été communiquées par les gestionnaires de la base SISE, les étudiants qui apparaissent comme inscrits dans ces trois disciplines dès la première année sont inscrits dans des Licences professionnelles qui correspondent à des diplômes paramédicaux (herboristerie...). Il s'agit d'un nombre limité de cursus Licence qui ne durent qu'une année et pour lesquels les étudiants sont inscrits automatiquement en niveau 1.

Effectifs par académies (2006-2016)

Académies	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
PARIS (01)	145 506	140 243	146 080	149 630	152 658	152 915	151 907	152 528	151 853	156 454	157 846
AIX- MARSEILLE (02)	63 089	60 962	62 970	63 926	63 462	62 352	60 907	58 991	59 844	62 484	63 613
BESANCON (03)	18 137	17 442	17 320	18 109	17 754	17 699	17 748	17 702	17 704	18 379	19 086
BORDEAUX (04)	65 294	64 344	64 090	65 928	65 256	65 729	66 094	67 578	69 132	71 595	72 953
CAEN (05)	23 159	22 160	22 401	22 262	21 918	21 367	21 603	22 117	23 383	25 746	26 855
CLERMONT- FERRAND(06)	24 948	24 766	25 626	26 777	27 401	28 993	27 223	27 881	27 571	28 114	28 556
DIJON (07)	23 719	23 307	24 357	24 385	23 572	23 873	23 567	24 173	24 570	25 056	26 681
GRENOBLE (08)	48 515	47 375	48 915	49 430	48 861	48 363	48 911	50 481	52 743	55 090	56 181
LILLE (09)	89 061	86 840	89 848	92 804	89 930	89 649	91 170	93 562	93 036	97 935	98 560
LYON (10)	83 820	82 129	84 831	88 312	89 138	90 466	93 421	94 135	95 857	100 808	103 164
MONTPELLIER (11)	57 815	56 462	57 937	61 086	62 606	64 127	65 403	67 606	67 860	69 826	71 384
NANCY-METZ (12)	48 182	46 511	46 528	47 840	47 718	47 422	47 318	48 250	50 485	53 558	55 921
POITIERS (13)	28 848	27 564	28 549	28 735	28 082	27 884	26 960	28 167	29 018	30 951	32 051
RENNES (14)	62 754	60 521	60 624	62 917	63 147	64 935	66 023	67 757	69 789	74 116	74 993
STRASBOURG (15)	43 907	42 816	44 316	45 323	45 641	46 118	46 842	47 392	48 975	50 987	53 327
TOULOUSE (16)	68 060	65 561	65 615	67 577	68 676	69 638	72 117	74 324	77 683	79 961	81 437
NANTES (17)	52 675	52 028	54 654	57 262	56 771	56 582	57 871	60 825	62 828	64 525	67 193
ORLEANS- TOURS (18)	34 851	33 723	34 120	34 719	33 479	34 481	34 360	35 053	36 635	38 825	40 973
REIMS (19)	19 343	18 893	20 003	20 355	19 514	19 262	19 392	20 576	21 702	23 340	22 148
AMIENS (20)	19 100	18 557	19 983	20 549	20 320	20 862	21 599	22 105	23 346	25 448	25 870
ROUEN (21)	27 389	26 156	27 259	28 596	28 324	28 390	29 345	30 476	31 086	32 841	34 823
LIMOGES (22)	12 811	12 423	12 485	12 503	12 380	12 316	12 883	13 349	13 780	13 920	14 381
NICE (23)	31 923	30 607	33 179	33 463	32 344	31 772	30 835	32 182	33 557	35 798	36 753
CRETEIL (24)	70 252	69 026	73 196	76 915	74 959	76 608	76 745	77 584	77 703	79 649	80 778
VERSAILLES (25)	83 973	82 833	85 982	88 994	86 987	89 874	90 946	93 972	93 372	96 058	89 689
TOTAL 25 ACADEMIES	1 247 131	1 213 249	1 250 868	1 288 397	1 280 898	1 291 677	1 301 190	1 328 766	1 353 512	1 411 464	1 435 216

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur.

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Note de lecture : La première colonne indique l'effectif de chacune des académies en 2006. Ainsi, celui-ci était de 145 506 pour l'académie de Paris et de 1 247 131 au total.

**Parts respectives des quatre groupes sociaux suivant les académies
(moyenne 2006-2016, en%)**

Académies	Effectifs	Très favorisé	Favorisé	Intermédiaire	Défavorisé
PARIS (01)	150 693	42,9%	5,4%	23,5%	28,1%
AIX-MARSEILLE (02)	62 055	31,5%	6,8%	30,7%	31,0%
BESANCON (03)	17 916	23,7%	9,3%	29,1%	37,9%
BORDEAUX (04)	67 090	32,1%	7,6%	31,2%	29,1%
CAEN (05)	22 997	26,5%	7,9%	33,2%	32,4%
CLERMONT-FERRAND(06)	27 078	26,5%	8,0%	30,5%	35,0%
DIJON (07)	24 296	26,9%	8,5%	34,0%	30,6%
GRENOBLE (08)	50 442	32,5%	7,8%	30,3%	29,5%
LILLE (09)	92 036	29,7%	6,5%	27,5%	36,3%
LYON (10)	91 462	34,6%	6,5%	27,2%	31,7%
MONTPELLIER (11)	63 828	25,3%	6,7%	28,6%	39,4%
NANCY-METZ (12)	49 067	28,3%	7,7%	28,5%	35,5%
POITIERS (13)	28 801	26,2%	9,1%	32,5%	32,2%
RENNES (14)	66 143	28,9%	8,4%	32,2%	30,5%
STRASBOURG (15)	46 877	31,1%	6,6%	24,8%	37,5%
TOULOUSE (16)	71 877	33,4%	8,0%	31,2%	27,4%
NANTES (17)	58 474	29,7%	7,8%	28,8%	33,8%
ORLEANS-TOURS (18)	35 565	27,2%	8,9%	32,7%	31,2%
REIMS (19)	20 412	22,6%	7,7%	30,4%	39,3%
AMIENS (20)	21 613	24,8%	7,8%	29,2%	38,2%
ROUEN (21)	29 517	27,3%	8,2%	30,4%	34,1%
LIMOGES (22)	13 021	28,1%	7,0%	35,0%	30,0%
NICE (23)	32 947	26,1%	6,3%	31,8%	35,8%
CRETEIL (24)	75 765	25,3%	6,1%	28,3%	40,3%
VERSAILLES (25)	89 335	38,1%	5,7%	26,0%	30,1%
TOTAL 25 ACADEMIES	1 309 306	31,4%	7,1%	28,8%	32,7%

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur

Note de lecture : La première colonne indique le pourcentage d'étudiants très favorisés en moyenne sur la période 2006-2016. Ainsi, celui-ci était de 42,9% pour l'académie de Paris et de 31,4% en moyenne en France métropolitaine.

**Parts des quatre groupes sociaux suivant les académies et le niveau d'études
(moyenne 2006-2016, en%)**

Académies	Première année de Licence					Licence					Master				
	Effectifs	T	F	I	D	Effectifs	T	F	I	D	Effectifs	T	F	I	D
PARIS (01)	44 522	42%	5%	24%	29%	96 749	44%	6%	24%	27%	53 944	42%	5%	23%	30%
AIX-MARSEILLE (02)	21 054	28%	7%	32%	33%	42 771	30%	7%	32%	31%	19 283	34%	6%	29%	31%
BESANCON (03)	6 457	24%	9%	30%	37%	13 614	24%	9%	30%	37%	4 302	23%	9%	26%	42%
BORDEAUX (04)	22 620	32%	8%	33%	28%	46 662	32%	8%	33%	27%	20 428	31%	7%	28%	33%
CAEN (05)	8 473	26%	8%	34%	31%	16 953	27%	8%	34%	31%	6 045	26%	7%	30%	37%
CLERMONT-FERRAND(06)	9 019	26%	8%	32%	33%	19 421	27%	8%	32%	33%	7 657	26%	7%	27%	40%
DIJON (07)	8 270	25%	9%	34%	31%	17 466	27%	9%	35%	30%	6 831	27%	8%	32%	33%
GRENOBLE (08)	16 423	33%	8%	31%	28%	35 815	33%	8%	31%	28%	14 627	32%	7%	28%	33%
LILLE (09)	31 204	27%	7%	27%	40%	65 470	29%	7%	28%	37%	26 566	32%	6%	27%	35%
LYON (10)	28 709	34%	6%	27%	33%	64 272	35%	7%	28%	31%	27 190	34%	6%	26%	34%
MONTPELLIER (11)	21 031	24%	7%	30%	39%	43 446	26%	7%	31%	36%	20 382	24%	5%	24%	47%
NANCY-METZ (12)	16 407	25%	8%	29%	38%	34 973	27%	8%	29%	35%	14 093	31%	7%	26%	36%
POITIERS (13)	9 277	25%	9%	34%	32%	19 911	26%	9%	35%	30%	8 890	26%	8%	28%	38%
RENNES (14)	22 102	28%	8%	33%	31%	48 041	29%	8%	33%	30%	18 102	30%	8%	29%	33%
STRASBOURG (15)	14 751	28%	6%	25%	41%	31 845	30%	7%	26%	38%	15 032	34%	7%	23%	37%
TOULOUSE (16)	23 273	31%	8%	31%	29%	50 557	33%	8%	32%	27%	21 320	34%	8%	30%	28%
NANTES (17)	19 469	29%	8%	29%	34%	42 125	30%	8%	30%	33%	16 349	30%	8%	27%	36%
ORLEANS-TOURS (18)	12 682	27%	10%	34%	30%	26 204	27%	9%	34%	29%	9 362	27%	8%	29%	37%
REIMS (19)	7 792	21%	8%	31%	41%	15 835	22%	8%	31%	39%	4 577	24%	7%	28%	40%
AMIENS (20)	8 386	25%	8%	30%	38%	16 164	25%	8%	30%	37%	5 448	24%	7%	27%	42%
ROUEN (21)	11 206	26%	8%	31%	35%	22 368	27%	8%	31%	34%	7 149	28%	8%	29%	35%
LIMOGES (22)	4 673	28%	8%	35%	30%	9 954	28%	7%	36%	29%	3 067	29%	6%	33%	32%
NICE (23)	11 990	25%	7%	34%	34%	23 755	26%	7%	34%	33%	9 191	26%	5%	27%	42%
CRETEIL (24)	24 491	24%	6%	30%	39%	53 122	25%	6%	29%	39%	22 643	26%	6%	26%	43%
VERSAILLES (25)	27 836	39%	6%	27%	27%	62 300	39%	6%	27%	28%	27 034	36%	5%	24%	34%
TOTAL 25 ACADEMIES	432 116	30%	7%	30%	33%	919 794	31%	7%	30%	32%	389 512	32%	6%	26%	35%

Champ : ensemble du système universitaire (étudiants de France métropolitaine hors doctorat).

Source : Base SISE-Inscrits 2006-2016 (Diffuseur : Adisp-Progedo) et calculs de l'auteur

Note de lecture : « TF » fait référence aux étudiants issus du groupe très favorisé, « F » à ceux issus du groupe favorisé, I à ceux provenant du groupe « intermédiaire » et D à ceux qui sont issus du groupe « défavorisé ». La première colonne indique l'effectif moyen de l'académie sur la période 2006-2016 pour a première année de Licence. Ainsi, celui-ci était de 44 522 étudiants pour l'académie de Paris. Parmi ceux-ci, 42% avaient une origine sociale très favorisée, 5% favorisée, 24% intermédiaire et 29% défavorisée.

**Annexe 7 : les 13 PCS du panel bacheliers 2014, regroupement en 3 groupes sociaux
et correspondance avec les PCS de la base SISE**

Dans le Panel bacheliers 2014, une partie du questionnaire concerne la PCS des parents. Afin de renseigner la profession des deux parents, il existe 13 possibilités décrites ci-dessous. Selon nous, il n'est pas possible de regrouper les 32 PCS utilisées dans la base SISE dans ces 13 PCS. En l'absence d'information supplémentaire, il n'existe pas de notre point de vue de solution simple pour regrouper ces 13 PCS en quatre groupes sociaux sur le modèle de ce qui est fait pour la base SISE. En revanche, un regroupement en trois catégories de PCS nous semble possible.

Si les deux groupes intermédiaires de la base SISE sont regroupés en un seul, il est alors possible d'établir une correspondance entre les 3 groupes nouvellement obtenus et ceux du panel.

Groupe social	Code PCS	Désignation de la profession
Intermédiaire	01	Agriculteur exploitant
Intermédiaire	02	Artisan, Commerçant, Chef d'entreprise de moins de 10 salariés
Très favorisé	03	Chef d'entreprise de 10 salariés ou plus, profession libérale
Très favorisé	04	Professeur de l'enseignement supérieur ou secondaire et assimilé
Très favorisé	05	Chercheur de la recherche publique ou privée
Très favorisé	06	Cadre et profession intellectuelle supérieure
Très favorisé	07	Professeur des écoles, de Collège ou de Lycée professionnel, instituteur
Intermédiaire	08	Profession intermédiaire
Intermédiaire	09	Employé
Défavorisé	10	Ouvrier qualifié, ouvrier non qualifié ou ouvrier agricole
Défavorisé	11	Inactif n'ayant jamais travaillé
Défavorisé	12	Non concerné
Défavorisé	99	Non réponse

Source : Panel bacheliers 2014 (SIES-MESRI). Définition des groupes sociaux : auteur.

Il convient de noter que la catégorie non concerné correspond probablement aux retraités. Toutefois, nous ne disposons pas à ce stade de davantage de précision concernant cette catégorie.

Annexe 8 : Résultats des calculs des indices

Sauf indication différente, les indices calculés ci-dessous sont corrigés du biais lié aux petites unités.

Les indices calculés au moyen d'une méthode de simulation (*Bootstrap*) sont présentés après les indices corrigés du biais lié aux petites unités.

Indices de mixité sociale pour l'Université hors doctorat (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,034	0,034	0,035	0,033	0,036	0,028	0,056	0,078	0,031	0,030	0,024	0,043	0,041
2007	0,034	0,034	0,035	0,032	0,036	0,028	0,055	0,077	0,031	0,030	0,024	0,043	0,042
2008	0,034	0,035	0,035	0,033	0,036	0,028	0,056	0,078	0,031	0,030	0,024	0,043	0,041
2009	0,032	0,033	0,033	0,031	0,034	0,027	0,054	0,075	0,030	0,029	0,023	0,044	0,039
2010	0,035	0,035	0,035	0,032	0,037	0,029	0,056	0,078	0,033	0,032	0,025	0,048	0,045
2011	0,033	0,033	0,033	0,030	0,036	0,028	0,056	0,077	0,032	0,031	0,024	0,047	0,044
2012	0,032	0,033	0,033	0,029	0,035	0,027	0,055	0,076	0,031	0,030	0,023	0,046	0,043
2013	0,031	0,031	0,031	0,028	0,033	0,026	0,054	0,075	0,030	0,029	0,022	0,045	0,040
2014	0,030	0,031	0,031	0,028	0,033	0,026	0,054	0,074	0,030	0,029	0,022	0,044	0,039
2015	0,031	0,031	0,031	0,028	0,033	0,026	0,055	0,075	0,030	0,029	0,022	0,044	0,040
2016	0,030	0,030	0,030	0,027	0,032	0,025	0,054	0,073	0,029	0,028	0,021	0,044	0,038

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure 2006 de 2006 à 2016.

Les abréviations de la première ligne désignent les indices suivants :

A : Atkinson (structure de l'année courante)

A_2006 : Atkinson (structure 2006)

A_2016 : Atkinson (structure 2016)

As : Atkinson symétrique

M : Indice d'Information mutuelle (entropie)

H : Theil (indice d'Entropie Normalisé)

D : Indice de dissimilarité

G : Gini multigroupes

E : Indice d'Exposition Multigroupe

DR : Indice de diversité relative

C : Indice du coef. de variation au carré

E_tfav : Indice d'Exposition Très Favorisés

E_def : Indice d'Exposition Defavorisés

Indices de mixité sociale pour l'ensemble de la Licence (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,038	0,039	0,039	0,037	0,040	0,031	0,060	0,084	0,034	0,033	0,027	0,047	0,046
2007	0,038	0,039	0,039	0,036	0,039	0,031	0,060	0,083	0,033	0,033	0,027	0,048	0,045
2008	0,038	0,038	0,038	0,037	0,039	0,031	0,060	0,083	0,033	0,033	0,027	0,048	0,045
2009	0,036	0,037	0,037	0,035	0,038	0,030	0,058	0,081	0,033	0,032	0,026	0,048	0,045
2010	0,038	0,038	0,039	0,036	0,040	0,032	0,060	0,083	0,036	0,035	0,027	0,052	0,050
2011	0,037	0,037	0,038	0,035	0,040	0,031	0,059	0,083	0,035	0,034	0,027	0,050	0,049
2012	0,036	0,037	0,037	0,034	0,039	0,030	0,058	0,081	0,035	0,034	0,026	0,050	0,048
2013	0,035	0,036	0,036	0,033	0,038	0,029	0,058	0,081	0,033	0,032	0,025	0,049	0,045
2014	0,034	0,035	0,035	0,032	0,036	0,029	0,056	0,079	0,032	0,031	0,025	0,048	0,044
2015	0,035	0,036	0,036	0,033	0,037	0,029	0,058	0,080	0,033	0,032	0,025	0,049	0,045
2016	0,033	0,034	0,035	0,032	0,036	0,028	0,057	0,078	0,031	0,030	0,024	0,047	0,041

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour le Master (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,040	0,040	0,040	0,040	0,042	0,033	0,060	0,083	0,037	0,036	0,029	0,042	0,054
2007	0,043	0,043	0,044	0,042	0,045	0,035	0,061	0,085	0,040	0,038	0,030	0,044	0,060
2008	0,043	0,043	0,043	0,042	0,045	0,035	0,062	0,086	0,039	0,037	0,030	0,044	0,058
2009	0,040	0,040	0,040	0,039	0,042	0,033	0,059	0,082	0,036	0,035	0,029	0,046	0,051
2010	0,042	0,042	0,043	0,039	0,044	0,035	0,060	0,085	0,040	0,039	0,029	0,052	0,057
2011	0,040	0,040	0,040	0,036	0,042	0,033	0,061	0,084	0,039	0,038	0,028	0,051	0,056
2012	0,040	0,040	0,040	0,036	0,042	0,033	0,061	0,084	0,039	0,037	0,028	0,051	0,055
2013	0,038	0,037	0,037	0,034	0,040	0,032	0,060	0,082	0,037	0,036	0,026	0,049	0,052
2014	0,037	0,036	0,037	0,033	0,038	0,031	0,059	0,080	0,036	0,035	0,026	0,048	0,051
2015	0,036	0,036	0,036	0,032	0,038	0,030	0,059	0,080	0,036	0,034	0,025	0,047	0,051
2016	0,036	0,036	0,036	0,032	0,039	0,031	0,058	0,079	0,037	0,035	0,026	0,048	0,052

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la première année de Licence (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,053	0,053	0,054	0,050	0,054	0,042	0,075	0,105	0,046	0,045	0,037	0,056	0,071
2007	0,052	0,052	0,052	0,048	0,053	0,041	0,073	0,103	0,045	0,044	0,036	0,056	0,071
2008	0,054	0,054	0,054	0,051	0,055	0,043	0,075	0,105	0,047	0,046	0,037	0,058	0,074
2009	0,051	0,051	0,051	0,048	0,052	0,041	0,072	0,101	0,046	0,045	0,035	0,057	0,070
2010	0,054	0,053	0,053	0,049	0,056	0,044	0,075	0,105	0,050	0,049	0,037	0,061	0,077
2011	0,052	0,051	0,051	0,047	0,053	0,041	0,074	0,102	0,048	0,046	0,035	0,060	0,073
2012	0,051	0,051	0,051	0,046	0,052	0,041	0,073	0,102	0,047	0,045	0,035	0,059	0,072
2013	0,050	0,050	0,050	0,046	0,051	0,040	0,073	0,101	0,046	0,044	0,035	0,060	0,069
2014	0,049	0,048	0,048	0,044	0,050	0,039	0,071	0,099	0,044	0,043	0,033	0,057	0,068
2015	0,049	0,049	0,049	0,045	0,050	0,039	0,072	0,100	0,044	0,043	0,034	0,058	0,067
2016	0,046	0,046	0,046	0,043	0,046	0,036	0,068	0,096	0,040	0,039	0,031	0,055	0,060

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la deuxième année de Licence (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,043	0,043	0,043	0,043	0,045	0,035	0,061	0,086	0,037	0,036	0,030	0,054	0,048
2007	0,044	0,044	0,044	0,043	0,044	0,034	0,061	0,085	0,037	0,036	0,030	0,053	0,047
2008	0,038	0,038	0,038	0,037	0,039	0,030	0,058	0,080	0,033	0,032	0,026	0,051	0,037
2009	0,038	0,038	0,038	0,038	0,038	0,030	0,057	0,079	0,032	0,031	0,026	0,050	0,036
2010	0,038	0,038	0,038	0,037	0,038	0,030	0,057	0,079	0,033	0,032	0,026	0,052	0,039
2011	0,038	0,038	0,038	0,037	0,039	0,031	0,058	0,080	0,034	0,033	0,027	0,053	0,043
2012	0,037	0,037	0,037	0,036	0,039	0,030	0,057	0,079	0,033	0,032	0,026	0,052	0,040
2013	0,037	0,037	0,037	0,036	0,038	0,030	0,057	0,079	0,033	0,032	0,025	0,051	0,039
2014	0,039	0,038	0,039	0,037	0,041	0,032	0,057	0,080	0,035	0,034	0,028	0,052	0,045
2015	0,039	0,038	0,039	0,037	0,039	0,031	0,057	0,080	0,034	0,033	0,027	0,052	0,045
2016	0,039	0,038	0,039	0,037	0,039	0,031	0,058	0,081	0,034	0,033	0,027	0,051	0,043

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la troisième année de Licence (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	DR	C	E_tfav	E_def
2006	0,043	0,043	0,043	0,044	0,044	0,034	0,061	0,087	0,036	0,035	0,030	0,047	0,048
2007	0,045	0,045	0,045	0,045	0,045	0,035	0,063	0,088	0,037	0,036	0,030	0,050	0,049
2008	0,047	0,047	0,047	0,048	0,047	0,037	0,065	0,091	0,039	0,038	0,032	0,053	0,052
2009	0,044	0,044	0,044	0,043	0,045	0,035	0,061	0,087	0,039	0,038	0,030	0,053	0,053
2010	0,046	0,046	0,046	0,045	0,047	0,037	0,064	0,090	0,040	0,039	0,031	0,056	0,055
2011	0,047	0,047	0,047	0,046	0,049	0,039	0,065	0,091	0,043	0,042	0,033	0,054	0,061
2012	0,047	0,046	0,047	0,045	0,048	0,038	0,064	0,090	0,042	0,041	0,032	0,053	0,061
2013	0,043	0,043	0,043	0,042	0,044	0,035	0,060	0,085	0,039	0,037	0,029	0,050	0,054
2014	0,042	0,041	0,042	0,040	0,042	0,033	0,059	0,083	0,037	0,036	0,028	0,050	0,050
2015	0,044	0,043	0,043	0,041	0,044	0,035	0,061	0,086	0,039	0,038	0,030	0,050	0,056
2016	0,042	0,041	0,042	0,039	0,042	0,033	0,059	0,084	0,037	0,036	0,028	0,048	0,053

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source: SISE Université (Diffuseur : Adisp-Progedo). Calculs de l'auteur.

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la première année de Licence calculés au moyen d'une méthode de *Bootstrap* (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	E_tfav	E_def	DR	C
2006	0,003	0,003	0,002	0,003	0,002	0,002	0,017	0,024	0,002	0,002	0,002	0,002	0,002
2007	0,003	0,003	0,003	0,003	0,002	0,002	0,017	0,025	0,002	0,002	0,002	0,002	0,002
2008	0,003	0,003	0,003	0,004	0,002	0,002	0,017	0,025	0,002	0,002	0,002	0,002	0,002
2009	0,003	0,003	0,003	0,004	0,002	0,002	0,016	0,024	0,002	0,002	0,002	0,002	0,002
2010	0,003	0,003	0,003	0,004	0,002	0,002	0,016	0,024	0,002	0,002	0,002	0,002	0,002
2011	0,003	0,003	0,003	0,004	0,002	0,002	0,016	0,023	0,002	0,002	0,002	0,002	0,002
2012	0,003	0,003	0,003	0,004	0,002	0,002	0,016	0,023	0,001	0,001	0,001	0,001	0,001
2013	0,002	0,003	0,003	0,003	0,002	0,002	0,015	0,023	0,001	0,001	0,001	0,001	0,001
2014	0,003	0,003	0,003	0,004	0,002	0,002	0,015	0,023	0,001	0,001	0,001	0,001	0,001
2015	0,002	0,003	0,002	0,003	0,002	0,002	0,015	0,022	0,001	0,001	0,001	0,001	0,001
2016	0,002	0,003	0,002	0,003	0,002	0,002	0,015	0,022	0,001	0,001	0,001	0,001	0,001

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat. Calculs de l'auteur au moyen d'une méthode de *Bootstrap* à partir d'une répartition aléatoire des étudiants (série de 100 simulations) au sein des formations observées réellement chaque année à partir de la base SISE Université (Diffuseur : Adisp-Progedo).

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la deuxième année de Licence calculés au moyen d'une méthode de *Bootstrap* (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	E_tfav	E_def	DR	C
2006	0,005	0,005	0,005	0,006	0,004	0,003	0,022	0,033	0,003	0,003	0,003	0,003	0,003
2007	0,005	0,005	0,005	0,006	0,004	0,003	0,022	0,033	0,003	0,003	0,003	0,003	0,003
2008	0,005	0,005	0,005	0,007	0,004	0,003	0,022	0,033	0,003	0,003	0,003	0,003	0,003
2009	0,005	0,005	0,005	0,007	0,004	0,003	0,022	0,033	0,003	0,003	0,003	0,003	0,003
2010	0,005	0,005	0,005	0,007	0,004	0,003	0,022	0,032	0,003	0,003	0,003	0,003	0,003
2011	0,005	0,005	0,005	0,007	0,004	0,003	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2012	0,005	0,005	0,005	0,007	0,004	0,003	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2013	0,005	0,005	0,005	0,007	0,004	0,003	0,021	0,031	0,003	0,003	0,003	0,003	0,003
2014	0,005	0,005	0,005	0,006	0,004	0,003	0,021	0,031	0,003	0,003	0,003	0,003	0,003
2015	0,005	0,005	0,005	0,006	0,004	0,003	0,021	0,031	0,003	0,003	0,003	0,003	0,003
2016	0,005	0,005	0,005	0,006	0,004	0,003	0,02	0,03	0,002	0,002	0,002	0,002	0,002

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat. Calculs de l'auteur au moyen d'une méthode de *Bootstrap* à partir d'une répartition aléatoire des étudiants (série de 100 simulations) au sein des formations observées réellement chaque année à partir de la base SISE Université (Diffuseur : Adisp-Progedo).

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la troisième année de Licence calculés au moyen d'une méthode de *Bootstrap* (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	E_tfav	E_def	DR	C
2006	0,005	0,005	0,004	0,006	0,004	0,003	0,021	0,031	0,003	0,003	0,003	0,003	0,003
2007	0,005	0,005	0,004	0,006	0,004	0,003	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2008	0,005	0,005	0,005	0,006	0,004	0,003	0,022	0,032	0,003	0,003	0,003	0,003	0,003
2009	0,005	0,005	0,005	0,007	0,004	0,003	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2010	0,005	0,006	0,005	0,007	0,004	0,003	0,021	0,031	0,003	0,003	0,003	0,003	0,003
2011	0,005	0,006	0,005	0,007	0,004	0,003	0,02	0,03	0,003	0,003	0,003	0,003	0,003
2012	0,005	0,006	0,005	0,007	0,004	0,003	0,02	0,03	0,003	0,003	0,003	0,003	0,003
2013	0,005	0,006	0,005	0,007	0,004	0,003	0,02	0,03	0,003	0,003	0,003	0,003	0,003
2014	0,005	0,006	0,005	0,007	0,004	0,003	0,02	0,03	0,003	0,003	0,003	0,003	0,003
2015	0,005	0,006	0,005	0,007	0,004	0,003	0,02	0,03	0,003	0,003	0,003	0,003	0,003
2016	0,005	0,005	0,005	0,007	0,004	0,003	0,02	0,029	0,002	0,002	0,003	0,002	0,002

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat. Calculs de l'auteur au moyen d'une méthode de *Bootstrap* à partir d'une répartition aléatoire des étudiants (série de 100 simulations) au sein des formations observées réellement chaque année à partir de la base SISE Université (Diffuseur : Adisp-Progedo).

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la première année de Master calculés au moyen d'une méthode de *Bootstrap* (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	E_tfav	E_def	DR	C
2006	0,0063	0,0063	0,0061	0,008	0,005	0,0039	0,0229	0,0341	0,0033	0,0033	0,0033	0,0033	0,0033
2007	0,0067	0,0067	0,0065	0,0084	0,0052	0,004	0,0232	0,0347	0,0034	0,0034	0,0034	0,0034	0,0034
2008	0,006	0,006	0,0058	0,0076	0,0047	0,0036	0,0219	0,0328	0,0031	0,0031	0,0031	0,0031	0,0031
2009	0,0057	0,0058	0,0056	0,0075	0,0044	0,0034	0,0209	0,0314	0,0029	0,0029	0,0029	0,0029	0,0029
2010	0,0069	0,0071	0,0068	0,0089	0,0048	0,0038	0,0219	0,0328	0,0032	0,0032	0,0032	0,0032	0,0032
2011	0,007	0,0072	0,007	0,009	0,0048	0,0038	0,0215	0,0322	0,0031	0,0031	0,0031	0,0031	0,0031
2012	0,007	0,0072	0,007	0,009	0,0047	0,0037	0,0213	0,0319	0,0031	0,0031	0,0031	0,0031	0,003
2013	0,007	0,0072	0,007	0,0088	0,0045	0,0036	0,0199	0,0304	0,0029	0,0029	0,0029	0,0029	0,0029
2014	0,0067	0,0069	0,0067	0,0086	0,0044	0,0035	0,0196	0,0301	0,0029	0,0028	0,0029	0,0029	0,0029
2015	0,0066	0,0069	0,0067	0,0085	0,0043	0,0034	0,0193	0,0296	0,0028	0,0028	0,0028	0,0028	0,0028
2016	0,0066	0,0068	0,0066	0,0085	0,0043	0,0034	0,0193	0,0298	0,0028	0,0028	0,0028	0,0028	0,0028

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat. Calculs de l'auteur au moyen d'une méthode de *Bootstrap* à partir d'une répartition aléatoire des étudiants (série de 100 simulations) au sein des formations observées réellement chaque année à partir de la base SISE Université (Diffuseur : Adisp-Progedo).

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure de l'année courante de 2006 à 2016.

Indices de mixité sociale pour la deuxième année de Master calculés au moyen d'une méthode de *Bootstrap* (2006-2016)

	A	A_2006	A_2016	As	M	H	D	G	E	E_tfav	E_def	DR	C
2006	0,009	0,009	0,009	0,011	0,006	0,005	0,024	0,036	0,004	0,004	0,004	0,004	0,004
2007	0,009	0,009	0,009	0,011	0,006	0,005	0,024	0,035	0,004	0,004	0,004	0,004	0,004
2008	0,008	0,008	0,008	0,01	0,006	0,004	0,023	0,034	0,004	0,004	0,004	0,004	0,004
2009	0,008	0,008	0,008	0,01	0,005	0,004	0,023	0,034	0,004	0,004	0,004	0,004	0,004
2010	0,008	0,008	0,008	0,01	0,005	0,004	0,022	0,033	0,003	0,003	0,003	0,003	0,003
2011	0,008	0,008	0,008	0,01	0,005	0,004	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2012	0,008	0,009	0,008	0,01	0,005	0,004	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2013	0,008	0,008	0,008	0,01	0,005	0,004	0,021	0,032	0,003	0,003	0,003	0,003	0,003
2014	0,008	0,008	0,008	0,01	0,005	0,004	0,019	0,03	0,003	0,003	0,003	0,003	0,003
2015	0,008	0,008	0,008	0,01	0,004	0,004	0,018	0,029	0,003	0,003	0,003	0,003	0,003
2016	0,008	0,008	0,008	0,01	0,004	0,004	0,018	0,029	0,003	0,003	0,003	0,003	0,003

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat. Calculs de l'auteur au moyen d'une méthode de *Bootstrap* à partir d'une répartition aléatoire des étudiants (série de 100 simulations) au sein des formations observées réellement chaque année à partir de la base SISE Université (Diffuseur : Adisp-Progedo).

Note de lecture : La première colonne fournit les valeurs de l'indice d'Atkinson en structure de l'année courante de 2006 à 2016.

Annexe 9 : décomposition de l'indice d'information mutuelle

Décomposition de l'indice d'information mutuelle suivant la discipline et les zones académiques. 2006-2016

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Indice M	0,036	0,036	0,036	0,034	0,037	0,036	0,035	0,033	0,033	0,033	0,032
Inter-AC	0,01	0,009	0,009	0,009	0,009	0,009	0,009	0,009	0,009	0,009	0,008
Intra-AC et inter-DIS	0,017	0,018	0,018	0,017	0,02	0,019	0,019	0,018	0,018	0,018	0,017
Intra-AC, intra-DIS, inter-UNIV	0,009	0,009	0,009	0,008	0,008	0,007	0,007	0,007	0,006	0,006	0,007
Inter-DIS	0,011	0,012	0,012	0,011	0,013	0,012	0,012	0,012	0,011	0,011	0,011
Intra-DIS, Inter-UNIV	0,025	0,024	0,024	0,023	0,024	0,023	0,023	0,022	0,021	0,022	0,021

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source : SISE (Diffuseur : Adisp-Progedo) – Calculs de l'auteur.

Lecture: La première ligne indique la valeur de l'indice chaque année. En 2006, la valeur de celui-ci est de 0,036. La deuxième ligne indique la valeur de l'indice au niveau national entre les académies, comme si l'unité d'observation était non plus la formation, mais l'académie. En 2006, la valeur de celui-ci est de 0,01.

**Décomposition de l'indice d'information mutuelle calculé par niveau d'études
suivant les zones académiques et les disciplines (moyenne 2006-2016)**

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
L1	0,054	0,053	0,055	0,052	0,056	0,053	0,052	0,051	0,050	0,050	0,046
Inter-AC	26%	23%	23%	24%	26%	26%	29%	28%	27%	28%	28%
Intra-AC et inter-DIS	49%	53%	53%	52%	51%	54%	52%	53%	55%	55%	55%
Intra-AC, intra-DIS, inter-UNIV	25%	24%	24%	24%	23%	20%	20%	19%	18%	17%	17%
Inter-DIS	33%	37%	36%	38%	36%	39%	36%	37%	39%	37%	39%
Intra-DIS, Inter-UNIV	67%	63%	64%	62%	64%	61%	64%	63%	61%	63%	61%
L2	0,045	0,044	0,039	0,038	0,038	0,039	0,039	0,038	0,041	0,039	0,039
Inter-AC	29%	28%	29%	31%	31%	33%	32%	34%	30%	29%	30%
Intra-AC et inter-DIS	49%	51%	47%	47%	45%	42%	44%	43%	46%	47%	49%
Intra-AC, intra-DIS, inter-UNIV	22%	21%	24%	22%	24%	25%	24%	23%	23%	24%	22%
Inter-DIS	36%	37%	36%	34%	33%	31%	32%	31%	28%	28%	30%
Intra-DIS, Inter-UNIV	64%	63%	64%	66%	67%	69%	68%	69%	72%	72%	70%
L3	0,044	0,045	0,047	0,045	0,047	0,049	0,048	0,044	0,042	0,044	0,042
Inter-AC	26%	24%	24%	23%	26%	23%	22%	21%	23%	24%	24%
Intra-AC et inter-DIS	47%	49%	48%	48%	47%	52%	52%	52%	50%	50%	48%
Intra-AC, intra-DIS, inter-UNIV	28%	28%	28%	29%	27%	25%	26%	26%	27%	26%	28%
Inter-DIS	29%	30%	31%	32%	30%	27%	25%	28%	28%	27%	25%
Intra-DIS, Inter-UNIV	71%	70%	69%	68%	70%	73%	75%	72%	72%	73%	75%
M	0,042	0,045	0,045	0,042	0,044	0,042	0,042	0,040	0,038	0,038	0,039
Inter-AC	24%	22%	23%	23%	24%	24%	24%	26%	28%	30%	25%
Intra-AC et inter-DIS	50%	52%	55%	56%	57%	55%	55%	55%	56%	56%	57%
Intra-AC, intra-DIS, inter-UNIV	26%	26%	22%	20%	20%	21%	21%	19%	16%	14%	17%
Inter-DIS	25%	26%	29%	31%	32%	31%	31%	32%	32%	32%	34%
Intra-DIS, Inter-UNIV	75%	74%	71%	69%	68%	69%	69%	68%	68%	68%	66%
UNIV	0,036	0,036	0,036	0,034	0,037	0,036	0,035	0,033	0,033	0,033	0,032
Inter-AC	24%	22%	23%	23%	24%	24%	24%	26%	28%	30%	25%
Intra-AC et inter-DIS	50%	52%	55%	56%	57%	55%	55%	55%	56%	56%	57%
Intra-AC, intra-DIS, inter-UNIV	26%	26%	22%	20%	20%	21%	21%	19%	16%	14%	17%
Inter-DIS	25%	26%	29%	31%	32%	31%	31%	32%	32%	32%	34%
Intra-DIS, Inter-UNIV	75%	74%	71%	69%	68%	69%	69%	68%	68%	68%	66%

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source : SISE (Diffuseur : Adisp-Progedo) – Calculs de l'auteur.

Lecture: La première ligne indique la valeur de l'indice chaque année pour la L1. En 2006, la valeur de celui-ci est de 0,054. La deuxième ligne indique la part de ce niveau d'ensemble qui est lié aux différences entre académies, comme si l'unité d'observation était non plus la formation, mais l'académie. En 2006, cette part était d'environ 26% du total.

**Décomposition de l'indice d'information mutuelle calculé pour Paris et la Province
suivant les zones académiques et les disciplines (moyenne 2006-2016)**

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Région parisienne	0,038	0,037	0,037	0,029	0,038	0,040	0,038	0,038	0,037	0,039	0,039
Inter-AC	17,65%	16,98%	19,33%	23,17%	21,88%	24,02%	24,93%	20,64%	20,39%	20,59%	20,18%
Intra-AC et inter-DIS	37,81%	40,28%	40,84%	44,35%	48,76%	44,48%	47,18%	49,09%	50,72%	48,39%	39,65%
Intra-AC, intra-DIS, inter-UNIV	44,54%	42,74%	39,83%	32,48%	29,36%	31,50%	27,89%	30,27%	28,89%	31,02%	40,17%
Inter-DIS	29,55%	33,81%	35,56%	37,71%	42,76%	39,81%	41,89%	41,63%	38,92%	36,17%	33,56%
Intra-DIS, Inter-UNIV	70,45%	66,19%	64,44%	62,29%	57,24%	60,19%	58,11%	58,37%	61,08%	63,83%	66,44%
France (hors région parisienne)	0,032	0,033	0,032	0,030	0,033	0,029	0,029	0,029	0,028	0,029	0,028
Inter-AC	22,02%	20,90%	20,37%	18,50%	18,32%	18,40%	18,74%	20,30%	19,95%	21,95%	19,93%
Intra-AC et inter-DIS	56,86%	59,05%	59,31%	58,50%	60,72%	63,05%	62,12%	62,24%	63,10%	61,89%	62,43%
Intra-AC, intra-DIS, inter-UNIV	21,12%	20,05%	20,32%	23,00%	20,96%	18,55%	19,14%	17,46%	16,95%	16,16%	17,64%
Inter-DIS	39,69%	42,07%	42,34%	43,02%	44,33%	46,06%	45,52%	45,82%	45,52%	42,41%	43,13%
Intra-DIS, Inter-UNIV	60,31%	57,93%	57,66%	56,98%	55,67%	53,94%	54,48%	54,18%	54,48%	57,59%	56,87%

Champ : ensemble des étudiants de France métropolitaine inscrits à l'Université hors doctorat.

Source : SISE (Diffuseur : Adisp-Progedo) – Calculs de l'auteur.

Lecture: La première ligne indique la valeur de l'indice chaque année pour la région parisienne. En 2006, la valeur de celui-ci est de 0,038. La deuxième ligne indique la part de ce niveau d'ensemble qui est liée aux différences entre académies. En 2006, cette part était d'environ 17,6% du total.

**Annexe 10 : taux de passage, de réorientation et d'abandon
à l'issue de la Licence 1 par discipline**

Discipline	Taux de passage	Taux de redoublement	Taux de changement d'orientation	Taux de sortie de l'Université
Ensemble	40,8%	29,0%	2,5%	27,7%
DROIT - SCIENCES POLITIQUES	40,2%	35,9%	2,1%	21,9%
SCIENCES ECONOMIQUES - GESTION (HORS A.E.S.)	38,3%	28,5%	3,2%	29,9%
ADMINISTRATION ECONOMIQUE ET SOCIALE (A.E.S.)	25,1%	32,2%	2,7%	40,0%
LETTRES - SCIENCES DU LANGAGE - ARTS	47,6%	22,3%	1,4%	28,7%
LANGUES	39,9%	27,3%	1,8%	31,0%
SCIENCES HUMAINES ET SOCIALES	44,6%	24,4%	2,0%	29,0%
SCIENCES DE LA VIE, DE LA TERRE ET DE L'UNIVERS	39,7%	28,7%	5,0%	26,6%
S.T.A.P.S.	39,1%	31,9%	2,0%	27,0%
SCIENCES FONDAMENTALES ET APPLICATIONS	42,3%	28,8%	4,5%	24,5%

Champ : néobacheliers entrés en 1ère année de Licence en 2014-2015. Universités et établissements assimilés, France entière hors Nouvelle Calédonie (données indisponibles en 2015-2016)

Source: MESR-DGESIP-DGRI-SIES (<https://www.enseignementsup-recherche.gouv.fr/fr/parcours-et-reussite-en-Licence-et-en-paces-les-resultats-de-la-session-2016-47793>)

Annexe 11 : liste des acronymes

APB : Admission post-bac

ADISP : Archives de Données Issues de la Statistique Publique

BTS : brevet de technicien supérieur

BCS : bourses d'enseignement supérieur sur critères sociaux

CPGE : Classe Préparatoire aux Grandes Ecoles.

DEPP : Direction de l'évaluation, de la prospective et de la performance

DUT : diplôme universitaire technologique

EESPIG : établissement d'enseignement supérieur privé d'intérêt général

EPSCP : établissements publics à caractère scientifique, culturel et professionnel

IAE : Instituts d'Administration des Entreprises

IEN : Indice d'exposition normalisé

INP : Institut National Polytechnique

IPAG : Instituts de Préparation à l'Administration Générale

IUT : Institut Universitaire de Technologie

Loi LRU : loi relative aux libertés et responsabilités des Universités

MESRI : Ministère de l'Enseignement Supérieur et de la Recherche

PCS : professions et catégories socioprofessionnelles

PISA : programme international pour le suivi des acquis des élèves

PROGEDO - Production et gestion des données en sciences sociales

RAVEL : recensement automatisé des vœux des élèves

SISE : Système d'information de Suivi de l'étudiant

STS : sections de techniciens supérieurs

UFR : Unités de Formation et de Recherche

ZEP : Zone d'éducation prioritaire

Bibliographie indicative

Abraham Y (2007), Du souci scolaire au sérieux managérial, ou comment devenir un HEC. *Revue française de sociologie*, 48 (1): 37-66.

Albouy V, Tavan C (2007), Accès à l'enseignement supérieur en France : une démocratisation réelle mais de faible ampleur. In: *Economie et statistique*, n°410, 2007: 3-22.

Algan Y, Huillery É, Prost C (2018), Confiance, coopération et autonomie : pour une école du XXI^{ème} siècle, Notes du conseil d'analyse économique, 2018/3 (n° 48) :1-12.

Allen M, Carter C (2007), Academic success determinants for undergraduate real estate students - *Journal of Real Estate Practice and Education*.

Arulampalam W, Naylor R, Smith J (2004). Factors affecting the probability of first year medical student dropout in the UK: A logistic analysis for intake cohorts of 1980-92. *Medical Education*, 38, 492-503.

Atkinson AB (1970), On the measurement of economic inequality. *Journal of Economic Theory*, 2 (3), : 244-263.

Avouac R, Harari-Kermadec H (2021), French Universities – A Melting Pot or a Hotbed of Social Segregation? A Measure of Polarisation within the French University System (2007-2015). *Economie et Statistique*, 528-529, 68–83.

Bailey N (2012), How spatial segregation changes over time: sorting out the sorting processes, *Environment and Planning A*, 44, 705-722.

Baron M (2005), Les migrations étudiantes dans le système universitaire français au début des années 90. *Revue d'Economie Regionale Urbaine*, (2), 281-300.

Beaubrun-Diant K, Maury TP, (2021), Income segregation and social housing in France, preprint, hal-02526776.

Bechichi N, Grenet J, Thebault G (2021), Ségrégation à l'entrée des études supérieures en France et en région parisienne : quels effets du passage à Parcoursup ? INSEE, Document de travail, N° 2021-003.

Beffy M, Fougère D, Maurel A (2009), L'impact du travail salarié des étudiants sur la réussite et la poursuite des études universitaires. In: *Economie et statistique*, n°422, 2009. : 1-50.

Bluntz C (2018), Choix d'orientation en terminale et mobilité géographique. *Note d'information Enseignement sup.* – Recherche, Publication de la DEPP.

Bodson X (1999), La vie d'étudiant : Analyse des manières d'être et de leurs conséquences sur la réussite en première candidature à l'UCL. Thèse de doctorat non publiée, Faculté des sciences économiques, sociales et politiques, Université Catholique de Louvain, Louvain-la-Neuve.

Bonneau C et S Grobon (2021): Enseignement supérieur : un accès inégal selon le revenu des parents, Focus du CAE, n° 76-2021, décembre.

Bonneau C, Charoussat P, Grenet J, Thebault G, (2022), Grandes écoles: des politiques d'ouverture sociale en échec, *Revue Éducation & formations*, n°103.

- Bourdieu P (1989) *La Noblesse d'État. Grandes Ecoles et esprit de corps*, Les Editions de Minuit, Paris.
- Bourguignon (1979), Decomposable income inequality measures. *Econometrica*, 47(4): 901–920.
- Boutchénik B, Coron C, Grobon S, Goffette C et Vallet L-A (2015), Quantifier l'influence totale de la famille d'origine sur le devenir scolaire et professionnel des individus. *Economie et Statistique*, 477, 5–23.
- Boutchenik B, Maillard S (2018), Peer Effects with Peer and Student Heterogeneity: An Assessment for French Baccalauréat, Actes des XIIIes Journées de Méthodologie Statistique de l'INSEE, 12-14 juin 2018.
- Boutchenik B, Givord P, Monso O (2018), La ségrégation sociale entre Collèges : un reflet de la ségrégation résidentielle nettement amplifié par les choix des familles, notamment vers l'enseignement privé. INSEE ANALYSES numéro 40, septembre 2018.
- Boutchenik B, Givord P, Monso O (2020) How do restrictive zoning and parental choices impact social diversity in schools?, *Science Po LIEPP Working Paper*, n°105, 2020-02-11.
- Brinbaum Y, Huguée C, Poullaouec T (2018) 50% to the bachelor's degree... but how? Young people from working-class families at university in France. *Economie et Statistique*, 499: 79–105.
- Brodaty T, Jacotin G (2016), Peut-on comparer les performances des Universités?. *Revue économique*, 67(2), 315-336.
- Burke MA, Sass T R, (2008), Classroom peer effects and student achievement. *Calder Urban Institute Working Paper 18*.
- Burke MA, Sass TR (2013), Classroom Peer Effects and Student Achievement, *Journal of Labor Economics*, vol. 31, n° 1, p. 51-82.
- Carlson, Susan M (1992), Trends in Race/Sex Occupational Inequality: Conceptual and Measurement Issues. *Social Problems* 39:269–90.
- Carrell SE, Fullerton RL, West JE, (2009), Does your cohort matter? Measuring peer effects in college achievement. *J. Labor Econ.* 27 (3).
- Chéron A, Courtioux P (2018), Les bénéfices socioéconomiques des diplômes du supérieur. *Position Paper EDHEC*, May.
- Courtioux P (2011), L'origine sociale joue-t-elle sur le rendement des études supérieures ?. *Position Paper EDHEC*, Novembre.
- Courtioux P, Maury T-P (2018) Public, privé et éducation prioritaire : une analyse comparative de la mixité sociale à l'école. *Formation-Emploi*, 144 : 133-154.
- Courtioux P, Maury T-P (2020), Private and public schools: A spatial analysis of social segregation in France, *Urban Studies*, 57(4), 865-882.
- Courtioux P, Maury T-P, Seux J, (2022) Ségrégation sociale à l'Université : Des disparités académiques sur la période 2006-2016 », *Revue D'économie Régionale et Urbaine* (3/2022), :385-417.
- Davezies L (2005), Influence des caractéristiques des pairs sur la scolarité, *Educations & formations*, n°72, MENESR-DEPP, : 171-199.

Degrave F, Martou F (1996), Efficacité de l'enseignement: Une question de mesure à propos de l'enseignement en Communauté Française de Belgique. In *Education et Formation, facteurs de compétitivité* (12ème Congrès des économistes belges de langue française, Commission 4, pp. 143-167). Charleroi: Centre Inter-universitaire de Formation Permanente.

Duguet A (2014), Les pratiques pédagogiques en première année universitaire : description et analyse de leurs implications sur la scolarité des étudiants. Dijon : thèse de doctorat, Université de Bourgogne.

Fabre J, Pawlowski E (2019), Aller étudier ailleurs après le Baccalauréat : entre effet de la géographie et de l'offre de formation, *INSEE Première*, n° 1727.

Fack G, Grenet J (2009), Sectorisation des collèges et prix immobiliers à Paris, Post-Print hal-02464106, HAL.

Fack G, Grenet J (2013), « Improving College access and success for low-income students: evidence from a large need-based grant program », *Paris School of Economics Working paper* n°2013-33.

Fack G, Huillery É (2021), « Enseignement supérieur : pour un investissement plus juste et plus efficace », *Notes du conseil d'analyse économique*, 2021/8 (n° 68), p. 1-12.

Felouzis G (2000), Repenser les inégalités à l'Université. Des inégalités sociales aux inégalités locales dans trois disciplines universitaires. In: *Sociétés contemporaines* N°38, : 67-97.

Foster G (2006), It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *J. Public Econ.* 90, 1455–1475.

Fougère D, Kiefer N, Monso O, Pirus C (2017), La concentration des enfants étrangers dans les classes de Collèges, *Éducation & formations*, n°95, MEN-DEPP, : 139-172.

Fougère D, Givord P, Monso O, Pirus C (2018) Les Effets De Pairss En Éducation : Comment Sont-Ils Mesurés Et Pour Quelles Conclusions ?, *Journées de méthodologie statistique de l'INSEE*, Juillet 2018.

Frankel DM, Volij O (2011) Measuring School Segregation. *Journal of Economic Theory*, 146(1): 1-38.

Frouillou L (2016) *Les mécanismes d'une ségrégation universitaire francilienne : carte universitaire et sens du placement étudiant*. Thèse de géographie. Université Paris I Panthéon-Sorbonne.

Galand B, Frenay M, Bourgeois E (2004), Facteurs de réussite en 1° candidature à la Faculté des Sciences et à l'Institut d'Éducation Physique et de Réadaptation. Communication orale à la Journée d'étude de la Chaire de pédagogie universitaire, Louvain-la-Neuve, Belgique.

Galand B, Neuville S, Frenay M (2005), L'échec à l'Université en Communauté Française de Belgique : Comprendre pour mieux prévenir ? – *Cahiers de Recherche en Éducation et Formation* 39 (5-17).

Geiss R, Keslassy E, Laurent D (2006) Ouvrir les Grandes Ecoles à la diversité. *Rapport de l'Institut Montaigne*.

Gibbons S, Telhal S (2016), Peer Effects: Evidence from Secondary School Transition in England , *Oxford Bulletin of Economics and Statistics*, vol. 78, n°4, p. 548-575.

- Giret JF et Morlaix S (dir.) (2019), *Regards croisés sur les conditions de réussite dans l'enseignement supérieur Français*. Paris. Cnesco.
- Givord P, Guillerm M, Monso O, Murat F, Afsa C (2015) La ségrégation sociale entre les Collèges. Une application à la composition sociale des Collèges français, *Actes des XII^{ème} Journées de Méthodologie Statistique*, Paris, 31 mars-2 avril 2015.
- Givord P, Guillerm M, Monso O, Murat F (2016), Comment mesurer la ségrégation dans le système éducatif ? Une étude de la composition sociale des Collèges français. *Education et Formations* 91: 21-51.
- Goodman L, William HK (1954). Measures of Association for Cross- Classifications. *Journal of the American Statistical Association* 49:732– 64.
- Grenet et Souidi (2021), Renforcer la mixité sociale au Collège : une évaluation des secteurs multi-Collèges à Paris, *Rapport IPP n°31 – Février 2021*.
- Grignon C, Gruel L (1999), *La Vie étudiante*, Paris PUF.
- Grignon C (2000), *Rapport À Monsieur Le Ministre De L'éducation Nationale, Les Étudiants En Difficulté Pauvreté Et Précarité*, OVE, mai 2000.
- Guyon N, Huillery E, (2016) Biased Aspirations and Social Inequality at School: Evidence from French Teenagers, *Sciences Po publications* 44, Sciences Po.
- Hauser RM, Wong R (1989), Sibling Resemblance and Inter-Sibling Effects in Educational Attainment, *Sociology of Education*, vol. 62, n° 3, pp. 149-171.
- Hoel J, Parker J, Rivenburg J (2005), A test for classmate peer effects in higher education. *Reed College* (September 2005).
- Hoxby C, Weingarth G, (2005), Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects, *Working paper Harvard University*.
- Hubbard P (2008), Regulating the social impacts of studentification: a Loughborough case study. *Environment and Planning A* 40(2): 323-341.
- Hutchens R (2001), Numerical measures of segregation: desirable properties and their implications. *Mathematical Social Sciences* 42(1): 13–29.
- INSEE (2020) statistique ethnique, lien internet publié le 27/08/2020 et consulté le 14 octobre 2022.
- Ireson J, Hallam S (2001), *Ability Grouping in Education*. Londres : Sage.
- Jahn, Julius, Schmid, Calvin F, Shrag, Clarence (1947), The Measurement of Ecological Segregation. *American Sociological Review*, 12, 293–303.
- James D, Taeuber K (1985), Measures of Segregation. *Sociological Methodology* 15:1-32.
- Kaye N (2020), Evaluating the role of bursaries in widening participation in higher education: a review of the literature and evidence. *Educational Review*. 73. 1-23.
- Kottmann A, Vossensteyn JJ, Kolster R, Veidemane A, Blasko Z, Biagi F, Sánchez-Barrioluengo F(2019), *Social Inclusion Policies in Higher Education: Evidence from the EU. Overview of major widening participation policies applied in the EU 28*, EUR 29801 EN, Publications Office of the European Union.
- Kuzmanic D, Valenzuela JP, Villalobos C (2021), Socioeconomic Segregation in Higher Education: Evidence for Chile (2009–2017). *High Educ Policy* (2021).

Lambert-Le Mener M (2012), La performance académique des étudiants en première année universitaire : influence des capacités cognitives et de la motivation (Thèse de doctorat). Université de Bourgogne, Dijon.

Lemaire S (2000), Les facteurs de réussite dans les deux premières années d'enseignement supérieur (Deug, DUT, BTS) – *Note d'information 00-25*.

Lima L, Nakhili N, Le Hénaff B (2019), Être étudiant de première génération en France, cela fait-il une différence ? In Giret JF, Belghith F, Tenret E (Eds.), *Regards croisés sur les expériences étudiantes* (p. 23-37). Paris: La Documentation Française.

Ly ST, Maurin E et Riegert A (2014), La mixité sociale et scolaire en Île-de-France : le rôle des établissements. *Rapport de l'Institut des Politiques Publiques* n°4.

Ly ST, Riegert A (2015), Mixité sociale et scolaire, ségrégation inter et intra établissement dans les Collèges et Lycées français. *Rapport du Conseil National d'Évaluation du Système Scolaire*. La Documentation Française, Paris.

Manski C (1993), Identification of Endogenous Social Effects: The Reflection Problem, *The Review of Economic Studies*, vol. 60, n°3 : 531-542.

Merle P (2011), « La carte scolaire et son assouplissement. Politique de mixité sociale ou de ghettoïsation des établissements ? », in Presses Universitaires de France, « Sociologie ».

Merle P (2012), La ségrégation scolaire. La Découverte, Repères , 128 pages.

MESRI (2006-2016) Système d'information de Suivi de l'Étudiant (SISE). Producteur : MESRI. Diffuseur : ADISP-PROGEDO. <https://doi.org/10.34724/CASD.484.3358.V1>

MESRI (2017), *Note flash du SIES N° 18*, « Parcours et réussite en Licence et en PACES : les résultats de la session 2016 », Novembre 2017.

MESRI (2017), *Repères et Références Statistiques*, Edition 2017.

MESRI (2017), *L'état de l'Enseignement supérieur et de la Recherche en France*, édition 2017, MESRI

MESRI (2018), *L'état de l'Enseignement supérieur et de la Recherche en France*, édition 2018, MESRI

Michaut C (2000), L'influence du contexte universitaire sur la réussite des étudiants. Dijon : Thèse de doctorat, Université de Bourgogne.

Michaut C (2012), Réussite, échec et abandon des études dans l'enseignement supérieur français : quarante ans de recherche – in : M. Romainville et C. Michaut (éds.) Réussite, échec et abandon dans l'enseignement supérieur (53-68). Bruxelles : De Boeck.

Monso O, Givord P, Guillerm M, Murat F, Afsa C (2015), Quels outils pour mesurer la ségrégation dans le système éducatif ? Une application à la composition des Collèges français, *Document de travail DEPP-INSEE*.

Monso O, Boutchenik B, Givord P (2018) How do restrictive zoning and parental school choices impact social diversity in schools? An empirical evaluation in France, *13èmes Journées de méthodologie statistique de l'INSEE*, 12-14 Juin.

Monso O, Fougère D, Givord P, Pirus C (2019), Les camarades influencent-ils la réussite et le parcours des élèves ?, *Sciences Po LIEPP Working Paper* n°86, 2019-04-04.

- Mora R, Ruiz-Castillo J (2003) Additively decomposable segregation indices: the case of gender segregation by occupations and human capital levels in Spain. *The Journal of Economic Inequality* 1: 147–179.
- Mueller CW, Parcel TL (1981), Measures of Socioeconomic Status: Alternatives and Recommendations. *Child. Develop.* 52 (1), 13–20.
- Oberti M, Savina Y (2019) Urban and school segregation in Paris: The complexity of contextual effects on school achievement: The case of middle schools in Paris metropolitan area. *Urban Studies* 56 (15): 3117-3142.
- O’Connell P, Mc Coy S, Clancy D (2006) Who went to College? Socio-Economic Inequality in Entry to Higher education in the Republic of Ireland in 2004. *Higher Education Quarterly* 60 (4): 312-332.
- OECD (2019) Balancing School Choice and Equity: An International Perspective Based on PISA, *OECD Publishing*, Paris.
- Perret C (2007), Quitter sa région pour entrer à l’Université: quels sont les facteurs explicatifs de la mobilité géographique des bacheliers?. *L’orientation scolaire et professionnelle*, (36/3), 313-342.
- Perret C, Berthaud J, Pichon L (2014), Proposer des révisions intersessions aux étudiants de 1ère année : quelles relations avec la réussite aux examens ? – Les Sciences de l’éducation - *Pour l’Ère Nouvelle* 47, 1 (37-65).
- Pierrel A (2015) « Réussite scolaire, barrière économique. Des boursiers et leur famille face aux frais de scolarité des Grandes Ecoles de commerce », *Sociologie*, 2015/3 (Vol. 6), p. 225-240.
- Place D, Vincent B (2009), L’influence des caractéristiques sociodémographiques sur les diplômes et les compétences. In: *Economie et statistique*, n°424-425, 2009. Mesurer les compétences des adultes avec l’enquête Information et vie quotidienne (IVQ) pp. 125-147.
- Piketty T, Valdenaire M (2006), L’impact de la taille des classes sur la réussite scolaire dans les écoles, Collèges et Lycées français. Estimations à partir du panel primaire 1997, *Les dossiers évaluations et statistiques*, MEN-DEP, n°173.
- Proust Marcel, *A la Recherche du Temps Perdu*.
- Reardon S (1998), Measures of Racial Diversity and Segregation in Multigroup and Hierarchically Structured Populations. Presented at the annual meeting of the Eastern Sociological Society, Philadelphia, PA.
- Reardon S, Yun J, Eitle T (2000), The changing structure of school segregation: Measurement and evidence of multiracial metropolitan-area school segregation, 1989–1995, *Demography*, Springer; *Population Association of America* (PAA), vol. 37(3), pages 351-364, August.
- Reardon S, Firebaugh G (2002) Measures of Multigroup Segregation. *Sociological Methodology* 32 (1): 33-67.
- Reardon S, O’Sullivan D (2004) Measures of Spatial Segregation. *Sociological Methodology* 34 (1): 121-162.
- Robbins SB, Lauver K, Le H, Davis D, Langley R (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261-288.

- Rocher T (2016), Construction d'un indice de position sociale des élèves. *Éducation & formations*, DEPP, 2016: 5-27.
- Romainville M (2000), L'échec dans l'Université de masse. Paris : L'Harmattan.
- Rouban L (2015), L'ENA ou 70 ans de paradoxe. CEVIPOF, Paris.
- Sacerdote B (2011), Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?, *Handbook of the Economics of Education*, vol. 3, p. 249-277.
- Selz M, Vallet LA (2006), La démocratisation de l'enseignement et son paradoxe apparent. INSEE, *Données sociales*, La société française, 101–107.
- Smith DP, Holt L (2007), Studentification and 'apprentice' gentrifiers within Britain's provincial towns and cities: extending the meaning of gentrification, *Environment and Planning A* 39(1): 142-161.
- Theil H (1967), *Economics and Information Theory*. Amsterdam: North-Holland.
- Vaughan-Whitehead D (ed.), 2016, *Europe's Disappearing Middle Class?*, Edward Elgar Publishing, number 17301.
- Zilloniz S (2017), L'activité rémunérée des étudiants et ses liens avec la réussite des études, *Travail et Emploi*, 152 | octobre-décembre 2017.
- Zimmerman DJ, Winston G (2004), Peer effects in higher education. In: Hoxby, C. (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*. *University of Chicago Press*.