



PRISE EN COMPTE DU CARACTÈRE ORDONNÉ DES CATÉGORIES SOCIALES DANS LA MESURE DE LA SÉGRÉGATION

Fabrice MURAT ()*

() DEPP, Sous-direction des évaluations et de la performance*

fabrice.murat@education.gouv.fr

Mots-clés (6 maximum) : Inégalités, ségrégation, économétrie, éducation

Domaine concerné : Indicateurs, échelles, indices - Statistique descriptive, analyse factorielle

Résumé

D'un point de vue statistique, la ségrégation peut être définie comme la propension qu'ont les situations locales à s'écarter de la situation moyenne. L'absence de ségrégation se traduit par une distribution aléatoire de la population dans l'espace étudié, aboutissant à des répartitions spatiales semblables pour les différents groupes sociaux. Ces dernières années, une attention particulière a été accordée à cette question, dans le domaine de l'éducation, notamment en ce qui concerne les écarts sociaux entre collèges.

Le choix entre les différents indicateurs de ségrégation disponibles dépend en partie de la nature de la variable décrivant le milieu social. Longtemps, les travaux sur la ségrégation ont porté sur des distinctions dichotomiques (favorisés/défavorisés ; noirs/blancs) et l'indice de Duncan et, dans une moindre mesure, l'indice d'exposition normalisé ont été privilégiés. Ces indices se relient assez facilement à des indicateurs de dispersion (l'écart absolu moyen et la variance).

Cependant, résumer le milieu social à une distinction dichotomique est réducteur et des variables plus fines sont maintenant utilisées (distinction en 4 groupes selon la profession des parents dans les données de l'éducation nationale ; tranches de revenu dans les données de l'Insee). Pour ce type de variable, c'est l'indice d'entropie qui semble le plus pertinent.

Cet indicateur ne tient toutefois pas compte du caractère ordonné des variables, s'il existe. Cela conduit à mettre sur le même plan des ségrégations d'une nature assez différente. Pour autant, on peut souhaiter déterminer dans les unités qui scolarisent le plus d'individus « très défavorisés », si les autres individus sont majoritairement « moyennement défavorisés » ou au contraire « plutôt favorisés », ce qui n'appelle pas les mêmes actions politiques et pédagogiques.

Pour tenir compte de cela, deux possibilités existent actuellement. Un indice d'entropie tenant compte du caractère ordonné des variables a été par exemple utilisé par Jean-Michel Floch pour étudier la ségrégation des revenus entre territoires. Une autre approche consiste à se placer dans un cadre complètement quantitatif (par exemple quand on étudie le revenu). L'indicateur le plus naturel est alors le R^2 , le rapport entre la variance entre unités et la variance totale au niveau individuel. Les catégorisations sociales utilisant la profession n'étant pas fondées sur une variable quantitative, cette solution n'est pas immédiate. Attribuer des valeurs arbitraires (1 aux plus défavorisés, 2 à la catégorie suivante...) est une solution simple, mais dont la robustesse est incertaine. Il est sans doute préférable d'utiliser une quantification fine des situations sociales fondée sur un modèle théorique explicite, comme l'Indice de Position Sociale développé à la DEPP.

Nous allons dans cet article proposer une autre approche, que l'on peut relier à la vision économétrique sous-jacente à l'utilisation du R^2 , en montrant le lien entre la mesure de la ségrégation et les indicateurs de corrélation. Une ségrégation importante implique que connaître l'unité d'appartenance d'un individu nous permet de prédire, avec une marge d'erreur assez raisonnable, sa catégorie sociale.

Dans le cas dichotomique, l'étude de la ségrégation prend alors la forme d'une régression logistique du fait d'être défavorisé par les indicatrices d'appartenance aux unités. Dans ce cadre, l'indice d'entropie est le pseudo- R^2 de MacFadden de ce modèle. Ce résultat peut être prolongé à une catégorisation sociale en G groupes : l'indice d'entropie multigroupe est le pseudo- R^2 de MacFadden de la régression polytomique **non ordonnée** de la catégorie sociale par les indicatrices d'appartenance aux différentes unités.

Le lien entre régressions logistiques et indices de ségrégation étant établi, nous cherchons alors à tenir compte du caractère ordonné de la catégorie sociale en utilisant une régression polytomique **ordonnée**. Le pseudo- R^2 de MacFadden de ce modèle paraît alors un indicateur pertinent de la ségrégation tenant compte du caractère ordonné de la catégorie sociale. La régression ordonnée étant un modèle plus simple que la régression non ordonnée (chaque unité est caractérisée par un seul coefficient et non $G-1$), l'indice ainsi obtenu est toujours plus petit que l'indice d'entropie habituel : l'écart entre les deux valeurs permet de tester si la ségrégation est ordonnée (les deux valeurs sont proches) ou non (les valeurs sont assez éloignées : on ne peut déduire dans une unité simplement la répartition dans les $G-1$ premiers groupes en connaissant la proportion du type G ; la ségrégation est « multidimensionnelle »).

Une partie de ce travail est enfin consacrée à la mise en œuvre de ces méthodes d'abord sur des données simulées, afin de montrer dans quels cas les indicateurs divergent. Les nouveaux indicateurs sont confrontés à l'indice d'entropie classique, non ordonné, mais aussi à celui utilisé par Floch, ainsi qu'à une mesure de la ségrégation avec une variable quantitative. L'intérêt des modèles multiniveaux à effets aléatoires est aussi présenté. Le même travail est ensuite mené sur des données réelles (milieu social des élèves de 6^e). L'indice d'entropie ordonné apparaît alors assez proche de l'indice d'entropie habituel et les classements des départements selon ces deux indicateurs sont très cohérents. Il y a cependant un décalage, d'environ 20 % entre les valeurs, un peu plus grand dans certains départements, ce que l'on peut en partie expliquer par un comportement particulier des collèges privés.

Abstract

This article focuses on the measurement of segregation, in particular in the educational field. First, we analyze the link between segregation indices and other statistical fields: the measurement of dispersion and especially indicators of correlations. This second perspective will be extended in particular by showing the contribution of econometric models. Secondly, we will use this link between segregation indices and econometric methods to propose indices taking into account the metric nature of the variable of interest. Indeed, depending on whether this variable is dichotomous, categorical, ordinal, quantitative, the possibilities are not the same and it seemed relevant to us to expand the range of statistical tools for the ordinal case. Finally, in the last part, we will compare the different methods presented on real data (the social background of pupils entering 6th grade) and will show what an analysis that takes into account the ordered nature of the social category brings.

L'analyse de la mixité sociale ou de son contraire, la ségrégation sociale, entre établissements scolaires a donné lieu ces dernières années à de nombreux travaux tant par des chercheurs que par des membres de la statistique publique (Givord et alii, 2016). Cet intérêt se justifie par deux types d'attente concernant la mixité sociale. D'une part, la mixité sociale est souhaitable en termes de cohésion sociale. Il est important que tous les individus, en particulier dès le plus jeune âge, en tant qu'élèves, se côtoient, quelles que soient leurs origines et conditions de vie. D'autre part, la répartition des élèves dans les établissements a sans doute un effet sur les résultats des élèves, tant

en termes de résultats moyens que d'inégalités. Une vaste littérature théorique et empirique a porté sur les « effets de pairs » (Monso et alii, 2019). Cette perspective est alors un peu « asymétrique » : c'est la forte concentration d'élèves défavorisés dans certains établissements qui est un sujet de préoccupation, plus que la recherche de l'entre-soi des familles très favorisées.

Pour mener à bien ces études et répondre à des questions diverses (évolution de la ségrégation en tenant compte des évolutions structurelles de la population, prise en compte de la segmentation entre secteur public et secteur privé, lien avec la ségrégation résidentielle), de nombreuses investigations méthodologiques ont été menées, en s'appuyant sur la vaste littérature théorique et empirique concernant les nombreux indices de ségrégation (Givord et alii, 2016). Le présent article s'inscrit dans le prolongement de ces travaux, en explorant deux voies convergentes :

- dans un premier temps, nous analyserons le lien entre les indices de ségrégations et d'autres domaines statistiques : la mesure de dispersion et surtout l'étude des corrélations. Cette deuxième perspective sera notamment prolongée en montrant l'apport des techniques économétriques.
- dans un deuxième temps, nous utiliserons ce lien entre indices de ségrégation et méthodes économétriques pour proposer des indices tenant compte de la nature métrique de la variable d'intérêt. En effet, selon que cette variable est dichotomique, catégorielle, ordinale, quantitative, les possibilités ne sont pas les mêmes et il nous a semblé pertinent d'étoffer l'offre d'outils statistiques pour le cas ordinal.
- Enfin, dans une dernière partie, nous comparerons les différentes méthodes présentées sur des données réelles (le milieu social des élèves entrants en 6^e) et montrerons ce qu'apporte une analyse tenant compte du caractère ordonné de la catégorie sociale.

1. Comment mesurer la ségrégation sociale ?

La définition statistique de la mixité et de la ségrégation sociale fait largement consensus (Givord et alii, 2016 ; Floch, 2017) : la « mixité sociale » désigne le degré avec lequel les unités regroupent en leur sein des individus de milieux sociaux différents ; l'opposé de cette notion, la « ségrégation sociale » désigne au contraire la tendance pour les unités à accueillir des individus d'un même groupe social.

1.1. Lien avec les indicateurs de dispersion

Avec une telle définition, il paraît assez naturel de relier les indicateurs de ségrégation et les indicateurs de dispersion. La ségrégation peut ainsi être considérée comme la mesure de la dispersion de la tonalité sociale entre les unités. Le fait que ces unités soient elles-mêmes des agrégats d'individus constitue l'une des particularités des indicateurs de ségrégation. L'autre spécificité, sur laquelle nous allons revenir, est que la variable caractérisant le milieu social au niveau individuel est généralement qualitative.

Il est très facile de relier l'indice de dissimilarité (ID) et l'indice d'exposition normalisé (IEN), deux indicateurs de ségrégation très utilisés, à des indicateurs de dispersion. En se limitant à une mesure du milieu social ne distinguant que deux situations (milieu favorisé et milieu défavorisé), en posant m le nombre d'unités, p la proportion d'individus défavorisés, $\pi_k = \frac{n_k}{n}$ le poids de l'unité k dans la population (rapport de son effectif sur l'effectif total) et p_k la proportion d'individus défavorisés dans l'unité k , alors :

$$ID = \frac{1}{2} \frac{\sum_{k=1}^m \pi_k |p_k - p|}{p(1-p)}$$

$$IEN = \frac{\sum_{k=1}^m \pi_k (p_k - p)^2}{p(1-p)}$$

L'indice de dissimilarité apparaît lié à l'écart absolu moyen des proportions d'individus défavorisés par unité, tandis que l'indice d'exposition normalisé est lié à la variance de ces valeurs. Notons que dans ce cas, la normalisation par $p(1-p)$ paraît assez naturelle puisqu'il s'agit de la variance sur l'ensemble des individus de l'indicatrice d'appartenance à la catégorie défavorisée. L'indice d'exposition normalisé est donc le rapport entre la variance inter-unité sur la variance totale. C'est pourquoi il est parfois désigné comme *variance ratio index* ou *correlation ratio*, ce qui indique un premier lien avec les mesures de corrélation que nous reprendrons plus loin.

Dans les deux cas, l'indice se décompose en une somme des écarts de chaque unité à la moyenne (en valeur absolue ou élevés au carré).

1.2. Cas d'une catégorie sociale à plus de 2 groupes

L'indice de dissimilarité et l'indice d'exposition normalisé sont bien adaptés à une mesure du milieu social dichotomique et se prêtent à des interprétations assez simples (l'indice de dissimilarité peut être ainsi être lié à la proportion de jeunes qu'il faudrait changer d'établissements pour aboutir à une égalité parfaite des répartitions). Cependant, ils se généralisent assez mal au cas d'une catégorisation sociale en plus de 2 groupes.

Dans ce cas, l'indice d'entropie va apparaître plus pertinent. Dans le cas dichotomique, sa formule est :

$$H = 1 - \frac{\sum_{k=1}^m \pi_k h(p_k)}{h(p)} = \sum_{k=1}^m \pi_k \frac{h(p) - h(p_k)}{h(p)}$$

avec $h(x) = -x \ln(x) - (1-x) \ln(1-x)$

La fonction $h(x)$ a une forme en cloche, vaut par convention 0 pour les valeurs 0 et 1 (en théorie, elle n'est pas définie, mais elle tend vers 0 pour ces valeurs) et a un maximum pour 0,5. Elle peut s'interpréter comme une mesure de la diversité au sein de l'établissement, maximale quand les deux groupes sociaux sont également représentés, minimale quand un seul groupe est présent (Reardon et Firebaugh, 2012).

Notons qu'en utilisant le principe de décomposition de la variance, l'indice d'exposition normalisé peut se présenter sous une forme assez proche :

$$IEN = 1 - \frac{\sum_{k=1}^m \pi_k p_k (1-p_k)}{p(1-p)}, \text{ la fonction } v(x) = x(1-x) \text{ ayant la même forme en cloche que } h(x).$$

La généralisation de l'indice d'entropie à une catégorisation sociale en G groupes, se fait en reprenant la formule ci-dessus et en définissant l'entropie pour une population se répartissant de la façon suivante dans les G groupes : $(q_1 \dots q_G)$:

$$h = - \sum_{j=1}^G q_j \ln(q_j)$$

Cette fonction est maximale quand la répartition est égalitaire dans les G groupes et minimale quand un seul groupe est représenté.

La deuxième partie de la formule de l'indice d'entropie suggère qu'il peut se décomposer comme la somme des écarts entre la situation de chaque unité et la situation moyenne, comme l'indice de dissimilarité ou l'indice d'exposition normalisé, mais c'est un peu trompeur. En effet, si une unité

ayant une entropie différente de l'entropie moyenne contribue bien à une forte ségrégation, l'inverse, en règle générale, n'est pas vrai : une unité peut avoir exactement la même entropie que la moyenne avec un profil social très différent (dans le cas dichotomique, si la proportion d'individus défavorisés sur l'ensemble de la population est 20 %, les unités en accueillant 80 % ont la même entropie). L'équivalence entre écart d'entropie à la moyenne et écart de profil social n'est possible que dans le cas particulier où les catégories regroupent le même nombre d'individus (ce qui peut arriver si ces catégories sont définies avec les quantiles d'une variable quantitative). Dans ce cas, l'entropie moyenne est au niveau maximal et ne peuvent s'en approcher que des unités qui ont profil social quasi équilibré. Notons que même dans ce cas, avec une même entropie très éloignée de la moyenne, on peut trouver des unités aux profils très différents (surreprésentation des défavorisés, des favorisés ou des individus dans la moyenne), mais cette remarque vaut aussi pour l'indice de dissimilarité et l'indice d'exposition normalisé dans le cas dichotomique (les unités avec une forte proportion d'individus défavorisés contribuent de la même façon que les unités avec une faible proportion).

1.3. Lien avec les mesures de corrélation

Si le lien entre indices de ségrégation et mesures de dispersion paraît assez naturel (on aurait pu aussi présenter l'usage de l'indice de Gini dans les deux domaines), il est aussi possible de relier les indices de ségrégation avec les mesures de corrélation, celle-ci pouvant être assimilées à la force de l'association entre l'appartenance à une unité et l'appartenance à un groupe social (Reardon et Firebaugh, 2002).

L'indice d'exposition normalisé va là encore fournir une illustration simple de ce principe. En effet, pour l'indicatrice d'appartenance au groupe défavorisé, comme il est le rapport entre la variance entre unités et la variance totale, il peut être considéré comme le R^2 de la régression de cette indicatrice par les indicatrices d'appartenance aux unités.

Reardon et Firebaugh (2002) montrent aussi que l'indice d'exposition normalisé est le V de Cramer (statistique dérivée du χ^2) du tableau croisant l'identifiant de l'unité et l'indicatrice d'appartenance au groupe défavorisé. Dans un cadre plus général, avec une catégorie sociale en plus de 2 groupes, Reardon et Firebaugh (2002) montrent que l'indice d'entropie peut être relié à l'indice G^2 , rapport de vraisemblance, proche du χ^2 et plus robuste sur de petits échantillons.

Le lien entre ségrégation et corrélation va apparaître plus clairement en considérant le cas où le critère social est de forme quantitative. Les études utilisant ce type de variable ne sont pas très fréquentes. Cependant, Reardon (2011) présente un certain nombre de travaux sur la ségrégation en termes de revenus, où la question a été traitée. Il précise que les avantages des différentes méthodes (notamment un retour aux indicateurs classiques de ségrégation sur variable qualitative par une mise en tranches des revenus) doivent être encore étudiés, mais il semble que les chercheurs aient souvent recours à une décomposition de la variance, qui rapproche cette méthode de l'indice d'exposition normalisé.

C'est le R^2 (ou sa racine carrée), rapport de la variance « inter » du revenu entre les unités par la variance du revenu entre individus qui quantifie la ségrégation. Comme nous l'avons signalé pour l'indice d'exposition normalisé, cela revient à effectuer la régression du revenu par les indicatrices d'appartenance aux unités.

Cette méthode peut être rapprochée des nombreux travaux en sciences de l'éducation sur l'« effet établissement » et l'« effet enseignant ». Pour mesurer l'influence de ces facteurs sur les résultats des élèves, on calcule la part de variance des scores expliquée par les indicatrices d'appartenance à tel établissement ou du fait d'avoir tel enseignant. Il ne s'agit pas véritablement d'une perspective de ségrégation, car les unités considérées sont censées jouer un rôle actif sur la caractéristique étudiée. Cependant, si au lieu de prendre les résultats à la fin de l'année scolaire ou de la scolarité, on utilise les caractéristiques initiales, ce type d'indicateur se rapproche des indicateurs de ségrégation.

Ce rapprochement suggère des développements dans l'étude de la ségrégation en s'inspirant des méthodologies courante pour étudier les « effets établissement » et les « effets enseignant » : les modèles multiniveaux semblent une piste intéressante, en particulier ceux utilisant les effets aléatoires, qui permettent de tenir compte de la petite taille de certaines unités, y compris pour des mesures du milieu social de nature qualitative (Paterson, 1992 ; Goldstein, Noden, 2003). L'un des intérêts de cette approche est, comme dans un modèle économétrique classique, d'introduire des variables de contrôle pour mieux mesurer la ségrégation¹.

Le travail qui va suivre poursuit un but assez proche, mais en restant dans le cadre d'une modélisation à effets fixes. L'objectif est de montrer la proximité entre les indicateurs de ségrégation et les indicateurs mesurant la qualité d'une régression logistique reliant l'indicatrice d'appartenance au milieu social défavorisé et les indicatrices d'appartenance aux différentes unités. Cette approche prolonge à la fois la similitude entre la ségrégation et le tableau croisé de ces deux variables et les techniques de régression linéaire naturellement utilisées pour des variables sociales quantitatives.

1.4. Ségrégation et régression logistique

On trouvera dans (Afsa, 2013) une description détaillée des principes et de la méthodologie associés à la régression logistique. Nous allons reprendre les principaux éléments dans le cas de la ségrégation, en utilisant le modèle suivant (très proche de celui proposé par (Goldstein et Noden, 2003) à ceci près qu'ils se placent dans un cadre multiniveau à effets aléatoires) :

$$\ln \frac{y_i}{1-y_i} = x_i' \beta = a + \sum_{k=1}^m \beta_k 1_{i \in k}$$

y_i est la probabilité que l'individu i soit défavorisé ; x_i est le vecteur de variables explicatives, ici les indicatrices d'appartenance aux unités, $1_{i \in k}$ pour $1 \leq k \leq m$; β est le vecteur de paramètres, qui prend la forme $(\beta_1 \dots \beta_m)$ auquel s'ajoute la constante a .

On a alors :

$$y_i = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$$

La vraisemblance du modèle s'écrit de la façon suivante :

$$\mathcal{L} = \prod_{i=1}^n P(y_i=1)^{y_i} P(y_i=0)^{1-y_i} = \prod_{i=1}^n \left(\frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right)^{y_i} \left(\frac{1}{1 + e^{x_i' \beta}} \right)^{1-y_i} = \prod_{i=1}^n \frac{(e^{x_i' \beta})^{y_i}}{1 + e^{x_i' \beta}}$$

La log-vraisemblance vaut :

$$L = \ln \mathcal{L} = \sum_{i=1}^n (y_i (x_i' \beta) - \ln (1 + e^{x_i' \beta}))$$

Et dans le modèle de cette note :

$$L = \sum_{k=1}^m \sum_{i \in k} (y_i (a + \beta_k) - \ln (1 + e^{a + \beta_k}))$$

Pour estimer les paramètres β_k , on va dériver cette expression pour chacun d'entre eux :

$$\frac{dL}{d\beta_k} = \sum_{i \in k} \left(y_i - \frac{e^{a + \beta_k}}{1 + e^{a + \beta_k}} \right) = n_k \left(p_k - \frac{e^{a + \beta_k}}{1 + e^{a + \beta_k}} \right)$$

Et en annulant cette dérivée, on obtient :

¹ On peut aussi citer (Kalter, 2000) et (Spriggs et Williams, 1996), qui ont la particularité de chercher à appliquer une méthodologie économétrique à l'indice de dissimilarité.

$$e^{a+\beta_k} = \frac{p_k}{1-p_k}$$

$$\beta_k = \ln \frac{p_k}{1-p_k} - a$$

Comme dans tout modèle logistique, il y a une indétermination (dans notre modèle, dériver en fonction de a ne conduit pas à une équation pertinente). Il faut donc ajouter une contrainte identifiante. Cela peut être une somme des paramètres β_k à 0 ou, comme souvent, en donnant cette valeur au coefficient de la dernière unité de la liste : $\beta_m = 0$ et donc $a = \ln \frac{p_m}{1-p_m}$. Chaque unité est

donc caractérisée par le logit de la proportion d'individus défavorisés (à une constante près), qui peut s'interpréter comme la propension de chaque unité à accueillir des individus défavorisés.

La log-vraisemblance s'exprime alors ainsi :

$$L = \sum_{k=1}^m n_k \left(p_k \ln \frac{p_k}{1-p_k} - \ln \left(1 + \frac{p_k}{1-p_k} \right) \right) = \sum_{k=1}^m n_k \left(p_k \ln p_k - p_k \ln \frac{1}{1-p_k} - \ln \left(\frac{1}{1-p_k} \right) \right)$$

$$L = n \sum_{k=1}^m \frac{n_k}{n} \left(p_k \ln p_k + (1-p_k) \ln (1-p_k) \right) = -n \sum_{k=1}^m \pi_k h(p_k)$$

Cette formule établit une relation entre la log-vraisemblance du modèle et l'entropie, qui a déjà été présentée : Hauser (1977), notamment, a montré le lien entre les modélisations logistiques et la théorie de l'information, à la base de l'utilisation de l'entropie. L'originalité de notre travail va être de faire le lien avec les indicateurs de ségrégation².

1.5. Pseudo-R² et indicateurs de ségrégation

Nous allons maintenant suivre l'analogie entre le R² mesurant la ségrégation d'une variable quantitative en fonction des unités d'appartenance, dans une régression classique et les pseudo-R² associés à la régression logistique d'une variable qualitative par ces unités d'appartenance.

Il existe un grand nombre de pseudo-R² : (Veal, Zimmerman, 1996) décrit leur principe et résume les simulations de Monte Carlo qui ont cherché à juger de leur robustesse pour mesurer la qualité d'un modèle.

Plusieurs pseudo-R² sont fondés sur la log-vraisemblance que nous venons de calculer. Plus précisément, ils utilisent la différence de log-vraisemblance entre le modèle effectivement estimé et le modèle nul n'utilisant aucune variable : $L_0 = -nh(p)$ et en définissant la déviance par $\Delta = 2(L - L_0) = 2nh(p)H = 2nM$ (M étant appelé l'indice d'information mutuel). On envisage aussi le cas où le modèle est parfait et dans ce cas, la log-vraisemblance est nulle et $\Delta^0 = -2L_0$.

Le pseudo-R² de Mc Fadden, l'un des pseudo-R² les plus utilisés, est le rapport entre Δ et Δ^0 , soit ici :

$$PsR^2_{McF} = \frac{\Delta}{\Delta^0} = \frac{2(L - L_0)}{-2L_0} = 1 - \sum_{k=1}^m \frac{n_k}{n} \frac{h(p_k)}{h(p)} = H$$

L'indice d'entropie est donc le pseudo-R² de Mc Fadden de la régression logistique de l'indicatrice repérant le milieu défavorisé par les indicatrices d'appartenance aux différentes unités.

² La mise en relation de l'indice d'entropie et de la modalisation économétrique ouvre aussi la possibilité de procéder à des tests statistiques, pour déterminer si l'indice de ségrégation est significativement différent de 0 (si la ségrégation n'est pas simplement due à la répartition aléatoire, forcément pas tout à fait uniforme des individus dans des unités de taille finie et souvent même réduite) ou si deux valeurs sont significativement différentes.

On trouvera en annexe A une analyse pour un certain nombre d'autres pseudo-R², en reliant plusieurs d'entre eux avec des indices de ségrégation et donnant pour les autres des idées de nouveaux indices de ségrégation.

1.6. Indice multigroupe et régression polytomique non ordonnée

Nous allons maintenant montrer que l'étude de la ségrégation avec une catégorisation sociale en plus de 2 groupes peut être assimilée à une régression polytomique **non ordonnée**. Un des groupes sociaux va être pris en référence (G , le dernier dans la suite) et on va déterminer la probabilité d'appartenir à chaque autre groupe social au lieu du groupe de référence en fonction de l'unité d'appartenance.

$$\ln \frac{P(C_i=j)}{P(C_i=G)} = x_i' \beta_j = a_j + \sum_{k=1}^m \beta_j^k 1_{i \in k}$$

x_i est le vecteur de variables explicatives, ici les indicatrices d'appartenance aux unités ; β_j est le vecteur de paramètres, avec notamment une constante. Chaque unité va être caractérisée par $G-1$ paramètres indiquant sa tendance à accueillir telle ou telle catégorie sociale (la part du G^e groupe étant déterminée par le complément à 100 %).

La vraisemblance du modèle s'écrit de la façon suivante :

$$\mathcal{L} = \prod_{i=1}^n \prod_{j=1}^G P(C_i=j)^{I(C_i=j)}$$

La log-vraisemblance vaut :

$$L = \ln \mathcal{L} = \sum_{i=1}^n \sum_{j=1}^G (C_i=j) \ln (P(C_i=j))$$

$$L = \sum_{i=1}^n \left(\sum_{j=1}^{G-1} (C_i=j) (x_i' \beta_j + \ln (P(C_i=G))) + (C_i=G) \ln (P(C_i=G)) \right)$$

$$L = \sum_{i=1}^n \left(\sum_{j=1}^{G-1} (C_i=j) (x_i' \beta_j) + \ln (P(C_i=G)) \right)$$

$$\text{Or } \sum_{j=1}^{G-1} P(C_i=j) = \sum_{j=1}^{G-1} e^{x_i' \beta_j} P(C_i=G) \text{ et donc } 1 - P(C_i=G) = P(C_i=G) \sum_{j=1}^{G-1} e^{x_i' \beta_j} \text{ et :}$$

$$P(C_i=G) = \frac{1}{1 + \sum_{j=1}^{G-1} e^{x_i' \beta_j}}$$

La log-vraisemblance s'écrit alors ainsi :

$$L = \sum_{i=1}^n \left(\sum_{j=1}^{G-1} (C_i=j) (x_i' \beta_j) - \ln \left(1 + \sum_{t=1}^{G-1} e^{x_i' \beta_t} \right) \right)$$

Dans notre étude :

$$L = \sum_{k=1}^m \sum_{i \in k} \left(\sum_{j=1}^{G-1} (C_i=j) (a_j + \beta_j^k) - \ln \left(1 + \sum_{t=1}^{G-1} e^{a_j + \beta_t^k} \right) \right)$$

Pour estimer les paramètres β_j^k , on va dériver cette expression pour chacun d'entre eux :

$$\frac{dL}{d\beta_j^k} = \sum_{i \in k} ((C_i = j) - \frac{e^{a_j + \beta_j^k}}{1 + \sum_{t=1}^{G-1} e^{a_t + \beta_t^k}})$$

Et en annulant cette dérivée, on obtient, si p_j^k est la proportion d'individus de la catégorie j dans l'unité k :

$$p_j^k = \frac{e^{a_j + \beta_j^k}}{1 + \sum_{t=1}^{G-1} e^{a_t + \beta_t^k}}$$

Comme pour le cas dichotomique, il faut imposer des contraintes pour obtenir l'unicité des solutions. On pose pour chaque catégorie j , $\beta_j^m = 0$.

Comme la somme des proportions des catégories dans un collège fait 100 %, on peut écrire :

$$p_G^k = 1 - \sum_{j=1}^{G-1} p_j^k = 1 - \frac{\sum_{j=1}^{G-1} e^{a_j + \beta_j^k}}{1 + \sum_{t=1}^{G-1} e^{a_t + \beta_t^k}} = \frac{1}{1 + \sum_{t=1}^{G-1} e^{a_t + \beta_t^k}} \text{ et donc } 1 + \sum_{t=1}^{G-1} e^{a_t + \beta_t^k} = \frac{1}{p_G^k}$$

$$\text{Soit } \frac{p_j^k}{p_G^k} = e^{a_j - \beta_j^k} \text{ et } \beta_j^k = \ln \frac{p_j^k}{p_G^k} - a_j$$

La série de contraintes sur les coefficients de la dernière unité donne : $\beta_j^m = \ln \frac{p_j^m}{p_G^m} - a_j = 0$ soit

$$a_j = \ln \frac{p_j^m}{p_G^m}$$

Finalement, $\beta_j^k = \ln \frac{p_j^k}{p_G^k} - \ln \frac{p_j^m}{p_G^m}$: chaque unité est caractérisée (à une constante près) par un ensemble de $G-1$ contrastes logistiques entre les $G-1$ premières catégories sociales et la dernière prise en référence. Une valeur positive indique que la catégorie est surreprésentée ; une valeur négative indique l'inverse.

La log-vraisemblance s'exprime alors ainsi :

$$L = \sum_{k=1}^m \left(\sum_{j=1}^{G-1} (C_i = j) \left(\ln \frac{p_j^k}{p_G^k} \right) - \ln \left(\frac{1}{p_G^k} \right) \right) = \sum_{k=1}^m n_k \left(\sum_{j=1}^{G-1} p_j^k \ln p_j^k - \ln p_G^k \sum_{j=1}^{G-1} p_j^k + \ln p_G^k \right)$$

$$L = \sum_{k=1}^m n_k \left(\sum_{j=1}^G p_j^k \ln p_j^k \right) = -n \sum_{k=1}^m \frac{n_k}{n} \left(\sum_{j=1}^G p_j^k \ln p_j^k \right) = -n \sum_{k=1}^m \frac{n_k}{n} h_k$$

en retenant pour h_k la définition habituelle de l'entropie par établissement.

Le raisonnement conduisant à faire le lien entre l'indice d'entropie et le pseudo- R^2 de Mac Fadden peut être alors reproduit.

Après être passé d'une régression linéaire d'une indicatrice d'appartenance à un groupe défavorisé par des indicatrices d'appartenance à des unités (méthode qui sous-tend l'indice d'exposition normalisé), à une régression logistique (reliée à l'indice d'entropie dichotomique), puis à une régression polytomique non ordonnée (reliée à l'indice d'entropie multigroupe), nous allons envisager la pertinence et l'intérêt de passer à une régression polytomique **ordonnée**, quand la catégorisation sociale comporte une hiérarchisation.

2. Comment prendre en compte la nature de la catégorie sociale ?

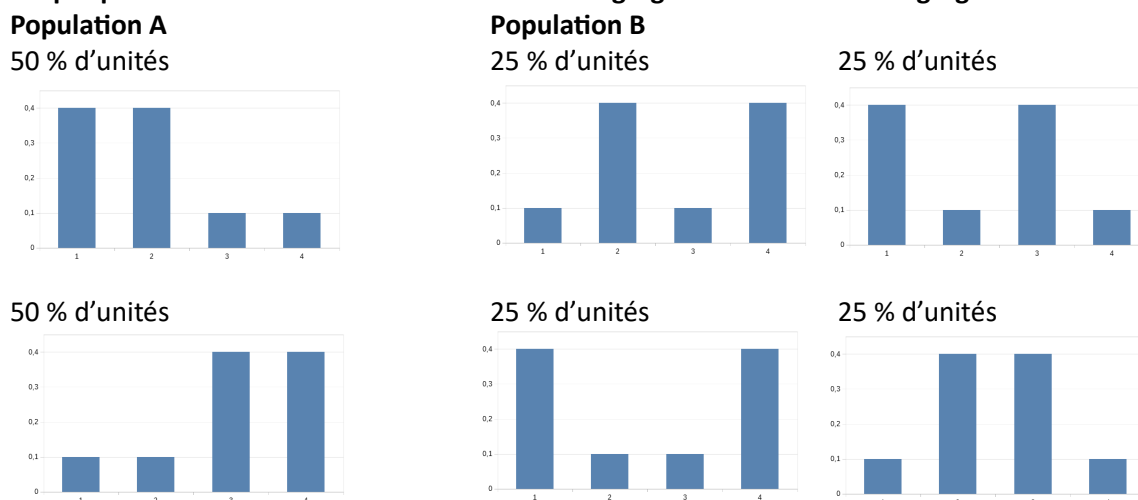
Parmi les nombreux indicateurs de ségrégation disponibles, le choix doit bien sûr être guidé par des principes théoriques et la vision que l'on a de la ségrégation : Massey et Denton (1988) ont ainsi distingué cinq dimensions théoriques dans les indicateurs de ségrégation, en particulier dans la façon dont ils prennent en compte l'aspect géographique. Les propriétés mathématiques d'un indicateur sont aussi des éléments intéressants, par exemple la possibilité de le décomposer par sous-populations (Frankel et Volij, 2011). Comme on l'a déjà dit, la nature de la catégorisation sociale doit également être prise en compte : certains indicateurs sont pertinents pour une distinction dichotomique, mais moins pour une catégorisation en plus de 2 groupes.

Ces éléments ne sont pas indépendants : la vision que l'on a de la ségrégation et la définition de la catégorisation sociale peuvent être fortement liées. Ainsi, si l'on accorde un intérêt particulier à la ségrégation des individus les plus défavorisés, il paraît logique de choisir une distinction dichotomique les repérant, permettant de repérer des unités concentrant une part importante d'entre eux, sans se préoccuper de l'existence d'unités où, par un mécanisme de recherche de l'entre-soi différent de ce que l'on pourrait appeler une « ségrégation subie », se regroupent beaucoup d'individus très favorisés, excluant même les classes moyennes.

L'utilisation d'une catégorisation en plus de 2 groupes se fonde donc plutôt sur une définition assez « ouverte » de la ségrégation : il s'agit de repérer toutes les situations où les unités se distinguent les unes des autres dans la répartition par catégories sociales, sans accorder d'importances différentes à ces catégories et sans tenir compte de leurs proximités. Cette approche est particulièrement pertinente, quand il n'existe pas de hiérarchisation entre les catégories, mais si celles-ci peuvent être ordonnées, on peut aussi souhaiter en tenir compte dans la mesure de la ségrégation.

L'exemple fictif ci-dessous l'illustre. Dans les deux populations, les individus se répartissent de la même façon dans la catégorie sociale en 4 groupes (25 % dans chaque). L'entropie dans chaque établissement est identique (1,19 environ), car il s'agit de permutations des mêmes proportions (10 %, 10 %, 40 %, 40 %), ce qui conduit, on l'a indiqué plus haut aux mêmes valeurs d'entropie. Par conséquent, l'indice d'entropie, découlant de la moyenne de ces valeurs constantes, est identique dans les deux populations.

Graphique 1 : illustration de la différence entre ségrégation ordonnée et ségrégation non ordonnée



Cette conclusion ne pose pas trop de problème si la catégorie sociale n'est pas ordonnée (même si qualitativement, les deux ségrégations ont une forme un peu différente). Même si la catégorie sociale est ordonnée, on peut ne pas vouloir en tenir compte, en considérant que la ségrégation est un problème global et que dans une unité qui regroupe beaucoup d'individus très défavorisés, peu importe que les autres soient plus souvent un peu défavorisés ou très favorisés. A l'inverse, cela

entre plutôt en contradiction avec une vision « unidimensionnelle » de la ségrégation. Dans ce cas, la population A est sans doute plus ségréguée, car elle oppose des unités très défavorisées à des unités très favorisées, diagnostic moins évident dans la population B.

La vision « unidimensionnelle » peut être justifiée si les préoccupations concernant la ségrégation sont principalement dues à la crainte d'un effet négatif de la concentration des individus défavorisés (par exemple sur les résultats scolaires, par des effets de pairs défavorables et la difficulté à attirer des personnels très qualifiés). La population A paraît alors beaucoup plus concernée que la population B. Dans ce cas, il faut alors être sûr que dans les unités accueillant 40 % d'individus très défavorisés et 40 % d'individus favorisés, la situation est moins bonne, en particulier pour les individus très défavorisés, par rapport aux unités accueillant 40 % d'individus très défavorisés et 40 % d'individus très favorisés, que dans ces dernières unités, la « polarisation » n'a pas d'autres effets pervers³. Le choix de la mesure de la ségrégation dépend toujours de considérations théoriques d'ordre sociologique et économique.

2.1. Indice d'entropie ordonnée

Pour tenir compte du caractère ordonné de la catégorie sociale dans la mesure de la ségrégation, Floch (2017) propose une méthodologie récente, qu'il met en œuvre sur la ségrégation des revenus entre quartiers au sein de différentes aires urbaines.

Cette approche paraît particulièrement pertinente dans le cas où, comme ici avec les revenus, on peut ajuster la catégorisation, tant en nombre (5 groupes ou 10 groupes) que de répartition (même si une équi-répartition au niveau global par « quintiles » ou « déciles » est généralement choisie).

Dans une première étape, on calcule les indices d'entropie dichotomiques pour des regroupements croissants de catégories : dans le cas du revenu, cela revient à construire les indicatrices sous des seuils croissants (et correspondant généralement à des quantiles) ; dans le cas plus général, on repère d'abord la catégorie la plus défavorisée, puis les deux catégories les plus défavorisées, etc.

L'indice d'entropie globale ordonnée est une moyenne de ces indices élémentaires⁴, en pondérant

l'indice d'entropie pour les individus au mieux dans la catégorie i par
$$\frac{h(p_i)}{\sum_{j=1}^G h(p_j)}$$

Cette méthode fournit une mesure globale de la ségrégation tenant compte du caractère hiérarchisé des données, mais son principal intérêt est sans doute dans la présentation des indices intermédiaires, qui donnent une information plus fine, un « profil » de la ségrégation. Dans l'étude de Floch, cela permet notamment de montrer que la ségrégation entre quartiers est plus forte pour les catégories de revenus les plus élevés, de façon plus ou moins marquée selon les aires urbaines (Le centre de Paris relève nettement de cette situation, alors que dans la banlieue de Nantes, la ségrégation des plus pauvres est identique à celle des très riches).

2.2. Indicateurs de ségrégation sur données quantitatives

Une autre possibilité consiste à aller au bout de la logique de classement et à utiliser une spécification quantitative. Cela revient par exemple à travailler directement sur le revenu, plutôt que sur une répartition en tranches. Cette perspective, déjà évoquée dans la partie 1.3, peut se présenter de la façon suivante :

³ Pour certains indicateurs ordonnés qui vont être présentés, une population composée d'unités avec comme répartitions (10 %, 40 %, 40 %, 10 %) ou (40 %, 10 %, 10 %, 40 %), sera considérée comme absolument non ségréguée : il n'y a que des unités « moyennes » (homogènes ou hétérogènes). Cela illustre encore que le choix d'un indicateur ordonné ne doit pas être automatique et suppose par exemple que la situation n'est pas jugée préoccupante dans les unités polarisés, par rapport à des unités du type (40 %, 40 %, 10 %, 10%),.

⁴ Notons que l'indice d'entropie multigroupe habituel s'exprime lui comme la moyenne des indices d'entropie d'appartenance à chaque catégorie (et non en-dessous d'une certaine catégorie, ce qui n'aurait d'ailleurs pas de sens pour une catégorisation sociale non ordonnée).

$$Q_i = x_i' \beta + \varepsilon_i = a + \sum_{k=1}^m \beta_k 1_{i \in k} + \varepsilon_i$$

où Q_i est la variable quantitative, β_k les paramètres associés à chaque unité, indiquant dans quelle mesure elle « prédit » la catégorie sociale, ε_i le terme aléatoire tenant compte du fait que l'appartenance à une unité ne permet pas de déterminer parfaitement la catégorie sociale.

Dans cette modélisation très simple, il s'agit d'une simple décomposition de la variance et l'indicateur de ségrégation le plus pertinent est le R^2 . Cela établit un lien avec l'indice d'exposition normalisé (R^2 de la régression linéaire d'une indicatrice) et l'indice d'entropie (pseudo- R^2 de Mac Fadden d'une régression logistique non ordonnée d'une variable catégorielle).

Ce modèle peut être complété de deux façons (voir (Goldstein et Noden, 2003) même s'ils utilisent la modélisation logistique d'une indicatrice) :

- d'une part, il est possible de rajouter des variables explicatives au modèle et par exemple de mesurer la ségrégation entre établissements en termes de performance scolaire, en tenant compte de la ségrégation en termes de milieu social (il suffit de rajouter cette variable dans le modèle).
- d'autre part, les termes β_k ici considérés comme des effets fixes peuvent être pris comme des paramètres aléatoires. L'utilisation de cette modélisation permet, comme indiqué en annexe B, de mieux tenir compte de la « ségrégation » aléatoire non négligeable provoquée par la taille finie, voire petite, des unités.

Notons qu'il ne s'agit bien sûr pas de modèles causaux. Le fait pour un élève d'appartenir à tel ou tel collègue a très peu d'influence sur la profession de ses parents (l'inverse est sans doute plus plausible). L'approche est entièrement descriptive. La vision causale correspond plutôt aux recherches sur l'« effet établissement », l'« effet classe » ou l'« effet enseignant », dans quelle mesure ces unités ont une influence sur la scolarité des élèves, notamment en termes d'évolution des compétences. Dans ce cas, la variable expliquée est généralement un niveau final, que l'on va modéliser par l'appartenance à telle ou telle unité et un certain nombre de variables de contrôle (le niveau initial, notamment). Cela signifie qu'il faut éviter d'étudier la ségrégation en fonction d'une donnée qui a pu être influencée par l'appartenance à telle ou telle unité : la dispersion des notes au DNB par collègue par exemple rend compte d'une possible ségrégation scolaire, mais aussi d'une efficacité différente d'un établissement à l'autre.

Dernière remarque générale, sur laquelle nous reviendrons plus tard : le modèle ci-dessus porte sur la situation moyenne. Il suppose que l'appartenance à une unité a un impact sur l'espérance de la variable sociale, mais pas sur d'autres aspects de sa distribution. Il pourrait être pertinent d'utiliser des régressions quantiles : cela rendrait compte du fait que deux unités peuvent se distinguer dans le niveau des $x\%$ les plus défavorisés, alors qu'elles ont le même niveau moyen. Or l'étude de Floch (2017) a montré que les phénomènes de ségrégation ne sont pas uniformes. Dans l'approche, en moyenne, comme dans l'utilisation de l'indice ordonné global présenté par Floch (2017), cette information est perdue, perte habituelle dans tout exercice de synthèse.

En termes d'application à une variable ordinale, trois situations sont possibles :

- Si comme dans l'étude de Floch (2017), la variable catégorielle est en fait une segmentation d'une variable quantitative, il est possible d'utiliser directement cette variable quantitative, avec les limitations qui viennent d'être évoquées.
- Il est possible de transformer la variable catégorielle en variable quantitative, en la « projetant » de façon plus moins simple, sur une donnée de cette nature : pour « quantifier » les professions, on peut les remplacer par le salaire moyen observé des individus de cette profession, par leur nombre moyen d'années d'études... Rocher (2016) propose une méthodologie, avec en préalable une phase de synthèse, par une analyse de données, pour tenir compte d'un nombre important de caractéristiques.

- Une telle « quantification » peut ne pas être possible, faute de données avec la catégorie sociale et une variable quantitative pertinente. Dans ce cas, il est toujours possible de quantifier de manière simple les modalités : 1 pour la plus défavorisée, 2 pour la suivante... C'est une pratique courante en psychologie dans le traitement des variables d'opinions (échelle de Likert), mais il y a bien sûr une part d'arbitraire qui incite à tester des variantes.

2.3. Utilisation d'une régression polytomique ordonnée

Nous allons maintenant tirer profit du lien établi dans la première partie entre indice d'entropie et régression logistique pour proposer une autre solution. L'idée est simplement de remplacer la régression polytomique non ordonnée par une régression polytomique ordonnée et de prendre le pseudo-R² de Mac Fadden de cette régression comme indicateur de ségrégation ordonnée.

Cette perspective peut être reliée à ce que nous venons de décrire pour une variable quantitative. On suppose que la catégorie sociale observée est la réalisation d'une segmentation d'une variable quantitative latente :

$$C_i = \begin{cases} 1 & \text{si } Q_i^0 < s_1 \\ j & \text{si } s_{j-1} \leq Q_i^0 < s_j \\ G & \text{si } Q_i^0 \geq s_{G-1} \end{cases}$$

où C_i est la catégorie sociale observée, avec G modalités, Q_i la variable latente sous-tendant cette catégorie, avec les seuils $(s_1 \dots s_{G-1})$.

On posera ensuite, comme dans le cas quantitatif,

$$Q_i^0 = x_i' \beta + \varepsilon_i = a + \sum_{k=1}^m \beta_k 1_{i \in k} + \varepsilon_i$$

Dans ce cas

$$P(C_i \leq j) = P(Q_i^0 \leq s_j) = P(x_i' \beta - s_j \leq -\varepsilon_i)$$

Et en supposant que le résidu suit une loi logistique, on aboutira au modèle logistique ordonné cumulatif :

$$\ln\left(\frac{P(C_i \leq j)}{1 - P(C_i \leq j)}\right) = x_i' \beta - s_j$$

Par rapport au modèle polytomique non ordonné, ce modèle se distingue par le fait que les paramètres β ne dépendent pas de j . On parle d'égalité des pentes. Ceci tient au fait que le modèle sur la variable latente porte seulement sur la moyenne, alors qu'on pourrait supposer une relation différente à certains niveaux de la distribution. Il est possible de procéder à une modélisation sans cette hypothèse d'égalité des paramètres et de tester la perte de qualité en la faisant. Ainsi, le modèle ordonné et le modèle non ordonné se comparent, dans le calcul du pseudo-R², à la même log-vraisemblance « vide », mais comme le modèle ordonné prend en compte moins de paramètres, l'indice proposé ici sera toujours inférieur à l'indice d'entropie habituel⁵. Le rapport entre l'indice ordonné et l'indice d'entropie, donc compris entre 0 et 1, donne une mesure du caractère plus ou moins ordonné de la ségrégation.

Par ailleurs, au lieu de résumer la ségrégation dans chaque unité par $G-1$ paramètres, on n'en retient qu'un, qui résume la propension de l'unité à accueillir des individus plus ou moins favorisés. Le pseudo-R² de Mac Fadden de ce modèle paraît un indicateur pertinent de la relation entre l'appartenance aux unités et l'appartenance aux catégories sociales, en tenant compte du caractère ordonné de celle-ci. Notons qu'il n'est pas possible d'établir une formule simple du pseudo-R² en

⁵ On peut envisager de contourner l'effet mécanique du plus grand nombre de coefficients, en utilisant un pseudo-R² ajusté, comme il existe des R² ajustés. Cela permet aussi de tenir compte du nombre plus ou moins grand d'unités par rapport au nombre d'individus, comme proposé en annexe B.

fonction des proportions : il doit faire l'objet d'une estimation numérique. C'est aussi le cas des paramètres des unités. En théorie, le paramètre pour chaque unité dépend alors non seulement de ses caractéristiques (répartition par catégories sociales), mais aussi de celles des autres unités. Cependant, des simulations sur données réelles et fictives ont montré que, hors cas très particuliers (avec une grosse variation de la répartition d'ensemble et des unités très « polarisées »), l'ordre des unités selon leur propension à accueillir des individus plus ou moins favorisés est très stable quand on change de population de référence. Cet indicateur pourrait fournir une mesure synthétique du profil social moyen de chaque unité.

Enfin, le lien entre étude d'une variable ordinale et étude d'une variable quantitative permet d'envisager l'utilisation des modèles à effets aléatoires. En effet, dans les études sur la ségrégation utilisant les modèles à effets aléatoires, la variable sociale est soit quantitative, soit dichotomique. L'usage d'effets aléatoires pour modéliser une variable non ordonnée ne paraît pas courant : il est alors nécessaire de postuler différents traits latents, corrélés, ce qui rend la modélisation assez complexe à estimer et interpréter. En revanche, l'utilisation d'un modèle polytomique ordonné, avec un seul trait latent, se prête bien à la prise en compte d'effets aléatoires, ce que nous essayerons dans la suite.

2.4. Comparaison des indicateurs sur données simulées

Avant de mettre en œuvre les différentes méthodologies sur des données réelles, nous allons les appliquer à des données fictives judicieusement simulées. Au total, nous retiendrons 8 indicateurs :

- H, l'indicateur d'entropie habituel sur la variable catégorielle
- H_h , l'indicateur d'entropie hiérarchisé utilisé par Floch (2017)
- $H_{\log EF}$, l'indicateur issu d'une régression logistique ordonnée avec effets **fixes**
- $H_{\log EA}$, l'indicateur issu d'une régression logistique ordonnée avec effets **aléatoires**
- R^2_{EF} , s'il existe une variable quantitative à l'origine de la variable catégorielle, la part de variance expliquée par le niveau unité dans un modèle à effets **fixes**
- R^2_{EA} , s'il existe une variable quantitative à l'origine de la variable catégorielle, la part de variance expliquée par le niveau unité dans un modèle à effets **aléatoires**
- R^2_{EFC} , la part de variance expliquée par le niveau unité dans un modèle à effets **fixes** en transformant la variable catégorielle en variable quantitative de façon simple (1 pour la catégorie la plus défavorisée, 2 pour la suivante, etc.)
- R^2_{EAC} , la part de variance expliquée par le niveau unité dans un modèle à effets **aléatoires** en transformant la variable catégorielle en variable quantitative de façon simple (1 pour la catégorie la plus défavorisée, 2 pour la suivante, etc.)

Tableau 1 : différences entre des indicateurs de ségrégations sur des cas-types

	Population A	Population B	Population C
H	0,14	0,14	0,14
H_h	0,18	0,07	0,07
$H_{\log EF}$	0,12	0,01	0,00
$H_{\log EA}$	0,27	0,03	0,00
R^2_{EF}	Pas de variable quantitative		
R^2_{EA}	Pas de variable quantitative		
R^2_{EFC}	0,29	0,04	0,00
R^2_{EAC}	0,28	0,03	0,00

Dans un premier temps, nous allons calculer ces indicateurs pour les populations fictives A et B présentées plus haut, ainsi que pour la population C, très polarisée, décrite dans la note 3, avec une moitié d'unités de répartition (10 %, 40 %, 40 %, 10 %) et une autre moitié de répartition (40 %, 10 %, 10 %, 40 %).

Alors que comme indiqué plus haut, l'indice d'entropie habituel a la même valeur pour les trois populations (0,14), tous les autres indices donnent des valeurs plus faibles pour les populations B et C par rapport à la population A, car la ségrégation y est moins ordonnée. Le phénomène est plus net pour les indices fondés sur des principes économétriques, avec des valeurs divisées par environ 10 pour la population B (de 0,12 à 0,01, avec l'indice fondé sur une régression polytomique ordonnée à effets fixes) et une valeur exactement nulle pour la population C. L'indice présenté par Floch (2017) a un effet moins marqué, mais non négligeable (de 0,18 pour la population A à 0,07 à la population B). Par ailleurs, il accorde la même valeur aux populations B et C. Enfin, les valeurs pour la modélisation quantitative arbitraire (il n'y a pas dans cet exemple de variable quantitative observée sous-jacente), sont proches de la modélisation polytomique ordonnée à effets aléatoires (celles à effets fixes donnant des valeurs plus faibles).

Nous allons procéder à une comparaison dans un cadre plus général, en construisant des populations plus ou moins ségréguées et plus ou moins polarisées, pour mettre en évidence des différences entre les indicateurs.

- Pour 50 unités, on simule un profil social quantitatif, qui visera une moyenne m_i tirée dans une loi normale (0,1) et un écart-type $e_i = c_i \frac{100-S}{S}$ où c_i est tiré dans une loi normale (1,P) jusqu'à obtenir une valeur comprise en 0 et 2. S est le coefficient de « ségrégation ordonnée » (il correspond à la variance inter-unité sur la variance totale) ; P est le coefficient de « ségrégation polarisée ».
- Dans chaque unité, on tire des valeurs du profil social quantitatif pour 500 individus dans la loi normale (m_i, c_i).
- Une variable catégorielle est créé avec une répartition arbitraire 20 %, 50 %, 20 %, 10 % sur l'ensemble de la population.

Le nombre assez restreint d'unités crée des perturbations : plutôt qu'augmenter ce nombre (ce qui conduit dans le calcul de certains indicateurs à des temps très longs), nous avons procédé à 20 simulations par valeur de (S, P). On a fait varier S de 5% à 30 % (ce qui correspond aux niveaux observés de ségrégation entre collègues) et P de 0 à 2 (la valeur 0 correspond à une absence de polarisation).

Tableau 2 : lien entre des indicateurs de ségrégations sur des populations simulées

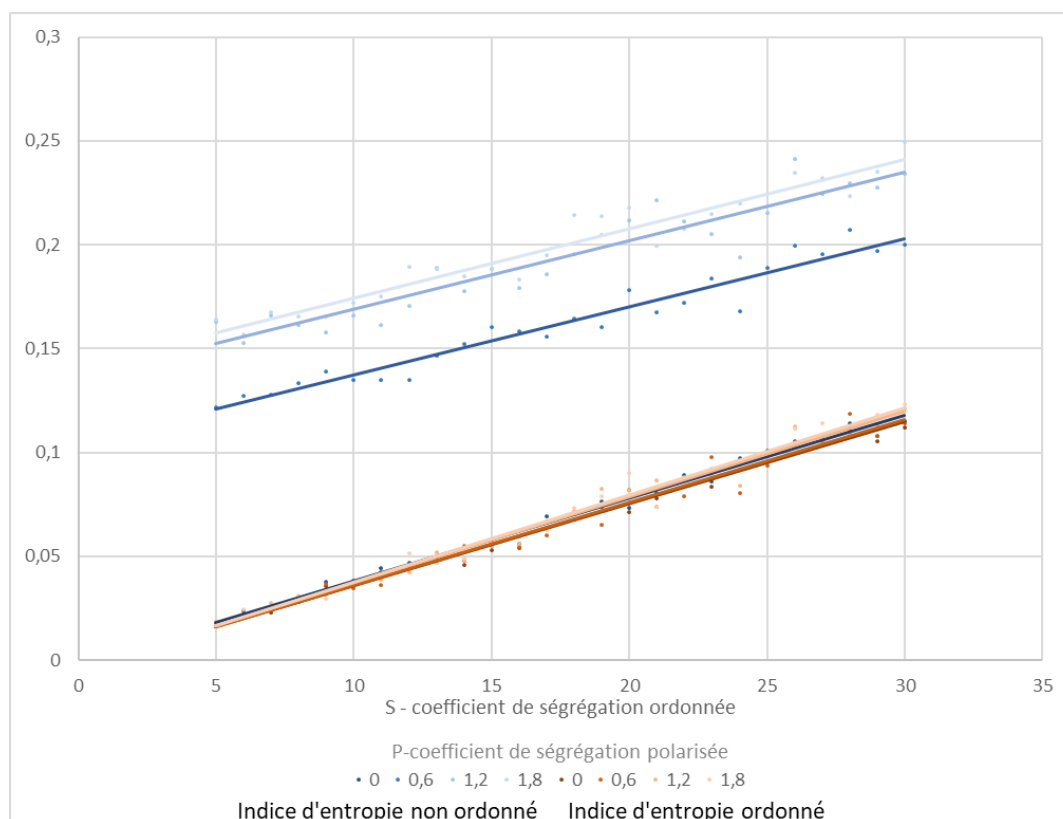
Variable	H	H _h	H _{logEF}	H _{logEA}	R ² _{EF}	R ² _{EA}	R ² _{EFC}	R ² _{EAC}
H	1							
H _h	0,96	1						
H _{logEF}	0,56	0,76	1					
H _{logEA}	0,56	0,76	0,98	1				
R ² _{EF}	0,38	0,61	0,97	0,93	1			
R ² _{EA}	0,38	0,61	0,97	0,93	1,00	1		
R ² _{EFC}	0,52	0,73	0,99	0,97	0,98	0,98	1	
R ² _{EAC}	0,52	0,73	0,99	0,97	0,98	0,98	1,00	1

Il y a une très forte corrélation entre les indicateurs fondés sur des principes économétriques. Dans une perspective quantitative, la corrélation entre les R^2 avec les modèles à effets fixes et ceux avec effets aléatoires est quasi parfaite, que l'on utilise la variable quantitative d'origine ou la catégorie sociale quantifiée de façon arbitraire avec les valeurs (1, 2, ...). Cela tient à ce que les deux méthodes se distinguent surtout dans le traitement de la ségrégation aléatoire, plus ou moins grande selon le nombre d'unités et d'individus par unité. Or ici, ces paramètres sont fixes, ce qui explique l'absence de divergence entre ces indicateurs. Signalons que les indicateurs sur la variable quantitative d'origine ou sur la catégorie sociale quantifiée sont aussi fortement liés entre eux (0,98). Mais ce résultat ne peut être extrapolé, car il est sans doute possible de simuler des populations, où le codage arbitraire (1, 2,...) ne correspondant pas à la variable quantitative d'origine, les deux perspectives seront moins convergentes.

Les corrélations entre les modèles logistiques ordonnés sur la catégorie sociale en 4 groupes et les modèles quantitatifs sont aussi très bonnes : 0,99 par exemple entre le modèle logistique à effets fixes et les modèles sur la catégorie sociale quantifiée ; 0,93 entre le modèle logistique à effets aléatoires et le modèle sur la variable quantitative d'origine (moins bonne corrélation). Le lien entre modèles à effets fixes et modèles à effets aléatoires est, contrairement à ce que nous avons dit sur les modèles quantitatifs, imparfait pour le modèle logistique ordonné, tout en restant très élevé (0,98).

En revanche, l'indice d'entropie habituel est beaucoup moins bien corrélé avec les indicateurs fondés sur des modèles économétriques : la corrélation reste correcte avec les modèles logistiques ordonnés (0,56), mais est beaucoup moins satisfaisante avec les modèles sur la variable quantitative d'origine (0,38). C'est le coefficient de polarisation P qui provoque ces divergences. Si P vaut 0, à S (ségrégation ordonnée) fixé, l'indice d'entropie habituel et l'indice issu d'une modélisation logistique ordonnée sont extrêmement proches (graphique 2). Quand on fait varier le coefficient P , cela n'a pas d'impact sur l'indice issu d'une modélisation logistique ordonnée, dont les valeurs à S fixé, restent quasi constantes. En revanche, l'indice d'entropie augmente sensiblement, intégrant cet aspect de polarisation dans sa mesure de la ségrégation, ce qui provoque des décalages avec l'autre indice.

Graphique 2 : lien entre l'indice d'entropie habituel et l'indice issu d'une modélisation polytomique ordonnée à effets fixes

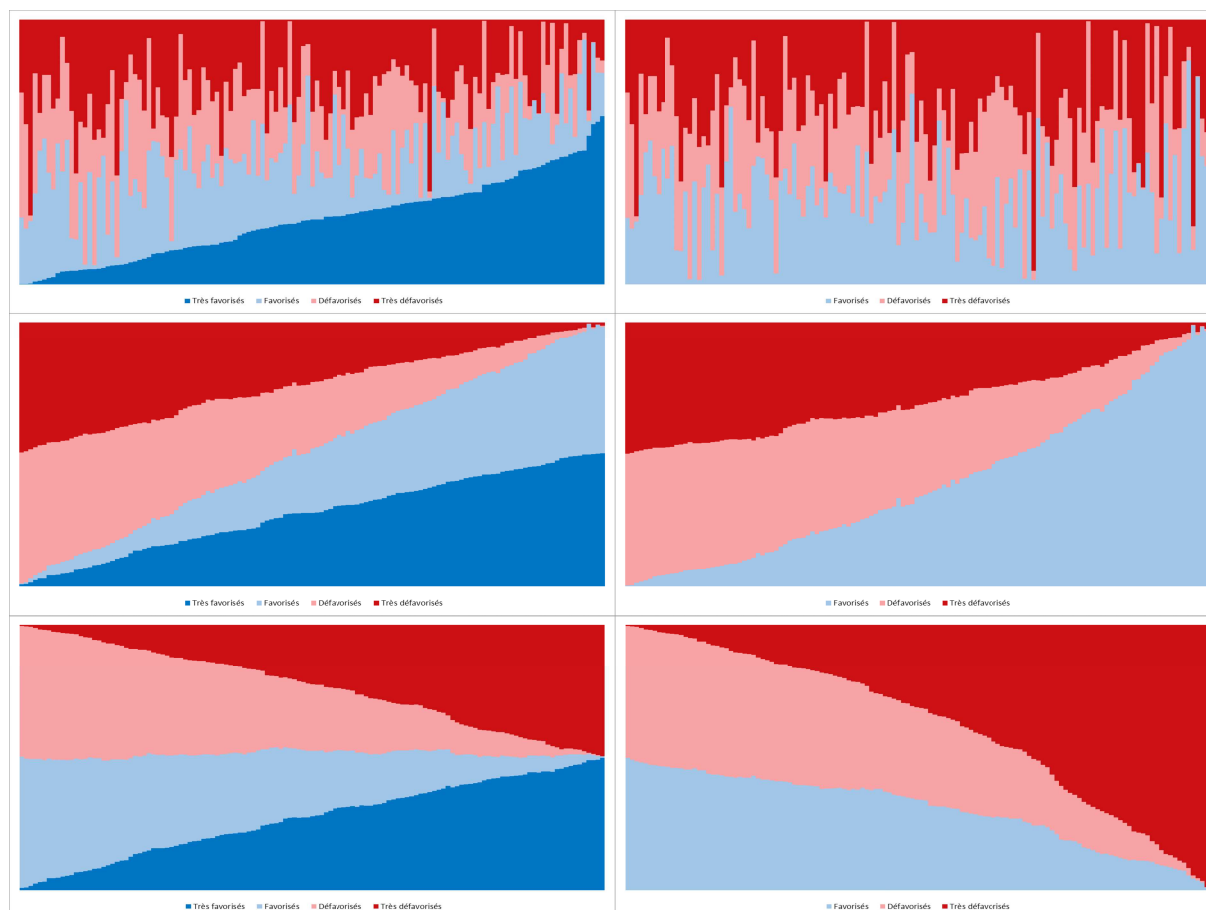


Enfin, l'indice d'entropie hiérarchisé présenté par Floch (2017), se trouve dans une situation intermédiaire, fortement corrélé avec l'indice d'entropie habituel (0,96), mais présentant des corrélations plus élevées avec les indices économétriques (0,76 avec les indices issus des modélisations polytomiques ordonnées, 0,61 avec les indices sur la variable quantitative d'origine).

Ainsi, la modélisation du comportement des unités par un seul paramètre, propension à accueillir des individus plus ou moins favorisés, peut mal rendre compte des données. On peut en tenir compte ou pas selon que sa vision de la ségrégation accorde une plus ou moins grande importance à l'aspect ordonné de la ségrégation. Dans une approche descriptive, on peut aussi envisager d'enrichir le modèle : en ajoutant un paramètre de « polarisation » ou un paramètre pour une catégorie sociale particulière, si l'on pense qu'elle est à l'origine de la divergence par rapport à modèle purement ordonné (par exemple la répartition des agriculteurs qui doit obéir à une logique particulière). Dans la suite, nous aborderons une démarche plus pragmatique en revenant aux données brutes et en les travaillant pour comprendre l'origine de la « ségrégation non ordonnée ».

Dans le cas où le nombre d'unités est assez restreint (quelques centaines de collèges par département dans la suite), il est possible de représenter les résultats pour l'ensemble des unités sous forme de graphiques assez lisibles. Selon que la ségrégation est non ordonnée, ordonnée ou polarisée, ces graphiques se différencient de façon très nette.

Graphique 2 : répartition d'unités par profil social sur 3 populations simulées

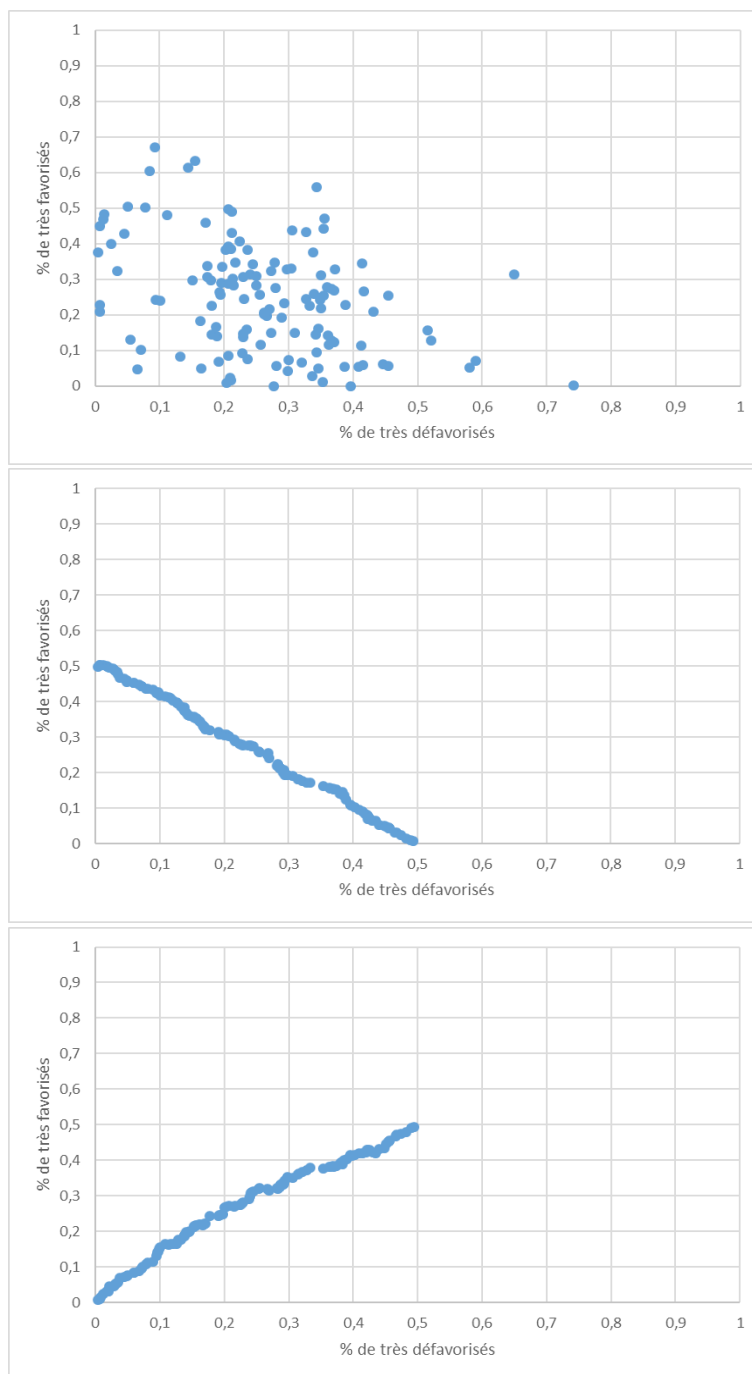


Chaque graphique ci-dessus donne des informations pour 130 unités, en colonnes, les couleurs indiquant la répartition dans les 4 catégories sociales hiérarchisées. Les unités ont été triées de celle accueillant le moins d'individus très favorisés à gauche à celle en accueillant le plus à droite. Le premier graphique présente le cas d'une ségrégation non hiérarchisée. Notons d'abord que la ségrégation est assez forte, avec de nombreuses unités où les individus très favorisés représentent plus de 30 % de la population et d'autres où ils sont moins de 5 %. On ferait le même constat en triant

par l'une des trois autres catégories. Cependant, ces ségrégations semblent « indépendantes » : les unités étant triés selon la proportion d'individus très favorisés, on n'observe pas de logique dans la répartition des trois autres catégories. Cela apparaît encore plus nettement, quand on retire la catégorie très favorisée de la répartition, en conservant le même ordre pour les unités (graphique de droite) : les proportions dans les trois couleurs varient d'une colonne à l'ordre sans aucune tendance.

La situation est différente dans la population en-dessous, avec une ségrégation très ordonnée : dans les unités avec le plus d'individus très favorisés, il y a essentiellement des individus favorisés et très peu d'individus (très) défavorisés. La part de ces derniers croît régulièrement au fur et à mesure que la part des individus (très) favorisés décroît. Ce que l'on voit encore plus nettement dans le graphique de droite, sans la part des très favorisés.

Graphique 3 : lien entre les proportions des catégories sociales par unité sur 3 populations simulées



La dernière population présente une situation très théorique, avec une polarisation extrême : la part des individus très défavorisés croît avec celle des individus très favorisés, celles des deux classes moyennes se réduisant d'autant. Si on attribue les valeurs (1, 2, 3 et 4) aux catégories sociales, la valeur moyenne de chaque unité varie très peu autour de 2,5, et les indicateurs de ségrégation ordonnés sont logiquement très faibles.

Une autre façon de distinguer les différentes formes de ségrégation est de croiser les proportions. Dans les graphiques ci-dessus, on présente un nuage de points où l'abscisse est la proportion d'individus très défavorisés dans l'unité et l'ordonnée la proportion d'individus très favorisés. Le premier graphique présente la situation d'une ségrégation non ordonnée : aucune relation nette n'apparaît entre les deux variables. Dans le deuxième cas, une corrélation négative rend compte que quand l'une des proportions est très élevée, l'autre est très basse, signe d'une ségrégation ordonnée ; en revanche, dans le dernier cas, la corrélation est positive et illustre la forte polarisation des unités dans cette population. Signalons qu'il y a une contrainte sur la répartition, induisant une dépendance négative : dans un unité où 90 % des individus sont très défavorisés, les très favorisés ne peuvent représenter plus de 10 % des effectifs. On peut contourner cette difficulté en calculant la proportion d'individus très favorisés, en excluant les individus très défavorisés.

Cette étude des corrélations entre proportions peut être généralisée, pour prendre en compte toutes les catégories sociales, en utilisant des techniques d'analyses de données : l'Analyse Factorielle des Correspondances (AFC) paraît alors l'outil le plus indiqué, comme présenté en annexe C.

3. La ségrégation entre collègues

Nous allons maintenant appliquer la méthodologie décrite ci-dessus à des données réelles. Il va s'agir d'étudier la ségrégation sociale entre collègues pour les élèves à l'entrée en 6^e (y compris SEGPA). Pour assurer une certaine robustesse, nous avons empilé toutes les années scolaires de 2003-2004 à 2020-2021. Nous ne conservons que les établissements qui ont scolarisé au moins 500 élèves sur cette période et où le taux de PCS renseignées est supérieur à 10 %. Cela représente 6 367 collèges (sur 7 728 au départ) pour 13 059 186 élèves (pour 13 983 216 au départ). Les professions des deux responsables ont été utilisées pour calculer l'indice de position sociale (Rocher, 2016). Une variable catégorielle en 4 groupes égaux a été construite à partir de cet indice.

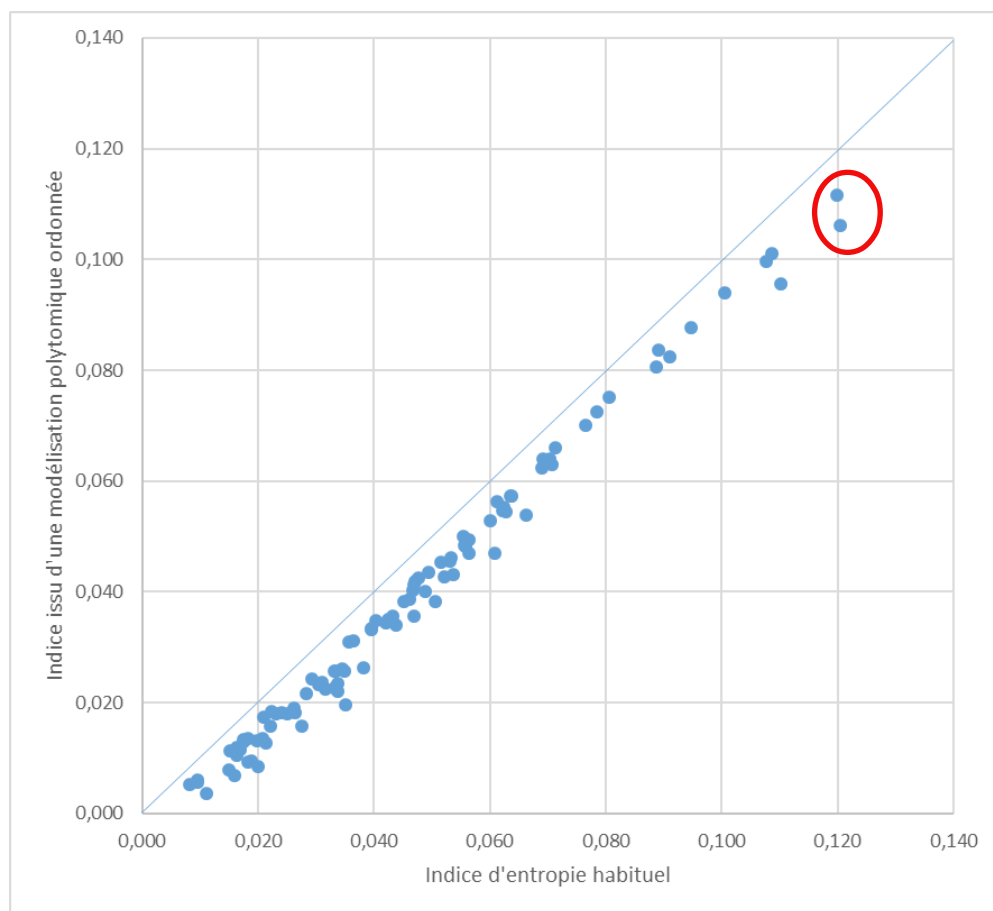
Pour déterminer ce qu'apportent les indicateurs ordonnés, nous allons calculer les différents indicateurs par département, car il a été montré qu'il y avait une forte variabilité de la ségrégation entre départements, en particulier en fonction de leur caractère plus ou moins urbain (Givord et alii, 2016). La matrice de corrélations entre ces indicateurs (tableau 3) montre une très forte cohérence des classements (on dépasse 0,98).

Tableau 3 : lien entre des indicateurs départementaux de ségrégation entre collègues

Variable	Moyenne	H	H _h	H _{logEF}	H _{logEA}	R ² _{EF}	R ² _{EA}	R ² _{EFC}	R ² _{EAC}
H	0,046	1							
H _h	0,060	0,997	1						
H _{logEF}	0,039	0,996	0,999	1					
H _{logEA}	0,103	0,984	0,983	0,986	1				
R ² _{EF}	0,106	0,989	0,995	0,995	0,983	1			
R ² _{EA}	0,107	0,982	0,983	0,985	0,994	0,986	1		
R ² _{EFC}	0,097	0,991	0,997	0,997	0,982	0,998	0,983	1	
R ² _{EAC}	0,098	0,985	0,987	0,988	0,997	0,986	0,998	0,986	1

Intéressons nous aux valeurs moyennes, qui vont donner quelques indications sur l'ensemble de la France (nous ne présentons pas ici des valeurs plus directes, car avec plus de 6 000 groupes, les modèles logistiques, en particulier le modèle ordonné, ne parviennent pas à converger en un temps raisonnable). Le R^2 de l'IPS, variable quantitative de référence de cette étude, est d'un peu plus de 10 %, que l'on utilise un modèle à effets fixes ou un modèle à effets aléatoires. On obtient des valeurs proches avec la catégorie sociale quantifiée ou le modèle polytomique ordonné à effets aléatoire. En revanche, le modèle polytomique ordonné à effets fixes donne une valeur plus faible (0,039). Ce n'est pas une surprise, les expertises sur les pseudo- R^2 (Veal, Zimmerman, 1996) montrant que le pseudo- R^2 de Mac Fadden de la régression d'une variable catégorielle a tendance, dans un cas plus général, à sous-estimer le R^2 de la variable quantitative ayant servi à construire la variable catégorielle. D'autres pseudo- R^2 ont un meilleur comportement, mais ils n'ont pas de lien avec les indicateurs de ségrégation. Le plus important n'est pas tant la valeur de l'indice de ségrégation que la façon dont il varie dans le temps ou entre des populations différentes.

Graphique 4 : lien entre l'indice d'entropie et l'indice issu de la modélisation polytomique ordonnée



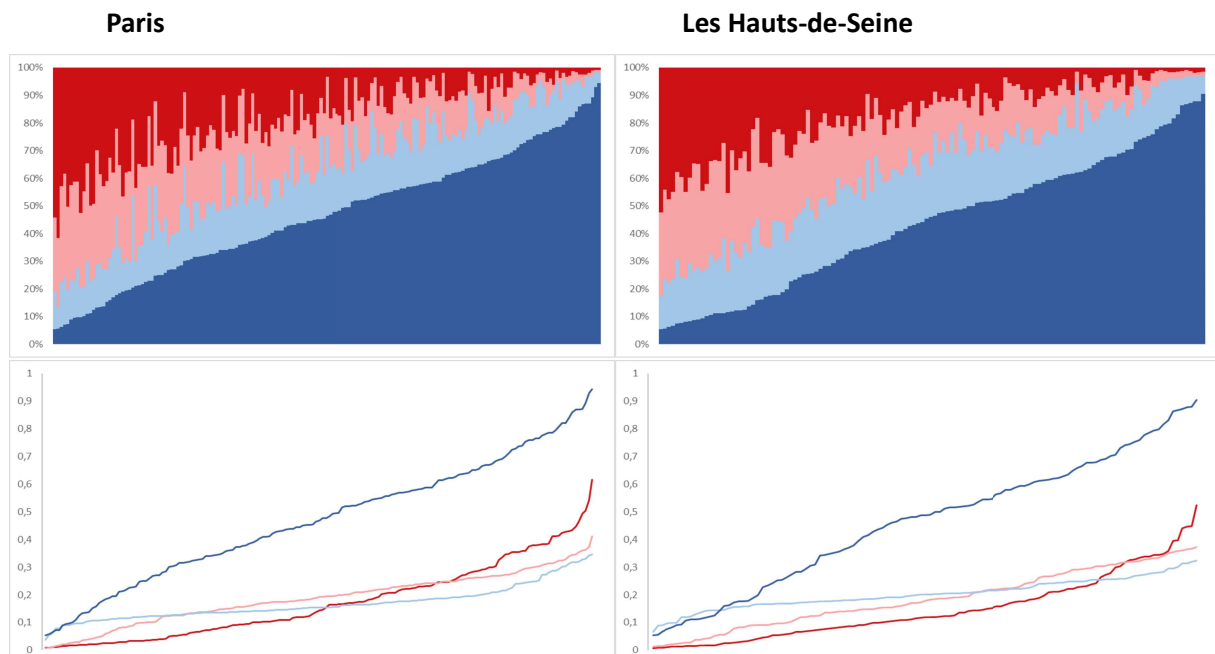
Il est plus intéressant de comparer l'indice du modèle polytomique ordonné à effets fixes avec l'indice d'entropie habituel. En effet, nous avons montré que le premier était forcément plus petit que le second et que le rapport indiquait dans quelle mesure on pouvait considérer la ségrégation comme ordonnée. Ici, le rapport est de 85 % (0,039/0,046), ce qui suggère donc que la ségrégation sociale entre collègues a un caractère ordonné très marqué, sans être parfait. De façon plus précise, en considérant l'ensemble des valeurs départementales (graphique 4), un décalage de 0,006 apparaît de façon à peu près systématique à tous les niveaux de ségrégation, à la fois dans les départements où la ségrégation est faible (avec des indices de 0,02) que dans ceux où elle est forte (indices à plus de 0,1). Cela signifie, en faisant le rapport entre les deux indices, que la ségrégation paraît moins « ordonnée » dans les départements où son niveau est faible. Cela peut venir du fait que la ségrégation ainsi mesurée est un peu « aléatoire » (même si en empilant les données, avec des unités

à plus de 500 élèves, nous éliminons une grande partie de l'effet lié au problème de taille) et non ordonnée.

Même si la corrélation entre les deux indices est très forte, il apparaît quelques décalages, dont l'analyse peut fournir des informations intéressantes. Nous allons prendre le cas des deux départements les plus ségrégués, par les deux indices d'ailleurs : Paris et les Hauts-de-Seine. Alors que le niveau indiqué par l'indice d'entropie habituel est identique dans les deux départements (0,12), l'indice fondé sur le modèle polytomique ordonné est un peu plus fort dans les Hauts-de-Seine, (0,112 contre 0,106) suggérant que la ségrégation y est plus ordonnée. Pour le visualiser, nous avons présenté la répartition par catégories sociales de tous les collèges de ces deux départements (graphique 5). Le caractère très ségrégué de ces académies apparaît nettement : à Paris, un peu plus d'un cinquième des collèges accueillent moins de 30 % d'élèves très favorisés et un sixième en accueillent plus de 70 % (l'indice d'entropie dichotomique est de 0,10). Le niveau de ségrégation est aussi élevé pour les élèves très défavorisés (indice d'entropie de 0,12), mais avec une proportion globale d'élèves de cette catégorie plus basse : 38 % des collèges accueillent moins de 10 % d'élèves défavorisés, alors que 8 % en accueillent plus de 40 % (partie basse du graphique 5 : la courbe en bleu reprend la courbe au-dessus en classant par proportion d'élèves très favorisés, les autres courbes correspondent à la même illustration en triant par les autres catégories sociales). La ségrégation pour les catégories moyennes est sensiblement plus faible (indice d'entropie autour de 0,02).

Le caractère moins ordonné de la ségrégation à Paris se manifeste par le fait qu'à proportion d'élèves très favorisés fixée, il est moins facile de prédire la proportion de chacune des autres catégories et en particulier des élèves très défavorisés : même si la relation est nettement négative et ordonnée (plus la proportion d'élèves très favorisés augmente et plus celle des élèves défavorisés baisse et cela de façon plus nette que la proportion d'élèves favorisés), la courbe rouge est plus « dentelée » que dans les Hauts-de-Seine, signalant de nombreuses « irrégularités ».

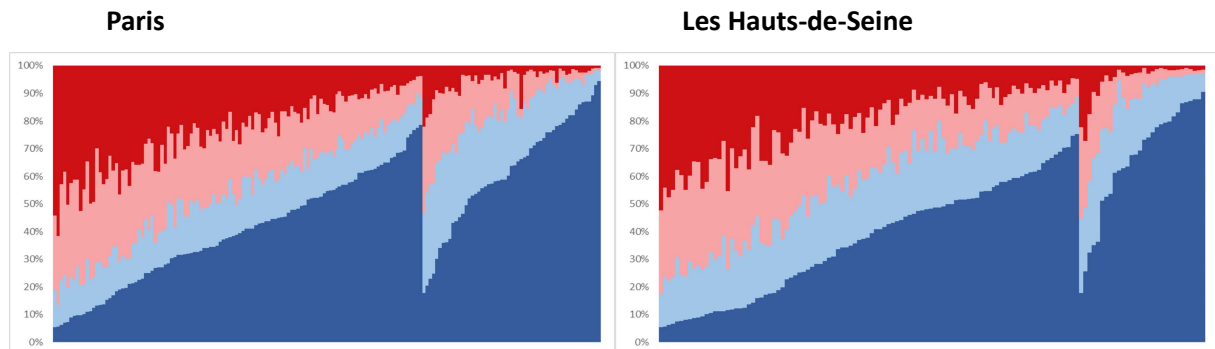
Graphique 5 : répartition des collèges par profil social à Paris et dans les Hauts-de-Seine



Un élément d'explication a pu être trouvé en distinguant secteur public et secteur privé : dans le graphique 6, avant de trier par proportion d'élèves très favorisés, on a trié par secteur, avec les collèges publics en premier. A Paris, les deux secteurs se distinguent de façon très nette, avec une tendance dans le privé à scolariser plus d'élèves très favorisés (la courbe en bleu est plus haute) et moins d'élèves très défavorisés (les barres rouges y sont nettement plus petites). Il apparaît aussi que

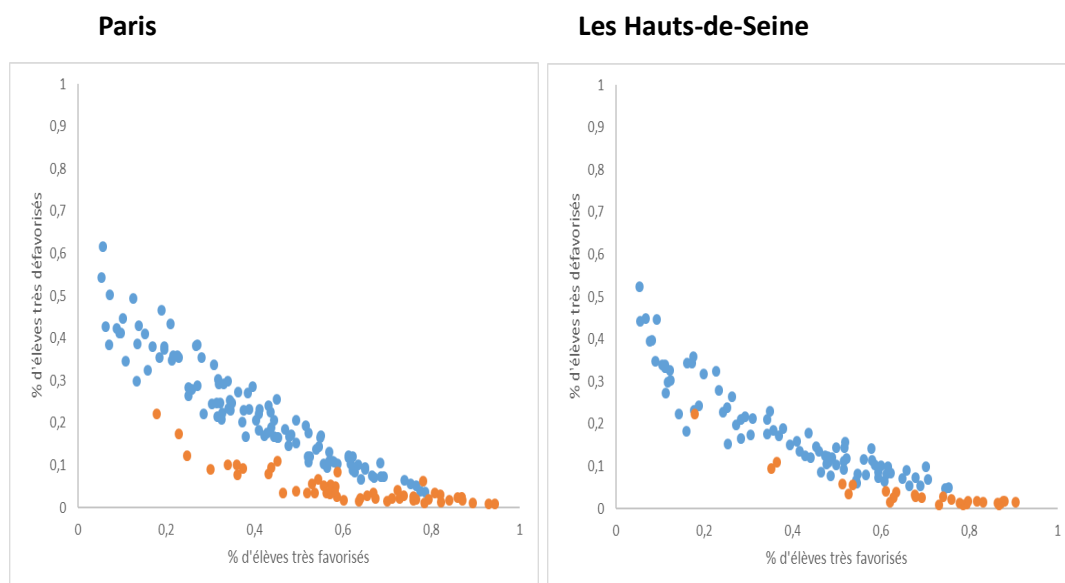
tant dans le public que dans le privé, la courbe rouge est moins « dentelée » que sur l'ensemble des collèges (avec juste un point aberrant dans le privé, en milieu de distribution). Ainsi, alors que sur l'ensemble des collèges, l'indice ordonné vaut 88 % de l'indice d'entropie habituel (0,106/0,12), l'écart se resserre dans le public (0,068/0,747=91%) et dans le secteur privé (0,092/0,102=91 %), avec des valeurs proches des Hauts-de-Seine. Le caractère moins ordonné des écarts entre collèges à Paris s'explique donc en partie par un « écart moins ordonné » entre public et privé : la surreprésentation des élèves très favorisés dans le privé est moins forte que la sous-représentation des élèves très défavorisés.

Graphique 5 : répartition des collèges par profil social à Paris et dans les Hauts-de-Seine, en distinguant les deux secteurs



Ce résultat peut aussi être illustré en présentant le croisement des deux proportions (graphique 5). A Paris, comme dans les Haut-de-Seine, les collèges privés (en orange) se concentrent dans la partie droite et basse du croisement, cumulant une forte proportion d'élèves très favorisés et une faible proportion d'élèves très défavorisés. En particulier, les collèges accueillant plus de 70 % d'élèves très favorisés sont presque tous privés. Les deux départements vont se distinguer à des concentrations plus faibles d'élèves très favorisés : à Paris, les collèges privés se différencient nettement des collèges publics accueillant la même proportion d'élèves très favorisés, en scolarisant souvent 10 points de moins d'élèves très défavorisés (alors que dans les Hauts-de-Seine, le décalage vers le bas est beaucoup moins net).

Graphique 5 : lien entre la proportion d'élèves très défavorisés par collège et la proportion d'élèves très favorisés, à Paris et dans les Hauts-de-Seine, en distinguant les deux secteurs



Le secteur privé parisien se caractérise donc surtout par une sous-représentation des élèves très défavorisés, alors que dans les Hauts-de-Seine, le caractère favorisé du secteur privé est plus global. Dans l'annexe D, d'autres exemples sont présentés : les départementaux où la ségrégation paraît moins ordonnée qu'ailleurs, sont souvent des départements de l'ouest de la France, avec un phénomène proche de la situation parisienne : les collèges privés y scolarisent à peu près autant d'élèves très favorisés que les collèges publics, mais ils accueillent beaucoup moins d'élèves très défavorisés, semblant viser un public de classes moyennes. C'est une forme de ségrégation partiellement non ordonnée.

4. Conclusion

Une part importante de ce travail a été consacré au lien entre indices de ségrégation et modélisation économétrique. Cela nous semblait avoir un intérêt pédagogique, car les propositions précédentes, fondées sur des modèles à effets aléatoires, relativement abstraits (Goldstein, Noden, 2003), ont parfois été reçues avec réticence par des chercheurs attachés à une vision plus descriptive (Gorard, 2004). En montrant le lien entre les indices de ségrégation et les modèles à effets fixes, par l'équivalence entre différents indices de ségrégation existant et des pseudo- R^2 , ce travail établit un lien intermédiaire. Cependant les modèles à effets aléatoires semblent avoir l'avantage de mieux prendre en compte la ségrégation créée « artificiellement » par la petite taille de certaines unités. L'inconvénient est qu'il est alors difficile de présenter la situation des différentes unités individuellement (ce qui est parfois souhaité, avec une représentation cartographique).

Concernant la question de prendre en compte le caractère ordonné de la catégorie sociale et de la ségrégation, plusieurs éléments doivent être rappelés :

- Même si la catégorie sociale est ordonnée (ce qui n'est pas toujours le cas), il n'est pas obligatoire de vouloir le prendre en compte dans la mesure de la ségrégation. On le fera si l'on juge que la concentration de publics en difficulté sociale est particulièrement préoccupante ; on ne le fera pas si l'on juge au contraire que des unités très « polarisées » ne sont pas souhaitables ; on évitera aussi de le faire si l'on a une vision globale de la ségrégation, fondée sur la nécessité d'interactions précoces entre des individus d'origines diverses.

- Il est aussi possible d'avoir une approche « statistique » et « descriptive », sans principes théoriques marqués : dans ce cas, la confrontation des indices ordonnés et des indices non ordonnés peut apporter des informations intéressantes, repérer des situations où une analyse fine est souhaitable.

- Pour faire une mesure ordonnée de la ségrégation, trois approches peuvent être distinguées :

- * Une modélisation linéaire simple, avec une synthèse par une part de variance inter-unité, de la variable sous forme quantitative paraît la solution la plus logique (avec effets aléatoires ou effets fixes, selon les objectifs et contraintes de l'étude) : il est ainsi possible de prendre en compte toute l'information disponible et l'indicateur (R^2) est statistiquement assez simple à présenter. S'il n'y a pas de variable quantitative sous-jacente, une quantification arbitraire (1, 2, ...) semble donner des résultats assez robustes.

- * Une modélisation polytomique ordonnée permet de proposer une solution sans quantification arbitraire (mais le modèle repose tout de même sur certaines hypothèses) : l'indicateur que l'on obtient peut être comparé à l'indice d'entropie habituel, ne prenant pas en compte le caractère ordonné de la variable, ce qui est une façon de mesurer le caractère ordonné de la ségrégation.

- * La proposition de Floch (2017) a aussi un grand intérêt : plus que l'indicateur synthétique, qui semble très proche de l'indice d'entropie habituel, c'est le profil de la ségrégation à différents niveaux de la distribution de la variable sociale qui apporte de l'information. Il permet de s'extraire de l'hypothèse d'uniformité sur l'ensemble de la distribution de la catégorie sociale de la corrélation en fonction de l'appartenance aux unités. Il permet aussi de tester d'une autre façon que la deuxième proposition, si la ségrégation est ordonnée ou non.

Il conviendra de mettre à l'épreuve des données cette démarche. L'article de Floch (2017) a permis de montrer que la ségrégation des revenus entre quartiers dans les aires urbaines, prenait des formes diverses et souvent non uniformes : elle tendait à être plus forte pour les ménages à revenus très élevés que pour les ménages pauvres.

L'étude faite dans cette note, dans le domaine scolaire, montre au contraire une forte cohérence entre les indices ordonnés et les indices non ordonnés, quant à la mesure de la ségrégation sociale entre collèges. Cela tient en partie au caractère très ordonné de la ségrégation dans la plupart des départements. Dans quelques cas, une petite divergence est apparue : elle a pu être en partie expliquée par des comportements différents entre secteurs public et privé, qui ne relèvent pas d'une logique tout à fait ordonnée ; dans certains cas, plus que par une surreprésentation des élèves très favorisés, les collèges privés se caractérisent par l'éviction des élèves très défavorisés, au profit des classes moyennes. Il n'est pas impossible que d'autres facteurs interviennent : la répartition des agriculteurs dans les collèges par exemple suit sans doute une logique plus géographique que sociale.

Ces résultats confirment à la fois le caractère assez ordonné des classes sociales, comme a pu le montrer Thierry Rocher, il y a quelques années et dans une contribution à cette même session (Rocher, 2022) et le caractère ordonné de la ségrégation sociale entre collèges. Cependant, le fatalisme n'est pas inévitable : le niveau de ségrégation entre collèges paraît très variable d'un département à l'autre ; il est logiquement plus faible dans les départements moins urbains, avec un réseau de collèges moins denses, provoquant moins de disparités. Cela ne semble pas empêcher ces départements d'obtenir des performances scolaires tout à fait satisfaisantes (Ben Ayed, 2006).

Bibliographie

- [1] Afsa C., 2013, « Le modèle logit – Théorie et approche », *Document de travail de la DEPP – série Méthode*, n°2013-M02.
- [2] Ben Ayed C., Broccolichi S., Trancart D., Mathey-Pierre C., 2006, « Fragmentations territoriales et inégalités scolaires : des relations complexes entre la distribution spatiale, les conditions de scolarisation et la réussite des élèves », *Éducation et formations*, n° 74.
- [3] Givord P., Guillermin M., Monso O., Murat F. (2016). « Comment mesurer la ségrégation dans le système éducatif ? », *Éducation & formations*, n° 91, pp 21-51, septembre 2016.
- [4] Danhier J., Martin E., Alarcon-Henriquez A., Kaelen R., Jacobs D. (2017). « Une ségrégation peut en cacher une autre. La répartition des élèves entre classes à prendre au sérieux », *Revue française de pédagogie*, n°199, pp 117-138.
- [5] D'Haultfœuille X. Rathelot R., 2017, "Measuring segregation on small units: A partial identification analysis", *Quantitative Economics*, Vol. 8, No. 1, pp 39–73.
- [6] Floch, J.-M. (2017). « Niveau de vie et ségrégation dans douze métropoles françaises » *Economie et Statistique / Economics and Statistics*, 497-498, pp 73-96.
- [7] Frankel D.M., Volij O., 2011, « Measuring school segregation », *Journal of Economic Theory*, n°146, pp. 1-38.
- [8] Goldstein H, Noden P., 2003, « Modelling Social Segregation », *Oxford Review of Education*, 30:3, 225-237
- [9] Gorard S., 2004, « Comments on 'Modelling Social Segregation' by Goldstein and Noden », *Oxford Review of Education*, 29:2, 435-440.
- [10] Hauser, J.R., 1977, « Testing the Accuracy, Usefulness and Significance of Probabilistic Choice Models: An Information Theoretic Approach », *Operations Research*, 26, pp. 406-421
- [11] Kalter, F., 2000, « Measuring Segregation and Controlling for Independent Variables », *Arbeitspapiere - Mannheimer Zentrum für Europäische Sozialforschung*, n°19.
- [12] Massey, D., & Denton, N., 1988, « The Dimensions of Residential Segregation », *Social Forces*, 67(2), pp.281-315.
- [13] Monso O., Fougère F., Givord P., Pirus C., (2019) « Les camarades influencent-ils la réussite et le parcours des élèves ? Les effets de pairs dans l'enseignement primaire et secondaire », *Éducation & formations*, n° 100, pp 23-52.
- [14] Paterson L., 1992, « Multilevel Modeling and Segregation Indices », *Multilevel Modelling Newsletter*, n°4-1.
- [15] Poupeau F., François J.-F., 2008, *Le sens du placement*, Raisons d'agir édition
- [16] Rathelot R., 2012, "Measuring Segregation When Units are Small: A Parametric Approach", *Journal of Business and Economic Statistics*, 30(4), p. 546-553.
- Xavier D'Haultfœuille
- [17] Reardon S.F., Firebaugh G., 2002, « Measures of Multigroup Segregation », *Sociological methodology*, N°32, pp. 33-67.
- [18] Reardon S. F., 2011, « Measures of Income Segregation », *CEPA Working Papers*.
- [19] Rocher T, 2022, « Une échelle unidimensionnelle des PCS : revisiter la méthodologie de construction de l'IPS (indice de position sociale) », Journées de la méthodologie statistique 2022
- [20] Spriggs W. E., Williams R. M., 1996, « A Logit Decomposition Analysis of Occupational Segregation : Results for the 1970s and 1980s », *The Review of Economics and Statistics*, Vol. 78, N°2, pp. 348-355
- [21] Rocher T. (2016) « Construction d'un indice de position sociale », *Éducation & formations*, n° 90, pp 5-27.
- [22] Veall M. R., Zimmerman K. F., « Pseudo-R2 Measures for Some Common Limited Dependent Variable Models », *Sonderforschungsbereich*, 386, Paper 18

Annexe A : pseudo-R² et indices de ségrégation

Dans le texte, nous avons montré l'équivalence entre l'indice d'entropie et le pseudo-R² de Mac Fadden de la régression logistique de l'appartenance à une catégorie sociale par les indicatrices d'appartenance aux unités. Nous allons ici, en nous limitant à une catégorisation sociale dichotomique, étudier le lien entre les autres pseudo-R² proposés dans la littérature et les indicateurs de ségrégation.

Pseudo R² de Cox&Snell et pseudo R² de Nagelkerke

Le pseudo R² de Cox&Snell (ou de Maddala) est l'une de ces propositions. Il s'exprime ainsi :

$$PsR^2_{CetS} = 1 - e^{\frac{-\Delta}{n}} = 1 - e^{\frac{-2(L-L_0)}{n}} = 1 - \left(\frac{L_0}{L}\right)^{\frac{2}{n}}$$

Où L_0 est la vraisemblance du modèle nul.

Ce pseudo-R² a été construit en transposant le lien entre le R² d'un modèle linéaire et la vraisemblance du modèle, en retenant une spécification normale. Il a l'inconvénient de ne pas être compris entre 0 et 1 et d'avoir un maximum : $1 - (L_0)^{\frac{2}{n}}$. C'est pourquoi Nagelkerke (et de façon indépendante Cragg & Uhler) a proposé un autre pseudo-R² :

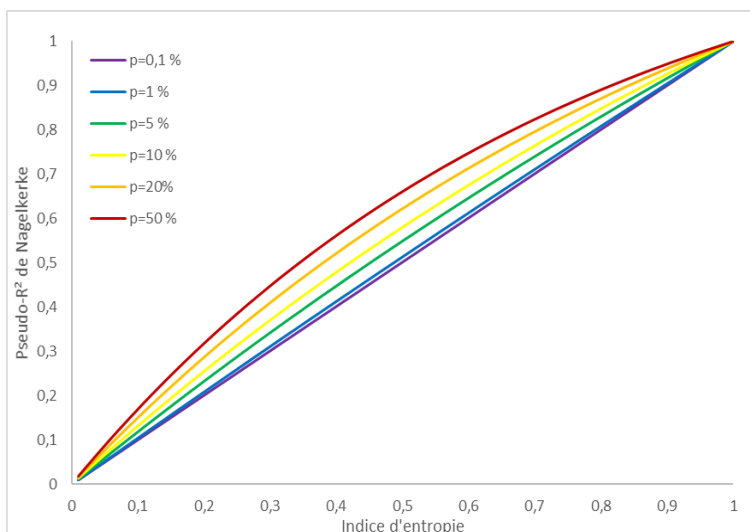
$$PsR^2_N = \frac{1 - \left(\frac{L_0}{L}\right)^{\frac{2}{n}}}{1 - (L_0)^{\frac{2}{n}}}$$

Dans le cas de notre modèle,

$$PsR^2_{CetS} = 1 - e^{-2h(p)H}$$

$$PsR^2_N = \frac{1 - e^{-2h(p)H}}{1 - e^{-2h(p)}}$$

Graphique 6 : lien entre le pseudo-R² de Nagelkerke et l'indice d'entropie



Le graphique ci-dessus illustre la relation entre l'entropie et le pseudo-R² de Nagelkerke pour différentes proportions d'individus défavorisés. Les différences sont surtout sensibles pour les proportions proches de 50 %, le pseudo-R² de Nagelkerke étant alors sensiblement plus haut que

l'indice d'entropie. Si l'on compare des zones (pays ou académie) très différentes en termes de proportion d'individus défavorisés, le classement selon le pseudo-R² de Nagelkerke et celui selon l'indice d'entropie peuvent sensiblement différer.

Pseudo R² d'Aldrich et Nelson et pseudo R² de Veall et Zimmerman

Aldrich et Nelson ont proposé un pseudo-R² s'inspirant du coefficient de contingence d'un tableau croisé, qui peut être considéré comme l'équivalent du coefficient de corrélation de Pearson pour une variable quantitative et donc de la racine carrée du R².

$$PsR^2_{AN} = \frac{\Delta}{\Delta+n}$$

Cet indicateur a l'inconvénient de ne pas avoir un maximum de 1. Veall et Zimmerman ont donc proposé une version standardisée :

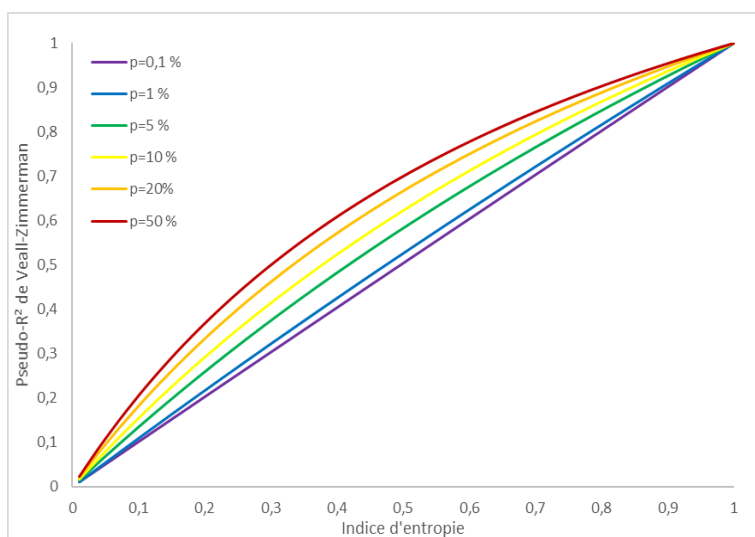
$$PsR^2_{AN} = \frac{\frac{\Delta}{\Delta+n}}{\frac{-2L_0}{-2L_0+n}}$$

Dans le cas de notre modèle :

$$PsR^2_{AN} = \frac{2h(p)H}{2h(p)H+1} = \frac{\frac{2h(p)H}{2h(p)+1}}{2h(p)+1}$$

La figure ci-dessous illustre la relation entre le pseudo-R² de Veall-Zimmerman et l'indice d'entropie. La divergence pour des proportions proches de 50 % est encore plus marquée qu'avec l'indice de Nagelkerke.

Graphique 7 : lien entre le pseudo-R² de Veall-Zimmerman et l'indice d'entropie



Pseudo R² d'Efron

Les indicateurs que nous allons maintenant présenter sont fondés sur une comparaison entre valeurs observées et valeurs prédites, reprenant le lien entre cette comparaison et le R² qui peut être fait dans une régression linéaire.

Le pseudo R² d'Efron va confronter l'indicatrice d'appartenance au milieu défavorisé à la probabilité estimée \hat{y}_i par le modèle. Cela conduit à poser

$$PsR^2_E = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{p(1-p)}$$

Dans cette note, on montre facilement que $\hat{y}_i = p_k$, k étant l'unité à laquelle appartient l'individu i .

$$PsR^2_E = 1 - \frac{\sum_{k=1}^m \frac{n_k}{n} \frac{1}{n_k} \sum_{i=1}^{n_k} (y_i - p_k)^2}{p(1-p)}$$

On retrouve la formule de la décomposition de la variance (ici mettant en évidence la variance intra) et donc le pseudo R² d'Efron est l'indice d'exposition normalisé.

Pseudo R² de Mac Kelvey et Zavoina

Le pseudo-R² de Mac Kelvey et Zavoina part d'une idée similaire, mais applique le principe de comparaison entre valeur prédite (\hat{y}^L) et valeur « réelle » à la variable latente (y^L). La valeur « réelle » n'est pas observée, pas plus que le résidu (ε), mais dans le cadre d'une régression logistique, ce résidu a une variance fixée : $\frac{\pi^2}{3}$.

$$PsR^2_{MKZ} = 1 - \frac{\text{Var}(\hat{y}^L)}{\text{Var}(y^L)} = 1 - \frac{\text{Var}(\hat{y}^L)}{\text{Var}(\hat{y}^L) + \text{Var}(\varepsilon)} = 1 - \frac{\text{Var}(\hat{y}^L)}{\text{Var}(\hat{y}^L) + \frac{\pi^2}{3}}$$

Des simulations de Monte Carlo ont montré que cet indicateur a de bonnes propriétés dans des situations classiques, mais est parfois délicat à calculer. Dans notre modèle, il va poser de gros soucis quand des unités ne comportent aucun individu défavorisé ou que des individus de ce type. En effet, \hat{y}^L correspond ici, à une constante près, au logit de la proportion d'individus défavorisés et donc si cette proportion est nulle ou égale à 1, le logit n'est pas défini. Sous SAS, la procédure LOGISTIC va tout de même fournir une estimation (on pourrait aussi en « bricoler » une, en rajoutant un individu défavorisé dans une unité où il n'y en a aucun), mais cette estimation est bien sûr très fragile.

D de Somers

Le dernier indicateur que nous allons présenter n'est pas considéré comme un pseudo-R², même s'il en a la forme (il est compris entre 0 et 1). Il est fondé sur la comparaison entre valeur observée et une prédiction fondée sur la probabilité estimée par le modèle (\hat{y}_i). On va prédire l'appartenance au milieu défavorisé si cette probabilité est supérieure à 0,5. Ensuite, on compare cette prédiction à la situation observée et on distingue le nombre de cas où les deux valeurs sont concordantes (n_c) et nombre de cas où les deux valeurs sont discordantes (n_d). L'indice D de Somers est défini de la façon

suivante (il existe des variantes tenant compte des situations où la probabilité vaut exactement 0,5 pour certains individus) :

$$D = \frac{n_c - n_d}{\frac{n(n-1)}{2}}$$

Il a été montré que cet indicateur est égal à l'indice de Gini.

Récapitulatif

Le tableau ci-dessous présente les indices de ségrégation et les pseudo-R² les plus classiques. Ce tableau rappelle le lien entre certains indicateurs de ségrégation et des indicateurs de qualité des régressions logistiques (indice d'entropie et pseudo-R² de Mc Fadden ; D de Somers et indice de Gini). Il montre aussi que l'indice de dissimilarité et l'indice d'Atkinson n'ont pas d'équivalents parmi les pseudo-R² (pour l'indice de dissimilarité, le lien fort avec l'indice de Gini permettrait sans doute de trouver quelque chose). Nous ne chercherons pas à établir ici un lien dans ce sens. En revanche, il nous paraît intéressant de noter que trois pseudo-R², plutôt bien notés dans les simulations sur le sujet, n'ont pas actuellement d'équivalents parmi les indices de ségrégation. L'utilisation de modèle probit pour proposer de nouveaux indicateurs pourrait aussi être explorée.

Tableau 4 : lien entre indices de ségrégation et pseudo-R²

Ségrégation	Pseudo-R ²
Indice de dissimilarité	Nd
Indice de Gini	D de Somers
Indice d'exposition normalisé	Pseudo-R ² d'Efron
Indice d'entropie	Pseudo-R ² de Mc Fadden
Indice d'information mutuelle	Log-vraisemblance divisée par n
Indice d'Atkinson	Nd
Nd	Pseudo-R ² de Nagelkerke
Nd	Pseudo-R ² de Veall-Zimmerman
Nd	Pseudo-R ² de Mac Kelvey et Zavoina

Annexe B : influence du nombre d'unités et du nombre d'individus par unité sur les indicateurs de ségrégation

Lorsque l'on étudie la ségrégation, c'est-à-dire les écarts entre des unités, le fait que certaines d'entre elles soient petites (une vingtaine d'individus par exemple pour des classes) peut provoquer une « ségrégation aléatoire », qui n'a pas forcément d'origine précise (mais qui peut avoir des conséquences réelles : les fluctuations aléatoires de contexte sont utilisées dans certaines études sur les effets de pairs, pour estimer ces effets, en neutralisant l'impact des facteurs inobservés possiblement liés aux regroupements non aléatoires des pairs). Ce phénomène est bien connu et il existe plusieurs solutions à ce problème (Rathelot, 2012 ; D'Haultfoeuille et Rathelot, 2017).

Nous allons ici profiter à la fois de l'énorme base de données réelles dont nous disposons et du processus de simulation des données, pour tester la robustesse des différentes méthodes de calcul des indicateurs à la présence de petites unités. Cette question est a priori indépendante de la proposition centrale de notre article, un indice de ségrégation issu d'une modélisation polytomique ordonnée et comme le calcul de cet indice est très gourmand en temps, nous ne le retiendrons pas ici. En définitive, nous allons comparer 5 indicateurs :

- l'indice d'entropie habituel

- l'indice d'entropie « ajusté » : comme l'indice d'entropie habituel est le pseudo- R^2 de Mac Faden de la régression logistique non ordonnée de la catégorie sociale par les indicatrices d'unités, on peut lui appliquer une correction tenant compte du grand nombre de degrés de liberté utilisé dans ce modèles :

$$H_{ajust} = 1 - \frac{-n \sum_{k=1}^m \pi_k h(k) - (m-1)(G-1)}{-n h(tot)} = 1 - \frac{n(1-H)h(tot) + (m-1)(G-1)}{n h(tot)} = H - \frac{(m-1)(G-1)}{n h(tot)}$$

- le R^2 de la régression de la variable sociale quantitative par les indicatrices d'unités.

- le R^2 ajusté correspondant : $1 - \frac{(1-R^2)(n-1)}{n-m-1}$

- Le rapport de la variance entre unités sur la variance totale dans un modèle multiniveaux à effets aléatoires.

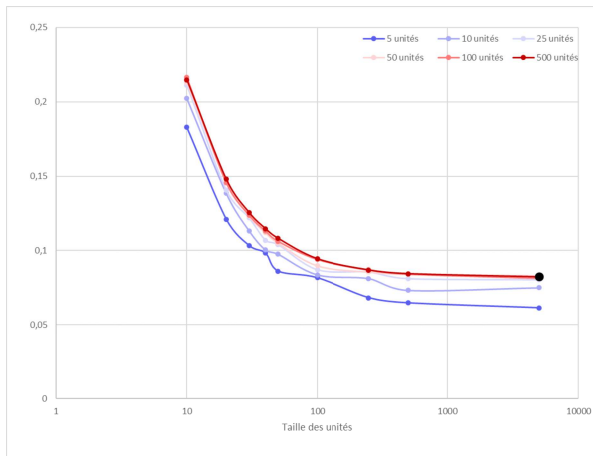
Ces cinq indicateurs ont d'abord été comparés sur des données simulées : on a retenu un coefficient de ségrégation ordonné de 20 (donc en visant un rapport entre variance entre unités et variance totale de 20 %) et une valeur nulle pour le coefficient de polarisation. On a ainsi simulé une population de 500 unités avec 5 000 individus dans chaque unité, qui sera la population de référence. Au sein de cette population, on va construire des sous-populations en tirant au hasard un certain nombre d'unités et un certain nombre d'individus par unité, ce processus étant répété plusieurs fois en particulier pour les faibles effectifs, pour tenir compte des fortes variations que cela provoque. Les cinq indicateurs ont été calculés pour chacun de ces tirages et on va étudier la moyenne à nombre d'unités et taille des unités fixés.

Les indicateurs sont d'une sensibilité assez différente aux variations de nombre d'unités et de taille d'unités (graphique 7). L'indice d'entropie est très sensible à la taille des unités : alors que la valeur de référence sur la population globale est 0,082, quand on ne retient que 10 individus par unités, l'indice d'entropie est d'environ 0,2, quel que soit le nombre d'unités retenu ; le décrochage commence à partir d'une taille de 100 individus (0,094) et s'accroît à partir de 30 (0,124). L'indice est aussi sensible au nombre d'unités, dans un sens inverse : plus ce nombre diminue, à taille d'unités donnée,

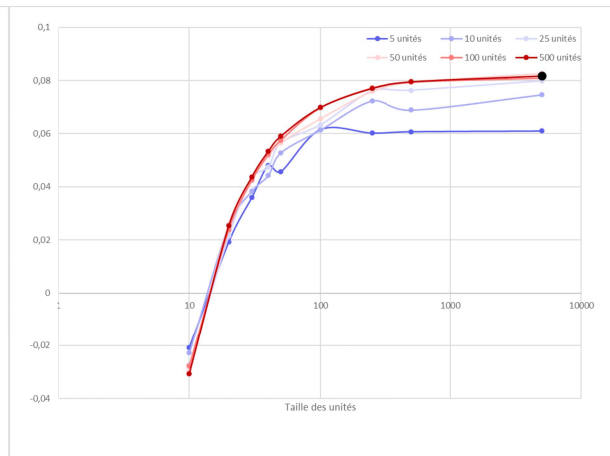
plus l'indice baisse (entre 0,02 et 0,03 point d'indice) ; faire baisser le nombre d'unités rend plus difficile, la mise en évidence de disparités entre elles.

Graphique 7 : indices de ségrégation en fonction du nombre d'unité et de leur taille

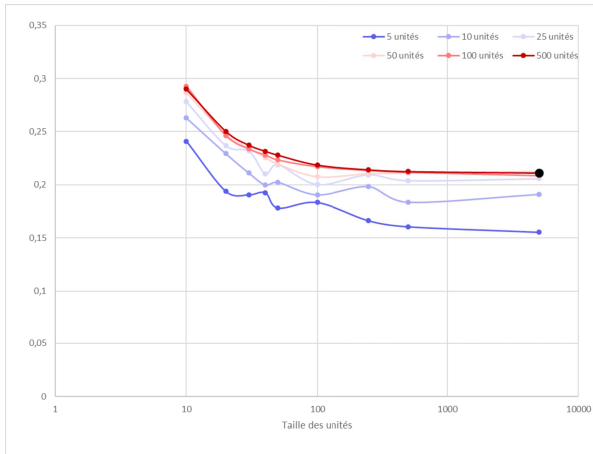
Indice d'entropie



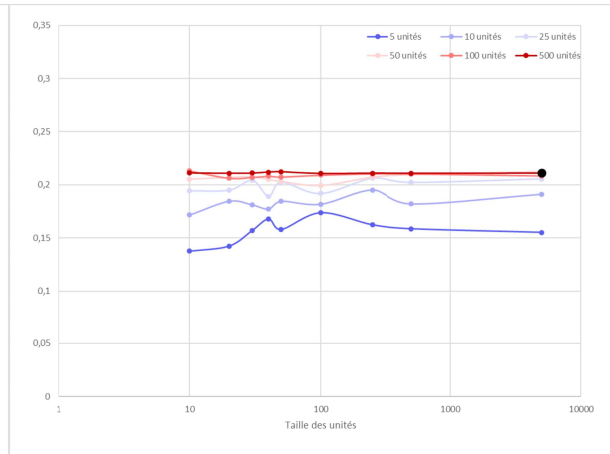
Indice d'entropie ajusté



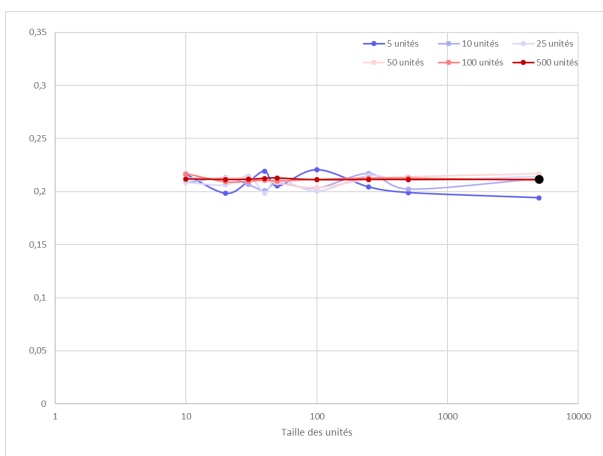
R²



R² ajusté



Part de variance inter dans un modèle à effets aléatoires



L'utilisation d'un indice dérivé du pseudo-R² ajusté ne fait qu'aggraver problème : le décrochage à partir 100 est tellement important que les valeurs deviennent négatives pour une taille d'unités de 10 individus.

Le constat est différent quand on utilise une variable quantitative pour mesurer le milieu social : le R^2 montre une tendance à augmenter pour les faibles tailles d'unités, mais elle semble moins marquée que pour l'indice d'entropie. Ainsi, pour une valeur de référence de 21,1 %, on atteint 35 % pour une taille de 10 individus par unité et un nombre d'unités de 500 ; de façon à première vue paradoxale, le résultat est plus proche de la référence avec la même taille, mais seulement 5 unités (24 %). Cela tient au fait qu'un biais inverse, intervient alors, comme pour l'indice d'entropie : à taille d'unité équivalente, passer de 500 à 5 unités fait en moyenne baisser le R^2 de 4 à 5 points. Ce biais apparaît aussi pour le R^2 ajusté. En revanche, cet indicateur semble très bien prendre en compte le problème posé par la faible taille des unités : à nombre d'unités fixés, il y a très peu de variations selon la taille des unités et pour 500 unités, par exemple, la valeur avec 10 individus par unités est quasi identique à celle de référence. Cependant, c'est l'indice issu d'une modélisation à effets aléatoires qui semble donner les meilleurs résultats : non seulement, il y a peu de variation selon la taille des unités, mais le nombre d'unités ne joue pas non plus sur les résultats.

Tableau 5 : variation des indicateurs pour un nombre d'unités et une taille d'unité fixés

Nombre d'unités	Taille des unités	Entropie	Entropie ajusté	R^2	R^2 ajusté	Effets aléatoires
5	10	0,062	0,067	0,123	0,139	0,158
5	50	0,044	0,046	0,101	0,104	0,128
5	500	0,036	0,036	0,087	0,087	0,109
10	10	0,046	0,049	0,093	0,104	0,111
10	50	0,032	0,033	0,075	0,077	0,085
10	500	0,031	0,031	0,073	0,074	0,082
25	10	0,030	0,031	0,058	0,064	0,066
25	50	0,023	0,023	0,051	0,052	0,054
25	500	0,018	0,018	0,044	0,044	0,046
100	10	0,015	0,017	0,033	0,036	0,037
100	50	0,010	0,010	0,023	0,023	0,023
100	500	0,008	0,008	0,019	0,019	0,019

Rappelons qu'il s'agit de valeurs moyennes obtenues en répétant les tirages plusieurs dizaines, voire centaines de fois, à nombre d'unités et taille d'unité fixés. Les variations autour de ces valeurs peuvent être très fortes, y compris pour les indicateurs qui se comportent le mieux en moyenne. Ainsi, pour 5 unités de 10 individus, si la valeur moyenne de l'indice d'entropie est de 0,182 (nettement plus élevée que le 0,082 de référence), l'écart-type est de 0,062. Même pour l'indice fondé sur le modèle à effets aléatoires qui donne de meilleurs résultats en moyenne (21,6 % contre 21,1 % dans la population de référence), les variations sont fortes, avec un écart-type de 15,8 points, ce qui surprendra peu les personnes familières avec la théorie des sondages (un échantillon de 5 grappes de 10 individus a peu de chance d'être précis, même s'il est sans biais).

Nous avons procédé aux mêmes analyses en partant d'un échantillon de 500 collègues ayant scolarisé au moins 3 000 élèves de 6^e sur la période étudiée et les constats sont quasi identiques à ceux qui viennent d'être présentés.

Annexe C : apport de l'analyse de données (AFC) à l'analyse de la ségrégation

Dans ce travail, nous avons surtout essayé de prolonger le lien entre indicateurs de ségrégation et mesure des corrélations, par l'utilisation de modélisations économétriques. Cependant, un autre prolongement peut être envisagé par l'utilisation des techniques d'analyses de données.

En effet, l'analyse de la ségrégation passe par l'étude du tableau croisé de l'appartenance aux unités par l'appartenance à une catégorie sociale. Or ce type de données se prête bien à la mise en œuvre d'une AFC (Analyse factorielle des correspondances), comme l'ont fait Poupeau et François (2008) pour étudier la ségrégation résidentielle, en croisant appartenance à une IRIS (regroupement de logements à des fins statistiques) et PCS.

Dans un premier temps, nous avons appliqué une AFC dans chaque département au croisement entre l'appartenance aux collèges et la catégorisation en 4 groupes d'IPS. Sur les 100 départements, l'inertie du nuage de points correspondant est très fortement corrélée (0,99) avec l'indice d'entropie : en effet, l'inertie est en fait un dérivé du Khi^2 , dont nous nous avons montré le lien avec les indices de ségrégation. Cependant, c'est l'analyse du premier axe qui donne les résultats les plus intéressants : dans tous les départements, il oppose de façon claire les catégories très défavorisées et dans une moindre mesure simplement défavorisées, aux catégories très favorisées et dans une moindre mesure favorisées. La part de cet axe dans l'inertie totale est élevée, suggérant une forte unidimensionnalité : en moyenne de 76 %, variant de 54 % à 94 %. Mais surtout, la corrélation entre cette force explicative du premier axe et le rapport entre l'indice ordonné et l'indice non ordonné (mesure du caractère ordonné de la ségrégation proposée dans le texte) est forte : 0,70 sur l'ensemble des départements, 0,88 pour les départements où le niveau de ségrégation est non négligeable (indice d'entropie supérieur à 0,02). Cela confirme, avec une méthodologie différente, le caractère très ordonné de la ségrégation entre collèges.

Dans un deuxième temps, nous avons suivi une démarche plus proche de celle de (Poupeau et François, 2008) : nous avons effectué une AFC sur le croisement entre l'appartenance aux collèges et la PCS du responsable, dans la nomenclature la plus détaillée (32 postes), sur l'ensemble du territoire français. En accord avec le principe d'une AFC, cette analyse va avoir deux intérêts :

- classer les PCS en fonction de leur tendance à se trouver dans les mêmes établissements ;
- classer les collèges en fonction de leur tendance à accueillir des élèves appartenant aux mêmes PCS.

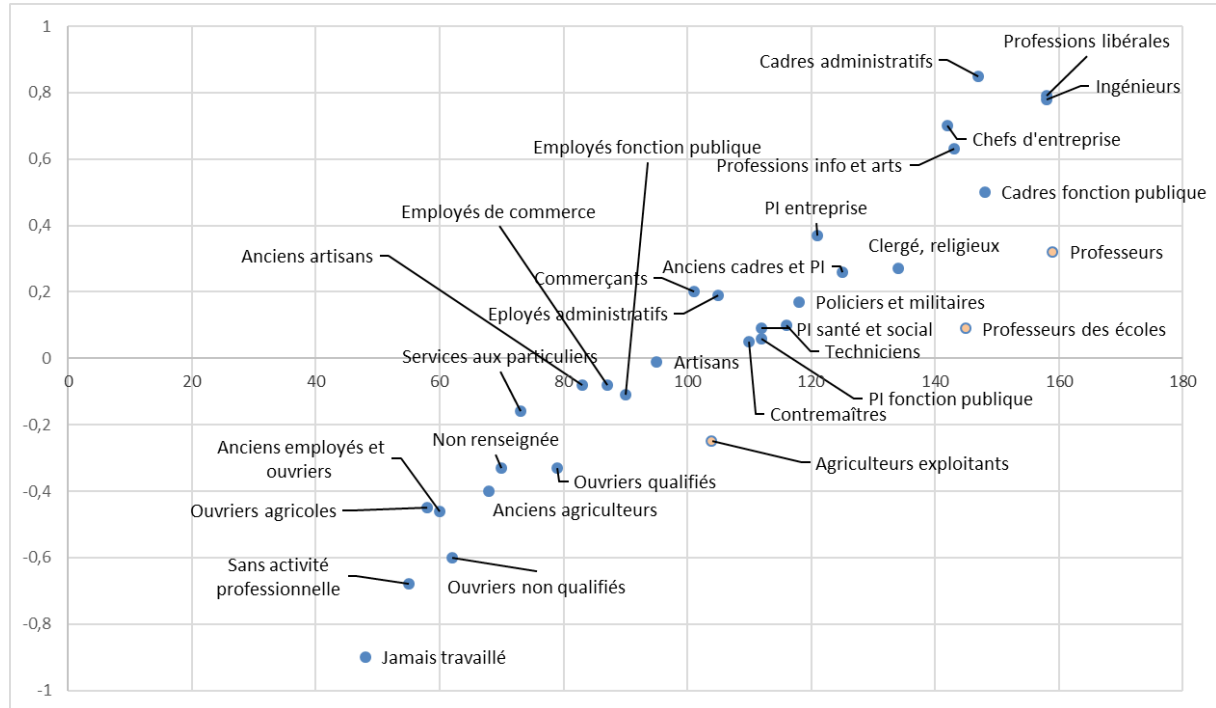
Le premier objectif va fournir une vision alternative au travail ayant conduit à l'IPS, pour établir des rapprochements entre PCS. Le deuxième objectif renvoie plus directement aux problématiques de ségrégation.

Le premier axe de l'AFC rend compte de 32 % de l'inertie totale et le second de 12 %. La prépondérance du premier axe est moins forte que dans la première analyse, car les données à analyser sont plus complexes (32 PCS contre 4 groupes d'IPS) : le comportement particulier de certaines catégories sociales en matière de scolarisation donne lieu à des axes secondaires (parfois parce que ces catégories sont peu fréquentes, phénomène fréquent dans une AFC du fait de l'usage d'une distance du Khi^2). Ceci dit, la confrontation entre les coordonnées des PCS sur le premier axe et l'indice de position sociale donne une interprétation très claire de ce premier comme mesure de la hiérarchie sociale (graphique 8) : la corrélation entre les deux séries est très forte (0,93) et les cas qui s'écartent le plus de la tendance s'expliquent assez facilement :

- Les agriculteurs, dont on peut effectivement penser que la répartition par établissement suit une logique particulière, plus géographique que sociale ; d'ailleurs le deuxième axe est très fortement lié à cette catégorie sociale (s'opposant à des catégories un peu floues : PCS non renseignées, chômeurs n'ayant jamais travaillé, sans activité professionnelle).

- Les enfants d'enseignants, qui sont moins bien classés avec la méthode AFC que par l'IPS : en effet, les autres catégories favorisées se regroupent nettement, car elles ont toutes une certaine tendance à aller dans des établissements privés (où il n'y aura pas d'élèves défavorisés, ce qui crée une forte cohérence sur le premier axe). C'est moins net pour les enseignants, qui recourent peu au privé.

Graphique 8 : lien entre l'IPS et le premier axe de l'AFC



Il est assez frappant de constater que le classement des PCS en fonction de leur tendance à aller dans les mêmes établissements scolaires est très proche du classement obtenu sur des critères liés aux conditions de vie et aux pratiques culturelles. On est tenté d'y voir une double validation : d'une part, la confirmation, après les travaux sur l'IPS, d'une forte segmentation et même hiérarchie des classes sociales en France ; d'autre part, l'importance de la ségrégation scolaire (sans doute en partie d'origine résidentielle) et le fait que cette ségrégation est fortement ordonnée.

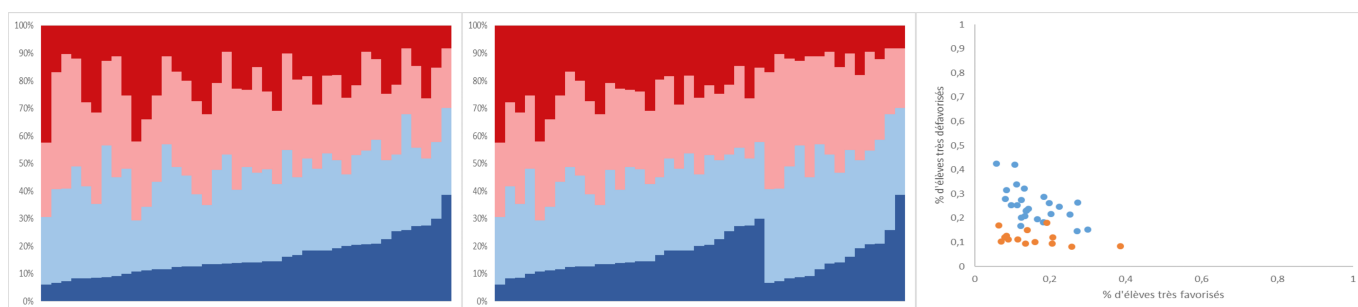
Le deuxième objectif de l'AFC est justement de classer les collèges sur la même structure factorielle. Les coordonnées des collèges sur le premier axe peuvent donc fournir une image de leur caractère plus ou moins favorisé et cela sans avoir émis d'hypothèse préalable sur la hiérarchisation des PCS. Cette piste pourrait être explorée pour proposer des indicateurs de ségrégation tenant compte d'une description très fine du milieu social des élèves.

Annexe D : illustration de la ségrégation dans quelques départements.

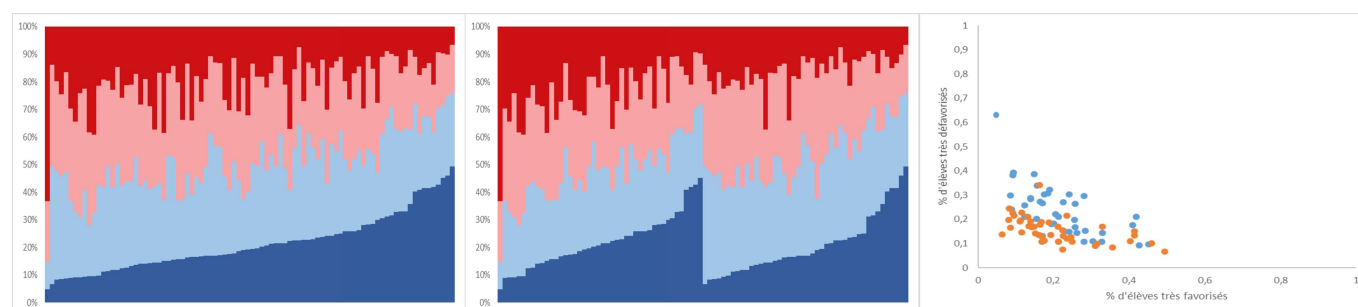
En complément des exemples de Paris et des Hauts-de-Seine, détaillés dans le texte, nous allons présenter la situation de quelques autres départements, en proposant 4 cas de ségrégations peu ordonnées et 4 cas de ségrégations ordonnées. On donnera à chaque fois l'histogramme de répartition des collèges par catégorie sociale, triés par ordre croissant de proportion d'élèves très favorisés ; le même graphique en distinguant public et privé ; le croisement entre la proportion d'élèves très favorisés dans un collège et la proportion d'élèves très défavorisés (collèges publics en bleu, collèges privés en orange).

Graphique 9 : profil social des collèges dans 8 départements

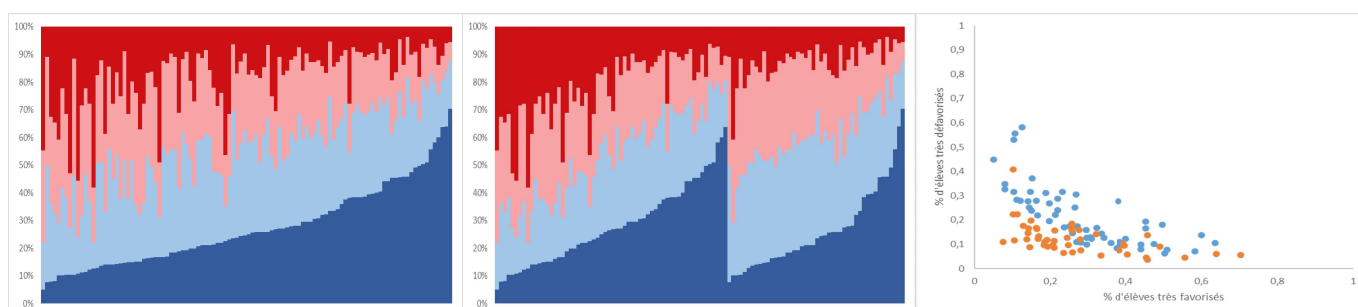
Mayenne (53)



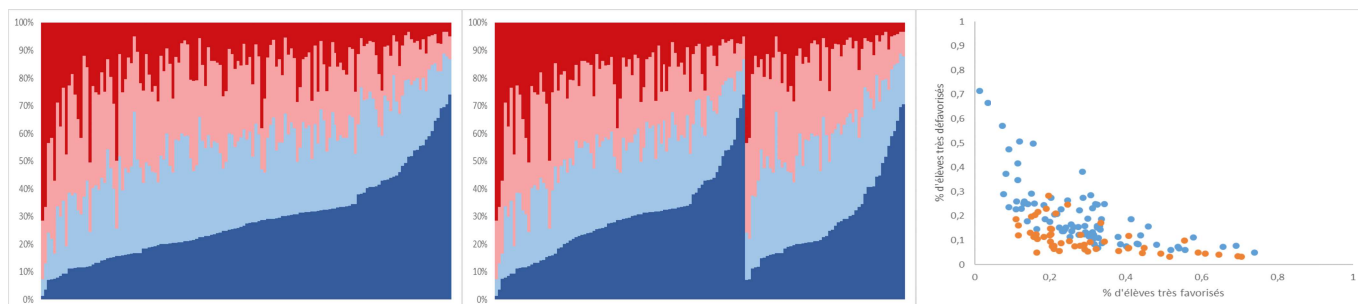
Morbihan (56)



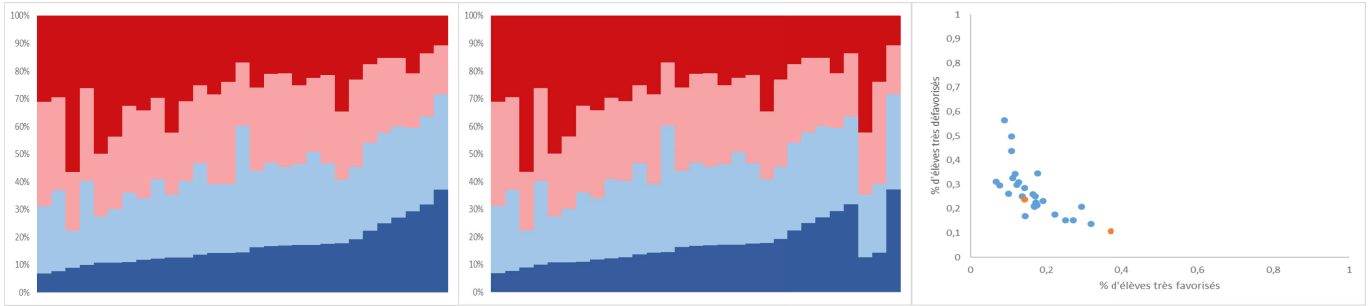
Ille-et-Vilaine (35)



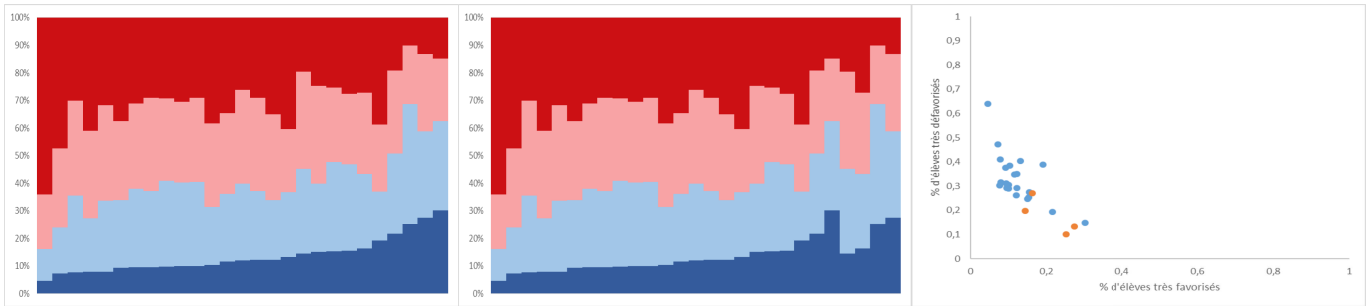
Loire-Atlantique (44)



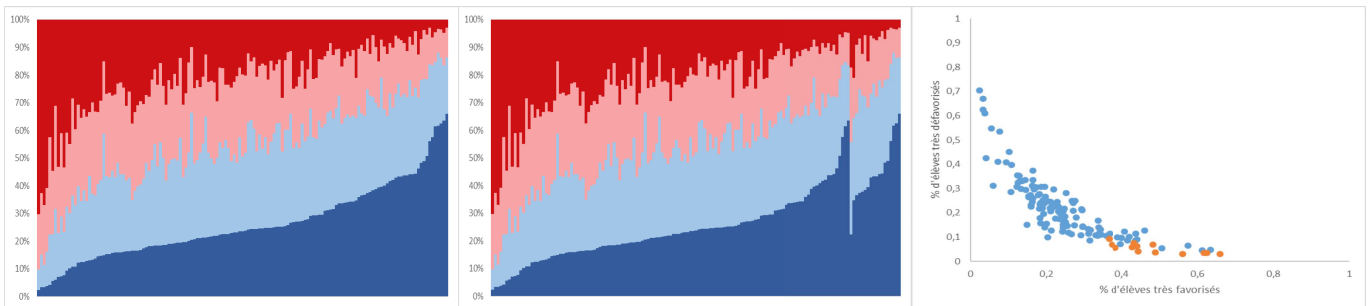
Cher (18)



Département Haute-Marne (52)



Seine-et-Marne (77)



Nord (59)

