

Les inégalités de compétences à la fin des études

Fabrice Murat

Annexe en ligne / Online Appendix

Quelques précisions méthodologiques sur la mesure des inégalités scolaires

Modélisations alternatives du niveau d'études

La modélisation du niveau d'études s'est fondée sur une quantification par les écarts de compétences. Ce choix peut conduire à une convergence un peu artificielle entre le modèle sur le niveau d'études et celui sur les compétences. C'est pourquoi, il paraît utile d'envisager des méthodes alternatives. L'ensemble des résultats concernant ces modélisations est donné dans un fichier en complément de l'article.

Une première possibilité, assez naturelle, consiste à attribuer à chaque niveau d'études le nombre d'années nécessaires pour l'obtenir. Dans un premier temps, ce sont les moyennes observées (la dernière colonne du tableau 2) qui ont été utilisées. La modélisation de cette nouvelle quantification du niveau d'études par les mêmes variables que dans le tableau 3 donne un R^2 légèrement inférieur à la quantification par les scores moyens (26.4 % contre 27.5 %). La comparaison entre ce modèle et celui sur l'âge de fin d'études effectivement observée peut être intéressante : elle indiquera les catégories pour lesquelles les études ont duré plus ou moins que ce que le niveau d'études atteint laisse attendre. La fréquence plus grande des redoublements dans certaines catégories est une première explication. Un début de scolarité plus tardif peut aussi être en jeu : cela va être notamment le cas pour les personnes nées à l'étranger qui ont eu des études plus longues que les personnes nées en France (0.73 an les séparent en termes d'année d'études observées contre 0.23 pour le nombre d'années d'études prédit).

Pour comparer cette modélisation avec celle utilisant la quantification par les scores moyens, la quantification par l'âge de fin d'études a été standardisée pour avoir la même variance que le score global. Les coefficients sont alors très proches avec ceux du tableau 3. Si l'on exclut les catégories peu représentées et particulières signalant l'absence d'information sur le diplôme ou la profession des parents, l'écart le plus grand concerne les personnes dont le père avait un diplôme du secondaire (baccalauréat, CAP ou BEP) : le coefficient est de -0.15 dans la quantification par les scores moyens contre -0.10 dans la quantification par l'âge de fin d'études. Dans les deux cas, le coefficient est non significativement différent de 0. Nous avons aussi utilisé une modélisation par un âge de fin d'études a priori (24 ans pour un niveau Bac+5, 21 ans pour un Bac+3/4, 19 ans pour un Bac+2, 17 ans pour un BAC, 15 pour un CAP ou un arrêt au mieux en 2nde GT ; un an supplémentaire est accordé quand le diplôme est réussi). Les résultats sont très proches du modèle avec l'âge de fin d'études observé (après standardisation sur la variance du score global, les écarts entre les coefficients des deux modèles sont inférieurs à 0.05).

Une autre possibilité est de se rapprocher de la méthode de Place & Vincent (2009) en conservant le caractère qualitatif du niveau d'études, avec un aspect hiérarchisé. L'ordre du tableau 2 donne la hiérarchie des niveaux d'études, à ceci près que les trois baccalauréats ont été regroupés. En effet, le baccalauréat général pourrait être considéré comme plus élevé dans une échelle scolaire, car il permet plus facilement de poursuivre des études supérieures et donc de progresser dans la hiérarchie des diplômes. Mais ici, cette catégorie regroupe les jeunes qui n'ont justement pas fait d'études après ce diplôme. Il paraît peu pertinent de leur attribuer un niveau d'études plus élevé qu'aux jeunes qui ont fait le choix d'un baccalauréat professionnel sans aller plus loin. Une régression logistique polytomique ordonnée a ensuite été effectuée sur cette variable en fonction des données du tableau 3. Pour rendre les coefficients comparables, on a tenu compte du fait que par construction, la propension à obtenir un haut diplôme sous-jacente au modèle a une variance égale à la somme de variance de la propension prédite par les variables explicatives (que l'on peut calculer en sortie de la régression) et de la variance du terme d'erreur, qui pour une régression logistique vaut $\pi^2/3$. On a recalé cette variance sur celle du score global. Il y a un peu plus de divergences entre ce modèle et celui du tableau 3, mais globalement, pour la plupart des variables, les coefficients des deux modèles sont proches : les divergences sont inférieures à 10 points d'écart-type.

Confrontation des inégalités de compétences et des inégalités de niveau d'études

Revenant à une quantification du niveau d'études par le score moyen, une autre question à clarifier est la confrontation des deux modèles. Dans le texte, nous avons comparé les coefficients, après avoir standardisé les variances. Même si nous avons surtout mis en avant les divergences les plus fortes, cette analyse revient à peu près à étudier un modèle qui ferait la régression de la différence entre le score global et le niveau d'études quantifié par les variables explicatives, ce qui peut sembler un peu délicat.

Une autre approche est possible en reproduisant les modèles du tableau 3 pour chaque indicateur de réussite scolaire, en ajoutant l'autre indicateur dans la régression : d'une part, le score en fonction des caractéristiques du jeune et du niveau d'études quantifié, d'autre part le niveau d'études quantifié en fonction du milieu social et du score.

Les inégalités de compétences à la fin des études

Fabrice Murat

Annexe en ligne / Online Appendix

Le premier intérêt de cette démarche est de juger globalement dans quelle mesure les inégalités sur les deux indicateurs de réussite sont redondantes. En effet, même si le modèle concernant les compétences et celui concernant le niveau d'études divergent sur quelques points (le sexe ou le pays de naissance notamment), ils semblent très proches sur des variables déterminantes comme le diplôme et la profession des parents. C'est cohérent avec la nette corrélation entre les deux indicateurs de réussite scolaire ($R^2=33\%$). On pourrait imaginer que les inégalités de compétences expliquent entièrement les inégalités de niveau d'études : dans ce cas, la prise en compte des caractéristiques du jeune n'apporterait rien à l'étude du niveau d'études, si l'on contrôle les résultats aux évaluations et le R^2 de ce modèle serait de 33 % (corrélation simple entre les 2 indicateurs de réussite). À l'inverse, on pourrait imaginer le cas, assez improbable, où l'information apportée par les caractéristiques du jeune serait complètement indépendante de celle apportée par les compétences : le R^2 du modèle global serait de $33+27.5=60.5\%$ (27.5 % correspondant au R^2 du modèle du niveau d'études par les caractéristiques du jeune). En fait, on se trouve dans une situation intermédiaire avec un R^2 de 44.7 %, soit un gain de 11.7 % par rapport au modèle avec uniquement le score (mais inférieur aux 27.5 % du lien entre caractéristiques du jeune et niveau d'études). À l'inverse, le modèle global (niveau d'études + caractéristiques du jeune) explique 39 % de la variance du score, soit 6 % de plus par rapport au modèle sans les caractéristiques du jeune (mais un gain moindre que les 19.7 % du lien entre caractéristiques du jeune et score).

Les inégalités de compétences et celles de niveau d'études se recoupent donc bien en grande partie, mais il subsiste des écarts importants pour chacun des deux indicateurs quand l'autre est contrôlé. Les coefficients du modèle permettent de préciser ce résultat, en confirmant des résultats donnés dans le texte : le coefficient associé au fait d'avoir vécu avec ses deux parents est positif dans le modèle du score et négatif dans le modèle du niveau d'études, ce qui rend compte du fait que, dans les modèles du tableau 3, il n'y avait pas d'écart selon cette catégorie pour le score, mais un coefficient négatif pour le niveau d'études quantifié. Les deux indicateurs de réussite scolaire donnent une image un peu différente de leur situation scolaire. Pour d'autres variables, les modèles globaux sont plus compliqués à interpréter. Ainsi par rapport aux personnes dont les parents sont diplômés du supérieur, ceux dont les parents n'ont aucun diplôme ont des coefficients négatifs dans les deux modèles. Cela tient sans doute au fait que ces individus ont des caractéristiques inobservées, dont certaines, à compétences données, jouent négativement sur le niveau d'études (par exemple un niveau moindre d'aspiration des parents) et d'autres qui, à niveau d'études fixé, sont liées négativement avec les compétences (il existe une hétérogénéité à un niveau d'études donné, probablement liée aux caractéristiques du jeune).