



HAL
open science

L'effet de l'homogamie éducative sur les inégalités : une note méthodologique avec une illustration pour le cas français

Pierre Courtioux, Vincent Lignon

► To cite this version:

Pierre Courtioux, Vincent Lignon. L'effet de l'homogamie éducative sur les inégalités : une note méthodologique avec une illustration pour le cas français. 2015. halshs-01254789

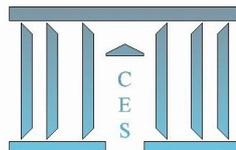
HAL Id: halshs-01254789

<https://shs.hal.science/halshs-01254789v1>

Submitted on 12 Jan 2016

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

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



**L'effet de l'homogamie éducative sur les inégalités :
une note méthodologique avec une illustration
pour le cas français**

Pierre COURTIoux, Vincent LIGNON

2015.91



L'effet de l'homogamie éducative sur les inégalités : une note méthodologique avec une illustration pour le cas français

Pierre Courtioux

(EDHEC Business School, Centre d'Economie de la Sorbonne)

Vincent Lignon

(CNAF, Centre d'Economie de la Sorbonne)

Résumé

Dans la lignée de travaux économiques récents, nous cherchons à quantifier l'effet de l'homogamie (*assortative mating*) et plus particulièrement de l'homogamie éducative (*educational homogamy*) sur les inégalités de niveau de vie mesurées par l'indice de Gini. Nos résultats montrent que dans le cas français, l'effet est faible. Nous montrons également que l'utilisation d'une variable éducation peu détaillée conduit à sous-estimer cet effet. Selon la méthode retenue, on peut imputer à l'homogamie une augmentation des inégalités de l'ordre de 0,8 à 2,5%, c'est-à-dire à un impact beaucoup plus faible que l'effet du système socio-fiscal français qui diminue les inégalités d'environ 16%.

Mots clés : homogamie ; inégalité ; éducation

Abstract

Following recent economic literature on the subject, the article estimates the impact of assortative mating and more particularly of educational homogamy on inequality measured by the Gini index of equivalized income. The results show that in the French case, this impact is small. They also show that an education variable has to be detailed enough in order to avoid an under-estimation of educational homogamy. Depending on the method used the impact of educational homogamy on inequality range from 0.8 to 2.5%, i.e. a much lower level than the French tax and benefit system which decreases inequality by almost 16%.

Key words : assortative mating ; inequality ; education

JEL classification: D31, J12, I2.

Le modèle de microsimulation GAMEO utilisé dans cet article est développé par le Pôle économie de l'EDHEC Business School, Stéphane Gregoir et Dede Houeto ont également contribué au développement du modèle. Vincent Lignon a effectué ce travail alors qu'il était doctorant au Centre d'Economie de la Sorbonne et à l'Ined. Des versions préliminaires de cet article ont fait l'objet d'une présentation au Séminaire Economie des Institutions (Centre d'Economie de la Sorbonne) et au Séminaire Café de l'Ined. Les auteurs remercient particulièrement pour leurs remarques et leurs commentaires : Nicolas Frémeaux, Dominique Meurs, Marion Leturcq, Delphine Remillon, Laurent Toulemon, ainsi que les participants à ces séminaires.

Introduction

Les effets de l'homogamie (*assortative mating*) et plus particulièrement de l'homogamie éducative sur les inégalités ont fait l'objet de récentes investigations. Les travaux disponibles sur le sujet portent principalement sur le cas des Etats-Unis (Greenwood *et al.* [2014], Eika *et al.* [2014], Hryshko *et al.* [2014], Harmenberg [2014]) de la Norvège (Eika *et al.* [2014]) et plus récemment de la France (Courtioux et Lignon [2015], Frémeaux et Lefranc [2015]). Les résultats obtenus diffèrent selon les pays mais également selon les méthodes utilisées.

En suivant la distinction d'Harmenberg [2014] il est possible de distinguer deux types de méthodologie dans la construction du contrefactuel d'une population dans laquelle il n'y aurait pas d'homogamie éducative. La première méthode, qu'il appelle *méthode additive* (ou *méthode comptable*) consiste à considérer que les revenus d'un individu donné sont totalement indépendants de ceux de son conjoint. Avec cette méthode la construction d'un contrefactuel passe par *l'appariement aléatoire* des individus de sexe opposé pour composer des couples contrefactuels. Les individus conservent leurs propres caractéristiques productives (leur revenu individuel reste constant) ; le revenu des couples du contrefactuel est alors obtenu en sommant les revenus des deux individus qui le composent. La deuxième méthode, qu'il appelle *méthode par imputation*, considère que le revenu observé du ménage dépend de diverses caractéristiques (par exemple la combinaison de diplômes des deux conjoints, le nombre d'enfants, etc.). Le contrefactuel est obtenu sur la base d'hypothèses d'identification (au moins implicites) des déterminants des revenus du ménage et une modification de la distribution de ces déterminants dans la population étudiée. En économie, on considère généralement que la première méthode ne tient pas compte du fait que l'offre de travail (au moins *via* son résultat observé, le salaire individuel) n'est pas exogène à la formation du couple.

Les résultats obtenus par Greenwood *et al.* [2014] à l'aide d'une méthode par imputation concluent qu'aux Etats-Unis, l'homogamie éducative contribuerait à augmenter les inégalités mesurées par l'indice de Gini de 26%. Pour ce même pays, les résultats obtenus sur la base de méthodes comptables sont beaucoup plus faibles : l'homogamie n'augmenterait les inégalités que de 2 à 3% (Hryshko *et al.* [2014]). Cependant, toujours pour les Etats-Unis et sur la base d'une méthode par imputation, Eika *et al.* [2014] ainsi qu'Harmenberg [2014] trouvent des résultats de l'ordre de 3 à 4%, c'est-à-dire beaucoup plus faibles que ceux de Greenwood *et al.* [2014]. Dans le cas Français, les premiers résultats ne traitent pas des inégalités de niveau de vie mais des inégalités de revenus : Courtioux et Lignon [2015] sur la base d'un modèle de microsimulation dynamique¹ estiment l'impact de l'homogamie sur les inégalités de revenus inter-temporels à 5% ; Frémeaux et Lefranc [2015] qui testent plusieurs méthodes de construction du contrefactuel estiment que cet effet est de l'ordre de 3 à 10% selon le type de méthode utilisé.

L'objectif de cet article est d'analyser le cas français en nous concentrant sur un indicateur standardisé des inégalités : l'indice de Gini des niveaux de vie (*equivalized income*). Par ailleurs, nous cherchons plus particulièrement à documenter la dimension *éducative* de l'homogamie. En effet, si la dimension éducative n'est pas nécessaire pour l'utilisation des méthodes comptables, c'est généralement cette dimension qui est privilégiée dans la construction des différentes variantes des méthodes par imputation. Or le choix de la variable éducative n'est pas sans implication sur la caractérisation du niveau d'homogamie éducative : en France par exemple, à même niveau de diplôme, les diplômés d'un master professionnel à l'Université et les diplômés d'écoles de commerce

¹ Ce modèle simule conjointement la carrière salariale et la mise en couple. De ce point de vue, ils relèvent plutôt des méthodes d'imputation.

n'ont pas la même propension à l'homogamie; les diplômés d'écoles de commerce sont plus homogames et ont en moyenne de meilleurs salaires, toutes choses égales par ailleurs (Courtioux et Lignon [2014]).

Pour analyser plus en détail l'effet de l'homogamie éducative sur les inégalités dans le cas français, nous commençons par discuter des avantages et des limites des deux méthodes que nous présentons dans leur version la plus simple ; nous discutons également de la mesure dans laquelle des résultats obtenus à partir d'un modèle de microsimulation dynamique permettent de compléter ces méthodes. Puis dans un second temps, nous présentons la base de données que nous utilisons. Enfin, dans une dernière partie, nous présentons et discutons les résultats obtenus.

I. Mesurer l'ampleur de l'effet propre à l'homogamie éducative

Pour mesurer l'ampleur de l'effet propre à l'homogamie éducative sur une population donnée, il convient de construire une population contrefactuelle dans laquelle les unions seraient aléatoires. C'est dans la structuration de cet aléa et dans ses implications sous-jacentes que les méthodes d'estimation diffèrent. En suivant Harmenberg [2014], on peut introduire une première distinction entre *méthode comptable* et *méthode par imputation*.

La méthode comptable

La méthode comptable est certainement la méthode la plus intuitive pour construire une population contrefactuelle. Elle nécessite d'identifier la contribution marginale du conjoint au revenu du couple, c'est-à-dire au sein de chaque ménage observé composé d'un couple, le revenu de l'homme (R_h) et le revenu de la femme (R_f).

Dans sa version la plus simple, pour chacun des hommes qui composent les N couples observés, on construit N ménages contrefactuels avec les N femmes qui composent les N couples observés (soit une population contrefactuelle composée de N^2 couples). La pondération de chacun de ces couples contrefactuels (w_c) vérifie $w_c = N \times p_c$ avec $p_c = w_{ih}/N \times w_{if}/N$ et où w_{ih} correspond à la pondération du ménage observé d'où est issu l'homme et w_{if} celle du ménage observé d'où est issue la femme du ménage contrefactuel. Le revenu de chacun des couples (R_c) est obtenu en sommant les revenus de l'homme et ceux de la femme pour la population observée comme pour la population contrefactuelle : $R_c = R_h + R_f$. Ces couples contrefactuels peuvent être ajoutés aux autres ménages observés (célibataires, ménages complexes, ménages avec couple d'individus de même sexe) et constituer ainsi la population contrefactuelle totale sur laquelle il est possible de calculer des indicateurs d'inégalité de revenu. Par construction, cette méthode aboutit à ce que nous pouvons qualifier d'un *contrôle complet de l'homogamie en coupe au sein de la population*. Au-delà de son aspect formel, ce contrôle complet pose cinq limites importantes à l'analyse des inégalités qu'il convient de discuter : (i) l'interprétation du contrôle de l'homogamie générationnelle au sein d'une population, (ii) l'absence de cohérence budgétaire des couples contrefactuels, (iii) l'absence de cohérence familiale des couples contrefactuels, (iv) l'existence d'un potentiel biais statistique à la hausse du Gini de la population contrefactuelle, (v) l'absence d'ajustement des comportements individuels à une mise en couple aléatoire.

(i) Le contrôle de l'homogamie générationnelle au sein d'une population est une conséquence du contrôle complet de l'homogamie en coupe qui caractérise la méthode comptable. Par construction la population contrefactuelle apparie des individus observés issus de « jeunes couples » avec des individus observés issus de « couples plus vieux ». Cette méthode revient à supposer qu'il existe un marché des mariages *neutre* c'est-à-dire un espace d'échange homogène dans lequel l'âge ou la génération de l'individu n'a pas d'importance pour la mise en couple. De ce point de vue et de manière plus générale, la population contrefactuelle a une cohérence en termes de logique combinatoire qui reste difficile à interpréter d'un point de vue sociologique ou économique : on ne sait pas comment les individus peuvent se rencontrer pour former un couple sur ce marché des mariages.

(ii) L'absence de cohérence budgétaire des couples contrefactuels existe dès lors que les revenus individuels analysés (R_h et R_f) ne sont pas strictement individuels. Dans l'absolu, cette méthode va poser problème pour analyser le revenu disponible, c'est-à-dire un revenu composé de revenus primaires et d'un ensemble de taxes et de transferts. En effet, certains impôts et certains transferts pris en compte ne dépendent pas directement des revenus du couple mais de la situation familiale ou des revenus du ménage (impôt sur le revenu, prestations familiales, etc.). Il convient de noter que cette limite s'étend aussi dans une certaine mesure aux revenus directs et indirects tirés du marché du travail : pour les fonctionnaires, le supplément familial de traitement (SFT)² même versé sur un traitement individuel correspond à une situation familiale particulière ; de même, pour les retraités, la partie des pensions correspondant à la réversion (par exemple pour les caisses complémentaires) n'est pas directement imputable à l'individu mais correspond à des droits relevant de sa trajectoire conjugale passée.

(iii) L'absence de cohérence familiale des couples contrefactuels complète par certains aspects les limites (i) et (ii). Même dans un cas où la contribution marginale de l'individu au revenu du couple serait identifiable, il reste encore à imputer au couple contrefactuel une structure familiale. En effet, cette structure familiale entre directement en compte dans le calcul du niveau de vie (*equivalized income*) qui applique au revenu disponible une échelle d'équivalence dépendant de la taille du ménage pour rendre compte des économies d'échelle. De ce point de vue, il est possible d'utiliser comme revenu individuel le niveau de vie (E_h et E_f). Ceci revient à poser l'hypothèse que la structure du couple contrefactuel correspond à une moyenne de la taille des ménages observés dont sont issus chacun de ses membres. Si T_h et T_f représentent respectivement la taille des ménages dont sont issus l'homme et la femme du couple contrefactuels dont le revenu disponible sont respectivement I_h et I_f alors $R_c = I_h/T_h + I_f/T_f$. Rien ne garantit donc que la structure familiale sous-jacente au couple contrefactuel soit cohérente.

² Pour mémoire, le SFT est une prime, croissante avec le nombre d'enfants, ajoutée au traitement du fonctionnaire ; dans le cas où les deux conjoints sont fonctionnaire seul l'un des deux touche le SFT.

(iv) Il convient également de noter un problème purement statistique lié à l'utilisation de l'indice de Gini avec la méthode comptable. Le passage de N à N^2 observations fait que la distribution de revenu de la population contrefactuelle et de la population observée ne sont pas directement comparables : le degré de précision (mesuré par le nombre d'observations) du contrefactuel est beaucoup plus important, ce qui risque de biaiser à la hausse le Gini.

(v) Enfin, le mode d'identification de l'effet de l'homogamie sur les inégalités ne constitue pas une « expérience naturelle ». Une mise en couple aléatoire (imposée par exemple dans le cadre d'un régime totalitaire par un planificateur central) ne garantit pas une stabilité des comportements individuels. De ce point de vue, les effets mesurés correspondent à un « effet du lendemain », c'est-à-dire avant ajustement des comportements individuels.

Les cinq points énoncés limitent-ils l'utilisation de la méthode comptable pour analyser les inégalités de niveau de vie ? On commencera par noter que le point (iv) est une limite purement statistique qu'il est facile de surmonter : l'utilisation d'une technique de *bootstrap* comme celle présentée dans l'annexe 1 permet d'estimer l'effet propre du nombre d'observations sur le coefficient de Gini et donc de présenter des indicateurs comparables pour la population observée et la population contrefactuelle.

L'absence de cohérence budgétaire (ii) et de cohérence familiale (iii) est un problème plus compliqué à résoudre pour l'analyse des inégalités de niveaux de vie. En effet, pour attribuer un revenu disponible propre à chacun des individus qui constitue le ménage observé, il est nécessaire de fixer une convention d'attribution des différents transferts et taxes qui portent sur les revenus du ménage (impôt sur le revenu, taxe d'habitation, allocations familiales, etc.) mais également des enfants. Dans le cas français, si un mode de répartition des taxes et transferts pour chacun des conjoints est envisageable (sur la base par exemple du taux d'effort moyen du ménage), la reconstitution du revenu des ménages contrefactuels est problématique dans la mesure où rien ne garantit que ce revenu soit cohérent avec la législation socio-fiscale en vigueur ou même qu'il soit effectivement possible (par exemple pour le législateur) de proposer un système socio-fiscal qui permette d'aboutir à ce contrefactuel³. On peut également faire remarquer que certains revenus individuels comme les retraites (notamment les pensions de réversion) constituent un revenu individuel qui n'est pas exempt de l'histoire conjugale et donc indirectement de l'homogamie éducative.

De plus, ce mode d'appariement de la méthode comptable ne permet pas de réaffecter les enfants des ménages observés aux ménages contrefactuels sans hypothèses complémentaires. La manière la plus intuitive évoquée précédemment consiste à faire transiter l'effet propre des enfants *via* le revenu que l'on affecte à chacun des conjoints lors de l'appariement aléatoire (c'est-à-dire en utilisant le niveau de vie et pas le revenu disponible comme mesure de la contribution marginale à la production de « bien-être » du ménage des individus composant le couple). Là encore, la reconstitution des ménages contrefactuels lors de l'appariement aléatoire ne garantit pas la « cohérence familiale » du couple contrefactuel. La structure familiale du ménage contrefactuel est

³ Ceci veut dire que la mesure de l'effet de l'homogamie éducative par cette méthode ne permet pas d'éclairer le législateur sur les mesures permettant la correction des inégalités mesurées.

alors par nature composite : chacun des individus observés apportant au ménage du couple contrefactuel la moitié des enfants de son ménage originel.

Pour essayer de dépasser la limite (i), le principe de base de la *méthode comptable* peut être amendé : on peut par exemple décider que l'appariement aléatoire est structuré par âge afin d'éviter que par construction les ménages contrefactuels soient constitués de cas peu probables d'individus « très jeunes » en couple avec des individus « très vieux ». Pour ce faire, on peut décider par exemple, pour un individu observé donné de corriger le poids des couples contrefactuel par un coefficient correspondant à une loi de répartition des différences d'âge au sein des couples qui composent la population observée. Au-delà des effets d'homogamie d'âge (c'est-à-dire la tendance à se mettre en couple avec des individus d'âge proche) ce mode de structuration du contrefactuel peut permettre de contrôler des effets démographiques de long terme (comme la tendance à l'accroissement du niveau moyen d'éducation). Néanmoins, cette structuration des ménages contrefactuels pose l'hypothèse implicite que le « marché » ou « l'espace » des mariages est structuré par l'âge, or on peut considérer que les choix des conjoints dont l'âge observé est une caractéristique observable n'est que la résultante d'un espace de choix contraint par d'autres variables non observées (par exemple, la priorité peut être de se marier dans « sa classe sociale » avec quelqu'un ayant de « bonnes perspectives de carrière »). Dans ce cas, l'amendement de la méthode comptable *a priori* purement « combinatoire » nécessite de poser des hypothèses complémentaires fortes sur les comportements sur le marché des mariages et la manière dont se forment les couples⁴. Or comme l'indique la limite (v), cette méthode ne prend pas en compte les ajustements de comportement qui pourraient résulter d'une « véritable expérience d'appariement aléatoire ». De manière générale en économie, on reproche à la méthode comptable de ne pas tenir compte des ajustements de comportements des conjoints sur le marché du travail : si les individus se mettent en couple avec un nouveau partenaire, leur répartition entre travail salarié et travail domestique se réajustera selon leur avantage comparatif pour ces différents types de tâche au sein du couple. Un diplômé qui quitte un conjoint diplômé pour un conjoint moins diplômé aura un avantage comparatif plus important à réduire son travail domestique et augmenter son offre de travail salarié. Cette critique qui relève de la théorie du capital humain appliqué au sein du couple, peut être étendue à un ensemble d'autres comportements (potentiellement endogènes à l'offre de travail) : la probabilité de séparation du couple, le calendrier des naissances, voire les choix éducatifs dans la mesure où ils sont eux-mêmes les éléments d'une stratégie matrimoniale.

De ce point de vue, il nous semble que l'intérêt principal de la méthode comptable pour construire une population contrefactuelle réside dans son caractère purement combinatoire et non expérimental. Cette méthode est plus adaptée pour les revenus individuels des personnes n'ayant pas encore subi des forts taux de mortalité. C'est pourquoi en général, les travaux qui utilisent cette méthode se concentrent sur les revenus tirés du marché du travail et limitent le champ de l'analyse à des ménages d'âge actif⁵.

⁴ De manière très simple, ces hypothèses peuvent par exemple correspondre dans une approche sociologique aux enchainements des différents âges de la vie et la place que prend la mise en couple dans cet enchainement.

⁵ C'est ce que nous faisons dans la suite de ce travail, cf. section II.

La méthode par imputation

La méthode par imputation est souvent présentée comme une manière de dépasser l'absence de prise en compte de l'ajustement de l'offre de travail à une mise en couple aléatoire (cf. *infra* limite (v)). Nous présentons la méthode et discutons de la mesure dans laquelle elle permet de dépasser l'ensemble des limites de la méthode comptable.

La méthode par imputation, contrairement à la méthode comptable, consiste à construire la population de couples contrefactuels, sur la base des couples observés, en appliquant un coefficient de correction à leur poids dans la population. Pour être mise en œuvre, elle nécessite une variable de revenu au niveau ménage (R_m) – la contribution marginale de chacun des individus composant le couple n'est pas une information nécessaire -, ainsi qu'une variable qualitative de contrôle de la structure d'homogamie - par exemple le niveau d'éducation - renseigné pour chacun des individus composant le couple (E_h et E_f). Le mode de calcul du coefficient de correction des pondérations est expliqué en détail dans Greenwood *et al.* [2014]⁶. Dans ce cadre, construire une population contrefactuelle de couples exempte d'homogamie nécessite de s'appuyer sur les deux variables qui servent à contrôler l'homogamie (E_h et E_f) pour produire une probabilité théorique d'appariement aléatoire de chaque type d'appariement en fonction du nombre d'hommes et de femmes présents dans chacun des items de la variable éducation (p_r) et de corriger les pondérations (c_r) sur la base de cette distance relative entre probabilité théorique (p_t) et probabilité observée.

Il convient tout de suite de noter que, par rapport à la méthode comptable, la limite purement statistique (iv) évoquée *supra* n'existe pas avec la méthode par imputation : par construction, le nombre d'observations dans la population observée et la population contrefactuelle est le même (N).

De plus, s'appuyer sur des couples effectivement observés pour construire une population contrefactuelle à l'avantage de dépasser les limites (ii) et (iii) présentées *supra* : par définition la cohérence budgétaire et la cohérence familiale sont préservées au niveau du ménage. Pour l'analyse d'un système socio-fiscal complexe, l'avantage d'une telle méthode est évident : comme elle ne nécessite pas d'identifier la contribution marginale aux revenus du couple des deux individus qui le composent, les ménages de la population contrefactuelle sont cohérents quel que soit le type de revenus observés. De même, par construction, les revenus du couple et leur nombre d'enfants sont cohérents avec le système socio-fiscal en vigueur. D'une certaine manière la population observée des couples constitue l'ensemble des possibles et c'est la structure de la population qui est modifiée dans la construction du contrefactuel. De ce point de vue, l'hypothèse centrale qui permet de dépasser les limites (ii), (iii) et (iv) de la méthode comptable est l'hypothèse de *mimétisme des comportements des couples observés en coupe* pour construire le contrefactuel. Son adoption a deux conséquences :

- (a) L'absence d'une garantie de l'égalité entre les agrégats de la population contrefactuelle et les agrégats de la population observée, et notamment, la somme des revenus agrégés des ménages ($\sum_{m=1}^N Rc$).
- (b) L'absence d'une cohérence inter-temporelle du comportement d'ajustement des ménages.

⁶ Cf. également annexe 2.

La littérature économique considère généralement de manière implicite que la première conséquence (a) n'est pas une limite mais un résultat de l'analyse (Greenwood *et al* [2014], Harmenberg [2014]). Cette différence est généralement interprétée comme le résultat d'un ajustement de l'offre de travail qui aurait lieu dans le cas d'une mise en couple aléatoire. D'une certaine manière la littérature économique tend à considérer que cette conséquence permet de dépasser la limite (v) de la méthode comptable sur l'ajustement des comportements. Il nous semble qu'une appréciation du dépassement de la limite (v) ne peut être faite qu'au regard de la deuxième conséquence et de ses liens avec les limites en termes d'homogamie générationnelle (i) de la méthode comptable.

En effet, il convient de noter qu'avec la méthode par imputation comme dans le cas de la méthode comptable, la construction de la population contrefactuelle peut conduire à mettre en couple indifféremment des « individus très jeunes » et des « individus très vieux », et biaiser la tendance à l'homogamie par âge ou génération. Par ailleurs dans une analyse où la variable de contrôle d'homogamie est le niveau d'éducation, la méthode ne tient pas compte de la configuration familiale effective de la population observée : une famille peut se voir « retirer » ses enfants dans le contrefactuel ou au contraire se voir « affecter » plusieurs enfants. De ce point de vue, l'ajustement des comportements par « mimétisme en coupe » se fait au prix de la perte de la cohérence inter-temporelle des ménages dans la population contrefactuelle. Dans la mesure où la formation de la famille est un processus dynamique, la cohérence familiale (iii) est une limite qui n'est pas dépassée par la méthode par imputation. Bien évidemment, comme dans le cas de la méthode comptable, il est possible de complexifier la structure de l'appariement contrefactuel pour contrôler un certain nombre de biais. Par exemple, structurer p_t non plus uniquement en fonction du niveau d'éducation, mais par niveau d'éducation croisé avec le nombre d'enfants, permet de rendre compte du fait que les choix d'éducation, d'union et de devenir parents sont endogènes. Mais dans ce cas, les « jeunes couples » n'ayant pas encore eu d'enfant et les couples pour lesquels les enfants ont décohabité auront des comportements mimétiques dans la population contrefactuelle ; or d'un point de vue économique, toutes choses égales par ailleurs, il n'apparaît pas évident que les individus qui composent ces différents types de couples puissent obtenir le même type de revenus sur le marché du travail. Par ailleurs comme dans le cas de la méthode comptable, il est possible d'arguer que certains revenus individuels (comme par exemple les pensions de retraite ou les allocations chômage) sont le résultat de trajectoires salariales sur le marché du travail. Dans la mesure où le *mimétisme* comportemental de la méthode par imputation ne garantit pas une cohérence inter-temporelle des revenus des individus qui composent la population contrefactuelle, la limite de cohérence budgétaire (ii) n'est pas non plus dépassée.

De manière plus générale, il apparaît difficile de tenir compte du fait que les choix d'éducation, de carrière, d'âge d'union et de parentalité sont endogènes sur la base de données en coupe car *i*) les choix des individus ont des conséquences sur la temporalité de la carrière et la vitesse de formation de la famille *ii*) les données en coupe sont composées d'un empilement de génération à des âges différents dont les trajectoires à âge donné peuvent avoir été différentes. Par exemple, les études disponibles montrent une modification du calendrier et du nombre de naissances pour les générations récentes. En effet, l'âge au premier enfant n'a cessé de croître depuis les années 1970 (Davie, [2012]) et le niveau de fécondité a diminué depuis 40 ans même si cette baisse s'est

interrompue⁷ (Toulemon *et al.* [2008], Prioux *et al.* [2010]). Pour expliquer ces évolutions, plusieurs facteurs sont mis en lumière par la littérature. Les plus fréquemment évoqués sont l’allongement de la durée des études (Bac *et al.* [2005], Davie, [2012] Ní Brohlcháin et Beaujouan [2012]), l’augmentation de l’incertitude sur le marché du travail (Meron et Widmer [2002]), l’augmentation de l’activité féminine et le développement des méthodes de contraception (Toulemon *et al.* [2008]). Si ces évolutions semblent concerner les deux sexes, les calendriers masculins et féminins d’entrée dans la parentalité demeurent tout de même sensiblement différents. En effet, les femmes ont des enfants plus tôt que les hommes et concentrent leur vie reproductive sur un intervalle de temps plus réduit⁸. Au-delà de ces effets générationnels, au niveau individuel, plusieurs déterminants peuvent jouer sur la probabilité d’avoir un enfant : la situation sur le marché du travail (Meron et Widmer [2002] Pailhé et Solaz, [2007]), l’âge de fin d’étude (Ní Brohlcháin et Beaujouan [2012]), la situation conjugale des individus mais également les caractéristiques de l’union et la composition de la famille (Toulemon, [1994]).

Comme indiqué précédemment, il est théoriquement possible de construire la variable d’homogamie à un niveau très fin en croisant plusieurs variables disponibles pour essayer de contrôler au mieux ces effets. Dans ce cadre cependant, d’un point de vue statistique, plus on cherche à détailler et plus le nombre d’appariement théoriquement possibles mais non observés risque d’être important⁹. Or contrairement à la méthode comptable, l’hypothèse de *mimétisme* liée à la méthode par imputation fait que ces appariements théoriquement possibles mais non observés ne rentrent pas dans la construction du contrefactuel. De ce point de vue, le contrôle croisé des effets d’éducation, d’âge et de génération nécessite potentiellement des bases de données avec un nombre d’observation beaucoup plus important que celles généralement utilisées pour estimer les inégalités de niveau de vie.

Les méthodes de microsimulation

Dans leur version les plus simples, la méthode comptable et la méthode par imputation ne permettent de garantir ni la cohérence budgétaire inter-temporelle (ii), ni la cohérence familiale inter-temporelle (iii). Par ailleurs, le contrôle de l’homogamie générationnelle et ses implications en termes de caractéristiques (par exemple éducatives) des conjoints potentiels n’est pas traitée ; les ajustements de comportements sur le marché du travail (limite v) sont uniquement traités avec la méthode par imputation d’une manière dont nous avons montré *supra* qu’elle était peu satisfaisante.

Les méthodes de microsimulation dynamique portent une attention particulière à la cohérence inter-temporelle de ménages contrefactuels. Pour mémoire, les modèles de microsimulation statiques (parfois appelés comptables) visent à calculer (le plus souvent sur barème) les différents prélèvements et transferts pour chaque ménage à partir des revenus primaires observés : de ce point de vue, ces modèles garantissent la cohérence budgétaire des

⁷ Thévenon et Luci [2011] constatent même un regain de la fécondité sur les périodes récentes qui peut être attribué au développement des politiques qui permettent une meilleure conciliation entre vie professionnelle et vie familiale.

⁸ Cf. Annexe 3.

⁹ Par exemple, sur la base d’une variable de diplôme en 20 modalités (cf. *infra*, tableau 1), on obtient une quarantaine de combinaisons non observées dans l’ERFS 2011 (Annexe 2, tableau A2.2).

ménages (Blanchet, 2014). Les modèles de microsimulation dynamique ajoutent à cette cohérence budgétaire statique, une cohérence budgétaire inter-temporelle (les pensions de retraite par exemple sont calculées sur la base de la trajectoire salariale passée) et une cohérence de la trajectoire familiale.

Dans l'idéal, pour mesurer l'effet de l'homogamie éducative sur les inégalités, on aimerait pouvoir contrôler la limite (i) et contrôler le mieux possible les incohérences temporelles, c'est-à-dire contrôler que lors de la mise en couple :

- (A) la distribution de diplômes des conjoints potentiels n'est pas la même au fil des générations ;
- (B) le diplôme des individus et celui de leurs conjoints quand ils sont en couple vont influencer leurs trajectoires futures sur le marché du travail ;
- (C) le diplôme des individus et celui de leurs conjoints quand ils sont en couple vont influencer la suite de leur trajectoire familiale (séparation, calendrier des naissances, probabilité de se mettre à nouveau en couple).

Plusieurs méthodes sont envisageables pour contrôler ces effets à l'aide de la microsimulation¹⁰. Nous voudrions présenter ici les possibilités offertes par un modèle de microsimulation de cohorte pour construire une population contrefactuelle sans homogamie éducative. Nous prenons comme référence le modèle présenté dans Courtioux et Lignon [2015]¹¹. Il s'agit d'un modèle de microsimulation dynamique destiné à simuler les trajectoires d'une cohorte de naissance à la sortie du système éducatif. Deux questions sont successivement abordées ici : la mesure dans laquelle le modèle prend en compte les trois éléments présentés *supra*, puis la manière dont il est possible de reconstruire une population contrefactuelle à partir d'une cohorte d'individus.

Pour tenir compte du fait que la distribution en termes de formation initiale de conjoints potentiels n'est pas la même au fil du temps, Courtioux et Lignon [2014] modélisent sur la base de couples observés la probabilité pour le détenteur d'un diplôme donné d'avoir un conjoint de tel ou tel diplôme en introduisant une tendance temporelle pour chaque type de diplôme (A). En s'appuyant sur des données décrivant les transitions sur le marché du travail, ils modélisent également ces probabilités de transitions : dans la simulation de la carrière d'un individu et de son potentiel conjoint, à un moment donné du temps, le diplôme obtenu joue sur la probabilité de transition vers le chômage de manière directe et de manière indirecte - via la trajectoire d'emploi passée (B). Par ailleurs, la probabilité de transition vers telle ou telle configuration familiale (union, parentalité, etc.) est estimée en tenant compte des diplômes des deux conjoints et de l'impact que leur situation respective vis-à-vis du marché du travail peut avoir sur cette probabilité (C).

Au final, le modèle de microsimulation produit les trajectoires stylisées d'une cohorte de naissance donnée en préservant la cohérence inter-temporelle des individus. Il est alors possible de construire une cohorte contrefactuelle exempte d'homogamie éducative, c'est-à-dire en ayant modifié le point (A) mais en ne modifiant pas les lois de transitions conditionnellement au diplôme

¹⁰ Pour des exemples français de microsimulation dynamique, cf. notamment Blanchet *et al.* [2011], Allègre et Timbeau [2014], Lignon [2014], Courtioux et Lignon [2015].

¹¹ Pour calculer un niveau de vie, nous avons besoin de simuler le nombre d'enfants et leur âge tout au long de la trajectoire de vie des individus afin de pouvoir calculer les unités de consommation du ménage de l'individu de référence à chaque âge. Courtioux et Lignon [2015] ne simulent pas l'arrivée d'enfants ni les trajectoires de parentalité. C'est pourquoi dans notre exercice de microsimulation, nous avons dû avoir recours à un module démographique de parentalité qui est décrit de manière exhaustive dans l'annexe 3.

correspondant aux points (B) et (C). On dispose alors de deux cohortes : une cohorte stylisée et une cohorte stylisée contrefactuelle sans homogamie éducative.

La comparaison directe entre une population et une cohorte de naissance n'est pas possible *a priori*. Cependant, il est possible de transformer ces cohortes de naissance en population en posant une hypothèse de stationnarité : la population à un moment donné du temps est constituée de l'empilement de la même cohorte à des âges différents. A ce stade la population d'individus n'est pas directement comparable avec la population de ménages utilisée par les méthodes comptables ou par imputation. Pour passer de cette population d'individus à une population de ménages, il faut 1) conserver les individus qui ne sont pas en couples (selon le nombre d'enfants, ils constituent des ménages de célibataires ou des familles monoparentales), 2) pour les individus en couple, pour tenir compte du fait qu'ils peuvent être en couple avec quelqu'un de la cohorte ayant servi à construire la population ou avec quelqu'un d'une autre cohorte, il faut diviser leur pondération par deux.

II. Les bases de données utilisées

Il existe plusieurs bases de données permettant de mesurer de manière précise les revenus, la taille du ménage et l'éducation des individus qui composent le ménage. Dans la mesure où nous nous concentrons sur le cas français, nous avons choisi d'utiliser principalement l'Enquête Revenus Fiscaux et Sociaux pour l'année 2011 (ERFS 2011) qui apparie les ménages de l'enquête Emploi avec la ou les déclarations fiscales des individus qui composent ces ménages. Dans ce cadre, une partie de l'information portant sur les impôts et les transferts sociaux versés sont également récupérés (notamment l'information portant sur le montant de l'impôt sur le revenu et sur le versement des minima sociaux distribués par la Caisse d'Allocation Familiale). Une autre partie des impôts (notamment la taxe foncière), ainsi qu'une partie des revenus du patrimoine que la déclaration fiscale ne permet pas de bien appréhender, sont imputés aux ménages par le producteur de données (l'Insee) sur la base de modèles économétriques¹². Néanmoins les enquêtes revenus fiscaux (ERF) ont connues des ruptures de série, notamment avec le passage à l'enquête Emploi en continue en 2003 et la progressive extension du champ des revenus couverts par l'enquête. Ceci justifie donc de se concentrer dans un premier temps sur une année particulière pour éclairer les enjeux liés à l'analyse du cas français.

En termes de revenus, l'intérêt de l'ERFS 2011 est de fournir directement le niveau de vie des ménages (NIVVIEM) - c'est-à-dire le revenu disponible (REVDISP) par unité de consommation (UC)¹³ -, mais également le revenu déclaré du ménage (REVDECM) - c'est-à-dire les revenus nets de cotisations sociales avant taxes et transferts. Il convient de noter que d'un point de vue légal les taxes et les transferts portent sur des assiettes de revenus potentiellement différentes¹⁴ qui incluent

¹² Un avantage de cette enquête est qu'elle est réalisée annuellement depuis 1995 et peut donc permettre dans un second temps d'approfondir l'évolution intertemporelle de l'effet de l'homogamie éducative sur les inégalités, ce qui n'est pas possible avec des enquêtes comme EU-SILC qui ne permettent une comparaison internationale homogène que depuis 2004.

¹³ Nous retenons ici comme mode de calcul des UC l'échelle d'équivalence Insee-OCDE. En suivant cette échelle, le premier adulte compte pour 1, chaque personne additionnelle de plus de 14 ans compte pour 0,5 et chaque personne additionnelle de moins de 14 ans compte pour 0,3.

¹⁴ En effet, un « ménage » au sens de l'Insee peut regrouper plusieurs « familles » au sens de la CNAF ou plusieurs « foyers fiscaux » au sens de l'administration fiscale.

tout ou une partie des revenus des individus qui composent le ménage. De ce point de vue, après prise en compte du système socio-fiscal, il n'y a pas de règle légale d'affectation des taxes et transferts au revenu d'un individu. Cet aspect ne pose pas de problème aux méthodes par imputation qui se caractérisent par une modification de la pondération du ménage, mais il ne permet pas de mettre en œuvre les méthodes comptables sans hypothèse complémentaire (cf. *supra*). En effet, ces méthodes nécessitent d'avoir une variable de revenu différenciée pour les deux conjoints. L'avantage de l'ERFS 2011 est de fournir au niveau individuel une décomposition des revenus catégoriels. Il est possible de distinguer pour chaque individu les salaires déclarés (SALAIRES_I), les allocations chômage et de préretraite (CHOMAGE_I) ainsi que les revenus des indépendants - c'est-à-dire les revenus agricoles (RAG_I), les revenus industriels et commerciaux (RIC_I) et les revenus non commerciaux (RNC_I). Ces différents types de revenus peuvent donc être calculés au niveau du ménage et décomposés au niveau du couple ou d'un individu en particulier.

En termes d'éducation, il est possible d'identifier le diplôme de l'individu à un niveau assez fin. On utilise pour cela les variables, « Diplôme le plus élevé obtenu » (DIP) et « Diplôme le plus élevé obtenu dans l'enseignement supérieur » (DIPSUP). Le tableau 1 présente les découpages principaux qui ont été retenus. Un premier découpage retient 5 catégories assez proches de celles utilisées par Greenwood *et al.* [2014]. L'autre découpage est plus fin et retient 20 diplômes. Ce dernier niveau permet de rendre compte au mieux des particularités du système d'enseignement supérieur français où, pour un même niveau d'éducation (par exemple le niveau Bac+5), existent différents types de diplômes aux caractéristiques particulières (par exemple écoles de commerce, écoles d'ingénieurs, masters universitaires).

Un autre intérêt de cette base de données consiste à permettre une identification des couples au sein du ménage. Contrairement à la méthodologie de Greenwood *et al.* [2014], nous avons choisi de ne pas écarter *a priori* les ménages complexes (c'est-à-dire ceux qui peuvent réunir plusieurs couples en leur sein, par exemple un couple d'ascendants et un couple de descendants vivant sous le même toit). En effet, en suivant cette méthodologie, ils peuvent être traités de la même manière que les ménages composés d'un célibataire ou d'un parent isolé. Pour reconstruire les couples, nous avons choisi de ne pas nous appuyer sur la variable « Type de ménage » (TYPMEN5), mais plutôt sur la variable précisant pour un individu donné qui est son conjoint au sein du ménage (NOICON). Généralement au sein d'un ménage, la personne de référence a un numéro d'ordre (NOI) égal à 1 et son conjoint (s'il existe) égal à 2, mais ce n'est pas systématiquement le cas¹⁵. Chaque ménage pouvant être composé de plusieurs couples, nous avons choisi de définir le couple de référence du ménage comme celui formé par la première personne interrogée et son conjoint. Au sein des couples, nous ne retenons que les couples composés d'individus de sexe différent. En effet, seule cette stratégie permet la construction de contrefactuels sans homogamie éducative, du type de ceux présentés précédemment. Les couples composés d'individus de même sexe - ils représentent un peu moins d'un pourcent des observations de couples - sont traités de la même manière que les ménages qui ne sont pas composés de couples. Pour ces ménages non traités, bien que les caractéristiques de la personne de référence ne soient pas directement utiles pour calculer l'effet de l'homogamie éducative sur les inégalités en suivant la méthodologie de Greenwood *et al.* [2014], il apparaît important de l'identifier afin de déterminer si le ménage entre dans le champ des estimations (notamment en ce qui concerne l'âge de la personne de référence). En cohérence avec ce qui a été fait pour repérer le couple de référence au sein du ménage, nous

¹⁵ Pour certains ménages, le premier individu a un numéro d'ordre égal à 4 ; il peut également y avoir « inversion » entre le numéro d'ordre d'un enfant et celui du conjoint.

retenons comme personne de référence pour les ménages ne comprenant pas de couple, la personne dont le numéro d'ordre (NOI) est le plus petit.

Pour compléter cette information sur la structure familiale, nous utilisons le nombre d'unités de consommation du ménage qui est directement renseigné dans l'enquête (NB_UCI). De même, l'âge détaillé (AG) de l'individu de référence et celui de son éventuel conjoint sont renseignés et conservés pour déterminer s'ils appartiennent au champ de l'étude.

Tableau 1. Niveau d'éducation et diplôme : les catégories retenues

Niveau d'éducation	Diplôme
Inférieur aux études secondaires	Pas de diplôme
Etudes secondaires	CAP/BEP Bac général Bac professionnel Bac technique
Etudes tertiaires	Capacité en droit Licence 2 (deug) DUT/DUST BTS Technicien du supérieur Diplôme paramédical
Etudes tertiaires de niveau Bac+3 à Bac+5	Licence 3 Autres diplômes de niveau au moins Bac+3 Master 1 (maîtrise) Master recherche (DEA) Master professionnel (DESS) Ecole de commerce Ecole d'ingénieurs
Etudes tertiaires niveau doctorat	Doctorat (hors santé) Doctorat (santé)

La population issue de l'exercice microsimulation et celle issue l'enquête l'ERFS 2011 ne sont pas directement comparables. En effet, le modèle de microsimulation ne simule pas l'ensemble des revenus, ni l'ensemble du système socio-fiscal¹⁶. Néanmoins, les revenus tirés directement du travail et les allocations chômage sont présents dans les deux sources. Dans la simulation, les revenus tirés directement du travail sont simulés sur la base d'une équation de gain à la Mincer dont la variable expliquée est le log du salaire mensuel (SALRED) : on reconstruit un revenu annuel du travail en

¹⁶ Pour mémoire, à ce stade de développement du modèle, les revenus du patrimoine ne sont pas simulés. Par ailleurs, la composante familiale des transferts n'est pas non plus simulée (quotient familiale, allocations familiales, etc.).

multipliant ce salaire mensuel par 12¹⁷. De ce point de vue, pour simuler le revenu des indépendants, on fait l'hypothèse que le salaire potentiel constitue une bonne approximation. Par ailleurs, le modèle retient un pas annuel et ne permet pas de cumuler au sein d'une même année des périodes de chômage et des périodes d'emploi : dans ce cadre, la distribution de revenus annuels tirés directement du travail est lissée pour les individus qui connaissent plusieurs mois de chômage dans l'année.

Cette stylisation des trajectoires dans notre modèle de microsimulation devrait conduire à écraser la dispersion des queues de distribution lorsque l'on ne retient que les revenus tirés directement du marché du travail (salaires). Néanmoins, on peut penser que la prise en compte des allocations chômage devrait conduire à atténuer cet effet d'écrasement ; c'est pourquoi nous retenons pour notre comparaison entre la méthode par microsimulation et les deux autres méthodes la variable « revenu tirés directement et indirectement du travail » qui inclut également les allocations chômage. Pour rester sur une définition cohérente des inégalités, nous divisons ce revenu par le nombre d'unités de consommation.

Pour limiter les problèmes liés à l'absence de contrôle de l'homogamie générationnelle dans les méthodes comptables et les méthodes par imputation, nous nous concentrons sur les ménages d'âge actif que nous définissons comme les ménages dont la personne de référence et son conjoint ont tous les deux entre 25 et 54 ans inclus¹⁸. Pour la méthode par microsimulation dynamique qui nécessite de choisir les caractéristiques de formation initiale de la cohorte stylisée utilisée pour construire la population, nous choisissons les caractéristiques éducatives de la cohorte des individus nés en 1970¹⁹.

III. Les résultats

Selon la variable de revenu retenue, nous disposons de trois types d'indices de Gini de référence pour mesurer les inégalités entre ménages dont l'individu de référence et son conjoint potentiel ont entre 25 et 54 ans : l'indice de Gini pour les niveaux de vie, celui des revenus du travail ajustés à la taille du ménage et l'indice de Gini pour les revenus tirés du travail non ajustés. Ces trois types d'indices sont reportés dans chacune des colonnes du tableau 2. Les indices qui nous intéressent le plus sont ceux des deux premières colonnes car ils sont ajustés à la taille du ménage et correspondent à la définition des inégalités en termes de niveau de vie sur laquelle nous nous concentrons ici. La dernière colonne est produite à titre informatif afin de permettre la comparaison avec les résultats disponibles dans la littérature pour le cas français²⁰. Par ailleurs, dans la mesure où les résultats ne sont pas directement comparables, une séparation claire dans le tableau est effectuée entre ce qui relève d'un calcul sur l'ERFS 2011 et ce qui est issu de l'exercice de microsimulation. On notera que la colonne niveau de vie n'est pas renseignée pour les calculs issus du modèle de microsimulation : en l'état actuel de développement du modèle, ce dernier ne permet

¹⁷ Pour plus de détails sur l'estimation des équations de gain, cf. Courtioux *et al.* [2011].

¹⁸ A titre d'exemple, Greenwood *et al.* [2014] se concentrent sur les individus âgés de 25 à 64 ans.

¹⁹ Nous effectuons ce choix car le travail économétrique en amont du modèle de microsimulation s'appuie principalement sur des bases de données plus anciennes que 2011 pour lesquelles les individus de la cohorte de naissance 1970 constituent un âge médian (cf. Courtioux *et al.* [2011]).

²⁰ Notamment Courtioux et Lignon [2015], Frémeaux et Lefranc [2015].

pas de rendre compte du système socio-fiscal de manière aussi complète que l'enquête ERFS 2011 (cf. *supra*).

Dans le tableau 2, les éléments en gras correspondent aux calculs des indices de Gini de référence (c'est-à-dire qui tiennent compte des comportements d'homogamie des individus). Ces éléments descriptifs permettent déjà de donner un premier aperçu de l'impact de la taille du ménage et du système socio-fiscal sur les inégalités. La prise en compte de la taille du ménage contribue à réduire les inégalités de revenu du travail (salaires, traitements, revenus des indépendants et allocations chômage) de 8%. Ceci tend à indiquer que, dans l'ensemble, les ménages qui ont les revenus tirés du marché du travail les plus élevés sont également ceux qui ont la charge de famille la plus importante²¹. Ce résultat reflète en partie un effet cycle de vie : les individus les plus avancés dans leur carrière (qui sont également plus avancés en âge) ont des salaires plus élevés, mais également des enfants relativement plus nombreux et plus âgés que les individus en début de carrière. Par ailleurs, après prise en compte de l'ensemble des taxes et des transferts, les inégalités sont moins importantes que lorsque seuls les revenus tirés du travail sont pris en compte. De ce point de vue, le système socio-fiscal contribue quant à lui à réduire les inégalités de 16%.

C'est bien à l'aune de cet effet important que l'on peut apprécier l'effet propre de l'homogamie éducative sur les inégalités. Dans le tableau 2, il est important de bien distinguer entre les contrefactuels construits à partir de la méthode par imputation qui permettent de mesurer l'effet propre de l'homogamie éducative et les contrefactuels produits à l'aide de la méthode comptable qui permet de mesurer l'effet complet de l'homogamie. Les contrefactuels par imputation constituent le point central qui nous intéresse ici tandis que ceux construits par la méthode additive constituent quant à eux une référence indicative correspondant à un contrôle complet de l'homogamie.

Globalement, le tableau 2 montre que l'ordre de grandeur de l'effet d'une suppression de l'homogamie qu'elle soit complète ou seulement éducative est beaucoup moins important. Selon la méthode retenue cet effet propre de l'homogamie impacte les inégalités sur une échelle allant de 0,7 à 3% en valeur absolue²². On notera également que le signe de cet effet peut changer et dépend notamment de la méthode utilisée. Au-delà de cette première relativisation de l'effet de l'homogamie éducative, il convient maintenant de regarder plus en détail ce qui contribue à expliquer la magnitude de cet effet et son lien avec les différentes méthodes retenues

²¹ On notera que les résultats obtenus à partir du modèle de microsimulation qui cherche à corriger l'effet générationnel indiquent que cet effet est plus important pour la population stylisée correspondant à la cohorte née en 1970 : il atteint 13%.

²² Pour mémoire, si on s'intéresse aux inégalités de revenus du travail non corrigés par la taille du ménage (UC), l'effet de l'homogamie éducative est nettement plus élevé et peut atteindre 10-11% (cf. dernière colonne du tableau 2), c'est-à-dire des résultats plus proches de ceux de Courtioux et Lignon [2015] et de Frémeaux et Lefranc [2015].

Tableau 2. Indices de Gini

	niveau de vie ^a	revenus tirés du marché du travail (ajustés) ^b	revenus tirés du marché du travail ^c
enquête revenus fiscaux			
référence	0,2895	0,3447	0,3752
contrefactuel par imputation			
<i>niveau d'éducation</i>	0,2871	0,3389	0,3711
<i>diplôme fin</i>	0,2859	0,3370	0,3699
contrefactuel par méthode comptable avec <i>bootstrap</i>			
	0,2929	0,3656	0,3424
microsimulation*			
référence		0,3119	0,3582
contrefactuel par appariement aléatoire intragénérationnel		0,3043	0,3522

Source : Enquête Revenus Fiscaux et Sociaux 2011 (Insee), (*) GAMEO (EDHEC) – Calcul des auteurs.

Champ : ménages avec (^a) niveau de vie/ (^{b, c}) revenus tirés du marché du travail positif et dont la personne de référence et son conjoint éventuel ont entre 25 et 54 ans.

Note : (^a) revenu disponible par UC, (^b) salaires, traitements, revenus indépendants et allocations chômage par UC, (^c) salaires, traitements, revenus indépendants et allocations chômage.

La méthode par imputation qui constitue ici le cœur de notre analyse, se caractérise par la construction d'un coefficient de correction des pondérations (c_r) qui dépend de la variable d'éducation choisie pour définir l'homogamie éducative. Les différents coefficients de correction calculés sur les ménages dont la personne de référence et son conjoint ont entre 25 et 54 ans sont présentés en annexe (tableau A2.1 et A2.2). Ces résultats permettent d'illustrer les enjeux de la méthode de correction : les coefficients les plus fortement corrigés (c'est-à-dire ceux que l'on devrait observer beaucoup plus souvent si l'appariement des couples était aléatoire) sont les couples dans lesquels l'homme n'a pas fait d'étude secondaire et la femme a un niveau doctorat ($c_r = 7,5446$), suivi des couples où l'homme a un niveau doctorat et la femme a moins que des études secondaires ($c_r = 4,8481$). Corrigé à un niveau plus fin, les coefficients de correction les plus élevés concernent les couples dans lesquels l'homme n'a pas de diplôme et la femme a un doctorat dans une discipline non médicales ($c_r = 27,733$), ainsi que les couples dans lesquels l'homme a un doctorat dans une discipline non médicale et la femme un bac professionnel ($c_r = 12,23$). Ces résultats montrent que les coefficients de correction peuvent être beaucoup plus forts lorsque l'on se place à un niveau fin d'éducation. Il convient également de noter que lorsque l'on travaille à un niveau très fin, un certain nombre d'appariements qui ne sont pas observés²³ ne sont (par définition) pas corrigés.

Les résultats obtenus dépendent de la variable d'éducation retenue pour caractériser l'homogamie. Pour cette méthode nous avons donc décidé de présenter quatre variantes sans homogamie éducative en croisant deux types de revenus et deux variables éducatives se distinguant par leur degré de précision (cf. tableau 1 pour les variables d'éducation utilisées). Les résultats montrent que la variable d'éducation choisie affecte de manière non négligeable le résultat. Lorsque

²³ Par exemple, dans la nomenclature détaillée (cf. annexe 2, table A2.2) les couples composés d'un homme titulaire d'un Bac professionnel et d'une femme titulaire d'un Doctorat (hors secteur médical) ne sont pas observés.

l'on retient le niveau d'éducation pour caractériser l'homogamie, cette dernière contribue à augmenter les inégalités de revenus tirés du travail de 1,7%, et les inégalités de niveau de vie de 0,8%. Une caractérisation plus fine de l'homogamie éducative à l'aide de la variable diplôme indique une contribution de l'homogamie plus forte : de 2,3% pour les revenus tirés du marché du travail et de 1,3% pour les niveaux de vie.

Par rapport à ces résultats, la méthode comptable qui contrôle l'ensemble des différentes dimensions de l'homogamie obtient des effets moins importants. Sans homogamie complète, les inégalités de revenus du travail ajustés à la taille du ménage ne diminueraient que de 0,6%, les inégalités de niveau de vie seraient plus importantes de 2%²⁴. Ce dernier effet *a priori* non attendu au vu de la littérature économique sur le sujet, suggère que l'effet propre de la dimension éducative de l'homogamie corrélée positivement au niveau de vie est supplanté par une autre dimension de l'homogamie (générationnelle, religieuse, etc.) corrélée négativement avec le revenu que nous n'identifions pas ici²⁵. L'existence d'un effet net positif de l'homogamie complète sur les niveaux de vie malgré un effet négatif de l'homogamie éducative stricte apparaît d'autant plus plausible que les estimations des effets propres de l'homogamie éducative sont faibles. Par ailleurs, des tests effectués à l'aide de la méthode par imputation sur l'ensemble de la population et non plus sur le champ des 25-54 ans laissent à penser que la dimension âge de l'homogamie²⁶ est un candidat plausible pour expliquer le sens de l'effet net. En effet, l'extension du champ c'est-à-dire le relâchement de l'hypothèse d'homogamie par âge diminue l'impact négatif de l'homogamie éducative sur les inégalités de niveau de vie ; de plus, lorsque l'on utilise une variable détaillée de diplôme, les comportements homogames contribuent alors à faire baisser les inégalités²⁷. Il apparaît dès lors important de contrôler les effets âges et générationnels, ou pour le moins savoir dans quelle mesure la méthode retenue pour simuler un contrefactuel sans homogamie éducative en tient compte.

Comme expliqué précédemment, la microsimulation a pour objectif de construire une cohorte permettant de contrôler le mieux possible les effets de génération en termes d'éducation, d'âge de mise en couple, etc. A ce stade de développement, le modèle que nous utilisons ne nous permet pas encore de reconstruire complètement un niveau de vie. Il nous permet cependant d'éclairer certaines intuitions sur les effets d'âge et de génération sur un revenu intermédiaire que constituent les revenus du travail par unité de consommation²⁸. En effet, la construction d'une pseudo-population à partir d'une cohorte simulée permet de produire des indicateurs d'inégalités

²⁴ La mesure de cet effet n'est pas un artefact statistique lié au plus grand nombre d'observations obtenu à partir de cette méthode : nous avons estimé un effet moyen robuste à l'aide de technique de *bootstrap* (cf. annexe 1).

²⁵ Pour mémoire, les différentes dimensions de l'homogamie et leurs évolutions au cours du temps est un objet de recherche pour lequel il existe une abondante littérature en sociologie et en démographie (par exemple Kalminj [1991] qui analyse dans le cas américain le basculement d'une homogamie religieuse vers une homogamie éducative). A notre connaissance, la corrélation entre ces différentes dimensions de l'homogamie et le niveau de vie des couples est beaucoup moins documentée.

²⁶ Pour mémoire, à partir de la méthode par imputation il n'est pas possible de distinguer un effet d'âge d'un effet de génération.

²⁷ Le Gini obtenu sur l'ensemble de la population est de 0,3167 ; le contrefactuel sans homogamie éducative en niveau d'éducation (cf. tableau 1) se fixe à 0,3179 ; le contrefactuel avec un niveau de diplôme plus fin se fixe à 0,3164. Selon ces estimations l'homogamie en niveau d'éducation augmente les inégalités de 0,4%, mais l'homogamie éducative en diplôme diminue les inégalités de 0,1%.

²⁸ Pour ce faire, il faudrait notamment pouvoir construire un module de simulation d'accès aux différents modes de garde qui ont un impact sur le montant de l'impôt sur le revenu dû par les familles.

pour une population représentant une génération stylisée donnée (ici la cohorte de naissance 1970) en laissant inchangé les effets d'homogamie par âge. Le tableau 2, indique que c'est lorsque l'on contrôle ces effets de génération en termes d'accès à l'éducation que cet effet de l'homogamie éducative est le plus important. Dans ce cadre l'homogamie éducative augmente les inégalités de 2,5%, c'est-à-dire à un niveau légèrement plus élevé.

Conclusion

Dans cet article, nous avons présenté et discuté différentes méthodes permettant d'estimer l'impact de l'homogamie éducative sur les inégalités de niveau de vie des ménages et les avons mis en œuvre sur des données françaises portant sur l'année 2011.

D'un point de vue méthodologique, la méthode comptable ne permet pas de distinguer l'effet de l'homogamie éducative des autres dimensions de l'homogamie. La méthode par imputation ne permet pas de distinguer homogamie éducative et homogamie de génération : ceci se traduit par une incohérence temporelle dans l'ajustement du comportement des individus à un changement de conjoint, notamment en termes d'offre de travail. Dans ce cadre, nous avons montré comment un modèle de microsimulation de cohorte dédié permet d'estimer un effet propre de l'homogamie éducative indépendant des effets « âge » et « génération ».

Les résultats obtenus pour le cas français montrent que les effets de l'homogamie éducative sont faibles. Ils varient dans une fourchette allant de 0,8 à 2,5%, c'est-à-dire un impact beaucoup plus faible que l'effet du système socio-fiscal français qui diminue les inégalités d'environ 16%. De plus, ces résultats sont obtenus de manière plus nette lorsque l'on retient une définition fine de l'éducation pour qualifier l'homogamie et lorsque l'on sépare les effets de l'homogamie éducative de l'homogamie générationnelle.

D'un point de vue plus général, ce travail suggère que l'impact de l'homogamie éducative sur les inégalités n'est pas négligeable mais que sa mise en évidence nécessite de bien identifier les autres dimensions de l'homogamie (et notamment l'homogamie générationnelle) dans la mesure où elles ne jouent pas forcément dans le même sens sur les inégalités de niveau de vie.

Références

- Allègre G. et X Timbeau [2014], « Les prêts à remboursement dans le supérieur : plus redistributif que l'impôt ? Une perspective de cycle de vie », communication : *International Workshop on Income Contingent Loan : public policies for funding higher education and managing social risk*, 16 June 2014, Maison des Sciences Economiques (Paris).
<http://professoral.edhec.com/actualites/evenements/international-workshop-on-income-contingent-loans-public-policies-for-funding-higher-education-and-managing-social-risk-198859.kjsp>
- Bac C., Legendre F., Mahieu R. et F. Thibault [2005], « Fécondité et âge de fin d'études en France depuis 1975. L'évolution au fil des générations des facteurs traditionnels de la fécondité », *Recherches et Prévisions*, 79, 21-35.
- Blanchet D., Buffeteau S., Crenner E. et Le Minez S. [2011], « Le modèle de micro-simulation Destinie 2 : principales caractéristiques et premiers résultats », *Économie et Statistique*, 441, 101-121.
- Blanchet D. [2014], « La microsimulation dynamique : principes généraux et exemples en langage R », Document de travail, Insee, M 2014/01.
- Courtioux P., Gregoir S., et D. Houeto [2011], « Enseignement supérieur et durées de subvention individuelle implicite. Une analyse par microsimulation dynamique », *Revue économique*, 62 (5), 835-866.
- Courtioux P. et V. Lignon [2014], « Avoir un diplôme pour faire une bonne carrière ou un bon mariage », *Position Paper EDHEC*, Mai.
- Courtioux P. et V. Lignon [2015], « Homogamie éducative et inégalité de revenu salarial : une perspective de cycle de vie », *Economie et Statistique*, 481 & 482, 149-183..
- Davie E. [2012], « Un premier enfant à 28 ans », *Insee Première*, Insee, n°1419.
- Eika L., Mogstad M et B. Zafar [2014], « Educational Assortative Mating and Household Income Inequality », *NBER Working Paper*, n°20271.
- Frémeaux N. et A. Lefranc [2015], « Assortive mating and earning inequality in France », Working Paper.
- Greenwood J., Guner N., Kocharkov G. et C. Santis [2014], « Marry Your Like: Assortative Mating and Income Inequality », *American Economic Review : Papers and Proceedings*, 104 (5), 3-5.
- Harmenberg K. [2014], « A Note : The Effect of Assortative Mating on Income Inequality », Working paper.
- Hryshko D., Chinhui J. et K. McCue [2014], « Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important is Assortative Matching? », *CES Discussion Paper*, U.S. Census Bureau, n°15-04, January.
- Kalmijn M., [1991], « Trends in religious and educational homogamy », *American Sociological Review*, 56 (6), 786-800.
- Lignon V. [2014], *Transitions familiales, professionnelles et investissements éducatifs. Une analyse par microsimulation dynamique*, Thèse de doctorat en économie, Université Paris 1 Panthéon-Sorbonne.
- Meron M. et I. Widmer [2002], « Les femmes au chômage retardent l'arrivée du premier enfant », *Population*, 57 (2), 327-357.
- Ní Brohlcháin M. et E. Beaujouan, [2012], « Fertility postponement is largely due to rising education enrolment », *Population Studies*, 66 (3), 311-327.

- Pailhé A. et A. Solaz [2007], « Inflexions des trajectoires professionnelles des hommes et des femmes après la naissance d'enfants », *Recherches et Prévisions*, 90, 5-27.
- Prioux F., Mazuy M. et M. Barbieri [2010], « L'évolution démographique récente en France : les adultes vivent moins souvent en couple », *Population*, 65 (3), 421-474.
- Enquêtes Revenus fiscaux et sociaux (ERFS) [2011], INSEE [producteur], ADISP-CMH [diffuseur].
- Thévenon O. et A. Luci [2011], « La fécondité remonte dans les pays de l'OCDE : est-ce dû au progrès économique ? », *Population et sociétés*, n°481.
- Toulemon L. [1994], « La place des enfants dans l'histoire des couples », *Population*, 49 (6), 1321-1345.
- Toulemon L., Pailhé A. et C. Rossier [2008], « France : High and stable fertility », *Demographic Research*, 19 (16), 503-556.

Annexes 1 : L'effet propre du nombre d'observations sur l'indice de Gini

L'indice de Gini est un indicateur couramment utilisé pour mesurer les inégalités de répartition au sein d'une population. Son caractère synthétique (il résume par un indice les caractéristiques de la courbe de Lorenz dont il est tiré) permet théoriquement de l'utiliser pour comparer la répartition pour des populations et/ou des biens distribués qui n'ont rien en commun. Cependant, pour une population donnée, le degré de précision retenu (par exemple classer les individus par décile *versus* par percentile croissant pour tracer la courbe de Lorenz) peut influencer le niveau du Gini : plus le degré de précision est faible plus le Gini sera faible. L'importance du biais dépendra de la structure de la population et de sa répartition.

De ce point de vue, comparer un indice de Gini calculé sur l'ERFS 2011 et un indice de Gini calculé sur la population contrefactuelle obtenu à l'aide de la méthode comptable (cf. *supra*) pose problème. Dans le premier cas, nous disposons d'une population de N observations de couples auxquelles sont ajoutées les autres types de ménages (soit 56 486 observations) ; dans le second cas, nous disposons de N^2 observations de couples dont les poids relatifs sont plus faibles, auxquelles sont ajoutés les mêmes autres types de ménages (soit un total de 258 157 384 observations). Le degré de précision telle que l'on peut l'appréhender à travers le nombre d'observations est plus grand dans le cas de la population contrefactuelle, ce qui risque de biaiser à la hausse son indice de Gini et donc de sous-estimer l'impact de l'homogamie sur les inégalités.

Pour corriger notre indice de Gini calculé sur la population contrefactuelle de ce biais de précision, nous utilisons une méthode de *bootstrap* qui permet de reconstruire un ensemble de populations contrefactuelles avec un nombre d'observations comparable à celui de l'ERFS 2011. Pour effectuer le tirage d'une population contrefactuelle comparable en observations à l'ERFS au sein de la population contrefactuelle complète, on multiplie chacune des pondérations individuelles des ménages de la population contrefactuelle par le résultat d'un tirage aléatoire dans une loi uniforme ; puis on sélectionne les 56 486 observations ayant la valeur la plus élevée. Nous effectuons 100 tirages permettant de construire 100 populations contrefactuelles avec un degré de précision comparable à celui des ERFS 2011. Les résultats issus de la méthode additive correspondent à la moyenne des indices de Gini calculés sur ces 100 populations contrefactuelles. Les résultats présentés dans le tableau A1 montrent que l'effet propre du nombre d'observations sur le Gini est négligeable et ne remet pas en cause les résultats que nous présentons.

Tableau A1. Résultats de la méthode par *bootstrap*

	méthode additive simple (A)	méthode additive avec bootstrap		(A) - (B)
		moyenne (B)	écart-type	
Indices de Gini				
niveau de vie ^a	0,2930	0,2929	5,844E-09	0,0001
revenus tirés du marché du travail (ajustés) ^b	0,3656	0,3656	1,605E-09	0,0000
revenus tirés du marché du travail ^c	0,3425	0,3424	4,598E-09	0,0000
Nombre de tirages	-	100		
Nombre d'observations	258 157 384	56 486		

Source : Enquête Revenus Fiscaux et Sociaux 2011 (Insee) – Calcul des auteurs.

Champ : ménages dont la personne de référence et son conjoint éventuel ont entre 25 et 54 ans.

Annexes 2

Tableau A2.1 Coefficient de correction de la pondération ménage par niveau d'éducation

Niveau d'éducation du conjoint	Niveau d'éducation de la conjointe	probabilité observée (p_o)	probabilité aléatoire théorique (p_r)	coefficient de correction de la pondération (c_r)
Moins que des études secondaires	Etudes tertiaires niveau doctorat	0,0006	0,0045	7,5446
Etudes tertiaires niveau doctorat	Moins que des études secondaires	0,0010	0,0048	4,8481
Etudes tertiaires de niveau Bac+3 à Bac+5	Moins que des études secondaires	0,0074	0,0344	4,6425
Des études secondaires	Etudes tertiaires niveau doctorat	0,0023	0,0091	3,9149
Moins que des études secondaires	Etudes tertiaires de niveau Bac+3 à Bac+5	0,0130	0,0429	3,3057
Etudes tertiaires niveau doctorat	Des études secondaires	0,0031	0,0095	3,1159
Des études tertiaires	Moins que des études secondaires	0,0098	0,0275	2,8135
Etudes tertiaires de niveau Bac+3 à Bac+5	Des études secondaires	0,0270	0,0689	2,5522
Moins que des études secondaires	Des études tertiaires	0,0202	0,0395	1,9487
Des études secondaires	Etudes tertiaires de niveau Bac+3 à Bac+5	0,0442	0,0858	1,9430
Des études tertiaires	Etudes tertiaires niveau doctorat	0,0015	0,0027	1,7813
Etudes tertiaires niveau doctorat	Des études tertiaires	0,0032	0,0041	1,2886
Des études tertiaires	Des études secondaires	0,0445	0,0550	1,2347
Des études secondaires	Des études tertiaires	0,0730	0,0789	1,0805
Des études secondaires	Moins que des études secondaires	0,0891	0,0918	1,0296
Moins que des études secondaires	Des études secondaires	0,0937	0,0918	0,9798
Etudes tertiaires de niveau Bac+3 à Bac+5	Des études tertiaires	0,0356	0,0296	0,8323
Des études secondaires	Des études secondaires	0,2403	0,1834	0,7632
Des études tertiaires	Etudes tertiaires de niveau Bac+3 à Bac+5	0,0351	0,0257	0,7335
Etudes tertiaires niveau doctorat	Etudes tertiaires de niveau Bac+3 à Bac+5	0,0078	0,0045	0,5749
Des études tertiaires	Des études tertiaires	0,0437	0,0236	0,5413
Moins que des études secondaires	Moins que des études secondaires	0,0971	0,0459	0,4728
Etudes tertiaires de niveau Bac+3 à Bac+5	Etudes tertiaires niveau doctorat	0,0074	0,0034	0,4599
Etudes tertiaires de niveau Bac+3 à Bac+5	Etudes tertiaires de niveau Bac+3 à Bac+5	0,0912	0,0322	0,3534
Etudes tertiaires niveau doctorat	Etudes tertiaires niveau doctorat	0,0084	0,0005	0,0565

Source : Enquête Revenus Fiscaux et Sociaux 2011 (Insee) – Calcul des auteurs.

Champ : ménage dont la personne de référence et son conjoint ont entre 25 et 54 ans.

Note : les appariements sous la ligne horizontale sont sous-représentés par rapport à ce qui serait observé dans le cas d'un appariement aléatoire.

Tableau A2.2. Coefficient de correction de la pondération ménage par diplôme

Niveau d'éducation du conjoint	Niveau d'éducation de la conjointe	probabilité observée (p_o)	probabilité aléatoire théorique (p_r)	coefficient de correction de la pondération (c_r)
Pas de diplôme	Doctorat (hors santé)	0,000	0,002	27,733
Doctorat (hors santé)	Bac professionnel	0,000	0,001	12,230
CAP/BEP	Ecole de commerce	0,000	0,002	8,932
Pas de diplôme	Ecole de commerce	0,000	0,002	8,830
Pas de diplôme	Ecole d'ingénieurs	0,000	0,003	7,789
Ecole de commerce	CAP/BEP	0,000	0,002	7,679
Master professionnel (DESS)	Pas de diplôme	0,001	0,006	7,357
Technicien du supérieur	Bac technique	0,000	0,000	6,820
Doctorat (hors santé)	CAP/BEP	0,000	0,002	6,638
Master recherche (DEA)	CAP/BEP	0,000	0,003	6,592
Doctorat (santé)	Bac professionnel	0,000	0,001	6,268
CAP/BEP	Master recherche (DEA)	0,001	0,004	6,162
Licence 2 (deug)	Bac professionnel	0,000	0,000	6,089
Ecole d'ingénieurs	Bac professionnel	0,000	0,002	6,051
Master recherche (DEA)	Pas de diplôme	0,000	0,002	5,907
Autres diplômes de niveau au moins Bac+3	Pas de diplôme	0,001	0,003	5,532
Autres diplômes de niveau au moins Bac+3	Bac professionnel	0,000	0,001	5,487
Autres diplômes de niveau au moins Bac+3	CAP/BEP	0,001	0,003	5,318
Bac technique	Doctorat (santé)	0,000	0,000	5,299
Pas de diplôme	Doctorat (santé)	0,001	0,003	5,206
Master professionnel (DESS)	CAP/BEP	0,001	0,007	5,163
Doctorat (santé)	Pas de diplôme	0,000	0,002	5,089
Ecole de commerce	Pas de diplôme	0,000	0,002	5,060
Master 1 (maîtrise)	CAP/BEP	0,001	0,006	4,946
CAP/BEP	Master professionnel (DESS)	0,002	0,012	4,891
Master 1 (maîtrise)	Pas de diplôme	0,001	0,006	4,861
CAP/BEP	Doctorat (hors santé)	0,000	0,002	4,846
Ecole de commerce	Bac technique	0,000	0,001	4,798
CAP/BEP	Ecole d'ingénieurs	0,001	0,003	4,766
Ecole d'ingénieurs	CAP/BEP	0,002	0,009	4,733
Doctorat (hors santé)	Pas de diplôme	0,001	0,002	4,617
Diplôme paramédical	Pas de diplôme	0,000	0,002	4,594
Bac général	Doctorat (hors santé)	0,000	0,000	4,588
Bac professionnel	Master recherche (DEA)	0,000	0,001	4,557
Ecole de commerce	Bac professionnel	0,000	0,000	4,526
CAP/BEP	Doctorat (santé)	0,001	0,004	4,252
Pas de diplôme	Autres diplômes de niveau au moins Bac+3	0,001	0,003	4,241
Pas de diplôme	Master recherche (DEA)	0,001	0,003	4,110
Ecole d'ingénieurs	Pas de diplôme	0,002	0,008	4,061
Pas de diplôme	Master professionnel (DESS)	0,002	0,009	3,905
Bac technique	Master professionnel (DESS)	0,000	0,001	3,903
Bac technique	Doctorat (hors santé)	0,000	0,000	3,850
DUT/DUST	Ecole de commerce	0,000	0,000	3,821
Master recherche (DEA)	Doctorat (santé)	0,000	0,000	3,575
Doctorat (hors santé)	Autres diplômes de niveau au moins Bac+3	0,000	0,000	3,552
DUT/DUST	Pas de diplôme	0,001	0,005	3,534
Master professionnel (DESS)	Bac professionnel	0,000	0,002	3,426
Doctorat (hors santé)	Bac technique	0,000	0,001	3,373
Licence 3	CAP/BEP	0,002	0,007	3,363
Licence 3	Pas de diplôme	0,002	0,007	3,358
Diplôme paramédical	Ecole de commerce	0,000	0,000	3,349
Diplôme paramédical	CAP/BEP	0,001	0,002	3,347
Licence 2 (deug)	Doctorat (santé)	0,000	0,000	3,291
Doctorat (santé)	CAP/BEP	0,001	0,003	3,274
Doctorat (santé)	BTS	0,000	0,001	3,190
Pas de diplôme	Technicien du supérieur	0,000	0,001	3,040
Pas de diplôme	Master 1 (maîtrise)	0,003	0,010	2,948
Doctorat (santé)	Bac technique	0,000	0,001	2,935
Master professionnel (DESS)	Bac général	0,001	0,003	2,857
CAP/BEP	Autres diplômes de niveau au moins Bac+3	0,001	0,004	2,828
BTS	Pas de diplôme	0,006	0,018	2,809
Bac technique	Ecole de commerce	0,000	0,000	2,802
Master 1 (maîtrise)	Bac technique	0,001	0,002	2,777
BTS	Doctorat (hors santé)	0,000	0,001	2,773
Licence 2 (deug)	CAP/BEP	0,001	0,002	2,738
Pas de diplôme	Licence 3	0,005	0,014	2,551
Pas de diplôme	DUT/DUST	0,002	0,004	2,547
Bac professionnel	Doctorat (santé)	0,000	0,001	2,527
Technicien du supérieur	Pas de diplôme	0,000	0,001	2,518
CAP/BEP	Master 1 (maîtrise)	0,005	0,013	2,517
Bac général	Ecole d'ingénieurs	0,000	0,001	2,484
Doctorat (santé)	Ecole d'ingénieurs	0,000	0,000	2,421
Doctorat (santé)	Bac général	0,000	0,001	2,376
Bac professionnel	Technicien du supérieur	0,000	0,000	2,363
Diplôme paramédical	Bac technique	0,000	0,000	2,362
Master professionnel (DESS)	Bac technique	0,001	0,002	2,326
Bac professionnel	Ecole d'ingénieurs	0,000	0,001	2,303
Bac technique	Autres diplômes de niveau au moins Bac+3	0,000	0,000	2,280
Diplôme paramédical	Autres diplômes de niveau au moins Bac+3	0,000	0,000	2,277
Bac professionnel	Ecole de commerce	0,000	0,000	2,271

Master 1 (maîtrise)	Bac professionnel	0,001	0,001	2,165
Doctorat (santé)	DUT/DUST	0,000	0,000	2,123
DUT/DUST	Bac professionnel	0,001	0,001	2,091
Master recherche (DEA)	Bac général	0,000	0,001	2,074
CAP/BEP	Licence 3	0,009	0,019	2,069
Doctorat (hors santé)	BTS	0,000	0,001	2,041
Diplôme paramédical	BTS	0,000	0,001	2,028
Pas de diplôme	BTS	0,010	0,020	1,946
CAP/BEP	Capacité en droit	0,000	0,000	1,941
Bac général	CAP/BEP	0,007	0,013	1,854
Bac technique	Pas de diplôme	0,004	0,007	1,848
Ecole d'ingénieurs	Bac technique	0,001	0,002	1,842
Pas de diplôme	Diplôme paramédical	0,006	0,011	1,820
DUT/DUST	Doctorat (hors santé)	0,000	0,000	1,723
Bac général	Doctorat (santé)	0,000	0,001	1,711
Licence 3	Bac technique	0,001	0,002	1,674
Diplôme paramédical	Ecole d'ingénieurs	0,000	0,000	1,672
Licence 2 (deug)	Ecole d'ingénieurs	0,000	0,000	1,663
CAP/BEP	Technicien du supérieur	0,001	0,001	1,652
Ecole de commerce	Licence 2 (deug)	0,000	0,000	1,636
Technicien du supérieur	DUT/DUST	0,000	0,000	1,632
Technicien du supérieur	Master recherche (DEA)	0,000	0,000	1,626
Pas de diplôme	Licence 2 (deug)	0,002	0,003	1,623
Master recherche (DEA)	Bac technique	0,000	0,001	1,598
Ecole de commerce	DUT/DUST	0,000	0,000	1,591
Bac général	Bac professionnel	0,002	0,003	1,579
Licence 2 (deug)	Pas de diplôme	0,001	0,002	1,554
Bac professionnel	Pas de diplôme	0,007	0,011	1,524
BTS	CAP/BEP	0,012	0,019	1,521
CAP/BEP	DUT/DUST	0,003	0,005	1,519
DUT/DUST	CAP/BEP	0,004	0,005	1,510
CAP/BEP	Diplôme paramédical	0,010	0,015	1,495
Licence 3	Ecole d'ingénieurs	0,000	0,000	1,491
CAP/BEP	Licence 2 (deug)	0,003	0,004	1,487
BTS	Doctorat (santé)	0,001	0,001	1,475
Bac professionnel	Licence 2 (deug)	0,001	0,001	1,456
Bac professionnel	Master 1 (maîtrise)	0,002	0,002	1,435
Diplôme paramédical	Licence 2 (deug)	0,000	0,000	1,435
Bac professionnel	Master professionnel (DESS)	0,002	0,002	1,414
Bac général	Ecole de commerce	0,000	0,000	1,404
BTS	Ecole d'ingénieurs	0,001	0,001	1,398
Licence 3	Bac professionnel	0,001	0,002	1,398
BTS	Ecole de commerce	0,001	0,001	1,376
BTS	Master recherche (DEA)	0,001	0,001	1,351
Ecole d'ingénieurs	Bac général	0,003	0,003	1,340
Doctorat (santé)	Licence 3	0,001	0,001	1,325
Bac général	Pas de diplôme	0,009	0,012	1,296
Master 1 (maîtrise)	Doctorat (santé)	0,000	0,000	1,289
Ecole de commerce	Doctorat (hors santé)	0,000	0,000	1,274
Ecole de commerce	Licence 3	0,000	0,001	1,266
Diplôme paramédical	Bac général	0,001	0,001	1,239
Master 1 (maîtrise)	BTS	0,002	0,003	1,222
Bac général	Master professionnel (DESS)	0,002	0,002	1,211
Technicien du supérieur	CAP/BEP	0,001	0,001	1,207
Pas de diplôme	Bac technique	0,011	0,013	1,206
Technicien du supérieur	Doctorat (santé)	0,000	0,000	1,199
Master 1 (maîtrise)	Autres diplômes de niveau au moins Bac+3	0,000	0,000	1,198
CAP/BEP	BTS	0,023	0,027	1,172
CAP/BEP	Bac général	0,022	0,026	1,171
Ecole de commerce	BTS	0,001	0,001	1,168
Pas de diplôme	Bac général	0,017	0,019	1,166
Licence 3	Ecole de commerce	0,000	0,000	1,153
Licence 2 (deug)	Autres diplômes de niveau au moins Bac+3	0,000	0,000	1,150
Bac technique	Licence 2 (deug)	0,000	0,000	1,144
Master professionnel (DESS)	BTS	0,002	0,003	1,126
Bac professionnel	Autres diplômes de niveau au moins Bac+3	0,001	0,001	1,110
Licence 2 (deug)	Technicien du supérieur	0,000	0,000	1,104
BTS	Capacité en droit	0,000	0,000	1,097
Doctorat (hors santé)	Bac général	0,001	0,001	1,082
Bac général	Master recherche (DEA)	0,001	0,001	1,078
Bac professionnel	Bac général	0,004	0,005	1,076
Autres diplômes de niveau au moins Bac+3	Diplôme paramédical	0,001	0,001	1,068
Bac technique	CAP/BEP	0,007	0,007	1,059
Bac général	Bac technique	0,003	0,003	1,051
Autres diplômes de niveau au moins Bac+3	Bac technique	0,001	0,001	1,047
Bac professionnel	Licence 3	0,003	0,003	1,045
Diplôme paramédical	DUT/DUST	0,000	0,000	1,043
Master 1 (maîtrise)	Bac général	0,002	0,002	1,041
Ecole d'ingénieurs	BTS	0,003	0,004	1,039
Capacité en droit	CAP/BEP	0,000	0,000	1,031
BTS	Bac général	0,007	0,007	1,030
Bac technique	DUT/DUST	0,001	0,001	1,030
Licence 3	Bac général	0,003	0,003	1,025
Pas de diplôme	Bac professionnel	0,011	0,011	1,021
Licence 2 (deug)	BTS	0,001	0,001	1,013
Bac général	DUT/DUST	0,001	0,001	1,007
Master recherche (DEA)	BTS	0,001	0,001	1,004
BTS	Master professionnel (DESS)	0,003	0,003	0,996
BTS	Bac professionnel	0,004	0,004	0,974

Bac général	Licence 3	0,004	0,004	0,966
Doctorat (hors santé)	Licence 2 (deug)	0,000	0,000	0,954
Master professionnel (DESS)	Licence 2 (deug)	0,000	0,000	0,951
Bac professionnel	CAP/BEP	0,013	0,012	0,947
Doctorat (santé)	Master 1 (maîtrise)	0,001	0,001	0,946
Autres diplômes de niveau au moins Bac+3	Bac général	0,001	0,001	0,942
Licence 2 (deug)	Bac technique	0,001	0,001	0,938
Licence 3	DUT/DUST	0,001	0,001	0,928
DUT/DUST	Bac technique	0,002	0,001	0,917
Master professionnel (DESS)	DUT/DUST	0,001	0,001	0,913
Diplôme paramédical	Master 1 (maîtrise)	0,000	0,000	0,912
BTS	Autres diplômes de niveau au moins Bac+3	0,001	0,001	0,902
DUT/DUST	Bac général	0,002	0,002	0,899
Bac général	Diplôme paramédical	0,003	0,003	0,899
Bac général	Master 1 (maîtrise)	0,003	0,003	0,899
Master 1 (maîtrise)	Doctorat (hors santé)	0,000	0,000	0,897
CAP/BEP	Pas de diplôme	0,069	0,061	0,895
Bac professionnel	DUT/DUST	0,001	0,001	0,892
Ecole de commerce	Diplôme paramédical	0,001	0,000	0,882
Bac général	BTS	0,006	0,005	0,882
Bac technique	Master recherche (DEA)	0,001	0,000	0,881
Pas de diplôme	CAP/BEP	0,056	0,049	0,874
Licence 2 (deug)	Doctorat (hors santé)	0,000	0,000	0,870
Technicien du supérieur	Diplôme paramédical	0,000	0,000	0,870
Ecole de commerce	Bac général	0,001	0,001	0,869
Autres diplômes de niveau au moins Bac+3	Licence 2 (deug)	0,000	0,000	0,869
Bac technique	Master 1 (maîtrise)	0,002	0,001	0,868
Diplôme paramédical	Master professionnel (DESS)	0,000	0,000	0,862
Ecole d'ingénieurs	Licence 3	0,003	0,002	0,861
Autres diplômes de niveau au moins Bac+3	BTS	0,002	0,001	0,855
Doctorat (hors santé)	Diplôme paramédical	0,001	0,001	0,850
Bac technique	Licence 3	0,002	0,002	0,849
Ecole de commerce	Doctorat (santé)	0,000	0,000	0,845
Licence 3	Doctorat (santé)	0,000	0,000	0,845
Doctorat (hors santé)	Ecole de commerce	0,000	0,000	0,843
Bac professionnel	Diplôme paramédical	0,003	0,003	0,841
DUT/DUST	Master professionnel (DESS)	0,001	0,001	0,838
Licence 2 (deug)	Diplôme paramédical	0,001	0,000	0,838
Technicien du supérieur	BTS	0,001	0,001	0,832
Diplôme paramédical	Bac professionnel	0,001	0,000	0,831
BTS	Bac technique	0,006	0,005	0,830
Bac technique	Bac général	0,003	0,003	0,829
Bac technique	Bac professionnel	0,002	0,002	0,813
Ecole d'ingénieurs	Licence 2 (deug)	0,001	0,001	0,810
Ecole d'ingénieurs	Diplôme paramédical	0,002	0,002	0,808
BTS	Diplôme paramédical	0,005	0,004	0,804
Licence 3	BTS	0,004	0,003	0,803
Technicien du supérieur	Ecole de commerce	0,000	0,000	0,801
Bac général	Autres diplômes de niveau au moins Bac+3	0,001	0,001	0,798
CAP/BEP	Bac technique	0,021	0,017	0,797
Master recherche (DEA)	Diplôme paramédical	0,001	0,001	0,786
Bac technique	Diplôme paramédical	0,002	0,002	0,785
Bac technique	Ecole d'ingénieurs	0,000	0,000	0,780
Bac professionnel	Bac technique	0,004	0,003	0,772
Licence 3	Diplôme paramédical	0,002	0,002	0,762
Technicien du supérieur	Master 1 (maîtrise)	0,000	0,000	0,761
BTS	Licence 2 (deug)	0,002	0,001	0,760
CAP/BEP	Bac professionnel	0,019	0,015	0,752
Diplôme paramédical	Licence 3	0,001	0,001	0,742
Doctorat (santé)	Master professionnel (DESS)	0,001	0,000	0,738
Autres diplômes de niveau au moins Bac+3	Master 1 (maîtrise)	0,001	0,001	0,733
DUT/DUST	Diplôme paramédical	0,002	0,001	0,720
Ecole de commerce	Ecole d'ingénieurs	0,000	0,000	0,709
DUT/DUST	Master 1 (maîtrise)	0,001	0,001	0,708
Licence 2 (deug)	Master 1 (maîtrise)	0,001	0,000	0,696
Licence 2 (deug)	Master professionnel (DESS)	0,001	0,000	0,696
Ecole d'ingénieurs	DUT/DUST	0,001	0,001	0,690
Master professionnel (DESS)	Technicien du supérieur	0,000	0,000	0,686
Bac professionnel	BTS	0,007	0,005	0,682
Master 1 (maîtrise)	Technicien du supérieur	0,000	0,000	0,681
Master 1 (maîtrise)	Diplôme paramédical	0,002	0,001	0,677
Licence 3	Technicien du supérieur	0,000	0,000	0,672
DUT/DUST	Licence 3	0,002	0,002	0,652
BTS	Master 1 (maîtrise)	0,006	0,004	0,647
Technicien du supérieur	Licence 3	0,001	0,000	0,645
Bac technique	BTS	0,005	0,003	0,635
Autres diplômes de niveau au moins Bac+3	DUT/DUST	0,000	0,000	0,633
BTS	Technicien du supérieur	0,001	0,000	0,631
Master recherche (DEA)	Licence 2 (deug)	0,000	0,000	0,621
DUT/DUST	BTS	0,004	0,002	0,617
Doctorat (santé)	Autres diplômes de niveau au moins Bac+3	0,000	0,000	0,611
CAP/BEP	CAP/BEP	0,108	0,065	0,606
Master recherche (DEA)	Licence 3	0,001	0,001	0,603
Bac général	Licence 2 (deug)	0,001	0,001	0,602
DUT/DUST	Autres diplômes de niveau au moins Bac+3	0,001	0,000	0,600
Master professionnel (DESS)	Doctorat (santé)	0,001	0,000	0,592
Ecole d'ingénieurs	Master recherche (DEA)	0,001	0,001	0,586
Master professionnel (DESS)	Diplôme paramédical	0,003	0,002	0,584
Bac général	Technicien du supérieur	0,000	0,000	0,576

Licence 2 (deug)	Bac général	0,001	0,001	0,576
Master 1 (maîtrise)	Ecole d'ingénieurs	0,001	0,000	0,567
Ecole de commerce	Master professionnel (DESS)	0,001	0,000	0,564
Master recherche (DEA)	DUT/DUST	0,000	0,000	0,560
BTS	Licence 3	0,009	0,005	0,560
BTS	DUT/DUST	0,003	0,001	0,560
DUT/DUST	Ecole d'ingénieurs	0,000	0,000	0,559
Licence 3	Autres diplômes de niveau au moins Bac+3	0,001	0,000	0,559
Doctorat (hors santé)	Ecole d'ingénieurs	0,000	0,000	0,558
Technicien du supérieur	Licence 2 (deug)	0,000	0,000	0,556
Licence 3	Master 1 (maîtrise)	0,003	0,001	0,547
Licence 3	Master professionnel (DESS)	0,002	0,001	0,546
Doctorat (hors santé)	Licence 3	0,001	0,001	0,543
Diplôme paramédical	Master recherche (DEA)	0,000	0,000	0,540
Technicien du supérieur	Master professionnel (DESS)	0,000	0,000	0,532
Master professionnel (DESS)	Master 1 (maîtrise)	0,003	0,001	0,530
Autres diplômes de niveau au moins Bac+3	Licence 3	0,002	0,001	0,528
Bac professionnel	Capacité en droit	0,000	0,000	0,528
Autres diplômes de niveau au moins Bac+3	Master professionnel (DESS)	0,001	0,001	0,524
Bac technique	Bac technique	0,004	0,002	0,508
DUT/DUST	Licence 2 (deug)	0,001	0,000	0,504
Licence 2 (deug)	Licence 3	0,001	0,001	0,503
Master professionnel (DESS)	Licence 3	0,004	0,002	0,501
Technicien du supérieur	Bac général	0,001	0,000	0,499
Licence 3	Licence 2 (deug)	0,001	0,000	0,497
Master 1 (maîtrise)	Ecole de commerce	0,000	0,000	0,484
Master recherche (DEA)	Ecole de commerce	0,000	0,000	0,476
Autres diplômes de niveau au moins Bac+3	Master recherche (DEA)	0,000	0,000	0,474
Pas de diplôme	Pas de diplôme	0,097	0,046	0,473
Doctorat (hors santé)	Master 1 (maîtrise)	0,001	0,000	0,472
Autres diplômes de niveau au moins Bac+3	Ecole d'ingénieurs	0,000	0,000	0,468
DUT/DUST	Master recherche (DEA)	0,001	0,000	0,468
Master recherche (DEA)	Ecole d'ingénieurs	0,000	0,000	0,465
Ecole d'ingénieurs	Technicien du supérieur	0,000	0,000	0,464
Master 1 (maîtrise)	Master recherche (DEA)	0,001	0,000	0,454
BTS	BTS	0,017	0,008	0,454
Master 1 (maîtrise)	Licence 2 (deug)	0,001	0,000	0,451
Master recherche (DEA)	Technicien du supérieur	0,000	0,000	0,447
Licence 3	Doctorat (hors santé)	0,001	0,000	0,443
Ecole d'ingénieurs	Autres diplômes de niveau au moins Bac+3	0,001	0,001	0,442
DUT/DUST	Technicien du supérieur	0,000	0,000	0,438
Licence 2 (deug)	DUT/DUST	0,000	0,000	0,434
Master recherche (DEA)	Master 1 (maîtrise)	0,001	0,000	0,432
Bac professionnel	Bac professionnel	0,006	0,003	0,426
Doctorat (santé)	Diplôme paramédical	0,001	0,001	0,424
Master 1 (maîtrise)	DUT/DUST	0,001	0,000	0,420
Doctorat (santé)	Doctorat (hors santé)	0,000	0,000	0,418
Master 1 (maîtrise)	Licence 3	0,004	0,002	0,408
Diplôme paramédical	Doctorat (santé)	0,000	0,000	0,401
Master professionnel (DESS)	Ecole de commerce	0,001	0,000	0,395
Autres diplômes de niveau au moins Bac+3	Technicien du supérieur	0,000	0,000	0,395
Ecole de commerce	Master 1 (maîtrise)	0,001	0,000	0,387
Master recherche (DEA)	Autres diplômes de niveau au moins Bac+3	0,000	0,000	0,375
Licence 2 (deug)	Master recherche (DEA)	0,000	0,000	0,371
Ecole de commerce	Master recherche (DEA)	0,000	0,000	0,368
Bac général	Bac général	0,014	0,005	0,365
Ecole d'ingénieurs	Master 1 (maîtrise)	0,005	0,002	0,363
Doctorat (santé)	Master recherche (DEA)	0,000	0,000	0,360
Autres diplômes de niveau au moins Bac+3	Doctorat (hors santé)	0,000	0,000	0,358
Technicien du supérieur	Autres diplômes de niveau au moins Bac+3	0,000	0,000	0,353
Master professionnel (DESS)	Autres diplômes de niveau au moins Bac+3	0,001	0,000	0,351
Doctorat (hors santé)	Master professionnel (DESS)	0,001	0,000	0,351
Master professionnel (DESS)	Doctorat (hors santé)	0,001	0,000	0,350
Doctorat (santé)	Ecole de commerce	0,000	0,000	0,348
Ecole d'ingénieurs	Doctorat (santé)	0,001	0,001	0,347
Master professionnel (DESS)	Ecole d'ingénieurs	0,001	0,000	0,333
Autres diplômes de niveau au moins Bac+3	Doctorat (santé)	0,001	0,000	0,330
Master 1 (maîtrise)	Master professionnel (DESS)	0,004	0,001	0,316
Ecole d'ingénieurs	Master professionnel (DESS)	0,005	0,002	0,313
Licence 3	Master recherche (DEA)	0,002	0,000	0,302
Doctorat (hors santé)	Doctorat (santé)	0,000	0,000	0,292
Licence 3	Licence 3	0,007	0,002	0,279
Master professionnel (DESS)	Master recherche (DEA)	0,002	0,000	0,267
Licence 3	Capacité en droit	0,000	0,000	0,246
Ecole d'ingénieurs	Ecole de commerce	0,001	0,000	0,235
Master recherche (DEA)	Doctorat (hors santé)	0,000	0,000	0,229
Ecole de commerce	Technicien du supérieur	0,000	0,000	0,227
Master 1 (maîtrise)	Master 1 (maîtrise)	0,005	0,001	0,224
Capacité en droit	Bac technique	0,000	0,000	0,223
Master recherche (DEA)	Master professionnel (DESS)	0,002	0,000	0,210
Ecole d'ingénieurs	Doctorat (hors santé)	0,001	0,000	0,208
DUT/DUST	DUT/DUST	0,002	0,000	0,202
Licence 2 (deug)	Licence 2 (deug)	0,001	0,000	0,192
Ecole de commerce	Autres diplômes de niveau au moins Bac+3	0,001	0,000	0,174
Doctorat (hors santé)	Master recherche (DEA)	0,001	0,000	0,163
Capacité en droit	Bac général	0,000	0,000	0,160
Autres diplômes de niveau au moins Bac+3	Ecole de commerce	0,001	0,000	0,160
Master professionnel (DESS)	Master professionnel (DESS)	0,008	0,001	0,150
Diplôme paramédical	Diplôme paramédical	0,004	0,000	0,119

Bac général	Capacité en droit	0,000	0,000	0,113
Master recherche (DEA)	Master recherche (DEA)	0,002	0,000	0,102
Autres diplômes de niveau au moins Bac+3	Autres diplômes de niveau au moins Bac+3	0,002	0,000	0,092
Ecole d'ingénieurs	Ecole d'ingénieurs	0,005	0,000	0,092
Technicien du supérieur	Technicien du supérieur	0,000	0,000	0,064
Doctorat (hors santé)	Doctorat (hors santé)	0,003	0,000	0,034
Ecole de commerce	Ecole de commerce	0,002	0,000	0,032
Doctorat (santé)	Doctorat (santé)	0,005	0,000	0,030
Pas de diplôme	Capacité en droit	0,000	0,000	0,000
Bac professionnel	Doctorat (hors santé)	0,000	0,000	0,000
Bac technique	Capacité en droit	0,000	0,000	0,000
Bac technique	Technicien du supérieur	0,000	0,000	0,000
Capacité en droit	Pas de diplôme	0,000	0,000	0,000
Capacité en droit	Bac professionnel	0,000	0,000	0,000
Capacité en droit	Capacité en droit	0,00	0,00	0,00
Capacité en droit	Licence 2 (deug)	0,00	0,00	0,00
Capacité en droit	DUT/DUST	0,00	0,00	0,00
Capacité en droit	BTS	0,00	0,00	0,00
Capacité en droit	Technicien du supérieur	0,00	0,00	0,00
Capacité en droit	Diplôme paramédical	0,00	0,00	0,00
Capacité en droit	Licence 3	0,00	0,00	0,00
Capacité en droit	Autres diplômes de niveau au moins Bac+3	0,00	0,00	0,00
Capacité en droit	Master 1 (maîtrise)	0,00	0,00	0,00
Capacité en droit	Master recherche (DEA)	0,00	0,00	0,00
Capacité en droit	Master professionnel (DESS)	0,00	0,00	0,00
Capacité en droit	Ecole de commerce	0,00	0,00	0,00
Capacité en droit	Ecole d'ingénieurs	0,00	0,00	0,00
Capacité en droit	Doctorat (hors santé)	0,00	0,00	0,00
Capacité en droit	Doctorat (santé)	0,00	0,00	0,00
Licence 2 (deug)	Capacité en droit	0,000	0,000	0,000
Licence 2 (deug)	Ecole de commerce	0,000	0,000	0,000
DUT/DUST	Capacité en droit	0,000	0,000	0,000
DUT/DUST	Doctorat (santé)	0,000	0,000	0,000
Technicien du supérieur	Bac professionnel	0,000	0,000	0,000
Technicien du supérieur	Capacité en droit	0,000	0,000	0,000
Technicien du supérieur	Ecole d'ingénieurs	0,000	0,000	0,000
Technicien du supérieur	Doctorat (hors santé)	0,000	0,000	0,000
Diplôme paramédical	Capacité en droit	0,000	0,000	0,000
Diplôme paramédical	Technicien du supérieur	0,000	0,000	0,000
Diplôme paramédical	Doctorat (hors santé)	0,000	0,000	0,000
Autres diplômes de niveau au moins Bac+3	Capacité en droit	0,000	0,000	0,000
Master 1 (maîtrise)	Capacité en droit	0,000	0,000	0,000
Master recherche (DEA)	Bac professionnel	0,000	0,000	0,000
Master recherche (DEA)	Capacité en droit	0,000	0,000	0,000
Master professionnel (DESS)	Capacité en droit	0,000	0,000	0,000
Ecole de commerce	Capacité en droit	0,000	0,000	0,000
Ecole d'ingénieurs	Capacité en droit	0,000	0,000	0,000
Doctorat (hors santé)	Capacité en droit	0,000	0,000	0,000
Doctorat (hors santé)	DUT/DUST	0,000	0,000	0,000
Doctorat (hors santé)	Technicien du supérieur	0,000	0,000	0,000
Doctorat (santé)	Capacité en droit	0,000	0,000	0,000
Doctorat (santé)	Licence 2 (deug)	0,000	0,000	0,000
Doctorat (santé)	Technicien du supérieur	0,000	0,000	0,000

Source : Enquête Revenus Fiscaux et Sociaux 2011 (Insee) – Calcul des auteurs.

Champ : ménages dont la personne de référence et son conjoint ont entre 25 et 54 ans.

Note : les appariements sous la ligne horizontale sont sous-représentés par rapport à ce qui serait observé dans le cas d'un appariement aléatoire ; les appariements en grisé ne sont pas observés dans l'enquête ($p_0 = 0$).

Annexes 3 : Le mode d'imputation des enfants par microsimulation

D'un point de vue technique, la simulation de la situation parentale des individus est une sous-partie du module *démographique* (Courtioux et Lignon [2015], Lignon [2014]). Elle intervient juste après la simulation de la situation conjugale et suit exactement le même principe. Après avoir déterminé la situation conjugale des individus à l'âge $a+1$, nous estimons, pour chaque individu de la génération, la probabilité relative d'être parent compte tenu de certaines caractéristiques (sexe, diplôme, statut conjugal l'année précédente, nombre et âge des enfants, position sur le marché du travail, etc.). La position des individus est ensuite déterminée par un calage des probabilités individuelles sur une cale qui estime la part d'individus de la génération 70 ayant au moins un enfant à un âge donné. Par convention, on suppose que chaque nouvel enfant arrivant dans le ménage de l'individu de référence correspond à une naissance. Pour les individus déjà parents, on détermine la probabilité de connaître une nouvelle naissance. Dans la simulation, nous déterminons s'il y a ou non une nouvelle naissance sur la base d'un tirage aléatoire (sans effectuer de calage sur marge). Au terme de la simulation, il est donc possible d'identifier à chaque étape du cycle de vie si les individus rencontrent une naissance et, dans le cas où ils sont parents, de calculer l'âge de leurs enfants. Nous poursuivons la présentation de la simulation des trajectoires parentales en fournissant des précisions sur (i) la construction des cales générationnelles et (ii) l'estimation des probabilités relatives d'avoir des enfants.

Construction des cales générationnelles

Pour tenir compte des effets de cohorte dans notre simulation, nous construisons et utilisons des cales générationnelles. Comme pour la modélisation des taux agrégés par âge de mise en couple (cf. Courtioux et Lignon [2015]) nous nous basons sur les comportements matrimoniaux observés pour les générations nées sur la période 1910-1994 dans les enquêtes Emploi disponibles pour la période 1969-2010. L'objectif est de modéliser Y , le taux d'individus ayant au moins un enfant de moins de 18 ans, en fonction de l'âge et d'une tendance (*trend*) générationnel. De plus, nous introduisons dans le modèle le taux de chômage courant pour capter l'effet de l'incertitude sur le marché du travail. Nous différencions également nos estimations par sexe compte tenu des différences qui caractérisent les profils masculins et féminins.

Pour les hommes, la modélisation inclut l'âge ($t - g$) avec des moments d'ordre 1 à 4. Le *trend* générationnel est appréhendé par la différence ($g - 1970$). Nous croisons également l'âge et l'âge au carré avec ce *trend* pour capter l'effet de retard au fil des générations dont le seul *trend* ne rend pas compte. Nous contrôlons enfin par le taux de chômage courant annuel (u) sa forme quadratique. Le modèle estimé pour les hommes peut dès lors s'écrire comme suit :

$$\text{Log} (Y_{gt} / (1 - Y_{gt})) = \alpha + \beta (t-g) + \gamma (t-g)^2 + \delta (t-g)^3 + \phi (t-g)^4 + \mu (g - 1970) + \omega (t-g) (g - 1970) + \lambda (t-g)^2 (g - 1970) + u (u_t) + \psi (u_t)^2 + \varepsilon_{gt}$$

Où g caractérise une génération et t l'année de l'observation.

Pour les femmes, la modélisation inclut l'âge élevé aux ordres 1 à 3²⁹. Le *trend* générationnel est appréhendé par la différence ($g - 1970$). Nous croisons ensuite l'âge et l'âge au carré avec le *trend*. Nous contrôlons enfin nos estimations par le taux de chômage sans inclure sa forme quadratique car le coefficient associé n'est pas significatif. Le modèle retenu pour les femmes peut dès lors s'écrire comme suit :

$$\text{Log} (Y_{gt} / (1 - Y_{gt})) = \alpha + \beta (t-g) + \gamma (t-g)^2 + \delta (t-g)^3 + \mu (g - 1970) + \omega (t-g) (g - 1970) + \lambda (t-g)^2 (g - 1970) + u (u_t) + \varepsilon_{gt}$$

Les coefficients obtenus pour ces modèles sont présentés dans le tableau A3.1 et le graphique A3.2 représentent les cales utilisées dans la simulation.

²⁹ Le moment d'ordre 4 n'est pas significatif contrairement aux hommes.

Tableau A3.1. - Modélisation des taux agrégés de parentalité

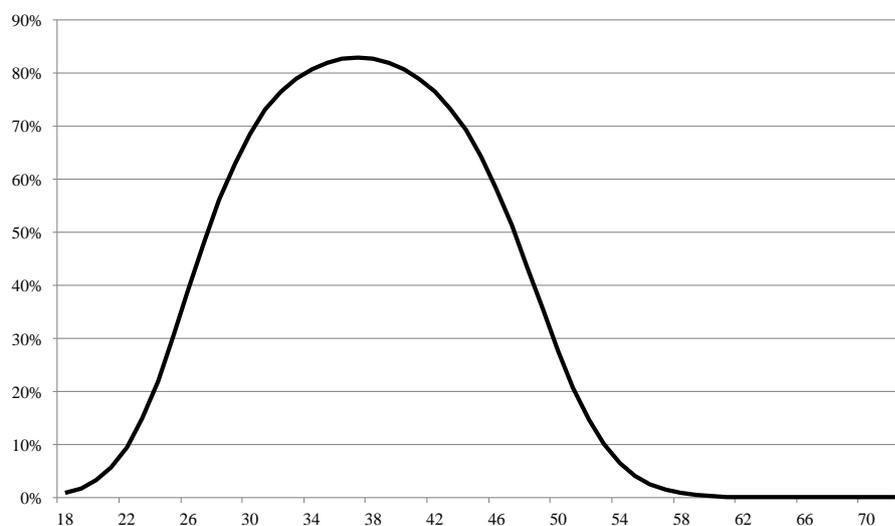
	Femmes	Hommes
Constante	-23,67	-31,52
Age	1,51	2,26
Age ²	-0,02	-0,05
Age ³	7,98E-05	4,84E-04
Age ⁴	-	-1,72E-06
Trend générationnel	-0,3	-0,13
Age*Trend générationnel	0,015	0,005
Age ² *Trend générationnel	-1,73E-04	-5,39E-05
Taux de chômage courant	-6,97E+00	-22,65
Taux de chômage courant au carré	-	108,91
<i>R</i> ²	0,96	0,94
<i>Obs.</i>	2 734	2 646

Source : enquêtes Emploi 1969-2010 (Insee) - calculs des auteurs.

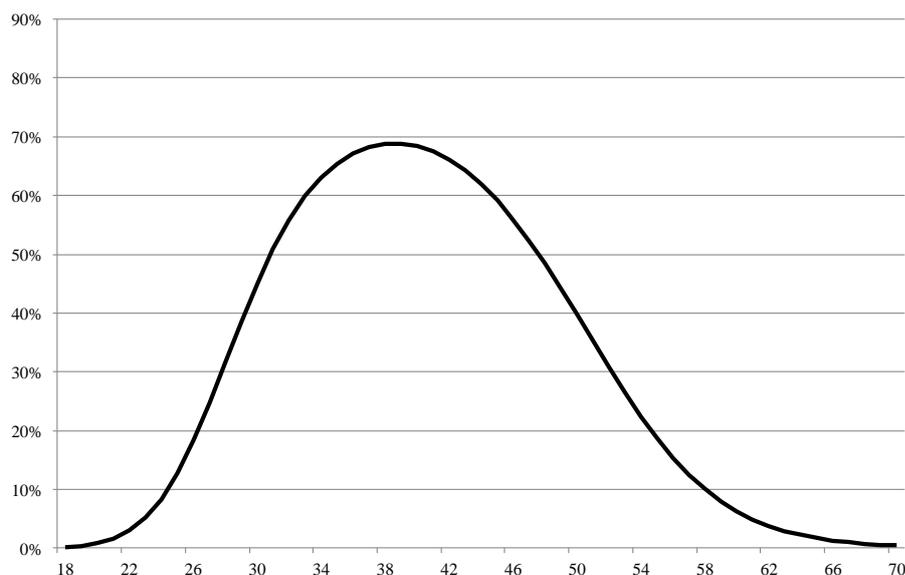
Note : tous les coefficients sont significatifs au seuil de 1% ; la génération de référence est celle née en 1970.

Graphique A3.2. - Part d'individus parents d'enfants de moins de 18 ans selon l'âge (simulation de cales agrégées pour la génération 1970 avec une hypothèse de taux de chômage de 8%)

Les femmes



Les hommes



Source : enquêtes Emploi 1969-2010 (Insee) - calculs des auteurs.

Estimation des probabilités de transition

Pour simuler les naissances, nous nous appuyons sur des modèles logistiques. Nous retenons trois types de modèles : *i*) un modèle qui estime la probabilité d'entrer en parentalité en $n+1$ pour les individus sans enfant en n (cf. tableau A3.3.), *ii*) un modèle qui estime la probabilité de rester parent en $n+1$ pour les pères et mères en n (cf. tableau A3.4), et *iii*) un modèle qui estime la probabilité de connaître une nouvelle naissance sachant que les individus sont déjà parents (cf. tableau A3.5).

Tableau A3.3. Estimation de la probabilité d'être parent en $n+1$ pour les individus non parents en n

	Femmes	Hommes
Constante	-10,39	-8,16
Age	0,46	0,32
Age²	-5,18E-03	-5,11E-03
Age*Fin des études	0,19	0,14
Age²*Fin des études	-5,88E-03	-2,49E-03
Situation conjugale (entre n et $n+1$) et professionnelle (en n)		
En couple		
<i>Individu et conjoint en emploi</i>	<i>réf.</i>	<i>réf.</i>
<i>Individu en emploi et conjoint hors emploi</i>	-0,05	0,12
<i>Individu hors emploi et conjoint en emploi</i>	0,38	-0,05
<i>Individu hors emploi et conjoint hors emploi</i>	-1,12	-0,95
Célibataire entre n et $n+1$		
<i>Emploi</i>	-2,44	-3,44
<i>Hors emploi</i>	-1,48	-2,21
Mise en couple	0,13	-1,78
Séparation	-2,29	-14,54
Capital humain	-0,03	-2,36E-03
<i>D de Somers</i>	0,89	0,86
<i>Paires concordantes (en %)</i>	93,8	91,6
<i>Obs.</i>	7965	6736

Source : enquêtes SRCV 2004-2009 (Insee) - calculs des auteurs.

Champ : ensemble des individus déclarant n'avoir aucun enfant de moins de 18 ans en n .

Note : les coefficients reproduits sont significatifs au seuil de 1%.

Dans nos estimations de probabilité de devenir père ou mère (cf. tableau A3.3.), nous introduisons l'âge et l'âge au carré afin de capter l'augmentation du niveau de fécondité sur la première partie du cycle de vie et sa décroissance sur la seconde partie. L'effet de la fin des études est appréhendé à partir d'un croisement entre l'âge et une indicatrice de fin d'étude. Pour ce modèle, il est possible de retenir une spécification relativement fine de l'interaction entre la situation matrimoniale des individus (en couple, célibataire, mise en couple ou séparation dans l'année) et leur situation professionnelle (y compris celle de leur éventuel conjoint). Nous créons donc 8 catégories les plus précises possible permettant d'apprécier ces effets. Enfin, nous introduisons une variable de capital humain³⁰ pour capter l'effet de la durée des études (y compris pour les étudiants). Nos résultats montrent que l'effet de l'âge est positif et croissant sur la première partie du cycle de vie mais décroît ensuite. Pour les femmes ayant terminé leurs études, la probabilité de devenir parent est, toutes choses égales par ailleurs, de 5% à 16 ans (6% pour les hommes), elle atteint un pic de 28% à 29 ans (pic de 25% pour les hommes à 31 ans) et décroît pour tendre vers 0 à partir de 50 ans (52 ans pour les hommes). En cours d'étude, les chances d'entrée en parentalité sont faibles. La proportion d'hommes concernés quel que soit l'âge est minime et ne dépasse jamais celle des individus qui ont terminé leurs études. Pour les femmes, la probabilité de devenir mère en cours d'étude augmente avec l'âge : elle reste en deçà de celle des femmes ayant terminé leurs études jusque 36 ans (peu de femmes sont cependant concernées au-delà de cet âge). Ce résultat traduit d'une part le retard des naissances avec l'allongement des études et d'autre part le rôle

³⁰ Correspondant à la différence entre l'âge d'entrée sur le marché du travail et l'âge légal de fin d'étude (16 ans).

prépondérant joué par la situation économique des hommes sur les décisions de fécondité au sein du couple (Robert-Bobée et Mazuy [2005] ; Ní Brohlcháin et Beaujouan [2012]). On remarque ensuite qu'être en emploi avec un conjoint en emploi a des effets positifs sur la probabilité d'être parent : nos estimations confirment l'idée selon laquelle la stabilité du couple et de la situation professionnelle sont des déterminants importants de l'entrée en parentalité (Pailhé et Solaz [2007]). Deux exceptions sont à noter cependant. Premièrement, pour les hommes, le fait d'être en emploi avec un conjoint hors emploi a une influence positive par rapport à la situation de référence : on peut penser que les conjointes ont une attitude de retrait du marché du travail parallèlement à l'arrivée d'un enfant (congé maternité, etc.). Cette supposition semble confirmée par le coefficient « individus hors emploi conjoint en emploi » des femmes. Deuxièmement, on observe que la mise en couple favorise l'entrée en parentalité des femmes. Ce coefficient est susceptible de capter un effet « remise en couple » : les femmes n'ayant pas eu d'enfants dans le cadre d'une première union ont rapidement un enfant lorsqu'elles se remettent en couple compte tenu des « contraintes biologiques » qui pèsent sur leur calendrier de naissance. On sait par ailleurs que la séparation est corrélée négativement avec l'entrée en parentalité (Toulemon [1994]), en particulier pour les hommes : si une naissance survient dans un contexte de rupture d'union, on peut penser que l'enfant sera confié en priorité à la mère. Enfin, un haut niveau d'éducation est corrélé à une plus faible probabilité de connaître une naissance indépendamment de la situation sur le marché du travail. Cet effet est surtout important pour les femmes conformément à ce que suggèrent les travaux de Prioux *et al.* [2010]. Le coefficient est également négatif pour les hommes mais de moindre intensité. Les effets de la situation professionnelle et du niveau d'éducation sont surtout importants en début de cycle de vie : ils diminuent très fortement par la suite, supplantés par l'effet d'âge.

Le tableau A3.4 présente les coefficients associés au modèle de parentalité pour les individus parents en n . Ce modèle cherche principalement à décrire la sortie de la parentalité en fonction de certaines variables explicatives. La sortie de la parentalité, telles que nous la définissons (ne plus avoir d'enfant dans le ménage), peut être due à l'émancipation des enfants (départ des enfants du foyer pour poursuivre des études ou travailler) et à des modifications de la situation conjugale des individus (la séparation peut par exemple impliquer une garde partagée ou exclusive). Dans notre modèle, nous introduisons donc l'âge du plus jeune enfant pour capter l'effet d'émancipation. Nous prenons en compte la situation conjugale en distinguant quatre configurations : vie en couple en n et $n+1$, célibat en n et $n+1$, mise en couple dans l'année, et séparation dans l'année. Ces variables ont pour but de saisir l'effet des transitions conjugales. Suite à une rupture, il est possible que la garde des enfants par l'un ou l'autre des conjoints dépende, au-delà du sexe, des situations professionnelles des individus du couple. Nous avons testé cette hypothèse : il en ressort que le statut d'activité est uniquement significatif pour les hommes. Nous différencions donc l'événement « séparation » pour les hommes suivant qu'il y a eu ou non une perte d'emploi. Enfin, nous introduisons une variable de capital humain pour capter l'effet du diplôme.

Tableau A3.4. Estimation de la probabilité d'être parent en $n+1$ pour les individus parents en n

	Femmes	Hommes
Constante	13,05	10,92
Age du plus jeune enfant	-0,72	-0,59
Situation conjugale entre n et $n+1$		
En couple	<i>réf.</i>	<i>réf.</i>
Célibataire	0,08	-1,32
Mise en couple	-5,54	-1,59
Séparation	-1,44	
<i>Avec perte d'emploi</i>		-10,30
<i>Sans perte d'emploi</i>		-7,95
Capital Humain	0,02	9,66E-03
<i>D de Somers</i>	0,92	0,94
<i>Paires concordantes (en %)</i>	95	96,2
<i>Obs.</i>	4567	4027

Source : enquête SRCV 2004-2009 (Insee) - calculs des auteurs.

Champ : ensemble des individus déclarant avoir au moins un enfant de moins de 18 ans en n .

Note : les coefficients reproduits sont significatifs au seuil de 1%.

Les coefficients estimés montrent tout d'abord que, toutes choses égales par ailleurs, la probabilité de ne plus avoir d'enfant dans le ménage dépend fortement de l'âge du plus jeune enfant. Pour les hommes en couple sur les deux périodes et avec au moins un enfant en n , la probabilité d'être encore père en $n+1$ est quasiment de 100% lorsque l'enfant le plus jeune a entre 0 et 11 ans³¹, puis commence à décroître pour la valeur 13 ans (96%), de 57% pour 17 ans (départs des enfants du ménage pour travailler ou étudier) et moins de 6 % pour 23 ans. Pour les femmes, l'évolution est quasi-similaire. Concernant les aspects familiaux, on remarque que la situation de couple a un effet positif comparé aux autres catégories à l'exception de la situation de célibataire pour les femmes. Toutes choses égales par ailleurs, on constate que les célibataires ont plus de chances de rester en parentalité mais cet effet est trop faible pour être interprété. La mise en couple a un effet négatif pour les deux sexes mais plus fort pour les femmes : on peut penser que les femmes qui se remettent en couple n'ont plus d'enfant dans le ménage compte tenu de l'influence négative de la présence d'enfant sur leur probabilité de mise en couple. En outre, on observe que la séparation diminue fortement la probabilité d'avoir un enfant dans le ménage, notamment pour les hommes qui perdent leur emploi. Si la naissance d'un enfant a lieu en même temps que la séparation de ses parents, alors on peut supposer qu'il sera gardé par sa mère, et ce, d'autant que la situation économique du père est instable. Enfin, un haut niveau de diplôme est associé à une plus forte probabilité de rester parents. De ce point de vue, on peut supposer que les enfants d'individus diplômés poursuivent davantage d'études que les autres et restent donc plus longtemps dans leur foyer. Dans ce cadre, il est envisageable que nous captions ici un effet « reproduction sociale ».

Notre troisième et dernier modèle sur les transitions parentales cherche à modéliser la probabilité de survenue d'une nouvelle naissance pour les individus déjà parents l'année précédente. Comme pour le modèle du tableau A3.3, nous introduisons l'âge et l'âge au carré. L'impact de l'allongement de la durée des études a été testé par un croisement de l'âge avec une indicatrice de fin d'étude. Ce croisement s'est avéré non significatif comme le suggèrent les travaux de Toulemon *et*

³¹ Dans ce cas, on peut supposer que les sorties à un niveau aussi faible correspondent principalement à des décès.

al. [2008] : le retard de premières naissances lié à l'augmentation du niveau d'étude ne se traduit pas par un retard sur les naissances de rang supérieur. La variable « nombre d'enfants de moins de 18 ans » cherche à saisir l'effet du rang de naissance. Pour appréhender la question relative à l'espacement entre les naissances, nous intégrons dans notre modèle la variable d'âge du plus jeune enfant et sa forme quadratique. Concernant, la situation professionnelle, une spécification précise conduit à des coefficients non significatifs exceptés pour les catégories « chômage en n et $n+1$ » pour les hommes et les comportements de reprise ou retrait d'activité pour les femmes. Nous conservons donc ces modalités dans nos estimations. De même, pour la situation conjugale, l'effet est non significatif pour les hommes (aucun célibataire ne connaît de naissances nouvelles³²), tandis qu'il est significatif pour les femmes en couple et celles qui se sont mises en couple dans l'année.

Tableau A3.5. Modèle de naissance pour les individus parents en n

	Femmes	Hommes
Constante	-11,09	-4,33
Age	0,69	0,28
Age²	-0,01	-0,01
Nombre d'enfants de moins de 18 ans	-0,95	-1,06
Age du plus jeune enfant	0,18	0,25
Age du plus jeune enfant au carré	-0,02	-0,03
Transitions professionnelles		
Chômage en n et $n+1$		-0,74
Retrait de l'activité	0,83	
Reprise de l'activité	-1,63	
Situation Conjugales		
En couple en n	1,06	
En couple en n suivi d'une mise en couple	0,97	
Capital humain		
Capital humain	-0,10	0,01
Capital humain au carré	6,90E-03	5,98E-04
<i>D de Somers</i>	0,7	0,67
<i>Paires concordantes (en %)</i>	84,6	83,1
<i>Obs.</i>	4567	3873

Source : enquête SRCV 2004-2009 (Insee) - calculs des auteurs.

Champ : ensemble des individus déclarant avoir au moins un enfant de moins de 18 ans en n . Les hommes célibataires sont exclus de ces modélisations.

Note : les coefficients reproduits sont significatifs au seuil de 1%.

Le tableau A3.5 montre que la probabilité de naissance de rang supérieur est fortement conditionnée par l'âge des individus : elle croît sur le début du cycle de vie et décroît ensuite. Le nombre d'enfants préalablement présents dans le foyer a une influence négative sur la probabilité d'une nouvelle naissance. Toutes choses égales par ailleurs, à 30 ans, avoir 2 enfants dans le ménage diminue la probabilité de naissance de 10 points pour les hommes (8 points pour les femmes) par rapport au fait de n'avoir qu'un enfant. Ce résultat est cohérent avec les chiffres disponibles sur les taux conjoncturels de fécondité fournis par l'Insee. On constate ensuite que l'âge du plus jeune enfant a un effet positif croissant jusque 6 ans et décroissant au-delà de ce seuil quel que soit le sexe. Ce résultat suggère l'existence de comportements d'espacement des naissances : les individus auraient

³²Nous excluons donc cette possibilité dans la simulation.

tendance à attendre que le dernier né grandisse avant d'avoir un nouvel enfant sans pour autant souhaiter une différence d'âge importante. Si on s'intéresse à l'effet de l'activité professionnelle, on remarque qu'une situation de chômage durable défavorise la probabilité de naissance de rang supérieur pour les hommes. Ce résultat est cohérent avec celui de Pailhé et Solaz [2007] qui montrent que, pour les hommes, les périodes d'instabilité sur le marché du travail coïncident rarement avec la naissance d'un enfant (ceci en raison de leur rôle de premier pourvoyeur de revenu au sein du couple). Pour les femmes, on constate que le coefficient associé au retrait de l'activité est positif et inversement pour celui qui concerne la reprise d'activité. A ce titre, Pailhé et Solaz [2007] montrent que, suite à une naissance, les femmes ont tendance à stopper leur activité professionnelle (pour s'occuper de l'enfant, etc.). Par extension, on peut penser qu'elles reprennent une activité lorsqu'elles n'ont plus de projet d'enfant ou quand le dernier né est pris en charge par le système éducatif (Thévenon [2009]). Enfin, on remarque que l'effet du capital humain est positif pour les hommes et suit une forme en U pour les femmes. Pour les hommes, ce coefficient peut refléter la sécurité économique procurée par un niveau de diplôme élevé. Pour les femmes, on retrouve ici encore l'ambiguïté du statut socioéconomique (diplôme et activité).