

Communication aux XVIèmes Journées d'Etude
« Les données longitudinales dans l'analyse du marché du travail »
4 et 5 juin 2009, Paris

L'analyse socio-économique des cheminements longs : l'apport de la microsimulation dynamique dans le champ de l'éducation

Pierre Courtioux
EDHEC Business School
Délégation de Paris
12 bis rue de la Victoire
75009 Paris
Tel : +33(0)1 53 32 76 44
pierre.courtioux@edhec.edu

Dede Houeto
Paris School of Economics
48 boulevard Jourdan
75014 Paris
Tel : +33(0)1 43 13 63 22
houeto@pse-ens.fr

Juin 2009

Résumé :

Cet article montre comment il est possible, de produire une description stylisée de la diversité des trajectoires individuelles d'une génération en fonction du diplôme des individus qui la composent à l'aide d'un modèle de microsimulation dynamique. Ce cadre permet de produire une description de la structure des risques de transition entre système éducatif et marché du travail. Dans une perspective exploratoire, on développe une analyse de la distribution des taux de rendement interne individuels. Elle nous conduit à confirmer certains résultats sur les rendements différenciés de l'éducation à un niveau donné selon la filière suivie et d'opposer ainsi les diplômes universitaires à ceux des écoles (BTS, école de commerce, école d'ingénieur). Elle nous permet également de discuter de la structure incitative sous-jacente aux liens entre le marché du travail et le système éducatif : pour les lycéens issus des familles modestes plus sensibles au rendement monétaire de leur formation, le choix de s'arrêter au Bac apparaît plus rentable même s'il s'avère plus risqué.

Codes JEL : C15, I2

1-Introduction

En général, l'analyse socio-économique des cheminements longs insiste sur l'importance de l'accès à des données de panel ; le coût financier et le temps incompressible lié à la construction de telles données limitent de fait le développement de l'analyse empirique des trajectoires individuelles sur longue période. Or, dans le champ de l'éducation, la perspective qui conduit à s'intéresser à l'ensemble de la « trajectoire de vie » (*life course*)¹ est intéressante dans la mesure où les faits stylisés indiquent l'existence d'un lien entre le niveau de diplôme et le niveau de salaire, le risque chômage, mais également l'espérance de vie. De ce point de vue, une analyse en coupe ne permet de rendre qu'imparfaitement compte des liens entre ces différentes dimensions.

L'insistance sur l'importance des données de panel est également une des caractéristiques de la littérature portant sur les méthodes dites de microsimulation. Par exemple, Legendre *et alii* (2003) indiquent que si les données en coupe instantanée peuvent constituer le point de départ d'un modèle de microsimulation, dans la mesure où la simulation cherche à retracer des trajectoires individuelles, il faut disposer de données pour lesquels les individus sont observés plusieurs fois : les données de panel. Le choix d'un champ d'analyse pour la microsimulation n'est pas sans effet sur la modélisation. D'un point de vue méthodologique, Blanchet (1998) rappelle que compte tenu de l'équilibre nécessaire entre le « souci du détail » et le caractère gérable du modèle, « l'ambition [d'une telle modélisation] ne peut être de déboucher sur une modélisation « totale » visant à reproduire exhaustivement la diversité des situations et comportements individuels et le va-et-vient entre ce comportement individuel et les contraintes macroéconomiques ou macrosociales ». Ceci indique que la structure d'un modèle de microsimulation doit correspondre à la dimension particulière que l'on souhaite étudier. En France, la microsimulation dynamique, c'est-à-dire celle qui est particulièrement intéressée par les trajectoires longues, s'est principalement développée dans le champ des retraites : le modèle Destinie de l'Insee² et le modèle Prisme³ de la Cnav en constituent deux exemples.

Dans ce cadre, notre contribution vise à montrer comment une approche par microsimulation dynamique fournit une méthode intermédiaire d'analyse des trajectoires longues dans le champ de l'éducation : sans diminuer l'intérêt de la constitution et de l'analyse de données de panel, la simulation de la diversité de trajectoires individuelles permet de dépasser une simple analyse en coupe en fournissant des indicateurs synthétiques en termes de risques sur les

¹ Pour une présentation plus générale de l'approche en termes de trajectoires de vie, voir notamment Anxo et Erhel (2008).

² Voir Insee (1999) pour une présentation.

³ Voir Poubelle *et alii* (2006) pour une présentation

trajectoires individuelles qui permettent de mettre en relation plusieurs dimensions des trajectoires longues dans une perspective descriptive. L'originalité du modèle de microsimulation que nous proposons ici tient principalement au champ traité et à la perspective retenue : le champ porte sur les liens entre éducation et marché du travail ; la perspective est générationnelle alors que les modèles de microsimulation dynamique précédemment cités traitent de l'ensemble de la population.

Dans ce papier on cherche à caractériser les risques de transition entre système éducatif et marché du travail. On s'intéresse alors principalement aux risques de carrières liés aux choix éducatifs dont la distribution des rendements internes individuels de l'éducation constitue un indicateur synthétique – la part des rendements individuels négatifs apparaît alors comme un indicateur original central pour rendre compte de ces risques. De ce point de vue, notre démarche s'inscrit dans la recherche d'indicateurs pertinents pour identifier et gérer les risques sociaux dans une perspective du type de celle des *marchés transitionnels de l'emploi* – pour une présentation voir par exemple Gazier et Schmid (2002), Schmid (2006). Cette perspective insiste sur l'importance des transitions sur le marché du travail appréhendées tout au long de la trajectoire de vie. Elle insiste également sur la nécessité de prendre en compte les différentes institutions qui soutiennent ces transitions. Dans ce cadre, pour rendre compte de la spécificité du système éducatif français et de ses liens avec le marché du travail, on retient une différenciation des diplômes aussi précise que possible permettant notamment de distinguer à même niveau d'éducation les diplômes universitaires de ceux délivrés par des écoles (BTS, écoles d'ingénieur, etc.). Le choix de cette perspective consiste alors à prolonger les analyses socio-économiques en termes d'*effet sociétal*⁴ quant à l'articulation du marché du travail et du système éducatif en introduisant la problématique des risques de carrières comme risques sociaux. Ce choix nous a conduit à privilégier deux dimensions techniques : (i) la maniabilité du modèle et (ii) la prise en compte la plus fine possible de l'hétérogénéité individuelle en termes de diplôme. Ceci se traduit par l'utilisation d'un « minimum » de données : on s'appuie alors principalement sur les données de l'enquête Emploi et sur les tables de mortalité disponibles.

La première section explique la méthode utilisée pour simuler la diversité des trajectoires de vie. La deuxième section illustre la perspective de gestion des risques sociaux ouverte par cette méthodologie à partir de la construction d'indicateurs synthétiques de la diversité des trajectoires de vie.

⁴ Au sens de Maurice *et alii* (1982).

2-Simuler la diversité des trajectoires pour une génération à partir de données en coupes

Le modèle de que nous utilisons vise à simuler la diversité des chroniques individuelles de revenu tout au long de la vie pour une génération donnée sur la base d'un petit nombre de caractéristiques individuelles connues. Pour rendre compte du lien entre formation initiale et emploi, nous avons choisi de différencier les individus sur la base de trois variables : le sexe, le diplôme et l'âge d'entrée sur le marché du travail. Nous avons retenu la nomenclature de diplôme la plus fine disponible dans l'enquête Emploi 2003-2005 : elle comprend vingt-et-un postes. La génération née en 1970 a été choisie pour illustrer cette hétérogénéité de départ et la diversité des trajectoires qui s'en suit. Le tableau 1 rappelle les caractéristiques de cette génération telles que l'enquête Emploi en Continu permet d'en rendre compte.

De manière générale, notre méthode de simulation repose sur la mobilisation de différents modèles (transition, salaire, etc.), ce qui nous conduit à distinguer l'*estimation* de ces modèles et la *simulation* sur la base de ces modèles. En effet, certaines variables sont introduites dans les estimations pour identifier l'effet des autres variables toutes choses égales par ailleurs ; ces variables ne sont pas forcément introduite dans la simulation. Les raisons qui sous-tendent ces choix sont alors expliquées au fil du texte.

Dans un premier temps, on précise comment appréhender les trajectoires sur le marché du travail. Dans un deuxième temps, on explique la manière dont il est possible de prendre en compte les différentiels de mortalité selon le diplôme.

Tableau 1 : Distribution du niveau de diplôme

Diplôme	Part
Pas de Bac	53.4%
Pas de diplôme	25.5%
CAP/BEP	27.9%
Niveau Bac	14.5%
Bac Général	6.6%
Bac Professionnel	3.0%
Bac Technique	4.9%
Capacité en Droit	0.1%
Bac +2	15.3%
DEUG	1.5%
DUT/DEUST	2.0%
BTS	8.9%
Autre diplôme de technicien sup.	0.6%
Diplôme paramédical	2.3%
Bac +3	5.6%
Licence	3.9%
Autre diplôme Bac+3	1.6%
Bac +4	3.9%
Maîtrise	3.9%
Bac +5	6.0%
DEA	1.2%
DESS	1.9%
Ecoles de commerce	0.9%
Ecoles d'ingénieurs	2.0%
Plus de Bac +5	1.4%
Doctorat (sauf Médecine)	0.7%
Doctorat (en Médecine)	0.7%
Ensemble	100%

Source : enquête Emploi 2003-2005 (Insee) –calcul des auteurs.

Champ : individus nés entre 1968 et 1972.

2-1-Diversité des trajectoires sur le marché du travail

Par rapport au marché du travail, la trajectoire d'un individu se caractérise par un enchaînement de positions (emploi, inactivité, chômage, retraite, etc.) et de revenus liés à ces positions ainsi qu'aux positions et aux revenus passés. Par exemple, pour bénéficier de l'allocation chômage, il faut être au chômage et avoir travaillé précédemment durant une période suffisamment longue ; le montant de l'allocation est calculé sur la base des salaires passés. Afin de décrire la diversité de ces trajectoires pour la génération 1970, on suppose qu'il existe un profil générationnel identifiable des taux d'emploi et des taux d'activités en fonction de l'âge, et que les effets des caractéristiques socio-économiques sur les probabilités

différentielles de transition entre différentes positions peuvent être appréhendées sur la base de taux de transitions annuels issus de l'enquête Emploi. Les revenus liés à ces positions sont calculés sur la base d'équations de salaire pour les personnes en emploi et d'une imputation sur barème des allocations de chômage et des pensions de retraite.

Taux d'activité et taux d'emploi par âge

L'estimation d'un profil par âge des taux d'activité et de chômage selon le sexe constitue le canevas dont nous nous servons pour appréhender la diversité des trajectoires individuelles au sein d'une génération. L'idée est de pouvoir fournir le taux de chômage et le taux d'activité à chaque âge selon le sexe pour une génération donnée. Ces taux peuvent ne pas avoir encore été observés. Par exemple, pour la génération née en 1970, le taux de chômage à 55 ans ne sera effectivement observable qu'en 2025. Pour l'estimer, on s'appuie sur une modélisation de ces taux à partir des taux observés pour les générations passées que l'on peut appréhender avec les enquêtes Emploi 1968-2005. En effet, pour la génération 1950 par exemple, on dispose de ces taux pour les âges qui vont de 18 à 55 ans, et pour la génération 1960 pour les âges qui vont de 16 à 45 ans. Par ailleurs, pour tenir compte de l'effet de la conjoncture sur la relation entre âge et taux de chômage ou d'activité, il est possible à partir des enquêtes emploi de calculer un taux de chômage annuel courant. Le modèle retenu pour estimer les taux d'activité est alors spécifié comme suit :

$$\log\left(\frac{P_{gt}}{1-P_{gt}}\right) = \mathbf{a} + \mathbf{b}(t-g) + \mathbf{c}(t-g)^2 + \mathbf{d}(t-g)^3 + \mathbf{j}(t-g)^4 + U_t + D_g \quad (1)$$

où P_{gt} est le taux d'activité de la génération g pour l'année t , U le taux de chômage courant et D une indicatrice générationnelle. On spécifie le taux de chômage comme suit :

$$\log\left(\frac{U_{gt}}{1-U_{gt}}\right) = \mathbf{a} + \mathbf{b}(t-g) + \mathbf{c}(t-g)^2 + \mathbf{d}(t-g)^3 + U_t + D_g \quad (2)$$

où U_{gt} est le taux de chômage d'une génération g pour l'année t , U le taux de chômage courant et D une indicatrice générationnelle. Les résultats de ces estimations sont reproduits dans le tableau 2.

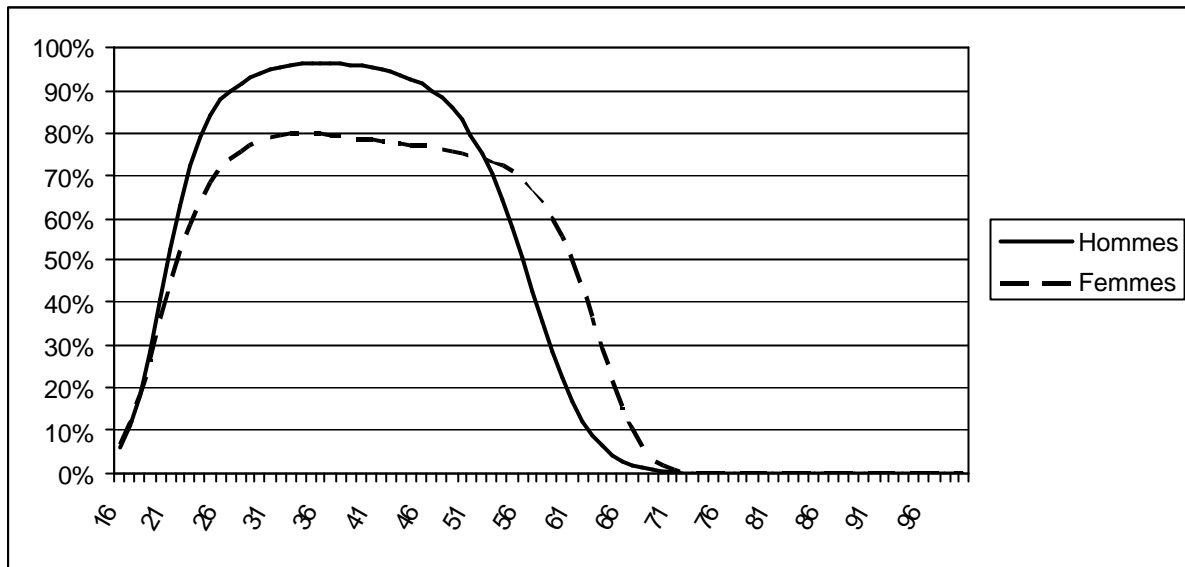
Table 2. Les déterminants du taux d'activité et du taux de chômage par âge

	Transformation logistique du taux d'activité		Transformation logistique du taux de chômage	
	Hommes	Femmes	Hommes	Femmes
Constante	-21.25 *	-27.11 *	3.53 *	3.25 *
Age	1.75 *	2.74 *	-0.56 *	-0.43 *
Age ²	-0.04 *	-0.10 *	0.01 *	0.01 *
Age ³	0.00042 *	0.00153 *	-0.00010 *	-0.00007 *
Age ⁴	-0.000002 *	-0.00001 *		
Taux de chômage courant	-3.44 *	-1.81 *	0.26 *	0.03 *
R2	0.97	0.92	0.82	0.82

Source: enquête Emploi 1968-2005 (Insee) – calcul des auteurs.

Note: l'année 1970 sert de référence pour ce modèle qui est estimé avec des indicatrices pour chaque génération dont les résultats ne sont pas reproduits dans ce tableau ; (*) pour significatif au seuil de 1% .

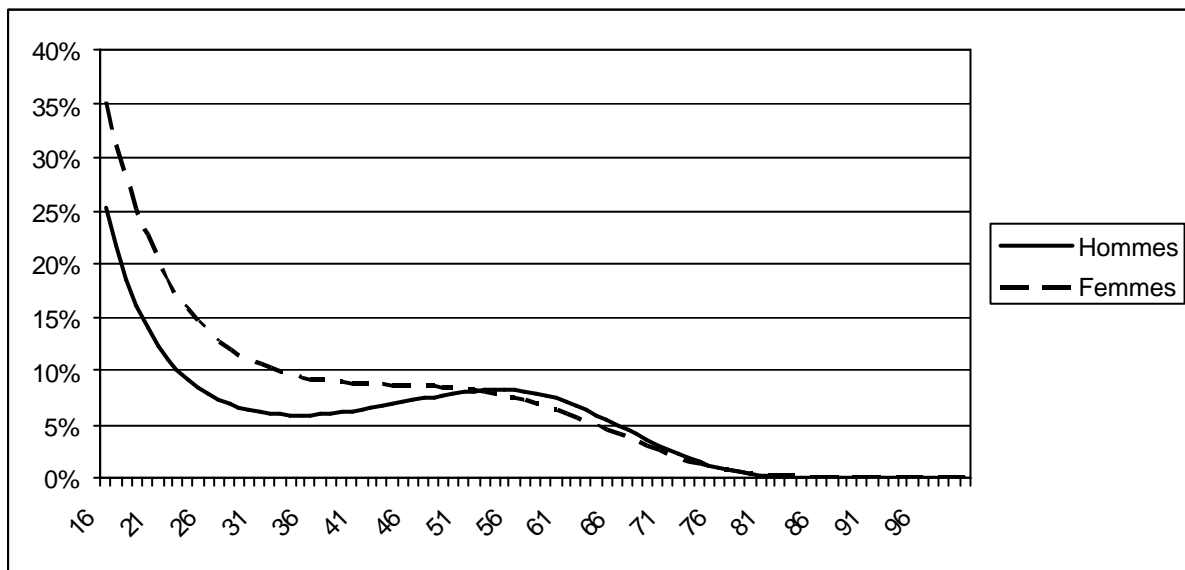
Les graphiques 1 et 2 constituent une description stylisée des profils d'activité et de chômage de la génération 1970 tels que l'on peut les recalculer à partir du tableau 2 en prenant comme hypothèse que ces individus ont connu, tout au long de leur vie, un marché du travail caractérisé par un taux de chômage courant à 8%. Dans la mesure où ne cherchons pas à simuler la carrière observée de la génération née en 1970 mais à en fournir une représentation stylisée, c'est ce profil qui est utilisé pour la simulation.

Graphique 1. Le taux d'activité tout au long de la vie

Source: calcul des auteurs.

Champ : génération 1970

Note: la simulation est basée sur une hypothèse de taux de chômage courant à 8% durant toute la durée de la vie.

Graphique 2. Le taux de chômage tout au long de la vie

Source: calcul des auteurs.

Champ : génération 1970

Note: la simulation est basée sur une hypothèse de taux de chômage courant à 8% durant toute la durée de la vie.

Modéliser les transitions sur le marché du travail

Pour modéliser les changements d'états sur le marché du travail, on s'appuie sur les profils générationnels présentés précédemment et sur une modélisation des différentiels de probabilité individuelle de transition que nous présentons ici. L'objectif principal de cette

modélisation est de rendre compte des différentiels de probabilité de transition selon le diplôme des individus.

La modélisation retient cinq états : l'inactivité, le travail indépendant, l'emploi public, l'emploi privé et le chômage. Les probabilités individuelles de transition entre ces états sont estimées sur la base des enquêtes Emploi 2003-2005⁵, par un jeu de logit binomiaux⁶. On modélise alors les probabilités conditionnelles suivantes : $P(a_{it} = 1)$ où a_{it} est une indicatrice d'activité à la période t , $P(s_{it} = 1/a_{it} = 1)$ où s_{it} est une indicatrice de travailleur indépendant à la période t , $P(p_{it} = 1/a_{it} = 1, s_{it} = 0)$ où p_{it} est une indicatrice de salarié du secteur public à la période t , $P(e_{it} = 1/a_{it} = 1, s_{it} = 0, p_{it} = 0)$ où e_{it} est une indicatrice de salarié dans le secteur privé à la période t .

Les variables utilisées dans cette modélisation des transitions comprennent notamment la position vis-à-vis du marché du travail l'année précédente -elle explique une part importante de ces probabilités différentielles, voir tableau 3- et des variables décrivant le statut socio-économique et le diplôme. Les équations estimées ont la forme suivante :

$$Y_{it} = \mathbf{a} + \mathbf{b}.Y_{i(t-1)} + \mathbf{d}.S_{it} + \mathbf{j}.D_{it} \quad (3)$$

où Y_i est la position de l'individu i à la période t , S est une matrice de variables socio-économiques et D le diplôme. Les résultats des estimations sont présentés dans le tableau 3.

⁵ Nous n'avons pas utilisé l'enquête Emploi avant 2003 car la nomenclature de diplôme est moins détaillée que dans l'enquête Emploi en continu.

⁶ Cette modélisation a été préférée à l'utilisation de logit multinomiaux, car elle est plus aisée à manipuler pour les simulations, notamment quant à l'ajustement des effectifs aux cibles issues des profils par âge des taux d'emploi et des taux de chômage.

Tableau 3 : Les déterminants des transitions (modèles logit dichotomique)

	Transition vers l'activité	Transition vers l'emploi (indépendant)	Transition vers l'emploi (secteur public)	Transition vers l'emploi (secteur privé)
Constante	0.261 *	2.196 *	1.434 *	2.108 *
Statut précédent				
Inactif	<i>ref</i>	-5.438 *	-5.336 *	-1.834 *
Chômeur	2.489 *	-6.610 *	-6.341 *	-3.091 *
Indépendant	5.057 *	<i>ref</i>	-5.555 *	-1.834 *
Salarié du secteur public	4.166 *	-10.020 *	<i>ref</i>	-1.874 *
Salarié du secteur privé	3.794 *	-8.122 *	-6.326 *	<i>ref</i>
Statut socio-économique				
Femme	-0.230 *	-0.601 *	0.548 *	-0.072 *
Nbr d'enfant(s) -pour les femmes	-0.089 *	0.066 *	0.008 *	-0.091 *
Présence d'un jeune enfant -pour les femmes	-1.579 *	0.126 *	0.143 *	0.101 *
Âge 55 et plus	-1.491 *			
Âge 60 et plus	-1.373 *			
Âge 65 et plus	-0.352 *			
Années d'expérience	-0.018 *	0.040 *	0.067 *	0.045 *
Années d'expérience (au carré)		-0.00003 *	-0.001 *	-0.001 *
Durée d'inactivité (en années)	-0.365 *			
Chômeur de longue durée			-13.459 **	-16.551 *
Diplôme				
Pas de diplôme du supérieur				
CAP/BEP	0.354 *	-0.054 *	0.553 *	0.300 *
Bac Général	0.262 *	0.359 *	0.500 *	0.337 *
Bac Professionnel	0.914 *	0.295 *	0.083 *	0.653 *
Bac Technique	0.495 *	0.056 *	0.593 *	0.451 *
Capacité en Droit	0.934 *	0.163 **	-0.841 *	-0.396 *
Bac+2				
DEUG	0.078 *	1.036 *	0.470 *	0.124 *
DUT/DEUST	0.772 *	-0.285 *	0.275 *	0.848 *
BTS	0.617 *	0.471 *	0.008 ns	0.681 *
Autre diplôme de technicien sup.	0.068 *	0.691 *	0.110 *	-0.042 *
Diplôme paramédical	0.376 *	2.421 *	3.606 *	1.121 *
Bac+3				
Licence	0.139 *	0.576 *	1.212 *	0.272 *
Autre diplôme Bac+3	0.739 *	0.905 *	0.302 *	0.613 *
Bac+4				
Maîtrise	0.314 *	0.637 *	0.844 *	0.202 *
Bac +5				
DEA	0.511 *	0.242 *	0.826 *	0.190 *
DESS	0.859 *	-0.112 *	0.429 *	0.391 *
Ecoles de commerce	1.164 *	-0.464 *	-0.626 *	0.529 *
Ecoles d'ingénieurs	0.827 *	0.671 *	0.151 *	0.634 *
Plus de Bac+5				
Doctorat (sauf Médecine)	0.935 *	0.344 *	1.362 *	0.422 *
Doctorat (en Médecine)	0.694 *	0.796 *	3.223 *	1.205 *
D de Sommer	0.955	0.958	0.911	0.72
P. Conc.	97.7	97.6	95.2	85.6
P. Disc.	2.2	1.9	4.2	13.5
P. Joint	0.1	0.5	0.6	0.9

Source: enquête Emploi 2003-2005 (Insee) – calcul des auteurs.

Note: (*) pour significatif au seuil de 1% ; (**) pour significatif au seuil de 5% ; (ns) pour non significatif

La simulation des transitions sur le marché du travail combine les profils générationnels d'activité et de chômage par âge selon le sexe et les différentiels de probabilité individuelle de transitions. La simulation des transitions tout au long de la vie répète la même opération pour chaque année. Tout d'abord, la probabilité de transition des individus vers l'activité est

calculée, puis la transition est déterminée en comparant sa probabilité avec une variable aléatoire individuelle. Lors de la simulation, le volume total des transitions aux différents âges est ajusté⁷ pour correspondre effectivement aux taux d'activité de la génération à l'âge considéré (voir *supra*). Puis, sachant que l'individu est actif, la probabilité d'être travailleur indépendant est calculée et le statut déterminé selon le même procédé⁸. Ensuite, sachant que l'individu est actif mais pas indépendant, on calcule sa probabilité de transiter vers l'emploi public, puis on détermine ce statut. Dans la procédure d'ajustement, on suppose que l'emploi public correspond à 20% de l'emploi quel que soit l'âge –ce taux correspond sensiblement à la moyenne observée sur les enquêtes Emploi 2003-2005. Ensuite, sachant que l'individu est actif mais ni indépendant ni salarié du secteur public, on calcule sa probabilité de transition vers l'emploi salarié et on détermine ce statut selon le même procédé. Les personnes au chômage sont les personnes qui n'ont pas été affectées à un autre état lors du processus. On notera que les variables de statut socio-économique concernant la famille ne sont pas incluses dans le calcul des probabilités individuelles de transition lors de la simulation, car notre appréhension des parcours longs ne cherche pas à prendre en compte cette dimension. Cependant ces variables sont incluses dans les estimations pour capturer les autres effets toutes choses égales par ailleurs.

Modéliser les revenus

La modélisation des revenus retenue ici comprend deux aspects distincts et complémentaires : (i) les revenus directement tirés du marché du travail, (ii) et les revenus provenant indirectement du travail à savoir les allocations chômage et les pensions de retraite.

La modélisation des revenus tirés du marché du travail utilisée dans la simulation se réduit à une modélisation du salaire. En effet, comme dans les enquêtes Emploi nous ne disposons pas de variables permettant d'appréhender facilement le revenu des indépendants, on pose comme hypothèse que la modélisation du salaire à laquelle on procède constitue une approximation acceptable.

Afin de modéliser les salaires, nous avons estimé séparément pour chaque diplôme des équations de gain à la Mincer. L'objectif de cette spécification est de prendre en compte un double aspect de l'hétérogénéité individuelle. Tout d'abord l'hétérogénéité des carrières

⁷ Dans la littérature anglo-saxonne, cette procédure d'ajustement de la simulation des situations individuelles à une cible agrégée est désignée sous le terme *alignment*.

⁸ Pour l'ajustement, on s'appuie sur une modélisation de la part d'indépendants en emploi pour la génération 1970, semblable à celle qui ont été présentées pour les taux d'activité et de chômage. Cette modélisation n'a pas été reproduite ici car elle n'a techniquement pas d'effet, dans la mesure où nous approximations le revenu des indépendants par celui des salariés de mêmes caractéristiques.

salariales moyennes pour un même niveau d'éducation. Il s'agit alors de différencier par exemple les profils de salaire des Bac +2 titulaires d'un diplôme universitaire de ceux titulaires d'un BTS. Ensuite, il s'agit pour les titulaires d'un même diplôme de pouvoir rendre compte de l'hétérogénéité des carrières salariales. De ce point de vue, on ne cherche pas à identifier l'effet propre du diplôme toutes choses égales par ailleurs –ce qui nécessiterait de contrôler les différents biais d'endogénéité -, mais à estimer des distributions de salaires : la partie résiduelle de l'estimation des équations de salaire constitue alors une information importante que l'on veut pouvoir conserver et introduire dans la simulation.

Du fait du faible nombre d'observations pour certains diplômes, il a fallu procéder à des regroupements pour les estimations. Ce regroupement s'est appuyé sur une analyse de données des liens entre diplôme et marché du travail mettant en œuvre des classifications ascendantes hiérarchique⁹, dans le cas où cette analyse de données se révélait insuffisante pour trancher, on a procédé à des regroupements par niveau de diplôme. On peut présenter les équations estimées comme suit :

$$\log(w_{id}) = \mathbf{a}_d + \mathbf{b}_d \cdot e_{id} + \mathbf{d}_d \cdot e_{id}^2 + \mathbf{j}_d \cdot X_{id} + \mathbf{e}_{id} \quad (4)$$

La variable expliquée est le logarithme du salaire horaire net (w_{id}), tel qu'il est possible de l'appréhender à partir des enquêtes Emploi 2003-2005, pour un individu i ayant un diplôme d ; e_{id} est le nombre d'années d'expérience et X une matrice de variables qui caractérisent les individus (genre, taux de chômage des jeunes lors du début de la carrière) et le type d'emploi (emploi public, secteur d'activité, etc.). La modélisation cherche principalement à capturer des profils de revenus selon le nombre d'années d'expériences tout au long de la carrière, différenciés selon le diplôme : le traditionnel effet de l'expérience est ici estimé par diplôme. Les résultats de ces estimations sont présentés dans le tableau 4.

⁹ Cette analyse de données est disponible sur demande auprès des auteurs.

Tableau 4 : Les déterminants du logarithme de salaire horaire selon le diplôme

	Sans diplôme ou titulaires d'un CAP/BEP)	Bac Pro.	Bac gé., Bac Tech. Capacité en droit et DEUG	DUT/DEUST	BTS, Techniciens du supérieur, Diplôme Paramédical	Licence, Autre Bac+3 et Maîtrise	DESS, DEA, Ecoles de com., doctorat	Ecole d'ingénieur (1)
Constante	2.2 *	2.3 *	2.2 *	2.2 *	2.1 *	2.7 *	2.6 *	3.0 *
Années d'expérience	0.012 *	0.030 *	0.020 *	0.030 *	0.026 *	0.037 *	0.040 *	0.055 *
Année d'expérience (au carré)	-0.0001 *	-0.0005 *	-0.0001 *	-0.0002 ns	-0.0003 *	-0.0005 *	-0.0006 *	-0.0010 *
Femme	-0.14 *	-0.12 *	-0.14 *	-0.16 *	-0.11 *	-0.13 *	-0.12 *	-0.16 *
Secteur Public	0.18 *	0.05 **	0.13 *	0.05 **	0.07 *	0.17 *	0.01 ns	0.13 *
Nombre d'heures (par mois)	-0.001 *	-0.001 *	-0.001 *	-0.002 *	-0.001 *	-0.003 *	-0.001 *	-0.003 *
Taux de chômage des jeunes (2)	-0.08 ns	-1.54 *	-0.04 ns	1.27 ns	0.90 *	0.04 ns	0.19 ns	-1.85 *
Secteur d'activité								
Secteur manufacturier et BTP	0.003 ns	0.030 **	0.007 ns	0.037 **	0.033 *	0.025 ns	0.110 *	0.026 ns
Energie	0.254 *	0.195 *	0.230 *	0.082 *	0.166 *	0.198 *	0.201 *	0.150 *
Finance	0.117 *	0.090 *	0.063 *	0.081 *	0.063 *	0.080 *	0.071 *	0.051 ns
Services aux entreprises	-0.025 *	-0.018 ns	-0.015 **	0.039 *	0.004 ns	0.029 **	0.041 *	0.034 ***
Services aux particuliers	-0.166 *	-0.122 *	-0.145 *	-0.114 *	-0.181 *	-0.152 *	-0.253 *	-0.140 *
Administration	-0.134 *	-0.100 *	-0.078 *	-0.070 *	-0.058 *	-0.095 *	-0.086 *	-0.242 *
Autre secteur	-0.049 *	-0.074 *	-0.063 *	-0.055 *	-0.028 *	-0.086 *	-0.084 *	0.120 *
Diplôme								
Sans diplôme	-0.06 *							
CAP/BEP	ref.							
Bac Général			ref.					
Bac Professionnel								
Bac Technique				-0.04 *				
Capacité en Droit				-0.10 **				
DEUG				0.10 *				
DUT/DEUST								
BTS					ref.			
Autre diplôme de technicien sup.					0.00 ns			
Diplôme paramédical					0.17 *			
Licence						-0.08 *		
Autre diplôme Bac+3						0.13 *		
Maîtrise						ref.		
DEA							-0.01 ns	
DESS							ref.	
Ecole de commerce							0.07 *	
Ecole d'ingénieur								
Doctorat (sauf médecine)								0.06 *
Doctorat (médecine)								0.29 *
R2	0.27	0.36	0.41	0.51	0.47	0.49	0.45	0.55

Source: enquête Emploi 2003-2005 (Insee) – calcul des auteurs.

Note: (*) pour significatif au seuil de 1% ; (**) pour significatif au seuil de 5% ; (***) pour significatif au seuil de 10% ; (ns) pour non significatif.

Les variables additionnelles correspondant à la matrice X sont introduites pour identifier cet effet de l'expérience toutes choses égales par ailleurs. Pour capturer un potentiel effet de génération lié à la conjoncture lors de l'entrée sur le marché du travail, on contrôle ces estimations par le taux de chômage des moins de 25 ans¹⁰. Pour capturer l'effet spécifique des carrières féminines sur le salaire, nous contrôlons par l'introduction d'une indicatrice de genre dans l'équation de salaire. Cet effet n'est pas introduit dans la simulation du salaire individuel. En effet, dans la mesure où l'estimation s'effectue sur données en coupe, ce différentiel peut s'expliquer par des différences de profil de carrière entre les hommes et les femmes. Or nous introduisons déjà ce différentiel de carrière dans la simulation – voir *supra*. L'introduction de l'effet genre risquerait alors de conduire à un « double compte ». En choisissant de ne pas introduire l'effet de genre dans la simulation des salaires, nous supposons que les différences en termes de profil par âge des taux d'activité et des taux de chômage selon le genre introduits dans la simulation, rendent compte de l'intégralité de l'effet

¹⁰ Lors de la simulation, on fait l'hypothèse que ce taux est constant sur la période et se fixe à 8%.

estimé du genre sur les salaires. Pour capturer la spécificité des salaires dans l'emploi public, on introduit une indicatrice d'emploi public qui est utilisée pour la simulation. Pour capturer l'effet spécifique du secteur d'activité du salaire, une série d'indicatrices est également introduite dans les estimations. Dans la simulation, nous supposons que les individus font leur carrière dans le même secteur d'activité. Ce secteur est imputé de manière aléatoire sur la base d'une répartition des emplois par secteur d'activité pour les moins de trente ans selon le diplôme et le sexe issue de l'enquête Emploi 2003-2005. Enfin, le temps de travail est introduit comme variable de contrôle. Pour les simulations, on suppose que les emplois correspondent à des emplois à temps plein et nous fixons alors arbitrairement le nombre d'heures travaillées par mois à 150.

Pour procéder à une simulation des salaires correspondant aux données sur lesquelles ils sont estimés, la distribution de la part non expliquée des salaires est conservée –l'élément e_{id} de l'équation (4). On dispose donc d'une distribution de résidus par types de diplômes¹¹. Dans le modèle de microsimulation, le salaire des individus est alors composé d'une partie expliquée calculée sur la base de la modélisation et d'une partie résiduelle affectée de manière aléatoire. D'un point de vue pratique, le résidu individuel est tiré une première fois, puis il est conservé tant que la personne reste en emploi ; si l'individu transite par une période de chômage, on tire aléatoirement un autre résidu pour calculer le salaire de l'emploi suivant. Ceci revient à poser l'hypothèse restrictive suivante : la partie non expliquée du salaire est uniquement due à l'emploi occupé et non aux caractéristiques non observées de l'individu (talent, réseaux sociaux, etc.).

La prise en compte des prélèvements et des prestations correspond à ce qui est généralement fait dans les modèles statiques ou arithmétiques de microsimulation : une imputation sur barème¹². Elle consiste simplement à généraliser ce procédé sur l'ensemble de la trajectoire de vie. Notre modèle de microsimulation dynamique simule les principaux transferts et prélèvements du système socio-fiscal français liés au travail, à savoir : les allocations chômage et les pensions de retraite. Ces transferts et prélèvements sont calculés sur barème à partir de la trajectoire salariale simulée. Le calcul du montant des allocations chômage et des pensions de retraite nécessite de connaître le salaire brut, on suppose par convention qu'il représente 1,2% du salaire net simulé. Pour les allocations chômage, seule l'ARE (l'Allocation d'Aide au retour à l'Emploi) est simulée dans une version simplifiée calée sur un pas annuel. En effet, dans notre modèle, les individus ne peuvent pas être au chômage

¹¹ Ceci permet notamment de ne pas imposer une loi à la distribution des résidus simulés.

¹² Pour une présentation générale, voir par exemple Blanchet (1998), Legendre *et alii* (2003).

moins d'une année ; lorsqu'ils arrivent en fin de droit au cours d'une année, on considère alors que les derniers mois ne sont pas indemnisés. Au vu de leur chronique de salaire et de leur durée d'emploi passée, les individus se voient affecter l'ARE durant leurs années passées au chômage. Du côté retraite, les trois principales composantes du système français sont simulées : pensions de base, retraites complémentaires et régime fonctionnaire. On considère que les individus arrivés à 65 ans qui ne sont plus en emploi liquident leurs droits à la retraite. La retraite de base est calculée à partir des 25 meilleures années. Les retraites complémentaires (AGIRC, ARRCO) sont simulées en fonction du statut : on considère qu'à partir de cinq années d'études après le Bac, les individus sont cadres et bénéficient du système correspondant (AGIRC). Les retraites complémentaires sont calculées sur la base des cotisations retraites acquittées tout au long de la carrière et de la valeur du point. Le régime fonctionnaire est également simulé : il est imputé aux individus qui ont travaillé plus de 41 ans dans le secteur public, leur pension de retraite correspond alors à 75% de leur dernier salaire.

Au terme de cet exercice, nous disposons d'une simulation de la trajectoire des individus composant une base de données artificielles représentative de la génération née en 1970 entre 16 et 100 ans. Il nous reste à corriger ces trajectoires de la probabilité d'atteindre ces différents âges pour les individus.

2-2-Différentiels d'espérance de vie et diplôme

L'existence d'un lien entre le niveau d'éducation et l'espérance de vie est largement connu et commenté. Par exemple, Kunst et Mackenbach (1994) mesurent et comparent la taille du différentiel de mortalité selon le niveau d'éducation pour différents pays. Leur étude montre que la France fait partie du groupe de pays pour lesquels ce différentiel est le plus important. De plus, la tendance générale à l'augmentation de l'espérance de vie ne fait pas disparaître ces inégalités. Par exemple, Mackenbach *et alii* (2003, 2008) montrent que ces inégalités ont persisté voire augmenté au cours du temps : l'augmentation de l'espérance de vie apparaît d'autant plus faible que le niveau d'éducation est faible. Ce constat est confirmé dans le cas français, notamment par les travaux de Menvielle *et alii* (2007).

L'ampleur de ce lien est difficile à mesurer, notamment car le niveau d'éducation est corrélé avec d'autres variables (comme le niveau de revenu) qui peuvent également avoir un impact sur la santé et l'espérance de vie. Dans ce cadre, quand on contrôle la mesure de l'effet de l'éducation sur la santé par d'autres variables comme le revenu ou les caractéristiques

familiales, on trouve généralement que cet effet a tendance à diminuer mais reste significatif – par exemple Schnittker (2004), Cutler et Lleras-Muney (2006).

Ces différentiels de mortalité selon le niveau d'éducation peuvent avoir plusieurs sources. Le cadre d'analyse économique de la santé comme capital humain proposé par Grossman (2006) permet de présenter la première de ces sources. On peut tout d'abord considérer que le niveau d'éducation a un impact sur la santé et donc l'espérance de vie. Ceci correspond à l'idée que les individus produisent leur propre santé en combinant différents facteurs de production ; le niveau d'éducation permet alors d'augmenter la productivité et donc la santé et l'espérance de vie.

On peut également considérer que les personnes de santé plus fragile ont plus de difficulté à réussir leurs études ce qui peut conduire à renverser la causalité présente dans le cadre de Grossman (2006). Dans ce cas c'est le niveau d'éducation et l'espérance de vie qui dépendent de la santé dans l'enfance.

Il existe également une troisième source. En effet, on peut également considérer que les choix individuels en termes de niveau d'éducation, d'investissement dans la santé et donc d'espérance de vie reflètent des caractéristiques individuelles en amont comme la préférence pour le présent. Les individus ayant une préférence forte pour le futur vont alors investir à la fois dans l'éducation et dans la santé ce qui se traduira par une corrélation entre niveau d'éducation et espérance de vie.

A ces trois explications possibles de la corrélation entre éducation et espérance de vie s'ajoutent des remarques d'ordre socio-économique sur les liens entre d'une part le niveau d'éducation et les caractéristiques socio-économiques des individus et d'autre part ces caractéristiques et l'état de santé et l'espérance de vie. Il apparaît généralement qu'une meilleure éducation conduit à une vie dans des conditions matérielles plus aisées qui permettent un meilleur accès à la santé –par exemple Cutler et Lleras-Muney (2006). Par ailleurs, certains travaux de sociologues et d'ergonomes montrent que les emplois occupés tout au long de la carrière et leur degré de pénibilité peuvent jouer sur la santé et sur l'espérance de vie -Gollac et Volkof (2007).

Dans le cas de la France, les différentiels de mortalité par niveau d'éducation ne sont pas disponibles. Cependant, on peut obtenir des tables de mortalité par âge, sexe et catégorie socio-professionnelle en coupe –notamment dans Vallin et Meslé (2001) et Robert-Bobée (2006). Pour estimer les taux de mortalité par âge, sexe et diplôme pour la génération 1970, on procède en deux étapes : on commence par effectuer une conversion catégorie socio-professionnelle/diplôme sur la base des données de l'enquête Emploi, puis on projette ces

tables de mortalité pour qu'elles correspondent à celle que connaîtra la génération 1970 tout au long de sa trajectoire de vie.

La mortalité selon le diplôme

Dans la première étape, on commence par transformer les tables de mortalité par âge, genre et catégorie socio-professionnelle en tables de mortalité par âge, genre et diplôme. On suit en cela une méthode déjà utilisée sur données françaises pour le modèle de microsimulation Destinie –voir Insee (1999). Robert-Bobée et Monteil (2005) donnent des tables de mortalité par catégorie socio-professionnelle pour la période 1991-1999. Pour transformer ces tables, on impute le taux de mortalité aux individus présents dans l'enquête Emploi 2003-2005 sur la base de leur âge, sexe et catégorie socio-professionnelle. A partir des individus de l'enquête Emploi on peut alors compiler des tables de mortalité moyenne par diplôme. Pour mieux correspondre aux tables de mortalité utilisée dans cette transformation, on aurait pu utiliser les enquêtes Emploi 1991-1999 ; cependant cette option n'a pas été retenue car la nomenclature de diplôme utilisée pour cette période est moins fine que celle des enquêtes Emploi 2003-2005 or il s'agit d'une dimension de l'analyse que nous privilégions. Cependant, la structure catégorie socio-professionnelle/diplôme varie un peu ce qui peut contribuer à introduire un léger biais dans le sens d'une compression du différentiel d'espérance de vie¹³.

Par ailleurs, cette base est trop petite pour produire des tables de mortalité par âge à un niveau d'éducation très fin. Il a donc fallu procéder à des regroupements de diplômes. L'objectif étant de rendre compte le plus précisément possible des particularités du système éducatif français, dans la mesure du possible, nous avons cherché à maintenir une différenciation par type d'éducation : par exemple, on a distingué les Bac selon qu'il s'agit de Bac général, professionnel ou technique ; nous avons également différencié les diplômes de l'éducation du supérieur selon leur spécificité –grandes écoles *versus* universités. Au total, on retient onze catégories de diplôme.

- 1- Les individus sans diplômes.
- 2-Les titulaires d'un CAP ou d'un BEP.
- 3-Les titulaires d'un Bac général.
- 4-Les titulaires d'un Bac professionnel, d'un Bac technologique ou d'une capacité en droit.
- 5-Les diplômés paramédicaux.

¹³ Les résultats de ces différences de structure CSP/éducation sont disponibles sur demande auprès des auteurs.

- 6-Les autres diplômés à Bac +2 (DEUG, DUT, BTS, autres diplôme technique du supérieur).
- 7- Les diplômés de niveau Bac +3.
- 8-Les diplômés de niveau Bac +4.
- 9-Les diplômés de niveau Bac +5 issus de l'Université.
- 10-Les écoles d'ingénieur et les écoles de commerce.
- 11-Les titulaires d'un doctorat.

Pour modéliser la mortalité à âge et diplôme donnés, sur la base des tables de mortalité par diplôme dont nous disposons, nous estimons une fonction qui distingue un effet lié aux taux de mortalité moyenne de la génération (\mathbf{b}) d'un effet lié à l'âge (\mathbf{g} et \mathbf{d}) et une constante (\mathbf{a}). Chacun de ces effets est décomposé en deux effets : un effet moyen et un effet propre au diplôme. Nous estimons donc la fonction suivante de manière séparée pour les hommes et les femmes :

$$\log \left[\frac{M_{ae}}{1 - M_{ae}} \right] = \mathbf{a}_e + \mathbf{b}_e \cdot \log \left[\frac{M_a}{1 - M_a} \right] + \mathbf{g}_e \cdot a + \mathbf{d}_e \cdot a^2 \quad (5)$$

où :

$$\mathbf{a}_e = \mathbf{a}_1 + \mathbf{a}_2 \cdot e \quad (5.1)$$

$$\mathbf{b}_e = \mathbf{b}_1 + \mathbf{b}_2 \cdot e \quad (5.2)$$

$$\mathbf{g}_e = \mathbf{g}_1 + \mathbf{g}_2 \cdot e \quad (5.3)$$

$$\mathbf{d}_e = \mathbf{d}_1 + \mathbf{d}_2 \cdot e \quad (5.4)$$

où M est le taux de mortalité, a l'âge et e le niveau d'éducation appréhendé par le diplôme. La variable expliquée est issue de nos tables de mortalité transformées ; tandis que la mortalité moyenne (M_a) est reprise de Vallin et Meslé (2001). Cette spécification s'inspire de celle utilisée pour le modèle Destinie –voir Insee (1999)- mais nous avons ajouté un terme d'interaction avec l'âge et l'âge au carré pour tenir compte du fait que l'impact du niveau d'éducation sur la mortalité peut potentiellement décroître avec l'avancée en âge. Cet ajout se justifie par les résultats de Robert-Bobée et Cadot (2007) qui montrent que malgré la

persistance d'un différentiel de mortalité aux âges élevés, ce différentiel tend à s'atténuer. Les résultats de ces estimations sont présentés dans le tableau 5.

Tableau 5 : Les déterminants du taux de mortalité par âge selon le sexe

Variable	Hommes	Femmes
Constante (a_1)	1.1832 *	-1.221 *
Mortality moyenne par âge (β_1)	1.0997 *	0.8383 *
Age (γ_1)	-0.0116 *	-0.0053 *
Age ² (d_1)	0.0000 ns	0.0002 *
Education (a_2)		
CAP – BEP	-0.5125 *	-0.3404 *
Bac général	1.148 *	-0.7679 *
Bac professionnel et technologique, capacité en droit	-0.4375 ns	-0.3675 *
Diplôme paramédical	-7.6262 *	-1.267 *
Autre diplôme Bac +2	1.2109 *	-0.5218 *
Bac +3	-1.4988 *	-0.941 *
Bac +4	5.1688 *	-1.3131 *
Bac +5 (Université)	-2.3255 *	0.0131 ns
Ecoles d'ingénieur, écoles de commerce	-1.3314 *	-2.4456 *
Doctorat	7.1069 *	-2.7472 *
Interaction entre Education et Mortalité moyenne par âge (β_2)		
CAP – BEP	-0.0363 ***	-0.0309 *
Bac général	0.0558 ns	-0.0799 *
Bac professionnel et technologique, capacité en droit	-0.153 *	-0.0364 **
Diplôme paramédical	-1.0102 *	-0.0958 *
Autre diplôme Bac +2	0.071 **	-0.0267 *
Bac +3	-0.2511 *	-0.0684 *
Bac +4	0.5859 *	-0.106 *
Bac +5 (Université)	-0.3388 *	0.0704 *
Ecoles d'ingénieur, écoles de commerce	-0.1213 **	-0.1822 *
Doctorat	0.8084 *	-0.2468 *
Interaction entre l'éducation et l'âge (γ_2)		
CAP – BEP	0.005 *	0.0000 *
Bac général	-0.0368 *	0.0001 *
Bac professionnel et technologique, capacité en droit	-0.0347 *	0.0001 *
Diplôme paramédical	0.0075 ns	0.0001 *
Autre diplôme Bac +2	-0.0435 *	0.0000 ns
Bac +3	-0.0249 *	0.0001 *
Bac +4	-0.0555 *	0.0001 *
Bac +5 (Université)	-0.0197 *	0.0000 *
Ecoles d'ingénieur, écoles de commerce	-0.0097 *	0.0000 *
Doctorat	-0.0791 *	0.0002 *
Interaction entre l'éducation et l'âge au carré (d_2)		
CAP – BEP	0.0000 ns	-0.001 *
Bac général	0.0003 *	-0.0027 *
Bac professionnel et technologique, capacité en droit	0.0004 *	-0.0064 *
Diplôme paramédical	0.0007 *	0.0046 *
Autre diplôme Bac +2	0.0004 *	0.0002 *
Bac +3	0.0004 *	0.0012 *
Bac +4	0.0001 **	0.0022 *
Bac +5 (Université)	0.0004 *	0.0036 *
Ecoles d'ingénieur, écoles de commerce	0.0002 *	0.0184 *
Doctorat	0.0001 *	0.0096 *
R2	0.9978	0.9998

Source: enquête Emploi 2003-2005 (Insee), Vallin et Meslé (2001), Robert-Bobée et Monteil (2005) –calcul des auteurs.

Note: les “sans diplôme” constituent la catégorie de référence.

Projection des tables de mortalité

Jusqu'à présent, nous avons proposé une méthode pour estimer un différentiel de mortalité par diplôme à un point donné du temps. Pour simuler la trajectoire de vie d'une génération, il convient de projeter ces taux de mortalité à différents points du temps. Pour illustrer ce problème, si on considère la génération née en 1970, pour notre simulation, nous avons besoin de connaître le taux de survie des individus à 30 ans en 2000 que l'on peut raisonnablement tirer des estimations présentées précédemment qui concernent la période 2003-2004, mais on en a également besoin pour des périodes beaucoup plus tardives lorsque cette génération avancera en âge : par exemple, on aura notamment besoin des taux de survie à 40 ans en 2010. Les taux de mortalité ont eu tendance à décroître en France tout au long du dernier siècle et demi – voir par exemple Pison (2005). L'extrapolation que nous utilisons suppose que la mortalité diminue de manière linéaire au cours du temps. Cette hypothèse semble raisonnable au vu des résultats disponibles pour la France¹⁴. Pour simuler une décroissance linéaire nous avons besoin de deux points dans le temps. Cependant, nous ne disposons pas d'une table de mortalité par diplôme, mais seulement d'une table de mortalité par âge et sexe en 2049 issue des projections de l'Insee et disponible dans Vallin et Meslé (2001). Nous faisons alors l'hypothèse que les tables de mortalité par diplôme connaîtront la même évolution que la table de mortalité moyenne.

Il s'agit donc d'une hypothèse conservatrice qui ne correspond pas aux résultats disponibles par ailleurs sur l'hétérogénéité de la baisse de mortalité au cours du temps selon le niveau d'éducation, présentée par exemple dans Menvielle *et alii* (2007). Cependant, cette hypothèse reste cohérente avec les autres éléments de la microsimulation, notamment la simulation des salaires qui ne prend pas en compte l'impact propre d'une modification de la répartition des niveaux d'éducation sur le rendement de certains diplômes. De manière similaire, on ne modélise pas d'autres éléments de contexte susceptibles de modifier cette disparité des taux de mortalité selon le diplôme comme la facilité d'accès au soin ou le caractère plus ou moins redistributif du système de santé. De ce point de vue et dans l'optique qui est la nôtre, il apparaît cohérent de choisir cette hypothèse conservatrice pour modéliser la mortalité.

Pour projeter les taux de mortalité au cours du temps, nous utilisons donc les tables de mortalité de l'année 1995 –approximée par la table de mortalité 1991-1999 dont nous disposons- et les tables de l'année 2049 comme suit :

¹⁴ Par exemple, Vallin et Meslé (2001), ou la base de données EcoSanté de l'OCDE.

$$M_{a,e}(n) = \exp \left[\log(M_{a,e(1995)}) + (n - 1995) * \frac{\log(M_{a(2049)}) - \log(M_{a(1995)})}{54} \right] \quad (6)$$

où $M_{ae}(n)$ est le taux de mortalité par âge (a) et niveau d'éducation (e) pour l'année n .

Au terme de cet exercice, nous disposons donc d'une fonction de survie différenciée selon le sexe et le diplôme. Dans la simulation, on applique cette fonction de survie à la pondération des individus à 16 ans, afin de calculer une nouvelle pondération pour chacun des âges simulés.

3-Le rendement des diplômes comme indicateur des risques de carrière

La section précédente nous a précisé comment, à partir d'un modèle de microsimulation, il est possible de produire une description stylisée de la diversité des trajectoires individuelles pour une génération dont on connaît les caractéristiques éducatives. A partir de ces résultats, il est possible d'envisager de calculer des indicateurs synthétiques individuels de carrière. Pour illustrer l'intérêt de notre perspective générationnelle d'analyse des risques de transition entre éducation et marché du travail sur la base de microsimulations, nous retenons un indicateur purement monétaire et le discutons: le taux de rendement interne d'un diplôme. En économie, la théorie du capital humain a largement diffusé l'analyse de l'éducation comme un choix d'investissement. Dans ce cadre, les rendements de l'éducation ont d'abord été appréhendés à l'aide d'équation de gain –voir notamment Mincer (1974). Les développements plus récents soulignent l'insuffisance de cette méthode dans la prise en compte des coûts d'opportunité et du contrôle du caractère endogène de certaines décisions –pour une synthèse récente, voir par exemple Heckman *et alii* (2006).

Notre objectif est de produire une distribution des rendements individuels en imposant le minimum de contraintes à la loi suivie par cette distribution. Il ne correspond pas à celui d'Heckman *et alii* (2006) qui identifient le rendement privé d'une année d'étude marginale sur données américaine. Cependant, dans une analyse des risques de carrière liés à l'éducation, notre perspective complète ces travaux. En commentant à chaque fois les résultats obtenus, on commence par présenter les hypothèses posées pour construire un taux de rendement interne individuel des diplômes à seize ans, puis celle pour identifier la part de ce rendement directement liée au différentiel d'espérance de vie.

3-1-La distribution des rendements de l'éducation

Nous avons simulé la carrière de la génération née en 1970, sur la base d'une population artificielle de 34 643 individus représentatifs de cette génération en termes de croisement diplôme/genre/année de fin d'étude telle que l'enquête Emploi 2003-2005 permet de l'appréhender.

Dans ce cadre, on définit t_i le taux de rendement interne de l'individu i à l'âge de 16 ans comme suit :

$$\sum_{a=1}^{a=84} F_{ia} (1+t_i)^{-a} = 0 \quad (7)$$

Où F_{ia} est le flux financier net lié à l'investissement dans l'éducation de l'individu i à l'âge $(15+a)$. On suppose par ailleurs que les individus ne peuvent pas vivre au-delà de cent ans. Si les résultats du modèle de microsimulation de carrière sur lequel nous nous appuyons permettent d'identifier les revenus d'un individu tout au long de sa vie, ces revenus ne peuvent pas être considérés comme liés uniquement à l'investissement dans l'éducation. Ils comprennent aussi les gains liés à l'éducation primaire et secondaire avant seize ans. Pour identifier les flux financiers nets liés à l'éducation, nous avons dû poser des hypothèses complémentaires sur les flux financiers nets, que l'on peut formuler comme suit :

$$F_{ia} = (W_{ia} + D_{ia} + R_{ia}) - O_{ia} - \bar{G}d0_a \quad (8)$$

Où W_{ia} est le salaire de l'individu i à l'âge $(a+15)$, D le montant de l'allocation chômage, R le montant de la pension de retraite. L'élément $\bar{G}d0_a$ sert à identifier le revenu net lié au diplôme obtenu ; il se définit comme le revenu moyen des individus qui n'ont pas de diplôme au même âge. L'élément O_{ia} représente le coût d'opportunité de l'éducation de l'individu i à l'âge $(15+a)$. Quand l'individu i avec un diplôme de niveau dx est encore dans le système éducatif, son coût d'opportunité (O_{dxai}) est déterminé comme suit :

$$\begin{aligned} O_{dxai} &= \text{MAX} \{ \bar{G}LF_{d0a}, \bar{G}LF_{d1a}, \dots, \bar{G}LF_{dxa} \} \quad \text{if } ALF_i < a \\ O_{dxai} &= 0 \quad \text{if } ALF_i > a \end{aligned} \quad (9)$$

où ALF_i est l'âge d'entrée sur le marché du travail de l'individu i , et \overline{GLF}_{dx} la moyenne des gains à l'âge $(a+15)$ des individus actifs avec un niveau d'éducation dx .

La distribution des taux de rendement interne individuels pour différents diplômes ainsi calculée peut être lue comme un indicateur des avantages et des risques monétaires à suivre une formation tels qu'ils se posent à un individu de seize ans ayant le choix entre quitter le système éducatif et entrer sur le marché du travail. On suppose alors par ailleurs que cet individu ne sait pas *a priori* où il se situe dans l'échelle des « talents » valorisés par le système éducatif et qu'il ne connaît pas non plus la manière dont ses caractéristiques familiales (capital culturel, réseaux sociaux, etc.) vont jouer sur les emplois qu'il pourra obtenir, ni même les interactions possibles entre ces différentes dimensions. Dans l'approche en termes de risques qui est celle que nous privilégions un indicateur nous semble particulièrement important : il s'agit de la part de rendement négatif d'un diplôme. Cet indicateur peut s'interpréter comme la part des individus ayant suivi une formation sans que cela se traduise par un avantage monétaire par rapport à la trajectoire moyenne d'un individu sans diplôme. Le tableau 6 présente les résultats obtenus¹⁵.

¹⁵ Certains diplômes ne sont pas repris dans les tableaux de résultats car ils représentent une part trop faible de la génération (capacité en droit) ou parce que leurs revenus restent mal appréhendés par l'enquête Emploi (doctorat en médecine).

Tableau 6 : Les rendements des diplômes calculés à 16 ans

Diplôme	Moyenne	Médiane	Part de rendements négatifs
Niveau Bac	16%	10%	5.6%
Bac Général	15%	9%	6.2%
Bac Professionnel	16%	11%	4.9%
Bac Technique	16%	10%	5.1%
Bac +2	11%	9%	4.8%
DEUG	8%	7%	10.0%
DUT/DEUST	12%	11%	3.3%
BTS	11%	10%	4.0%
Autre diplôme de technicien sup.	8%	7%	6.8%
Diplôme paramédical	10%	9%	5.1%
Bac +3	8%	7%	7.1%
Licence	7%	7%	8.4%
Autre diplôme Bac+3	10%	9%	3.9%
Bac +4	7%	7%	6.9%
Maîtrise	7%	7%	6.9%
Bac +5	9%	9%	3.1%
DEA	7%	7%	4.1%
DESS	8%	8%	4.0%
Ecoles de commerce	11%	11%	2.4%
Ecoles d'ingénieurs	11%	11%	2.0%
Plus de Bac +5	6%	5%	2.6%
Doctorat (sauf Médecine)	6%	5%	2.6%

Source : calcul des auteurs.

Champ : génération née en 1970.

Note : la simulation est effectuée sur la base de 34 643 observations représentatives de la génération 1970. Le rendement des écoles de commerce et de certaines autres écoles est surévalué du fait de la non prise en compte des coûts de scolarité propres à l'école.

Les résultats du tableau 6 confirment un certain nombre de résultats déjà connus sur les rendements de l'éducation. A savoir que pour un même niveau d'éducation les rendements se différencient selon la filière suivie –voir Goux et Maurin (1994). Par exemple, à niveau Bac +2, les rendements pour les titulaires d'un BTS apparaissent meilleurs que ceux pour les titulaires d'un DEUG, que l'on considère des valeurs centrales telles que la moyenne et la médiane ou des indicateurs de risques tels que la part de rendements négatifs. De même, à niveau Bac +5 la situation des diplômés des écoles d'ingénieur et des écoles de commerce apparaît plus favorable que celle des titulaires d'un diplôme universitaire à vocation professionnelle (DESS) ; cette dernière apparaît cependant légèrement plus favorable que la situation des titulaires d'un diplôme universitaire à vocation généraliste.

Un autre résultat intéressant est obtenu en comparant la situation des Bac à certains diplômes du supérieur. En effet, la situation des titulaires d'un Bac apparaît plus favorable que la situation des titulaires d'un diplôme Bac +2 lorsque l'on considère les indicateurs de valeur centrale : le taux de rendement moyen des Bac est de 16% alors que celui des Bac +2 est de 11% (il varie entre 8 et 12% selon le diplôme considéré) ; le taux de rendement médian est de 10% pour les Bac (il varie entre 9 et 11% selon le Bac considéré) alors qu'il est de 9% pour les Bac +2 (il varie entre 7 et 11% selon le diplôme considéré). L'indicateur de risque que constitue la part de rendement négatif vient pondérer ce jugement : la part des rendements négatifs est plus faible pour les Bac +2 que pour les Bac (4,8% *versus* 5,6%). Cependant cette assertion n'est pas vraie lorsque l'on décompose selon le type de diplôme obtenu : la situation des titulaires d'un DEUG apparaît moins favorable que celle des titulaires d'un Bac, qu'il soit général, professionnel ou technique, au regard de l'ensemble des indicateurs présentés : rendement moyen, médian et part des rendements négatifs. De plus cette position plus favorable du Bac au regard de ces trois indicateurs se retrouvent pour d'autres diplômes du supérieur : les licences et les maîtrises.

De manière plus générale, un résultat important consiste à souligner que ce qui différencie sensiblement les études du supérieur du Bac, ce n'est pas le rendement médian ou moyen mais plutôt la diminution d'un risque de rendement négatif. En effet, poursuivre des études conduit à renoncer à un salaire ce qui s'avère coûteux en termes actualisés.

Bien entendu, ce cadre d'analyse ne reflète qu'imparfaitement la structure de choix qui s'offre à un lycéen de seize ans. On peut arguer que la décision de poursuivre ses études peut être poussée par d'autres motifs qu'un argument monétaire : le goût pour des travaux intellectuels, le plaisir d'apprendre, la reconnaissance sociale, etc. Par ailleurs, même si l'on ne considère que la dimension monétaire, le choix définitif de l'étudiant ne s'effectue pas à seize ans. Il suit plutôt un processus séquentiel : une fois passé son Bac, il se demande s'il est rentable de continuer ses études encore une année, de même lorsqu'il a obtenu son DEUG, il se demande s'il est rentable de continuer une année, etc. Ceci conduit à réviser le calcul du taux de rendement qui préside au choix de continuer ou non ses études en prenant comme coût d'opportunité non plus la situation des sans diplôme, mais celle des personnes qui arrêtent leurs études au niveau qu'a atteint l'étudiant. Dans ce cadre, l'analyse proposée par Mincer et révisée par Heckman *et alii* (2006) conserve toute sa pertinence. Néanmoins l'approche que nous proposons constitue un complément pour discuter des choix individuels d'éducatons en insistant sur la distribution des taux de rendement internes.

En effet, on peut penser que la structure des incitations monétaires qui s'impose à un lycéen peut jouer sur sa motivation à « investir » dans le système éducatif. On peut également penser que le lien entre incitation monétaire et motivation est sûrement plus important pour les lycéens issus de familles modestes pour lesquelles il est difficile de les soutenir financièrement. Or, dans un processus en partie auto-réalisateur, l'investissement précoce dans le système éducatif constitue un des éléments important pour réussir ses études dans le supérieur. Dans ce cadre, la structure des incitations monétaires pour des lycéens issus de familles modestes est en défaveur des études supérieures lorsque l'on considère le taux de rendement médian qui est plus élevé que les autres diplômes considérés (si l'on exclut les écoles de commerce, les écoles d'ingénieur, les DUT et les BTS). Le rendement moyen est quant à lui largement supérieur à tous les autres diplômes : cette différence importante entre le revenu moyen et le revenu médian indique une hétérogénéité des rendements plus importante pour les Bac que pour les autres diplômes. Cet indice d'un plus grand risque est confirmé par le fait que la part de rendement négatif des titulaires d'un Bac reste supérieure à celle des diplômes qui ont un taux de rendement médian comparable (2- 4% versus 4,9-6,2%).

De ce point de vue, l'absence d'un investissement dans le système éducatif nécessaire aux études supérieures pour des lycéens sensibles aux rendements monétaires de leurs études est soutenue par l'espérance de rendement monétaire élevée du Bac ; elle dénote cependant une plus faible aversion aux risques.

3-2-L'estimation de la part liée au différentiel d'espérance de vie

L'avantage de la méthode par microsimulation est qu'elle permet de faire une décomposition des différentes composantes du rendement. Pour compléter notre illustration, nous montrons comment la différence d'espérance de vie moyenne entre différents diplômes contribue à renforcer ou non les constats précédant. Là encore notre objectif n'est pas de fournir une évaluation de l'utilité d'une année de vie, mais plutôt de fournir un indicateur monétaire permettant de discuter de la structure des risques de rendement de l'éducation et des incitations sous-jacentes.

Dans notre cadre d'analyse, il est possible d'identifier la part des rendements de l'éducation directement imputable au différentiel d'espérance de vie selon le diplôme. En effet, la part de rendement liée à l'espérance de vie d'un individu i avec un diplôme p peut se calculer de la manière suivante :

$$p_{di} = 1 - \frac{t_{0i}}{t_{di}} \quad (10)$$

où t_{di} est le taux de rendement interne de i avec l'espérance de vie moyenne correspondant à son diplôme, et t_{0i} le rendement interne de i avec l'espérance de vie moyenne d'un individu sans diplôme. Les résultats de ces estimations sont présentés dans le tableau 7.

Tableau 7 : Part des rendements de l'éducation liée à la différence d'espérance de vie d'avec les sans diplômes

Diplôme	Part de la moyenne	Part de la Médiane	Différence en point de la part de rendements négatifs
Niveau Bac	0.2%	1.0%	-0.002
Bac Général	0.3%	1.1%	-0.005
Bac Professionnel	0.3%	0.0%	0.000
Bac Technique	0.2%	1.0%	0.000
Bac +2	0.5%	1.1%	-0.001
DEUG	0.7%	0.0%	-0.002
DUT/DEUST	0.4%	0.0%	-0.001
BTS	0.5%	0.0%	-0.001
Autre diplôme de technicien sup.	0.9%	1.5%	-0.002
Diplôme paramédical	0.3%	0.0%	-0.001
Bac +3	0.8%	1.4%	-0.001
Licence	0.9%	1.5%	-0.002
Autre diplôme Bac+3	0.5%	0.0%	0.000
Bac +4	0.9%	0.0%	-0.002
Maîtrise	0.9%	0.0%	-0.002
Bac +5	0.8%	1.1%	0.000
DEA	1.1%	1.5%	0.000
DESS	0.8%	1.2%	-0.001
Ecoles de commerce	0.5%	0.0%	0.000
Ecoles d'ingénieurs	0.7%	0.0%	0.000
Plus de Bac +5	2.0%	1.9%	0.000
Doctorat (sauf Médecine)	2.0%	1.9%	0.000

Source : calcul des auteurs.

Champ : génération née en 1970.

Note : la simulation est effectuée sur la base de 34 643 observations représentatives de la génération 1970. Le rendement des écoles de commerce et de certaines autres écoles est surévalué du fait de la non prise en compte des coûts de scolarité propres à l'école.

Nos résultats montrent qu'en moyenne, la part des rendements de l'éducation liée directement à l'espérance de vie est plutôt croissante avec le niveau d'éducation. Ce résultat général masque cependant une forte hétérogénéité selon le diplôme considéré.

Par ailleurs, il apparaît que la part des rendements directement liée à l'espérance de vie est relativement faible. Elle représente entre 0,2% et 2% des rendements moyens –de 0+ à 1,9% pour les rendements médians. Par ailleurs, la part de personnes dont le rendement devient positif du fait d'une augmentation de l'espérance de vie est quasi nulle quel que soit le diplôme considéré : elle atteint au maximum cinq centièmes de points de pourcentage.

Bien entendu, cette analyse comprend certaines limites. En effet, comme nous l'avons rappelé dans la section 1.2, les études sur les déterminants de la santé soulignent qu'il existe un impact propre du revenu, indépendant du niveau d'éducation. Sur la base des données dont nous disposons, il n'est pas possible d'appréhender ce lien. On peut alors penser que pour les titulaires d'un Bac dont l'hétérogénéité en termes de rendement monétaire de l'éducation est forte, l'application d'une fonction de survie moyenne propre au diplôme masque en fait une hétérogénéité plus grande des rendements de l'éducation que celle qu'il nous est possible d'appréhender. Néanmoins, l'ampleur des résultats sur la part des rendements l'éducation directement lié à l'avantage en termes d'espérance de vie des diplômés indique qu'il est peu probable que ces effets remettent en cause les conclusions établies dans la section 2.1.

4-Conclusion

Cet article montre comment il est possible, de produire une description stylisée de la diversité des trajectoires individuelles d'une génération en fonction du diplôme des individus qui la composent à l'aide d'un modèle de microsimulation dynamique. Ce cadre permet de produire une description de la structure des risques de transition entre système éducatif et marché du travail. Une analyse de la distribution des taux de rendement interne individuels nous conduit à confirmer certains résultats sur les rendements différenciés de l'éducation à un niveau donné selon la filière suivie et d'opposer ainsi les diplômes universitaires à ceux des écoles (BTS, école de commerce, école d'ingénieur). Elle permet également de discuter de la structure incitative sous-jacente aux liens entre le marché du travail et le système éducatif : pour les lycéens issus des familles modestes plus sensibles au rendement monétaire de leur formation, le choix de s'arrêter au Bac apparaît plus rentable même s'il s'avère plus risqué.

L'analyse de cette structure incitative permet de formuler une hypothèse sur son influence pour expliquer la faible participation des étudiants issus des familles modestes à l'éducation du supérieur. Cette structure incitative peut conduire à nourrir un sous-investissement des lycéens issus des familles modestes dans l'enseignement secondaire. Dans la mesure où l'investissement dans les études secondaires s'avère nécessaire à la réussite des études

supérieures cette structure du risque de rendement de l'éducation peut s'avérer partiellement auto-réalisatrice.

Du point de vue des recommandations de politique économique, notre analyse des risques monétaires de carrière indique qu'une politique volontariste de développement de l'éducation supérieure comme celle prônée par la Stratégie de Lisbonne, risque de buter dans son extension aux étudiants issus des familles les plus modestes, à des contraintes de rentabilité de la formation. La réussite de cette stratégie passe alors par un desserrement de la contrainte budgétaire pour ces étudiants, dans une mesure qu'il conviendra d'explorer plus avant.

Références

Anxo Dominique, Erhel Christine (2008), « Irréversibilité du temps, réversibilité des choix ? Les fondements des 'marchés transitionnels' en termes de trajectoires de vie », *Revue Française de Socio-économie*, n°1, p.199-219.

Blanchet Didier (1998), « La microsimulation appliquée à l'analyse des politiques sociales », *Economie et Statistique*, n°315, p.29-34.

Cutler David, Lleras-Muney Adriana (2006), "Education and Health: Evaluating theories and evidence", *NBER working paper*, n° 12352.

Gazier Bernard, Schmid Günther (eds) (2002), *The dynamic of Full Employment. Social Integration by Transitional Labour Markets*, Edward Elgar, Cheltenham.

Gollac Michel, Volkoff Serge (2007), *les conditions de travail*, La découverte.

Goux Dominique, Maurin Eric (1994), « Education, expérience et salaire : tendances récentes et évolution de long terme », *Economie et Prévision*, n° 116, 1994-5, p.155-178.

Grossman Michael (2006), « Education and Nonmarket Outcomes », in Hanushek Eric, Welch Finis (eds), *Handbook of the Economics of Education*, vol. 1, Elsevier, p.577-633.

Heckman James J., Lochner Lance J., Todd Petra E. (2006), "Earning functions, rates of return and Treatment Effects: The Mincer equation and beyond, in Hanushek Eric, Welch Finis (eds), *Handbook of the Economics of Education*, vol. 1, Elsevier, p.307-458.

Insee (1999), « Le modèle de microsimulation Destinie », document de travail Insee, division redistribution et politique sociale.

Kunst A., Mackenbach J. (1994), "The Size of Mortality Differences Associated with Educational Level in Nine Industrialised Countries", *American Journal of Public Health*, 84 (6), p. 932-937.

Legendre François, Lorgnet Jean-Paul, Thibault Florence (2003), « Que peut-on retenir de l'expérience française en matière de microsimulation ? » *Economie et prévision*, n°160-161, p.I-XV.

Mackenbach Johan P., Bos Vivian, Andersen Otto, Cardano Mario, Costa Giuseppe, Harding Seeromanie, Reid Alison, Hemström Örjan, Valkonen Tapani, Kunst E. Anton (2003)

« Widening socioeconomic inequalities in mortality in six Western European Countries », *International Journal of Epidemiology*, 32 (5), p.380-387.

Mackenbach Johan P., Stirbu Irina, Roskam R. Albert-Jan, Schaap M. Maartje, Menvielle Gwenn, Leinsalu Mall, Kunst E. Anton (2008), « socioeconomic Inequalities in Health in 22 European countries », *The New England Journal of Medicine*, 358, p.2468-2481.

Maurice Marc, Sellier François, Silvestre Jean-Jacques (1982), *Politique d'éducation et organisation industrielle en France et en Allemagne. Essai d'analyse sociétale*, Presse Universitaires de France.

Menvielle Gwenn, Chastang Jean-François, Luce Danièle, Leclerc Annette, (2007), « Evolution temporelle des inégalités sociales de mortalité en France entre 1968 et 1996. Etude en fonction du niveau d'études par cause de décès », *Revue d'Epidémiologie et de Santé Publique*, 55, p.97-105.

Mincer Jacob (1974), *Schooling, Experience and Earning*, Columbia University Press, New-York.

Pison Gilles (2005), « France 2004 : l'espérance de vie franchit le seuil des 80 ans », *Population et Sociétés*, n° 410.

Poubelle Vincent, Albert Christophe, Beurnier Paul, Couhin Julie, Grave Nathanaël (2006), « Prisme, le modèle de la Cnav », *Retraite et Société*, n°48, p.202-215.

Robert-Bobée Isabelle (2006), « Projection de population 2005-2050 pour la France métropolitaine », Document de travail Insee, Méthode et résultats, n° F0603.

Robert-Bobée Isabelle, Cadot O. (2005), « Mortalité aux grands âges : encore des écarts selon le diplôme et la catégorie sociale », *Insee Première*, n° 1122.

Robert-Bobée Isabelle, Monteil Christian (2005), « Quelles évolutions des différentiels sociaux de mortalité pour les femmes et les hommes ? Tables de mortalité par catégorie sociale en 1975, 1982 et 1990 et indicateurs standardisés de mortalité en 1975, 1982, 1990 et 1999 », Document de travail Insee, n° F0506.

Schmid Günther (2006), « Sharing Risk Management through Transitional Labour Markets », *Socio-economic Review*, 4, p.1-33.

Schnittker Jason (2004), « Education and the Changing Shape of the Income Gradient in Health », *Journal of Health and Social Behavior*, 45, p.286-305.

Vallin J., Meslé F. (2001), *Tables de mortalité françaises pour les XIX^e et XX^e siècles et projections pour le XX^e siècle*, INED.