

La mesure de la singularité des trajectoires

Spécificités individuelles et histoire sociale dans la probabilité de connaître le chômage dans la seconde moitié du XXe siècle

*Thomas Soubiran**

Cette communication a pour objet l'analyse des transitions vers le chômage en France au cours de la trajectoire professionnelle. Il s'agit d'inscrire ce phénomène dans une double historicité, celle de la trajectoire des salariés et celle des espaces sociaux dans lesquels ils évoluent et ce, afin de dégager certaines interactions. Les histoires individuelles sont faites d'évolutions et de ruptures dont les effets positifs ou négatifs peuvent être ressentis plusieurs années après leur occurrence. Mais ces inflexions ne trouvent pas nécessairement leur origine dans l'histoire individuelle proprement dit mais peuvent découler du contexte social, économique ou historique. Ainsi, la seconde moitié du XX^e siècle a été marquée par une rupture au cours des années soixante-dix où, à une période de plein emploi succède une période de croissance puis de maintien du chômage à des niveaux élevés. Le « choc » pétrolier d'octobre 1973 a ouvert une nouvelle période rendant le chômage globalement plus probable au cours de la trajectoire professionnelle. Cependant, si le contexte historique détermine des tendances, il constitue aussi (avec l'histoire individuelle) une source de variabilité des trajectoires les unes par rapport aux autres mais aussi au sein même de chaque trajectoire. À partir des données de l'enquête Histoire de vie - Construction des Identités, l'analyse se propose donc de mettre en lumière trois types d'effets en les considérant dans leurs interactions : les effets de période (en insistant sur la rupture de l'année 1973), les effets de rémanence de la trajectoire professionnelle et l'appréhension des différences intra et interindividuelle. De par la période qu'elle couvre, l'enquête Hdv permet ainsi d'envisager l'endogénéité des trajectoires et la variation de ses effets dans les temps individuels et historique. De plus, l'utilisation de modèles linéaires généralisés mixtes permet d'étendre ces interactions aux différences intra et interindividuelles et de montrer le caractère social et historique de l'individualité.

La communication sera organisée comme suit. La première partie développera les avantages de l'analyse au long cours et ce, d'un point de vue général et relativement à l'étude des transitions vers le chômage. Ces développements serviront de fondement à l'explicitation des propriétés du modèle utilisé. Les deux dernières parties procéderont, elles à la description des données et à la présentation des résultats, respectivement.

1. Les apports des analyses dans le temps long

Comme on pu le noter en leur temps des auteurs comme Karl Marx ou Max Weber, les phénomènes sociaux ont un caractère intrinsèquement historique. En effet, si les individus et les organisations sont marqués par des traits distinctifs (et constitutifs), ceux-ci ont un caractère plus ou moins pérenne. Les individus changent au fur et à mesure qu'ils avancent en âge en même temps que les contextes de leur trajectoire évoluent. Et, si les choses peuvent se voir dotées d'une certaine substance, celle-ci peut se trouver à son tour substantiellement altérée par la succession des événements. De ce point de vue, un des apports de l'analyse longitudinale est de pouvoir saisir comment des événements antérieurs influent le cours de la trajectoire, que ces événements soient relatifs à l'individu lui-même ou au contexte dans lequel il s'inscrit. Les trajectoires présentent ainsi un caractère fortement endogène. La probabilité de connaître le chômage à un moment donné est, entre autres facteurs, d'autant plus grande si l'individu en a déjà fait l'expérience précédemment. Les transitions peuvent aussi être influencées par d'autres événements, comme la mobilité professionnelle, ou par des événements provenant de dimensions de la trajectoire autres que la trajectoire professionnelle.

* Chargé d'études, centre associé au Céreq pour la région Île-de-France, thomas.soubiran@univ-paris1.fr.

L'analyse longitudinale permet, de plus, de distinguer empiriquement trois types de variations interindividuelles qui se trouvent confondues en coupe synchronique : les effets d'âge, les effets de cohorte et les effets de période. Cette distinction revêt une grande importance dans l'étude de phénomènes tels que le chômage. D'une part, le marché du travail français est marqué, notamment, par la concentration du chômage parmi les actifs les plus jeunes et les plus âgés. D'autre part, le premier choc pétrolier d'octobre 1973 consécutif de la guerre du Kippour ouvre une période où, au quasi plein emploi de l'après-guerre, succède une forte croissance du chômage. Cette phase de croissance fut elle-même suivie par d'autres phases où le taux de chômage varie mais demeure globalement élevé. Si l'augmentation du prix du pétrole décrétée unilatéralement par l'OPEP à la suite du conflit fut sans doute plus le déclencheur d'une crise latente qu'elle n'en fut la cause, cet événement géopolitique permet de distinguer une césure à partir de laquelle les conditions dans lesquelles se déroulent les trajectoires professionnelles changent radicalement. Ces changements impliquent donc des effets de période mais aussi de cohorte, du fait du changement des conditions dans lesquelles se déroule la socialisation des individus. Un autre apport de l'analyse longitudinale tient aussi dans la possibilité de caractériser la singularité des trajectoires individuelles. Car, si les caractéristiques sociales dessinent des communautés de destins, aucune trajectoire n'est strictement identique à une autre. Chaque individu s'inscrit dans une histoire qui lui est propre. Néanmoins, cette hétérogénéité individuelle n'en est pas pour autant a-sociale (Elias 1969). La notion d'individu telle qu'elle est comprise actuellement renvoie aux processus historiques qui ont vu la constitution de l'État moderne et de l'économie de marché ainsi que l'urbanisation progressive de la population. L'individualité est aussi le produit des normes juridiques qui encadrent l'activité humaine. Certains marchés d'emploi sont par exemple contrôlés par des normes qui font largement dépendre la carrière de critères univoques, comme le diplôme ou l'ancienneté, qui contribuent à réduire les différences entre les individus, qu'ils soient considérés comme membres de catégories ou comme individus proprement dits. À titre d'exemple, les trajectoires dans la fonction publique étant en partie déterminées par le diplôme, les différences salariales entre hommes et femmes y sont moins marquées. On peut ainsi envisager que certaines caractéristiques sociales contribuent tant à l'hétérogénéité interindividuelle qu'intra-individuelle. Une des conséquences de la crise fut la mise en œuvre de réformes des normes régulant le marché du travail dans le sens d'une plus grande « flexibilité » permettant un roulement de la main-d'œuvre plus important. Les facilités faites aux entreprises pour se séparer de leurs salariés a pour corollaire une plus grande dépendance des transitions vers le chômage de situations spécifiques relatives à l'entreprise et à la position du salarié en son sein.

2. La modélisation statistique dans le temps long

D'un point de vue statistique, l'analyse longitudinale comporte plusieurs avantages. Tout d'abord, elle permet d'augmenter la puissance des tests (Hedeker et Gibbons 2006), c'est-à-dire qu'un nombre moins important d'observations est nécessaire pour obtenir la même probabilité de rejeter une hypothèse nulle erronée (minimisation des erreurs de type II). De plus, lorsque l'on dispose de plusieurs observations pour un même sujet, chaque sujet devient, en quelque sorte, son propre contrôle. Comme certaines caractéristiques des individus changent dans le temps alors que d'autres demeurent constantes, l'analyse se rapproche de l'analyse expérimentale (« l'expérimentateur » n'ayant cependant pas la possibilité de choisir ce qui varie et ce qui ne varie pas et à quelle moment). Néanmoins, la répétition de mesure sur un même individu apporte aussi son lot de difficultés. Les théorèmes usuels de l'inférence statistique sont fondés sur le postulat que les observations ont été tirées aléatoirement et sont donc indépendantes, ce qui n'est plus cas lors de l'analyse de données provenant d'observations réitérées sur un même sujet. Les analyses se doivent a minima de corriger l'effet de grappe induit par la structure hiérarchique des données.

2.1. La structure du modèle

Différentes méthodes ont été proposées pour analyser ce type de données. D'un point de vue heuristique, on peut distinguer deux approches : une considérant la variance inter-individuelle comme une perturbation à corriger (GEE - Generalized Estimating Equations) (Hardin et Hilbe 2007) et une autre visant à caractériser cette variance ((G)LMM - (Generalized) Linear Mixed Models) (Snijders et Bosker 1999 ; McCulloch *et al.* 2006). Ces deux approches peuvent être considérées comme des extensions des modèles linéaires généralisés

(McCullagh et Nelder 1989 qui constituent eux-mêmes une généralisation de la régression multiple (LM - Linear Model) à des distributions d'erreurs non-gaussiennes (distributions binomiale, multinomiale, Poisson, gamma, etc.). C'est la seconde de ces approches qui a été adoptée ici. Elle présente en effet l'intérêt de pouvoir faire le lien avec certaines théories sociologiques des trajectoires et de fournir le moyen de leur opérationnalisation empirique. Dans plusieurs de ses travaux, P. Bourdieu (Bourdieu 1974 et Bourdieu 1979) mobilise la distinction suivante. Les trajectoires, entendues comme la succession des différentes positions occupées par un individu dans un espace social, peuvent être conçues comme le produit de l'interaction entre deux composantes : la trajectoire sociale et ce que l'on nommera ici la trajectoire singulière. La première composante découle de l'effet des caractéristiques sociales sur les positions occupées telles que le sexe, la cohorte, la condition sociale, l'origine géographique, etc. La seconde découle, elle, d'éléments plus ineffables de l'histoire de l'individu comme son rapport au monde, son histoire familiale ou son environnement immédiat. En utilisant la notation des GLMM, cette distinction peut être formalisée de la façon suivante :

$$\eta_i = \mathbf{X}_i\beta + \Gamma\mathbf{v}_i \quad (1)$$

η_i désigne la fonction (qui sera définie plus loin) liant les états connus par un individu i en fonction de ses caractéristiques, \mathbf{X}_i désigne ses caractéristiques sociale, β le vecteur de paramètres de régression qui leur sont associés (aussi appelés effets fixes), \mathbf{v}_i les paramètres des pentes spécifiques à chaque individu et Γ la matrice liant ces paramètres à d'autres caractéristiques de l'individu.

Les GLMM se distinguent donc des GLM par l'ajout de paramètres \mathbf{v}_i propres à chaque individu. Si les \mathbf{v}_i étaient observés, les paramètres du modèle seraient estimables au moyen des méthodes utilisées pour les (G)LM ce qui n'est, en pratique, pas le cas. Ce problème peut néanmoins trouver une solution en considérant que les \mathbf{v}_i sont des variables latentes suivant une distribution spécifique $g(\mathbf{v}; \Theta)$ de paramètre Θ :

$$L(\mathbf{Y}_i) = \int f(\mathbf{Y}_i | \mathbf{v}_i) g(\mathbf{v}_i; \Theta) d\mathbf{v}_i \quad (2)$$

$f(\mathbf{Y}_i | \mathbf{v})$ désigne ici la fonction de densité liant le prédicteur linéaire et la variable dépendante (cf. éq. 1). Dans ce cas, le modèle n'estime pas les \mathbf{v} correspondant aux effets individuels mais les paramètres contrôlant leur distribution. Ils quantifient donc le montant de la variance interindividuelle des trajectoires¹. Dans le cas le plus simple, le modèle ne compte qu'une seule variable latente, mais d'autres peuvent être ajoutées, par exemple pour prendre en compte les fluctuations de la variance interindividuelle dans le temps. Le choix le plus courant pour $g(\mathbf{v}; \Theta)$ est une distribution normale d'espérance nulle avec Σ_q comme matrice de covariance (q désigne ici le nombre de variables latentes introduites dans le modèle)² :

$$L(\mathbf{Y}_i) = \int f(\mathbf{Y}_i | \mathbf{v}_i) \phi_q(\mathbf{0}; \Sigma_q) d\mathbf{v}_i \quad (3)$$

On notera que dans le cas linéaire (LMM), c'est-à-dire lorsque $f()$ est aussi supposée obéir à une loi normale, cette expression a une solution analytique du fait que $f()$ est le produit d'une distribution normale multivariée. Mais cette situation est davantage l'exception que la règle et, dans la plus part des cas, l'analyse doit recourir à une approximation de $g()$ ³.

La conception des trajectoires mise-en-œuvre par ce modèle est la suivante. Les trajectoires observées sont la réalisation de la conjonction (et, on le verra, de l'interaction) entre la trajectoire sociale et la trajectoire singulière. Les caractéristiques sociales déterminent la direction générale prise par la trajectoire et les caractéristiques singulières la font plus ou moins substantiellement dévier de cette direction. La trajectoire singulière correspond donc à l'écart à la trajectoire probable imputable aux individus eux-mêmes.

¹ Les \mathbf{v}_i peuvent néanmoins être estimés à partir de ces paramètres en utilisant le théorème de Bayes.

² D'autres distributions sont néanmoins envisageables.

³ Les modèles ont été estimés au moyen des quadratures de Gauss-Hermite (Pinheiro et Bates 1995).

2.2. Définition du lien entre les états constituant la trajectoire et les caractéristiques sociales

Le paradigme des GLMM permet d'articuler les composantes sociales et singulières des trajectoires. Dans ce cadre, différentes solutions sont envisageables pour définir le lien entre les caractéristiques individuelles et leurs trajectoires. La relation entre les caractéristiques sociales d'un individu et les états par lesquels il transite peut notamment être définie au moyen de modèle de durée (Courgeau et Lelièvre 1989; Blossfeld et Rohwer 2002). Ces modèles peuvent prendre différentes formes. La plus courante est celle dite des taux de transition proportionnels (*Proportional Hazard Models*):

$$\lambda_i(t | \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i \beta} \quad (4)$$

$\lambda(t)$ désigne ici la fonction mesurant le taux de transition instantané (et λ_0 désigne la même fonction pour un groupe servant de référence). $\lambda(t)$ correspond à la limite du rapport entre la probabilité qu'un événement survienne entre t et $t + dt$ conditionnellement au fait qu'il ne soit pas survenu au préalable et l'intervalle dt :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{Pr(t < T \leq t + dt | T > t)}{dt} \quad (5)$$

Cette relation peut être reformulée comme le rapport entre la probabilité jointe que t appartienne à $(t, t + dt)$ avec $T > t$ et la probabilité que $T > t$:

$$\lambda(t) = \frac{f(t)}{S(t)} \quad (6)$$

En d'autres termes, $\lambda(t)$ est égal à la densité d'un événement en t rapportée à la probabilité $S(t)$ d'avoir « survécu » jusque là. Si on postule que t suit une distribution particulière $f(t)$ ayant pour fonction de probabilité cumulée $F(t)$, $S(t)$ peut être définie comme :

$$\begin{aligned} S(t) &= Pr(T > t) \\ &= 1 - F(t) \\ &= \int_t^{\infty} f(x) dx \end{aligned} \quad (7)$$

On notera enfin que $\lambda(t)$ est la dérivée de $S(t)$. Ces définitions amènent plusieurs remarques. Tout d'abord, ces modèles analysent la distribution du nombre de transitions observées en fonction du temps. Ils se focalisent donc sur les transitions et, plus spécifiquement, sur les rythmes de transition d'un état à un autre (soit ce qui fait varier le temps écoulé jusqu'à l'événement). Surtout, ils postulent que les transitions auront lieu nécessairement à un moment ou un autre (et ce, du fait que $\lambda(t)$ est la dérivée de $S(t)$ et que $S(\infty)=1$, soit la probabilité de connaître l'événement est égale à un lorsque le temps écoulé tend vers l'infini). De façon réciproque, le modèle postule que tous les individus sont exposés au « risque » de connaître l'événement et ne se distinguent que par le temps qui les sépare de l'occurrence de celui-ci.

Ce postulat pose un problème connu sous le nom de censure à droite. En pratique, les périodes d'observation sont limitées dans le temps et une partie de l'échantillon peut ne pas connaître l'événement. Différentes méthodes peuvent être utilisées pour prendre en compte les phénomènes de censure à droite, la plus simple consistant à considérer le phénomène comme aléatoire. Les modèles de durée sont appropriés pour l'étude de certains aspects de phénomènes comme la nuptialité ou la mobilité professionnelle. Cependant, le postulat que la transition interviendra nécessairement est, pour le cas qui nous occupe, difficilement tenable. S'il ne peut pas être considéré comme un événement rare, le chômage est néanmoins un état qui ne touche potentiellement qu'une partie de la population active. Ainsi, dans l'enquête *Hdv*, 24% des enquêtés ont déclaré au moins un épisode de chômage. À cela s'ajoute le fait que les individus peuvent connaître plusieurs transitions par le chômage à différents moments de leur trajectoire.

C'est pourquoi il a semblé nécessaire de reformuler la question et de ne pas tant chercher à savoir quand, mais qui connaît la transition à un moment donné de la période observée. Pour ce faire, les trajectoires ont été analysées au moyen d'un modèle binomial :

$$\Pr(Y_{ij}) = p_{ij}^{Y_{ij}} (1 - p_{ij})^{1-Y_{ij}} \quad (8)$$

$$\text{avec } p_{ij} = \Pr(Y_{ij} = 1) = \frac{1}{(1 + e^{-\eta_{ij}})}$$

Le modèle mesure donc la probabilité de faire la transition vers le chômage et, le cas, échéant, la probabilité d'y rester d'une année sur l'autre. Ce modèle présente en effet l'avantage de pouvoir intégrer l'historique de la trajectoire sous la forme de variables renseignant des événements antérieurs et d'appréhender le processus comme une chaîne de Markov. La formulation en termes de modèle logit (ou probit) présente aussi l'intérêt de pouvoir faire le lien avec une autre catégorie de modèles (IRT -*Item Response Theory*). S'ils n'ont pas été développés pour l'analyse longitudinale mais dans une perspective psychométrique, ces modèles reposent sur des fondements statistiques similaires (Molenberghs et Verbeke 2006). Ce lien permet de mobiliser les concepts et les méthodes de l'IRT dans le cadre de l'analyse longitudinale. L'objet du modèle de Rasch est de mesurer des capacités (« habiletés », etc.) à partir d'une batterie de tests en contrôlant par les effets marginaux (en l'occurrence, la probabilité de donner une réponse correcte à un item est globalement plus ou moins grande en fonction de la difficulté, quelle que soit la capacité du répondant). Il permet notamment d'envisager les GLMM comme des modèles de mesure permettant de dégager les facteurs spécifiques aux individus qui leur font connaître le chômage en contrôlant par les effets marginaux (le taux de chômage à un moment donné) et par les caractéristiques des individus. Ils permettent de plus d'élargir la gamme des modèles utilisables dans le cadre de l'analyse longitudinale.

2.3. Variance intra et interindividuelle

Dans le modèle qui vient d'être esquissé, si caractéristiques sociales et singulières agissent en conjonction, elles ne s'influencent pas mutuellement. La variance interindividuelle est constante quel que soient les propriétés sociales des individus ou le contexte dans lequel il évolue. Cette hypothèse apparaît cependant restrictive. On a vu précédemment que les régulations du marché du travail s'étaient affaiblies donnant plus de poids aux configurations immédiates dans lesquelles s'inscrit l'individu. De ce fait, on peut envisager que la variance interindividuelle ait pu s'accroître après 1973. Plus généralement, les caractéristiques sociales, en tant que propriétés sociales, confèrent à leurs détenteurs une emprise variable sur les choses. De par les asymétries de distribution des capitaux sociaux, les individus sont structurellement inégaux quant à leur possibilité de contrôler l'aléa du cours des choses et de maîtriser leur destin.

Dans le cadre du modèle linéaire, ces phénomènes d'interaction sont désignés par le terme d'hétéroscédasticité de la variance (Snijders et Bosker 1999). Dans l'IRT, ils ont pour nom fonctionnement différentiel de l'item (DIFF) (de Boeck et Wilson 2004). Ils peuvent être pris en compte en ajoutant un effet d'interaction entre la ou les variable(s) soupçonnées d'interférences et la variable latente. Si cette variable a deux modalités, l'interaction a pour forme :

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} v_i$$

Si l'individu i est caractérisé par la première modalité de la variable. A chaque catégorie correspond alors un paramètre de variance spécifique. Ces interactions peuvent être aussi appliquées à des variables continues et au cas multivarié (Goldstein 2003). De façon symétrique, on peut aussi envisager que les effets fixes peuvent aussi influencer la variance intra-individuelle, soit, non pas les différences entre les individus mais les différences des individus par rapport à eux-mêmes d'une année à l'autre. Cela est rendu possible par une simple modification de la fonction de lien du modèle. La formule généralement utilisée pour la distribution logistique correspond dans les faits à sa version standardisée (centrée réduite) qui est un cas particulier de l'expression plus générale :

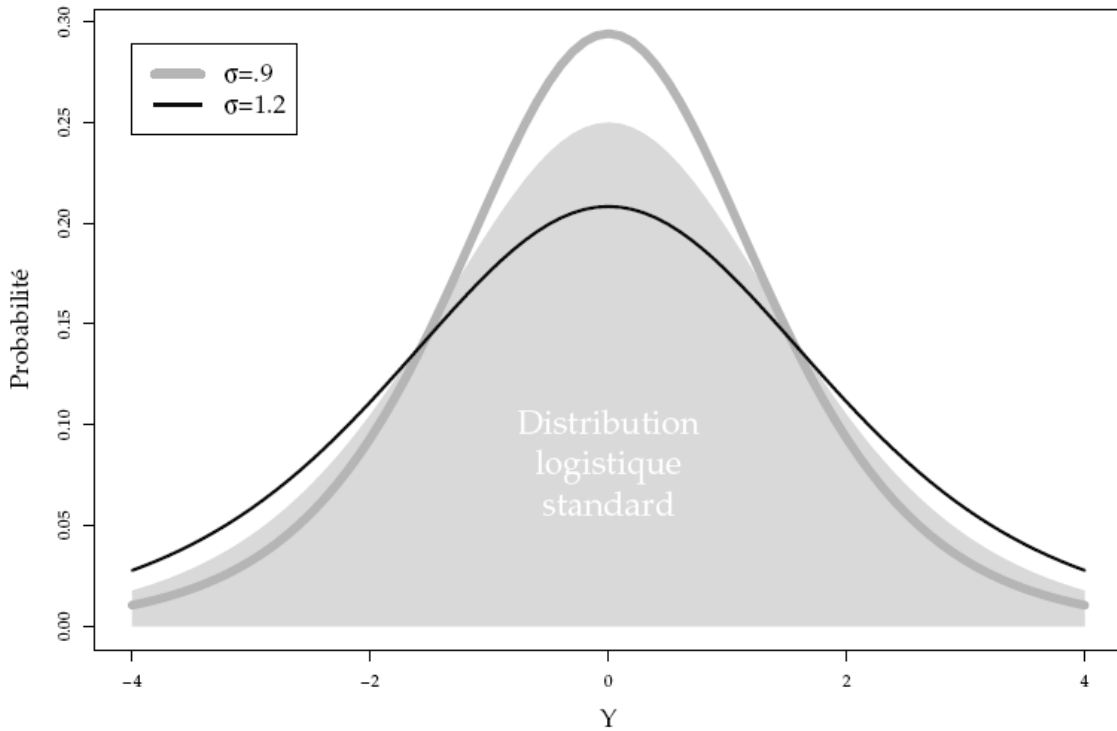
$$p_{ij} = Pr(Y_{ij} = 1) = \frac{1}{(1 + e^{-z_{ij}})} \quad (9)$$

$$\text{avec } z_{ij} = \frac{\eta_{ij} - \mu_{ij}}{\phi_{ij}}$$

μ_{ij} désigne le paramètre de position et ϕ_{ij} , le paramètre de dispersion de la distribution. Elle a pour espérance μ_{ij} et pour variance, $(\pi^2/3)\phi^2$. La figure 1 montre la forme de la courbe pour différentes valeurs de ϕ . Plus la courbe est évasée et plus la variance intra-individuelle est importante. Le paramètre de dispersion permet donc d'intégrer la variance intra-individuelle dans le modèle en divisant le prédicteur linéaire (cf. éq. 1) par ϕ (Hedeker & al. 2006) :

$$\eta_i = \frac{\mathbf{X}_i\boldsymbol{\beta} + \Gamma\mathbf{v}_i}{\phi_i} \quad (10)$$

FIGURE 1 – Distribution logistique en fonction de différentes valeurs de ϕ_i



En outre, cette reformulation permet de mesurer l'importance relative que prend la variance interindividuelle dans le phénomène observé. Cette mesure est couramment utilisée dans le cadre du modèle linéaire où, dans le cas d'une seule variable latente, la variance totale de Y est égale à :

$$V(\mathbf{Y}) = \sigma_v^2 + \sigma_\epsilon^2 \quad (11)$$

σ_v^2 désigne la variance inter-individuelle et σ_ϵ^2 désigne ici la variance intra-individuelle.

En exprimant la covariance en termes de corrélation, on obtient une mesure connue sous le nom de coefficient de corrélation intra-classe (ICC) ou VPC (*Variance Partition Coefficient*) :

$$ICC = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2} \quad (12)$$

Cette mesure, issue de l'analyse de la variance, indique la proportion de la variance inexpiquée par les effets fixes du modèle attribuable aux individus. À l'instar de l'espérance, la dispersion de la variable dépendante peut être modélisée en fonction des caractéristiques de l'individu :

$$\eta_i = \frac{\mathbf{X}_i \boldsymbol{\beta}_\mu + \boldsymbol{\Gamma} \mathbf{v}_i}{\exp(\mathbf{X}_{v,i} \boldsymbol{\beta}_\epsilon)} \quad (13)$$

La combinaison de ce modèle avec le modèle hétéroscédastique permet de calculer l'ICC pour chacune des variables. Cela peut être réalisé en remplaçant, dans la formule 12, σ_v^2 par $\sigma_{v,x}^2$ (soit la variance correspondant à l'interaction entre cette variable et la variable latente) et en remplaçant σ_ϵ^2 par $\sigma_{\epsilon,x}^2$, $\sigma_{\epsilon,x}^2 = (\pi^2/3)\phi^2$ avec $\phi = \exp(\beta_x)$ (soit la variance intra-individuelle induite par la variable x).

3. Présentation des données

3.1. Données

3.1.1. L'enquête Histoire de vie

L'enquête Hdv a été réalisée en 2003 par l'Insee. Elle est le produit de la collaboration de six établissements public : l'Institut national d'études démographiques (Ined), quatre services statistiques ministériels - la Direction (Drees) et (Dares) du ministère du travail, la Division des études et de la prospective (Dep) du ministère de la Culture, (Div), ainsi que l'Institut National de la santé de la recherche médicale (Inserm). L'enquête proprement dite a été réalisée par l'Insee. Le questionnaire se compose de deux grandes parties : une portant sur « l'identité » des individus et l'autre portant sur leur trajectoire. Les analyses s'appuieront essentiellement sur cette dernière partie. Elle renseigne l'occurrence de différents types d'événements :

- Familiaux : mise(s) en couple, séparation(s), départ de chez les parents.
- Migratoires : mobilité géographique inter et intra-nationale, acquisition de la nationalité
- Professionnels : statut d'activité, profession exercée (qualifiée avec la PCS à deux chiffres)

Les événements ont été recueillis au moyen d'une grille biographique complétée par des questions insérées dans le questionnaire.

3.1.2. Plan de sondage

Le recueil des données a suivi un protocole différent des enquêtes habituellement réalisées par l'Insee. L'Insee procède généralement à des tirages stratifiés dans l'EM (Echantillon-Maître) selon la méthode Kish. La base de sondage individuelle utilisée pour l'enquête Hdv est issue d'échantillons constitués pour d'autres enquêtes - les enquêtes Histoire familiale (EHF) et Vie quotidienne et santé (VQS). D'après la documentation de l'enquête, « *Le recours à cette base de sondage était justifié par la nécessité de procéder à des surreprésentations de certaines populations (les personnes nées à l'étranger, celles de parents nés à l'étranger et les personnes handicapées de moins de 60 ans). En effet ces catégories d'individus présentent un intérêt particulier en termes d'identité et doivent pouvoir faire l'objet d'études spécifiques dans le cadre de cette enquête. Or, ces groupes sont faiblement représentés dans la population générale. Il était donc nécessaire de les surreprésenter afin d'améliorer la précision des analyses les concernant. Les informations nécessaires à cette sur représentation sont absentes du recensement de la population. Les enquêtes Histoire familiale (EHF) et Vie quotidienne et santé (VQS) qui ont porté sur environ 1/10e de la population recensée,*

comprenaient des informations sur les lieux de naissance de l'individu et de ses parents (pour l'EHF) et le degré de handicap (pour VQS). Elles pouvaient donc constituer une base de sondage pour l'enquête Histoire de vie. »

Au final, la base de l'enquête Hdv décrit les trajectoires de 8 403 individus représentant au total 419 525 années de vie.

3.2. Constitution de l'échantillon analysé

Périodisation des temporalités historiques et individuelles

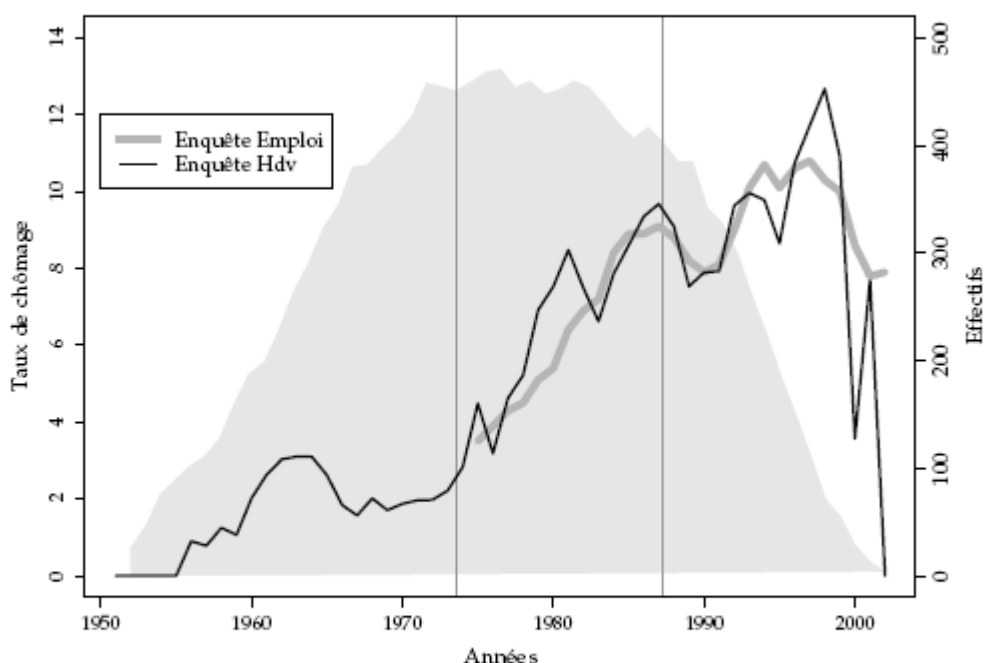
Du fait que l'échantillon a été construit à partir de la population française de 2003, toutes les années renseignées ne sont pas uniformément représentées dans les grilles biographiques. La distribution des années observées est ainsi marquée par la structure par âge de la population française à la date de l'enquête. L'individu le plus âgé de l'enquête est né en 1906. Les plus jeunes sont nés en 1985 (âge minimum de 18 ans à l'enquête). 90% des individus sont nés après 1930 et la médiane des naissances se situe en 1955. Le nombre d'années-vie augmente donc progressivement jusqu'aux années 1970 où il se stabilise puis décroît à partir du milieu des années 1980. Du fait que les années sont très inégalement représentées, les analyses ne pouvaient porter sur l'intégralité de la période couverte par l'enquête. La période d'observation a donc été limitée aux années 1950-2003. La borne inférieure a été établie en fonction du nombre d'années-vie et des changements de contexte économique. Les années trente et quarante du siècle dernier ont en effet été marquées successivement par la crise économique puis par la guerre avec des restrictions qui ont perduré jusqu'en 1947 (date à laquelle le rationnement a été interrompu). Les indices de production ont considérablement chuté pendant cette période. De plus, le marché du travail durant la guerre se trouve dans une situation particulière. Près d'un million et demi d'hommes sont retenus prisonniers en Allemagne. À cela s'ajoute la main-d'œuvre réquisitionnée dans le cadre du STO (Service du travail obligatoire) instauré par le régime de Vichy en 1942 à l'instigation de l'occupant. De plus, la production a été partiellement réorientée pour soutenir l'effort de guerre allemand. D'autre part, la population française est encore majoritairement rurale et l'activité économique s'articule largement autour de la production agricole. Les particularités de ces périodes combinées à faiblesse des effectifs disponibles dans l'enquête pour les années leur correspondant ont conduit à fixer le point de départ de l'analyse en 1950. Cette date intervient en effet entre la mise en œuvre du plan Marshall et avant la création de la CECA (Communauté économique de l'acier et du charbon) et marque l'ouverture d'une nouvelle période qui fut nommée rétrospectivement les Trente Glorieuses par assonance avec les journées de juillet 1830.

Les trajectoires ont été analysées en fonction du nombre d'années effectivement travaillées, soit le nombre d'années depuis la première transition vers le marché du travail (comme actif ou chômeur) défalquées du nombre d'années d'interruption. Cette métrique a été retenue parce qu'elle permet de synchroniser les trajectoires selon l'idée que le nombre d'années d'exercice comme actif constitue une procuration de l'expérience acquise. L'hypothèse est que l'accumulation de cette dernière accroît la valeur d'usage d'un individu et le protège contre le chômage. En outre, l'intérêt de la substitution du nombre d'années d'activité à l'âge est qu'elle permet de lever les problèmes d'identification liées au modèle ACP (Age, Cohorte, Période) sans avoir à contraindre la matrice des prédicteurs en utilisant, par exemple, des méthodes telles que « l'estimateur intrinsèque » (Yang et al. 2004). Ces artifices sont rendus nécessaires par le fait que l'année de naissance, la période et l'âge sont parfaitement colinéaires. La matrice des prédicteurs est alors de plein rang moins un. Dans ce cas, le système d'équations n'est plus identifié et peut avoir une infinité de solutions. En comptant les années d'activité à partir de la première transition vers le marché du travail, la corrélation entre les trois variables s'en trouve relâchée et ce, d'autant plus que de nombreuses trajectoires de l'enquête ont été interrompues par un événement ou un autre (service militaire, maladie ou accident, transition vers le statut de femme au foyer, etc.).

Les trajectoires analysées sont donc celles des individus qui ont commencé à travailler à partir de 1950 (afin d'analyser des trajectoires complètes) et qui ont au moins eu dix années d'activité professionnelle (afin d'obtenir une certaine profondeur historique tout en incluant des cohortes qui n'ont travaillé qu'après 1973).

Si cette méthode permet d'identifier le modèle, elle comporte aussi des inconvénients. Du fait des critères de sélection utilisés, les années de début et de fin de période sont faiblement représentées ce qui conduit à une distorsion des probabilités de connaître le chômage durant ces années (cf. figure 2). De plus, elle conduit à surreprésenter les plus diplômés dans les cohortes les plus anciennes. En prolongeant leurs études, ces derniers retardent d'autant leur agrégation au marché du travail par rapport à ceux qui sont nés la même année mais qui ont interrompu leur cursus scolaire plus tôt. La réciproque est vraie pour les moins diplômés pour l'autre extrémité de la période analysée. De plus, n'ont été introduites dans l'analyse que les années où les enquêtés étaient potentiellement exposés au risque de connaître une transition vers le chômage, en l'occurrence les années où les individus ont été salariés du privé. Cette restriction a été motivée par le fait que les marchés d'emploi de la fonction publique⁴ et des indépendants fonctionnent selon des règles différentes qui posent la cessation momentanée de l'exercice d'une profession sans retrait du marché du travail dans des termes différents. Le statut de fonctionnaire rend ainsi le licenciement d'un titulaire exceptionnel. La question d'éventuels biais de sélection liés à l'orientation vers un segment par rapport à un autre a, pour le moment, été écartée.

FIGURE 2 – Évolution du taux de chômage en France [1950-2003] et effectifs de l'échantillon par année



Lecture : En 1975, le pourcentage de chômeur s'élève à 3.5% selon l'enquête *Emploi*, à 4.5% selon l'enquête *Hdv* et le nombre d'années renseignées est de 468.

Champ : voir texte.

Source : enquête *Hdv* et enquête *Emploi*.

Une autre source de biais de sélection, plus spécifique aux femmes, peut être trouvée dans la participation au marché du travail. L'analyse des trajectoires d'installation à partir de l'enquête *Hdv* (Soubiran 2006) montre que la probabilité de transiter par le statut de femme au foyer augmente considérablement parmi les femmes des cohortes qui constitue le gros de l'échantillon de l'enquête. En effet, les trajectoires présentent un caractère sexué très marqué. C'est pourquoi les analyses ont été réalisées séparément pour les hommes et les femmes. Les résultats présentés ici ne portent que sur les hommes.

⁴ On notera qu'il existe un flou dans l'enquête qui ne fait référence qu'au « statut dans la profession » sans distinguer le statut du contrat liant le salarié à un établissement et le statut de cet établissement. Depuis les années 1990, différents types d'établissements publics tels que la Poste font ainsi un recours de plus en plus massif à des contrats privés.

3.3. Variables

Outre le nombre d'années d'activité professionnelle, les variables introduites dans les modèles sont les suivantes :

Cohorte : regroupement des années de naissance en quatre classes (≤ 1940 , [1940-1949], [1950-1959], [1960-1969])

Période : découpage à partir de quatre seuils (1973, 1987, 1993, 1998). Afin d'accommoder les discontinuités observables entre ces intervalles, quatre variables ont été construites selon le principe suivant (Singer et Willett 2003). Jusqu'au seuil, la variable prend pour valeur zéro et les valeurs supérieures comptent le nombre d'années écoulées depuis ce moment-là. D'autres solutions ont été testées comme des splines cubiques mais elles conduisaient à une moindre qualité d'ajustement⁵.

Mobilité géographique : le plan de sondage visait explicitement à surreprésenter les migrants. L'analyse distingue donc à la fois le pays de naissance (France, Afrique, autre)⁶ et si l'individu travaillait en France l'année considérée.

Situation l'année précédente : cinq modalités (actif du privé, actif d'autres segments, chômeur, inactif pour reprise d'études, inactif pour une autre raison)

Dernière profession occupée : nomenclature PCS à un chiffre (aussi appelée « GS » pour groupes sociaux) plus une modalité pour les années non renseignées qui correspondent à des périodes d'instabilité où la personne interrogée change fréquemment d'emploi/profession, ces changements pouvant être accompagnés d'épisodes de chômage.

Reprise d'étude : variable dichotomique prolongeant la modalité reprise d'étude de la variable décrivant la situation

Mobilité professionnelle : changement de profession à partir de la PCS la plus détaillée disponible (le code PCS à deux chiffres) et le nombre d'années écoulées depuis lors (Note : l'enquête ne renseigne pas les changements d'emploi au sein de la même PCS).

Situation conjugale : variable dichotomique (en couple ou non, quel que soit le statut).

4 Résultats

4.1 Modèle de référence

Le modèle de référence inclut le nombre d'années d'activité, la cohorte, l'âge à la première transition vers l'activité, la période et la mobilité géographique. Au regard des paramètres, la probabilité de connaître le chômage décroît avec le nombre d'années d'activité. Chaque nouvelle cohorte a plus de chance de connaître le chômage. Plus l'âge de la transition vers l'activité augmente plus les chances de chômage diminuent.

L'individu de référence est un homme né en France dans les années quarante, en couple, ouvrier ayant commencé sa carrière avant 1973, actif occupé l'année précédente et qui a changé de profession au moins une fois. Les individus d'origine africaine ont une plus grande probabilité de chômage que les individus d'origine française mais leur probabilité d'être au chômage est plus importante avant leur migration, ce qui semble indiquer une motivation économique à la migration.

⁵ Une autre possibilité consisterait à recourir aux méthodes semi-paramétriques.

⁶ L'échantillon de l'enquête devrait néanmoins autoriser des catégories plus fines dans des analyses ultérieures.

Les tendances sont analogues pour les individus originaires d'autres pays mais dans des proportions moindres.

4.2. Effets des variables supplémentaires

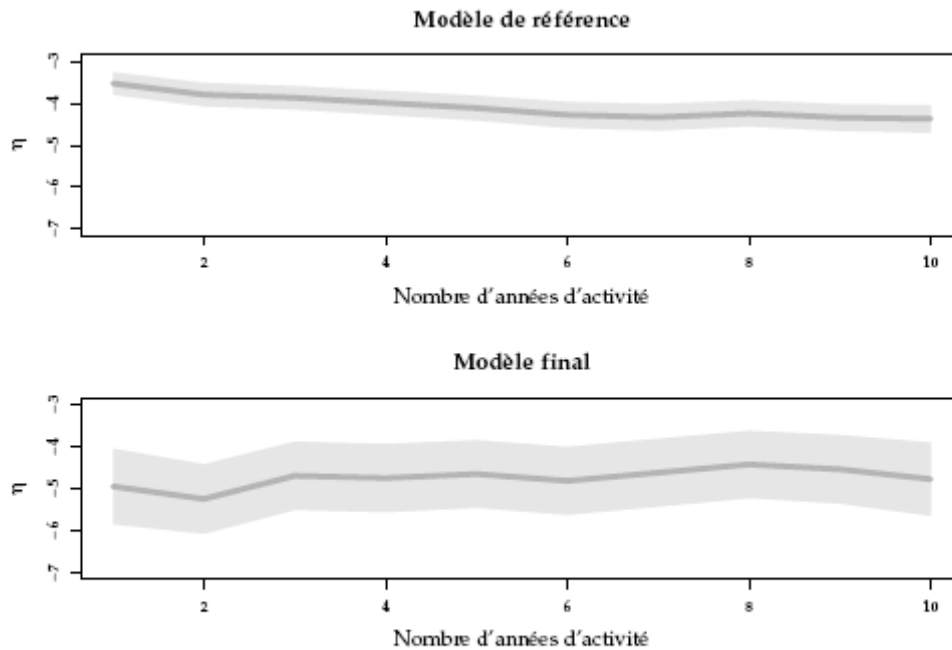
Le changement de profession en dehors des périodes d'instabilité a globalement un effet négatif sur la probabilité de connaître le chômage. Cet effet est sensible dès l'année suivant la transition et s'accroît avec le temps. Il est néanmoins tempéré par la période postérieure à 1973.

Être dans une situation autre que l'emploi en $t-1$ augmente la probabilité d'être au chômage, particulièrement pour ceux qui l'étaient déjà totalement ou partiellement l'année précédente. Le chômage dépend aussi de la dernière profession occupée. Les ouvriers et les employés ainsi que ceux dont la profession n'est pas renseignée du fait d'une situation d'instabilité professionnelle sont les plus exposés. Les professions intermédiaires et les « cadres » (GS=3) apparaissent plus protégés. Le croisement entre la situation antérieure et la dernière profession n'est pas significatif. Ces effets ne semblent pas plus varier dans le temps de la trajectoire, ou au fil des différentes périodes. Néanmoins, du fait de problèmes de codage apparents, les résultats relatifs aux groupes sociaux doivent être interprétés avec une certaine réserve.

L'année suivant une interruption d'activité pour une reprise d'étude est marquée par une probabilité de chômage plus importante. Par contre, la reprise d'étude éloigne du chômage avec le temps (interaction positive avec l'ancienneté de l'activité). L'effet de la situation antérieure varie en fonction de la période. Si les chômeurs de l'année précédente ont une probabilité globalement plus importante d'être au chômage, celle-ci augmente moins vite dans la période 74-87 pour eux, mais elle diminue aussi moins rapidement dans la période 88-93. La période 94-98 ne semble pas avoir d'effet à la différence de la dernière période où la probabilité d'être au chômage diminue plus rapidement pour eux. L'effet de la dernière profession ne semble pas, lui, varier dans le temps.

À ce stade, il est important de noter que l'introduction des deux variables caractérisant la situation professionnelles en $t-1$ a une incidence sur l'effet de l'ancienneté de l'activité. Comme le montre la figure 3, la courbe, de décroissante devient quasiment plate. Ainsi, une fois des effets de rémanence introduits dans le modèle, chaque année d'ancienneté supplémentaire ne diminuerait pas les chances de connaître le chômage.

FIGURE 3 – Logit de connaître le chômage en fonction du nombre d'années d'activité



Lecture : Au cours de la première année d'activité, pour la catégorie de référence du premier modèle, le logit η de connaître le chômage s'élève à -3.5.

Champ : voir texte.

Source : enquête *Hdv*.

Différents effets d'interaction ont été testés pour expliquer ce phénomène mais peu se sont révélés significatifs. L'effet de l'ancienneté ne semble ainsi pas varier en fonction de la cohorte, de la période, de la situation antérieure ou de la profession. En plus de l'interaction avec la reprise d'études mentionnée précédemment, l'effet de l'ancienneté varie cependant en fonction de la mobilité professionnelle. En soi, le fait d'avoir changé de profession réduit les chances de chômage. Par contre, l'ancienneté du changement ne les fait pas varier. L'effet de l'ancienneté de la transition vers une autre profession varie néanmoins avec le point de la trajectoire où elle est intervenue : plus elle intervient précocement dans la carrière et plus elle diminue la probabilité de connaître le chômage. Ces effets ne sont pas constants dans le temps historique et varient selon les périodes. L'ancienneté fait ainsi diminuer la probabilité de transition vers le chômage à un rythme moins rapide après 1973. Par contre, l'effet protecteur de la précocité de la transition s'accroît après cette date. Le clivage avant et après 1987 est strictement inverse. L'ancienneté de la transition professionnelle de cette transition réduit les chances de transition vers le chômage alors que la précocité augmente la probabilité de chômage. Il apparaît donc que la période d'augmentation du chômage (après 1973) favorise les transitions plus précoces alors que la période de diminution (après 1987) favorise, elle, l'ancienneté dans la profession proprement dite. De plus, l'ancienneté dans la profession interagit avec la situation passée où elle tempère les effets d'avoir été chômeur ou d'avoir interrompu son activité l'année précédente. Au regard de ces résultats, on peut formuler l'interprétation que l'effet protecteur de l'avancée dans la trajectoire professionnelle n'est pas le fait du nombre d'année d'activité en soi, mais plutôt l'ancienneté sur un segment spécifique des marchés d'emploi. Ces tendances pourraient être interprétées comme l'effet de l'accumulation de capitaux sociaux spécifiques à ce segment, qu'il s'agisse de savoir-faire, de compréhension du fonctionnement du segment ou de réseaux sociaux. Les variations correspondant aux périodes se laissent, elles, moins facilement interpréter et devront faire l'objet de plus amples analyses.

L'introduction d'effets d'interactions entre la mobilité géographique et la période fait apparaître une rupture entre avant et après 1973 pour les immigrés d'origine africaine. Avant cette date, leur probabilité de connaître le chômage en France ne se distingue pas de celles des ressortissants d'origine française (ils ont même des probabilités moindres mais non significatives). Par contre, après 1973, leur probabilité de connaître le chômage augmente beaucoup plus rapidement. Ceci pourrait s'expliquer par le fait que le recours à l'immigration (principalement nord-africaine) fut organisé par le gouvernement français pour pallier aux déficits de main-d'œuvre dans certains secteurs d'activité. Dans ce contexte, les migrants étaient quasiment assurés de trouver un emploi à leur arrivée. La crise commencée en 1973 a par la suite conduit à des restructurations dans des secteurs où la main-d'œuvre immigrée était surreprésentée. À cela s'ajoute sans doute des phénomènes de discrimination.

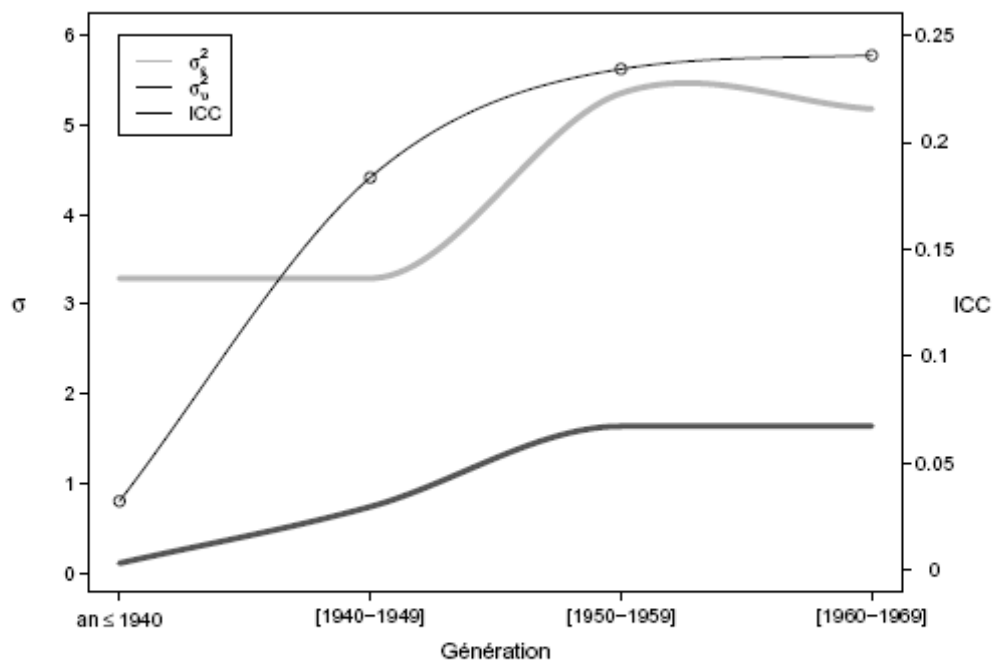
Enfin, être en couple (sans différencier les statuts conjugaux) diminue la probabilité d'être au chômage et cet effet ne varie ni en fonction du temps ni en fonction des cohortes.

Hétérogénéité intra et interindividuelle

Afin de procéder à un premier essai, les variances intra et interindividuelles ont été modélisées en fonction de la cohorte⁷. Il ressort des analyses que l'année de naissance contribue aux deux sources de dispersion. Sur la période observée, chaque nouvelle génération voit sa variance intra et interindividuelle augmenter (cf. figure 4). D'une part, en prenant pour référence les individus nés entre 1940 et 1949, le surcroît de la dispersion est quasiment nul pour la première cohorte (≤ 1940). Elle est à peu près identique pour les deux cohortes suivantes ([1950-1959] et [1960-1969]). D'autre part, l'ICC s'élève à 3% pour les individus nés avant 1940. Il monte à 18% pour la cohorte suivante et se stabilise à 24% pour les deux dernières. La trajectoire sociale semble donc perdre de l'importance au fil des générations dans l'explication des transitions vers le chômage. Il est intéressant de noter que les variances inter et intra-individuelle jouent peu pour la cohorte n'ayant pas été active après 1973 dans le sous-échantillon analysé alors qu'elles augmentent pour la cohorte qui a connu le choc pétrolier et ses conséquences après avoir entamé sa trajectoire et, qu'enfin, l'ICC se stabilise pour celle ayant intégralement travaillé après cette date. Avant de proposer une interprétation de ces résultats, il est important de souligner que ces proportions sont relatives au modèle estimé. L'analyse est en effet implicitement fondée sur le principe du tiers exclu. Autrement dit, l'introduction d'autres variables pourrait conduire à réduire la variance interindividuelle. L'évolution de l'ICC semble néanmoins étayer l'hypothèse que le processus de singularisation des trajectoires est un phénomène historique. Le sens de l'évolution vers plus de variation après 1973 pourrait trouver sa source dans la dégradation du rapport de force entre salariés et employeurs dans un contexte d'augmentation du chômage. Le niveau atteint par les dernières cohortes pourrait lui s'expliquer par les transformations opérées dans la régulation des marchés d'emploi et de l'organisation du travail à partir des années 1980.

⁷ Les résultats présentés n'ont donc pas encore été stabilisés. Ils ne sont donnés qu'à titre d'illustration.

FIGURE 4 – Variance intra/inter-individuelle et ICC pour chaque génération



Lecture : Pour les individus jusqu'en 1940, la variance intra-individuelle σ_c^2 est égale à 3.2, la variance inter-individuelle σ_u^2 est égale à .1 et l'ICC est égale à 3%.

Champ : voir texte.

Source : enquête Hdv.

5. Conclusion

L'analyse de la probabilité de connaître le chômage sur une longue durée permet de faire ressortir l'effet de la rémanence de la trajectoire sociale ainsi que de la segmentation des marchés d'emploi pour les trajectoires masculines. Le suivi sur la totalité de la seconde partie du XX^e siècle souligne de plus l'incidence du contexte sur les trajectoires individuelles. L'année 1973 marque ainsi une rupture à partir de laquelle les caractéristiques réduisant la probabilité de ne pas être en emploi telles que l'ancienneté dans la profession perdent de leur efficacité alors que les facteurs augmentant cette probabilité (comme avoir été au chômage l'année précédente) deviennent plus prégnants. La période fait aussi apparaître de nouveaux risques (par exemple pour les migrants provenant du continent africain). Il apparaît de surcroît que l'effet des transformations du contexte dans lequel se déroule les trajectoires professionnelles ne se résume pas à un effet de période et n'affectent pas uniformément les salariés selon leur génération. Les devenirs des individus ayant tracé les dix premières années de leur trajectoire pour partie ou intégralement après 1973 semblent ainsi manifester une hétérogénéité croissante au fil des années. Il ne s'agit cependant que de résultats provisoires. Dans le cadre des modèles linéaires généralisés mixtes, la qualité de la mesure de la variance interindividuelle dépend fortement de la spécification des effets fixes. Le modèle gagnerait probablement à introduire d'autres éléments de la trajectoire sociale et cette reformulation pourrait affecter la variance mesurée. De plus, la suite des analyses devra prendre en compte les différents processus de sélection qui conduisent à participer ou non aux marchés d'emploi et à s'orienter vers un segment ou un autre (public, privé ou indépendant).

Bibliographie

- Blossfeld H.P. et Rohwer G. (2002), *Techniques of event history modeling. New approaches to causal analysis*, Lawrence Erlbaum, Hillsdale.
- Boeck P. (de) et Wilson M. (2004), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, Springer, Berlin.
- Bourdieu P. (1974), « Avenir de classe et causalité du probable », *Revue française de sociologie*, XV(1).
- Bourdieu P. (1979), *La Distinction. Critique sociale du jugement*, Minuit, Paris.
- Courgeau D. et Lelièvre E. (1989), *Analyse démographique des biographies*, Éditions de l'Ined, Paris.
- Elias N. (1969), *La dynamique de l'Occident*, Minuit, Paris.
- Goldstein H. (2003), *Multilevel Statistical Models*, Hodder Arnold, London.
- Hardin J. et Hilbe J. (2007), *Generalized Linear Models and Extensions*, Stata Press, College Station.
- Hedeker D. et Gibbons R.D. (2006), *Longitudinal Data Analysis*, Wiley-Interscience, London.
- McCullagh P. et Nelder J. (1989), *Generalized Linear Models*, Chapman and Hall, London.
- McCulloch C.E., Searle S.R., Neuhaus J.M. (2006), *Generalized, Linear, and Mixed Models*, Wiley, London.
- Molenberghs G. et Verbeke G. (2006), *Models for Discrete Longitudinal Data*, Springer, Berlin.
- Pinheiro J.C. et Bates D.M. (1995), « Approximations to the Loglikelihood Function in the Nonlinear Mixed Effects Model », *Journal of Computational and Graphical Statistics*, IV(1).
- Singer J.D. et Willett J.B. (2003), *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, New York.
- Skrondal A. et Rabe-Hesketh S. (2004), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Chapman & Hall/CRC, Boca Raton.
- Snijders T.A. et Bosker R. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage, Thousand Oaks.
- Soubiran T. (2006), « La jeunesse dans tous ses états », in O. Samuel (resp. sc.), *Les catégories familiales et socioprofessionnelles à l'épreuve des identités et des trajectoires biographiques*, rapport à la Mire-Drees, p.73-117.
- Yang Y., Fu W.J. & Land K.C. (2004), « A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models », *Sociological Methodology*, XXXIV.