

# LES DÉTERMINANTS DE LA DURÉE PASSÉE AU CHÔMAGE

May 2021

Cécile BRISSARD, *Directed by* :  
Baptiste GOUMAIN, Philippe RAPOPORT et Eric DEFEBVRE  
Pauline PEYDIÈRE *Programming in* : SAS, Python

Université Paris 1 Panthéon-Sorbonne, France

## Résumé

Au travers d'une approche micro-économétrique, nous avons travaillé à l'élaboration d'une étude sur les déterminants de la durée du chômage en France. Après avoir étayé la littérature sur le sujet, et traité les données issues de l'enquête française Santé et Itinéraire Professionnel 2006, nous utilisons les outils statistiques à notre disposition et vues en cours de Microéconométrie des modèles qualitatifs. L'objectif principal de cet article est donc de mettre au service de la problématique du chômage la théorie des modèles de durée. Les principaux déterminants de la durée du chômage sont, d'après nos résultats, le sexe, l'âge, le niveau de diplôme, la santé physique et mentale de l'individu, ainsi que les activités pratiquées en dehors de l'emploi. La discussion sur la pertinence de notre modélisation et des variables, nous amène enfin sur une ouverture à d'autres déterminants non identifiés ici.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Littérature</b>	<b>4</b>
<b>3</b>	<b>Analyse de la base de données</b>	<b>5</b>
3.1	Échantillon et choix des variables . . . . .	5
3.2	Statistiques descriptives . . . . .	7
<b>4</b>	<b>Modélisation</b>	<b>8</b>
4.1	Méthode . . . . .	8
4.2	Résultats . . . . .	10
<b>5</b>	<b>Discussion</b>	<b>14</b>
5.1	Problématiques de notre approche . . . . .	14
5.2	Ouvertures et améliorations possibles . . . . .	14
<b>6</b>	<b>Annexes</b>	<b>16</b>

# 1 Introduction

Le chômage est défini par l'INSEE comme "l'ensemble des personnes de 15 ans et plus, privées d'emploi et en recherchant un." A ce titre, il ne constitue pas une représentativité de la population inactive mais bien de ceux qui, pourtant en âge de travailler, ne se voit pas offrir l'emploi qu'ils convoitent par la société. Ainsi depuis près de 50 ans, les gouvernements agissent à la compression de cette part de la population active. C'est principalement dans cette optique que de nombreux rapports et études se penchent sur la question de la durée du chômage : afin de comprendre les causes du chômage de longue durée et de l'endiguer.

Dans ce contexte, nous tenterons d'apporter, par l'étude statistique de la base de donnée SIP06, une réponse aux questions suivantes : La durée passée au chômage est-elle issue de déterminants individuels propres aux chercheurs d'emploi, qu'ils soit de nature socio-économiques ou démographiques ? Qu'en est-il du rôle des enjeux de santé économique et des politiques menées par l'Etat dans le sens d'une réinsertion professionnelle ?

La densité de la littérature sur le sujet, notamment dans le cadre économétrique, sera dans un premier temps passée en revue en section 2. La description de la base de donnée et de l'échantillon obtenu suite au traitement de celle-ci seront exposée dans la Section 3. Et, c'est dans cette même section que nous présenterons les premiers résultats des statistiques descriptives basiques essentielles à la bonne connaissance des variables explicatives qui s'offrent à nous. Par la suite, la Section 4 sera destinée à l'explication des choix de modélisation et au compte rendu des résultats en découlant. Résultats et études des variables à partir desquels, nous établissons en Section 5 une discussion sur les problématiques rencontrées et les améliorations possibles de nos modèles. Une ouverture où les déterminants non mentionnés jusqu'à lors seront également suggérés comme causes possibles mais manquantes à notre étude. Enfin, nous concluons en Section 6, ce projet axé sur la durée du chômage.

## 2 Littérature

Avant de construire notre étude sur la durée du chômage, il est nécessaire d'établir une revue de la littérature disponible sur le sujet à ce jour<sup>1</sup>. Quelques auteurs, comme Denis FOUGERE et Thierry TESTE, présentent plusieurs publications sur cette étude de la durée du chômage.

Tout d'abord, il ressort de la majorité des études que la durée du chômage est une variable généralement censurée à droite et à gauche : en effet, disposant de la durée en mois ou en année passée au chômage, nous ne connaissons pas nécessairement la date d'entrée des individus. Aussi, concernant la date de sortie, celle-ci n'est parfois par renseignée et on ne connaît alors pas le nombre d'années ou de mois passés au chômage mais l'on sait uniquement que le chômeur est toujours référencé comme tel auprès de Pôle emploi. Cette particularité que l'on retrouve dans les études présentées ici, implique l'utilisation des fonctions de survie, et de hasard avec par exemple l'utilisation du modèle de Weibull, ou de modèles paramétrique ou semi-paramétrique comme les modèles de Cox.

Chez Teste[3], ces modèles permettent la correction des biais de sélection dus à l'endogénéité des durées de chômage. L'auteur y choisit des variables binaires afin d'éviter une colinéarité entre celles-ci et la variable expliquée : ainsi elles sont toutes de nature qualitative et regroupées en modalités excepté le taux de chômage. La principale solution retenue dans l'article, pour une modélisation optimale de la probabilité de sortie du chômage selon les causes étudiées, est la maximisation d'une vraisemblance conditionnelle à la date d'entrée au chômage. Cette méthode permet de corriger automatiquement les biais selon l'ancienneté au chômage : dans le cas où il y a censure à gauche i.e. où l'on peut supposer que la date d'entrée au chômage est antérieure à la date du début de l'étude.

L'utilisation de ces modèles de durée et des log-vraisemblances pour déterminer le modèle le plus adapté aux données, est retrouvée dans l'étude de Fougère et Bonnal[2].

Enfin, dans l'étude publiée par l'ADRES[1] portant très précisément sur des durées de chômage censurées à droite, est développé un estimateur basé sur la technique des  $k$  voisins : l'estimation s'appuie sur les voisins non censurés les plus proches de la date  $t$  et qui sont pondérés par leurs poids. Ce dernier est déterminé à l'aide de l'estimateur de Kaplan-Meier. Le but de ce papier est donc une validation empirique et une comparaison des différents estimateurs disponibles avec celui qu'introduisent les auteurs. Ils développent donc la méthode du noyau qui consiste en une méthode de lissage, celle des  $k$  plus proches voisins mais aussi celle des histogrammes à pas aléatoires. L'objectif est de minimiser le Mean Integrated Square Error ainsi estimé à partir de données simulées puis de l'appliquer à une cohorte issue de données réelles. Cette dernière étape leur permet de produire une analyse de la probabilité de trouver un emploi selon le *Sexe* et également de rendre compte du *paradoxe du Mover – Stayer* qui apparaît lorsque certains facteurs ne sont pas également observable au sein de la population, il y a donc une hétérogénéité que le statisticien ne peut pas observer.

Une caractéristique des données de chômage également traitée par D. Fougère où il explique que le statisticien ne connaît pas certaines caractéristiques détectables par les employeurs comme la motivation ; ainsi les individus qui vont rester stables (i.e. chômeurs) sont indifférenciés des mobiles (i.e. ceux qui vont trouver un emploi). Dans ce cas précis, les déterminants sont donc non identifiés car il n'y a pas foncièrement de raison expliquant qu'un individu sorte du chômage tandis que l'autre soit un chômeur de longue durée.

Plus généralement, au niveau des déterminants de la durée du chômage étudiés dans les différents articles précités, on retrouve des facteurs dits "démographiques" :

- |                                            |                                           |
|--------------------------------------------|-------------------------------------------|
| > sexe,                                    | > état de santé,                          |
| > âge,                                     | > la situation familiale et matrimoniale, |
| > nationalité ou statut d'immigration,     | > le nombre d'enfants ;                   |
| > qualification ou niveau d'étude atteint, |                                           |

---

1. cf. Références

Et des facteurs socio-économique :

- |                                                                                                     |                                                                   |
|-----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| > catégorie socio-professionnelle                                                                   | > accès aux aides à l'insertion ou le nombre de mise en relation, |
| > revenus mensuels perçus,                                                                          |                                                                   |
| > lieu d'habitation,                                                                                | > durée et montant des droits à l'indemnisation,                  |
| > ancienneté au chômage et raison de l'entrée au chômage,                                           | > connaissances acquises indépendamment du niveau de diplôme,     |
| > caractéristique des périodes d'emploi (durée, type de contrats, périodes de chômage antérieures), | > emploi précédemment occupé et type d'emploi recherché.          |
| > accès à la mobilité,                                                                              |                                                                   |

### 3 Analyse de la base de données

La base de données sur laquelle nous avons travaillé est issue de l'enquête française Santé et Itinéraire Professionnel de 2006. Cette étude est menée sur un ensemble d'individus interrogés deux fois à 4 années d'intervalle (en 2006 puis en 2010) par la DREES et la DARES, et a pour objectif de construire un échantillon représentatif de la population concernant le travail, l'emploi et la construction ou l'altération de la santé. Cette base contient les réponses de près de 14 000 individus à un questionnaire traitant de la santé mentale, la santé physique, l'environnement familial, la carrière, l'enfance et des activités sociales.

A partir de la base de données initiale, nous avons hérité d'un échantillon d'individus ayant entre 19 et 74 ans. La base initiale comprend 13 648 individus parmi lesquels 7 453 femmes et 6 195 hommes (soit 54.61% et 45.39% respectivement). Après traitement des valeurs manquantes et ré-encodage de certaines variables sur le logiciel Python, nous disposons de 65 variables.

#### 3.1 Échantillon et choix des variables

En premier lieu, il est important de noter que la population d'intérêt étant uniquement les individus au chômage ou ceux dont la durée passée au chômage a été d'au moins 1 ans au cours de leur vie professionnelle, toutes les autres observations seront omises de la base sur laquelle nous modélisons.

Afin d'obtenir la base *SIP06\_Cleanst* les valeurs manquantes sont recodées en 0, les chaînes de caractères associées à des modalités ou ajoutées sous forme de dummies. Par ailleurs, parmi les 65 variables ainsi recensées, nous nous devons d'effectuer un choix des plus pertinentes pour la modélisation. Ainsi, présentons les variables :

La **variable d'intérêt** *duree\_chomage* représente le nombre d'année passée au chômage pour un individu. Ces durées vont de 0 à 27 ans : la variable est donc discontinue, ce à quoi nous ferons attention lors de la modélisation en utilisant l'approximation d'Efron. Les données sont censurées à gauche puisque la date de début de l'étude est nécessairement postérieure à la date de début de chômage de certains individus. Cependant, ne disposant pas des dates de début, ni de fin de chômage, mais uniquement de la durée, nous ne pourrions pas gérer cette caractéristique de nos données. Nous nous concentrons donc sur la gestion de la censure à droite qui est matérialisée ici par les 9999 : une valeur qui concerne les individus encore au chômage et pour lesquels on ne peut donc pas rendre compte de la durée.

#### Les variables démographiques :

- *ident\_ind* représente l'identifiant de chaque individu, cette variable étant inutile à la modélisation, sera retirées de la table finale.
- *sexenq* donne le sexe de l'individu, après encodage elle prend la valeur 0 pour masculin et la valeur 1 pour féminin.
- *agenq* donne l'âge de l'individu. L'estimation des ratios de hasard par le modèle de Cox<sup>2</sup> nécessitant des modalités à comparables entre elles, cette variable est répartie en 5 classes avec la procédure SAS *PROC RANK* qui permet d'obtenir une répartition équilibrée des classes. Absente de la table finale, on retrouve cette variable sous le nom de *classage*.

---

2. Défini dans la section 4.1. Méthode

- *flnais*, *fnivdip*, *fsitua* constituant la nationalité de l'individu, le niveau de diplôme et la situation vis-à-vis de l'emploi respectivement, sont gardées conforme à leur définition dans le dictionnaire des variables.
- *fnelev* donnant le nombre d'enfants à charge par individu est encodé : le chiffre 4 représente la modalité d'avoir 4 enfants et plus, ce qui permet d'en avoir une répartition plus équilibrée.
- Les variables de santé :
  - > *duree\_handi* indique la durée passée en situation de handicap en année. Tout comme notre variable d'intérêt, elle est donc discontinue, et afin d'en obtenir des ratios de hasard comparables, nous la séparons en 4 modalités. *Duree\_handi* prend la valeur 0 pour la durée correspondante, 1 pour une durée positive et inférieure à 10 ans, 2 pour une durée supérieure à 10 ans (et pouvant donc aller jusqu'à 57ans) et enfin, 3 pour la modalité de censure : lorsque la valeur est 9999 et que la période de handicap n'est pas terminée.
  - > *sq1g*; *sq2g*; *sq3g*, respectivement, l'état de santé, la présence ou l'absence de maladie chronique, et d'un problème de santé depuis 6mois au moins limitant les activités quotidiennes, sont : pour la première, supprimée car c'est une variables subjectives, à l'appréciation de l'individu ; pour la deuxième, conservée et recodée en 1 pour "OUI", en 0 pour "NON" ; pour la troisième, supprimée car le lien reste ambigu et l'information ajoutée faible, on peut supposer des arrêts maladies ou l'évaluation d'un taux d'invalidité correspondant donc à un handicap (information déjà contenue dans *duree\_handi*).
  - > *maladie\_chronique*, donne le type de maladie qui touche l'individu porteur. La variable étant en chaîne de caractère, une dummy prenant la valeur 1 lorsque le sujet est porteur, 0 sinon, a été créé pour chaque type de maladie. Malheureusement, la faible représentativité, et le fait que l'information contenue dans *sq2g* soit suffisante, nous conduit à ne pas prendre en compte ces dummies.
  - > enfin, après recherche, *EDM* soit épisodes dépressifs majeurs et *TAG* soit troubles anxieux généralisés, sont des indicateurs de santé mentale issus du Mini International Neuropsychiatric Interview (MINI), s'appuyant sur le Diagnostic and Statistical manual of Mental disorders (DSM-IV). Elles ont été recodées en 1 pour "OUI", 0 pour "NON".

### Les variables socio-économiques :

- Le dernier revenu mensuel perçu - *zremen*, est comme vu précédemment avec la variable âge, répartie en classes à partir de la table finale (qui ne prend en compte que la population d'intérêt). Elle ne sera donc pas utilisée pour la modélisation directement mais sous la forme de *classrev*, son équivalent en 3 modalités obtenues avec *PROC RANK*.
- Les variables traitant des conditions de travail, informant sur la carrière des individus et sur la capacité à concilier travail et vie privée :
  - > *elip* est le pourcentage des périodes d'emplois longs dans l'itinéraire professionnel. C'est une variable normalement corrélée négativement à la durée du chômage ; elle peut en être un déterminant. En effet, un individu avec des temps d'emplois courts (plusieurs petits postes ou emplois précaires de type CDD ou intérim) a plus de chance de se retrouver au chômage et d'avoir du mal à en sortir vers un emploi long. A l'inverse et dans une moindre mesure, une seule période d'emploi long suivie d'un licenciement peut engendrer un chômage de longue durée : les recruteurs auront tendance à estimer que l'individu, par manque d'expériences diverses, peut avoir des difficultés à s'adapter/être trop spécialisé. Cette variable est continue et ne peut donc être utilisée dans l'interprétation des ratios de hasards de Cox ; ce pour quoi elle est remplacée par *classelip*, une variable comportant 5 classes, avec *PROC RANK*.
  - > *eaacol* ; *eaefam* ; *eaethi* ; *eaepur* ; *eahum* sont des variables qui portent sur le vécu du travailleur lors de ses périodes d'emploi, tant sur son ressenti éthique, émotionnel que sur les relations induites avec ses collègues ou sa famille. Très intéressantes mais subjectives (à l'appréciation propre de l'interrogé) et trop peu étroitement liées à notre variable expliquée, nous ne les avons pas sélectionnées.
  - > *ealati* ; *eaenv* ; *eaabri* ; *ealour* ; *eaexi* ; *eaerre* ; *eadepl* ; *eahsem* sont des variables qui portent sur les conditions de travail donc peu pertinentes pour expliquer la durée du chômage. Elles peuvent cependant expliquer le développement de maladies chroniques ou la survenance d'accidents du travail amenant un handicap ; mais disposant d'ores et déjà de telles variables et devant l'ambiguïté du lien avec la durée de chômage, nous ne les prendrons pas en compte.

- Les variables des conditions de vie durant l'enfance :
  - > *fpermer* permet de connaître par qui l'individu a été élevé au cours de son enfance, nous l'avons donc conservée telle que décrite dans le dictionnaire des variables.
  - > *fnaim* ; *fnaip*, le lieu où l'enfant a été élevé seul par sa mère ou son père, respectivement. Nous n'avons souhaiter garder que la nationalité de la mère car le fait d'être élevé par son père seul représente peu d'individus, contre 10% pour "mère seule"<sup>2</sup>. De plus, on compte 20% d'"à l'étranger" dans *fnaim*<sup>2</sup>, ce qui permet d'opposer les deux modalités pour rendre compte de leur impact sur le chômage.
  - > *fgue* ; *fviopr* ; *fepr* ; *fdrog* ; *fcofam* ; *fsep* ; *fdec* ; *fsafa* ; *fhand* représente un ensemble de conditions dans lesquelles l'individu a été élevé. Malgré les récents travaux sur le chômage qui montre sa dépendance aux variables psychologiques dues à des événements passés, ou le reflet de difficultés de développement, leur impact nous apparaît trop lointain. Aussi devant le nombre de variables disponibles, elles ne nous paraissent pas comme être les plus significatives : nous les excluons de la modélisation.
- Les variables portant sur les activités externes à l'emploi :
  - > *zart\_1*, *zben\_1*, *zspo\_1*, *zsync\_1*, *zpol\_1*, *zreli\_1* indiquent respectivement si l'individu pratique ou non une activité artistique, bénévole, sportive, syndicale, s'il est engagé politiquement ou religieusement. Agissant comme des lieux de socialisation, ces activités peuvent rendre compte d'une non-nécessité de retour à l'emploi, ou bien d'un réseau qui permettra au contraire une mise en relation plus rapide ou de compétences acquises/développées autrement que par la formation favorisant un retour vers l'emploi.

Une fois les variables sélectionnées dans une table finale ; nous choisissons de créer deux dummies *chomeur* et *chomeur\_actu*, qui prennent la valeur 1, respectivement, quand la durée du chômage est supérieure à 0 et quand *fsitua* prend pour valeur la modalité 4 ; et 0 dans les cas contraire. Ces deux nouvelles variables nous permettent de ne sélectionner que la population ayant été affectée au moins une fois par le chômage. Une fois notre population de chômeur ainsi isolées, nous élaborons la variable de censure (notée *censure*), afin de traiter le problème de censure à droite de la variable expliquée. Les données sont supposées censurées lorsque l'on ne connaît pas la durée du chômage, c'est donc le cas des 9999. Nous définissons *censure* = 0 pour *duree\_chomage* > 27<sup>3</sup> et *censure* = 1 sinon. Cette variable nous sera utile pour la modélisation, comme nous le verrons dans la section suivante.

## 3.2 Statistiques descriptives

Notre base maintenant complète et ciblée sur la population d'intérêt, nous allons formuler quelques statistiques descriptives sur celle-ci. *SIP* est composée de 27 variables et 2 841 observations, dont 39% d'individus de sexe masculin contre 61% de sexe féminin. Il en ressort, d'après la population initiale citée plus haut, que la répartition des chômeurs est largement en faveur des femmes, même si elles étaient déjà plus nombreuses dans la base *SIP06\_Clean*. Près de 22% des individus de la distribution ont moins de 34ans et 29.83% ont plus de 55ans ; de plus, si on se base sur l'âge moyen de départ en retraite (situé approximativement à 63ans), près de 10% de la distribution est plus susceptible d'appartenir à la catégorie "retraités" qu'à celle des chômeurs. En effet, *fsitua* nous permet d'affirmer une part de 14.43% de retraités parmi ceux ayant déjà rencontrés une période de chômage. A ce sujet, les deux variables *chomeur* et *chomeur\_actu* créées nous permettent de connaître la part des chômeurs de moins de 1 ans. En effet, lorsque *chomeur* = 0 mais qu'il n'a pas été supprimé de la table c'est que l'individu a indiqué être chômage (*fsitua* = 4). On peut donc en déduire que 12.32% des chômeurs sont dits "de courte durée". Concernant, cette population qui s'est déclarée au chômage, elle représente 32.42% du total. Pour finir, la dernière caractéristique directement liée à la variable expliquée est *censure* : elle représente le nombre de personnes étant actuellement au chômage sans qu'on sache depuis combien de temps, soit 16.61% des observations. Il n'est pas étonnant que la part des chômeurs s'étant définis comme tel soit supérieure à la part de censure puisque la modalité *fsitua* = 4 ne comprend pas que les chômeurs inscrits à l'ANPE. Par ailleurs, 42.77% déclarent avoir un emploi au moment de l'enquête, contre 59.4% dans la cohorte de *SIP06\_Clean*.

Au niveau du revenu, près de la moitié de la population a déclaré ne pas avoir eu de revenu le mois précédent l'enquête, avec 48.72% de 0. Et, concernant le pourcentage d'emploi long au cours de la carrière *elip*, 28.82% de la distribution n'ont pas eu accès à ce type d'emploi puisque leur ratio est de 0 quand, par comparaison, environ 20% de nos chômeurs ont un ratio d'au moins 74/100. De même, quand la population générale affiche un taux de près de 23% de poursuite d'étude après le Baccalauréat, ce n'est le cas que pour 17.88% de notre échantillon, où de même

2. Donnée issues de *SIP06\_Cleanst* donc toutes populations confondues

3. La durée la plus élevée de la distribution soit *censure* = 0 pour *duree\_chomage* = 9999

la part des titulaires d'un CAP ou BEP est plus importante de 4 points de pourcentage que dans la population globale. Ce qui nous permet de conclure que le niveau de diplôme est moindre parmi les chômeurs.

Dans la description de la cohorte de chômeur, l'âge, le sexe, le niveau de diplôme, l'accès à l'emploi de longue durée donnent jusqu'à maintenant des résultats peut surprenant même si les différences avec le reste de la population ne sont pas conséquentes. Qu'en est-il de la nationalité? La population d'étranger est légèrement plus importante dans notre cohorte, soit 8.17%. Lorsque l'enfant a été élevé uniquement par sa mère - ce qui représente 10% des interrogés au sein des deux bases de données - avec plus de 20% de celles-ci, ils ont vécu ailleurs qu'en France.

En revanche, à propos de vie familiale, un plus grand nombre d'individus ont un ou deux enfants pour les deux populations, même si la répartition est différente avec davantage de 2 enfants pour la population globale (32 contre 28 pour 1000) et de 1 enfant au sein des chômeurs (19% versus 17%). Un grand nombre d'enfants ne semble donc pas être lié avec une difficulté à concilier vies familiale et professionnelle engendrant la survenue d'une période de chômage. Que peut-on alors dire sur les autres formes d'activité externes mentionnées dans la base? La part des bénévoles est plus grande pour la population globale, de 4 points de pourcentage, que chez les chômeurs; de même qu'ils sont moins nombreux à participer à des activités syndicales - 2% et 3% respectivement. L'activité la plus pratiquée est ainsi le sport pour les deux cohortes et est exercée par 33.5% de ceux étant ou ayant été au chômage, suivi de la participation aux activités bénévoles (21%) et de l'engagement dans des activités artistiques (16.02%).

Les événements de santé comme le fait d'avoir une maladie chronique, vérifie également les hypothèses, 37.1% ont indiqué être souffrant contre 34% lorsque les non-chômeurs sont compris. Pour le reste, 86.2% des répondants ne sont pas en situation de handicap ou le sont depuis moins d'un an. De plus, près de 4% des handicapés semblent en être sortie puisque nous disposons de la durée du handicap. Ce qui laisse une proportion de porteurs d'un handicap de 9.82/100. Pour la prise en compte des événements de santé mentale, les épisodes dépressifs majeurs *EDM* surviennent chez 11.58% des chômeurs, et, les troubles anxieux généralisés *TAG* se manifestent chez 9.15%.

Le contenu de la base de donnée étant désormais précisé, et nos chômeurs quelque peu dépeints avec lui, nous nous appuyons sur la littérature et ce que nous savons des variables pour définir la méthode de modélisation adéquate.

## 4 Modélisation

Avant de débiter la modélisation, permettant d'estimer les déterminants de la durée du chômage, rappelons à ce propos la branche statistique correspondante. Nous nous situons dans l'analyse de survie, aussi appelée analyse de la durée en économie, qui rassemble les modélisations du temps restant avant la mort ou avant un événement. Ici, ce qu'on veut obtenir c'est la probabilité de sortie du chômage donc le temps restant avant de trouver un nouvel emploi. C'est également pour cette raison que nous avons extrait de notre table de base, la seule population à risque concernée : ceux qui ont connu au moins une période de chômage au cours de leur vie.

### 4.1 Méthode

La première étape de l'analyse de survie consiste en la représentation des données sous la forme graphique d'une courbe de survie. Ceci implique donc la modélisation du facteur temps dans la probabilité d'occurrence des événements. Si ces événements d'intérêt ne se sont pas produits (i.e. si on n'observe pas une sortie du chômage) avant la fin de la période d'observation, on dira qu'il y a une censure de la série de données. La fonction de survie, que l'on note  $S$ , donne la probabilité que l'individu ne sorte pas du chômage avant la durée  $t$ . Elle est définie par :

$$S(t) = \mathbb{P}(T > t) \quad (1)$$

où le moment de perte du statut de chômeur est exprimée par  $T$ , et la fonction de probabilité par  $\mathbb{P}$ .

Afin de modéliser la courbe de survie de nos données censurées à droite, nous utilisons l'estimateur de *Kaplan - Meier*. Cette estimation du maximum de vraisemblance non-paramétrique de  $S(t)$ , est donné par le produit suivant :

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \quad (2)$$

Le nombre de survivants juste avant le temps  $t_i$  est donné par  $n_i$ , lorsqu'il n'y a pas de censure. Puisque dans notre cas il y a censure,  $d_i$  représente le nombre de perte à chaque avancée dans le temps. Il en découle que l'estimateur



exposé ci-dessus permet la formation d'une courbe de *Kaplan – Meier* qui commence à 100% lorsque tous les individus sont présents et considérés comme "à risque", allant jusqu'à 0% s'ils sortent tous du statut de chômeur. Dans le cas de non-censure, les courbes de survie et de *Kaplan – Meier* sont équivalentes.

Une fois la courbe de survie obtenue sur SAS avec la *PROC LIFETEST*, nous cherchons à connaître la loi qui modélise au mieux selon la fonction de répartition de nos données. Pour ce faire, nous comparons les quatre lois suivantes :

- > la loi exponentielle ;
- > la loi de *Weibull*, qui est très utilisé dans notre littérature ;
- > la loi log-normale ;
- > et la loi de *Burr*, aussi connu comme loi de Singh-Maddala, ou encore loi log-logistique généralisée.

Les paramètres de chaque loi sont ajustés aux données par régression ou bien par maximum de vraisemblance, c'est ainsi que SAS détermine le modèle correspondant le plus à notre distribution.

On s'attend à ce que la *PROC SEVERITY* permettant de la déterminer, choisisse une loi de *Weibull*. En effet, elle est souvent utilisée, notamment parce qu'elle permet l'obtention de valeurs positives, qui, dans le cadre de modèles de durée, sont les seules à avoir du sens. De plus, elle permet une diversité de profils du taux de défaillance  $\lambda$  et s'approche de la loi normale lorsque le paramètre  $p$  est compris entre 3 et 4. Les profils de survenu de l'évènement "sortir du chômage" sont alors :

- >  $\lambda$  décroissants si  $p < 1$ ,
- >  $\lambda$  constants si  $p = 1$ ,
- >  $\lambda$  croissants si  $p > 1$ ,

Nous verrons qu'il résulte de ce processus de détermination de la loi, que notre modèle suit une loi de *exponentielle*, définie comme une loi paramétrique d'espérance  $\lambda$ , permettant de déterminer la forme de la fonction de hasard. Elle donne la durée de vie d'un système durant sa phase d'exploitation, et implique que même si l'évènement a déjà eu lieu cela n'influence pas le fait qu'il ait pu se produire à nouveau, ni la durée de l'évènement. On dit du phénomène étudié qu'il est "sans mémoire".

Connaissant la loi paramétrique, il ne nous est pas nécessaire de recourir à l'approximation de celle-ci au travers d'un modèle semi-paramétrique. Cependant, notre littérature mentionnant la loi de Cox et celle-ci représentant une bonne approximation de la loi exponentielle, nous chercherons les ratios de hasard via le modèle de Cox.

Cox, classé comme modèle semi-paramétrique à risques proportionnels, est un modèle de régression en temps continu. Son objectif est, par la régression des variables explicatives sur la variable d'intérêt, de modéliser le logarithme du risque instantané. Il nous permet d'approximer les modèles paramétriques, sans hypothèse particulièrement restrictive sur la distribution des risques  $h(t)$ , comme sur celle de la courbe de survie  $S(t)$ . Afin d'obtenir les ratios de hasard, il nous faut, là encore, déterminer une variable de durée : *duree\_chomage*, ainsi qu'une variable de censure qui prend la valeur 0 en cas de données censurées et 1 pour la partie où l'évènement attendu a eu lieu. Rappelons le, lorsque nous disposons du temps passé au chômage pour un individu, c'est que le retour à l'emploi, ou à un autre statut est survenu.

Comme l'ensemble de nos facteurs explicatifs n'évoluent pas dans le temps, nous effectuons simplement une régression linéaire où les sujets censurés ne sont pas considérés directement, grâce à la procédure SAS suivante :

```
PROC PHREG DATA=Micro.SIP;  
CLASS fnivdip(ref="4");  
MODEL duree_chomage*censure(0) = fnivdip / TIES=EFRON ;  
RUN;
```

Si le but reste de trouver le meilleur modèle par rapport aux données observées, cela revient à trouver la combinaison des valeurs des paramètres qui maximise la log-vraisemblance. On parle ici d'une vraisemblance partielle, puisque l'estimation ne concerne que la part de la population subissant l'évènement.

Le problème est que nos données ne sont pas continues et qu'il y a des égalités entre les durées, ce qui nécessite l'un des traitements suivant :

- La méthode exacte, qui considère que l'apparente égalité des durées est due au manque de précision (discretisation) des mesures - ce qui est notre cas, puisque pour deux durée de 3 ans, l'un des chômeurs peut être sortie à 3ans et 6mois, tandis que l'autre seulement à 3ans et 142jours. Cependant ceci implique d'inclure dans le calcul toutes les permutations possibles des égalités, ce qui est long et complexe.
- La méthode de Breslow, qui suppose qu'en cas de durées égales, tous les événements ont la même probabilité d'apparition que le premier événement.
- La méthode d'Efron calcule un risque moyen pour tous les événements survenus dans le même temps. En pratique, lorsque beaucoup d'individus quittent la période de chômage au même moment, la méthode d'Efron est souvent plus performante que celle de Breslow.

Pour suivre l'intuition économique de la théorie du chômage, la méthode Breslow semble être plus adaptée à notre étude.

Enfin, nous précisons la *CLASS* de la variable à comparer aux autres modalités, ce qui nécessite de n'avoir que les variables polytomiques ou dichotomiques. Généralement, la modalité choisie est celle concernant la plus grande part de la population.

La *PROC PHREG* sera appliquée sur le modèle complet, puis sur une série de sous-modèles ne contenant que quelques, voire une, variables afin d'en connaître l'effet ciblé.

## 4.2 Résultats

Suivant la méthode qui vient d'être énoncée, nous obtenons les résultats décrits ci-dessous, suivis des interprétations correspondantes.

### Quelle loi ?

Tout d'abord, comme avancé dans la méthode, les tests d'ajustement que nous avons destinés à la détermination de la loi que suivent les données, permettent la sélection du modèle exponentiel. Ainsi, c'est cette même loi, qui permet d'optimiser les critères de sélection de modèle, comme visible dans le tableau ci-dessous.

Sélection d'un modèle : tests d'ajustement							
Distribution	-2 Log-vraisemblance	AIC	AICC	BIC	KS	AD	CvM
<b><i>Exponentiel</i></b>	<b>1658*</b>	<b>1740*</b>	<b>1741*</b>	<b>1984*</b>	<b>6.57004*</b>	<b>5981*</b>	<b>640.11862*</b>
<i>Weibull</i>	133580	133664	133666	133914	6.57004	19084	640.11862
<i>Log-normal</i>	128576	128660	128661	128910	6.57004	14689	640.11862
<i>Burr</i>	129960	130046	130047	130302	6.57004	46289	640.11862
<i>Note : l'astérisque (*) marque le meilleur modèle d'après le critère de chaque colonne.</i>							

La procédure *SEVERITY* du logiciel SAS note avec une astérisque les meilleurs résultats associés à chacun des critères de sélection : -2 log-vraisemblance, AIC, AICC, BIC... tous sont minimisés avec la loi exponentielle. Ce résultat n'était pas forcément attendu, tant les auteurs que nous avons consultés associent au modèle de durée du chômage la loi de Weibull. Cependant, cette dernière n'est qu'un prolongement de la loi exponentielle et le résultat reste satisfaisant puisqu'il confirme la possibilité d'estimer la log-vraisemblance par un modèle de Cox.

Mais, avant de pouvoir interpréter les ratios de hasard, fournis par la régression de Cox, nous souhaitons observer la durée du chômage, stratifiée ou non, sur diverses fonctions de Kaplan-Meier.

### Fonction de survie

Ainsi, lors de notre première tentative, les données censurées - apparaissant malgré notre modélisation de la censure à droite via la variable dichotomique *censure* - nous nous devons de remplacer les 9999 par des valeurs manquantes (ou . sous SAS) afin que celles-ci n'emportent pas la distribution vers la droite, rendant les graphiques illisibles. Ceci fait, nous obtenons les fonctions de survie appelées FIGURES 1 à 4<sup>4</sup>.

La première figure nous octroie la possibilité d'observer l'évolution du taux de survie de la variable expliquée seule, sans l'impact spécifique des variables explicatives. En 0, 100% de la distribution est au chômage, puis à chaque

4. c.f. Annexes

"marche" des individus meurent, ici cela signifie qu'ils sont sortis du chômage. Ainsi, moins de 20% des individus considérés reste au chômage pendant plus de 5 ans et, plus de 80% ont connu un chômage de longue durée (i.e. supérieur à 1 an). A partir de 10 années de chômage, on visualise un écrasement de la courbe, tant la proportion des observations où la variable *duree\_chomage* prend une valeur supérieure est faible. Le temps de chômage le plus long est de 27 ans, mais ne concerne que très peu de personnes.

Les graphiques suivants, nous permettent d'avoir un aperçu des déterminants de cette durée de chômage. De fait, la stratification par le sexe rend compte de la discrimination à l'emploi des femmes. Quant la durée maximale de chômage masculin est de 21 ans, celle des femmes est de 27 ans. De plus, les hommes quittent ce statut plus rapidement que leurs homologues du sexe opposé, puisqu'au bout de 2 ans près de la moitié des hommes sont sortis du chômage, ce qui n'est le cas que pour 40% de la cohorte féminine.

Nous comparons ensuite le fait d'être sujet à une maladie chronique au fait de ne pas en avoir développé<sup>5</sup>, par son impact sur le chômage. Là encore, le résultat attendu est observé : la durée passée au chômage s'allonge pour un individu porteur et le pourcentage d'individu quittant le chômage est plus élevé à chaque "marche" si ceux-ci ne sont pas malades - par exemple à l'année 3 il ne reste que 20% de chômeur pour lesquels *sq2g* prend la valeur 0 soit "NON" contre 30% de 1 soit "OUI".

Nous souhaitons également stratifier par la nationalité ; cependant, la non significativité du paramètre implique le non-rejet de  $H_0$ . En d'autres termes, les trois strates ont la même fonction de survie. Il en est de même pour la durée de handicap, où la variable *duree\_handi* est découpée en 3 strates qui donnent toutes trois la même fonction de Kaplan-Meier ; ainsi que pour le fait d'être bénévole ou non.

Pour ces trois variables, la stratification ne permet pas de rendre compte de divergence dans la durée du chômage, et elles ne semblent donc pas en être des déterminants. Finalement, le fait de pratiquer ou non une activité sportive, est bien significatif et peut donc être interprété comme un déterminant de la durée passée au chômage. On rejette donc  $H_0$ , et un individu sportif voit sa durée de chômage écourtée par rapport aux individus se déclarant non pratiquants. Nous supposons que le sport permet d'acquérir un cercle social favorable à la mise en relation débouchant sur un emploi.

### Modèle de Cox : Ratios de hasard

Suite à cet ensemble de résultat, nous procédons à l'estimation des ratios de hasard par le modèle de Cox. Dans un premier temps, toutes les variables sélectionnées sont incluses dans le modèle et sans gestion de la discontinuité de la variable d'intérêt. Cette méthode nous donne en résultat le tableau affiché en FIGURE 5, où nous constatons que seul les variables *classrev*, *sexenq* et *classage* sont significatives. Ainsi, pour toutes les autres il n'y a pas d'impact divergents selon les différentes modalités sur la durée de chômage. Cependant, en analysant plus précisément la significativité par modalité, nous obtenons les résultats interprétables suivants :

Analyse des valeurs estimées du maximum de vraisemblance						
Paramètre	Valeur estimée des paramètres	Erreur type	khi-2	Pr >khi-2	Rapport de risque	Libellé
<i>fnivdip</i>	0.20318	0.08327	5.9533	0.0147	<b>1.225</b>	fnivdip 7
<i>fnaim</i>	-0.15591	0.07081	4.8482	0.0277	<b>0.856</b>	fnaim 2
<i>classrev</i>	0.10982	0.05733	3.6689	0.0554	<b>1.116</b>	Rang pour la variable zremen 2
<i>sexenq</i>	0.13656	0.04625	8.7168	0.0032	<b>1.146</b>	sexenq 0
<i>classage</i>	0.65191	0.07149	83.1629	<.0001	<b>1.919</b>	Rang pour la variable agenq 0
<i>classage</i>	-0.33858	0.08040	17.7365	<.0001	<b>0.713</b>	Rang pour la variable agenq 4

Le premier résultat, qui recoupe celui vu précédemment avec les fonctions de survie, concerne la discrimination à l'embauche par le sexe. Avec comme modalité de référence le fait d'être une femme, le rapport de risque lorsque *sexenq* = 0 est de 1.146 : les individus de sexe masculin ont donc 14.6% de chance en plus que leur durée de chômage soit écourtée, que leurs homologues féminines. Ce résultat attendu sera retrouvé à chaque modélisation et s'explique notamment par la position de la femme dans la société actuelle. A diplôme, âge, revenu égaux, elles vont pourtant rester au chômage plus longtemps : un héritage de la place de la femme au foyer qui reste empreint dans la carrière professionnelle féminine, notamment pour des questions d'équilibre entre vie familiale et professionnelle (prise en charge des enfants, congé maternité, absentéisme...).

Sont également significatives deux classes d'âge, celle des jeunes (entre 20 et 36 ans) et des aînés (58 et 74 ans). Ainsi, avec des rapports de risque de 1.92 et 0.71 respectivement, ce résultat souligne l'importance du chômage de longue durée chez les plus de 55 ans. Par opposition, le modèle ne fait pas état d'un chômage des jeunes qui durerait

5. FIGURE 3

plus longtemps que la moyenne (soit 2 ans), puisque ceux-ci ont 91.9% de chance d'en sortir plus rapidement que la classe de référence - i.e. la deuxième classe d'âge comprenant les individus âgés entre 42 et 50 ans. Au niveau de la stratification par âge, le résultat attendu d'une difficulté d'embauche de ceux en fin de carrière, pour une productivité qui s'abaisse, un coût plus élevé ou l'attente d'arrêt maladie plus fréquents par exemple, est illustré par les rapports de risque.

Ces ratios permettent également de rendre compte de l'importance du niveau de diplôme dans la détermination de la durée du chômage. En effet, avec un niveau de diplôme *BAC* + 2 (modalité 7), l'individu a 22.5% de chance supplémentaire de sortir du chômage par rapport à celui qui aurait un *CAP* ou un *BEP* : autrement dit, plus le niveau de diplôme est élevé, moins la période de chômage sera importante, un déterminant qui n'est pas surprenant par rapport aux hypothèses de la littérature.

Enfin, lorsque l'interrogé a vécu à l'étranger uniquement avec sa mère, ses chances de sorties du chômage sont multipliées par 0.856 ; on peut donc penser qu'un individu n'ayant pas grandi en France, une fois au chômage, aura des difficultés à quitter ce statut dans le même temps qu'un individu qui aurait toujours habité le pays. Les recensements ethniques n'étant pas disponible en France, c'est au travers de cette forme de variable que peuvent être observées les discriminations à l'embauche d'ordre ethnique (couleur de peau, origine du nom...). Et, concernant la variable de revenu, par rapport à la modalité 0, qui comprends ceux n'ayant déclaré aucun revenu le mois dernier, la modalité 2 (ceux ayant déclaré un revenu mensuel supérieur à 1000 euros) offre 1.116 fois plus de chance de sortir du chômage plus rapidement. Un résultat difficile à interpréter tant la distribution des revenus de notre base est déséquilibrée avec 48% de 0.

Afin de poursuivre notre analyse, nous effectuons la même régression en utilisant l'approximation d'Efron pour corriger la discontinuité de la variable de durée. La gestion de ce problème, permet une significativité d'un plus grand nombre de variables, visible dans le tableau en Annexe<sup>6</sup>. Ainsi, faire partie de la modalité 1 de la variable *classage* octroie 21.8% de chance en plus d'écourter sa durée de chômage. Aussi, l'activité sportive devient significative, et le fait d'être sportif octroie une probabilité de sortie rapide du chômage supérieure au non sportifs de 9.4%. Nous remarquons également que les variables précédemment interprétées prennent des ratios de hasard amplifiés mais allant toujours dans le sens vu ci-dessus. L'approximation d'Efron permet donc un gain de précision.

Pour conclure cette analyse, et devant la faible significativité de nombreuses variables, nous interprétons les modèles partiels par [groupe de] variable[s] afin de connaître l'effet des déterminants supposés dans la littérature.

Ainsi, en prenant comme référence le fait d'avoir deux enfants, on cherche à savoir l'impact du nombre d'enfants sur la durée passée au chômage. Cependant, la seule modalité significative est celle de ne pas avoir d'enfant. Nous ne pouvons qu'affirmer qu'une personne sans enfant a 33.8% de chance supplémentaire de sortir du chômage par rapport à un individu qui aurait 2 enfants. Le fait ou non d'avoir des enfants apparaît donc bien comme un déterminant de cette durée, qui ne nous permet pourtant pas d'affirmer que plus le nombre d'enfants est grand, plus la durée du chômage est longue.

Par la suite, nous voulons approfondir l'effet du niveau de diplôme, en ne faisant apparaître que cette variable : les modalités 1, 2, 5, 7 et 8, deviennent significatives. Les résultats obtenus confirment qu'à niveau de diplôme inférieur au *BEP*, *CAP* la durée de chômage est allongée, et à niveau supérieur elle est écourtée. Il est, par ailleurs, surprenant que détenir un *BAC* + 2 pourvoit plus de chance de sortir du chômage, que si l'individu a un niveau supérieur à *BAC* + 2. Il y aurait donc un effet de seuil, à partir d'un certain diplôme, que l'on peut supposer dû à des attentes salariales en conséquence trop élevées ou un niveau de spécification ne correspondant pas aux besoins du marché du travail. De même, les agents sans diplôme ont paradoxalement une durée de chômage plus courte que ceux munis d'un certificat d'étude professionnell (*CEP*) : un phénomène difficile à expliquer mais il est probable que l'absence de diplôme soit compensée par l'expérience.

Même isolée, la variable *fpermer* reste non significative ; ainsi l'environnement familiale ne semble pas être un déterminant de la durée du chômage.

Nous rassemblons, par la suite les variables *sexeng* (1.14 pour "masculin"), *sq2g* (1.20 si "NON"), *duree\_handi* et *zspo\_1* (1.20 pour "OUI") dans un modèle, ce qui permet la significativité de la variable de maladie chronique, pour laquelle le ratio de hasard confirme les résultats obtenus avec la fonction de survie<sup>7</sup>. *A contrario*, la mesure de durée d'handicap n'est pas significative ; alors le fait d'avoir été handicapé depuis moins ou plus de 10 ans, ou de l'être encore n'a pas d'impact divergent sur la durée du chômage. Dans la continuité des variables de santé,

6. Analyses des valeurs estimées du maximum de vraisemblance - Efron

7. FIGURE 3

on regroupe ensuite *EDM* et *TAG* dans un sous modèle, dont le résultat admet une différence avec le fait d'être atteint d'épisode dépressifs majeurs. Ainsi, une personne ayant des problèmes de santé mentale à un ratio de 0.85, par rapport à une personne non atteinte : sa durée de chômage est susceptible d'être plus longue que la moyenne. Ceci impacte donc l'emploi autant que les maladies chroniques.

Pour en revenir aux activités permettant de garder le contact social et donc associées, d'après les hypothèses, aux contacts permettant la mise en relation vers un emploi : regroupées dans une régression, seules *zart\_1* et *zspo\_1* sont significatives. Dans les deux cas, le fait d'y participer, permettrait de réduire sa durée de chômage : les ratios de hasard sont de 1.135 et 1.195, respectivement.

Finalement, on souhaite évaluer l'impact de la nationalité, *fnais* ; de l'environnement familial, *fpermer* ; et du lieu de vie avec la mère, *fnaim*, le cas échéant, dans un modèle conjoint. La nationalité n'offre toujours pas de significativité et n'est donc pas considérée comme un déterminant de la durée du chômage. Par rapport à la modalité 1, le fait de vivre avec père et mère ensembles, seules les modalités 5 - par des membres de la famille proche - et 7 - autre cas (foyer, institution...) sont significatives et impliquent une moindre chance de sortie du chômage à hauteur de 0.764 et 0.676 respectivement. Quant à la nationalité de la mère, c'est un déterminant qui agit toujours dans le même sens qu'exposé précédemment (0.87).

Conscients de la particularité qu'engendre le choix d'étudier les variables une à une, nous procédons dans la partie suivante à l'exposé des limites de notre étude.

## 5 Discussion

### 5.1 Problématiques de notre approche

Emettons un regard critique sur les modèles qui ont été présentés précédemment, quelques éléments sont importants à mentionner.

Tout d’abord la gestion des variables manquantes laisse à désirer. En effet, nous sommes conscients que les observations dont la durée de chômage prend la valeur 0 peuvent indiquer deux éléments contradictoires : puisqu’elles concernent à la fois ceux qui ont vécu une période de chômage inférieure à un an et ceux qui n’ont jamais été au chômage. Cependant, le fait de recouper avec la variable *chomeur\_actu* (issue de *fsitua* = 4) nous a assurément permis de contrôler ce problème. Ainsi, la population à risque est considérée correctement définie.

Concernant la variable d’intérêt, l’un de nos plus gros soucis est la nature des données censurées. De fait, la nature des données ne permet que de prendre en compte une censure à droite pour ceux dont on sait qu’ils sont toujours au chômage mais dont on ne dispose pas de la durée, et non de la censure à gauche discutée dans la littérature. Il en est de même pour la variable *duree\_handi*, où le regroupement en 3 modalités imposées par la discontinuité de la variable et sa censure n’est pas idéal pour juger de son impact sur la variable expliquée. De plus, la gestion de la discontinuité de cette même variable ne vaut pas la présence d’une variable continue. Notre pouvoir explicatif est donc limité en raison de la nature des données fournies.

Enfin, le modèle globale, permet bien de rendre compte de l’effet positif de l’approximation d’Efron, puisque celle-ci nous permet d’obtenir davantage de significativité. Par contre, le nombre de variables implique que, toujours les mêmes, à savoir *sexeng*, *classage*, *classrev*, *flnaim*, *fnivdip* et *zspo\_1* semblent contenir tout le pouvoir explicatif du modèle. Ce fait, nous contraint à tenter de trouver d’autres déterminants par le biais de sous-modèles ; nous avons conscience de leur incomplétude.

### 5.2 Ouvertures et améliorations possibles

Conformément aux soucis évoqués ci-dessus, différentes variables explicatives plus informatives sur le sujet peuvent être incluses. Notamment, si nous avions eu les dates de début et de fin de chômage sous forme de deux variables supplémentaires, nous aurions pu, d’une part, éviter la discontinuité de la variable, et d’autre part avoir une meilleure gestion de la censure à droite et surtout à gauche. De même, si la durée de handicap elle-même avait pris cette forme nous aurions certainement pu juger de l’impact de la discrimination à l’emploi par le handicap sur la durée de chômage.

Afin de traiter toutes les formes de discriminations, nous manquent de réelles variables ethniques. De fait, étudier les discriminations raciales à l’emploi avec comme déterminant de la durée du chômage seulement la nationalité ou le lieu de vie passé d’un individu n’est pas suffisant. Statistiquement, ce type de données - non-disponibles en France - consiste donc un réel manque à notre modèle.

Par ailleurs, le fait que nos données suivent une loi exponentielle implique une fonction de hasard constante dans le temps. Il est surprenant d’avoir ce résultat concernant des durées, car il suppose une indépendance au temps. Nous supposons que ce fait est également dû à la discontinuité de la variable. Si les données avaient suivies, comme présumé, une loi de Weibull, notre fonction de hasard aurait évolué dans le temps, selon les valeurs de  $p$  - impliquant nécessairement d’autres résultats.

Autrement, l’offre de variables psychologiques, présentes dans l’enquête, mérite notre attention. Le choix de les évincées parce qu’elles restent subjectives, a aussi été influencé par le nombre important de variables dont nous disposons. Les principales (seuls déterminants que nous avons pu affirmer) prenant tout le pouvoir explicatif, il aurait été compliqué d’établir l’impact des variables psychologiques sur la durée de chômage. Cependant, avec une meilleure offre de ces variables, nous pourrions établir un modèle spécifiquement destiné à donner la relation entre celles-ci et la durée du chômage.

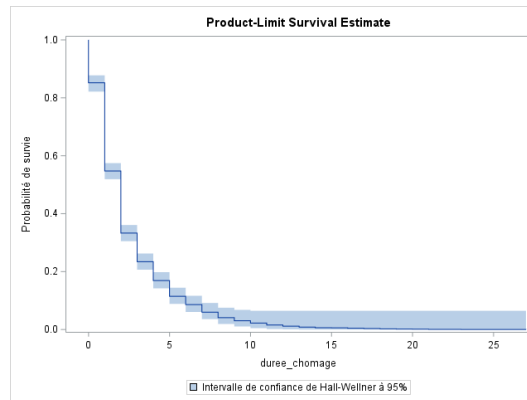
Enfin par rapport aux variables étudiées dans la littérature, d’autres déterminants importants auraient pu être inclus comme le lieu d’habitation, les modalités de déplacement et le type d’emploi recherché.

## Références

- [1] M. BONNEU A. NASSIRI, M. DELECROIX. "Estimation non paramétrique du taux de hasard : Application aux durées de chômage censurées à droite". *Annales d'Économie et de Statistique*, (n°58) :pp 215–232, 2000.
- [2] Liliane BONNAL Denis FOUGERE. "Les déterminants individuels de la durée du chômage". *Économie prévision*, (n°96) :pp 45–82, 1990-5.
- [3] Thierry TESTE. "Les déterminants individuels de la durée de chômage : conséquence de l'oubli du biais d'échantillonnage". *[Rapport de recherche] Laboratoire d'analyse et de techniques économiques(LATEC)*, (n°9717), 1997.

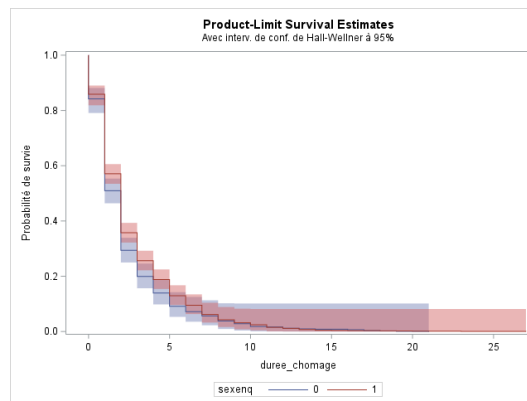
## 6 Annexes

FIGURE 1 – Fonction de Survie par Kaplan-Meier.



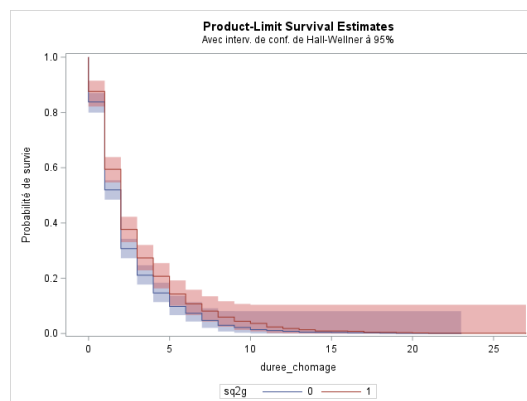
*Note : 472 observations with invalid time or censoring values were deleted.*

FIGURE 2 – Fonction de Survie stratifiée par *sexenq* - Kaplan-Meier.



*Note : 472 observations with invalid time, censoring, or strata values were deleted.*

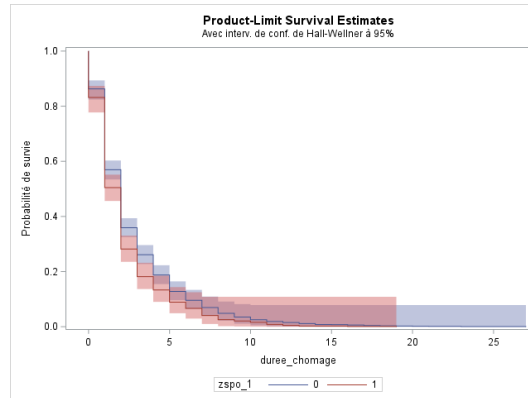
FIGURE 3 – Fonction de Survie stratifiée par *sq2g* - Kaplan-Meier.



*Note : 472 observations with invalid time, censoring, or strata values were deleted.*



FIGURE 4 – Fonction de Survie stratifiée par *zspo\_1* - Kaplan-Meier.



Note : 472 observations with invalid time, censoring, or strata values were deleted.

FIGURE 5 – Significativités des paramètres- modèle de Cox.

Tests Type 3			
Effet	DDL	Khi-2 de Wald	Pr > khi-2
fnelev_2	4	1.4621	0.8333
fnivdip	8	11.6529	0.1674
flnais	2	0.7351	0.6924
fnaim	2	5.2149	0.0737
fpermer	5	8.1754	0.1468
sq2g	1	0.0737	0.7861
classrev	2	11.2641	0.0036
sexenq	1	8.7168	0.0032
classage	4	153.8816	<.0001
zben_1	1	0.5519	0.4575
zsyn_1	1	0.3262	0.5679
zpol_1	1	0.2496	0.6173
zreli_1	1	0.8643	0.3525
zart_1	1	1.8514	0.1736
zspo_1	1	2.2145	0.1367
duree_handi	3	3.0197	0.3886
EDM	1	1.2174	0.2699
TAG	1	0.1114	0.7385

#### Analyse des valeurs estimées du maximum de vraisemblance - Efron

Paramètre	Valeur estimée des paramètres	Erreur type	khi-2	Pr >khi-2	Rapport de risque	Libellé
<i>fnivdip</i>	0.26737	0.08370	10.2034	0.0014	<b>1.307</b>	fnivdip 7
<i>fnaim</i>	-0.18812	0.07084	7.0516	0.0079	<b>0.829</b>	fnaim 2
<i>classrev</i>	0.16115	0.05761	7.8246	0.0052	<b>1.175</b>	Rang pour la variable zremen 2
<i>sexenq</i>	0.17223	0.04646	13.7442	0.0002	<b>1.188</b>	sexenq 0
<i>classage</i>	0.78842	0.07142	121.8629	<.0001	<b>2.200</b>	Rang pour la variable agenq 0
<i>classage</i>	0.19701	0.06627	8.8379	0.0030	<b>1.218</b>	Rang pour la variable agenq 1
<i>classage</i>	-0.37842	0.08084	21.9112	<.0001	<b>0.685</b>	Rang pour la variable agenq 4
<i>zspo_1</i>	0.08941	0.04646	3.7027	0.0543	<b>1.094</b>	zspo_1 1