

L'ANALYSE DE LA TRANSITION
FORMATION-EMPLOI :
LE CAS DU GRAND DUCHE DE LUXEMBOURG

Contribution to the 2nd Annual Report in the
Frame of EC PETRA Program

Laure-Suzanne ALLEGREZZA-CARVOYEUR[1]
Jean-Luc KOP[2]

Walferdange, juillet 1991

[1] Université de NANCY II
[2] CEPS/INSTEAD et Université de NANCY II.

CESIJE ASBL
Centre d'Études sur la Situation
des Jeunes en Europe
LUXEMBOURG

RESUME

La durée de non emploi que subissent les jeunes à la fin de leurs études peut-elle être un indicateur de leurs difficultés d'insertion professionnelle ? En d'autres termes ceux qui ne trouvent pas rapidement un emploi ont-ils des caractéristiques différentes de ceux qui s'insèrent rapidement dans le marché du travail ? C'est à cette question que se propose de répondre l'étude présentée. Pour ce faire, elle s'appuie sur les données micro-économiques de l'Enquête par panel auprès des ménages luxembourgeois (PSELL). 78 individus, sur plus de 6000 que compte le panel, ont été recensés comme ayant quitté l'école à la fin de l'année scolaire 1985. Les techniques économétriques de l'histoire événementielle sont utilisées afin de modéliser les chances qu'ont les individus de trouver un emploi, une fois sortis du système scolaire. Les conclusions sont les suivantes. Un certain nombre de caractéristiques dont, à priori, on attendait un effet se révèlent non significatives. Ainsi, les femmes n'ont pas plus de difficultés que les hommes, les étrangers pas plus que les nationaux. Plus encore, que les études suivies soient de type général ou qu'elles soient de type technique, les élèves n'ont pas un meilleur accès à l'emploi. Au total, seuls l'âge de fin d'études et le niveau scolaire atteint sont apparus influencer la durée de non emploi. Ces quelques remarques doivent être prises avec réserve en raison de la faiblesse numérique de l'échantillon retenu.

L'ANALYSE DE LA TRANSITION FORMATION-EMPLOI :
LE CAS DU GRAND DUCHE DE LUXEMBOURG.

SOMMAIRE

INTRODUCTION

PRESENTATION GENERALE DE L'ETUDE

ASPECTS METHODOLOGIQUES : LES SOLUTIONS RETENUES
POUR ANALYSER LA TRANSITION FORMATION-EMPLOI

LES RESULTATS

CONCLUSION

BIBLIOGRAPHIE

ANNEXE

INTRODUCTION

A l'époque où en Europe le chômage des jeunes est une question d'actualité, il peut être intéressant de mesurer les difficultés d'insertion professionnelle des sortants du système scolaire. C'est ce que nous nous proposons de voir dans le cas du Luxembourg. Ce petit pays dont l'avantage est d'avoir un taux de chômage de la population très faible (environ 2 %) offre peut être le modèle à suivre pour l'insertion professionnelle des jeunes.

Nous nous sommes plus particulièrement intéressés à l'analyse de la durée du non emploi des "sortants" du système scolaire résidant au Grand Duché de Luxembourg. De ce fait sont retenus aussi bien les jeunes qui recherchent effectivement un emploi que ceux qui n'en recherchent pas (les inactifs). A ces deux catégories il faut en ajouter une troisième : ceux qui ont une durée de non emploi nulle.

L'analyse présentée ci-dessous est basée sur des données micro-économiques qui sont issues du panel socio-économique luxembourgeois (PSELL), et plus précisément des vagues 1985 et 1986. Cette analyse est sujette à certaines limites. La première concerne la taille de l'échantillon : seuls 78 jeunes scolarisés en 1984-1985 ont été observés comme quittant le système scolaire en fin d'année scolaire 1985. Une deuxième limite découlant de la première a trait à l'absence de distinction^[1] au sein de la population des "non-emplois" du type d'individus :

- s'agit-il d'individus connaissant un chômage continu ?
- sont-ce des individus qui passent de chômeurs à inactifs, auquel cas il faudrait les classer parmi les travailleurs que les économistes appellent "découragés" ?
- ou inversement s'agit-il d'individus qui passent d'inactifs à chômeurs ?

[1] Compte tenu de la faiblesse numérique de l'échantillon, nous avons été contraints de limiter les catégories d'individus.

Enfin la brièveté de la période étudiée (8 mois - de septembre 1985 à avril 1986) limite la portée de nos résultats. Cette étude doit donc être considérée comme provisoire. L'ajout des vagues successives du PSELL permettra, en accroissant la taille de l'échantillon et la longueur de la période d'observation, d'infirmier ou de confirmer les tendances apparues dans ce rapport.

Les considérations économiques suggèrent que les caractéristiques socio-démo-économiques peuvent avoir un effet négatif sur la durée de non emploi de certains sortants, positif pour d'autres ; ainsi le sexe féminin peut avoir un effet positif sur cette durée, et le sexe masculin un effet négatif. Quant au type de formation suivie, on s'attend à ce que les sortants de l'enseignement professionnel aient une durée de non emploi plus courte que ceux qui sortent de l'enseignement général, car, en principe, la formation technique est plus adaptée aux caractéristiques du marché du travail.

Pour la mesure de ces effets, nous utilisons des techniques économétriques particulières, notamment celles de l'histoire événementielle.

Le déroulement de notre travail est le suivant. Nous commençons par une présentation générale de l'étude (I). Cette partie est suivie par un exposé de la méthodologie mise en oeuvre pour analyser la transition formation-emploi (II). Enfin nos résultats sont indiqués (III).

I - PRESENTATION GENERALE DE L'ETUDE

L'intérêt d'une étude sur la transition entre la formation et l'emploi nous semble suffisamment évident pour qu'il ne soit pas nécessaire d'y insister longuement. Pour le politique, de telles études peuvent servir par exemple à vérifier dans quelle mesure il y a adéquation entre le système de formation et le marché de l'emploi ou encore à reconsidérer certains types d'enseignement pour lesquels la transition s'effectue difficilement. Dans cette perspective, il est important du point de vue scientifique d'aborder ce problème avec une visée à la fois descriptive et explicative. Autrement dit, il s'agit dans un premier temps de décrire les caractéristiques des individus qui quittent le système scolaire chaque année. Ensuite, il s'agit de construire un modèle qui permette de prédire de la manière la plus fiable possible la qualité de la transition ; encore faut-il s'interroger sur la pertinence des indicateurs à utiliser pour mesurer cette qualité. Plus spécifiquement, on essaiera de mettre en évidence les caractéristiques de la formation et/ou de l'environnement du jeune qui sort du système scolaire qui ont le plus d'influence sur la qualité de la transition ; corrélativement, on essaiera d'insister parmi les variables explicatives sur celles qui sont potentiellement les plus susceptibles d'être modifiées par des actions externes et qui sont, de fait, les plus intéressantes dans une perspective appliquée. Par exemple, montrer que le sexe ou la nationalité d'origine ont une influence sur la durée de la transition peut suggérer une politique éducative qui vise à réduire les inégalités liées à ces caractéristiques. Bien entendu, les variables d'enseignement doivent avoir, dans cette perspective, une importance primordiale ; montrer la longueur de la transition (i.e. la difficulté de l'insertion) pour un certain type d'enseignement doit interroger les pouvoirs publics sur l'intérêt de conserver ce type d'enseignement.

Dans ce premier chapitre, nous allons essayer de donner une vision globale de notre étude en exposant ses caractéristiques principales, en explicitant les définitions utilisées et en faisant la liste des variables retenues pour les différentes analyses. Il s'agit en fait de donner au lecteur les informations suffisantes pour comprendre et justifier le point de vue particulier qui a été adopté ici.

1) Les caractéristiques générales de l'étude

Comme nous l'avons vu dans l'introduction, les analyses présentées ici se proposent d'étudier certains des aspects de la transition entre l'école et l'emploi. Le fait de préciser que seuls certains des aspects de cette transition sont envisagés dans ce rapport n'est pas trivial par rapport à notre démarche et nous allons voir que plusieurs raisons expliquent cette restriction.

Les analyses menées pour ce rapport se basent sur les données du panel socio-économique des ménages au Luxembourg. Il ne s'agit donc pas, comme nous l'avons déjà dit, d'une recherche fondamentale où l'on tenterait de vérifier empiriquement un modèle théorique. Nous nous situons délibérément dans une recherche appliquée dans laquelle seules les variables qui appartiennent à la banque des variables du panel luxembourgeois peuvent être utilisées. Nous sommes donc amenés à faire des choix quant aux variables à prendre en compte en fonction de leur pertinence par rapport à notre objet d'étude. Il est donc bien évident que nous ne pouvons tester des hypothèses spécifiques quant à l'influence (facilitante ou inhibitrice) de certaines variables (qui ne figurent pas dans la banque des variables du panel) sur la durée de la transition ; c'est dans ce sens que nous parlons d'une étude limitée à certains aspects de la transition école-emploi. Il faut noter également que le travail présenté ici n'est qu'une première analyse où toute la richesse des informations disponibles n'a pas été exploitée. Pour l'avenir, d'autres aspects doivent être étudiés comme l'influence des carrières (ou trajectoires) scolaires sur les carrières de la vie active. Ces analyses n'ont pu être menées, notamment en raison de la faible durée d'observation (2 années).

Ces analyses s'appuient sur un long et fastidieux travail préalable d'épuration des variables d'enseignement du panel. En effet, on s'est aperçu que les informations concernant la scolarité des enfants, récoltées le plus souvent auprès de leur famille, étaient entachées de nombreuses incohérences qu'il était nécessaire de corriger. Le temps passé pour effectuer ces corrections explique que deux années seulement ont pu être prises en compte pour cette première étude. Ces corrections ont fait l'objet d'une note technique interne (KOP & ALLEGREZZA-CARVOYEUR 1991) et ne sont donc pas évoquées ici mises à part les informations nécessaires à la compréhension globale de notre démarche.

2) Les définitions utilisées

Il est nécessaire, dans toute étude, de préciser les définitions adoptées par les auteurs lorsque celles-ci ne sont pas transparentes. Or, on pourrait penser que dans la scolarité et dans l'emploi, ces définitions ne posent pas de problèmes particuliers : cela serait oublier les problèmes de formation complémentaire (i.e. formation non initiale^[2]) ou de travail à temps partiel (i.e. à partir de combien d'heures de travail par semaine considère-t-on qu'un individu travaille ?) pour ne citer que deux exemples.

Les définitions que nous allons adopter résultent toutes d'un choix délibéré et pertinent par rapport à notre problématique. Il n'est pas question pour nous d'imposer ces définitions comme des définitions normatives et figées. Elles servent simplement à circonscrire notre domaine d'études et à préciser très exactement quels sont les individus concernés. Cette précision nous paraît nécessaire afin de permettre à n'importe quel autre chercheur d'effectuer de nouvelles analyses sur les données traitées ici ou encore sur des données d'autres pays dans une perspective comparative. Nous sommes cependant pleinement conscients du caractère imparfait et souvent réducteur de ces définitions et des critiques qu'elles ne manqueront pas de soulever. Néanmoins, rien n'empêche de reprendre ces définitions, de les remodeler et de recommencer éventuellement les analyses.

A - Les individus scolarisés

La définition des individus scolarisés est très importante car c'est elle qui nous permet de déterminer quels sont ceux qui font partie de notre échantillon.

On considère qu'un individu scolarisé est un individu :

- qui est en formation à plein temps
- dont l'activité principale est constituée par cette formation
- qui n'est pas handicapé au point de devoir suivre un enseignement dans un établissement spécialisé

[2] Formations entreprises après une interruption de la scolarité (ex : cours du soir, stages de formation en entreprises, etc.).

Commentons cette définition en insistant sur les individus qui se trouvent de ce fait exclus de notre groupe d'enquête. Nous ne considérons pas comme scolarisés les sujets qui suivent des formations complémentaires, des stages professionnels ou des cours du soir. Cette position peut être justifiée par les raisons suivantes :

- il ne s'agit pas d'enseignements à temps plein ; on peut supposer par ailleurs qu'ils sont suivis en grande majorité, par des individus qui ont une activité professionnelle ;
- il ne s'agit pas de formation initiale mais de formation complémentaire qui est très dépendante de la situation professionnelle actuelle des individus.

Par conséquent, on va trouver une grande diversité de cas de figure parmi ces formations sans qu'il soit possible d'en établir une liste exhaustive ni même une typologie (on peut envisager par exemple, en cours du soir, tant des cours d'initiation à l'informatique ou à une langue que des cours dispensés en vue de l'obtention d'un diplôme technique de l'enseignement supérieur).

Sans nier l'influence potentielle de ces formations sur la carrière professionnelle, leur non-prise en compte n'est pas contradictoire avec les objectifs de cette étude.

De la même façon sont exclus les individus handicapés physiques ou mentaux qui fréquentent des établissements spécialisés. Leur nombre est très faible^[3] et leur situation par rapport au marché du travail est de toute manière très spécifique (ateliers protégés, etc.). Seule une étude sur un échantillon composé de nombreux handicapés serait significative si l'on désire s'intéresser de manière plus fine à ce problème. Il faut noter aussi que cette exclusion ne concerne que les handicaps les plus graves (e.g. infirmes moteurs-cérébraux) ; les handicapés légers ayant de plus en plus tendance à être intégrés dans le système scolaire ordinaire.

[3] 9 individus seulement sur l'ensemble des scolarisés des deux premières vagues d'enquête sont dans des établissements pour handicapés.

B - Les apprentis

La situation par rapport à l'apprentissage est intéressante pour notre étude. On peut en effet penser que le fait d'avoir suivi un apprentissage a un effet facilitant l'intégration dans le monde professionnel. C'est avec un soin tout particulier que nous avons vérifié les informations concernant l'apprentissage d'autant que l'on a pu noter beaucoup d'erreurs dans les réponses des familles.

On considère donc comme apprenti tout individu qui est scolarisé en enseignement professionnel en cycle moyen et qui possède un contrat d'apprentissage qui l'oblige à suivre une partie de sa formation en entreprise. Il découle de cette définition que le statut d'apprenti est lié de façon stricte à un certain nombre de caractéristiques présentes dans le questionnaire. Un apprenti doit :

- être scolarisé
- en lycée technique II ou en lycée technique III
- en 10^e, 11^e ou 12^e année
- toucher un salaire d'apprentissage (le salaire touché doit être substantiellement plus bas que le salaire d'une personne ayant une activité salariée)

Cette définition nous a permis à la fois de considérer comme non-apprentis des individus qui ne respectent pas ces conditions mais également de considérer comme apprentis des individus qui se déclarent non-apprentis mais dont la situation correspondait strictement à la définition (pour plus de détail, cf KOP & ALLEGREZZA-CARVOYEUR 1991).

C - L'emploi

Pour définir la situation par rapport à l'emploi, nous avons eu recours à des définitions strictement opérationnelles en fonction des réponses (après épuration) au tableau d'orientation des questionnaires du panel luxembourgeois (ce tableau d'orientation renseigne sur la situation des individus mois par mois sur les 12 mois qui s'écoulent entre deux enquêtes). On peut notamment dégager 3 grands cas de figure :

- les individus sont considérés comme ayant un emploi s'ils déclarent travailler au moins 10 heures par semaine
- les individus sont considérés comme étant au chômage s'ils déclarent rechercher un emploi et être inscrits ou non à l'agence pour la recherche d'emploi.
- les individus sont considérés comme étant inactifs s'ils sont dans une autre situation (non scolarisés, non chômeurs et ne travaillant pas)^[4].

[4] Le cas des retraités n'a pas de sens dans notre étude.

3) La liste des variables retenues pour l'étude

On distingue classiquement les variables dépendantes (variables à expliquer) et les variables explicatives (variables éventuellement indépendantes) ; c'est en fonction de cette classification que sera organisé ce paragraphe. Le lecteur peut trouver en annexe les statistiques descriptives de ces variables calculées pour l'ensemble des individus de notre échantillon (n = 78).

A - Les variables dépendantes

Dans cette première étude, nous cherchons à expliquer la durée de non emploi après la sortie du système scolaire. Pour pouvoir calculer cette durée, il est nécessaire de pouvoir disposer d'informations précises quant à la situation des individus ; ces variables sont fournies par le tableau d'orientation dont on a déjà parlé plus haut. Il faut signaler à ce sujet que des erreurs relativement systématiques ont été mises en évidence sur les informations de ce tableau justement pour les individus qui quittent le système scolaire entre deux enquêtes consécutives. En effet, on constate que très souvent la situation des individus est constante sur les 12 mois. Etant donné que les enquêtes ont lieu au mois de mai de chaque année, cela revient à considérer que des individus sont scolarisés jusqu'en avril 85 et changent de situation à partir de mai (travaillent, recherchent un emploi, etc.) ce qui n'est pas cohérent avec le système scolaire qui fixe la fin de l'année aux alentours de la mi-juillet. Nous avons donc été amenés à construire deux modèles à partir de ces informations, modèles qui vont conduire à calculer des durées différentes de non emploi.

a) Le modèle 1

Dans un premier modèle (modèle 1) on considère que les sujets scolarisés l'année scolaire 1984-85 le sont jusqu'en août 1985 quelles que soient les réponses données quant à la situation pour les mois de mai 1985 à août 1985. Ce n'est donc qu'à partir de septembre 1985 que l'on prend en compte les informations concernant la situation, complétées éventuellement par la date à laquelle les sujets déclarent commencer à travailler dans l'entreprise qui les emploie^[5].

[5] Cette information n'est bien entendu disponible que pour les individus qui travaillent à la date de la seconde enquête.

b) Le modèle 2

Dans un second modèle (modèle 2) on se réfère de façon stricte aux informations fournies par les questionnaires concernant la situation professionnelle (les informations du tableau d'orientation sont donc considérées comme fiables).

c) Les différentes situations possibles

Quel que soit le modèle, les individus peuvent être dans l'une des 8 grandes catégories suivantes (ces catégories sont exhaustives et exclusives) :

- les sujets sont scolarisés et apprentis ;
- les sujets sont en études rémunérées ; ce sont des individus qui touchent un salaire ou une bourse qui ne peuvent être considérés comme salaire d'apprentissage car les formations dispensées sont des formations supérieures (ce sont par exemple les élèves d'écoles d'infirmières qui, dans le cadre de leur formation, sont employés dans des hôpitaux qui les rémunèrent) ;
- les sujets sont les autres scolarisés ;
- les individus déclarent tenir le ménage ; ils ne sont donc pas scolarisés, ne travaillent pas et ne recherchent pas d'emploi ;
- les individus disent explicitement rechercher un emploi ;
- les individus travaillent ;
- ils déclarent être dans une autre situation. Cette catégorie n'est pas une catégorie résiduelle construite mais elle correspond à l'une des possibilités de réponse offerte par le questionnaire. Aussi, ne pouvons-nous pas détailler les situations qui correspondent à cette catégorie.

d) Le calcul des durées de non emploi

Munis de ces informations, il devient alors très facile de calculer des durées de non emploi, selon les deux modèles décrits supra, en comptant le nombre de mois s'écoulant entre le dernier mois de scolarisation et le premier mois où l'individu travaille.

A ces deux variables de durée de non emploi se rajoutent deux indicateurs de censure correspondant à ces deux modèles. Une durée est dite censurée lorsqu'elle est connue avec incertitude ; c'est le cas pour tous les individus qui n'ont pas pris d'emploi pendant la période d'observation.

Les durées calculées sont effectivement des durées de non emploi et pas seulement des durées de chômage ; il ne deviendra possible de distinguer entre les individus au chômage et les individus inactifs que lorsque le nombre de sujets de l'échantillon sera plus élevé, autrement dit lorsque de nouvelles vagues d'enquête seront prises en compte.

Notons enfin qu'il revient au chercheur de déterminer les variables dépendantes auxquelles il s'intéresse et que le choix a été fait dans cette première étude de ne retenir que la durée de non emploi ; il sera important de rajouter à cette variable d'autres indicateurs de la qualité de la transition comme le niveau du premier salaire ou la durée passée dans le premier emploi. Seules, à nouveau des contraintes tenant à la fois au faible effectif de l'échantillon et à la courte durée d'observation nous empêchent de le faire.

B - Les variables explicatives.

Un certain nombre de variables ont été retenues pour essayer de construire un modèle explicatif de la durée de non emploi. Ces variables ont été choisies parmi l'ensemble des variables disponibles dans le panel luxembourgeois et en fonction de leur pertinence par rapport à l'objet d'étude : ce sont les variables qui sont considérées souvent comme ayant le plus de poids sur le déroulement de la scolarité et/ou de la vie professionnelle^[6].

Voici la liste de ces variables regroupées par thème et accompagnées si nécessaire d'explications concernant leur provenance ou leur mode de calcul.

Les variables au niveau individuel

- Les variables démographiques

- * Le sexe des individus
- * L'âge des individus en 1985
- * L'ancrage des individus : nous avons pris comme variable d'ancrage un indicateur composite d'intégration qui varie de 0 à 100 : 0 représente une intégration très faible et 100 une intégration maximale. Neuf variables rentrent dans la composition de cet indicateur : la nationalité, le pays de naissance, la langue maternelle et la maîtrise des 3 langues utilisées au Luxembourg (luxembourgeois, français et allemand). Cet indicateur a été élaborée par TOURNOIS [1990].

[6] Notons que le choix a été établi en concertation entre les différents membres du projet PETRA.

- Les variables de scolarité en 1985

- * L'année d'enseignement en 1985 : pour cette variable, nous avons tout simplement compter le nombre d'années à partir de la première année d'enseignement primaire^[7] (sans compter les redoublements éventuels).
- * Le type d'enseignement suivi en 1985 (i.e. lycée technique, enseignement général, supérieur universitaire, etc.)

Les variables au niveau du ménage

- Les variables de composition du ménage

- * Le nombre total de personnes dans le ménage
- * Le nombre d'adultes dans le ménage (i.e. les individus qui ont atteints ou dépassés l'âge légal de la majorité ; 18 ans)
- * Le nombre d'enfants dans le ménage (i.e. les individus qui n'ont pas encore atteints l'âge légal de la majorité ; 18 ans)
- * Le nombre de personnes avec un emploi dans le ménage
- * Le nombre de personnes "pouvant être actives" dans le ménage (sont considérées comme personnes actives dans le PSELL (1) les hommes de 25 à 64 ans et les femmes de 25 à 59 ans quel que soit leur statut vis-à-vis de l'emploi (actifs occupés, chômeurs ou inactifs) (2) les jeunes de moins de 25 ans qui ont un emploi ou qui perçoivent le chômage ou qui ont déjà travaillé.
- * Une variable de typologie des ménages (cette typologie a été établie en fonction des différentes variables de composition du ménage énoncées plus haut).

- La situation géographique des ménages par cantons

- Le statut économique et social des ménages

*Le statut social du ménage

La formation achevée du chef de ménage

- * Le statut économique des ménages ; pour plus de détails sur le mode de calcul de ces différents indicateurs, on peut consulter TOURNOIS [1988].

[7] Par exemple, la classe de première en enseignement général est recodée en treizième année, la première année universitaire en quatorzième année.

Les revenus totaux du ménage

- Le total des revenus individuels mensuels des membres du ménage en 1985 (c'est-à-dire essentiellement des revenus de l'activité)
- Le total des revenus mensuels du ménage en 1985 (on tient compte ici, en plus, des revenus non individualisables, comme les prestations familiales)

Les quotients familiaux

- Le quotient familial système C.E.S. (ce quotient familial tient compte du nombre de personnes dans le ménage et de leur âge).
- Le quotient familial système L.I.S. (il ne tient compte que du nombre de personnes dans le ménage)

* Les indicateurs d'intégration au niveau du ménage (pour tous ces indicateurs, cf TOURNOIS [1990] ; ils se réfèrent à la même logique que ceux construits au niveau individuel)

- Un indicateur d'intégration du ménage (cet indicateur qui varie de 0 à 100 est tout simplement la moyenne des scores individuels d'intégration (cf supra) des personnes de plus de 12 ans du ménage).
- Un indicateur de l'hétérogénéité de l'intégration au sein du ménage (l'hétérogénéité de l'intégration est appréciée par la somme, pour chaque ménage, des différences élevées au carré entre le score individuel d'intégration et l'indicateur d'intégration du ménage ; cette somme est pondérée par le nombre total de personnes de plus de 12 ans du ménage).
- Une typologie des ménages en fonction de leur intégration et de leur hétérogénéité.

Avant d'aborder la partie empirique de ce travail qui se propose de mettre en évidence la relation entre ces variables et la durée du non emploi, nous allons exposer dans le prochain chapitre les choix méthodologiques retenus pour ces analyses.

II - ASPECTS METHODOLOGIQUES : LES SOLUTIONS RETENUES POUR ANALYSER LA TRANSITION FORMATION-EMPLOI.

Pour tester les difficultés d'insertion professionnelle des sortants du système scolaire, nous développons un modèle qui permet l'estimation de la probabilité, pour un sortant, d'être dans un état donné à chaque point du temps d'un horizon infini. En d'autres termes nous voulons reconstituer des trajectoires individuelles.

Alors que l'analyse descriptive peut être considérée comme une approche préliminaire, l'estimation dynamique va gagner un certain intérêt dans le contexte des modèles de durée. En effet, la prise en compte d'une troncature - une incertitude sur la dimension de la période étudiée - va limiter les biais inhérents à l'approche descriptive. Pour ce faire, nous faisons l'hypothèse que ce sont les décisions des individus qui déterminent la durée qu'ils vont passer dans le statut de non emploi.

C'est pourquoi notre modèle empirique du comportement doit permettre (a) d'expliquer la décision de participation ou non à l'emploi

et (b) d'expliquer la durée passée dans le statut de non emploi, et la transition vers le statut d'emploi.

1) Les concepts utilisés

A - Le choix du statut.

La théorie économique postule, le plus souvent, que l'individu choisit. Il faut donc préciser (a) quelle est la nature des différents choix qui s'offrent à lui, (b) comment ces choix peuvent être modélisés.

a) La nature des choix

Il s'agit de spécifier la nature (la mesure) de la décision (variable) que l'on cherche à expliquer. En effet la participation au marché du travail recouvre deux notions : la participation effective à une activité rémunérée et la participation à la force de travail (occupe effectivement un emploi ou est à la recherche d'un emploi).

La décision de participer au marché du travail peut être mesurée de plusieurs façons :

- de façon continue, en utilisant comme variable dépendante expliquée, le nombre d'heures travaillées ou les revenus tirés de l'activité

- ou de façon discontinue en utilisant une variable catégorielle : binaire (par exemple : participe - ne participe pas), ou à modalités multiples (exemple : participe à plein temps, à temps partiel ou ne participe pas).

Compte tenu de nos objectifs (conditionnés en partie par l'information dont nous disposons), nous modélisons le binôme participe - ne participe pas à une activité rémunérée. Ce binôme donne lieu à une variable binaire qui prend la valeur 1 quand le sortant occupe un emploi rémunéré (même s'il s'agit d'un emploi à temps partiel) sur la période d'observation.

La décision est supposée prise à tout moment. Donc, à tout moment, t , un sortant est soit avec emploi, soit sans emploi. Ceci peut être représenté par une variable, $y(t)$.

b) la modélisation des choix

La probabilité d'état est le concept de base dans l'analyse de la transition. A $y(t)$, variable observée, correspond la variable aléatoire latente $Y(t)$, fonction continue, strictement monotone et qui décrit un processus dépendant du temps. Chaque statut peut alors être défini comme une probabilité d'état, la probabilité que l'individu i occupe l'état a au temps t : $P(t) = \Pr[Y(t) = a]$. On a : $\forall t, \sum_y p_y(t) = 1$, ce qui revient à écrire que la liste des états est exhaustive. L'estimation empirique de la probabilité d'état s'écrit :

$$p_y(t) = \frac{\wedge \text{ Nombre d'unités pour lesquelles } Y(t) = a}{\text{Nombre total d'unités observées}}$$

L'estimation des probabilités d'état dépend, de façon évidente, du temps particulier t choisi pour les calculer. Ainsi, à 70 ans, les individus de l'échantillon seront probablement tous sans emploi. La probabilité d'être sans emploi à cet âge ou au-delà est constante et égale à l'unité, celle d'avoir un emploi égale à zéro. Quand la probabilité d'état ne varie plus avec le temps, le processus a atteint l'équilibre^[8].

[8] Nous avons, infra, une autre probabilité qui est d'autant plus grande que la période d'observation est longue : c'est la probabilité d'observer un changement de statut.

B - les changements d'état

L'analyse précédente peut être améliorée grâce à deux nouveaux concepts qui permettent de mesurer les changements d'état ; il s'agit (a) de la probabilité instantanée de transition et (b) de la durée passée dans un état.

a) la probabilité instantanée de transition

Aux changements d'état correspondent deux probabilités de transition : la probabilité de transition entre deux points du temps et la probabilité de transition instantanée.

1. La probabilité de transition entre deux points du temps.

Considérons Δt un petit intervalle de temps. On exprime la probabilité de transition entre deux points du temps par l'équation : $P_{ijk}(t, t + \Delta t) = \Pr[Y(t + \Delta t) = k \mid Y(t) = j]$.

C'est, par exemple, la probabilité pour un individu d'avoir un emploi en t_1 sachant qu'en t_0 il n'en avait pas. Cette probabilité de transition dépend de t et de $t + \Delta t$. C'est une probabilité conditionnelle, la probabilité d'être dans l'état k en $t + \Delta t$, conditionnelle à celle d'être dans l'état j en t . On a comme estimation empirique de la probabilité de transition :

$$P_{jk} = \frac{\wedge \text{ Nombre d'unités qui sont en } j \text{ en } t \text{ et en } k \text{ en } t + \Delta t}{\text{Nombre d'unités qui sont dans l'état } j \text{ en } t}$$

$$= \frac{\Delta y}{\Delta t} ; \text{ c'est la probabilité d'un changement particulier dans } y \text{ entre deux points du temps.}$$

Si $P_{ijk}(t)$ est la probabilité de transition de l'état j vers l'état k , pour l'individu i ^[9], alors $p_{ij}(t)$ est la probabilité que le sortant i soit dans l'état j en t .

[9] C'est la probabilité que l'individu soit dans l'état k en t , sachant qu'il était dans l'état j en $t-1$.

2. La probabilité de transition instantanée est appelée taux de hasard^[10]

Notre analyse est axée sur la variation des proportions de changements de statut au cours du temps, parmi un groupe de sortants et sous des conditions spécifiées. C'est pourquoi nous avons recours au taux de hasard. Ce taux est défini comme la probabilité de transition sur une unité de temps, quand l'unité de temps est infinitésimale (et $\Delta t > 0$). C'est une façon de définir la dépendance de durée^[11]. Ce taux n'est pas, en fait, une probabilité car il peut être supérieur à l'unité. Plus formellement le taux entre deux états j et k est égal à :

$$r_{jk}(t) = \lim_{\Delta t \rightarrow 0^+} \frac{p_{jk}(t, t + \Delta t)}{\Delta t} \quad \forall j \neq k.$$

Comme p_{jk} est une probabilité et Δt est positif, le taux de transition n'est jamais négatif, bien que sans limite supérieure. Notons que si $r_{jk}(t) = r_{jk}$, on a des taux de transition instantanée qui sont indépendants du temps.

Les taux de hasard ne sont directement pas observables bien qu'ils soient utilisés comme variable dépendante latente dans l'analyse dynamique de variables discrètes. Mais, contrairement à ce qu'affirme ALLISON [1984], ils ont deux significations intuitives particulières : la durée moyenne passée dans chaque état et la probabilité qu'un événement ait lieu dans un intervalle de temps. Nous aurons l'occasion de voir le lien entre l'espérance mathématique de la durée dans l'état j et le taux de passage de l'état de non emploi, j , à celui d'emploi, k : $E(u_j) = 1/r_{jk}$. Si le changement d'état est conçu comme un événement, alors $E(u_j)$ peut être interprété comme le temps d'attente moyen jusqu'à l'événement. Enlevons les indices j et k pour simplifier l'écriture. Il reste :

$$r(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\text{Probabilité de transition}}{\Delta t}$$

[10] Nous nous référons à la terminologie employée par TASSI et alii [1990] (E.N.S.A.E.)

[11] On aura une dépendance de durée positive si la dérivée du taux de hasard par rapport au temps est positive, c'est-à-dire si $dr(t)/dt > 0$. Si tel est le cas, la probabilité de voir une sortie du non emploi augmente quand la longueur de la période augmente. Inversement, on aura une dépendance de durée négative si $dr(t)/dt < 0$.

ou, de façon plus formelle :

$$r(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{dy}{dt}$$

C'est la probabilité, pour l'individu, d'être dans un certain état^[12] en $t + \Delta t$, compte tenu du fait qu'il était déjà dans cet état en t . En d'autres termes, c'est la probabilité de ne pas avoir connu l'événement en $t + \Delta t$ sachant qu'il n'a pas été connu auparavant. C'est, par exemple, la probabilité pour une femme sortie du système scolaire de ne pas avoir pris d'emploi en t_2 , sachant qu'elle était déjà sans emploi en t_1 (avec $\Delta t = 1$, un jour, une semaine, un mois, etc.).

b) Présentation du concept de durée dans un état.

Nous cherchons à décrire comment nous allons modéliser la longueur de l'intervalle entre deux événements, c'est-à-dire comment tenir compte du temps passé dans chaque état. Pour ce faire, nous faisons appel à la probabilité de transition instantanée. Nous avons vu précédemment qu'il s'agit de la probabilité de transition sur une unité de temps, sachant que cette unité de temps est infinitésimale. Nous utilisons aussi la fonction de distribution des temps d'attente, qui mesure la longueur de l'intervalle entre deux événements.

Les variations de Y peuvent être décrites en termes de distribution des temps d'attente, c'est-à-dire en termes de longueur de l'intervalle entre deux événements. On a : $F_j(t) = \Pr(T \leq t)$. $F_j(t) = 1 - G_j(t)$ est la probabilité que l'événement j ait lieu au plus tôt en t . Son complément à l'unité, $G_j(t)$, s'appelle la fonction de survie. Il est égal à $G_j(t) = \Pr(T \geq t)$. En d'autres termes, on a $F_j(u) = \Pr[U < u]$ ^[13], la probabilité que le changement d'état ait lieu au bout de u unités de temps.

La fonction de distribution cumulée, la fonction de survie et le taux de hasard sont reliés entre eux. En effet, en temps continu, la fonction de hasard $r(t)$ spécifie le taux instantané d'échec en $T = t$, conditionnel à la survie au temps t . On a :

[12] En terminologie événementielle d'être "survivant"

[13] Elle est le complément, à l'unité, de la fonction de survie. Cette dernière exprime la probabilité d'un événement à chaque point du temps. Elle se note $G_j(u) = \Pr[U \geq u]$, et se lit : probabilité que l'individu soit encore dans le même état après u unités de temps (u mois dans notre cas).

$$r(t) = \frac{f(t)[14]}{1 - F(t)} = \frac{f(t)[15]}{G(t)}$$

De ce fait :

$$r(t) = - \frac{dG(t)}{G(t)} \cdot \frac{1}{dt} ;$$

Comme $dG(t)/G(t)$ est la dérivée de $\text{Log } G(t)$, on a $r(t) = -d\text{Log}G(t)/dt$. En intégrant et en utilisant $G(0) = 1$, on obtient :

* $G(t) = \exp \left(- \int_0^t r(u) du \right)$ qui est l'expression de la fonction de survie par rapport à la fonction de hasard.

Elle donne la probabilité que l'individu soit encore sans emploi au bout de t mois.

* $F(t) = 1 - \exp \left(- \int_0^t r(u) du \right)$ qui est la fonction de distribution des temps d'attente

* $f(t) = r(t) \cdot \exp \left(- \int_0^t r(u) du \right)$, dérivée de la précédente qui est la fonction de densité de probabilité.

2) - Les méthodes d'analyse

A - Le modèle probabiliste dans un cadre statique : une méthode de mesure de la probabilité de trouver un emploi.

Il existe plusieurs types de méthodes d'estimation applicables au choix du statut et à celui de l'état : des méthodes linéaires et des méthodes non linéaires. Nous survolons les premières dans le paragraphe suivant et nous développons ensuite les secondes.

Les méthodes d'estimation linéaires comme les moindres carrés ordinaires peuvent donner des estimateurs non convergents,

[14] $f(t)$, fonction de densité de probabilité, est égale à :

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\text{Pr}[t \leq T < t + \Delta t]}{\Delta t}$$

[15] or $f(t) = dF(t)/dt = -dG(t)/dt = d(1 - G(t))/dt$

car les résidus sont corrélés avec la variable dépendante. L'utilisation des moindres carrés pondérés qui rendent constante la variance des erreurs remédie à ce problème. Il reste toutefois des inconvénients, surtout parce que les valeurs estimées de la variable dépendante peuvent sortir de l'intervalle $[0,1]$. C'est pourquoi nous avons choisi d'employer une méthode non linéaire, à savoir le logit binomial.

Notre spécification d'un modèle de probabilité non linéaire repose sur l'usage d'une fonction logistique car la variable dépendante est catégorielle^[16] : elle a la valeur 0 ou 1. La fonction logistique binomiale s'écrit :

$$F(Z) = \frac{\exp(Z)}{1 + \exp(Z)} = P(Y = 1) \quad [17]$$

c'est-à-dire, si $Z = \beta'X_i$, où β est le vecteur des coefficients associés aux variables explicatives X_i :

$$F(\beta'X_i) = \frac{\exp(\beta'X_i)}{1 + \exp(\beta'X_i)} = \frac{1}{1 + \exp(-\beta'X_i)}$$

$P(Y = 1)$, probabilité que les ex-étudiants occupent un emploi, est une fonction non linéaire des variables explicatives. Si l'on veut prédire les effets de changements de l'une des variables explicatives sur la probabilité qu'un événement ait lieu, il est nécessaire de connaître la dérivée partielle de la fonction ci-dessus par rapport à chacune des variables explicatives :

$$\begin{aligned} \frac{\delta}{\delta X_i} F(\beta'X_i) &= \frac{\exp(\beta'X_i)}{[1 + \exp(\beta'X_i)]^2} * \beta_i \\ &= \beta_i * P(Y = 1) * P(Y = 0). \end{aligned}$$

β_i mesure l'effet d'un changement de X_i sur la variable dépendante latente (non observée) et non sur la variable dichotomique observée. Il apparaît comme un facteur multiplicatif et c'est lui qui détermine en fait le signe de l'effet, puisque l'autre facteur de l'équation est toujours positif. Cet effet n'est pas constant : il varie avec les X_i . En conséquence, il est nécessaire, dans les modèles logit, pour avoir une idée de l'intervalle de variation des changements résultant dans les proba-

[16] Nous faisons alors implicitement l'hypothèse que la distribution de probabilités des résidus de nos équations suit une loi particulière, qui est de type logistique.

[17] Donc : $F(-Z) = \text{Prob}(Y = 0) = 1 - \text{Prob}(Y = 1)$

bilités, de calculer la valeur de la variable dépendante à différents niveaux des variables explicatives. Quand ces niveaux sont faibles, le taux de changement est simplement la dérivée^[18].

Somme toute, en évaluant les effets des variables explicatives sur la probabilité de trouver un emploi, on gagne en lisibilité dans l'interprétation des coefficients. Cette étape est nécessaire pour pouvoir comparer le modèle probabiliste statique et le modèle probabiliste dynamique que nous avons mis en oeuvre. Nous nous penchons maintenant vers ce dernier.

B - Le modèle probabiliste dynamique ou modèle de durée : une méthode empirique d'analyse de la transition école - emploi, en présence de données censurées.

Nous voulons mesurer la durée passée dans l'état de non emploi. Or celle-ci est parfois connue avec incertitude (censurée^[19]) : les sortants du système scolaire ne sont pas tous observés comme prenant un emploi sur la période générale d'observation. Or une des techniques économétriques qui permet d'analyser des variables continues dans un échantillon incluant des données censurées repose sur le recours aux modèles de hasard. Cette technique incorpore explicitement le temps dans l'analyse. Elle permet d'étudier le temps que met un événement à se produire (par exemple, la durée de l'épisode avant qu'un jeune trouve un emploi). Elle utilise l'histoire passée de l'individu. On fait alors l'hypothèse que les choix établis à un moment donné ne sont pas innés, mais qu'ils se modifient suite à l'expérience personnelle. Cette prise en compte de l'histoire passée est rarement faite par les techniques traditionnelles d'analyse statique qui, le plus souvent, ignorent la dimension temporelle de l'événement étudié^[20].

[18] On doit néanmoins remarquer que le logit d'un événement est une fonction linéaire des variables explicatives. En abrégé, on a $\log P / \log(1 - P) = -\beta'X$.

[19] Certains parlent de données tronquées. En fait, il existe deux façons principales de qualifier les durées incomplètes. On parle de données tronquées à gauche (respectivement à droite) quand la variable d'intérêt n'est pas observée quand elle est inférieure (respectivement supérieure) à un certain seuil (BROESBEKE et alii [1989]). On parle de données censurées quand la variable d'intérêt n'est pas observée suffisamment longtemps : la valeur vraie de la variable est supérieure ou inférieure à la valeur observée.

[20] C'est le cas, par exemple, des modèles logit et probit, des tables de contingences, des régressions avec variables explicatives muettes.

On voit donc, dès lors, que l'unité d'analyse qui nous intéresse, est l'histoire événementielle de l'individu, c'est-à-dire les séquences ordonnées où il se déplace, le temps que prend chaque séquence (la durée de l'épisode de vie) ainsi que la façon dont ces séquences s'enchaînent dans le temps (les transitions vers d'autres situations).

Nous commençons par évoquer le problème de la censure des données avant d'explicitier les différentes approches que nous utilisons.

a) Le problème de la censure des données^[21].

Il y a censure quand les durées observées sont connues avec incertitude. Si c'est l'origine temporelle de l'événement qui est inconnue, on parle de censure à gauche ; sinon de censure à droite. Nos données concernant les sortants du système scolaire en 1985 ne sont pas censurées à gauche puisque, par construction, la période d'observation débute le premier mois de fin de scolarité^[22]. De plus, pour tous les individus, nous disposons d'une information sur la période scolaire (type d'études, niveau d'études, etc.).

La censure à droite concerne une proportion relativement faible d'individus : 15 %. C'est ~~pas~~ préoccupant dès l'instant où la durée maximale observée est censurée^[23], ce qui est notre cas. Il existe en fait trois types de censure à droite^[24], dont deux types sont présents dans notre échantillon : la censure dite "fixe" et la censure dite "aléatoire". Nous sommes concernés par la censure fixe de type I, parce que la longueur de la période générale d'observation est la même pour chaque individu de l'échantillon. Nous ne sommes pas concernés par la censure fixe de type II parce qu'il n'y a pas fin de l'observation après un nombre fixe d'événements. Mais nous avons un cas de censure aléatoire. Citons à ce propos P.D. ALLISON [1984], p. 29, "Random

[21] cf RAY et alii [1986] p. 191 et suivantes.

[22] Procéder de la sorte introduit un biais de sélection car les individus ayant quitté l'enseignement avant 1985 et n'ayant jamais encore travaillé auraient dû être retenus dans une étude sur le non emploi. Mais ces individus n'auraient pu être comparés à ceux de leur cohorte ayant trouvé un emploi avant 1985.

[23] Sinon il y a sous-estimation du temps moyen.

[24] HUBER [1989] répertorie deux types supplémentaires de censure, qui ne sont que des variantes de la censure aléatoire ; ce sont la censure progressive de type I et la censure progressive de type II.

censoring also includes designs in which observation ends at the same time for all individuals, but begins at different times".

Or la censure aléatoire pose problème : certains individus ne changent pas d'état simplement parce que nous ne les avons pas observés suffisamment longtemps. Se pose alors la question de savoir si l'on peut admettre que les périodes censurées sont indépendantes des durées observées. Mais comme nous n'avons que peu d'individus soumis à ce type de censure, nous ne pensons pas qu'il y a problème et qu'il faille vérifier les hypothèses d'indépendance.

b) Une méthode d'estimation pour un modèle avec hétérogénéité.

Les modèles de hasard sont une façon aisée d'analyser des transitions. L'analyse des transitions instantanées se fait d'au moins deux façons :

- d'abord par une analyse dite non paramétrique sur une population (ou des sous populations) homogène(s) dont on suppose a priori que tous les membres ont un comportement semblable.

- ensuite par une analyse de type paramétrique ou par une analyse de type semi-paramétrique^[25] permettant la mise en relation de la durée observée dans un statut et un vecteur de variables dont on suppose qu'elles peuvent influencer cette durée^[26].

Nous avons utilisé les deux types de méthodes d'estimation car nous avons procédé de la façon suivante. Pour trouver une forme fonctionnelle à la dépendance temporelle de $r(t)$, nous avons eu recours à la méthode, non paramétrique, actuarielle, de l'estimateur de KAPLAN-MEIER [1958]. Pour chercher les variations du taux de hasard, $r(t)$, qui dépendent de variables exogènes et pour construire un modèle de $r(t)$ qui dépende du temps, t , nous avons utilisé le modèle, semi-paramétrique, des hasards proportionnels de COX et différents modèles paramétriques. Nous avons commencé par le modèle de Weibull afin de vérifier si la liaison du taux de transition au temps est linéaire, auquel cas nous aurions une distribution exponentielle des taux de transition. Dans notre cas l'usage de modèles log-linéaires est limité par deux raisons : d'abord l'existence d'un grand nombre de durées nulles empêche de calculer le logarithme de cette durée, ensuite la plupart des distributions log-linéaires sont des distributions non défectueuses, ce qui suppose que tous les individus de notre

[25] Le premier type d'analyse, à la différence du second, suppose des hypothèses sur l'allure de la fonction de hasard.

[26] Il est aussi possible de faire des analyses paramétriques sans variable explicative.

échantillon vont dans un terme plus ou moins loin prendre un emploi ; or l'on sait, d'après l'Histoire passée, qu'il n'en est rien : certains ne vont jamais occuper d'emploi au cours de leur vie potentiellement active.

Nous développons maintenant le fonctionnement de ces différentes méthodes, en mettant l'accent sur les raisons qui nous ont poussés à les utiliser.

1. L'estimateur non paramétrique de KAPLAN et MEIER [1958] : fondement, mise en oeuvre et limites.

KAPLAN et MEIER sont les premiers à avoir développé un estimateur permettant de tenir compte des observations censurées à droite. Ils ont écrit une fonction de survie adaptée aux cas où un seul événement peut survenir.

Mais si l'estimateur de KAPLAN-MEIER est un estimateur non paramétrique, puisqu'il ne fait pas d'hypothèse sur la forme fonctionnelle de la distribution des événements étudiés, il n'est pas, en revanche et à proprement parler, un estimateur permettant de tenir compte de l'hétérogénéité de la population. C'est l'utilisation qui en est faite qui permet de vérifier si cette population est ou non hétérogène. C'est pourquoi nous commençons par développer le fondement mathématique de cet estimateur, puis son utilisation et enfin ses limites.

1.1 Le fondement de l'estimateur de KAPLAN et MEIER[27].

L'estimateur de KAPLAN-MEIER suppose que les dates où les événements ont lieu peuvent être ordonnées : $t_{(1)} < t_{(2)} < \dots < t_{(k)}$. En général, ces dates sont prises de façon à être également espacées : $[t_{(i)} - t_{(i-1)}] = c$.

On a $n_{(i)}$ nombre d'individus, au temps $t_{(i)}$ ($i = 1, \dots, k$),
 k périodes distinctes ($k \leq n$),
 $d_{(i)}$ nombre de "morts", au temps $t_{(i)}$ ($i = 1, \dots, k$),
 $m_{(i)}$ nombre de périodes censurées,
 $G(t)$ la fonction de survie.

Alors une estimation non paramétrique de la fonction de survie est fournie par le produit de tous les ratios, à chaque période, du nombre de survivants en fin de période au nombre d'individus en début de période :

[27] Il est aussi appelé estimateur du produit limite.

$$\hat{G}(t) = \prod_{i:t(i) < t} \frac{n(i) - d(i)}{n(i)}$$

$n(i)$ est le nombre d'individus n'ayant pas encore connu l'événement au temps juste antérieur à $t(i)$ (c'est le nombre d'individus pour lesquels l'événement pourrait surgir).

Deux statistiques, la variance et la moyenne, permettent de juger de la qualité de l'estimation.

La variance asymptotique de $\hat{G}(t)$ est calculée en utilisant la formule de GREENWOOD^[28] :

$$\text{Var}[\hat{G}(t)] = \sum_{i:t(i) < t} \frac{d(i)}{n(i)[n(i)-d(i)]} \cdot \hat{G}^2(t(i))$$

La moyenne de $\hat{G}(t)$ est égale à (avec $t_0 = 0$) :

$$E[\hat{G}(t)] = \sum_{i:t(i) < t} \hat{G}(t(i)) \cdot (t(i) - t(i-1)).$$

En calculant l'estimateur de KAPLAN-MEIER, on est sûr que la probabilité conditionnelle de prise d'un emploi, à chaque période $t(i)$, correspond exactement à la fréquence relative conditionnelle observée des échecs en $t(i)$, fréquence donnée par $d(i)/n(i) = \hat{r}(i)$. En d'autres termes, $\hat{r}(i)$ est la fonction de hasard. D'où :

$$\hat{G}(t) = \prod_{(i)} (1 - \hat{r}(i)).$$

L'estimateur de KAPLAN-MEIER est un estimateur, par le maximum de vraisemblance, de la fonction de survie, en l'absence d'hypothèse sur la forme fonctionnelle de cette fonction.

1.2 La mise en oeuvre de l'estimateur de KAPLAN-MEIER.

Nous avons utilisé le logiciel LIMDEP, créé par GREENE [1988]. Pour commencer, il est établie une table de survie basée

[28] La formule de GREENWOOD est :
 $\text{Var}[s(t)] = [S(t)]^2 \cdot \sum_{i < j} d_i / [n_i(n_i - d_i)]$.

sur la méthode de CUTLER et EDENER^[29]. Pour chaque intervalle, sont calculés les éléments suivants :

- le nombre d'observations, n_j
- la taille du domaine du risque^[30], $R_j = n_j - m_j/2$
où m_j est le nombre d'observations censurées ; ceci revient à considérer que les individus qui ne prennent pas d'emploi sont à risque pendant la moitié de l'intervalle.
- le nombre d'individus qui prennent un emploi (qui sortent du domaine du risque), d_j
- la proportion de sortants (d'individus qui prennent un emploi), $q_j = d_j/R_j = d_j/[n_j - 0.5 \cdot m_j] = 1 - p_j$
- la proportion d'individus qui prennent un emploi (c'est-à-dire la fonction de survie cumulée), $P_j = (1 - q_{j-1}) \cdot P_{j-1}$,
où $P_1 = 1$
- l'écart type pour l'estimation de P_j ^[31],

$$SE(P_j) = P_j \cdot \sum_{k=1}^{j-1} q_k / [(R_k \cdot (1 - q_k))]^{1/2}$$

- le taux de hasard cumulé, $r_j = 2 \cdot q_j / [1 \cdot (2 - q_j)]$, où 1 est la largeur de l'intervalle ; en conséquence, r_j est donné en 1^e d'unités ;
- l'écart type pour la fonction de hasard,
 $SE(h_j) = h_j [(1 - (1 \cdot h_j/2)^2) / (R_j \cdot q_j)]^{1/2}$.
- le hasard intégré = $-\text{Log}(P_j)$; c'est une sorte de résidu (LANCASTER [1985])
- la médiane.

En fait, la méthode de CUTLER et EDENER peut être considérée comme paramétrique, puisqu'elle fait intervenir, 1, la largeur de l'intervalle.

[29] L'intervalle de variation de la variable de durée est divisé en 10 intervalles égaux.

[30] Le nombre d'individus n'ayant pas encore pris d'emploi.

[31] La variance de P_j et celle de q_j sont respectivement :

$$\text{Var}(P_j) = P_j^2 \cdot \sum_{k=1}^{j-1} q_k / [(R_k \cdot (1 - q_k))] \text{ et } \text{Var}(q_j) = \text{Var}(p_j)$$

= $p_j q_j / R_j$. La connaissance de ces variances permet de juger de l'adéquation de l'ajustement.

1.3 Les limites de l'analyse non paramétrique

Ce sont, d'une part, celles inhérentes à la méthode de CUTLER et EDENER et d'autre part celles inhérentes à tous les estimateurs non paramétriques.

La méthode de CUTLER et EDENER ne raisonne pas sur la continuité du domaine du risque, mais raisonne sur dix intervalles égaux. Si cette façon de procéder n'altère pas profondément la forme générale de la courbe de survie, elle a, néanmoins, pour conséquence de placer l'estimation de la courbe de survie à un niveau différent de celui qu'elle aurait en présence d'un domaine du risque continu.

La courbe de survie estimée de $G(t)$ est un instrument descriptif. La représentation graphique de l'estimation de $\text{Log } G(t)$ est plus utile, car si la courbe obtenue est une droite, le taux de transition peut être considéré comme indépendant du temps. Dans le cas contraire, il est nécessaire de rechercher soit l'hétérogénéité au sein de la population, soit la dépendance temporelle du taux de transition. En fait, les graphiques du logarithme de la fonction de survie ne peuvent être que des instruments exploratoires, ils ne permettent pas de tests d'hypothèses, ni de mesure de l'effet des caractéristiques personnelles des individus sur la durée qu'ils passent dans le non emploi. C'est pourquoi il nous faut nous tourner vers des méthodes plus sophistiquées.

2. Une stratégie paramétrique permettant la prise en compte de la dépendance temporelle et de l'hétérogénéité au sein des sortants du système scolaire.

Précédemment nous avons fait deux hypothèses : d'abord, nous avons postulé l'absence de relation particulière dans la liaison du taux de transition au temps ; ensuite nous avons admis l'homogénéité de la population étudiée (ou de la sous-population étudiée) : en d'autres termes, nous avons supposé que le taux de transition est le même pour tous les individus. Or, si ce taux varie selon les caractéristiques de l'individu, X_i , et/ou selon le temps t où il se situe, il est important de pouvoir en tenir compte. En effet, la probabilité d'avoir un emploi n'est pas la même pour tous les individus. Elle est considérée notamment comme une fonction de la qualification.

Il existe deux stratégies, au moins, qui permettent de tenir compte de l'hétérogénéité^[32] de la population. Une première stratégie consiste à supposer que les paramètres fondamentaux du modèle sont fonction de variables explicatives observées et d'un terme d'erreur dont la distribution de probabilité est connue. C'est-à-dire que, si dans le modèle de base, on constate qu'il y a à la fois hétérogénéité de la population^[33] et dépendance temporelle, il peut être nécessaire d'introduire des variables explicatives dans le modèle. Une autre stratégie, complémentaire de la première, consiste à retenir les différences de nature dans les transitions étudiées ; ainsi, par exemple, une transition hors de l'inactivité vers un travail à temps partiel pourra être de nature différente d'une transition hors de l'inactivité vers un travail à temps plein. C'est l'analyse en termes de risques alternatifs. Ces deux stratégies sont complémentaires l'une de l'autre. Compte tenu de la faiblesse de nos effectifs, nous serons amenés à limiter nos analyses à la première stratégie.

La dépendance temporelle est prise en compte en donnant une forme fonctionnelle à la relation du taux de transition au temps, l'hétérogénéité par l'introduction de variables explicatives.

Les variables explicatives peuvent être stables au cours du temps (par exemple, le sexe) ou évoluer (comme le revenu). Du fait de la longueur limitée de notre période d'observation, nous ne nous intéressons pas aux variables évolutives.

Le modèle général s'écrit : $r(t ; X) \equiv r(t ; X, \beta)$ pour le taux de hasard, où $X = (X_1, \dots, X_p)$ est un vecteur de p variables explicatives, qualitatives ou quantitatives, et β un vecteur de paramètres inconnus que l'on cherche à estimer.

La question que l'on se pose est : 'quelle est la nature de la relation entre le(s) taux de transition et les variables explicatives ? Ou comment rendre compte de l'effet des caractéristiques de l'individu sur sa durée de séjour dans le non emploi. S'il y a plusieurs façons de mettre en relation taux de transition et variables explicatives^[34], nous nous limitons à introduire les variables explicatives dans le vecteur indépendamment de la dépendance temporelle.

[32] L'hétérogénéité est ici définie par les variables mesurées ou non qui diffèrent entre les individus et qui peuvent différer, pour un même individu, au cours du temps.

[33] Suite à une analyse par sous-population, par exemple.

[34] TRUSSEL et RICHARDS [1981] ont montré que les conclusions de l'analyse sont influencées par le type de dépendance au temps des variables introduites.

Les modèles (semi-) paramétriques autorisent une forme fonctionnelle (imposée ou non) à la relation du taux de transition au temps et, éventuellement, permettent de tenir compte de l'hétérogénéité des individus étudiés.

2.1 Le modèle de COX avec prise en compte de l'hétérogénéité : un modèle n'imposant pas de forme fonctionnelle à la relation du taux de transition au temps.

Le modèle de COX est semi-paramétrique parce qu'il suppose que le taux de hasard est égal à : $r_j(t;X) = r(0;X) \cdot \exp(\beta_j X_j)$, où $r(0;X)$ est une fonction de nuisance non spécifiée (dépendante avec certitude du temps et qui affecte tous les individus j de la même façon)^[35] et où $\exp(\beta_j X_j)$ est une fonction de densité particulière. Cette fonction pourrait aussi être toute autre fonction positive $g(X)$.

Dans l'échantillon total avec I événements en $t(i)$, la vraisemblance partielle sera^[36] :

$$L_j = \left[\begin{array}{c} \text{Probabilité} \\ \text{événement ait lieu} \\ \text{exactement en } t_i \end{array} \right] \cdot \left[\begin{array}{c} \text{Probabilité événement ait} \\ \text{lieu pour l'individu } j \\ \text{étant donné un événement} \end{array} \right]$$

Dans la forme $r(t;X) = r(0;X) \cdot g(X)$, l'effet des variables explicatives est d'agir de façon multiplicative, et indépendamment du temps, sur la fonction de hasard. Ce type de modèle est généralement appelé modèle de hasards proportionnels ou modèle à risques proportionnels. Sa caractéristique est qu'entre deux individus le taux de transition instantanée est indépendant du temps. Il peut être considéré comme un modèle non paramétrique, puisque la forme du taux de base $r(0;X)$ n'est pas spécifiée.

La mesure de l'influence d'une variable explicative sur la probabilité de connaître une transition se fait par le raisonnement suivant. La variation, du taux de hasard consécutive à une modification ΔX_j de la variable X_j , toutes choses égales par ailleurs s'écrit^[37] :

-
- [35] On appelle aussi $r(0;X)$, taux instantané sous-jacent ou taux de hasard de base en $t = 0$; c'est une fonction de degré zéro, c'est-à-dire que si $r(0;X)$ est constant, le modèle est dit exponentiel.
- [36] Car la vraisemblance partielle est plutôt construite pour les événements que pour les individus.
- [37] Cette équation, appliquée à un individu ayant les caractéristiques X , se réduit au calcul de l'élasticité quand $\Delta X_j = 1$. On a : $\epsilon_{r/X} = (\delta r(t;X)/\delta X)/(X/r(t;X))$ ou encore : $\epsilon_{r/X} = [\exp(\alpha_j \Delta X_j) - 1] \cdot X$

Les modèles (semi-) paramétriques autorisent une forme fonctionnelle (imposée ou non) à la relation du taux de transition au temps et, éventuellement, permettent de tenir compte de l'hétérogénéité des individus étudiés.

2.1 Le modèle de COX avec prise en compte de l'hétérogénéité : un modèle n'imposant pas de forme fonctionnelle à la relation du taux de transition au temps.

Le modèle de COX est semi-paramétrique parce qu'il suppose que le taux de hasard est égal à : $r_j(t;X) = r(0;X) \cdot \exp(\beta_j X_j)$, où $r(0;X)$ est une fonction de nuisance non spécifiée (dépendante avec certitude du temps et qui affecte tous les individus j de la même façon)^[35] et où $\exp(\beta_j X_j)$ est une fonction de densité particulière. Cette fonction pourrait aussi être toute autre fonction positive $g(X)$.

Dans l'échantillon total avec I événements en $t(i)$, la vraisemblance partielle sera^[36] :

$$L_j = \left[\begin{array}{l} \text{Probabilité} \\ \text{événement ait lieu} \\ \text{exactement en } t_i \end{array} \right] \cdot \left[\begin{array}{l} \text{Probabilité événement ait} \\ \text{lieu pour l'individu } j \\ \text{étant donné un événement} \end{array} \right]$$

Dans la forme $r(t;X) = r(0;X) \cdot g(X)$, l'effet des variables explicatives est d'agir de façon multiplicative, et indépendamment du temps, sur la fonction de hasard. Ce type de modèle est généralement appelé modèle de hasards proportionnels ou modèle à risques proportionnels. Sa caractéristique est qu'entre deux individus le taux de transition instantané est indépendant du temps. Il peut être considéré comme un modèle non paramétrique, puisque la forme du taux de base $r(0;X)$ n'est pas spécifiée.

La mesure de l'influence d'une variable explicative sur la probabilité de connaître une transition se fait par le raisonnement suivant. La variation, du taux de hasard consécutive à une modification ΔX_j de la variable X_j , toutes choses égales par ailleurs s'écrit^[37] :

[35] On appelle aussi $r(0;X)$, taux instantané sous-jacent ou taux de hasard de base en $t = 0$; c'est une fonction de degré zéro, c'est-à-dire que si $r(0;X)$ est constant, le modèle est dit exponentiel.

[36] Car la vraisemblance partielle est plutôt construite pour les événements que pour les individus.

[37] Cette équation, appliquée à un individu ayant les caractéristiques X , se réduit au calcul de l'élasticité quand $\Delta X_j = 1$. On a : $\epsilon_{r/X} = (\delta r(t;X)/\delta X)/(X/r(t;X))$ ou encore : $\epsilon_{r/X} = [\exp(\alpha_j \Delta X_j) - 1] \cdot X$

$$\begin{aligned}\Delta r(t;X) &= \frac{r(0;X)\exp(\alpha X + \alpha_j(X_j + \Delta X_j)) - r(0;X)\exp(\alpha X + \alpha_j X_j)}{r(0;X)\exp(\alpha X + \alpha_j X_j)} * 100\% \\ &= [\exp(\alpha_j \cdot \Delta X_j) - 1] * 100\%\end{aligned}$$

Si β est négatif, alors $\exp(\beta)$ est inférieur à l'unité et en conséquence chaque accroissement de la variable X_j diminue la probabilité de connaître une transition ; inversement si β est positif, chaque accroissement de la variable X augmente la probabilité de connaître l'événement. Si β est nul, alors la variable n'a pas d'effet.

2.2 Le modèle de Weibull avec prise en compte de l'hétérogénéité : un modèle paramétrique permettant de vérifier l'indépendance temporelle du taux de transition.

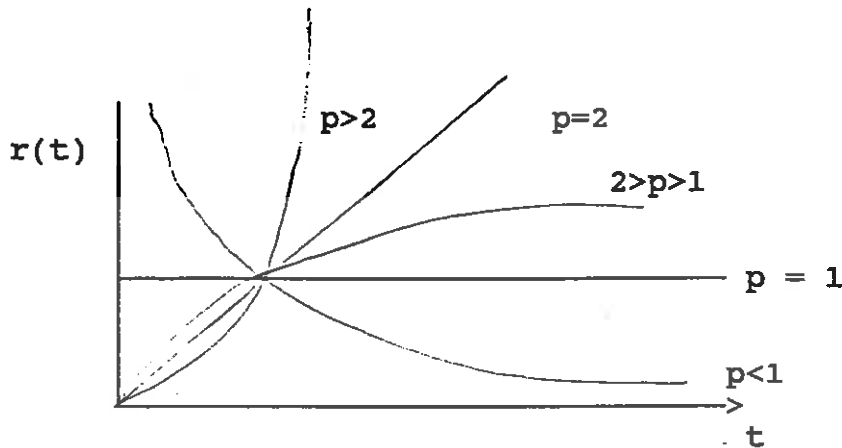
En alternative au modèle de COX, on peut spécifier une distribution particulière. D'emblée, certaines distributions sont inadaptées. C'est, par exemple, le cas de la distribution log-logistique car elle nécessite que la distribution des temps d'attente soit non défectueuse, c'est-à-dire que les probabilités de survie doivent tendre vers zéro. Or certains individus, plus vraisemblablement parmi les femmes, ne vont jamais prendre d'emploi. Dans ce cas la distribution des temps d'attente est dite défectueuse. En revanche, la distribution de Weibull n'est pas soumise à cette contrainte ; c'est pourquoi nous l'utilisons pour les analyses paramétriques.

Le hasard s'écrit, en présence de variables explicatives :

$$r(t;X) = \frac{p}{\exp(\alpha_0 + \alpha_i X_i)} \cdot \left[\frac{t}{\exp(\alpha_0 + \alpha_i X_i)} \right]^{p-1} \quad [38]$$

$r(t)$ est monotone décroissant pour $p < 1$ signalant une dépendance de durée négative. Dans un tel cas, plus un individu aura été observé dans l'état de non emploi, moins il aura de chance de prendre un emploi. Le hasard est croissant pour $p > 1$, ce qui indique une dépendance de durée positive. Ces différents cas se représentent graphiquement comme suit :

[38] si $p = 1$, alors on retrouve la distribution exponentielle, et donc l'absence de dépendance de durée.



Le modèle de Weibull peut aussi être considéré comme un modèle de hasard proportionnel puisqu'entre deux individus le rapport des taux de hasard est indépendant du temps. En effet si l'on considère deux individus, l'un ayant X_1 pour vecteur de variables explicatives, l'autre ayant X_2 et si l'on forme le rapport des taux de hasard entre ces deux individus, on obtient :

$$\frac{r(t, X_1)}{r(t, X_2)} = \exp[\alpha \cdot (X_1 - X_2)], \text{ qui est bien indépendant du temps } t.$$

Le temps moyen d'attente de l'événement est donné par la formule :

$$E(t) = \Gamma [(1 + p)/p] / r, \text{ où } \Gamma[.] \text{ est la fonction gamma.}$$

La variance est :

$$\text{Var}(t) = [\Gamma [(1 + p)/p] - \Gamma [(1 + p)/p]^2]^2 / r^2$$

Comme dans le modèle de COX, on peut étudier l'effet d'un changement de X_j en tenant constant l'ensemble des autres X . Cet effet montre de combien de pourcent le taux de hasard se modifie lors d'une variation ΔX_j de X_j .

Si $r(t; X) = p \cdot \exp(\alpha_0 + \alpha \cdot X + \alpha_j \cdot X_j)$, alors la variation du taux suite à une modification ΔX_j de X_j s'écrit^[39] :

$$\exp(+\alpha_j \cdot \Delta X_j p) - 1, \text{ où } \alpha_j = -\beta_j.$$

Quand on multiplie cette expression par 100, on obtient le pourcentage de changement de $r(t; X)$ accompagnant une variation de X_j .

[39] La démonstration de ce résultat est semblable à celle du modèle de COX.

On peut de plus connaître directement comment se modifie la durée moyenne, suite à une variation, ceteris paribus, d'une variable X_j . La variation de la durée moyenne s'écrit :

$$[E(t ; X, X_j + \Delta X_j) - E(t ; X, X_j)] / E(t ; X, X_j)$$

$$= [\exp(-\alpha_j \Delta X_j) - 1] * 100 \%, \text{ où } \alpha = -\beta$$

Nous achevons ici la présentation des modèles de régression que nous allons employer. Nous nous proposons maintenant d'examiner les tests que nous allons mettre en oeuvre pour vérifier l'adéquation de nos modèles de régression.

3) La mesure de l'adéquation des méthodes d'analyse utilisées

Les méthodes d'analyse présentées ci-dessus sont basées sur la maximisation de la vraisemblance ; dans ce contexte, la fiabilité des estimations peut être établie en regardant plusieurs statistiques. Certaines sont plus spécifiques à l'analyse en logit, d'autres sont davantage appropriées à l'analyse événementielle ; enfin une troisième catégorie s'emploie dans les deux types d'analyse.

A - Les tests communs à l'analyse de régressions logistiques et à l'analyse événementielle.

1. La première statistique que nous utilisons est celle du rapport de vraisemblance (RV).

Partant de E_0 , la vraisemblance pour le modèle avec k contraintes sur les paramètres (c'est l'hypothèse nulle) et de E_1 , celle pour le modèle sans contrainte sur les paramètres, on calcule : $RV = \max(E_0) / \max(E_1)$. Ce test est connu comme le test (la statistique) du ratio de vraisemblance. Il vérifie que $c = -2 \text{ Log}(RV)$ suit approximativement une distribution du χ^2 avec k degrés de liberté (il y a autant de degrés de liberté qu'il y a de paramètres) et teste l'hypothèse que tous les coefficients, sauf la constante, sont nuls. C'est une statistique semblable au F de la régression par les moindres carrés.

2. Par ailleurs, un pseudo R^2 peut être calculé.

Il est égal à $c / (N + c)$, où N représente la taille de l'échantillon, et c la statistique précédente du rapport de vraisem-

blance. Cette mesure s'approche de 1 quand la qualité de l'estimation s'améliore.

3. Le R^2 de MAC FADDEN, troisième statistique que nous employons, est une autre possibilité d'appréhension de la qualité de l'estimation. Il s'écrit :

$$R^2_{MF} = 1 - \frac{-2 \text{ LOG}(RV)}{-2 \text{ LOG}(0)} .$$

Il correspond au pourcentage du logarithme de la vraisemblance qui est gagné grâce à l'introduction de variables explicatives. La qualité de la régression est d'autant meilleure que ce test s'approche de l'unité.

B - Les tests spécifiques à l'analyse de régression logistique.

Une mesure de la qualité de l'estimation spécifique à l'analyse de régressions logistiques est basée sur la table de classification. Cette table compare les probabilités prédites et les probabilités réelles. Elle peut être assez arbitraire car les probabilités prédites sont continues. Plusieurs statistiques permettent de mesurer le pouvoir de prédiction du modèle, qui confirment une bonne qualité du modèle quand elles s'approchent de l'unité. Nous utilisons les suivantes :

1. La sensibilité de l'estimation est la proportion de sortants prenant un emploi que la méthode d'analyse détecte. En abrégé la sensibilité est la proportion de vrais positifs qui sont prédits positifs : c'est le nombre de personnes qui prennent effectivement un emploi, pour lesquelles les résultats de l'estimation ont prédit une prise d'emploi, divisé par le nombre de personnes pour lesquelles les résultats sont positifs.

2. La spécificité de l'estimation est la proportion de sortants ne prenant pas d'emploi que la méthode d'analyse classe parmi ceux ne prenant pas d'emploi. En abrégé la spécificité est la proportion de vrais négatifs qui sont prédits négatifs.

3. Le taux de faux positifs est la proportion de prédictions positives qui sont en fait négatives.

4. Le taux de faux négatifs est la proportion de prédictions négatives qui sont en fait positives.

Ces quatre tests sont calculés à partir du schéma suivant :

		prédictions		
		0	1	
observations	0	VN	FP	d
	1	FN	VP	c
		b	a	N

La signification des abréviations est la suivante : VP : vrais positifs ; FP : faux positifs ; FN : faux négatifs ; VN : vrais négatifs. Les définitions précédentes peuvent donc s'écrire formellement comme suit :

- sensibilité = $VP / (VP + FN) = VP / c$
- spécificité = $VN / (FP + VN) = VN / d$
- taux de faux positif = $FP / (VP + FP) = FP / a$
- taux de faux négatif = $FN / (FN + VN) = FN / b$

A ces quatre mesures on peut ajouter une mesure du risque qui se définit comme : $(VP + FN) / N$.

C - Le test du log-rank. Un test que nous utilisons spécifiquement dans l'analyse événementielle.

On peut tester l'homogénéité de la population au regard du taux de hasard grâce à plusieurs tests. Nous nous limitons au test du log-rank (appelé encore test de MANTEL-HAENSZEL). Ce test est semblable au test de WILCOXON aux pondérations près accordées aux différentes durées. Il donne des valeurs d'autant plus élevées que les distributions spécifiques à chaque sous-population diffèrent ; un test du Chi² permet de dire si l'écart peut être considéré comme significatif.

Le test du log-rank s'applique notamment quand la population totale est divisée en sous-populations. Dans ce cas, il peut être intéressant de vérifier si les fonctions de survie des populations se recouvrent ou se différencient de façon significative.

Les méthodes de comparaisons des courbes de survie entre elles ont été abondamment développées dans la littérature. Nous ne les reprendrons pas ici en détail. Nous nous limitons au cas simple de deux groupes à comparer et à l'approche de principe dans l'application du test de MANTEL-HAENSZEL.

Supposons que l'on veuille tester l'égalité des courbes de survie $G_A(t)$ et $G_B(t)$ pour deux populations p_A et p_B . Soient $t_{(1)} < t_{(2)} < \dots < t_{(n)}$ les temps d'échecs pour l'échantillon de taille n formé par les deux populations. En $t_{(j)}$, on a m_j échecs observés et n_j sujets à risque juste avant $t_{(j)}$ ($j = 1, 2, \dots, 7$). Pour l'ensemble des deux populations on a m_1 échecs.

Pour tester l'hypothèse nulle $H_0 : G_A(t) = G_B(t)$, on utilise la statistique suivante qui repose sur la différence entre le nombre de "morts" observées dans chaque population et une quantité qui peut être considérée comme le nombre moyen d'échecs sous l'hypothèse nulle :

$$X^2 = \frac{n \cdot (ad - bc)^2}{n_A \cdot n_B \cdot m_1 \cdot m_2} = \frac{n}{n-1} \cdot \left[\frac{a - E_0(a)}{[\text{Var}_0(a)]^{1/2}} \right]^2$$

où a = nombre de "morts" dans p_A ,
 b = nombre de "vivants" dans p_A ,
 c = nombre de "morts" dans p_B ,
 d = nombre de "vivants" dans p_B ,
 m_1 = nombre total de "morts" : $a + c$,
 m_2 = nombre total de "vivants" : $b + d$,
 $E_0(a) = n_A \cdot m_1 / n$
 et $\text{Var}_0(a) = \frac{[n_A \cdot n_B \cdot m_1 \cdot m_2]}{[n^2(n-1)]}$
 $= \frac{[n_A \cdot (n - n_A) \cdot m_1 \cdot (m - m_1)]}{[n^2(n-1)]}$

Pour K événements ordonnés, la statistique de MANTEL-HAENSZEL est appliquée K fois au domaine du risque. Dans sa forme générale, le test s'écrit :

$$MH = \frac{\sum_{i=1}^K w_i \cdot [a_i - E_0(a_i)]}{[\sum_{i=1}^K w_i \cdot \text{Var}_0(a_i)]^{1/2}}$$

Dans la statistique de MANTEL-HAENSZEL, le facteur de pondération w_i est égal à 1.

Cette section nous a permis d'entrevoir plusieurs façons de mesurer l'effet des caractéristiques individuelles sur la probabilité de connaître prendre un emploi à un moment donné et instantanément. Nous avons par ailleurs exposé les statistiques permettant de juger de la qualité des estimations et notamment de comparer les courbes de survie entre elles.

Il ne faut pas oublier que l'utilisation des modèles probabilistes dynamiques repose sur des hypothèses fortes, notamment pour les modèles concernant la distribution des taux de transition instantanés. Les modèles semi-paramétriques introduits par COX en 1972 permettent de lever ces hypothèses fortes en n'imposant pas de formalisation à l'effet de durée.

Concernant les tests d'adéquation, il nous faut noter qu'aucun n'est universellement reconnu comme mesurant la qualité de l'estimation. C'est leur convergence que nous utilisons comme critère de qualité.

Nous nous penchons maintenant sur les résultats de nos analyses.

III - LES RESULTATS

Notre présentation est la suivante ; nous commençons par décrire la population des jeunes sortants du système scolaire au regard d'un certain nombre de critères socio-démo-économiques et environnementaux. Ensuite, nous présentons les résultats des analyses effectuées pour mesurer l'influence de ces critères sur la durée du non emploi.

1) Les caractéristiques des sortants du système scolaire résidents au Grand Duché de Luxembourg.

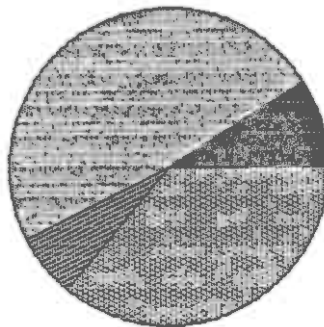
Il s'agit de présenter pour la population résidant au Grand Duché les caractéristiques des sortants du système scolaire. Un premier point à noter est que ces jeunes n'ont pas forcément effectué toute leur scolarité au Grand Duché. C'est notamment le cas des diplômés de l'enseignement supérieur : le pays n'offrant pas toutes les formations possibles. C'est aussi le cas d'un certain nombre de jeunes inscrits dans des écoles étrangères. Cependant devant la faiblesse des effectifs concernés, nous ne sommes pas à même de vérifier si les durées de non emploi de ces derniers sont différentes des durées de non emploi des sortants du système scolaire luxembourgeois.

Nous nous proposons de comparer deux populations : ceux qui ont trouvé un emploi et ceux qui n'en ont pas trouvé sur la période d'observation, période qui se situe de septembre 1985, date de fin de scolarité, à mars 1986, date de fin d'observation. La durée de non emploi est donc au maximum de huit mois.

- 66 jeunes sur les 78 recensés (soit 85 %) vont trouver un emploi sur la période d'étude. Pour la moitié environ, il y a un passage direct de l'école à l'emploi : 29 ont une durée de non emploi nulle. Ceci explique le faible pourcentage de "chômeurs" (= a recherché un emploi pendant au moins un mois) parmi ces 66 jeunes : seuls 9 % ont déclaré avoir recherché un emploi. A contrario, parmi ceux qui n'ont pas trouvé d'emploi sur la période d'observation, 8 sur 10 ont déclaré en rechercher un pendant au moins un mois ; ce qui signifie que 2 jeunes sans emploi sur dix ne sont pas en fait des chômeurs au sens du BIT : ils n'ont pas d'emploi mais ils n'en recherchent pas. Il nous faut noter que ces commentaires comme le seront les suivants sont à prendre avec précaution : les effectifs concernés étant très faibles, nous ne pouvons être sûrs que les différences constatées sont significatives.

On peut, a priori, supposer que ce sont surtout les élèves de sexe féminin qui ne recherchent pas d'emploi à la sortie du système scolaire.

graphique 1 : Sexe et indicateur d'emploi
(sans = sans emploi ; emploi = avec emploi)



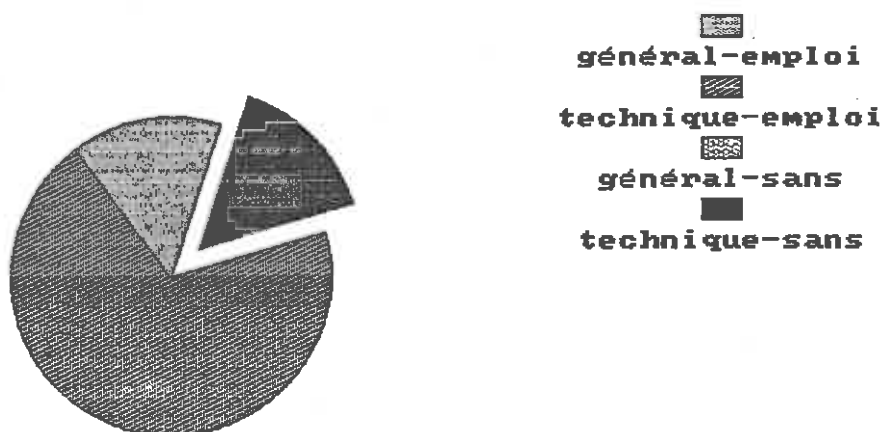
 homme-emploi
 homme-sans
 femme-emploi
 femme-sans

- Si on observe effectivement que les jeunes femmes sont davantage représentées parmi les personnes qui ne trouvent pas d'emploi (elles représentent 58 % de ceux qui ne trouvent pas d'emploi alors qu'elles ne représentent que 45 % de la population), on n'observe pas de différence significative quand on compare la population selon le sexe et selon le fait d'avoir cherché ou non un emploi (après avoir retiré les individus qui ont une durée de non emploi nulle).

- Le type d'enseignement oppose ceux qui ont suivi un enseignement général (études classiques ou modernes en secondaire général ou enseignement supérieur universitaire) aux autres (études en secondaire technique, en complémentaire, ou en enseignement supérieur non universitaire). Les résultats vont à l'encontre de nos attentes : ce n'est pas l'enseignement technique qui apparaît comme favorisant l'insertion professionnelle ; en effet tous les individus ayant suivi un enseignement de type général vont trouver un emploi sur la période d'observation ; ce qui n'est pas le cas de deux individus sur dix ayant suivi un enseignement de type technique. Deux explications s'offrent à nous : la première repose sur la grande variété des formations classées en technique, et il apparaît en fait que les individus ayant le plus de difficultés à trouver un emploi sont davantage

représentés parmi ceux ayant suivi un enseignement complémentaire ou les cours d'orientation et d'initiation professionnelle (COIP). Une autre explication est liée à la tertiariation de l'économie luxembourgeoise, ce qui valorise davantage les formations de type général car plus souples à adapter que les formations spécifiques.

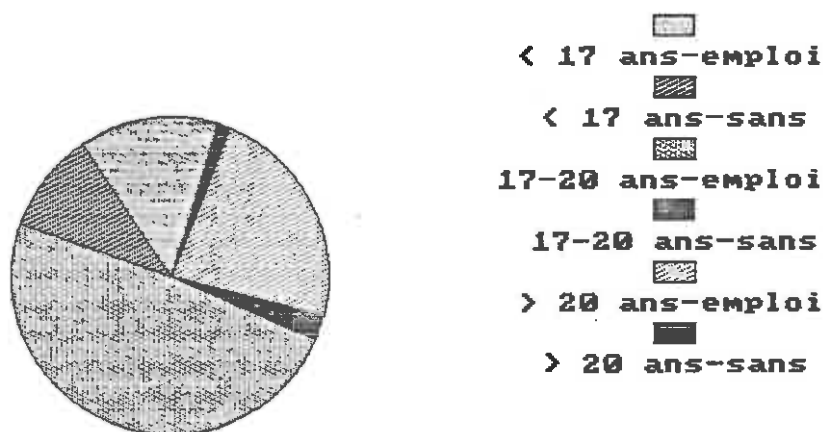
Graphique 2 : Type de formation suivie et indicateur d'emploi.



- Quand on regarde le nombre d'années de scolarité, en d'autres termes le niveau scolaire atteint, les résultats sont les suivants : ceux qui ne trouvent pas d'emploi atteignent en moyenne la 8e année (soit l'équivalent de deux années de secondaire) alors que les autres arrivent, en moyenne, en 12e année (soit l'équivalent de six années de lycée).

- L'âge, l'année de la sortie du système scolaire, est en conformité avec ces deux résultats : ceux qui ne trouvent pas d'emploi sont en moyenne trois ans plus jeunes que les autres (16 ans contre 19 ans)

Graphique 3 : Age et indicateur d'emploi



La comparaison des variables âge et niveau scolaire atteint est intéressante ; elle indique que ceux qui n'ont pas trouvé d'emploi ont accumulé un retard dans leur carrière scolaire beaucoup plus important que les autres : on est habituellement en 8e année à 13 ans, ce qui se traduit par trois ans de retard, en moyenne, pour ceux qui n'ont pas encore d'emploi au bout des huit mois étudiés. Ce retard n'est en moyenne que d'une année pour les autres jeunes.

- La seule variable d'environnement que nous possédons est la région d'habitation du jeune sortant du système scolaire : si les habitants du centre du pays et de la région d'Esch sur Alzette (sud) trouvent plus facilement un emploi, ce sont ceux des régions davantage agricoles (le nord et l'est) qui ont les proportions d'individus en non emploi les plus fortes.

- Les autres variables à notre disposition sont des variables concernant la famille de l'élève. On constate que :

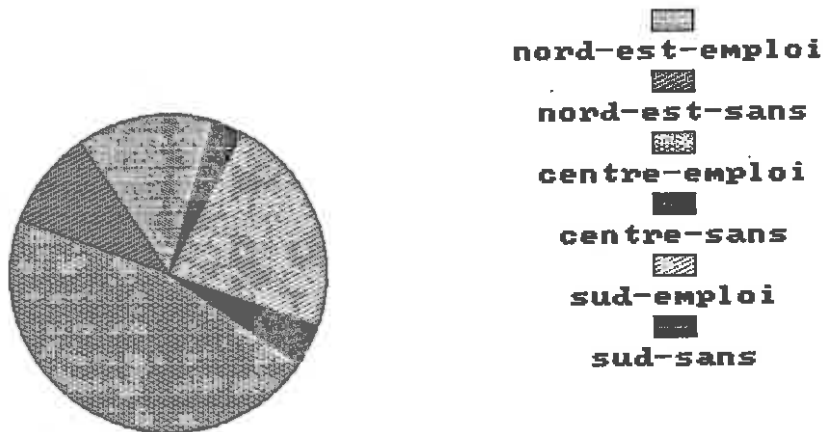
* la probabilité de trouver un emploi est d'autant plus forte que le quotient familial (revenu du ménage pondéré par la taille de ce dernier) est élevé (ce raisonnement n'est plus valable quand on s'intéresse au revenu total non pondéré) ;

* c'est dans les ménages où le chef a un niveau de formation faible que l'enfant a le moins de chance de trouver un emploi ;

* ce sont les ménages qui ont le plus de personnes (le plus d'adultes et le plus d'enfants) et le plus d'actifs que les chances de trouver un emploi sont les plus faibles.

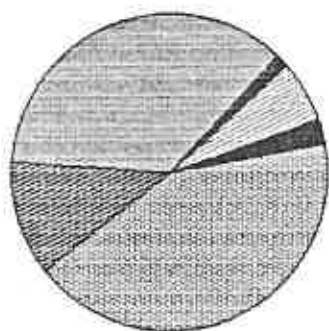
Cependant l'on doit noter que, de par la définition des concepts d'enfants et d'adultes dans le PSELL, l'élève dont on analyse la transition de l'école à l'emploi, est tantôt classé parmi les enfants (quand il a moins de 18 ans), tantôt parmi les adultes (quand il a plus de 18 ans). En revanche on ne remarque pas de différence de comportement selon le degré d'intégration du ménage dans le pays^[40].

Graphique 4 : Région d'origine et indicateur d'emploi



[40] cf TOURNOIS [1990]

graphique 5 : Niveau de formation du chef de ménage et indicateur d'emploi de l'enfant



■ inférieur-emploi
 ▨ inférieur-sans
 ▩ moyen-emploi
 ■ moyen-sans
 ▨ supérieur-emploi
 ■ supérieur-sans

- 2) - L'impact des caractéristiques socio-démo-économiques des sortants du système scolaire sur le fait de trouver ou non un emploi.

Nous reprenons ici la démarche exposée dans les aspects méthodologiques (chapitre précédent). Les résultats de l'analyse de la probabilité de trouver un emploi sont d'abord présentés (A) suivis par ceux de la probabilité instantanée de trouver un emploi (B).

A - La mesure de la probabilité de trouver un emploi : l'utilisation d'un modèle logistique.

Trois régressions logistiques sont présentées qui mesurent la probabilité pour un sortant du système scolaire d'occuper un emploi dans les huit mois qui suivent la fin de scolarité.

Dans la première régression (tableau A.1) un certain nombre d'indicateurs sont entrés pour mesurer cette probabilité. Ce sont des variables de contexte et des caractéristiques propres à l'individu. On supposait, a priori, que ces variables avaient une influence sur le fait de trouver ou non un emploi.

Parmi les variables de contexte^[41], on trouve :

- la formation du chef de ménage, CHEFINF ; dans cette variable, sont opposés les chefs n'ayant pas dépassé la 6^e année (valeur 1 de la variable) aux autres chefs de ménage (valeur 0) ;
- le nombre d'enfants à charge dans le ménage du sortant, NBENF ; le sortant est lui-même considéré comme enfant à charge ;
- un indicateur de l'hétérogénéité de l'intégration des membres du ménage, SD85 ;
- le nombre de personnes ayant un emploi dans le ménage, NBEMP ;
- le quotient familial du ménage calculé avec la pondération C.E.S., QF1.

A ces cinq variables ont été rajoutées des caractéristiques propres à l'individu sortant du système scolaire :

- le sexe, SEXE ; la variable prend la valeur 1 pour les hommes et la valeur 0 pour les femmes ;
- le fait d'avoir été (valeur 1) ou non (valeur 0) apprenti avant de quitter l'école, APPRENTI ;
- l'âge, AGE ; cette variable est aussi introduite au carré, AGE2 et au cube, AGE3^[42] : des analyses préliminaires ont montré que la probabilité de trouver un emploi ne variait pas de façon monotone avec l'âge.

Enfin, la dernière variable introduite est la durée du non emploi, DUR1, calculée selon le modèle 1 défini dans le premier chapitre. Cette durée est connue avec incertitude pour les individus n'ayant pas trouvé d'emploi.

Les résultats font apparaître qu'aucun indicateur ne révèle avoir un quelconque pouvoir explicatif de la probabilité d'occuper un emploi à la sortie du système scolaire. La raison en est l'existence de solutions multiples au problème d'optimisation.

[41] Une définition plus précise des variables se trouve au chapitre I.

[42] La variable AGE2 a été divisée par 10 et la variable AGE3 par 100 afin de rendre la lecture de leur coefficient plus accessible.

Tableau A.1

Maximum Likelihood Estimates						
	Log-Likelihood.....			-0.27482E-06		
	Restricted (Slopes=0) Log-L.			-33.487		
	Chi-Squared (10).....			66.974		
	Significance Level.....			0.27217E-12		
	R ² MF.....			0.00000		
Variable	Coeffic.	Std.Er.	T-ratio	Prob:t>x	Mean X	Std. X
DUR1	-19.9718	2412.66	-0.008	0.99340	2.5513	3.0179
AGE	20.1519	3428.92	0.006	0.99531	18.949	3.7827
AGE2	-7.34376	2162.23	-0.003	0.99729	37.318	17.525
AGE3	4.95615	3974.73	0.001	0.99901	7.7141	6.7170
SEXE	16.0850	7272.62	0.002	0.99824	1.4487	0.50058
APPRENTI	16.6854	8323.21	0.002	0.99840	0.2308	0.42405
CHEFINF	-26.0199	8604.60	-0.003	0.99759	0.4744	0.50257
NBENF	-11.8220	2742.43	-0.004	0.99656	1.1667	1.4180
SD85	-1.11243	308.733	-0.004	0.99713	9.6410	13.742
QF1	0.003460	0.97346	0.004	0.99716	21310.	6880.4
NBEMP	-33.8017	5670.87	-0.006	0.99524	1.6667	1.0403

Un certain nombre de régressions alternatives ont été testées qui montrent des résultats semblables. Dans ces régressions alternatives, d'autres variables ont été introduites qui ne présentaient pas non plus de caractère significatif (par exemple, le lieu de résidence) ou qui ne pouvaient l'être dans celle présentée dans le tableau 1 en raison des corrélations sous-jacentes : il en est ainsi du niveau scolaire atteint qui est corrélé à l'âge de fin d'études ($r = 0,733$). Nous ne présentons pas les résultats de la mise en oeuvre de ces régressions ici mais le lecteur intéressé peut les obtenir en s'adressant aux auteurs.

Devant la mauvaise qualité des résultats obtenus en introduisant les variables influençant, à notre avis^[43], le fait de trouver ou non un emploi, nous avons été amenés à limiter le nombre de variables explicatives. Nous souhaitons garder la variable DUR1 afin de vérifier une idée que nous avions. En effet, si l'on peut penser qu'au fur et à mesure que le temps s'écoule dans le non emploi, un individu a de moins en moins de chance d'occuper un emploi (effet négatif de DUR1 sur la probabilité de trouver un emploi), peut-on penser que le fait

[43] Et à celui des membres de notre équipe qui travaillent sur le projet PETRA.

d'occuper ou non un emploi est retracé par la durée du non emploi (effet négatif de la variable "a trouvé un emploi" sur la durée du non emploi). En d'autres termes, peut-on dire que la variable DUR1 et la variable "a trouve un emploi" sont endogènes l'une à l'autre.

Tableau A.2

Maximum Likelihood Estimates						
Log-Likelihood.....						-8.3640
Restricted (Slopes=0) Log-L.						-33.487
Chi-Squared (3).....						50.246
Significance Level.....						0.32173E-13
R ² MF.....						0.7502315
Variable	Coeffic.	Std. Er.	T-ratio	Prob:t>x	Mean X	Std. X
DUR1	-1.44640	0.532480	-2.716	0.00660	2.5513	3.0179
AGE	0.490610	0.528843	0.928	0.35356	18.949	3.7827
AGE2	0.199639	0.379654	0.526	0.59900	37.318	17.525
AGE3	-0.822002	0.748553	-1.098	0.27215	7.7141	6.7170
Frequencies of actual & predicted outcomes						
Predicted outcome has maximum probability.						
	Predicted					
Actual	TOTAL	0	1			spécificité : 92 %
Total	78	12	66			sensibilité : 98 %
0	12	11	1			
1	66	1	65			

La mise en oeuvre de ce test est retracée dans les régressions présentées dans les tableaux A.2 et A.3.

Dans la deuxième régression (tableau A.2), seules deux variables sont entrées : la durée du non emploi et l'âge de fin de scolarité, dans la forme présentée ci-dessus. Les résultats font apparaître que la durée du non emploi est négativement liée à la probabilité d'occuper un emploi : plus la durée du non emploi est longue, moins les individus sont susceptibles d'occuper un emploi. Si ceci est le résultat attendu, il n'en reste pas moins qu'on peut mettre en doute la pertinence de l'introduction de cette variable : son haut degré de significativité confirme son endogénéité potentielle, de même que l'amélioration de la vrai-

semblance qui résulte de son introduction ($R^2_{MF} = + 75 \%$, cf chapitre 2).

Tableau A.3

Maximum Likelihood Estimates						
Log-Likelihood.....						-27.438
Restricted (Slopes=0) Log-L.						-33.487
Chi-Squared (2).....						12.099
Significance Level.....						0.23588E-02
R^2_{MF}						0.1806373
Variable	Coeffic.	Std.Error	T-ratio	Prob:t>x	Mean X	Std.X
AGE	-0.680130	0.301124	-2.259	0.02391	18.949	3.7827
AGE2	0.664640	0.259495	2.561	0.01043	37.318	17.525
AGE3	-1.25629	0.487462	-2.577	0.00996	7.7141	6.7170
Frequencies of actual & predicted outcomes Predicted outcome has maximum probability.						
		Predicted				
Actual	TOTAL	0	1			spécificité : 100 %
Total	78	0	78			sensibilité : 0 %
0	12	0	12			
1	66	0	66			

C'est pourquoi, dans la troisième régression (tableau A.3), nous avons retiré cette variable de l'analyse : les résultats font apparaître un haut degré de significativité de la variable âge. De plus le pouvoir explicatif du modèle est assez bon : cette variable contribue à 18 % de l'amélioration de la vraisemblance. Cependant, si le modèle classe bien les individus prenant un emploi dans la catégorie des individus avec emploi (l'indicateur de spécificité est égal à 100 %), il n'en est pas de même des individus sans emploi qui sont tous prédit comme prenant un emploi (indicateur de sensibilité égal à 0%).

Somme toute, si l'on peut dire que l'âge de fin d'études a un certain pouvoir explicatif de la probabilité de trouver un

emploi^[44], les résultats doivent néanmoins être pris avec précaution. En effet le modèle n'est guère stable : de l'absence de significativité des variables explicatives mises en oeuvre dans la première régression, absence due à l'existence de solutions multiples, on passe à un modèle très significatif dans la seconde régression ($R^2_{MF} = 0.75$), alors que l'on a juste enlevé certaines variables, pour finir avec un modèle expliquant 18 % du phénomène de la prise d'emploi. Cette sensibilité du modèle aux variables introduites résulte, à n'en pas douter, de la faiblesse numérique de l'échantillon, faiblesse qui se traduit par une absence de diversité des situations rencontrées.

C'est pourquoi, nous prolongeons l'analyse en nous tournant vers d'autres méthodes qui permettent de palier certaines difficultés liées à l'approche ci-dessus, notamment en mettant l'accent sur la durée du non emploi et en tenant compte du fait que cette durée peut être connue avec incertitude.

B - La mesure de la probabilité instantanée de trouver un emploi : l'utilisation de modèles de hasard.

La probabilité instantanée de trouver un emploi trouve son corollaire dans la durée du non emploi (cf chapitre 2), objet de notre étude.

Pour essayer de cerner les déterminants de la durée de non emploi des sortants du système scolaire, nous utilisons des données mensuelles : ceci favorise une analyse plus fine et introduit une plus grande variance parmi les individus étudiés. Cependant le nombre de durées différentes est relativement faible par rapport au nombre de durées exactes recensées^[45] (10 %). Un avantage de ces données fines est de s'intégrer assez bien dans la réalité du marché du travail : les individus sont en général, lors de leur premier emploi, embauchés sur une base mensuelle (contrats de 3 mois, etc.) et payés sur une base mensuelle.

Pour préparer l'analyse de l'impact des caractéristiques personnelles et familiales sur la durée du non emploi, nous exécutons une analyse préliminaire par la méthode actuarielle de KAPLAN-MEIER^[46]. Nous complétons cette approche en nous livrant

[44] Des résultats semblables sont trouvés quand on introduit le niveau scolaire atteint.

[45] C'est-à-dire compte non tenu des durées censurées.

[46] Cette méthode suppose que le taux de hasard est le même pour tous les individus, mais qu'il varie selon la période.

à une analyse non paramétrique par sous-population : ceci nous permet d'appréhender le degré d'homogénéité de la population au regard de la durée du non emploi. Ensuite, nous avons recours à une analyse semi paramétrique et à une analyse paramétrique de la relation du taux de hasard au temps.

a) Une méthode non paramétrique d'analyse de la transition de l'école à l'emploi.

Deux types d'analyses sont réalisées. Dans la première, on examine l'influence du mode de calcul de la durée du non emploi. Dans la seconde, on tient compte des caractéristiques individuelles (en d'autres termes, l'hétérogénéité) sur la durée de non emploi.

1) Influence du mode de calcul de la durée de non emploi.

Dans le chapitre 1, nous avons introduit l'existence de deux modèles. Dans le premier modèle les caractéristiques individuelles ayant trait au système scolaire ont été corrigées de façon à rendre compte de la réalité de ce système. Dans le second modèle, aucune correction n'est faite.

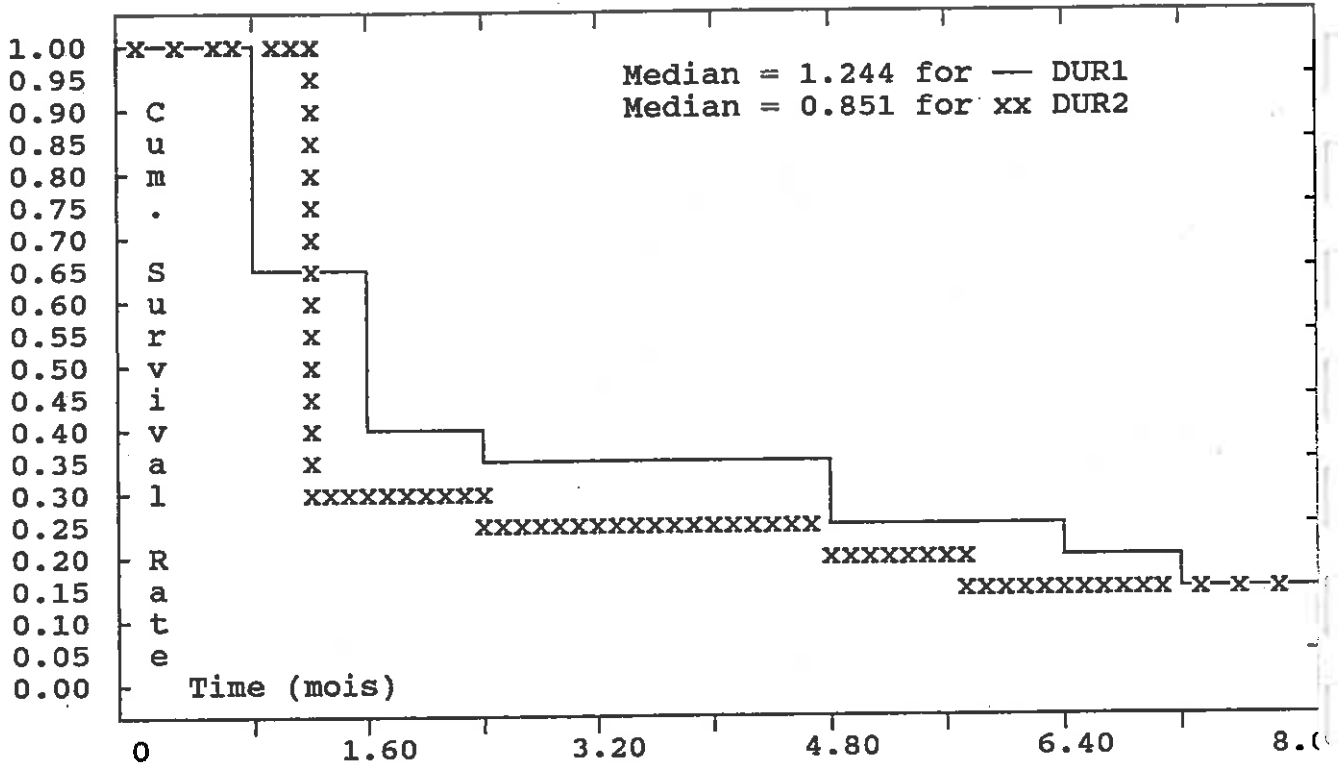
Nous avons alors tenté d'apprécier la nécessité de nos corrections. Si les deux analyses fournissent des résultats semblables, on peut en conclure à l'inutilité de modifier les réponses individuelles. Le graphique B.1.1 ci-dessous retrace l'évolution temporelle de la courbe de survie de chaque modèle.

D'emblée, il apparaît que nos corrections ne semblent pas inutiles, puisque les deux courbes de survie ne se superposent pas. On peut voir l'impact de nos modifications en comparant la position relative des deux courbes de survie ; on constate que nos corrections ont surtout eu pour effet de rallonger la période de non emploi : la courbe de DUR1 est, le plus souvent, supérieure à celle de DUR2, indiquant que dans le modèle 1 les individus sortent moins rapidement du non emploi.

Tableau B.1.1

Estimated Survival Function		
Duration variable is Status is given by variable	DUR1 CENS1	DUR2 CENS2
Number of Observations	78	78
- Exiting	66	67
- Censored	12	11

Graphique B.1.1 Durée du non emploi selon les deux modèles :
Représentation des courbes de survie estimée.



Les résultats présentés dorénavant ne mettent l'accent que sur le premier modèle. La variable dépendante est DUR1 et l'indicateur de censure est CENS1. Cet indicateur prend la valeur 1 quand la durée du non emploi est connue avec certitude.

2) L'analyse par sous-population permet de tester le degré d'homogénéité de la population selon la durée du non emploi^[47].

Nous avons partagé la population étudiée selon de nombreux critères. Seules deux variables se sont révélées significatives, en ce sens qu'elles indiquent que la population étudiée est hétérogène^[48]. Ce sont : l'âge de fin d'études (tableaux et

[47] Dans cette section, deux individus ont été retirés des analyses car présentant une non réponse sur l'une des variables utilisées.

[48] Les critères non significatifs sont notamment la localisation géographique de l'étudiant, le type d'études suivies, le niveau de vie de la famille, le degré d'intégration de la famille, le nombre d'adultes au foyer.

graphiques B.2.1 et B.2.2) et le niveau scolaire atteint (tableau et graphique B.2.3). Ces deux variables sont d'ailleurs fortement corrélées entre elles ($r = 0.7325$).

Un premier résultat est le suivant : les tests du log rank et de WILCOXON attestent que la durée du non emploi est liée à chacune de ces deux variables. Nous présentons leurs courbes de survie estimées, le résultat des tests et le partitionnement de la population selon que les observations sont ou non censurées et selon le critère de sélection de la population.

2.1 Durée du non emploi et âge de fin d'études.

L'âge de fin d'études a été partagé en trois classes correspondant à des situations particulières :

- ≤ 16 ans ; niveau inférieur,
- 17 - 20 ans ; niveau moyen,
- > 20 ans ; niveau supérieur.

Par rapport à l'âge de fin d'études, la prise d'emploi évolue dans le même sens : elle est d'autant plus rapide que l'âge est élevé. Ainsi, dans le graphique B.2.1, la courbe la plus basse s'applique aux individus les plus âgés, la courbe la plus haute se rapporte aux plus jeunes. Ce résultat est conforme à nos attentes.

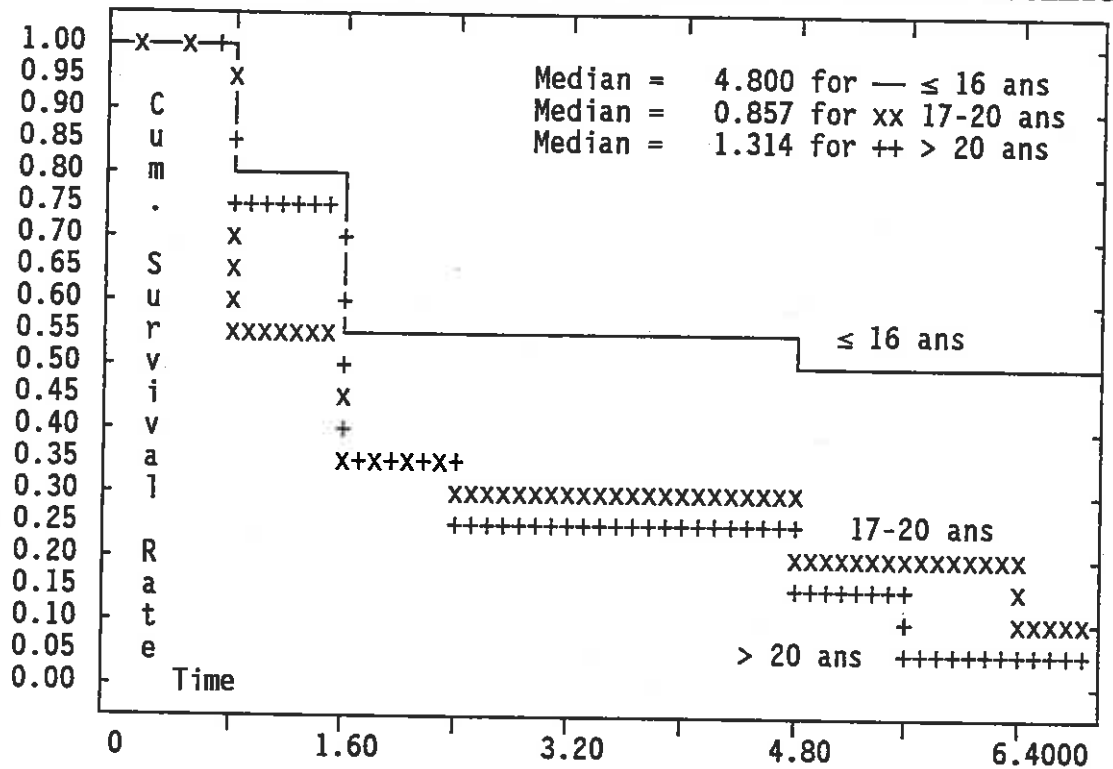
Tableau B.2.1 Durée du non emploi et âge de fin de scolarité :
Résultats de l'analyse par sous-populations.

Estimated Survival Function			
Duration variable is DUR1 Status is given by variable CENS1			
Stratification variable is CLAGE2 Number of strata is 3.			
	Stratum 1 ≤ 16 ans	Stratum 2 17-20 ans	Stratum 3 > 20 ans
Number of Observations	18 (24 %)	39 (51 %)	19 (25 %)
- Exiting	10 (56 %)	37 (95 %)	18 (95 %)
- Censored	8 (44 %)	2 (5 %)	1 (5 %)

Tableau B.2.2 Durée du non emploi et âge de fin de scolarité :
Table de survie estimée.

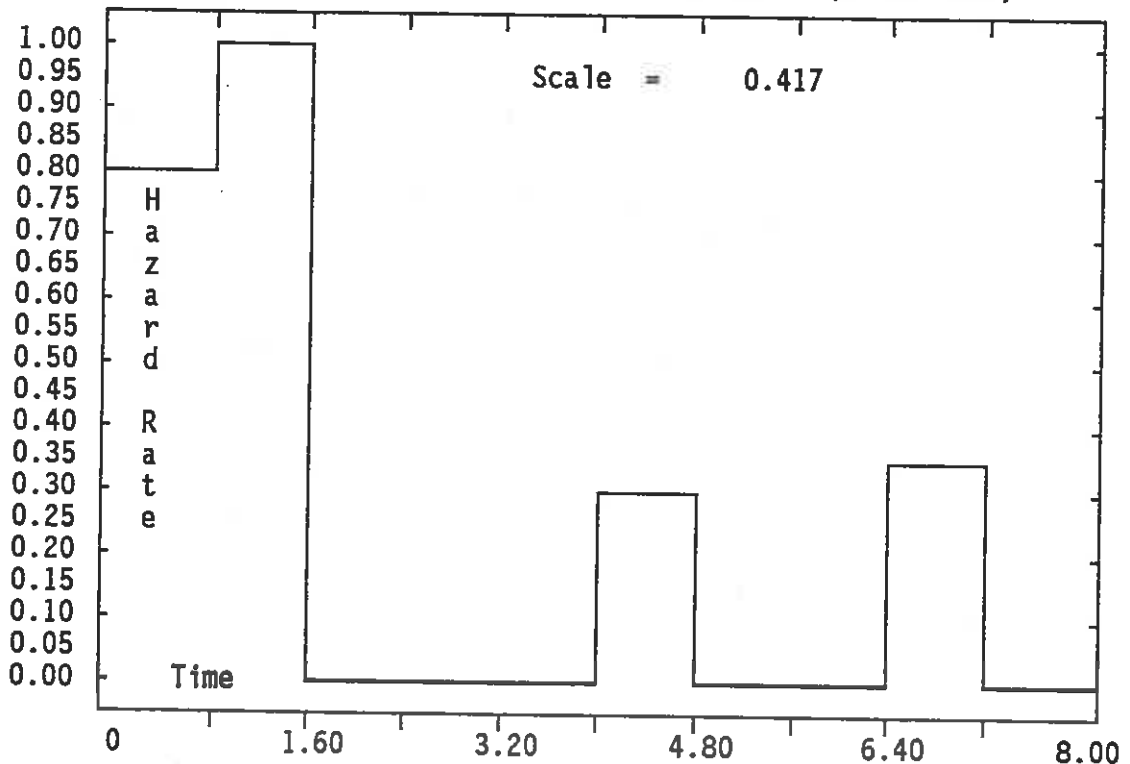
>>>>>>>>> Estimation results for stratum 1						
Duration	Enter	Cnsred	At Risk	Exiting	Survival (Std.Er.)	Hazd Rate (Std.Er.)
0.0-0.8	18	0	18	4 (0.2222)	1.0000 (0.0000)	0.3125 (0.1550)
0.8-1.6	14	0	14	4 (0.2857)	0.7778 (0.0980)	0.4167 (0.2054)
1.6-2.4	10	0	10	0 (0.0000)	0.5556 (0.1171)	0.0000 (0.0000)
2.4-3.2	10	0	10	0 (0.0000)	0.5556 (0.1171)	0.0000 (0.0000)
3.2-4.0	10	0	10	0 (0.0000)	0.5556 (0.1171)	0.0000 (0.0000)
4.0-4.8	10	0	10	1 (0.1000)	0.5556 (0.1171)	0.1316 (0.1314)
4.8-5.6	9	0	9	0 (0.0000)	0.5000 (0.1179)	0.0000 (0.0000)
5.6-6.4	9	0	9	0 (0.0000)	0.5000 (0.1179)	0.0000 (0.0000)
6.4-7.2	9	0	9	1 (0.1111)	0.5000 (0.1179)	0.1471 (0.1468)
7.2-8.0	8	8	4	0 (0.0000)	0.4444 (0.1171)	0.0000 (0.0000)
>>>>>>>>> Estimation results for stratum 2						
0.0-0.8	39	0	39	19 (0.487)	1.0000 (0.0000)	0.8051 (0.1749)
0.8-1.6	20	0	20	7 (0.350)	0.5128 (0.0800)	0.5303 (0.1959)
1.6-2.4	13	0	13	1 (0.077)	0.3333 (0.0755)	0.1000 (0.0999)
2.4-3.2	12	0	12	1 (0.083)	0.3077 (0.0739)	0.1087 (0.1086)
3.2-4.0	11	0	11	0 (0.000)	0.2821 (0.0721)	0.0000 (0.0000)
4.0-4.8	11	0	11	3 (0.273)	0.2821 (0.0721)	0.3947 (0.2250)
4.8-5.6	8	0	8	0 (0.000)	0.2051 (0.0647)	0.0000 (0.0000)
5.6-6.4	8	0	8	3 (0.375)	0.2051 (0.0647)	0.5769 (0.3241)
6.4-7.2	5	0	5	3 (0.600)	0.1282 (0.0535)	1.0714 (0.5589)
7.2-8.0	2	2	1	0 (0.000)	0.0513 (0.0353)	0.0000 (0.0000)

Graphique B.2.1 Durée du non emploi et âge de fin de scolarité :
Représentation des courbes de survie estimée.

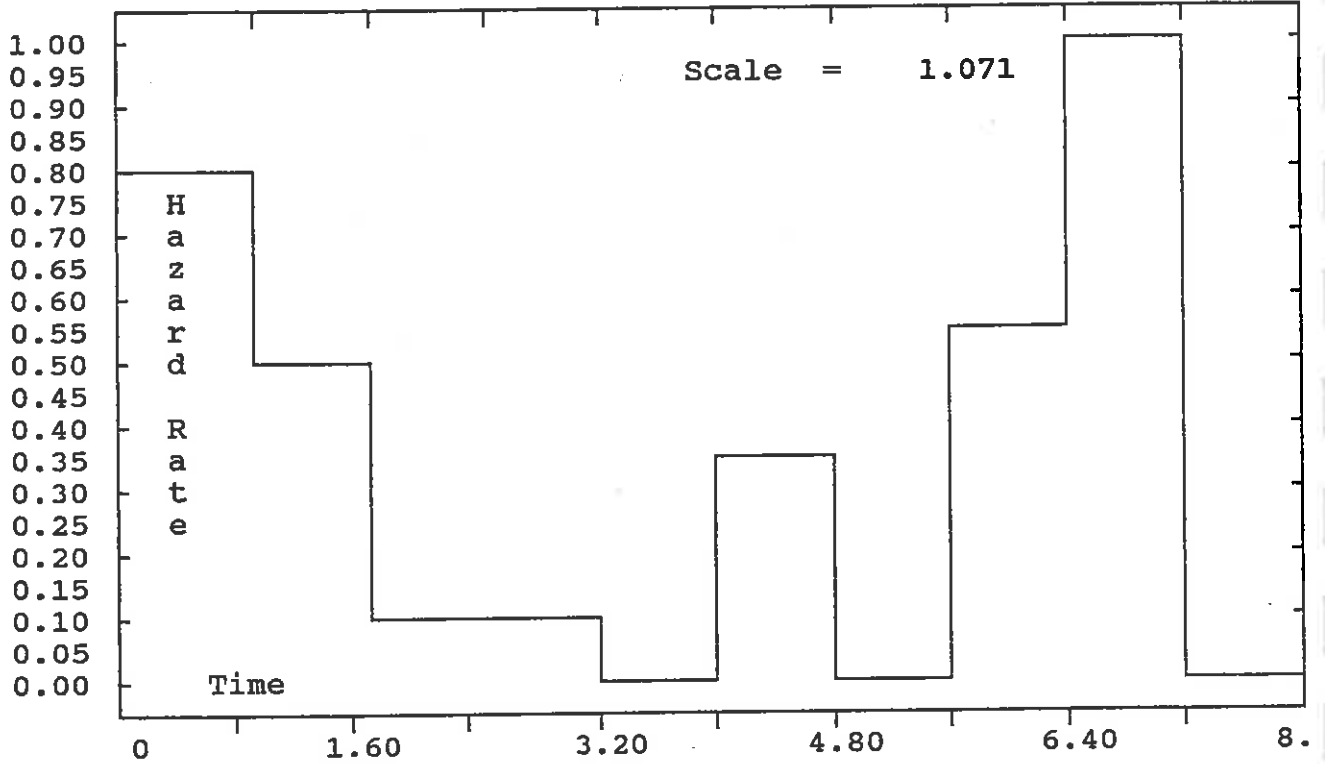


Graphique B.2.2 Représentations de l'évolution du taux de hasard au cours du temps.

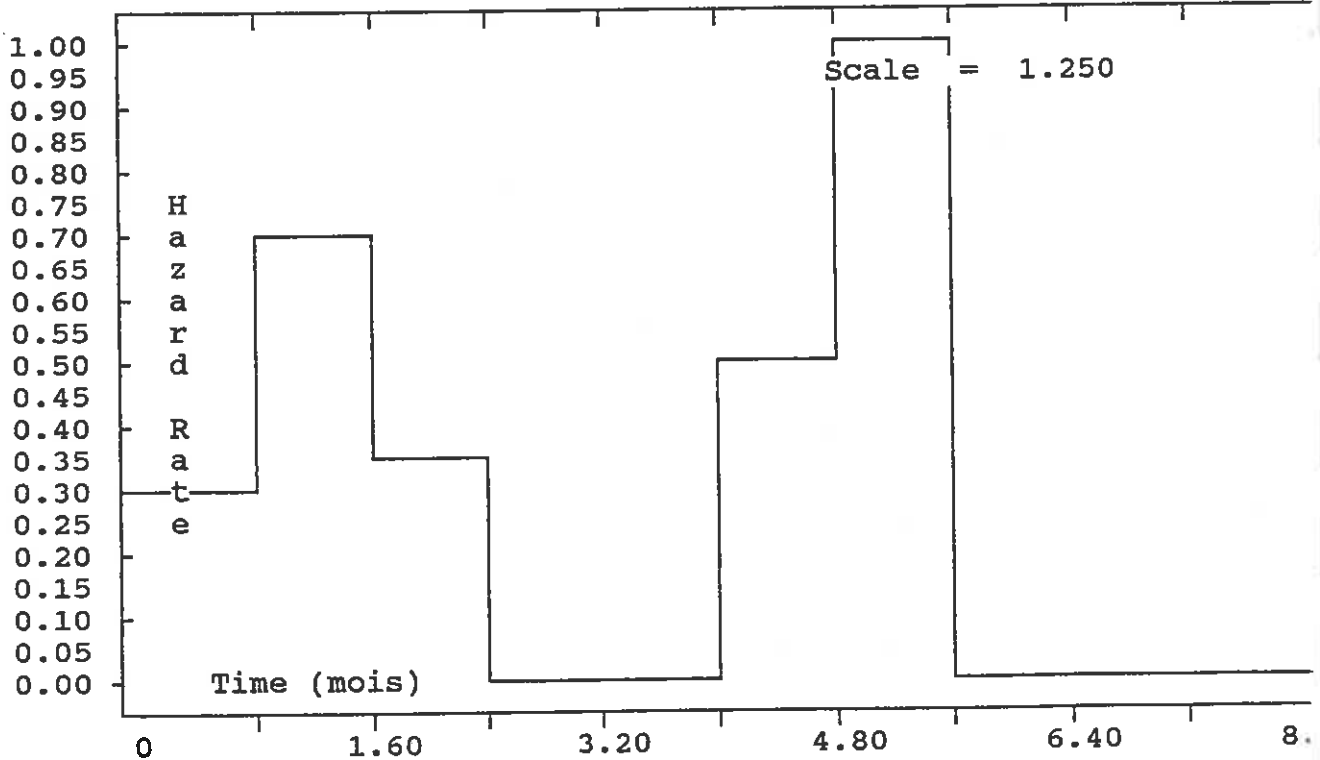
>>>>>>>>> Estimation results for stratum 1 (≤ 16 ans)



>>>>>>>>> Estimation results for stratum 2 (17 - 20 ans)



>>>>>>>>> Estimation results for stratum 3 (> 20 ans)



2.2 Durée du non emploi et niveau scolaire atteint

Le niveau scolaire de fin d'études a été partagé en trois classes qui correspondent à certains niveaux de qualification :

- < 12 années ; niveau inférieur
- 12 - 13 années ; niveau moyen correspondant à un niveau baccalauréat
- > 13 années ; niveau supérieur.

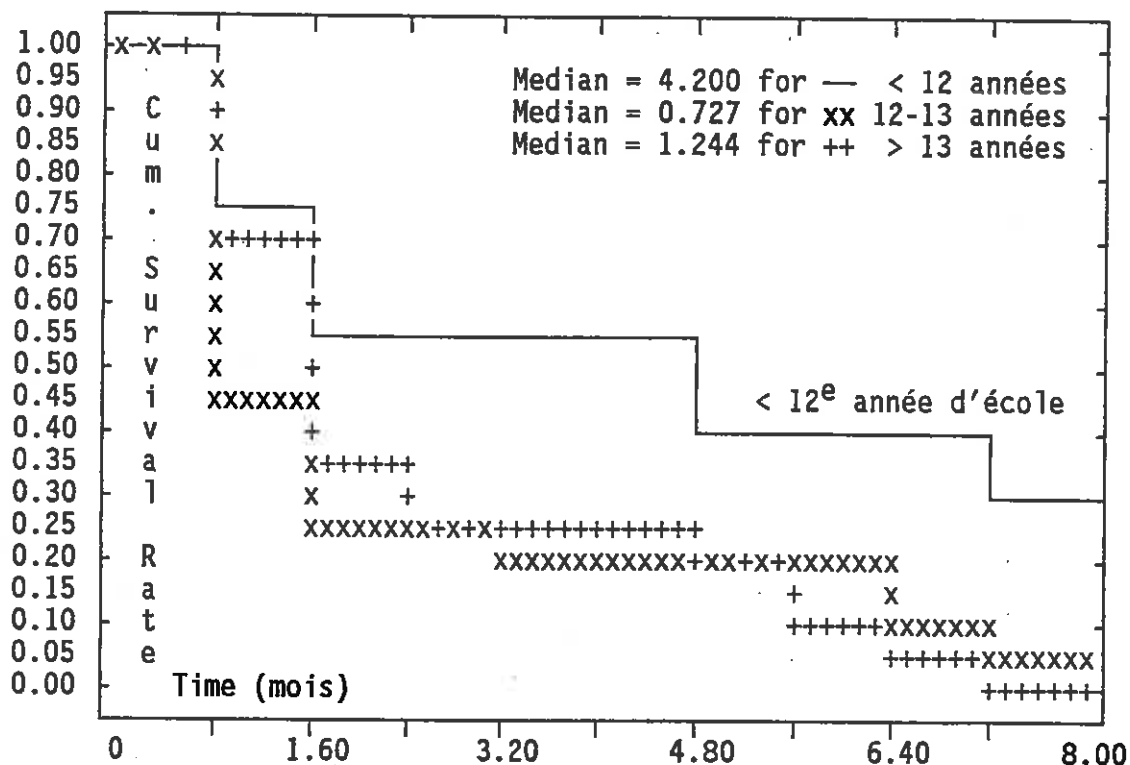
Sur le graphique B.2.2., il apparaît nettement que ce sont les individus qui ont le plus faible niveau d'études qui ont le plus de difficultés d'insertion professionnelle : leur courbe de survie, placée à un niveau supérieur, indique deux phénomènes : d'abord un moindre ancrage sur le marché de l'emploi, et ensuite un ancrage plus éloigné de la fin des études.

A contrario, si les individus les plus diplômés prennent le plus rapidement un emploi, ils le prennent aussi le plus souvent : plus de neuf sur dix des sortants ayant atteint la douzième année prennent un emploi contre seulement sept sur dix de ceux n'ayant pas dépassé la onzième année. Les conclusions sont dès lors semblables à celles concernant l'âge : une sortie anticipée du système scolaire peut être un révélateur d'une plus grande difficulté (ou d'une moins grande volonté de la part des parents) d'insertion sur le marché de l'emploi.

Tableau B.2.3 Durée du non emploi et niveau scolaire atteint : Résultats de l'analyse par sous-populations.

Estimated Survival Function			
Duration variable is DUR1 Status is given by variable CENS1			
Stratification variable is VAR01 Number of strata is 3.			
	Stratum 1 < 12 années	Stratum 2 12-13années	Stratum 3 > 13 années
Number of Observations	28 (36 %)	20 (26 %)	30 (38 %)
- Exiting	19 (68 %)	19 (95 %)	28 (93 %)
- Censored	9 (32 %)	1 (5 %)	2 (7 %)

Graphique B.2.3 Durée du non emploi et niveau scolaire atteint : Représentation des courbes de survie estimée.



Homogeneity tests : Degrees of freedom= 2			
Log-rank	=	9.3549	, Prob. 0.00930
Gen. Wilcoxon	=	6.1115	, Prob. 0.04709

La mise en oeuvre de ces analyses non paramétriques a permis d'entrevoir une certaine hétérogénéité de la population au regard de deux critères, l'âge et le niveau scolaire atteint. Elles n'ont pas permis de mesurer l'importance du rôle de ces deux variables sur la durée du non emploi. C'est ce que se proposent de faire les analyses suivantes, qu'elles soient semi-paramétriques ou paramétriques.

b) Une méthode semi-paramétrique d'analyse de la transition : le recours à la méthode de Cox n'impose pas de forme fonctionnelle dans la relation du taux de transition au temps.

Cette approche cherche à cerner les déterminants de la durée de non emploi : la variable expliquée est, comme précédemment, la

durée de non-emploi, durée connue avec certitude pour les seuls individus qui vont occuper un emploi sur la période d'observation. Des variables explicatives sont introduites : ce sont d'abord les variables significatives d'un comportement spécifique dans les analyses précédentes. Il s'agit de l'âge de fin de scolarité et du niveau de fin d'études. Il nous faut noter que ces variables ne peuvent pas être introduites simultanément en raison de leur forte corrélation : en général des âges élevés sont associés à des niveaux de fin d'études élevés. Sont ajoutées les variables dont on attend un effet significatif et qui dans les analyses précédentes se sont révélées sans effet ; ce sont le degré d'ancrage de la famille, la localisation géographique, le niveau de revenu de la famille, le nombre d'enfants dans la famille, le nombre de personnes ayant un emploi et le sexe de l'individu sortant du système scolaire.

Au total, trois analyses sont réalisées. La première prend en compte l'ensemble des variables dont on attend un effet, les deux suivantes ne retiennent que les variables significatives : d'abord le niveau scolaire atteint, ensuite l'âge de fin de scolarité.

1. Impact de différentes caractéristiques sur la durée du non emploi : une analyse semi-paramétrique de la liaison du risque d'emploi au temps.

Le modèle de Cox mis en oeuvre s'écrit $r(t;X) = r(0;X) \cdot \exp(\beta X \cdot t)$. Il suppose que tous les individus font face à un risque de base, $r(0; X)$, risque de base multiplié par une fonction exponentielle dépendant des caractéristiques des individus, $\exp(\beta X \cdot t)$. Le risque de base varie selon les individus. Cette variation est incorporée dans l'analyse en multipliant la fonction de hasard par une composante aléatoire spécifique à l'individu sortant du système scolaire.

Les résultats de l'analyse sont présentés dans le tableau B.2.4.

Selon le R^2_{TMF} [50] le modèle a un faible pouvoir explicatif (4,3 % d'amélioration de la vraisemblance) [51]. Il est de plus fortement non significatif (comme l'atteste le seuil de signification du χ^2 et comme le confirme le test du log-rank).

[50] Il s'agit d'une statistique inspirée de celle calculée par D. MacFadden et appliquée à l'analyse de survie. Elle donne l'amélioration de la vraisemblance qui résulte de l'introduction de variables explicatives.

[51] Ceci ne veut pas dire que c'est un mauvais modèle, mais que des facteurs essentiels n'ont pas été pris en compte.

Les autres résultats confirment totalement les tendances trouvées lors de l'analyse non paramétrique par sous-populations. Les individus ont un comportement d'emploi significativement différent selon leur âge en fin de scolarité et la liaison n'apparaît pas linéaire : le signe positif de la variable AGE et de la variable au cube, AGE3, et le signe négatif de la variable au carré, AGE2, confirment que ce sont les individus les plus jeunes, au moment où ils quittent le système scolaire, qui ont la plus faible probabilité d'emploi.

Tableau B.2.4

Cox Survival Regression Model						
Duration variable is DUR1						
Status is given by variable CENS1						
Total Number of Observations	=					76
Total Number of Observations Exiting	=					65
Total Number of Observations Censored	=					11
Total Number of Distinct Exit Times	=					7
Maximum Likelihood Estimates						
Log-Likelihood.....						-238.73
Restricted (Slopes=0) Log-L.						-249.48
Chi-Squared (15).....						21.505
Significance Level.....						0.12147
R ² TMF.....						0.04271
Log-rank test with 15 degrees of freedom:						
Chi-squared =	19.548					- probability = 0.1900
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std X
AGE	3.051960	1.436540	2.125	0.03363	18.987	3.8175
AGE2	-1.264450	0.594968	-2.125	0.03357	37.488	17.704
AGE3	0.165176	0.078499	2.104	0.03536	77.730	67.922
NBENF	-0.049750	0.148736	-0.334	0.73801	1.1447	1.3924
NBEMP	-0.239311	0.186100	-1.286	0.19847	1.6447	1.0159
SEXE	0.189238	0.294869	0.642	0.52102	0.55263	0.50053
APPRENTI	0.314525	0.333015	0.944	0.34493	0.22368	0.41948
SD85	0.019242	0.012038	1.598	0.10994	9.8026	13.884
CENTRE	0.082634	0.357195	0.231	0.81705	0.44737	0.50053
NORDEST	0.039539	0.395752	0.100	0.92042	0.31579	0.46792
CHEFINF	-0.282938	0.351030	-0.806	0.42023	0.47368	0.50262
CLQF1	0.066988	0.587974	0.114	0.90929	0.14474	0.35417
CLQF2	-0.300156	0.581911	-0.516	0.60599	0.19737	0.40066
CLQF3	0.050555	0.472683	0.107	0.91483	0.25000	0.43589
CLQF4	0.286595	0.450963	0.636	0.52509	0.23684	0.42797

Tableau B.2.5

Cox Survival Regression Model						
Duration variable is DUR1						
Status is given by variable CENS1						
Total Number of Observations	=					76
Total Number of Observations Exiting	=					65
Total Number of Observations Censored	=					11
Total Number of Distinct Exit Times	=					7
Maximum Likelihood Estimates						
Log-Likelihood.....						-244.12
Restricted (Slopes=0) Log-L.						-249.48
Chi-Squared (2).....						10.723
Significance Level.....						0.46949E-02
R ² MF.....						0.02148
Log-rank test with 2 degrees of freedom:						
Chi-squared =	8.7954					probability = 0.0123
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D. X
VAR01	1.023870	0.512213	1.999	0.04562	12.197	2.4332
VARCAR	-0.034849	0.019858	-1.755	0.07928	154.62	60.8750

2. Impact du niveau scolaire atteint sur la durée du non emploi : une analyse semi-paramétrique de la liaison du risque d'emploi au temps.

Les résultats de l'analyse sont présentés dans les tableaux B.2.6 et B.2.7.

Selon le R^2_{TMF} (tableau B.2.6) le modèle a, comme le précédent, un faible pouvoir explicatif (2,1 % d'amélioration de la vraisemblance). Mais il est fortement significatif (comme l'atteste le seuil de signification du χ^2 et comme le confirme le test du log-rank).

Pour les autres résultats, ce qui nous intéresse ici, c'est la mesure de l'impact du niveau scolaire sur la durée du non emploi. Pour ce faire, nous avons introduit la variable niveau scolaire atteint en 1985 (VAR01) et la variable au carré (VARCAR). Ces deux variables sont significatives au seuil de 5 % pour VAR01 et de 10 % pour VARCAR. Leur signe contrasté (négatif pour VARCAR, positif pour VAR01) laisse supposer que l'impact du niveau scolaire sur la durée du non emploi n'est pas monotone. Pour vérifier cette assertion, on peut quantifier l'impact d'une

variation unitaire du niveau scolaire sur la probabilité qu'ont les individus de prendre un emploi. Cet impact est mesuré sur un individu type. A titre d'exemple, nous partons de l'individu qui a les caractéristiques moyennes. C'est une étudiant qui a atteint la douzième année. S'il avait suivi une année supplémentaire d'études, sa probabilité de prendre un emploi augmenterait de 169 %^[52]. Si ce même individu avait suivi deux années supplémentaires de cours, sa probabilité de trouver un emploi se trouverait multipliée par 6,7 (+ 574 %^[53]). Le tableau B.2.7 résume différents effets d'une modification du niveau scolaire atteint. Les résultats trouvés y montrent l'effet non négligeable d'une variation du niveau scolaire sur la durée du non emploi.

Tableau B.2.7

Variation du niveau scolaire atteint (en années)	$\exp(\beta \Delta X)$	Variation de la probabilité de prendre un emploi
+ 1 an	2,69	+169 %
+ 2 ans	6,74	+574 %
+ 3 ans	15,77	+1477 %
+ 4 ans	34,39	+3339 %
- 1 an	0,35	-65 %
- 2 ans	0,11	-89 %
- 3 ans	0,03	-97 %
- 4 ans	0,01	-99 %

Enfin l'analyse de la médiane^[54] montre que plus de la moitié des sortants du système scolaire trouve un emploi dans un délai de 4 mois après la fin des études. Il ne faut pas oublier que, parmi les 50 % d'individus ne trouvant pas d'emploi dans ce délai, un certain nombre ne désire pas en acquérir un.

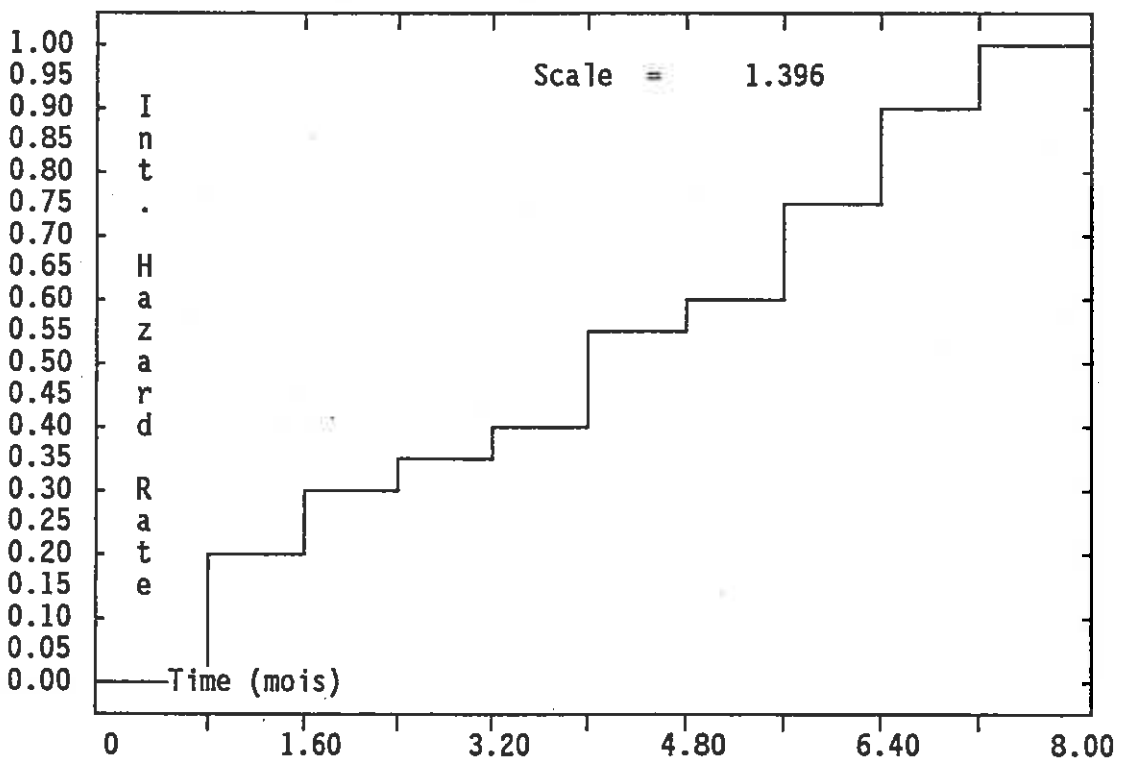
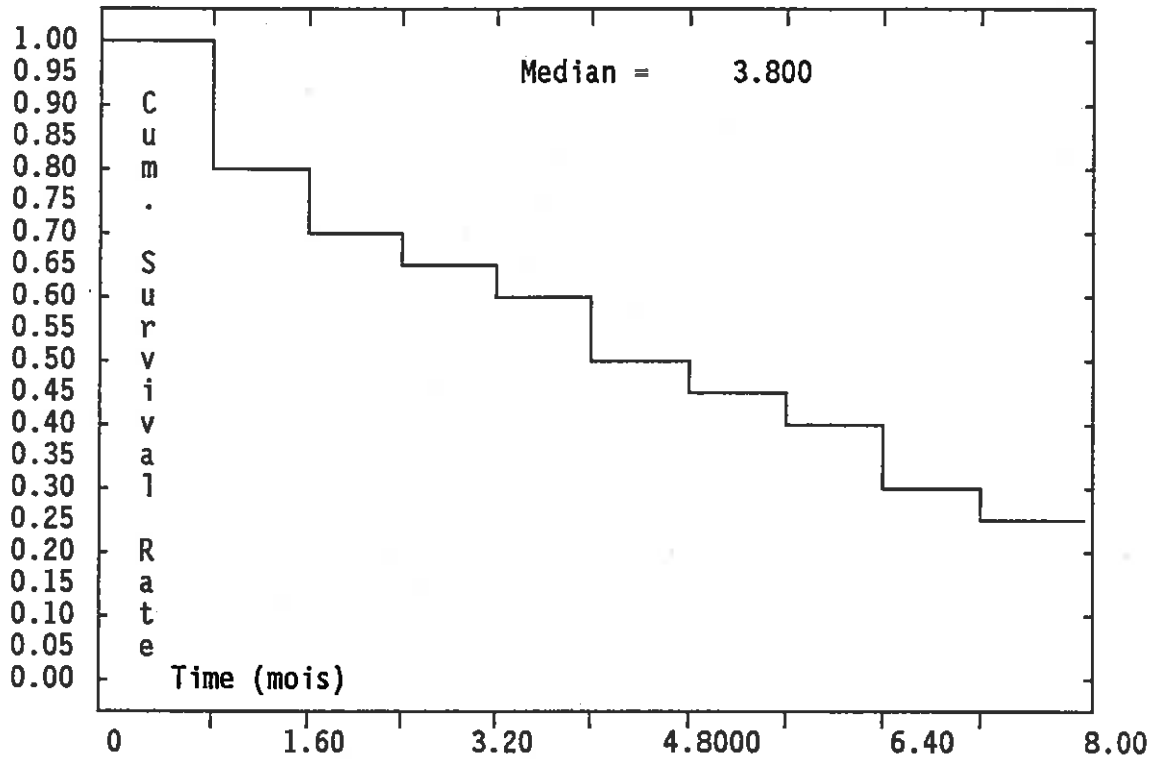
[52] Ce chiffre est égal à : $[\exp(1,02387-0,03485)-1]*100\%$

[53] Ce chiffre est égal à : $[\exp(1.02387*2-0.03485*2^2)-1]*100\%$.

[54] Ce chiffre médian est différent de celui trouvé lors des analyses univariées car ces dernières considèrent comme exactes les durées de non emploi connues avec incertitude.

Graphiques B.2.4 Représentations de la courbe de survie estimée et du taux de hasard intégré de la durée du non emploi selon le niveau scolaire atteint.

Survival curve is plotted at means of regressors.



3. Impact de l'âge de fin de scolarité sur la durée du non emploi : une analyse semi-paramétrique de la liaison du risque d'emploi au temps.

Les résultats de l'analyse sont présentés dans les tableaux B.2.8. et B.2.9.

Selon le R^2_{TMF} (tableau 3.1) le modèle a, comme précédemment un faible pouvoir explicatif (2,6 % d'amélioration de la vraisemblance), et il est toujours fortement significatif (cf le seuil de signification du χ^2 et le test du log-rank).

Tableau B.2.8

Cox Survival Regression Model						
Duration variable is DUR1						
Status is given by variable CENS1						
Total Number of Observations	=					76
Total Number of Observations Exiting	=					65
Total Number of Observations Censored	=					11
Total Number of Distinct Exit Times	=					7
Maximum Likelihood Estimates						
Log-Likelihood.....						-243.10
Restricted (Slopes=0) Log-L.						-249.48
Chi-Squared (3).....						12.771
Significance Level.....						0.51597E-02
R^2_{TMF}						0.02557
Log-rank test with 3 degrees of freedom:						
Chi-squared =	10.908					- probability = 0.0122
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D.X
AGE	3.24057	1.212830	2.672	0.00754	18.987	3.8175
AGE2	-1.30861	0.514370	-2.544	0.01096	37.488	17.704
AGE3	0.168451	0.068880	2.446	0.01446	77.730	67.922

Pour les autres résultats, ce qui nous intéresse ici, c'est la mesure de l'impact de l'âge de fin de scolarité sur la durée du non emploi. Pour ce faire, nous avons introduit la variable âge de fin de scolarité (AGE), la variable au carré (AGE2) et la variable au cube (AGE3). Ces trois variables sont significatives au seuil de 1 %. Leur signe contrasté (négatif pour AGE2, positif pour AGE et AGE3) laisse supposer que l'impact d'années supplémentaires d'études sur la durée du non emploi n'est pas

linéaire, mais plutôt exponentiel au début puis relativement constant.

Pour vérifier cette assertion, nous avons quantifié l'impact de variations de l'âge de fin de scolarité sur la probabilité qu'ont les individus de prendre un emploi. Nous partons d'un individu qui a des caractéristiques données, par exemple celui qui a les caractéristiques moyennes. C'est un étudiant qui 19 ans. S'il avait quitté l'école une année plus tôt, sa probabilité de prendre un emploi aurait diminué de 99 % (et devient donc voisine de zéro). Le tableau B.2.9 résume différents effets d'une modification du niveau scolaire atteint. Les résultats trouvés y montrent l'effet non négligeable de deux années supplémentaires d'études (l'effet de trois années et plus n'étant guère différent de celui de deux années).

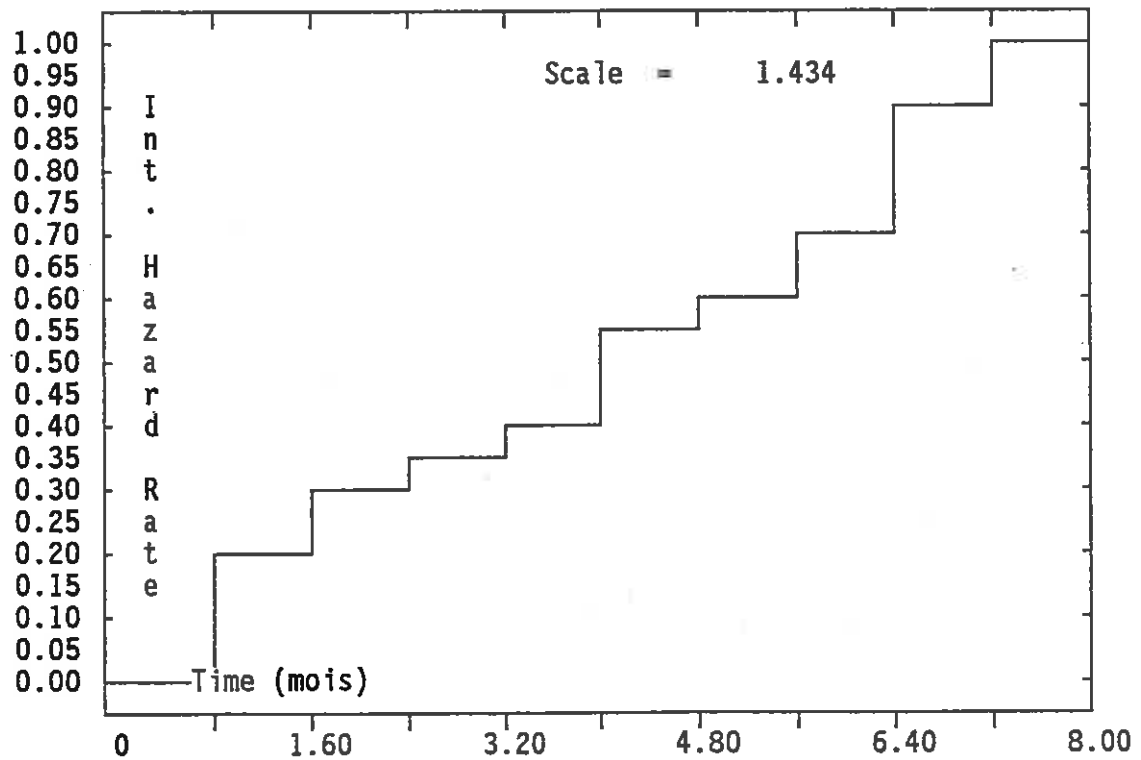
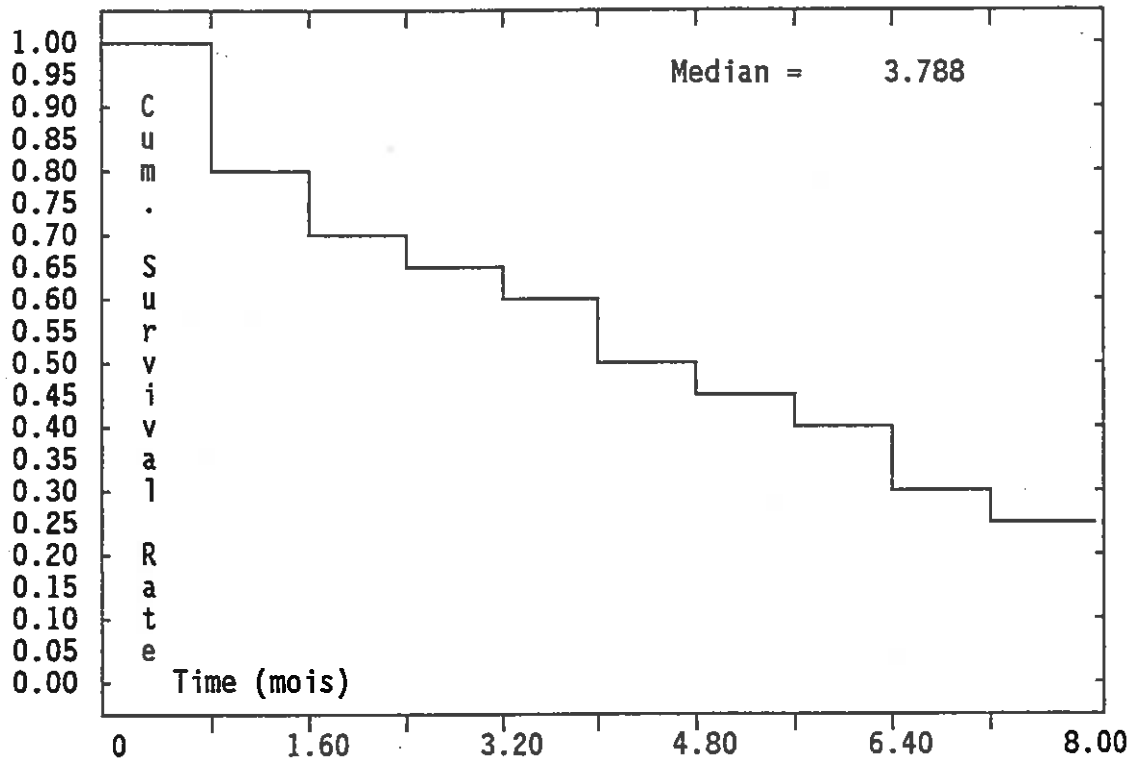
Tableau B.2.9

Variation l'âge de fin de scolarité (en années)	$\exp(\beta\Delta X)$	Variation de la probabilité de prendre un emploi
+ 1 an	8,17	+717 %
+ 2 ans	13,39	+1239 %
+ 3 ans	12,09	+1109 %
+ 4 ans	16,53	+1553 %
- 1 an	0,01	-99 %
- 2 ans	0,00	-100 %
- 3 ans	0,00	-100 %
- 4 ans	0,00	-100 %

Enfin l'analyse de la médiane montre, comme précédemment, que plus de la moitié des sortants du système scolaire trouve un emploi dans un délai de 4 mois après la fin des études.

Graphiques B.2.5 Représentations de la courbe de survie estimée et du taux de hasard intégré.

Survival curve is plotted at means of regressors.



Nous avons prolongé cette étude en utilisant un modèle paramétrique de WEIBULL. Les raisons en ont été développées au chapitre 2.

c) Les résultats de la mise en oeuvre de méthodes paramétriques d'analyse de la transition instantanée.

Toujours sous l'hypothèse de proportionnalité des hasards, nous avons testé un type de modèles alternatifs, qui impose une forme particulière dans la relation du taux de hasard au temps. Nous avons choisi le modèle général de WEIBULL^[55]. Le taux de hasard s'y exprime comme : $r(t;X) = \exp(-\alpha X)^p \cdot p \cdot t^{p-1}$. Il y a donc, comme dans le modèle de Cox, un effet proportionnel ($\exp(BX)$), mais, à la différence du modèle de Cox, la fonction de hasard de base $r(0;X) = p \cdot t^{p-1}$ dépend d'un paramètre inconnu, p . L'intérêt d'un tel modèle est que, si le paramètre $p (= 1/\sigma)$ est égal à l'unité, on est sûr d'avoir un taux de transition constant sur le temps. Mais ce n'est pas ce que l'on constate ; nous avons, dans l'ensemble des modèles mis en oeuvre, $\sigma = 3,2$ et donc $p = 0,3$. Et comme la statistique du T confirme que ce résultat est significatif, on est amené à envisager une certaine dépendance de durée négative : plus les individus sont observés sans emploi, moins ils ont de chances d'en occuper un.

Nous présentons les résultats de trois analyses. Leur caractéristique est d'avoir les mêmes variables explicatives que les trois analyses présentées dans le paragraphe précédent.

Un premier résultat fait apparaître une proximité des coefficients entre les modèles de COX et les modèles de WEIBULL. Cette proximité s'explique par le fait que, dans les deux cas, nous faisons l'hypothèse que le hasard est proportionnel entre deux individus.

Par ailleurs, on ne constate pas d'amélioration de la qualité de la régression : le R^2_{TMF} est de l'ordre de 3 %. Les variables significatives sont les mêmes que précédemment, à savoir l'âge, l'âge au carré et l'âge au cube d'une part, le niveau sco-

[55] Ces modèles sont estimés sous forme log-linéaire. Comme nous avons beaucoup de durées nulles, nous avons été amenés à les remplacer par 0.01, de façon à ce que leur logarithme devienne opérationnel.

laire atteint et ce niveau au carré d'autre part. L'opposition de signes entre les variables signale une relation non monotone. Quand on quantifie l'impact d'une modification, toutes choses égales par ailleurs, d'une variable explicative sur la probabilité, pour l'individu ayant les caractéristiques moyennes, de prendre un emploi, on obtient des effets de taille comparable à ceux obtenus par le modèle de COX.

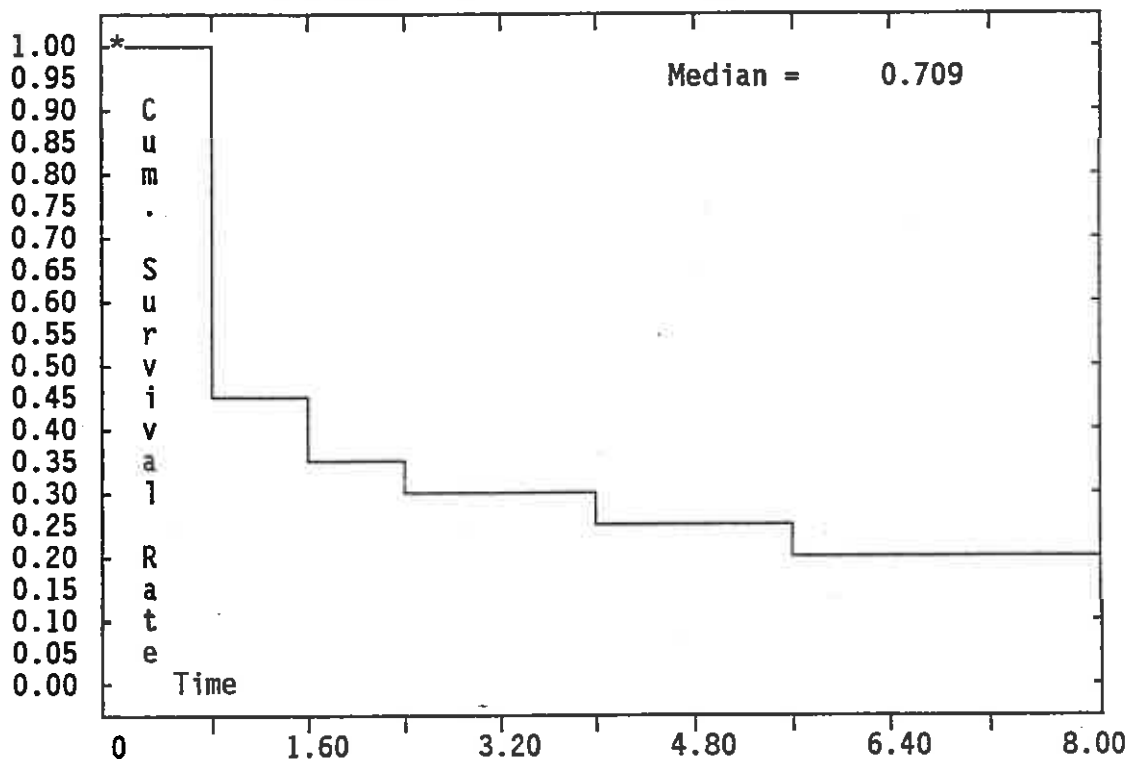
Tableau B.2.10

Log-linear survival regression model: Weibull						
Least squares is used to obtain starting values for MLE.						
Censoring status variable is CENS1						
=====						
<< Ordinary Least Squares Regression >>						
Dependent Variable	LOGT	Number of Observations				76
Mean of Dep. Var.	-1.87723	Std. Dev. of Dep. Var.				3.93269
Std. Error of Regr.	3.90698	Sum of Sqrd. Residuals				1114.31
R - squared	0.039354	Adjusted R - Squared				0.013035
Durbin Watson Stat.	0.000000	Estd. Autocorrelation				0.191039
Total Variation	1159.96	Regression Variation				45.6493
F(2, 73)	1.4953	Prob. Value for F				0.23097
Log-Likelihood	-209.879	Const. Log-L (for b=0)				-211.405
Akaike Information	2.76420	Amemiya Prediction				15.8670
Chi-squared[2]	3.051373	Prob Value =				0.217472
=====						
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D. X
VAR01	-2.21374	1.53398	-1.443	0.14898	12.197	2.4332
VARCAR	0.080774	0.06131	1.317	0.18771	154.62	60.875
ONE	12.6354	9.38079	1.347	0.17800	1.0000	0.0000
Maximum Likelihood Estimates						
Log-Likelihood.....						-194.31
R ² TMF.....						0.02528
=====						
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D. X
VAR01	-3.53700	2.06105	-1.716	0.08614	12.197	2.4332
VARCAR	0.12385	0.08215	1.508	0.13165	154.62	60.8750
ONE	24.37890	12.76780	1.909	0.05621	1.0000	0.0000
Sigma	3.28664	0.47105	6.977	0.00000	1.6447	1.0159
Parameter	Estimate	Std. Error	Confidence Interval			
Lambda	0.67961	0.35137	-0.0091 to 1.3683			
P	0.30426	0.04361	0.2188 to 0.3897			
Median	0.44115	0.22808	-0.0059 to 0.8882			
Percentiles of survival distribution:						
Survival	0.25	0.50	0.75	0.95		
Time	4.30	0.44	0.02	0.00		

Mais, à l'opposé du modèle de COX, la prise en compte simultanée d'un certain nombre de variables explicatives (cf tableau B.2.12) ne fournit aucune explication de la durée du non emploi.

Par ailleurs, une constante apparaît dans le modèle de WEIBULL qui n'était pas là dans celui de COX, car ces derniers sont homogènes de degré zéro, alors que les premiers sont homogènes de degré 1 ; de ce fait l'absence de constante dans les modèles de WEIBULL conduit à des résultats biaisés.

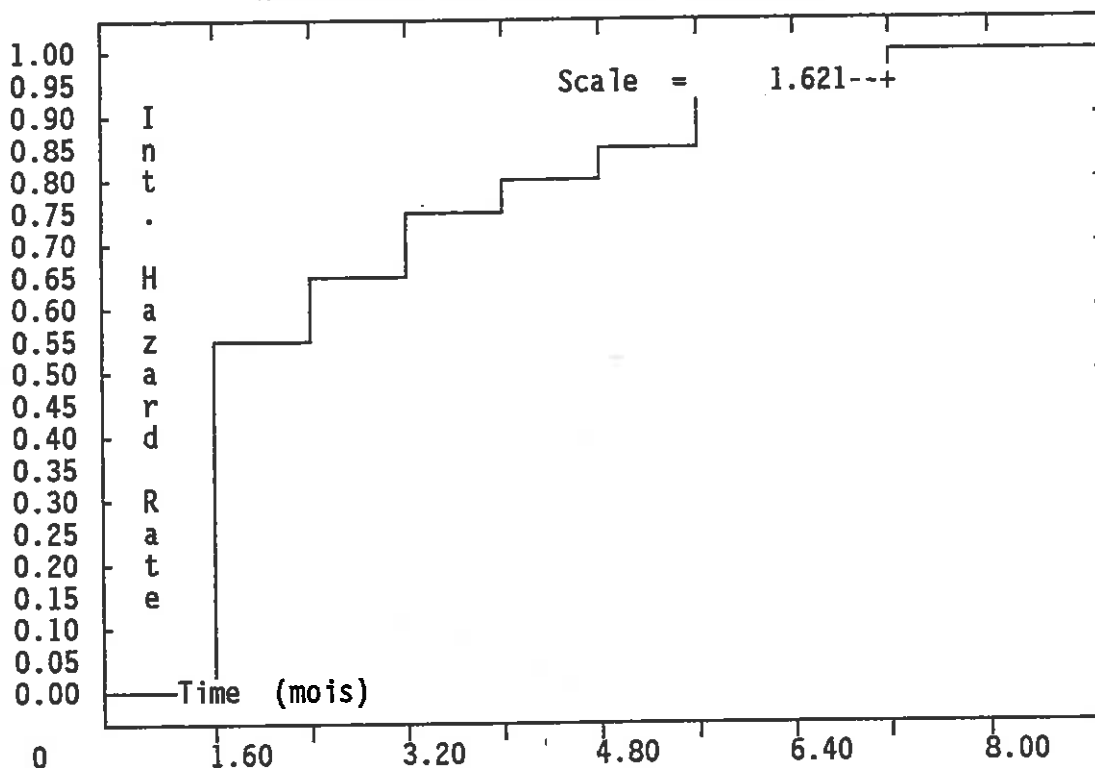
Graphique B.2.6 Courbe de survie estimée de la durée du non emploi quand seul le niveau scolaire atteint est introduit en variable explicative.



De plus l'analyse des quartiles produit des résultats différents entre les deux types de modèle : si un sortant sur quatre n'a aucune chance d'occuper un emploi dans les quatre mois (contre un sur deux précédemment), un sortant sur deux va occuper un emploi dans les quinze jours. Pour l'individu ayant les caractéristiques moyennes (19 ans, 12 années de scolarité) le temps avant d'occuper un emploi plutôt long, environ 14 mois. Ce temps de 14 mois peut étonner dans un échantillon où la durée maximale recensée est de 8 mois. L'explication est simple : pour les individus n'ayant pas encore trouvé d'emploi en fin de période d'observation, il a été calculé une durée de non em-

ploi et comme elle est nécessairement supérieure à 8 mois, cette durée estimée fait augmenter la durée moyenne de non emploi.

Graphique B.2.7 Taux de hasard cumulé de la durée du non emploi quand seul le niveau scolaire atteint est introduit en variable explicative.



On peut voir comment se modifie cette durée moyenne suite à une variation, *ceteris paribus*, d'une variable explicative. Si l'âge augmente d'une année, la durée moyenne du non emploi diminue de 91 % (cf Tableau B.2.11), pour passer à un mois ; si le niveau scolaire atteint augmente d'une année, la durée du non emploi diminue de 65 %^[56], soit d'environ 9 mois (tableau B.2.10).

[56] $\Delta E = (\exp[(-3,537+0,124)*0,30426] - 1)*100 \%$

Tableau B.2.11

Log-linear survival regression model: Weibull
Censoring status variable is CENS1

Maximum Likelihood Estimates

Log-Likelihood.....	-192.00
R ² MF.....	0.03687

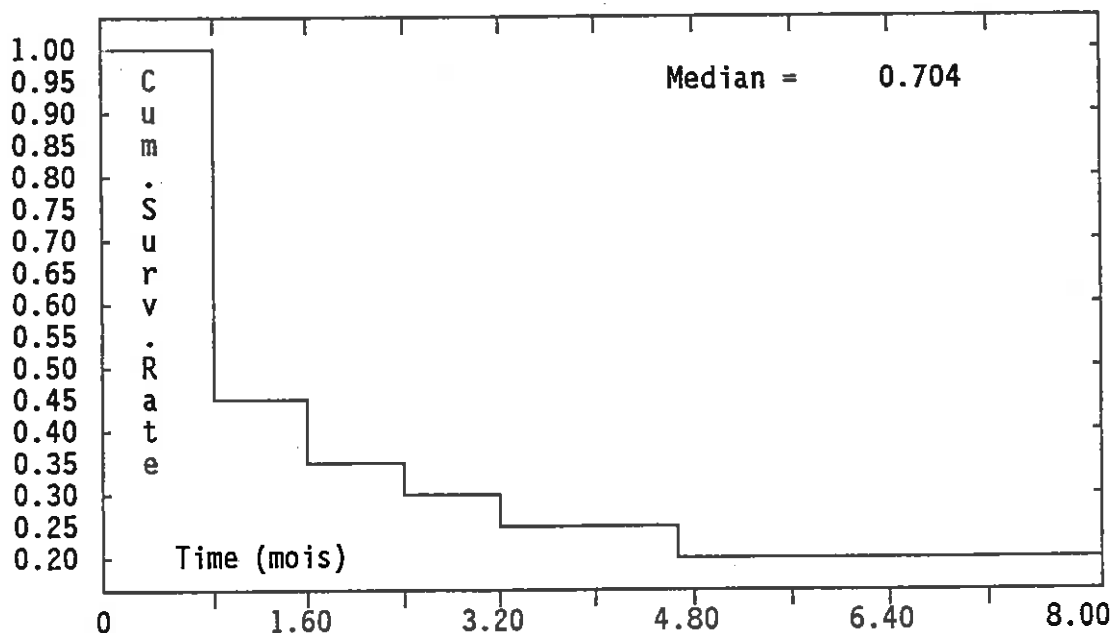
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D. X
ONE	93.9210	74.6033	1.259	0.20805	1.0000	0.0000
AGE	-12.2200	11.2387	-1.087	0.27690	18.987	3.8175
AGE2	5.1032	5.5102	0.926	0.35437	37.488	17.7040
AGE3	-0.6802	0.8747	-0.778	0.43675	77.730	67.9220
Sigma	3.2023	0.5154	6.213	0.00000	0.553	0.5005

Parameter	Estimate	Std. Error	Confidence Interval
Lambda	0.71305	0.43168	-0.1331 to 1.5591
P	0.31228	0.05026	0.2138 to 0.4108
Median	0.43367	0.26255	-0.0809 to 0.9483

Percentiles of survival distribution:

Survival	0.25	0.50	0.75	0.95
Time	3.99	0.43	0.03	0.00

Graphique B.2.8 Courbe de survie estimée de la durée du non emploi quand seul l'âge de fin d'études est introduit en variable explicative.



Graphique B.2.9 Taux de hasard cumulé de la durée du non emploi quand seul l'âge de fin de scolarité est introduit en variable explicative.

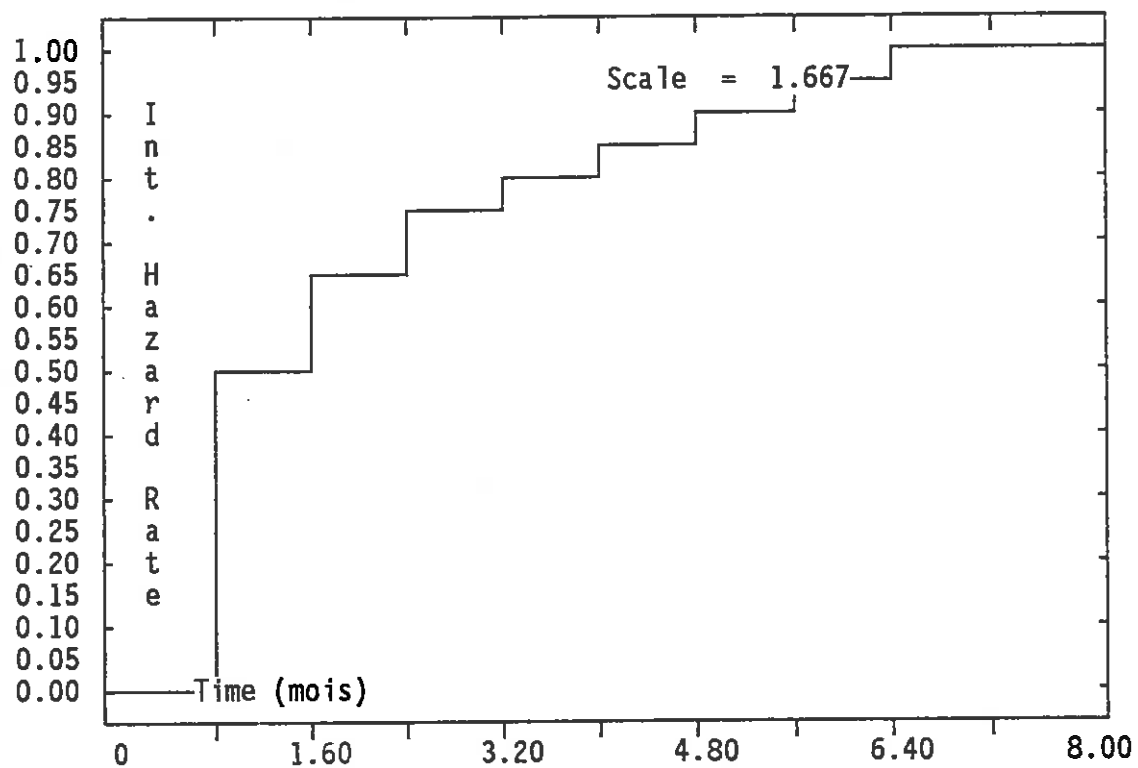


Tableau B.2.12

Log-linear survival regression model: Weibull						
Least squares is used to obtain starting values for MLE.						
Censoring status variable is CENS1						
=====						
Maximum Likelihood Estimates						
Log-Likelihood.....						-186.73
R ² TMF.....						0.06331
=====						
Variable	Coeffic.	Std.Err.	T-ratio	Prob:t>x	Mean X	Std.D. X

ONE	80.13750	73.2583	1.094	0.27400	1.0000	0.00000
AGE	-10.7521	10.8240	-0.993	0.32053	18.987	3.8175
AGE2	4.610770	5.26847	0.875	0.38149	37.488	17.704
AGE3	-0.624504	0.83742	-0.746	0.45582	77.730	67.922
SEXE	-0.870664	1.09957	-0.792	0.42846	0.5526	0.50053
APPRENTI	-1.128850	1.17599	-0.960	0.33710	0.2237	0.41948
CENTRE	-0.198207	1.27007	-0.156	0.87599	0.4474	0.50053
NORDEST	-0.468408	1.45055	-0.323	0.74676	0.3158	0.46792
CHEFINF	0.371979	1.30070	0.286	0.77489	0.4737	0.50262
CLQF1	-0.372853	2.38297	-0.156	0.87567	0.1447	0.35417
CLQF2	0.881826	2.28338	0.386	0.69935	0.1974	0.40066
CLQF3	-0.157551	1.74920	-0.090	0.92823	0.2500	0.43589
CLQF4	-1.185300	1.69149	-0.701	0.48346	0.2368	0.42797
NBENF	0.345383	0.65047	0.531	0.59544	1.1447	1.3924
SD85	-0.061277	0.04226	-1.450	0.14704	9.8026	13.884
NBEMP	0.704897	0.71864	0.981	0.32666	1.6447	1.0159
Sigma	3.059720	0.47985	6.376	0.00000	0.0000	0.0000

Parameter	Estimate	Std. Error	Confidence Interval			

Lambda	0.77457	0.54284	-0.2894 to		1.8385	
P	0.32683	0.05126	0.2264 to		0.4273	
Median	0.42064	0.29480	-0.1572 to		0.9984	

Percentiles of survival distribution:						
Survival	0.25	0.50	0.75	0.95		
Time	3.51	0.42	0.03	0.00		

CONCLUSION

Somme toute, quel que soit le modèle testé, les résultats interrogent. En effet, les femmes n'apparaissent pas avoir plus de difficultés à trouver un emploi que les hommes. De même être de nationalité luxembourgeoise ne facilite pas l'insertion sur le marché du travail. Plus encore, que les études suivies soient de type général ou qu'elles soient de type technique, les élèves n'ont pas un meilleur accès à l'emploi.

On peut s'interroger sur le pourquoi de l'absence de différence de comportement des individus selon les variables retenues. Car ces variables devraient avoir un effet : les travaux des économistes du "job search" l'attestent.

L'absence constatée de différence de comportement est-elle due :

- aux caractéristiques du marché luxembourgeois de l'emploi?
- à la faiblesse numérique de l'échantillon ?
- aux méthodes d'estimation employées ?
- à d'autres raisons ?

S'il est vrai qu'il semble, a priori, étonnant d'étudier la durée du non emploi dans une économie qui manque de main d'oeuvre et qui en conséquence a recours à une forte main d'oeuvre étrangère (43 % de la population active est étrangère et une grande partie sont des travailleurs non-résidents), on constate néanmoins un certain taux de chômage (environ 2 %). De plus les études effectuées ci-dessus font apparaître, sinon une difficulté à trouver un emploi, du moins, pour certains individus, un certain délai avant d'en trouver un.

Cependant ces difficultés ne sont pas liées, nous l'avons souligné supra, au type d'enseignement suivi. Sans doute, la tertiarisation de l'économie luxembourgeoise a-t-elle pour conséquence de favoriser l'insertion des élèves ayant suivi une formation générale par rapport à ceux ayant suivi une formation spécifique, ces dernières étant moins aisément adaptables aux besoins changeant des employeurs. Plus encore, l'étude fait apparaître que tous les individus ayant suivi une formation générale ont trouvé un emploi avant la fin de la période d'observation (8 mois).

D'autre part, l'étude a l'avantage de mettre en lumière qu'une proportion non négligeable de jeunes sortent de l'école (environ 15 %) avant la fin de la scolarité obligatoire. Interrogés par ce résultat, nous nous sommes penchés sur les autres caractéristiques de ces individus. Il est alors apparu qu'il s'agissait quasi-exclusivement de jeunes d'origine portugaise. Et il semble que certaines familles portugaises (peut-être d'immigration récente) retirent leurs enfants de l'école à l'âge de fin de scolarité obligatoire ayant cours au Portugal (14 ans).

Or ces phénomènes peuvent avoir des répercussions, sinon sur l'insertion professionnelle initiale, du moins sur le niveau de revenu des individus concernés et sur leur insertion professionnelle future : à supposer que l'économie luxembourgeoise doive, un jour, faire face à une compression de sa main d'oeuvre, les premiers touchés seront les non qualifiés et ils sont relativement nombreux à la sortie du système scolaire.

Cette disqualification des sortants du système scolaire luxembourgeois risque de trouver, lors de l'ouverture du grand marché européen en 1993, toute son importance face à la concurrence qui naîtra.

Ces quelques remarques doivent être prises avec réserve en raison de la faiblesse numérique de l'échantillon étudié. L'introduction des vagues successives de l'enquête par panel auprès des ménages luxembourgeois permettra soit de les infirmer, soit de les confirmer. Mais il n'y a pas, a priori, de raison de penser que les élèves sortis du système scolaire en 1985 (cohorte étudiée dans ce rapport) aient des caractéristiques fondamentalement différentes de ceux sortis les années suivantes. Le seul avantage d'un échantillon plus grand est de multiplier les cas de figures possibles et donc de ne plus faire reposer les résultats sur les deux seules variables qui sont apparues significatives dans cette étude, à savoir l'âge de fin de scolarité et le niveau atteint.

BIBLIOGRAPHIE

- ALLEGREZZA-CARVOYEUR L.S., KOP J.L., [1991] Les variables du fichier PETRA. Définition, mode de calcul et statistiques univariées, Document FEE, CEPS/INSTEAD, LUXEMBOURG, 20 p.
- ALLISON P.D., [1984], Event History Analysis : Regression for Longitudinal Event Data, SAGE, 84 p.
- AMEMIYA T., [1975], "Qualitative Response Models" Annals of Economic and Social Measurement, pp. 363-272.
- AMEMIYA T., [1981], "Qualitative Response Models : A Survey", Journal of Economic Literature, (19), pp. 1483-1536.
- CARROLL G.R., [1983], "Dynamic Analysis of Discrete Dependent Variables : A Didactic Essay" Quality and Quantity (17), pp.425- 460.
- DIEKMANN A., PREISENDOERFER P., [1988] : "Turnover and Employment Stability in a Large West Germany Company" European Sociological Review.
- ELANDT-JOHNSON R.C., JOHNSON N.L. [1980] : Survival Models and Data Analysis, NEW YORK, Wiley.
- ELBERS C., RIDDER G., [1982], "True and spurious duration dependence : the identifiability of the proportional hazard model", Review of Economic Studies, pp. 403-409.
- FOURNIER Ch., [1989], "Diplôme, trajectoire sociale et activité des jeunes femmes", Travail et Emploi, pp. 36-42.
- FRANK R.H., [1978], "How Long is a Spell of Unemployment ?", Econometrica, (46), pp. 285-302.
- HECKMAN J.J., SINGER B., [1984a], "Econometric Duration Analysis" Journal of Econometrics (24), pp. 63-132.
- HERPIN N., [1990], "La famille à l'épreuve du chômage", Economie et Statistiques, (235), pp. 31-42.
- IM, [1986], "Modèle log-linéaire et modèle de Cox dans l'analyse d'une table de contingence", Revue de Statistique Appliquée, pp. 5-16.
- KIEFER N., NEUMANN R.G., [1979], "An Empirical Job Search Model, with a Test of the Constant Reservation Wage Hypothesis", Journal of Political Economy, pp. 89-108.
- KIEFFER N.M., [1988], "Economic Duration Data and Hazard Functions" Journal of Economic Literature, (26), pp. 646-679.

- KOP J.L., ALLEGREZZA-CARVOYEUR L.S., [1991], Les variables d'enseignement et les variables connexes des deux premières vagues du panel luxembourgeois. Détection des anomalies et propositions de corrections, Document FEE, CEPS/INSTEAD, LUXEMBOURG, 21 p.
- LANCASTER T., [1979], "Econometric Methods for the Duration of Unemployment" Econometrica, (47)
- MCCALL J.J., [1970], "Economics of Information and Job Search", Quarterly Journal of Economics, (84), pp. 113-126.
- MORTENSEN D.T., [1986], "Job Search and Labor Market Analysis" in O. ASHENFELTER et R. LAYARD (eds), pp. 849-920.
- NICKELL S., [1979], "Estimating the Probability of Leaving Unemployment", Econometrica, pp. 1249-1266.
- RAY J.C., CARVOYEUR L.S., JEANDIDIER B., [1986], Transferts sociaux et modes de cohabitation. Le cas des femmes ayant des enfants à charge., Rapport de recherche pour le Commissariat Général du Plan et la MIRE, 376 p.
- RAY J.C., LIMAN TINGUIRI K.M., CARVOYEUR L.S., [1983], API et désincitation au travail., Rapport pour le Commissariat Général du Plan, 259 p.
- REES A., WATTS H.W., [1975], "An Overview of the Labor Supply Results", in J.A. PECHMAN et P.M. TAMPANE (eds), pp. 69-87.
- SAPSFORD D., TZANNATOS Z., [1990] "Labor Economics : An Overview of Some Recent Theoretical and Empirical Developments" in Current Issues in Labour Economics D.SAPSFORD et Z. ZANNATOS Z. (eds), pp. 1-7.
- SINGER B., HECKMAN J.J. (eds), [], Longitudinal Studies of the Labor Market, New York : Academic Press.
- TARONE R.E., WARE J., [1977], "On distributional free tests for equality of survival distributions" Biometrika, (64), pp. 156-160.
- TOIKKA R., [1976], "A Markovian Model of Labor Market Decisions by Workers" American Economic Review, (66), pp. 821-834.
- TOURNOIS J.
- TOURNOIS J.
- TUMA N.B., HANNAN M.T., [1984], Social Dynamics. Models and Methods, ORLANDO : Academic Press, 578 p..

**ANNEXES DU DOCUMENT PETRA : DEFINITION DES VARIABLES ET
FREQUENCES**

Cette annexe se propose de compléter la définition des variables qui a été faite dans le premier chapitre ; nous insistons surtout sur les variables qui ont donné lieu à des recodages en classes. On trouvera également les statistiques descriptives pour chacune des variables ; tableau de fréquences ou statistiques univariées de base selon la nature de la variable.

1. Les variables dépendantes

CENS1 et CENS2 : variables de censure correspondant respectivement aux modèles 1 et 2. Elles indiquent en fait si les individus ont trouvé en emploi pendant la période d'observation.

CENS1 INDICATEUR DE CENSURE MODELE 1

Value Label	Value	Frequency	Percent
CENSURE (n'a pas trouvé d'emploi)	.00	12	15.4
NON CENSURE (a trouvé un emploi)	1.00	66	84.6
		78	100.0
Total			

CENS2 INDICATEUR DE CENSURE MODELE 2

Value Label	Value	Frequency	Percent
CENSURE (n'a pas trouvé d'emploi)	.00	11	14.1
NON CENSURE (a trouvé un emploi)	1.00	67	85.9
		78	100.0
Total			

DUR1 et DUR2 : durées de non emploi correspondant respectivement aux modèles 1 et 2.

DUR1 DUREE DE NON EMPLOI EN MOIS : MODELE 1

Value Label	Value	Frequency	Percent
	.00	29	37.2
	1.00	18	23.1
	2.00	3	3.8
	3.00	1	1.3
	4.00	7	9.0
	5.00	2	2.6
	6.00	3	3.8
	7.00	4	5.1
	8.00	11	14.1
		-----	-----
	Total	78	100.0
Mean	2.551	Median	1.000
Mode	.000	Std dev	3.018
Minimum	.000	Maximum	8.000

DUR2 DUREE DE NON EMPLOI EN MOIS : MODELE 2

Value Label	Value	Frequency	Percent
	.00	45	57.7
	1.00	10	12.8
	2.00	2	2.6
	3.00	2	2.6
	4.00	3	3.8
	5.00	2	2.6
	6.00	1	1.3
	8.00	6	7.7
	9.00	1	1.3
	12.00	6	7.7
		-----	-----
	Total	78	100.0
Mean	2.269	Median	.000
Mode	.000	Std dev	3.775
Minimum	.000	Maximum	12.000

2. Les variables explicatives

2.1 Les variables au niveau individuel

2.1.1 Les variables démographiques

a) le sexe

SEXE

Value Label	Value	Frequency	Percent
MASCULIN	1	43	55.1
FEMININ	2	35	44.9
		-----	-----
	Total	78	100.0

b) l'âge des individus en 1985

AGE

Value Label	Value	Frequency	Percent
	14.00	7	9.0
	15.00	5	6.4
	16.00	7	9.0
	17.00	5	6.4
	18.00	16	20.5
	19.00	8	10.3
	20.00	11	14.1
	21.00	9	11.5
	22.00	2	2.6
	23.00	3	3.8
	24.00	1	1.3
	26.00	1	1.3
	29.00	2	2.6
	38.00	1	1.3
		-----	-----
	Total	78	100.0

Mean	18.949	Median	18.000
Mode	18.000	Std dev	3.783
Minimum	14.000	Maximum	38.000

N.B. Dans certaines analyses nous avons utilisé :

- l'âge élevé au carré : AGE2
- l'âge élevé au cube : AGE3
- l'âge recodé en classe : CLAGE2

Value Label	Value	Frequency	Percent
moins de 16 ans	1.00	19	24.4
de 17 à 20 ans	2.00	40	51.3
plus de 20 ans	3.00	19	24.4
		-----	-----
	Total	78	100.0

c) L'ancrage des individus

PIC100

Mean	94.038	Median	100.000
Mode	100.000	Std dev	13.528
Kurtosis	17.301	Skewness	-3.698
Minimum	13.000	Maximum	100.000
Valid cases	78	Missing cases	0

2.1.2 Les variables de scolarité en 1985

a) l'année d'enseignement

VAR01 ANNEE SCOLAIRE 85 RECODEE

Value Label	Value	Frequency	Percent
	7.00	2	2.6
	8.00	2	2.6
	9.00	6	7.7
	10.00	8	10.3
	11.00	10	12.8
	12.00	20	25.6
	13.00	8	10.3
	14.00	4	5.1
	15.00	9	11.5
	16.00	3	3.8
	17.00	5	6.4
	18.00	1	1.3
	Total	78	100.0

Mean	12.308	Median	12.000
Mode	12.000	Std dev	2.509
Minimum	7.000	Maximum	18.000

Dans certaines analyses, nous avons utilisé :

- l'année de scolarité élevée au carré : VARCAR
- l'année de scolarité recodée en classe : CLANNE

Value Label	Value	Frequency	Percent
< 12 années	1.00	28	35.9
12-13 années	2.00	28	35.9
> 13 années	3.00	22	28.2
	Total	78	100.0

b) le type d'enseignement en 1985

VAR02 TYPE ENSEIGNEMENT 85

Value Label	Value	Frequency	Percent
NE SAIT PAS	-9.00	3	3.8
ENSEIGNEMENT SECONDAIRE	6.00	4	5.1
LYCEE TECHNIQUE 1	7.00	12	15.4
LYCEE TECHNIQUE 2	8.00	16	20.5
LYCEE TECHNIQUE 3	9.00	9	11.5
ENS. COMPLEMENTAIRE	10.00	8	10.3
COIP	11.00	6	7.7
SUPERIEUR NON UNIVERSITAIRE	13.00	12	15.4
SUPERIEUR UNIVERSITAIRE	14.00	8	10.3
		-----	-----
Total		78	100.0

Pour certaines analyses, nous avons utilisé une variable binaire qui oppose l'enseignement général (enseignement secondaire et supérieur universitaire) à l'enseignement technique (lycée technique, complémentaire, COIP, supérieur non universitaire).

TYPENS TYPE ENSEIGNEMENT RECODE

Value Label	Value	Frequency	Percent
enseignement général	1.00	12	15.4
enseignement technique	2.00	66	84.6
		-----	-----
Total		78	100.0

N.B. Les données manquantes ont été recodées au mode.

c) l'apprentissage

Nous avons utilisé une variable dichotomique de situation par rapport à l'apprentissage.

APPRENTI

Value Label	Value	Frequency	Percent
non apprenti	.00	60	76.9
apprenti	1.00	18	23.1
		-----	-----
Total		78	100.0

2.2 Les variables au niveau du ménage

2.2.1 Les variables de composition du ménage

a) le nombre total de personnes dans le ménage

VCM002 NOMBRE TOTAL DE PERSONNES DANS LE MENAGE

Value Label	Value	Frequency	Percent
	1	1	1.3
	2	2	2.6
	3	13	16.7
	4	25	32.1
	5	15	19.2
	6	14	17.9
	7	2	2.6
	8	2	2.6
	9	3	3.8
	11	1	1.3
		-----	-----
	Total	78	100.0
Mean	4.756	Median	4.000
Mode	4.000	Std dev	1.730
Minimum	1.000	Maximum	11.000

b) le nombre de personnes avec un emploi dans le ménage

NBEMP nombre de personnes avec emploi dans le ménage

Value Label	Value	Frequency	Percent
	.00	6	7.7
	1.00	34	43.6
	2.00	24	30.8
	3.00	9	11.5
	4.00	4	5.1
	5.00	1	1.3
		-----	-----
	Total	78	100.0

2.2.2 La situation géographique des ménages par région

Deux variables ont été créées ; la première distingue entre la région du centre et les autres, la deuxième entre la région du nord-est et les autres.

CENTRE

Value Label	Value	Frequency	Percent
n'habite pas la région centre	.00	43	55.1
habite la région centre	1.00	35	44.9
		-----	-----
Total		78	100.0

NORDEST

Value Label	Value	Frequency	Percent
n'habite pas le nord-est	.00	53	67.9
habite le nord-est	1.00	25	32.1
		-----	-----
Total		78	100.0

2.2.3 Le statut social du ménage :

la formation achevée du chef de ménage

CHEFINF

Value Label	Value	Frequency	Percent
primaire ou inférieure	.00	41	52.6
supérieure à primaire	1.00	37	47.4
		-----	-----
Total		78	100.0

2.2.4 Le statut économique du ménage

a) le quotient familial système C.E.S.

QF1

QUOTIENT FAMILIAL C.E.S. (FL)

Mean	21309.974	Median	20799.000	
Mode	7576.000	Std dev	6880.445	
Kurtosis	.503	Skewness	.305	
Minimum	6903.000	Maximum	41488.000	Valid cases
78	Missing cases	0		

Pour certaines analyses le quotient familial a été recodé en classes.

CLQF1

Value Label	Value	Frequency	Percent
qf supérieur à 14000 FL	.00	67	85.9
qf inférieur à 14000 FL	1.00	11	14.1
		-----	-----
Total		78	100.0

CLQF2

Value Label	Value	Frequency	Percent
qf < 14000 FL ou > 19000 FL	.00	62	79.5
qf entre 14000 et 19000 FL	1.00	16	20.5
		-----	-----
Total		78	100.0

CLQF3

Value Label	Value	Frequency	Percent
qf < 19000 FL ou > 23000 FL	.00	58	74.4
qf entre 19000 et 23000 FL	1.00	20	25.6
		-----	-----
Total		78	100.0

CLQF4

Value Label	Value	Frequency	Percent
qf < 23000 FL ou > 27000 FL	.00	60	76.9
qf entre 23000 FL et 27000 FL	1.00	18	23.1
		-----	-----
Total		78	100.0

CLQF5

Value Label	Value	Frequency	Percent
qf inférieur à 27000 FL	.00	65	83.3
qf supérieur à 27000 FL	1.00	13	16.7
		-----	-----
Total		78	100.0

2.2.5 L'intégration du ménage :

l'hétérogénéité de l'intégration

SD85

Mean	9.641	Median	4.500	
Mode	.000	Std dev	13.742	
Kurtosis	.286	Skewness	1.327	
Minimum	.000	Maximum	44.000	Valid cases
Missing cases	0			78

TABLE DES MATIERES

SOMMAIRE.....	p.1
INTRODUCTION.....	p.2
I - PRESENTATION GENERALE DE L'ETUDE.....	p.5
1) Les caractéristiques générales de l'étude.....	p.6
2) Les définitions utilisées.....	p.6
A - Les individus scolarisés.....	p.6
B - Les apprentis.....	p.8
C - L'emploi	p.8
3) La liste des variables retenues pour l'étude.....	p.9
A - Les variables dépendantes.....	p.9
B - Les variables explicatives.....	p.11
II - ASPECTS METHODOLOGIQUES : LES SOLUTIONS RETENUES POUR ANALYSER LA TRANSITION FORMATION-EMPLOI.....	p.14
1) Les concepts utilisés.....	p.14
A - Le choix du statut.....	p.14
B - les changements d'état.....	p.16
2) Les méthodes d'analyse.....	p.20
A - Le modèle probabiliste dans un cadre statique : une méthode de mesure de la probabilité de trouver un emploi.....	p.20
B - Le modèle probabiliste dynamique ou modèle de durée : une méthode empirique d'analyse de la transition école - emploi, en présence de données censurées...	p.21
3) La mesure de l'adéquation des méthodes d'analyse utilisées.....	p.33
A - Les tests communs à l'analyse de régressions l ogistiques et à l'analyse événementielle.....	p.33
B - Les tests spécifiques à l'analyse de régression logistique.....	p.34