Cahiers du CREAD n°52, 2ème trimestre 2000, pages 61-78.

D. FERROUKHI(*) A. SID(**) L. LAICHE(***)

Modeles de recherche d'emploi des diplômes de l'enseignement superieurs

Première partie : caractérisation de la loi de durée de chômage des sortants de l'université

INTRODUCTION :

Cet article, en deux parties, fait la présentation des modèles de recherche d'emploi.

Dans la première partie, il s'agit de caractériser la loi de la variable de durée du chômage (premier) des diplômés de l'université sous-tendant le phénomène étudié, c'est à dire procéder au choix de la fonction de hazard la plus adaptée aux données dont on dispose.

La deuxième partie de cet article -qui sera publiée dans la prochaine livraison des Cahiers du Cread- procède à la modélisation, en partant des résultats précédents (première partie), du processus de recherche d'emploi en mettant en relation la durée du chômage (premier) et certaines variables individuelles (âge, sexe, CSP, etc...) et scolaires (type de formation, spécialité, etc), une application est réalisée sur les données issues de l'enquête sur l'insertion professionnelle des diplômés de l'enseignement supérieur réalisée par l'Unité de Recherche en Evaluation de la Formation Supérieure de l'INPS.

1. Données longitudinales: leur développement et leur utilisation dans les modèles de recherche d'emploi.

1.1. Constitution de banques de données pour une meilleure compréhension du phénomène d'insertion.

A la fin des années 70 et suite aux deux chocs pétroliers successifs une crise économique structurelle mondiale s'installe et on assiste dans l'ensemble des pays à une dégradation profonde et durable du niveau global de l'emploi et plus particulièrement celui des jeunes. De plus en plus de jeunes, munis ou pas d'un diplôme, éprouvent des difficultés à accéder à un emploi (premier) et la période d'insertion, quasi-instantanée en période de croissance, tend à s'allonger .

Les familles sont de plus en plus préoccupées par le devenir de leurs enfants qui, de mieux en mieux formés (allongement de la scolarité), éprouvent les pires difficultés à accéder à un emploi. Quant aux pouvoirs publics qui, jusque là, adoptaient une position assez neutre dans la régulation du marché de travail, en faisant confiance aux forces (offre et demande) en présence pour le rétablissement de l'équilibre et limitant leur rôle à la préparation de l'offre d'éducation et de formation en élargissant les capacités d'accueil et en favorisant le développement de nouvelles filières plus adaptées à la demande économique, la montée en puissance du chômage des jeunes et sa persistance vont les obliger à intervenir sur le marché du travail pour mieux réguler son fonctionnement.

Des dispositifs lourds sont alors mis en place pour mieux saisir et appréhender ce phénomène nouveau, pour ajuster les politiques les plus appropriées notamment en direction des catégories de population les plus vulnérables. On assiste alors, au cours des années récentes, à la multiplication de dispositifs d'aide à l'insertion et par voie de conséquence à la constitution de différentes sources de données individuelles longitudinales qui rendent possible la mise en oeuvre de nombreux modèles économétriques. La majeure partie des études économétriques fondées sur les données individuelles longitudinales (données de Panel) a été publiée au cours de ces 20 dernières années (Pederson, Westergard-Nielson, 1993). Les méthodes économétriques et statistiques nécessaires à l'obtention d'estimateurs fiables ont été mises au point au cours de cette même période.

1.2. Spécificité de l'analyse économétrique des données de durée.

Les relations théoriques utilisées par l'analyse des données individuelles sont essentiellement dérivées de la théorie relative à la recherche d'emploi. Les études qui utilisent des données individuelles accordent une très large place aux transitions entre différentes situations (ou états) au regard du marché du travail - notamment l'emploi, le chômage, et "hors de la population active", des subdivisions de ces catégories (emplois temporaires et permanents, contrats à durée déterminée (CDD) ou indéterminée (CDI), contrats d'interim, etc.).

La plupart des modèles de recherche d'emploi utilisent comme variable endogène (expliquée) la durée de chômage en relation avec une série de variables observées (individuelles, institutionnelles, conditions locales et/ou régionales du marché du travail, etc.).

L'analyse économétrique des données de durées présente les spécificités suivantes:

i) Une donnée de durée est une variable aléatoire réelle positive indiquant un temps passé dans un état (durée du mariage, durée du chômage) ou encore un intervalle entre deux états (intervalle intergénésique par exemple) (Feroukhi D. et Zemmamouche 1989).

ii) La loi de probabilité d'une variable aléatoire réelle est identifiée, en général, par sa fonction de distribution de probabilité, densité de probabilité ou sa fonction caractéristique. L'interprétation d'une variable en terme de durée va permettre de définir d'autres notions associées, dont les plus fondamentales sont:

- la fonction de "survie" notée S (t)

- la fonction de "hasard" notée H (t)

Ces deux fonctions permettent une représentation commode des durées de "vie" en termes de probabilité. La fonction de survie est, par définition, la probabilité qu'une épisode ne soit accompli à la date t, autrement dit c'est la probabilité que la sortie d'un état donné ou la transition d'un état vers un autre n'ait pas encore lieu. La fonction de hasard se définit comme le taux instantané (ou ² risque ²) de sortie d'un état donné ou de transition entre deux états à la date (t) étant donné qu'il a duré jusqu'à cette date là (t);

***iii*)** L'apport principal de ce type d'analyse consiste en la prise en charge de données incomplètes dites durées "censurées" qui fournissent une information sur la variable d'intérêt (durée de chômage par exemple) et donc sur le phénomène à étudier, sans quoi une sous-estimation des mesures de ce phénomène surviendrait;

iv) Une critique majeure peut être adressée aux modèles de "survie": ce sont des modèles qui supposent le choix "ad-hoc" de la spécification du taux de hasard. Le caractère aléatoire de la spécification choisie peut être renforcé par le choix, à priori, d'une loi statistique. La contribution majeure du travail présenté consiste justement à traiter cette dernière question.

2. Modèle Structurel Et Modèle Réduit De Recherche D'Emploi.

2.1. Introduction.

Un modèle structurel qui décrit la loi d'évolution de la durée du chômage conditionnellement à des variables explicatives devrait en toute logique et conformément à la théorie économique, constituer une forme simple mais cependant suffisamment représentative du marché du travail.

Les variables doivent, par conséquent, intégrer des informations concernant à la fois l'offre et la demande de travail mais aussi les mécanismes régissant la rencontre de ces deux forces sur le marché du travail.

En fait, très peu de modèles économétriques intégrant l'ensemble des facteurs régissant le marché du travail ont été proposés dans la littérature traitant de ce sujet, notamment au sein des recherches entreprises au niveau microéconomique. La plupart des tentatives se sont intéressées à la modélisation de l'offre de travail en étudiant plus particulièrement l'influence (ou pas) de certains facteurs sur la durée passée au chômage ou à la loi de transition entre l'état «chômage» et l'état «emploi» ou un autre état (inactivité par exemple). La loi de la durée passée dans l'état «chômage» est alors déduite d'hypothèses sur les comportements individuels, d'hypothèses élémentaires sur le processus d'arrivée des offres d'emplois ainsi que d'une règle élémentaire d'affection.

2.2. Modèle structurel de recherche d'emploi.

Le modèle structurel de base de recherche d'emploi proposé par Lancaster (1979) est utilisé comme référence dans de très nombreux travaux. Il s'intéresse à la loi de transition entre l'état «chômage» et l'état «emploi».

2.2.1. Description succinte du modèle structurel.

Les opportunités d'embauche se présentent à l'individu selon une séquence temporelle et l'option ouverte à l'intéressé (le chômeur en l'occurrence) est d'accepter l'offre d'emploi qui lui est faite ou de continuer à en chercher un autre. Le critère de choix consiste à maximiser l'espérance mathématique d'une fonction d'utilité propre à l'individu concerné dépendant du revenu (et du loisir éventuellement) sur une période allant du moment où il décide de se mettre à la recherche d'un emploi à un certain horizon (âge de la retraite par exemple).

A chaque instant (t), l'individu arrête un niveau de salaire $\bar{w}(t)$, appelé salaire de réserve, en dessous duquel il rejette les offres d'embauche qui lui sont proposées.

2.2.2. Formulation du modèle structurel.

a- la distribution des salaires offerts est désignée par F(w) (w :salaire);

b- le mécanisme (éventuellement stochastique) gouvernant l'offre d'emploi est connu. On désignera par λ(t).dt la probabilité qu'une opportunité d'embauche se présente au cours de l'intervalle de temps (t,t+dt), cest-à-dire l'offre (continue) d'un emploi obéit à un processus aléatoire.

Le niveau du salaire de réserve arrêté par l'individu en quête d'un emploi implique que la probabilité d'accepter une offre d'embauche au cours du laps de temps (t, t+dt) est :

$$\left(1 - F\left(\bar{w}(t)\right)\right).\lambda(t)dt = \pi(\mathrm{t}).\lambda(\mathrm{t})\mathrm{dt} = \theta(\mathrm{t})\mathrm{dt} \qquad \mathrm{(I)}$$

avec

$$\theta(t) = \left(1 - F\left(\bar{w}(t)\right)\right).\lambda(t) = \pi(t)\lambda(t)$$

Le modèle (I) est appelé modèle structurel de recherche d'emploi et θ (t) est appelée fonction de hasard qui donne, à chaque instant (t) la probabilité de transition de l'état de «chômage» à l'état «emploi». Dans ce modèle structurel, il s'agit d'estimer les fonctions λ et F sur la base des observations de la variable w.

2.2.3. Modèle réduit.

L'économétrie «structurelle» de recherche d'emploi s'appuie sur l'observation des salaires acceptés ou des salaires de réserve. Cependant, la séquence des salaires de réserve $\bar{w}(t)$ n'est pas toujours observable et l'estimation des paramètres du modèle structurel s'en trouve compromise car les probabilités d'accepter une offre et d'accepter cette même offre en dépendent. Une façon naturelle de procéder à la spécification d'un modèle de recherche d'emploi est d'utiliser les données de durée (de chômage) observées modélisant le comportement de la fonction de hasard θ (t) dans l'expression (I). Un modèle «réduit» tentera d'estimer globalement la fonction de hasard θ (t) au vu des durées de chômage. Cette deuxième approche (modèle réduit) permet de mesurer directement diverses variables (descriptives de l'environnement économique, individuelles, institutionnelles, etc.) sur la probabilité conditionnelle de transition.

L'utilisation des méthodes statistiques des données de durée soulève néanmoins plusieurs problèmes importants dont le premier est celui de la spécification de la forme fonctionnelle de la probabilité étudiée et que le prochain chapitre va développer.

3. Caractérisation d'une loi de variables de durée.

Avant de chercher à estimer un modèle économétrique de durée, il faut avoir au préalable une idée sur la distribution des données de durée dont on dispose.

3.1. Estimation non-paramètrique.

On cherche tout d'abord, et dans une première étape, à estimer la fonction de survie, ce qui permet de donner une idée approximative de la loi suivie par les données observées. L'estimateur de Kaplan-Meier est le plus fréquemment usité (Kaplan, Meir, 1958, p.457-481) et possède, sous certaines conditions, des propriétés asymptotiques intéressantes. Lorsque l'on trace la fonction de survie (contre les durées) ainsi que les autres fonctions qui lui sont liées (hasard, hasard intégré et logarithme du hasard intégré) leur allure permet d'avoir une interprétation simple et de se fixer approximativement les idées sur la loi de distribution du phénomène étudié (Kiefer, 1988, p.646-679).

Ainsi, si le graphe associé à la fonction de hasard d'un phénomène de durée est constant et qu'en même temps la fonction de hasard intégré présente une tendance linéaire par rapport au temps, on peut en déduire que la loi associée aux données dont on dispose est une fonction exponentielle.

Il y a lieu de signaler qu'à partir des graphes ainsi obtenus la déduction de la loi sous-tendant le phénomène n'est pas toujours aisée en raison du fait que l'on fait référence dans ce type d'approche, essentiellement aux lois les plus connues (Exponentielle, Weibull, Log-logistique, Log-normale) auxquelles les données présentes n'obéissent pas automatiquement .

3.2. Estimation semi-paramétrique.

A ce niveau de la recherche on suppose que le phénomène étudié obéit à une loi de probabilité donnée dont seul le paramètre θ demeure encore inconnu et qu'il y a lieu d'estimer afin d'arriver à la spécification complète de la distribution des données de durée dont on dispose.

Cette famille de distributions a pu être sélectionnée sur la base de la théorie économique sous-tendant le phénomène étudié, déduite des graphes obtenus dans la phase d'estimation non-paramétrique des données, etc. Dans ce qui suit, nous allons présenter la méthode d'estimation du paramètre θ de la distribution des données de durée.

Elle sera suivie d'une application aux données tirées de l'enquête insertion des diplômés de l'université (Ferroukhi, Fraihat, 1998).

3.2.1. Estimation par le Maximum de Vraisemblance.

Désignons par f(t, θ) la fonction de densité associée à la variable de durée T et spécifiée au paramètre près θ. Si les données sont complètes, c-à-d non censurées, la fonction de vraisemblance associée à l'échantillon s'écrit:

$$L(\theta) = \prod_{i=1}^{n} f(t_i, \theta) \qquad (II)$$

Lorsque certaines données sont censurées (à droite), on construit la vraisemblance de la façon suivante:

- Pour une donnée censurée t_i, affectée de la variable indicatrice $d_i=0$, la seule information dont on dispose est que la durée du séjour dans l'état considéré est au moins égale à t_i. Aussi, la contribution à la vraisemblance à partir de cette observation est la valeur de la fonction de survie $S(t_i, \theta)$, c-à-d la probabilité que la durée du séjour est supérieure à t_i ;

- Pour les éléments dont la durée est non censurée affectés de la variable indicatrice (d_i =1), la contribution à la vraisemblance est, évidemment, $f(t_i, \theta)$.

Ainsi, la vraisemblance associée à l'échantillon s'écrit:

$$L = \prod_{i=1}^{n} f(t_i)^{di} . S(t_i)^{1-di} \qquad (III)$$

La log-vraisemblance s'écrit alors :

$$LogL(\theta) = \sum_{i=1}^{n} d_i Log(f(t_i)) + \sum_{i=1}^{n} (1-d_i) Log(S(t_i, \theta)) \qquad (IV)$$

Cette dernière expression s'écrit, en tenant compte des relations qui lient S, h et Λ (1) :

$$LogL(\theta) = \sum_{i=1}^{n} d_i Log(\lambda(t_i, \theta)) - \sum_{i=1}^{n} \Lambda(t_i, \theta) \qquad (V)$$

Il s'agit d'estimer le paramètre θ par le maximum de vraisemblance

3.2.2. Méthodes de maximisation.

La maximisation de la Log-vraisemblance consiste à trouver la valeur θ^*

qui annule la dérivée première (condition nécessaire pour un optimum local), à savoir:

$$\frac{\partial Log L(\theta^*)}{\partial \theta} = 0 \qquad \text{(VI)}$$

i) Méthode de Newton-Raphson (Gourieroux, 1984):

La méthode la plus usitée est celle de Newton-Raphson. C'est une méthode itérative qui consiste, à partir d'une valeur donnée (θ_0) de départ du paramètre θ, à améliorer la solution à partir de la relation suivante:

$$\theta_{m+1} = \theta_m - \left[\frac{\partial^2 Log L(\theta)}{\partial \theta \partial \theta'}(\theta_m)\right]^{-1} . \frac{\partial Log L(\theta)}{\partial \theta}(\theta_m) \qquad \text{(VII)}$$

D'autres algorithmes dérivés de la précédente relation ont été développés .

ii) Méthode du score (Taheshi Amémiya, 1985):

Dans la formule de récurrence de Newton-Raphson (VI), la matrice $\left[-\frac{\partial^2 \mathrm{Log L}(\theta)}{\partial \theta \partial \theta'}\right]$ est remplacée par son espérance mathématique

E $E\left[-\frac{\partial^2 \mathrm{Log L}(\theta)}{\partial \theta \partial \theta'}\right]$ pour donner:

$$\theta_{m+1} = \theta_m + E\left[-\frac{\partial^2 Log L(\theta)}{\partial \theta \partial \theta'}\right]^{-1} . \frac{\partial Log L(\theta)}{\partial \theta} \qquad \text{(VIII)}$$

iii) Méthode de Brendt-Hall-Hall-Hausman (Taheshi Amémiya, 1985):

Pour passer de l'algorithme de Newton-Raphson à la méthode du score, on substitue à $E\left[-\frac{\partial^2 \mathrm{Log(L)}}{\partial \theta \partial \theta'}\right]$ l'expression suivante: $\left[-\frac{\partial^2 \mathrm{Log L}(\theta)}{\partial \theta \partial \theta'}\right]$. Les résultats de la statistique mathématique nous permettent ainsi d'écrire, et sous certaines conditions, que:

$$E\left[-\frac{\partial^2 Log(L)}{\partial\theta\partial\theta'}\right]=E\left[\frac{\partial LogL}{\partial\theta}.\frac{\partial LogL}{\partial\theta'}\right] \qquad \text{(IX)}$$

ce qui aboutit à l'algorithme suivant:

$$\hat{\theta}_{m+1}=\hat{\theta}_m+E\left[\frac{\partial LogL(\theta)}{\partial\theta}.\frac{\partial LogL(\theta)}{\partial\theta}\right]^{-1}.\left[\frac{\partial LogL(\theta)}{\partial\theta}\right] \qquad \text{(X)}$$ avec,

comme approximation à $E\left[\frac{\partial LogL(\theta)}{\partial\theta}.\frac{\partial LogL(\theta)}{\partial\theta'}\right]$ l'expression suivante:

$$\left[\frac{\partial LogL(\theta)}{\partial\theta}.\frac{\partial LogL(\theta)}{\partial\theta'}\right]$$

La convergence de ces algorithmes demeure tributaire de:

- la concavité de la Log-vraisemblance;

- du point de départ: la vitesse de convergence et la qualité de la solution (locale ou globale) dépendent en grande partie du choix de la valeur initiale θ_0.

Sous certaines conditions, l'estimateur du maximum de vraisemblance est asymptotiquement convergent et normal, de variance asymptotique:

$$V(\hat{\theta}_m)=-\left[\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right]^{-1}_{\hat{\theta}_m} \qquad \text{(XI)}$$

($\hat{\theta}_m$: estimateur du Maximum de vraisemblance)

A partir de cette loi asymptotique, des tests d'hypothèses peuvent être effectués sur le paramètre θ.

3.2.3. Cas particulier de quelques lois.

La démarche précédente a été appliquée à l'estimation des paramètres de quelques lois d'une très large utilisation dans la procédure de modélisation des données de durée. Nous allons présenter quelques unes d'entre elles en explicitant, par la même occasion, les dérivées premières et secondaires nécessaires à la mise en oeuvre de l'un des algorithmes présentés plus haut.

i) Fonction exponentielle:

Elle est caractérisée par un seul paramètre (δ) avec pour:

- fonction de hasard : $\lambda(t,\delta)=\delta$

- fonction de hasard cumulée: $\Lambda(t,\delta)=\delta t.$

En remplaçant λ et Λ par leur expression dans la relation (II) de la Log vraisemblance, on obtient la formule suivante:

$$LogL(\delta) = \sum_{i=1}^{n} d_i Log(\delta) - \delta \sum_{i=1}^{n} t_i \qquad (XII)$$

D'où la dérivée première:

$$\frac{\partial LogL(\vec{\delta})}{\partial \delta} = \frac{1}{\delta} \cdot \sum_{i=1}^{n} d_i - \sum_{i=1}^{n} t_i = 0 \qquad (XIII)$$

On en déduit directement l'estimateur de $\hat{\delta}$:

$$\hat{\delta} = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i} \qquad (XIV)$$

Ainsi donc, cet estimateur $\hat{\delta}$ peut être calculé sans avoir recours à l'algorithme de Newton Raphson, ce qui n'est pas le cas pour les deux prochaines autres lois (Weibull, Log Logistique) que nous allons présenter.

ii) Fonction de Weibull:

Cette fonction est caractérisée par deux paramètres $\theta = (\delta, \alpha)$ et dont le hasard et le hasard intégré sont:

- $\lambda(t, \theta) = \delta \alpha \; t^{\alpha - 1}$.

- $\Lambda(t, \theta) = \delta \, t^{\alpha}$.

En substituant λ et Λ par leur expression dans la relation (II) de la Log vraisemblance, on obtient:

$$LogL(\delta,\alpha) = \sum_{i=1}^{n} d_i Log(\delta) + \sum_{i=1}^{n} d_i Log(\alpha) + (\alpha - 1)\sum_{i=1}^{n} Log(t_i) \cdot d_i - \delta \sum_{i=1}^{n} t_i^{\alpha} \qquad (XV)$$

Les dérivées premières s'écrivent alors:

$$\begin{cases} \dfrac{\partial LogL(\delta,\alpha)}{\partial \delta} = \dfrac{1}{\delta} \displaystyle\sum_{i=1}^{n} d_i - \sum_{i=1}^{n} t_i^{\alpha} = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad (XV - a) \\ \dfrac{\partial LogL(\delta,\alpha)}{\partial \alpha} = \dfrac{1}{\alpha} \displaystyle\sum_{i=1}^{n} d_i + \sum_{i=1}^{n} d_i Log(t_i) - \delta \sum_{i=1}^{n} (Log(t_i)) t_i^{\alpha} = 0 \end{cases}$$

Les dérivées secondes sont les suivantes:

$$\begin{cases} \dfrac{\partial LogL(\delta,\alpha)}{\partial\delta^2} = -\dfrac{1}{\delta^2}\sum_{i=1}^{n} d_i \\ \dfrac{\partial LogL(\delta,\alpha)}{\partial\alpha^2} = -\dfrac{1}{\alpha^2}\sum_{i=1}^{n} d_i - \delta\sum_{i=1}^{n}\left(Log(t_i)\right)^2 t_i^{\alpha} \qquad (XV-b) \\ \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\delta\partial\alpha} = \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\alpha\partial\delta} = -\sum_{i=1}^{n}\left(Log(t_i)\right)t_i^{\alpha} \end{cases}$$

***iii*)** Fonction Log-Logistique :

Deux paramètres δ et α caractérisent cette famille:

$\theta = (\delta, \alpha)$

- $\lambda(t, \theta) = (\delta \alpha t^{\alpha-1}) / (1 + \delta t^{\alpha})$

- $\Lambda(t, \theta) = Log(1 + \delta t^{\alpha})$.

L'expression (II) devient, dans ce cas précis:

$$LogL(\delta,\alpha) = Log(\delta)\sum_{i=1}^{n} d_i + Log(\alpha)\sum_{i=1}^{n} d_i + (\alpha-1)\sum_{i=1}^{n} d_i Log(t_i) - \sum_{i=1}^{n} Log(1+\delta t_i^{\alpha})d_i - \sum_{i=1}^{n} Log(1+\delta t_i^{\alpha}) \qquad (XVI)$$

Les dérivées premières s'écrivent alors:

$$\begin{cases} \dfrac{\partial LogL(\delta,\alpha)}{\partial\delta} = \dfrac{1}{\delta}\sum_{i=1}^{n} d_i - \sum_{i=1}^{n}(1+d_i)\cdot\dfrac{t_i^{\alpha}}{1+\delta(t_i)^{\alpha}} \\ \qquad\qquad\qquad\qquad\qquad\qquad (XVI-a-) \\ \dfrac{\partial LogL(\delta,\alpha)}{\partial\alpha} = \dfrac{1}{\alpha}\sum_{i=1}^{n} d_i + \sum_{i=1}^{n} d_i Log(t_i) - \delta\sum_{i=1}^{n}(1+d_i)\dfrac{Log(t_i)\cdot t_i^{\alpha}}{\left(1+\delta(t_i)^{\alpha}\right)^2} \end{cases}$$

et les dérivées secondes :

$$\left\{\begin{array}{l} \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\delta^2} = -\dfrac{1}{\delta^2}\sum_{i=1}^{n} d_i \\ \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\alpha^2} = -\dfrac{1}{\alpha^2}\sum_{i=1}^{n} d_i - \sum_{i=1}^{n}\left[\delta(1+d_i).(Log(t_i))^2.(t_i)^{\alpha}\right]/(1+\delta(t_i)^{\alpha})^2 \\ \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\delta\partial\alpha} = \dfrac{\partial^2 LogL(\delta,\alpha)}{\partial\alpha\partial\delta} = -\sum_{i=1}^{n}(1+d_i).(Log t_i).(t_i^{\alpha}/(1+\delta(t_i^{\alpha}))^2) \end{array}\right. \qquad \textbf{(XVI-b-)}$$

Les dérivées premières «F» et secondes « J » pour les 2 lois (Weibull et Log Logistique) ont été introduites dans l'algorithme de Newton Raphson. Le tableau (I) donne la valeur des estimateurs de chacune des situations étudiées (par établissements et pour certaines filières) alors que les figures (I) à (III) retracent les courbes associées aux fonctions de distributions empiriques des données de durée de recherche d'emploi déduites de l'enquête «insertion professionnelle des diplômés de l'enseignement supérieur» (Ferroukhi, Fraihat, 1998).

4. Spécification de la loi des durées de chômage des diplômés de l'enseignement supérieur.

4.1. Test non-paramétrique d'une loi de probabilité.

Le but de cette partie est de tenter de dégager la loi suivie par chacune des séries de données de durée de recherche d'emploi des sortants des différents établissements retenus dans l'étude .Dans un premier temps, nous avons sélectionné une famille de trois lois parmi les plus usitées dans les modèles de recherche d'emploi. On procède par la suite à l'estimation de leurs paramètres respectifs à partir des données de durées de chômage des primo-demandeurs d'emploi(2). Il s'agit maintenant de savoir si la fonction empirique liée aux données s'ajuste à l'une des lois classiques présentées précédemment.

4.2. Test non-paramétrique et fonction de répartition: Définitions.

4.2.1. Un test non-paramétrique.

C'est un test d'hypothèses pour lequel il n'est pas nécessaire de spécifier la forme de la distribution de la population étudiée. Il exige généralement que les observations soient indépendantes les unes des autres.

4.2.2. Généralité sur la fonction de répartition (Lebart, Fénelon, p.157-158)

Considérons un échantillon tiré au hasard (x_1 , x_2 , ... , x_n) d'une variable continue **x** de densité f (x). Rappelons que la fonction de répartition de **x** est définie par:

$$F(x) = \int_{-\infty}^{x} f(u)\,.\,du.$$

On définit la fonction de répartition empirique liée à l'échantillon par:

$F_n(x)$ = {proportion des valeurs de $x_1, x_2, \ldots, x_n$

qui sont inférieures ou égales à x. }

$F_n(x)$ est une fonction en escalier présentant «n» sauts d'amplitude «1/n», se produisant aux n valeurs x_i observées.

Pour chaque variable x_i de l'échantillon, la fonction de répartition est :

$$\begin{cases} F(X) = P(X_i \leq X) \\ 0 < F(X) < 1 \end{cases}$$

Appelons succès l'événement pour lequel $x_i \leq x$ pour x fixé, alors F (x) est la probabilité d'un succès, et le nombre total de succès $n F_n(x)$ dans l'échantillon suit une loi binomiale de paramètre n et P = F (x).

On a, par conséquent:

$$E(F_n(x)) = F(x).$$

$$Var(F_n(x)) = \frac{F(x)[1-F(x)]}{n}$$

Par convergence des fréquences vers la probabilité (loi faible des grands nombres) ceci implique qu'en tout point x fixé $F_n(x)$ converge en probabilité vers F (x).

$$-\infty < x < +\infty \quad F_n(x) \xrightarrow{P} F(x).$$

En fait, on peut démontrer un théorème plus fort stipulant la convergence globale en tout point x de $F_n(x)$ vers $F_n(x)$ (Théorème de Glivenko-Cantelli):

$$\text{Si } n \to \infty, \text{Sup} \mid F_n(x) - F(x) \mid \xrightarrow{P=1} 0 \qquad \text{(XVII)}$$

Ainsi pour n assez grand, une fonction de répartition empirique doit donner une image fidèle de la répartition théorique, et $F_n(x)$ peut être utilisée comme estimation de F (x) par la quantité sup $|F_n(x) - F(x)|$:

Notons que cette quantité sup $|F_n(x) - F(x)|$ est une variable aléatoire quand on considère $F_n(x)$ comme fonction aléatoire de l'échantillon $(x_1, x_2, \ldots, x_n)$.

4.3. Test d'ajustement de Kolmogorov - Smirnov (Lebart, Fénelon)

.

Le test d'ajustement de Kolmogorov est un test non paramétrique en présence d'un seul échantillon aléatoire tiré d'une population dont la fonction de répartition inconnue est F (x).

Son but est de déterminer si la fonction de répartition inconnue F (x) est en fait une fonction de répartition spécifique connue F_o (x).

Soit F_o (x), la fonction de répartition d'une variable aléatoire continue X, et soit F_n la fonction de répartition empirique correspondant à un échantillon au hasard (x_1 , x_2 ,...,x_n).

On définit les statistiques suivantes :

$$D_n^+ = \sup_x [F_n(x) - F_o(x)]$$

$$D_n^- = \sup_x [F_n(x) - F_o(x)] \qquad \text{(XVIII)}$$

$$D_n = \sup_x | F_n(x) - F_o(x) | = \text{Max} (D_n^+, D_n^-).$$

Avec les justifications intuitives fournies par le théorème de Glivenko - Cantelli, on imagine que les écarts maxima entre la courbe continue F_o (x) et la courbe en escalier F_n (x) pouvant servir à tester l'hypothèse nulle suivante:

(H_o): La distribution hypothétique de x a pour répartition F_o (x) ou, autrement dit: F (x) = F_o (x) pour tout x.

L'alternative (donc la statistique) étant:

(H_1 Unilatéral) F (x) > F_o (x) pour toute valeur de x (testée avec la statistique D_n^+)

(H_1 Unilatéral) F (x) < F_o (x) pour toute valeur de x (testée avec la statistique D_n^-)

(H_1 Unilatéral) F (x) P F_o (x) pour au moins une valeur de x (testée avec la statistique D_n).

La règle de décision est la suivante : :
Rejeter H_o au seuil de signification (α) si D_n (ouD_n^+ ou D_n^-) est supérieur à la valeur de la table de Kolmogorov- Smirnov ayant pour paramètre n et (1- α) c.à.d si D_n (D_n^+ ouD_n^-) > $t_{n,(1-\alpha)}$.

Les résultats sont consignés dans les trois tableaux (I) à (III) qui suivent :

Tableau 1 : Loi spécifique à chaque établissement.

Etablis.	Nbre d'obs.	Les lois de probabilité	D_n calculé $D_n = Max (D_n^+, D_n^-)$	D_n (Tabulé) α=5%	D_n (Tabulé) α=1%	Résultat du test
USTHB	341	Exponentielle	0.10685	0.073648	0.0882694	Weibull **
		Weibull	0.092581			
		Log Logistique	0.10789			
Blida	104	Exponentielle	0.16838	0.1333589	0.159834	Log Logistique*
		Weibull	0.17953			
		Log Logistique	0.11671			
USTO	113	Exponentielle	0.15063	0.127938	0.1533375	Weibull*
		Weibull	0.043161			Log Logistique**
		Log Logistique	0.12835			

Si D_n> la valeur critique indiquée au tableau on rejette l'hypothèse :
H_0 : La distribution admet F(t) pour fonction de distribution théorique
*   *Acceptée à 5%*   *** Acceptée à 1%*

Tableau 2 : Loi spécifique à quelques spécialités (USTHB).

Filières	Nbre d'obs.	Les lois de probabilité	D_n calculé $D_n = Max (D_n^+, D_n^-)$	D_n (Tabulé) α=5%	D_n (Tabulé) α=1%	Résultat du test
Génie-Civil	115	Exponentielle	0.12322	0.12322	0.1519982	Exponentielle *
		Weibull	0.28849			Log Logistique
		Log Logistique	0.10682	0.10682		*
Biologie	50	Exponentielle	0.11346	0.19	0.23	Exponentielle *
		Weibull	0.26972			Log Logistique
		Log Logistique	0.94826			*
Electronique	73	Exponentielle	0.18089	0.16	0.19	Exponentielle**
		Weibull	0.16535			Log Logistique*
		Log Logistique	0.14391			

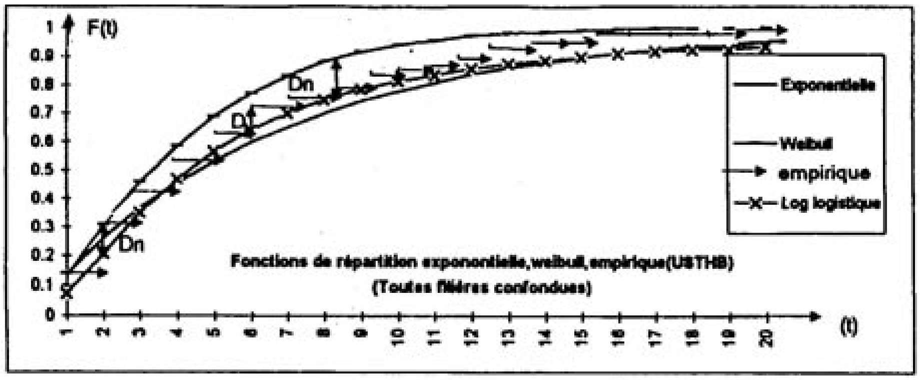
*   *Acceptée à 5%..*   **   *Acceptée à 1%..*

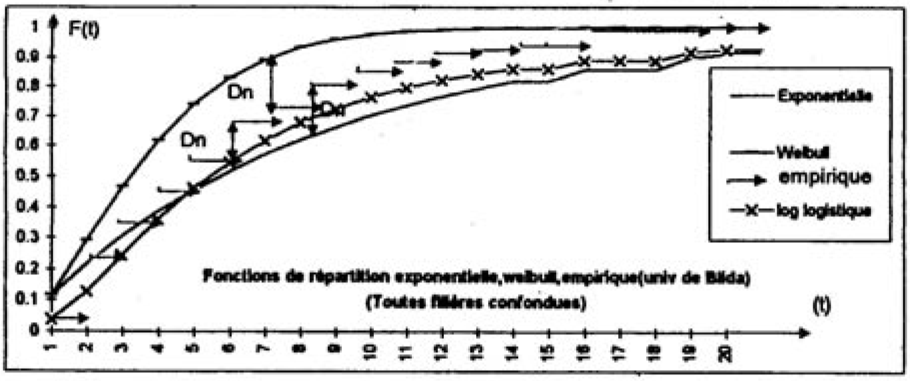
Tableau 3 : Loi spécifique à deux filières partagées par trois établissements (USTHB, BLIDA, USTO).

Spécialités	Nbre d'obs.	Les lois de probabilité	D_n calculé $D_n = Max (D_n^+, D_n^-)$	D_n (Tabulé) α=5%	D_n (Tabulé) α=1%	Résultat du test
Ensemble Génie-civil	87	Exponentielle Weibull Log Logistique	0.18089 0.16535 0.14543	0.15	0.18	Log Logistique *
Ensemble Electronique	149	Exponentielle Weibull Log Logistique	0.13048 0.26629 0.099625	0.1114155	0.1335348	Log Logistique*

*   *Acceptée à 5%*
** *Acceptée à 1%*

Figure I : Fonctions de distrbution par établissement (toutes filières confondues)





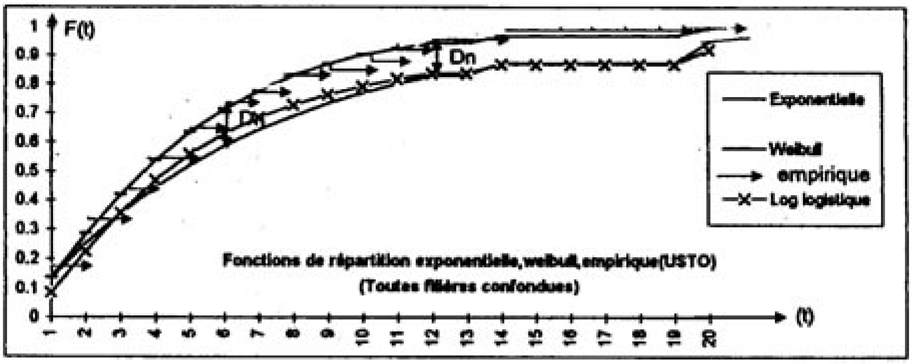
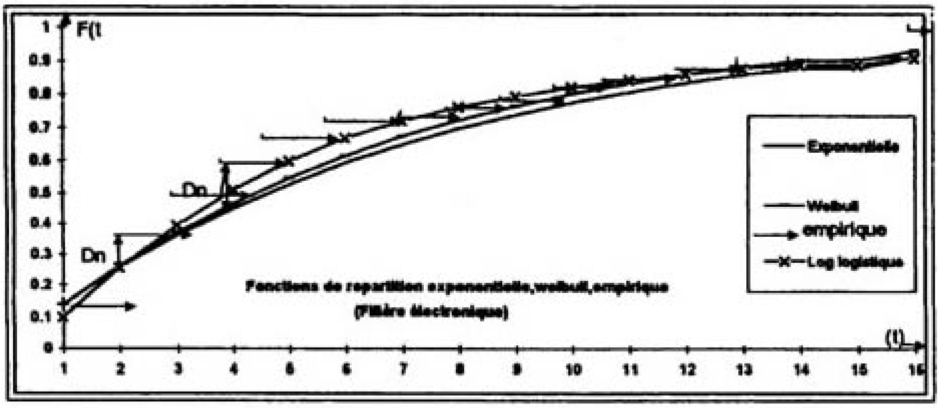
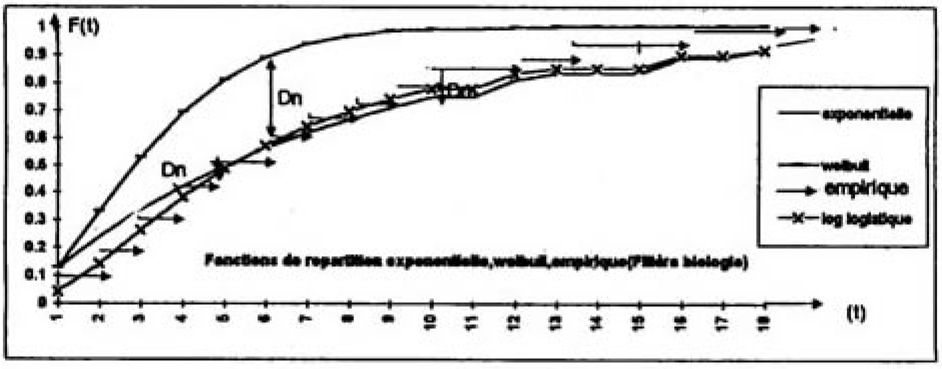


Figure II : Fonctions de distribution pour quelques filières de l'usthb





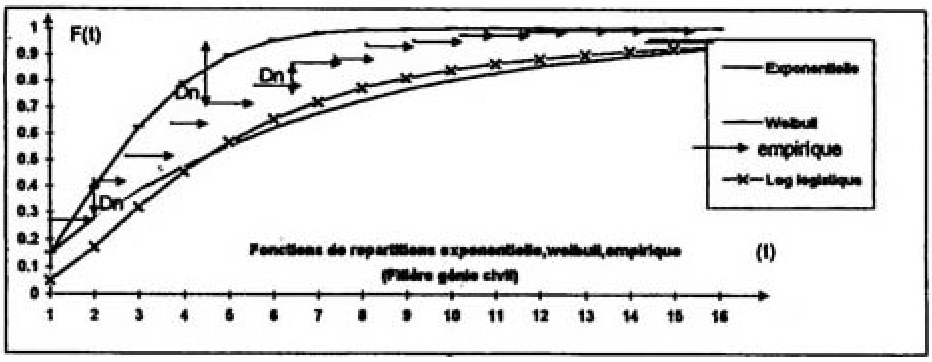
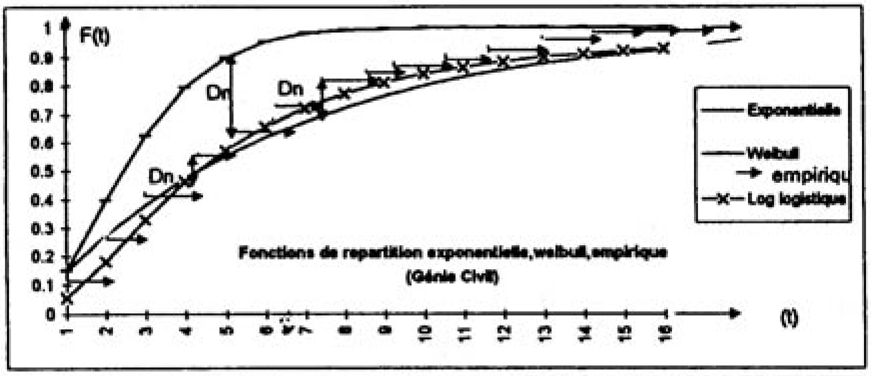
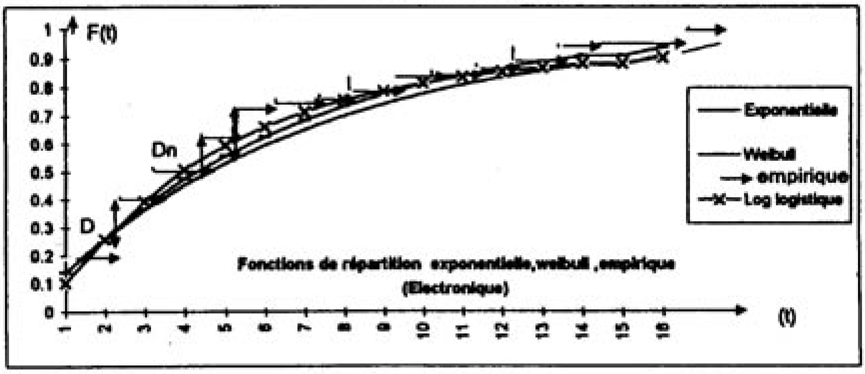


Figure III : tous établissements confondus





Références bibliographiques

FERROUKHI D. & ZEMMAMOUCHE S, 1989. *Econométrie des données de durée appliquée à l'étude de la fécondité*. Collection statistique n°51, ONS, Alger.

FERROUKHI D. & FRAIHAT S, 1998. *Insertion professionnelle des diplômés de l'enseignement supérieur*, Unité de recherche en évaluation. INPS, Alger.

GOURIEROUX C, 1984. *Econométrie des variables qualitatives*, Economica, Paris.

KAPLAN W.J. & MEIER P, 1958. Non paramétric estimation from incomplete observations. *In Journal of the American Statistical Association, 53*.

KIEFER N.M, 1988. Economic duration Data and Hazard Fonction. *In review of Economic littérature, vol XXVI, juin.*

LANCASTER T, 1979. Econométric methods for the duration of unemployement. *In Econometrica, 47 (4).*

LEBART L. & FÉNELON J, 1979. *Traitement des données statistiques*, Dunod, Paris.

PEDERSON P.J. & WESTERGARD-NIELSON N, 1993. Chômage: ce que montrent les données individuelles longitudinales. *In revue économique de l'OCDE, n° 20.*

TAHESHI AMÉMIYA, 1985. *Advanced econometrics.* Blackwell.

Notes

(*)Unité de recherche en évaluation de la formation supérieure, INPS Alger.

(**) Unité de recherche en évaluation de la formation supérieure, INPS Alger.

(***) Unité de recherche en évaluation de la formation supérieure, INPS Alger.

(1) H : fonction de hasard, L : fonction de hasard intégré.

(2) Conditionnellement à l'hypothèse, à chaque fois, que la variable durée suit une de ces lois.