Université de Montréal

Assurance chômage, hasard moral et estimation du modèle de hasard proportionnel via le principe d'entropie généralisée

par

Simon Couture

Département de Sciences Économiques
Faculté des Arts et des Sciences

Rapport de recherche présenté à la Faculté des études supérieures
en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en sciences économiques, option économétrie

Août, 2000

Remerciements

Je tiens d'une part à souligner la source d'inspiration que fût pour moi le sérieux avec lequel Claude Montmarquette, mon directeur de recherche, aborde l'étude des sciences économiques. D'autre part, je remercie la Faculté des études supérieures pour le support financier.

Sommaire

Bien qu'un programme d'assurance chômage ait potentiellement le double effet bénéfique de maintenir un certain niveau de vie à l'individu qui aurait récemment perdu son emploi (pourvu qu'il soit assuré) de même qu'à l'aider financière à supporter les frais de recherches d'un nouvel emploi, l'autre côté de la médaille est, entre autres, le traditionnel problème de hasard moral. Un tel soutien financier risque d'inciter le chômeur à substituer une fraction de son revenu de réemploi au profit des allocations, créant ainsi un effet désincitatif chez le bénéficiaire du programme à se trouver un nouvel emploi et retardant par le fait même sa réinsertion sur le marché du travail.

Sur la base d'études empiriques antérieures, on note qu'il semble effectivement y avoir un tel effet désincitatif à chercher pour du travail, effet qu'on comprend mieux après avoir présenté quelques modèles théoriques. Cependant, le dénominateur commun de ces études est l'arbitraire par lequel les auteurs contrôlent pour l'hétérogénéité non observée dans l'estimation de leur modèle de hasard proportionnel. Cette faiblesse motive le développement d'une nouvelle méthode d'estimation du modèle de hasard proportionnel; il s'agit de l'estimer en employant le principe d'entropie généralisée. Cette avenue offre potentiellement une alternative simple et efficace afin de contrôler pour l'hétérogénéité non observée.

Table des matières

Liste des figures

I- Introduction

Depuis ses premiers balbutiements en 1940, le programme d'assurance chômage (a-c)[1] au Canada s'est mérité des tapes dans le dos, puisqu'il permet aux individus temporairement sans revenu de travail de leur garantir un certain niveau de vie, de même qu'il permet d'éponger les frais associés à la recherche d'un nouvel emploi. Cependant, il a également su se faire taper sur les doigts; pour différentes raisons que nous allons étudier, on a réalisé qu'il génère potentiellement différents coûts pour la société.

Parmi ces coûts, on note le problème de hasard moral. Il semble qu'un individu bénéficiant des allocations d'a-c soit sujet à substituer une fraction de son revenu de réemploi au profit des allocations, créant un désincitatif à la réinsertion sur le marché du travail et augmentant ainsi inutilement la durée des épisodes de chômage. Cependant, Acemoglu et Shimer (1999a,b) montrent que les allocations d'a-c permettent au chercheur d'emploi de trouver un poste dans lequel il est susceptible d'être le plus productif. Or, sur la base d'un modèle théorique calibré pour l'économie américaine par Acemoglu et Shimer (1999b), il semble que ce dernier effet positif sur la productivité de l'appariement employé-employeur domine le coût associé à l'effet désincitatif à la réinsertion sur le marché du travail.

Néanmoins, sur la base d'études empiriques antérieures se consacrant uniquement à détecter la présence d'un effet désincitatif dans les données (donc, sans considérer le gain potentiel de productivité), il semble effectivement que les allocations d'a-c accroissent la durée des épisodes de chômage. Cependant, en passant en revue les études qui nous amènent à cette conclusion, on note une certaine faiblesse quant à la façon dont ils traitent l'hétérogénéité non observée dans l'estimation de leur modèle de hasard proportionnel. Or, cette faiblesse motive le développement d'une nouvelle méthode d'estimation du modèle de hasard proportionnel; il s'agit d'avoir recours au principe d'entropie généralisée. Il s'avère que cette méthode d'estimation a justement le potentiel d'offrir une méthode simple et efficace afin de contrôler pour l'hétérogénéité non observée.

Le reste du rapport se divise en quatre parties. En Partie II, on dresse le portrait du programme d'a-c en contrastant ce qu'il devrait être dans un monde idéal et ce qu'il est en réalité. C'est à ce moment qu'on donne l'intuition derrière différents avantages et inconvénients qu'il génère; la présentation du problème de hasard moral est particulièrement importante pour la suite du rapport. On voit alors intuitivement que les allocations d'a-c ont à la fois le potentiel de créer un désincitatif à la réinsertion sur le marché du travail et un gain de productivité. En Partie III, on formalise, via des modèles de recherche d'emploi et principal-agent, l'intuition donnée en Partie II concernant l'effet désincitatif de même que le gain potentiel de productivité en présentant les papiers de Hopenhayn et Nicolini (1997) et Acemoglu et Shimer (1999b). En Partie IV, on se consacre à l'économétrie de l'a-c, en révisant quelques études empiriques antérieures qui ont vérifié si effet désincitatif il y a. Puisqu'elles ont tous recours à l'économétrie des données de durée.

plus particulièrement au modèle de hasard proportionnel, nous révisons tout d'abord ces méthodes. Par la suite, le survol des études nous laisse entrevoir un certain arbitraire dans la façon dont les auteurs des études empiriques traitent l'hétérogénéité non observée. Je suggère alors une nouvelle méthode d'estimation des modèles de hasard proportionnel employant le principe d'entropie généralisée et qui offre une alternative assez simple et potentiellement efficace afin de contrôler pour cette hétérogénéité non observée. Je conclus en Partie V par un bref résumé et une discussion de directions à prendre dans le futur afin de rendre cette nouvelle méthode d'estimation pleinement opérationnelle.

II- Portrait de l'assurance chômage

La présentation qui suit permet de contraster entre ce que devrait idéalement être un programme d'a-c, section II-1), et ce qu'il est en réalité, section II-2); c'est à ce moment qu'on donne l'intuition derrière la possibilité qu'un programme d'a-c génère des effets désincitatifs à la réinsertion sur le marché du travail. Dans la section II-3), on regarde comment le programme affecte la dynamique sur le marché du travail de façon plus générale.

II-1) Le monde idéal[2]

Qu'on parle de chômage structurel, frictionnel, ou cyclique (et supposons pour les besoins de la cause, saisonnier), le chômage demeure un risque[3], c'est-à-dire qu'il existe une probabilité non nulle qu'un individu perde son emploi et se retrouve sans revenu de travail pour une période de durée non nulle (c'est-à-dire qu'il ne puisse recouvrer un nouvel emploi au même instant qu'il se fait licencier). Or, lorsqu'on parle de risque, on distingue généralement entre deux grandes catégories, soient le risque pur (ex: accident) et le risque spéculatif (ex: jeu de hasard).

Supposons qu'on adhère à l'idée selon laquelle le chômage soit un risque pur, c'est-à-dire que la probabilité de perte d'emploi soit en quelque sorte hors contrôle du travailleur. Face à cette possibilité de perdre son emploi, il peut soit épargner ou avoir recours à une assurance privée, ce afin de <u>*i-* se maintenir un *niveau de vie respectable* le cas échéant, mais surtout *ii- éponger les coûts de recherche d'un nouvel emploi*, contribuant à l'efficacité de résorption des déséquilibres sur le marché du travail</u> (dans la mesure où on perçoit le chômage comme un déséquilibre). L'assurance domine l'épargne pour deux raisons. Premièrement, un individu qui aurait épargné sans avoir subi de perte d'emploi aurait inutilement réduit sa consommation. Deuxièmement, des contraintes sur le marché des capitaux empêchent le parfait lissage de la consommation via l'épargne.

---

[1] Tout au long de ce rapport, j'emploie l'expression "assurance chômage" bien que depuis 1996 au Canada on parle plutôt "d'assurance emploi".

[2] La façon de présenter le problème est inspiré de Corak (1994).

[3] Le chômage saisonnier ne constitue pas un risque; la probabilité de perte de revenu d'emploi est égale à un.

Or, une telle assurance privée contre la perte temporaire de revenu d'emploi devrait assurer un individu contre le risque pur et devrait idéalement être caractérisée par les quatre conditions suivantes:

a) Absence de sélection adverse,

b) Absence de hasard moral,

c) Absence de corrélation entre les probabilités de perte de revenu de travail pour différents individus,

d) Les probabilités de perte sont calculables.

Malheureusement, on va rapidement se rendre compte qu'en pratique, cet idéal fait défaut en tout point.

II-2) La dure réalité

Premièrement, le chômage reflète une part de risque spéculatif. En effet, le choix de filière en éducation, le choix de la localité de résidence, etc... sont des paramètres sur lesquels un individu peut jouer et qui peuvent affecter la probabilité qu'il perde ou non son emploi. À moins de vouloir payer un prix exorbitant, aucun assureur acceptera de couvrir le risque spéculatif.

Deuxièmement, le système d'assurance est loin d'être aussi idéal que celui dont on vient de décrire; en particulier:

(a') La sélection adverse est un problème bien réel. Durant la première année du programme en 1941 (sanctionné en 1940), 42% de la population active en étaient assujettis. Dans ce contexte, la demande des emplois assurés augmente, contribuant ainsi à l'apparition du problème de sélection adverse et aux distorsions dans l'allocation des ressources qui en découle[1].

Or, Chiu et Karni (1998) montrent que le problème de sélection adverse est suffisant pour empêcher le secteur privé d'assurer les revenus d'emploi et justifient ainsi que l'État doit le faire, comme c'est le cas en pratique. Le problème est que dès lors où cela se produit, *i)* émerge le problème traditionnel de non-neutralité des interventions gouvernementales[2] et *ii)* il est fort possible de voir l'objectif d'efficacité du marché du travail passer derrière l'objectif plus général d'assurer un minimum de niveau de vie aux individus temporairement sans revenu d'emploi; en effet, comme on voit un peu plus loin, le gouvernement peut être tenté d'en faire un programme de redistribution des revenus. Ce dernier point prend tout son sens

---

[1] Dans la théorie des *contrats implicites*, on suppose que les employés sont davantage averse au risque à une fluctuation de leur revenu que l'employeur, de sorte qu'il existe un gain d'échange entre les deux partis à partager le risque. Dans ce contexte, la firme peut offrir une "assurance" contre les fluctuations du revenu de ses employés en leur allouant un salaire moyen plus faible, mais stable. Ce faisant, il n'y a plus de sélection adverse, tout les employés se retrouvant ainsi assurés.

[2] Peut-être moins important aujourd'hui car durant ses trente premières années d'existence, le programme d'a-c est financé via un mode tripartie, c'est-à-dire employés, employeurs et gouvernement. Cependant, depuis 1991, le programme est autofinancé; seules les cotisations des employés et employeurs amènent l'eau au moulin.

avec la réforme de 1971 qui mène à la *Loi de l'assurance chômage*; Lazar (1994, 37) observe que "the 1971 revisions moved (...) the UI program away from insurance principles and toward horizontal equity and income support."

Il n'est donc pas surprenant que suite à cette réforme, on observe une couverture presque universelle avec des critères d'admissibilité quasi laxistes en exigeant que 8 semaines d'emploi assurable durant les 52 semaines précédants la demande; le résultat est que 92% de la force de travail a un revenu d'emploi assuré en vertu du programme. Le bon côté de cette histoire est qu'avec un tel taux, le problème de sélection adverse devient pratiquement inexistant. Cependant, le programme prend plus l'allure d'un mode de redistribution de la richesse qu'autre chose, négligeant l'objectif de contribuer à une meilleure efficacité sur le marché du travail, auquel cas peuvent émerger d'autres distorsions, car à première vue il n'est pas le dessein d'un programme d'assurance de redistribuer la richesse; bien qu'un tel objectif puisse être légitime, y parvenir par le truchement du programme d'a-c n'est peut-être pas la meilleure idée. Par exemple, Kesselman (1986, 102) note une certaine prise de conscience à cet effet à l'intérieur du rapport de la Commission Royale sur l'Union Économique et Perspectives de Développement pour le Canada de 1985:

"The Commission rejected a reorientation of UI to make it an explicit redistributive program. (...) It feared that giving the program a greater redistributive cast would exacerbate its labour market distortions and incentives for unstable employment. Conversely, it felt that strengthening the social insurance orientation of UI would promote labour market adjustment and employment stability."

(b') Une autre réalité de l'a-c et qui est également sujette à créer des distorsions dans l'allocation des ressources est le problème de hasard moral, sur lequel nous insistons particulièrement pour les besoins du présent rapport.

Une fois le revenu d'emploi assuré via le programme d'a-c, autant l'employeur que l'employé est susceptible de modifier son comportement de façon à rendre un individu davantage sujet à en devenir et/ou demeurer bénéficiaire. Par exemple, en plus des coûts directs de financement à la caisse du programme qui peuvent nuire à la création d'emplois, une firme chez qui le revenu d'emploi de ses travailleurs sont assurés risque de modifier son comportement d'embauche. Feldstein (1976) a montré comment dans ce contexte l'employeur pouvait avoir recours à répétition aux mises à pied temporaires; il s'agit en quelque sorte d'une subvention qui lui permet de faire face aux périodes creuses[1]. De plus, protégés par l'ancienneté et/ou du capital humain spécifique, la plus part des travailleurs mis à pied savent qu'ils ont de bonnes chances d'être rappelé, suggérant ainsi que la durée et la fréquence des épisodes fassent partie du contrat de travail[2] et donc que le chômage soit de nature volontaire bien que les contractions de la demande agrégée soient involontaire.

[1] C'est pour obvier à ce phénomène que Feldstein (1976) milite en faveur de l'*experience rating*, c'est-à-dire exiger que les firmes chez qui les employés ont souvent recours aux prestations d'a-c cotisent davantage au financement de la caisse du programme.

[2] Un peu comme dans la théorie des contrats implicites, mais sans considérer l'aspect assurance.

Cependant, ce qui nous importe davantage est d'étudier en quoi l'a-c peut modifier le comportement de recherche d'emploi d'un individu sans revenu de travail; on s'attaque ici à son effet désincitatif. À la Partie III, on étudie en détail le modèle de recherche d'emploi, ce qui nous permet de mieux comprendre pourquoi le support financier dispensé par le programme afin de les aider à éponger les coûts associés à la recherche d'emploi (qui a le mérite, comme le montre Gruber (1997), de lisser la consommation du chômeur) implique un tel effet négatif. Mais pour l'instant, mentionnons simplement qu'une hausse du niveau d'allocation d'a-c augmente le salaire de réserve du chercheur d'emploi ce qui le rend plus capricieux face aux salaires qui lui sont offerts tout au long du processus de recherche, diminuant ainsi les probabilités qu'il se trouve un nouvel emploi et augmentant la durée des épisodes de chômage. Le chômage est donc encore une fois volontaire puisque l'individu pourrait, s'il le voulait, choisir la première offre qu'il reçoit[1].

Malgré tout, il ne faut pas voir dans ce qui précède qu'un aspect négatif. En effet, si aider un chômeur à payer les frais de recherche d'emploi accroît la durée des épisodes de chômage, alors on peut intuitivement entrevoir que des périodes plus longues de recherche d'emploi devraient conduire l'employé à se trouver un poste dans lequel il a le meilleur potentiel productif. Comme on le verra plus loin, cette idée est reprise par Acemoglu et Shimer (1999a,b) qui montrent qu'en présence d'individus averses au risque, les allocations d'a-c haussent la productivité et l'output national via une composition d'emplois plus productifs au sein de l'économie. Donc, certes les allocations d'a-c peuvent inciter le chômeur à retarder sa réinsertion sur le marché du travail, cependant elles conduisent potentiellement à un meilleur appariement entre les travailleurs et les firmes.

Ainsi, en regard à (a') et (b'), particulièrement (b'), retenons que le programme d'a-c semble favoriser l'apparition d'un chômage volontaire, volontaire au sens de Feldstein et du modèle de recherche d'emploi. De plus, il est bon de souligner que (a') et (b') est le propre de la plupart des régimes d'assurances privées. Cependant, deux choses intéressantes se produisent lorsqu'on relève les hypothèses (c) et (d); d'une part, le chômage semble alors de nature involontaire, et d'autre part, ce sont deux aspects qu'on ne trouve généralement pas en assurance privée.

(c') Malheureusement, il est fort possible que la perte d'emploi subie par deux individus soit corrélée et même autocorrélée. Par exemple, corrélée entre individus d'une même firme, d'une même industrie, entre classes de la société, etc..., ce en raison d'un choc négatif à la demande agrégée, du progrès technologique ou d'un effet d'hystérésis par exemple. Concernant ce dernier point, Lemieux et MacLeod (1998) estiment que ceux qui bénéficient d'allocations d'a-c ont une plus grande probabilité d'y avoir recours à nouveau dans le futur; Dellas (1997) montre que certains individus vont substituer l'acquisition de capital humain au profit d'allocations d'a-c, ce qui les rend également plus sujets à avoir de nouveau recours au soutien financier du programme dans le futur.

[1] Contrairement à l'idée de Feldstein où c'est l'employeur qui décide de la durée de l'épisode de chômage.

(d') De plus, en raison de l'impossibilité de prévoir les points de retournement du "cycle" d'affaire vu la complexité des phénomènes économiques, il est utopique de penser calculer les probabilités de pertes d'emplois dans le futur.

Ainsi, avec (c') et (d'), le chômage semble être de nature involontaire au sens où un individu peut perdre son emploi vu une contraction de la demande et ce tout en étant disposé à travailler pour un salaire inférieur (les rigidités du salaire nominal empêchant le réemploi à un tel niveau). Or, il s'agit en fait d'une bonne chose puisque, comme on a vu, c'est précisément ce type de chômage qui est souhaitable d'assurer (risque pur). Cependant, deux choses. *D'une part*, il y a tout lieu de se demander si le chômage résultant est vraiment involontaire. Il semble en effet être ardu de développer des fondements micro-économiques d'un chômage involontaire, qu'on pense à la théorie du salaire d'efficience ou celle de l'insider-outsider[1]. Ceci peut sans doute s'expliquer, comme l'écrit Lucas (1978, 354), par le fait que "there is an involuntary element in *all* unemployment, in the sense that no one chooses bad luck over good: there is also a voluntary element in all unemployment, in the sense that however miserable one's current work options, one can always choose to accept them."

D'autre part, que les probabilités de pertes d'emplois soient corrélées entre travailleurs et impossibles à prévoir rend dérisoire le recours à une assurance afin de protéger ses arrières financièrement. Qui serait assez fou pour assurer le revenu d'emploi de son voisin s'il n'a aucune idée de la probabilité que ce dernier perde son travail? Il s'agit en fait de deux autres facteurs qui justifient l'intervention de l'État afin d'assurer l'existence d'un tel programme. Cependant, lui également doit faire avec (c') et (d') ce qui explique pourquoi, comme on a vu en (b'), il pouvait être tenté d'en faire un programme de sécurité de revenu, étant impossible d'assurer le revenu d'emploi au sens propre. Ainsi, il ne peut pas se consacrer uniquement à l'objectif d'efficacité sur le marché du travail et le programme doit avoir un volet sécurité de revenu justifiant les effets redistributifs, la couverture du chômage saisonnier, "volontaire" et du risque spéculatif.

II-3) Impacts sur la dynamique du marché du travail

L'éventail des différents avantages et coûts potentiels découlant du programme d'a-c qu'on vient d'énumérer, bien que non exhaustif, témoigne de la complexité qui entoure la mise en place d'un tel système. On comprend alors qu'il soit périlleux en pratique de fixer les paramètres du programme en essayant de balancer les bons et mauvais côtés; pour s'en convaincre, on a qu'à regarder le nombre de réformes qu'a subit le programme depuis 1940[2].

On mentionne non exhaustif puisque pour les besoins de la cause, on a essentiellement considéré que les transitions chômeur-employé et employé-chômeur. Cependant, partant des trois statuts que peut

[1] Cahuc et Zilberberg (1996)
[2] Voir Dingledine (1980) et Lin (1998) pour un survol historique.

revête un individu sur le marché du travail, soient *i*-employé, *ii*-chômeur ou *iii*-inactif (hors de la population active), on remarque que ces dernières ne sont que deux de $3^2 = 9$ possibilités de transitions. À chaque jour, des individus vont entrer, demeurer ou sortir de chacun des statuts précités (Jones (1993)), transitions <u>interstatuts</u> qui vont cependant être exacerbées vu la présence d'un programme d'a-c[1].

Qui plus est, il y a également beaucoup d'hétérogénéité à l'intérieur de chacun des trois statuts, de sorte qu'il est possible d'assister à des mouvements <u>intrastatuts</u> induits par un programme d'a-c qui échapperont aux transitions interstatuts. Premièrement, le meilleur exemple d'hétérogénéité dans les *emplois* est la théorie de la segmentation du marché du travail (*dual labor market*). De plus, un travailleur peut être soit salarié (sans distinction ici aux différents modes de rémunération) ou travailler à son compte (on sait qu'au Canada, les travailleurs autonomes ne sont pas couverts par le programme d'a-c), etc... Deuxièmement, comme le soulignent Atkinson et Micklewright (1991, 1683), il faut s'entendre sur ce qu'on entend par *chômeur*:

"On the one hand, there are those actively seeking for work who are not eligible for or not claiming benefit; on the other hand, there are those in receipt of benefit who are not regarded as actively seeking for work. And there are "discouraged workers", not currently searching, who are missing from both sets of figure if they do not claim benefit, but who classify themselves as unemployed."

Troisièmement, parmi les *inactifs* qui sont susceptibles de transiter vers un autre statut vu la présence du programme d'a-c, notons les travailleurs de l'économie informelle, les étudiants, les bénévoles et ceux qui demeurent au foyer. Concernant ce dernier groupe, Cullen et Gruber (2000) montrent que, dans l'esprit du travailleur ajouté (*added worker*), il y a effet d'éviction chez l'offre de travail des femmes mariées lorsque le mari est bénéficiaire d'allocations d'a-c.

En plus de cette hétérogénéité propre à chaque statut du marché du travail, il y a bien entendu l'hétérogénéité de la population active. Le sexe, la race, l'âge, l'éducation, etc..., sont tous des attributs qui sont susceptibles d'induire différentes réponses chez les membres de la population suivant la mise en place ou la modification des paramètres du programme d'a-c.

Ayant décrit différents effets du programme d'a-c sur le marché du travail, nous nous concentrons maintenant que sur (b') de la section II-2). En Partie III, on veut en effet formaliser l'intuition qu'on a alors donné en présentant des modèles théoriques qui nous permettent de mieux comprendre, d'une part, les distorsions entraînées par le problème de hasard moral du côté du chômeur, et d'autre part, le gain potentiel de productivité.

---

[1] Voir Clark et Summers (1982) qui estiment les probabilités de transitions (processus de Markov) entre chacun de ces statuts causées par la présence d'un programme d'a-c.

III- Hasard moral et assurance chômage en théorie

L'objectif de cette Partie III est de mieux comprendre deux forces en opposition dont on a donné l'intuition à la section II-2); d'un côté, la présence du programme d'a-c créé un effet désincitatif du côté du chômeur à réinsérer le marché du travail et de l'autre côté, le programme a le potentiel d'accroître la productivité de l'appariement entre travailleurs et employeurs. Étant donné que l'effet désincitatif est le principal objet du présent rapport, les deux prochaines sections III-1) et III-2) le regardent de plus près via respectivement le modèle de recherche d'emploi et le cadre d'analyse principal-agent employé par Hopenhayn et Nicolini (1997). En section III-3), on se tourne vers Acemoglu et Shimer (1999b) qui étudient d'une part le gain potentiel de productivité vu la présence du programme via un modèle statique et d'autre part, discutent d'un modèle dynamique où ils incorporent le problème de hasard moral afin de juger quel effet est l'emporte.

III-1) Modèle de recherche d'emploi

Cette première section sert donc d'introduction théorique à l'étude des programmes d'a-c. Au point III-1.a), on présente les motivations qui ont menées au développement du modèle de recherche d'emploi, modèle qui est présenté au point III-1.b). Au point III-1.c), on discute de deux prédictions du modèle de recherche d'emploi reliées à l'effet désincitatif du programme d'a-c.

III-1.a) *Vers le modèle de recherche d'emploi*

Comme point de départ, il est utile de se placer dans le contexte de la théorie néoclassique d'offre de travail. On suppose ici qu'un individu maximise sa fonction d'utilité (fonction du loisir L et de la consommation C) sous contrainte de ressources et qu'il offre ses services productifs si et seulement si le salaire du marché est supérieur ou égal à son salaire de réserve, ce dernier définit comme étant le rapport des utilités marginales du loisir et de la consommation, évalué au point (L_0, R), où L_0 est la dotation en temps et R correspond aux niveau des ressources autres que le revenu de travail; voir figure 1. Autrement, l'individu est qualifié d'inactif, c'est-à-dire qu'il consacre toute sa dotation de temps au loisir si le salaire du marché, *w*, est inférieur à son salaire de réserve. Sur la figure 1, le salaire du marché (soit la pente de la contrainte budgétaire) est supérieur au salaire de réserve de l'individu et il travaille $(L' - L_0)$.

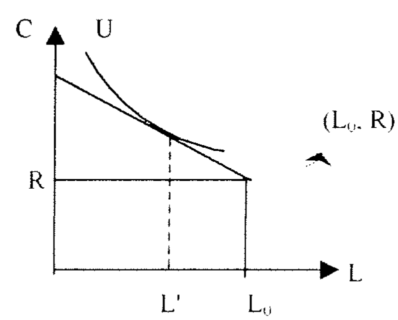


Figure 1: Offre de travail néoclassique

La grande difficulté avec cette théorie est qu'il est impossible d'observer du chômage en raison de cette dichotomie employé/inactif; un peu gênant lorsque c'est ce qu'on cherche à étudier. Un individu sans emploi va vraisemblablement affecter une fraction de sa dotation en loisir à se chercher un emploi. Or, la notion de "chercher" pour du travail dans la théorie néoclassique ne fait aucun sens puisque l'information y est supposée parfaite. Cette situation motive le développement d'une autre théorie d'offre de travail, celle de la théorie de la recherche d'emploi.

Ainsi, on suppose maintenant que l'information concernant les salaires offerts et l'emplacement des firmes offrantes n'est pas connue avec certitude. Dans ce contexte, la première question qui nous vient à l'esprit est de savoir quelle devrait être la méthode optimale pour l'individu au chômage de chercher pour du travail. En fait, Stigler (1962) proposait à l'origine un processus de recherche dit d'*optimal sample size*. Il voyait le travailleur comme sélectionnant un échantillon aléatoire de taille n de salaires offerts, à un coût de c\$ par salaire échantillonné. Le travailleur accepte alors d'offrir ses services pour la firme qui annonce le salaire le plus élevé à l'intérieur de l'échantillon. Le problème du travailleur est donc de choisir la taille optimale n sachant c.

Des développements nous amènent par la suite à emprunter une *stopping rule*, issue de la théorie de décision statistique. On perçoit maintenant le travailleur comme échantillonnant les offres salariales une à la fois et décidant d'arrêter ou de continuer sa recherche sur la base de l'échantillon qu'il aura observé jusqu'à ce moment. Un des premiers à formaliser cette méthode de recherche d'emploi à l'intérieur d'un cadre théorique d'offre de travail est Mortensen dans les années '70; Mortensen (1986) reprend l'essentiel de ses travaux et c'est cette référence qui sert d'inspiration au point III-1.b) dans la présentation du modèle de recherche d'emploi sous sa forme originelle.

III-1.b) *Modèle canonique*

Le principal outil dont on a besoin afin de trouver la règle optimale d'arrêt pour l'individu au chômage est le principe de Bellman[1], principe qui s'appuie sur le *contracting mapping theorem*. Dans le cas qui nous intéresse, ce théorème implique que si l'utilité espérée de revenus pour la prochaine période est bornée et que le taux d'escompte est strictement inférieur à un, alors l'utilité perdue en ignorant des périodes éloignées est inférieure à la somme des extrémités d'une série géométrique. En termes plus simples, ce principe stipule que la décision contemporaine à l'intérieur d'une séquence de décisions maximise les revenus nets contemporains plus les flux anticipés (anticipations conditionnelles à l'information courante) actualisés, sous l'hypothèse que les décisions futures soient optimales. Dans ce contexte, ce principe convertit un problème d'optimisation intertemporel en une séquence de problèmes d'optimisations d'une période.

[1] Voir Mas-Colell, Whinston et Green (1995).

Ainsi, soit $q(n, h)$ la probabilité qu'un chômeur reçoive n offres pendant une période de longueur h, où l'offre est entièrement caractérisée par w, c'est-à-dire le salaire offert. $V^e(w)$ est l'espérance d'utilité d'arrêter de chercher en acceptant w; $V^e(w)$ est aussi égale à w/r, c'est-à-dire la valeur présente du revenu anticipé au salaire w. Si le chômeur accepte l'offre, il demeurera à l'emploi de la firme offrante jusqu'à l'infini et recevra w pour toujours; sinon, il continu à chercher et a une espérance d'utilité V^u. De plus, soit $\beta_h = e^{-rh}$ le facteur d'escompte et $\xi \equiv [b - c]$ les bénéfices nets liés à l'activité de recherche d'emploi, où b et c sont respectivement l'ensemble des bénéfices (allocations d'a-c) et des coûts, constants tout au long de l'épisode de chômage. Donc:

$$V^u(\Omega) = \xi h + \beta_h E\{ \max [V^u(\Omega(t+h)), V^e(x)] \mid \Omega(t) = \Omega \}$$

où, E est l'opérateur d'espérance mathématique, x correspond à la meilleure offre observée durant la prochaine période de longueur h et Ω est l'ensemble d'information disponible au temps t.

La façon dont on traite l'information imparfaite sur le marché du travail dans ce modèle théorique de base est de supposer que le chômeur connaît seulement qu'une fonction de distribution (exogène) de l'ensemble des salaires w offerts sur le marché du travail. En supposant de plus cette fonction invariante sur le plan temporel, il en découle que le chômeur n'acquiert aucune information supplémentaire tout au long de son processus de recherche. Ceci implique que $V^u(\Omega) = V^u(\Omega(t+h))$; l'utilité espérée de continuer à chercher est constante à travers le temps. Soit donc $G(w, n)$ cette probabilité que la meilleure des n offres soit inférieure ou égale à w sachant $n \geq 1$ (on notera $G(w, 1) \equiv F(w)$), alors on trouve que:

$$V^u = \xi h + \beta_h \left[\sum_{n=1}^{\infty} q(n, h) \int_0^{\infty} \max(V^u, V^e(x))\, d\, G(x, n) + q(0, h)\, V^u \right],$$

ou de façon équivalente:

$$(1 - \beta_h)\, V^u = \xi h + \beta_h \left[\sum_{n=1}^{\infty} q(n, h) \int_0^{\infty} \max(0, V^e(x) - V^u)\, d\, G(x, n) \right]. \tag{1}$$

Cette dernière expression démontre qu'il existe une solution unique V^u (pourvu que la moyenne de la distribution des salaires offerts soit finie). Dans ce cas, la stratégie optimale du chômeur satisfait à la *reservation property*, c'est-à-dire qu'au moment où la richesse anticipée du travailleur maximise la stratégie optimale d'arrêt, il arrête de chercher et accepte l'emploi. En d'autres mots, il est optimal d'arrêter de chercher lorsque le salaire offert dans n'importe quelle période donnée (d'où l'utilité du principe de Bellman) est supérieur ou égale à la valeur critique appelée salaire de réserve, noté w^r. À ce salaire, on a donc que $V^e(w^r) = w^r/r = V^u$. Explicitons maintenant cette valeur critique.

Si on admet que les offres arrivent selon un processus Poisson avec paramètre δ, c'est-à-dire $q(n, h) = e^{-\delta h}(\delta h)^n / n!$, alors la probabilité de recevoir n offres dans l'intervalle h est indépendante du temps écoulé depuis la dernière offre reçue. Supposons $n = 1$, on montre que $\lim_{h\to 0} q(1, h)/h = \delta$ et $\lim_{h\to 0} (1 - \beta_h)/h = r$. Donc, en divisant les deux côtés de l'expression (1) par h et en laissant $h\to 0$, on trouve :

$$r\,V^u = \xi + \delta \int_0^\infty \max(0, V^e(x) - V^u)\, d\,F(x)$$

Puisque $V^e(w^r) = w^r/r = V^u$, le salaire de réserve prend la forme finale suivante:

$$w^r = \xi + (\delta/r) \int_{wr}^\infty (x - w^r)\, d\,F(x) \qquad (2)$$

Ce modèle de recherche d'emploi génère également quelques implications en regard à la distribution des durées des épisodes de chômage. Soit λh la probabilité qu'un chômeur se trouve un emploi dans l'intervalle de temps h. Celle-ci est le produit de la probabilité qu'une offre soit reçue dans l'intervalle de temps h et la probabilité que cette offre soit acceptée selon la règle optimale d'arrêt de l'individu; mathématiquement:

$$\lambda h = (\delta h + o(h)) \cdot \int_{w_r}^\infty f(w)\, dw$$

Laissons $h \to 0$ et divisons par h des deux côtés, on obtient:

$$\lambda = \delta \cdot \pi(w^r) \qquad (3)$$

où $\pi(w^r) = \int_{wr}^\infty f(w)\, dw$. Ceci correspond au taux de transition entre le chômage et l'emploi; on verra plus loin que cela correspond au taux de hasard.

III-1.c) *Prédictions*

À l'aide des expressions (2) et (3), il est possible de procéder à de la statique comparative afin de déterminer l'impact de certaines variables et paramètres sur le modèle. De par la nature du travail, on s'intéresse uniquement à b. Or, on trouve :

$$\frac{\partial w^r}{\partial b} = \frac{\partial w^r}{\partial \xi}\,\frac{\partial \xi}{\partial b} > 0. \qquad (4)$$

$$\partial \lambda \,/\, \partial b = - r\,\delta f(w^r) \,/\, (r + \lambda) < 0. \qquad (5)$$

Donc, (4) suggère que plus les allocations d'a-c augmentent, plus le salaire de réserve augmente; (5) nous dit que la probabilité de sortir du statut "chômeur" diminue si les allocations d'a-c augmentent. L'intuition derrière ces dérivées est la suivante: si les prestations d'a-c augmentent, le chercheur d'emploi va augmenter son salaire de réserve ce qui a pour conséquence de diminuer la probabilité que le chômeur reçoive une offre w telle que $w > w^r$. Ce faisant, l'individu risque de voir la durée de son épisode de chômage augmenter, ou de façon équivalente, il diminue ses chances de se trouver un emploi, $\partial\lambda \,/\, \partial b < 0$, d'où l'effet désincitatif.

III-2) Modèle de Principal- Agent

On note de multiples extensions du modèle de recherche d'emploi relevant certaines hypothèses jugées contraignantes (Devine et Kiefer (1991)). Par exemple, en regard à la section II-3), il est possible de généraliser au cas où l'individu à la possibilité de transiter entre les trois statuts sur le marché du travail, plutôt que le cadre dichotomique employé-chômeur comme on vient d'étudier. Mais en ce qui nous concerne, une autre hypothèse pas très réaliste est celle de stationnarité de l'environnement du chercheur d'emploi, c'est-à-dire la stationnarité de $F(w)$, r , b et δ. En fait, la présentation du modèle de recherche d'emploi non stationnaire ne nous intéresse pas vraiment. Cependant, dans l'article de Hopenhayn et Nicolini (1997), on s'intéresse beaucoup à la non-stationnarité de b et δ. En effet, ils ont recours au modèle principal-agent répété afin de dériver un contrat d'a-c optimal, où on exploite l'effet lissage de la consommation produit par les allocations d'a-c, comme on a vu à la section II-2), afin de réduire l'effet désincitatif de ces dernières.

Or, la non-stationnarité de δ vient du fait que l'agent (chômeur averse au risque) peut varier le niveau de l'effort "a" avec lequel il cherche un nouvel emploi, effort non observable par le principal (gouvernement neutre au risque), créant ainsi le problème de hasard moral[1, 2]. La non-stationnarité de b découle du contrat optimal qui suggère qu'afin de réduire ce problème de hasard moral, le niveau des allocations d'a-c doit diminuer au cours de la période de recherche d'emploi, suivie d'une taxe sur le salaire de réemploi, l'idée étant qu'afin de lui fournir des incitations intertemporelles pour qu'il maximise a, le contrat doit entraîner une diminution de la consommation de l'agent autant lorsqu'il est chômeur (allocations diminuent avec le temps) qu'employé (taxe sur le salaire).

Les hypothèses du modèle sont les suivantes. Les préférences de l'agent sont données par $E\sum_{t=0}^{\infty}\beta^t\,[u(c_t) - a_t]$, où $c_t \in \Re_+$ et $a_t \in A = [0,\cdot]$ sont respectivement la consommation et l'effort de recherche d'emploi, $u'_c > 0$ et $u''_c < 0$ [3], $\beta < 1$ est le facteur d'escompte (le même pour le principal) et E est l'opérateur d'espérance. La probabilité que l'agent trouve un emploi est donnée par $p_t = p(a_t)$, où $p'_a > 0$ et $p''_a < 0$ et $p(\cdot)$ satisfait aux conditions d'Inada. On suppose que l'agent n'a aucune autre source de revenu que celui d'emploi et que le principal est donc à même de contrôler la consommation de l'agent. Finalement, on fait l'hypothèse qu'il n'y a aucune hétérogénéité dans les emplois, ces derniers offrant un salaire constant et permanent w à l'agent qui est embauché.

---

[1] Dans le modèle de recherche d'emploi, ça se traduirait par une fonction $\delta = \delta(a)$, $\delta'_a > 0$.

[2] Si le principal pouvait observer a, alors il n'y aurait pas de problème de hasard moral et le contrat optimal ne serait qu'un problème standard de partage de risque où le principal supporte tout le risque: Mas-Colell, Whinston et Green (1995).

[3] On remarque que $u(\cdot)$ ne dépend pas de l'effort à fournir dans l'exercice de ses fonctions lorsqu'il est employé; sans conséquence pour la dérivation du contrat optimal.

Au temps $t = 0$, le principal offre un contrat à l'agent, spécifiant l'effort recommandé "a_t" que ce dernier doit fournir et la séquence de transferts nets, ce en fonction du statut de l'agent sur le marché du travail entre 0 et t; soit $\mathbf{h}_t$ le vecteur qui retrace l'histoire de l'agent jusqu'à la période t (dichotomie 0 : sans emploi , 1: employé)[1]. Donc, le contrat est une fonction τ: $\mathbf{h}_t \rightarrow \{a_t, z_t\}$, où la paire $\{a_t, z_t\}$ spécifie les termes du contrat, avec a_t et z_t correspondent respectivement au niveau d'effort recommandé afin d'inciter l'agent à chercher (seulement s'il est en chômage) et à la séquence des transferts nets. À chaque contrat τ est associé $V_0(\tau)$ et $C_0(\tau)$, respectivement l'utilité actualisée espérée de l'agent et le coût actualisé espéré des transferts nets épongés par le principal; explicitons successivement ces deux derniers éléments.

L'optimalité du contrat va évidemment dépendre de ce qu'on entend optimiser. Ici, l'idée est de minimiser $C_0(\tau)$ sous contrainte à $V_0(\tau) = V$, où V est un niveau minimum d'utilité actualisée escomptée garanti par le contrat offert à l'agent sans emploi à $t = 0$. Plus précisément, ce contrat lui assure une consommation courante c et précise l'effort à fournir a de même que des niveaux V^e et V^u d'utilité actualisée espérée respectivement s'il trouve ou ne trouve pas d'emploi à la fin de la période $t = 1$. On peut ainsi réécrire $V = u(c) - a + \beta \{p(a) V^e + [1 - p(a)] V^u\}$; il s'agit de la contrainte de participation. Cependant, on comprend bien que $a = 0$ si $V^e < V^u$. Puisque l'effort a n'est pas observable, il faut s'assurer que l'agent fournisse le niveau prescrit par le contrat, c'est-à-dire que $a \in \text{argmax}_{\hat{a} \in A} \; u(c) - \hat{a} + \beta \{p(\hat{a}) V^e + [1 - p(\hat{a})] V^u\}$. $p(a)$ étant strictement concave, a sera le niveau optimal si $\beta \, p'(a) \, (V^e - V^u) = 1$ (dérivée de l'expression précédente par rapport à a); il s'agit cette fois de la contrainte d'incitation.

De son côté, à $t = 0$, le principal doit à la fois prendre en compte le coût de garantir V^e et V^u. Tout d'abord, à partir du moment où l'agent se trouve un emploi, à $t = t'$, il n'y a plus de problème d'incitation entre ce dernier et le principal, de sorte qu'il est optimal de lui offrir un niveau de consommation constant c^e définie implicitement par $V^e = E \sum_{t=t'}^{\infty} \beta^t [u(c^e_t)] = u(c^e) / (1-\beta)$. Ceci implique un transfert net de $z = c^e - w$ par période que le principal continuera de lui verser. Actualisé, on écrit $W(V^e) = [u^{-1}((1-\beta)V^e) - w] / (1-\beta)$, avec $W'(V^e) = 1 / u'(c^e) < 0$. Or, si $c^e < w$, alors le principal prélèvera une <u>taxe</u> de $w - c^e$ par période de façon à ce que les coûts qu'il doit continuer de supporter après que l'agent se soit trouvé un emploi ne soient pas supérieurs à zéro.

Plaçons-nous maintenant au moment où l'agent est à la recherche d'un emploi, soit à $t = 0$. Le principal doit de façon équivalente supporter le coût d'assurer à l'agent un niveau V^u. Le transfert à l'agent durant cette période est $z = c^u$ et $C(V^u)$ représente le coût actualisé espéré de ce transfert pour le principal. De façon à inciter l'agent à fournir l'effort prescrit, le principal doit maintenant minimiser $C(V)$, soit le coût actualisé espéré du contrat. En ayant une fois de plus recours au principe de Bellman (voir le point III-1.b)), on a:

[1] $\mathbf{h}_t$ contient $t - 1$ zéros si l'agent est toujours sans emploi à la fin de la période t, ou bien composé de $t' - 1$ zéros suivi de $t - t'$ uns si l'agent s'est trouvé un emploi à la fin de $t' < t$.

$$C(V) = \min_{a,\, c,\, Ve,\, Vu} \; c + \beta \{p(a)\, W(V^e) + [1 - p(a)]\, C(V^u)\} \tag{6}$$

$$\text{s.c. :} \quad u(c) - a + \beta \{p(a)\, V^e + [1 - p(a)]\, V^u\} = V \qquad \text{et} \qquad \beta\, p'(a)\, (V^e - V^u) = 1.$$

c'est-à-dire qu'à chaque période, le principal minimise son coût de transfert actuel (c) et futur ($\{\cdot\}$) sous contraintes à ce que l'individu cherche et choisisse le niveau d'effort prescrit par le contrat. Après avoir formulé le Lagrangien et manipulé les conditions de premier ordre du principal, le contrat optimal est défini par:

$$p'(a)\, [W(V^e) - C(V^u)] = \eta\, p''(a)\, (V^e - V^u),$$

$$C'(V^u) = u'(c^u)^{-1} - \eta\, p'(a) \cdot [(1 - p'(a))]^{-1},$$

$$W'(V^e) = u'(c^u)^{-1} + \eta\, p'(a) \cdot [\, p(a)]^{-1},$$

où η est le multiplicateur de Lagrange associé à la contrainte d'incitation; la condition d'enveloppe est donnée par $C'(V) = p(a)\, W'(V^e) + [1 - p(a)]\, C'(V^u) = u'(c^u)^{-1}$.

À l'aide de ces résultats, on montre tout d'abord que le niveau des allocations d'a-c diminue au cours de l'épisode de recherche d'emploi. En effet, après avoir montré que $\eta > 0$[1], on trouve que $W'(V^e) - C'(V^u) = \eta\, p'(a) \cdot [\, (1 - p'(a))^{-1} + (p'(a))^{-1}] > 0$. Étant donné la condition d'enveloppe, ceci implique que $W'(V^e) > C'(V) > C'(V^u)$. Or, si la fonction de coût C est convexe[2], alors on obtient que $V > V^u$, c'est-à-dire que l'utilité actualisée espérée de la séquence de transferts durant la période de recherche d'emploi doit être inférieure à l'utilité actualisée espérée garanti dans le contrat. Vu la concavité de la fonction d'utilité de l'agent, ceci implique que sa consommation doit diminuer lorsqu'il cherche pour du travail.

De plus, s'il n'y avait pas de taxe sur le salaire de réemploi, cela signifierait que par $V^e_t = u(w) / (1-\beta)$, $\forall t$. Cependant, les auteurs montrent[1] que sous le contrat optimal, V^e_t ne peut pas demeurer constant pour toujours; il doit diminuer.

Donc, le contrat optimal suggère une séquence de transferts qui diminue en période de recherche d'emploi et une taxe sur le salaire de réemploi qui augmente avec la durée de l'épisode de recherche. L'idée étant qu'afin de fournir des incitations intertemporelles, le contrat doit punir l'agent pour toute période de recherche qui s'étire en réduisant sa consommation future, autant lorsqu'il est sans emploi qu'employé. Cependant, il faut réaliser que cette taxe sur le salaire de réemploi risque d'être désincitative. Or, encore une fois si C est convexe et que $-p''(a)/\, p'(a)^2$ est décroissant en a, alors l'effet "revenu permanent" domine ce potentiel effet désincitatif[1]

---

[1] Voir en annexe de leur article pour la preuve.

[2] Étant donné la non-linéarité des contraintes de (6), ces dernières définissent potentiellement un ensemble non-convexe, de sorte qu'il peut être difficile de garantir la convexité de la fonction C. Cependant, à travers leur exercice de calibration, ils trouvent que cette dernière est convexe.

III-3) Gain de productivité

Une autre hypothèse du modèle de recherche d'emploi qui laisse à désirer est qu'à travers sa règle optimale d'arrêt, le chômeur maximise ses revenus. Le problème avec cette hypothèse est qu'elle suppose l'individu neutre au risque. Si on laisse plutôt le chômeur maximiser une fonction d'utilité croissante en salaire, c'est-à-dire $u'(w) > 0$, la règle optimale d'arrêt s'écrit comme suit :

$$u(w^r) = (u(L_u)/u(L_e))\, u(b) + (\delta/r) \int_{w^r}^{\infty} (u(x) - u(w^r))\, d\,F(x).$$

où L_u et L_e sont respectivement le loisir lors des périodes de chômage et d'emploi. Ce faisant, on admet la possibilité que les chômeurs soient averses au risque. Dans ce contexte, on peut montrer que les prédictions du modèle de base ne sont pas affectées; de plus, le salaire de réserve évolue en relation inverse avec l'aversion au risque du chômeur.

Supposons une économie sans a-c. Ainsi, d'après ce qu'on vient de voir, plus un individu est averse au risque, plus son salaire de réserve est bas, ce qui implique qu'il cherchera parmi les emplois à bas salaires[1] (où il est plus facile à d'être embauchée) et peu productifs. La firme accepte d'embaucher un travailleur à un salaire plus faible, cependant, comme dans l'esprit de la théorie des contrats implicites, ce salaire plus faible reflète simplement une prime que l'employeur exige de l'employé afin de lui fournir une sorte d'assurance contre la possibilité de licenciement. Il en résulte une composition d'emplois inefficace. De l'autre côté, étant donnée (4), la présence d'a-c incite les chômeurs à chercher pour des emplois à salaires et productivités plus élevés, encourageant ainsi les chômeurs à encaisser une plus grande part de risque, conduisant à une composition d'emplois plus productifs dans l'économie.

Donc, en dépit des "pathologies" qui accompagnent un programme d'a-c (section II-2)), ce dernier a potentiellement le mérite d'améliorer la productivité de l'appariement entre employeurs et chercheurs d'emploi; Acemoglu et Shimer (1999a,b) nous offrent un cadre d'analyse théorique basé sur un modèle de recherche d'emploi d'équilibre[2] et d'appariement avec travailleurs averses au risque, afin de présenter la logique derrière cet effet. Pour fin de présentation, j'expose le modèle de Acemoglu et Shimer (1999b), version légèrement simplifiée de Acemoglu et Shimer (1999a). Les hypothèses du modèle sont les suivantes.

---

[1] Ce type de recherche où le chômeur échantillonne une section particulière de $F(w)$ est appelée *systématique*, comparativement à la recherche *aléatoire*, où il échantillonne les offres arbitrairement; c'est cette dernière qu'on a supposée dans la présentation du modèle canonique au point III-1.b).

[2] Le modèle de recherche d'emploi d'équilibre est une autre extension du modèle de base où $F(w)$ est endogène.

Côté offre de travail, on suppose qu'il existe un continuum de travailleurs identiques ayant chacun une fonction d'utilité à la von Neumann-Morgenstern $U(c,h)$, avec c et h représentent respectivement la consommation (soit la somme de ses actifs initiaux A et de son revenu net de travail y; s'il travaille, $y = (1 - \tau)\, w$, où τ est la taxe proportionnelle prélevée afin de financer le programme d'a-c et s'il ne travaille, $y = b$, où b correspond aux allocations d'a-c) et les heures travaillées et où $U(\cdot)$ est continuement différentiable, strictement croissante en c et faiblement concave; supposons de plus que le nombre d'heures travaillées soit fixe, de sorte que $u(c) = U(c,h)$.

Côté demande, il existe également un continuum de firmes potentielles, chacune ayant la même fonction de productions $g(\cdot)$ et offrant la possibilité de créer un emploi avec spécificité $\alpha \in [0,1]$ et un coût $\gamma > 0$; le cas échéant, le travailleur produit $g(\alpha)$ unités d'output, avec $g'_\alpha > 0$. Cependant, les emplois avec α élevé sont difficiles créer vu le nombre relativement faible de travailleurs ayant la formation et les compétences requises pour combler un poste où la spécificité est grande. On représente la probabilité que l'appariement entre les deux partis relativement à l'emploi de spécificité α soit élevée par la fonction $M(\alpha)$, avec $M'_\alpha < 0$.

Au début de chaque période, chaque firme j décide d'ouvrir un poste ou non. Si elle va de l'avant, elle supporte un coût $\gamma > 0$, choisi la spécificité α_j et affiche un salaire offert w_j. Les employeurs et travailleurs se rencontrent à travers un processus de recherche d'emploi, ces derniers cherchant parmi les emplois affichés (α_j, w_j); on suppose qu'ils préfèrent des salaires plus élevés et des emplois moins spécifiques. Du côté des firmes, on suppose qu'elles préfèrent évidemment payer des salaires plus faible possible, mais elles font face à un arbitrage puisqu'un emploi avec spécificité et salaire élevée améliore potentiellement la productivité de la firme, mais les travailleurs qui possèdent les atouts pour combler ce poste sont difficiles trouver.

Bien qu'il soit possible, comme le font Acemoglu et Shimer (1999a), de supposer qu'il y ait plus de compétition pour certains postes via une file d'attente pour l'emploi (α_j, w_j) (*queuing*), on suppose plutôt que chaque travailleur applique à un seul emploi à la fois et qu'il y a un travailleur appliquant à chaque poste affiché. Donc, c'est seulement après s'être rencontré que le couple nouvellement apparié observe si l'employé possède les qualités requises; dans l'affirmative, il en résulte une production $g(\alpha)$ et l'employé gagne $y = (1 - \tau)\, w$. Sinon, on doit dissoudre l'appariement et l'employé reçoit $y = b$.

Maintenant, à l'équilibre, la paire (α, w) doit maximiser l'utilité espérée du travailleur sous contrainte au profit nul de la firme. Concrètement, on a (l'interprétation est directe):

$$\max_{\alpha, w}\; M(\alpha)\, u(A + (1 - \tau)\, w) + (1 - M(\alpha))\, u(A + b) \quad \text{s.c.} \quad M(\alpha)(g(\alpha) - w) \geq \gamma \tag{7}$$

Graphiquement, le problème est représenté par une condition de tangence entre la courbe d'indifférence de l'individu et la condition de profit nul de la firme, représenté à la figure 2.

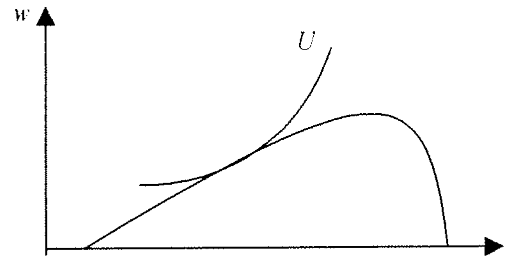


Figure 2: Condition de tangence entre la courbe d'indifférence et la condition de profit nul.

où la forme de U reflète les préférences des individus (des salaires plus élevés et des emplois moins spécifiques) et celle de la condition de profit nul de la firme reflète l'arbitrage précité face auquel cette dernière est confrontée; on peut maintenant procéder à la statique comparative sur b.

Puisque b n'intervient que dans la fonction objective de (7), seule la courbe d'indifférence de l'individu va se déplacer, en s'aplanissant vers la droite suivant une hausse de b comme on le montre sur la figure 3. En effet, selon (4), une hausse de b implique une hausse du salaire de réserve, incitant ainsi les chercheurs d'emploi à regarder pour des emplois offrant un salaire plus élevé, ceux-ci étant également plus productif (d'où la hausse de w et α), mais auxquels sont aussi associé une plus grande probabilité de perte d'emploi. Donc, en d'autres mots, la présence d'un programme d'a-c encourage les travailleurs à supporter un plus grand risque de perte d'emploi et il en résulte une compositions d'emplois plus productifs au sein de l'économie.

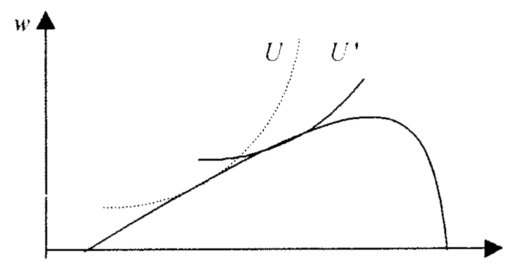


Figure 3: Déplacement du point de tangence suivant une hausse du niveau des allocations d'a-c.

Acemoglu et Shimer présentent par la suite un modèle dynamique d'équilibre général. Ce faisant, il est possible de modéliser à la fois le gain de productivité et le problème de hasard moral qui découlent du programme d'a-c. Cependant, ce modèle est légèrement plus complexe que le modèle statique et surtout, il ne possède pas de solution analytique. Pour ces raisons, on ne s'attardera pas à le dériver. Il est tout de même intéressant de souligner que leur résultats de calibration du modèle à l'économie américain semblent indiquer qu'accroître le niveau d'a-c augmente l'output et le bien-être de l'économie, suggérant ainsi, contrairement au problème de hasard moral prit isolément, des gains substantiels à hausser la générosité du programme.

Bien que la calibration du modèle dynamique d'équilibre général de Acemoglu et Shimer (1999b) laisse croire que l'effet désincitatif est dominé par un gain de productivité, on se tourne maintenant, et ce pour le reste du rapport, vers la recherche d'un effet désincitatif du programme d'a-c en pratique. Pour y parvenir, on passe en revue quelques études empiriques cherchant à vérifier si cet effet est bien réel ou non. Plus précisément, ces dernières cherchent à voir si, comme (5) nous enseigne, les allocations d'a-c accroissent la durée des épisodes de chômage. Or, le recours à l'économétrie des données de durée est très populaire à cette fin. La Partie IV est donc consacrée à la présentation théorique et pratique des méthodes économétriques qui nous permettent de juger de la présence de l'effet désincitatif; une nouvelle méthode d'estimation est proposée.

IV- Hasard moral et assurance chômage en pratique

La section IV-1) sert de présentation aux outils économétriques les plus fréquemment utilisés dans la vérification empirique de l'effet désincitatif; pour les besoins de la suite du rapport, il est important de s'attarder au modèle de hasard proportionnel et aux méthodes de contrôle de l'hétérogénéité non observée. En effet, les trois études qu'on passe en revue à la section IV-2), en l'occurrence Lancaster (1979), Moffitt (1985) et Meyer (1990), emploient ces techniques. C'est lors de cette revue qu'on souligne une faiblesse qui émane de ces études empiriques, relativement à la façon dont ils traitent l'hétérogénéité non observée, ouvrant ainsi la voie à la section IV-4), où on suggère une méthode alternative d'estimation du modèle de hasard proportionnel basée sur le principe d'entropie généralisé, principe qu'on aura présenté à la section IV-3). Cette nouvelle méthode d'estimation offre une avenue simple et potentiellement efficace afin de contrôler l'hétérogénéité non observée.

IV-1) L'économétrie des données de durée

Afin de vérifier s'il existe un effet désincitatif associé au programme d'a-c, on a besoin de données sur la durée des épisodes de chômage. Or, deux choses: *premièrement*, ces données sont souvent biaisées, ce pour différentes raisons qu'on étudie au point IV-1.a). *Deuxièmement*, les économètres ont importé de la biométrie les outils statistiques développés à l'étude des données de durée; on présente ces outils au point IV-1.b).

IV-1.a) *Biais dans les données de durées*

Il existe deux sources importantes de collectes de données relativement à la durée des épisodes de chômage, chacune avec des biais qui doivent être pris en compte lorsqu'on applique les méthodes économétriques associées aux données de durée qu'on étudie au point IV-1.b). Pour la première méthode, l'information est collectée seulement chez les individus en chômage au moment de l'enquête. Ceci implique deux grandes sources de biais: *i*) le *length-biased sampling* a trait au fait qu'on ne tiendra pas compte des épisodes de chômage comme celle représentée par t_2 sur la figure 4, car elles se situent entre les deux dates d'enquêtes A et B; il y aura donc sous-représentation de courts épisodes de chômage; *ii*) le problème de *right-censoring* découle du fait qu'au moment de la deuxième date de l'enquête, si un individu est au chômage, il est impossible de déterminer à quel moment il le quittera et quel sera son salaire de réemploi.

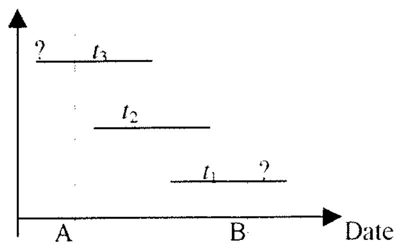


Figure 4 : Censure dans les données de durée.

Dans l'autre méthode de collecte, on sonde les individus indépendamment de leur statut sur le marché du travail et on leur demande rétrospectivement de l'information sur leur statut depuis une date "A". Bien que cette méthode nous permette de retrouver les épisodes de type t_2, elle soufre néanmoins elle aussi d'importants biais: *i*) une épisode de chômage comme t_3 est dite *left-censored* puisqu'on ne peut pas dire à partir de quel moment l'individu a débuté cette épisode de chômage; *ii*) il y a aussi problème de *length-biased sampling* étant donné que les épisodes en progrès au commencement de l'enquête sont davantage sujettes à être longues que courtes; *iii*) cette méthode d'enquête est de plus sujette aux *recall errors*, qu'on distingue en deux effets : *iii*-a) effet mémoire, c'est-à-dire lorsque les répondants oublient de reporter certaines épisodes de chômage et *iii*-b) *telescoping effect*, lorsque le répondant se souvient d'épisodes de chômage mais les reporte avec erreurs[1].

IV-1.b) *Concepts de base*

Afin de bien interpréter les résultats des études présentées à la section IV-2) et comprendre la méthode d'estimation proposée de la section IV-4), nous passons en revue l'économétrie des données de durée en deux temps. Dans un premier temps, nous présentons les concepts de base en supposant la population homogène. Dans un deuxième temps, nous relevons cette hypothèse d'homogénéité et voyons comment contrôler pour l'hétérogénéité, observé et non observée; c'est à ce moment qu'on présente le modèle de hasard proportionnel de même que des méthodes afin de contrôler l'hétérogénéité non observée, deux sujets qui motivent la section IV-4).

[1] Un type de *telescoping effect* qui retient beaucoup l'attention dans la littérature est le *heaping effect* (également appelé le *digit preference effect*), défini comme étant une concentration anormale des réponses à certains niveaux de duré, étant donné que les gens arrondissent les durées qu'ils rapportent. Torelli et Trivellato (1993) et Ryu et Slottje (2000) énoncent plusieurs méthodes afin de contrôler pour ce *heaping effect*.

(A) Homogénéité de la population

Soit $T > 0$ une variable aléatoire continue représentant la durée avant qu'un individu tirée d'une population homogène sorte d'un statut donné, et ayant une fonction de distribution $f(t)$. Afin de caractériser cette variable, on pourrait spécifier directement une distribution particulière à $f(t)$. On trouverait alors les fonctions cumulatives ($F(t)$) et de survie ($S(t)$):

$$F(t) = \int_0^t f(t) = \Pr(T \leq t) \quad \text{et} \quad S(t) = 1 - F(t) = \Pr(T \geq t).$$

Une méthode alternative est plutôt de modéliser la fonction de hasard, définie comme étant :

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \to 0} \Pr(t \leq T < t + \Delta t \mid T \geq t) / \Delta t \\ &= \lim_{\Delta t \to 0} [F(t + \Delta t) - F(t)] / S(t) = f(t) / S(t).\end{aligned}$$

La fonction de hasard spécifie donc le taux de sorti instantané au temps $T = t$ conditionnel au fait d'avoir survécu jusqu'à t. Or, on vérifie facilement que $\lambda(t)$ spécifie tout de même la distribution de T. En effet, sachant que $\lambda(t) = f(t) / S(t)$, on a que $\lambda(t) = - d \ln (S(t)) / dt$. En intégrant et en utilisant $S(0) \equiv 1$, on trouve :

$$S(t) = \exp\left(- \int_0^t \lambda(u)\, du \right) = \exp(- \Lambda(t)) ,$$

où $\Lambda(t)$ correspond à la fonction de hasard intégré (on fait l'hypothèse que $S(t) \to 0$ lorsque $t \to \infty$). Donc, on voit effectivement il est possible de retrouver la distribution de T puisque $f(t) = \lambda(t) \cdot \exp(- \Lambda(t))$.

Généralement, on opte pour la seconde approche en spécifiant directement une forme paramétrique particulière à $\lambda(t)$, l'allure de la fonction de hasard ayant la particularité de caractériser le processus stochastique sous-jascent. En effet, si $\partial\lambda(t) / dt > 0$ (<0), on dit que la dépendance en durée est positive (négative), c'est-à-dire que les chances de sortir d'un statut donné croissent (diminuent) avec le temps.

La question maintenant est de savoir, quelle paramétrisation doit-on attribuer à la fonction de hasard? Cette question n'est pas banale puisque chaque spécification implique une dépendance en durée qui lui est propre. Par exemple, la distribution exponentielle implique un hasard constant, c'est-à-dire $\partial\lambda(t) / dt = 0 \Rightarrow \lambda(t) = \lambda$. En effet, $- d \ln S(t) / (d\, t) = \lambda$ implique que $S(t) = \mathrm{K}e^{-\lambda t}$, où K est la constante d'intégration. En utilisant le fait que $S(0) \equiv 1$, on trouve que $\mathrm{K} = 1 \Rightarrow S(t) = e^{-\lambda t}$, c'est-à-dire la distribution exponentielle. D'autre part, si on suppose une fonction de survie Weibull, qui emboîte l'exponentielle, on a $\lambda(t) = \lambda p(\lambda t)^{p-1}$, où $p > 0$ (on voit effectivement que $\lambda(t) = \lambda$ si $p=1$). Or, cette formulation implique une dépendance en durée positive si $p > 1$ et négative si $p < 1$, où le hasard évolue de façon monotone. Cette monotonicité de la fonction de hasard disparaît si on suppose une survie soit du type log-normale ou log-logistique. Dans ce dernier cas, $\lambda(t) = \lambda p(\lambda t)^{p-1}/[1+(\lambda t)^p]$. Il y a bien sûr plusieurs autres paramétrisation possibles de la fonction de hasard avec leur spécification respective de la dépendance en durée; voir Kalbfleisch et Prentice (1980).

Supposons maintenant que la distribution des données (plutôt que la fonction de hasard) est connue à un vecteur de paramètres θ près. On a donc $f(t\,|\theta)$, la densité d'une observation de durée t. Comme on a vu au point IV-1.a), les données de durée risquent souvent d'être censurées. Lorsqu'une épisode est effectivement censurée à la durée t, la seule information disponible est que la durée est d'au moins t, c'est-à-dire $S(t \mid \theta)$. Soit $d = 1$ si l'épisode t est non-censurée, la fonction de vraisemblance s'écrit de la façon suivante :

$$L(\theta) = \prod_i [f(t_i\,|\theta)]^{d} \cdot [S(t_i \mid \theta)]^{(1-d)}.$$

On vient de voir qu'on peut passer de $f(\cdot)$ à $\lambda(\cdot)$, de sorte qu'en prenant le logarithme de la fonction de vraisemblance et en simplifiant pour les relations qu'on vient de dériver, on trouve :

$$\ln L(\theta) = \sum_i d_i \ln \lambda(t_i\,|\theta) - \sum_i \Lambda(t_i \mid \theta).$$

On estime donc le modèle en maximisant $\ln L(\theta)$ par rapport à θ.

On comprend bien que l'utilité de la précédente approche repose sur une bonne paramétrisation de la fonction de hasard, sinon les résultats sont quasi inutilisables. Donc, plutôt que d'imposer une structure particulière aux données, les chercheurs ont souvent recours (souvent comme analyse préliminaire) à l'estimateur non-paramétrique de Kaplan-Meier. Supposons que les épisodes de chômage soient placées en ordre croissant $t_1 < t_2 < \ldots < t_k$; on suppose que si $T = t$ et que $d = 1$, alors la censure s'opère immédiatement après t. Donc, soit b_i le nombre d'épisodes qui terminent à t et m_i le nombre d'épisodes censurées entre t_i et t_{i+1}, on défini alors l'ensemble à risque $n_i = \sum_{j\geq i} (m_j + b_i)$ comme les épisodes qui sont éligibles à se terminer en t_i. Un estimateur non-paramétrique naturel de la fonction de hasard est donc $\lambda^*(t_i) = b_i / n_i$. Il s'en suit un estimateur non-paramétrique de la fonction de survie $S^*(t_i) = \prod_{j=1}^{i} (n_j - b_j)/ n_i$.

(B) <u>Hétérogénéité de la population</u>

Jusqu'à maintenant, on supposait que la population d'où on observait l'individu avec la durée T était homogène. En d'autres mots, nous n'avions pas admis la possibilité que des facteurs externes puissent jouer un rôle dans la distribution de survie. On relève cette hypothèse en admettant la possibilité que les individus diffèrent entre eux selon des caractéristiques *observées*, ce qui non permet d'introduire la définition du modèle de hasard proportionnel, et *non observées*.

❖ Hétérogénéité observée

On note deux méthodes afin d'introduire des variables explicatives dans la fonction de hasard; paramétrique et semi-paramétrique. Le modèle semi-paramétrique de "hasard proportionnel" est des plus importants pour la suite du travail.

Paramétrique : La façon de généraliser les définitions qu'on a dériver précédemment est de spécifier la fonction de hasard comme:

$$\lambda(t \mid X) = \lim_{\Delta t \to 0} \Pr(t \leq T < t + \Delta t \mid T \geq t, X) \,/\, \Delta t\,,$$

où X correspond à l'ensemble des variables explicatives pertinentes. L'approche paramétrique standard est de supposer une forme spécifique $\phi(\cdot)$ pour la fonction de hasard, c'est-à-dire $\lambda(t \mid X) = \phi(X, \beta, t)$. Il suffit par la suite de construire la fonction de vraisemblance comme on a fait plus haut et maximiser par rapport à β. On remarque qu'en présence de censure, on peut retrouver les modèles de régressions censurées (Tobit).

Semi-Paramétrique : La méthode semi-paramétrique la plus fréquemment utilisée afin d'introduire des variables explicatives dans les modèles de durée est la méthode du hasard proportionnel[1], où on défini $\lambda(t \mid X) = \lambda_0(t)\,\phi(X\beta)$, où $\lambda_0(t)$ est appelé *hasard de base*. L'importante caractéristique de ce dernier modèle est que toute la dépendance en durée est capturée dans le hasard de base, c'est-à-dire que la différence dans $\lambda(t \mid X)$ entre individus à une durée donnée t ne dépend de t seulement qu'à travers les variations des régresseurs. Par exemple, soit i et j deux individus, on a :

$$\lambda_i(t^* \mid X) = a_{ij} \cdot \lambda_j(t^* \mid X),\ \text{où } a_{ij} = \phi(X_i\beta)/\phi(X_j\beta),\ \forall \text{ durée } t^*$$

L'intérêt de ce modèle économétrique est qu'il est possible d'obtenir une estimation de β sans avoir à spécifier une forme particulière au hasard de base, qui est traité comme un paramètre de nuisance. On parle alors de modèle de hasard proportionnel avec un <u>hasard de base non-paramétrique</u>. La méthode d'estimation d'un tel modèle est la *vraisemblance partielle*. L'idée de cette méthode est semblable à celle de l'estimateur de Chamberlain dans la cadre du logit en donnée de panel. En effet, on spécifie $\phi(X\beta) = \exp(X\beta)$ et on trouve une estimation de β en conditionnant la probabilité qu'un individu sorte d'un statut au temps T_i sur l'ensemble à risque R_i:

$$\Pr[t_i = T_i \mid R_i] = \frac{\lambda_0(t)\exp(X_i\beta)}{\lambda_0(t)\sum_{j\in R_i}\exp(X_j\beta)} = \frac{\exp(X_i\beta)}{\sum_{j\in R_i}\exp(X_j\beta)}$$

[1] Il y a également le modèle de vie accéléré (AFT) où on a $\lambda(t \mid X) = \lambda_0(t\,\phi(X\beta))\,\phi(X\beta)$. L'effet des variables explicatives est ici de ré-échelonner l'axe temporel plutôt que de faire changer de place la fonction de hasard, comme c'est le cas dans le modèle de hasard proportionnel. Voir Neumann (1997) qui discute du Generalized Accelerated Failure Time model, qui emboîte les modèles de hasard proportionnel et l'AFT

On constate donc que cette méthode revient à estimer un logit multinomial sur l'ensemble à risque[1]. Dans certain cas cependant (comme on verra chez Moffitt (1985)), il peut être important de retenir le hasard de base. Dans ce cas, la solution est de remplacer $\lambda_0(t)\exp(X\beta)$ dans la fonction de vraisemblance et de maximiser par rapport à λ_0 et β; c'est ce qu'on appel le *general maximum-likelihood*.

D'autres chercheurs préfèrent tout de même spécifier un <u>hasard de base paramétrique</u>. Le choix de la forme pour $\lambda_0(t)$ n'est pas arbitraire, comme le souligne Lancaster (1979), étant donné le choix de $\phi(X\beta) = \exp(X\beta)$. Une spécification qui revient souvent pour $\lambda_0(t)$ est la Weibull; on trouve alors que $\lambda(t \mid X) = \exp(X\beta)\cdot\lambda p(\lambda t)^{p-1}$. On trouve l'estimation de β en maximisant la fonction de vraisemblance.

Finalement, différents auteurs (comme on verra chez Meyer (1990)) ont développé des estimateurs <u>semi-paramétrique du hasard de base</u>. Par exemple, Han et Hausman (1990) traitent δ_t, le logarithme du hasard de base intégré

$$\ln \int_0^{t_i} \lambda_i(t|X)\, dt = X_i\beta + \varepsilon_i \qquad \text{et} \qquad \ln \int_0^{t} \lambda_0(t)\, dt = \delta_t\,,$$

comme constant à chaque période et à estimer en même temps que les β, $t = 1, \ldots, T$. Donc, la probabilité que l'individu i sorte d'un statut donné au temps t est:

$$\int_{\delta_{t-1} - X_i\beta}^{\delta_t - X_i\beta} f(\varepsilon)\, d\varepsilon = \Pr(\delta_{t-1} - X_i\beta < \varepsilon_i < \delta_t - X_i\beta).$$

Si on construisait la fonction de vraisemblance de ce modèle, en faisant l'hypothèse que ε suivre une loi normale (Weibull) et si on suppose qu'il n'y a pas d'observations censurées, alors on obtient un probit (logit) ordonné. C'est à l'aide de ce résultat qu'on présente à la section IV-4) une méthode alternative d'estimation des modèles de hasard proportionnel.

- Hétérogénéité non observée

La méthode la plus courante afin de contrôler pour l'hétérogénéité non observée est de supposer une forme paramétrique à l'effet individuel. Par exemple, supposons qu'on ait $S(t_i|v_i)$, où v correspond à certaines caractéristiques qui échappent à l'économètre. Nous, ce qui nous intéresse est la fonction de survie non-conditionnelle. La solution est de supposer une distribution particulière de l'hétérogénéité, $G(v_i)$, et d'intégré la fonction de survie sur v: $S(t) = \int_v S(t|v)\, g(v)\, dv$.

[1] Une autre méthode d'estimation du modèle de hasard proportionnel est l'approche bayesienne. Là encore, on traite $\lambda_0(t)$ comme un paramètre de nuisance. L'information à priori sur la forme de $\lambda_0(t)$ et de l'impact de certaines variables explicatives peuvent être incorporé formellement; Kalbfleisch et Prentice (1980) donnent plus de détails.

Heckman et Singer (1984) suggèrent une méthode (semi-paramétrique) alternative en avançant que celle qui précède sur-paramétrise la fonction de survie et peut conduire est des résultats erronés si on spécifie mal $G(v_i)$; ils supposent que $G(v)$ puisse être approximée par une distribution discrète avec un nombre fini de points, ces derniers étant les points de support pour la distribution de v. Les probabilités associées à ces points de supports sont estimées en même temps que les β dans la fonction de vraisemblance. Or, le gros désavantage de cette méthode, et un bref coup d'œil à leur article peut facilement nous en convaincre, est qu'elle est difficile à estimer. On va voir cependant qu'à la section IV-4), la méthode d'estimation alternative du modèle de hasard proportionnel qui est proposée implique potentiellement une façon assez simple afin de contrôler pour cet hétérogénéité non observée et qui s'inspire justement de Heckman et Singer (1984).

IV-2) Revue d'études empiriques

On se tourne maintenant vers trois études empiriques (employant malheureusement des données autres que canadiennes) qui ont recours à ces méthodes économétriques, particulièrement au modèle de hasard proportionnel, afin d'estimer si les allocations d'a-c sont susceptibles d'accroître la durée des épisodes de chômage, comme le suggère (5). On discute par la suite de la faiblesse des études.

IV-2.a) *Lancaster (1979)*

Lancaster (1979) est la bougie d'allumage dans l'application des méthodes précédentes à l'analyse des données sur la durée des épisodes de chômage. Ce dernier a recours à un modèle de hasard proportionnel en estimant quatre spécifications différentes, ce sur 479 travailleurs peu qualifiés d'Angleterre pour l'année 1973, où le hasard de base est *i)* constant, *ii)* Weibull, *iii)* constant avec hétérogénéité paramétrique et *iv)* Weibull avec hétérogénéité paramétrique; dans les deux derniers cas, la distribution de l'hétérogénéité ($G(v)$) est supposée de type gamma. Parmi les variables explicatives qu'il utilise, on note le taux de remplacement des allocations de chômage (allocation de chômage hebdomadaire totale divisé par le salaire net reçu lors du précédent emploi). Toutes les variables sont mesurées en logarithme de sorte que les estimations s'interprètent en terme d'élasticité. Ainsi, les élasticités estimées du taux de remplacement sont: *i)* 0.43, *ii)* 0.53, *iii)* 0.43 et *iv)* 0.6. Donc, bien que les écarts-types des coefficients estimés en *iii* et *iv* soient assez larges, on conclu néanmoins que les allocations de chômage augmentent la durée des épisodes de recherche d'emploi des individus.

IV-2.b) *Moffitt (1985)*

Moffitt (1985) débute l'étude de la durée (en semaines) des épisodes de chômage de 4628 hommes américains en calculant l'estimateur de Kaplan-Meier. L'inspection visuelle du graphique démontre une baisse du taux de hasard dans les (environs) huit premières semaines, après quoi le hasard est plus ou moins constant jusqu'à la vingtième semaine. Passé ce point, la fonction de hasard croit rapidement et est caractérisée par deux pointes aux durée 25 et 37. Il estime par la suite une autre estimation de Kaplan-Meier, mais cette fois-ci en subdivisant l'échantillon selon que l'individu est admissible à plus ou moins de vingt-six semaines en allocations d'a-c. Les résultats montrent qu'il y a très peu de différences entre les deux fonctions de hasard pour les vingt premières semaines, mais qu'à environ vingt-six semaines, la fonction de hasard pour ceux qui sont admissibles à des allocations sur une plus grande période demeure relativement constante.

Aux yeux de Moffitt, ceci suggère qu'il doit y avoir un effet dans la fonction de hasard associé à la durée d'admissibilité au régime d'a-c. Son objectif par la suite est d'essayer de déterminer si, en contrôlant pour cette durée potentielle d'admissibilité au régime, on arrive à une fonction de hasard plus ou moins constante sur tout l'horizon temporel. Pour y parvenir, il spécifie, contrairement à Lancaster, un hasard de base non-paramétrique: la fonction de hasard s'écrit comme $\lambda(t|X(t))= \lambda_0(t)\ \exp(X(t)\beta)$, où la dépendance temporelle de X est là pour indiquer que les allocations d'a-c ne sont pas stationnaire tout au long de la durée de la période de recherche d'emploi.

Or, puisque son objectif se traduit ici à voir si l'introduction d'une variable contrôlant pour la durée potentielle du régime (définie comme étant le nombre de semaine d'éligibilité au programme) dans X permet de lisser le hasard de base $\lambda_0(t)$, il ne peut pas se tourner vers l'emploi de la méthode de maximum de vraisemblance partielle, sinon le hasard de base disparaît. Il a donc recours au *general maximum-likelihood*. Comme on a vu au point IV-1.b), l'idée est de maximiser conjointement sur λ_0 et β. Règle générale, les résultats d'estimations pour diverses spécifications montrent que le coefficient estimé de la variable "duré potentielle" du régime est très peu significatif (il ne reporte pas les estimés du hasard de base) alors que le coefficient estimé associé au niveau de l'allocation d'a-c est négatif et significatif presque pour l'ensemble des modèles estimés. Donc, la probabilité de quitter le statut chômeur diminue à mesure qu'augmentent les allocations d'a-c. Ainsi, encore une fois, (5) tient.

IV-2.c) *Meyer (1990)*

À l'instar de Moffitt, Meyer (1990) commence par appliquer l'estimateur Kaplan-Meier sur la durée des épisodes de chômage de 3365 américains. En fait, puisqu'il travaille sur un sous-échantillon de la base de données que Moffitt (1985) a utilisé, il observe naturellement lui aussi d'importantes pointes à mesure qu'approche la date de terminaison des allocations d'a-c. Cependant, l'approche de Meyer se distingue de celle de Moffitt principalement pour deux raisons. Premièrement, plutôt que d'estimer un modèle de hasard proportionnel avec hasard de base non-paramétrique, il se tourne vers l'emploi d'un modèle de hasard proportionnel où le hasard de base est semi-paramétrique. Plus particulièrement, il défini $\lambda_i(t|X(t)) = \lambda_0(t)\exp(X_i(t)\beta)$, avec le hasard de base défini comme :

$$\ln \int_t^{t+1} \lambda_0(t)\, dt = \delta_t\,,$$

où δ_t est estimé en même temps que β. Deuxièmement, Meyer contrôle pour la durée potentielle de participation au régime d'a-c en ayant comme variables indépendantes des *time until benefit exhaustion spline* – plutôt que le nombre de semaines d'éligibilité au programme utilisé par Moffitt. Ces variables sont définies comme étant l'effet additionnel sur la fonction de hasard de s'approcher de x périodes de la date terminale de réception des allocations. Plus formellement, il défini les six variables *UI*1, *UI*2-5, ..., *UI*41-54; soit τ le nombre de semaines avant l'expiration du programme, alors *UI*1 = 1 si $\tau = 1$ et *UI*1 = 0 autrement. De la même façon, *UI*2-5= min(6-τ,4) si $\tau \leq 5$ et , *UI*2-5=0 autrement, etc....

Meyer rapporte également différentes spécifications de modèles estimés, où le niveau des allocations d'a-c apparaît dans chacune d'elles. Lorsqu'il ne corrige pas pour l'hétérogénéité non observée, il obtient une élasticité de -0.88 de la fonction de hasard par rapport au niveau des allocations de chômage (les variables étant mesuré en logarithme). Donc, la probabilité de quitter le statut chômeur diminue avec une hausse du niveau des allocation d'a-c, encore une fois cohérent avec (5). De plus, il est bon de souligner que les coefficients estimés des *time until benefit exhaustion spline* indiquent effectivement un effet significatif sur le hasard du fait de s'approcher de la date d'expiration du programme, contrairement aux résultats de l'approche de Moffitt (1985). En effet, les coefficients estimés des variables *UI*6-10, *UI*2-5 et *UI*1 sont respectivement 0.005, 0.185 et 0.67. Donc, la probabilité de se trouver un emploi (en supposant qu'il n'y a que deux statuts, travailleur et chômeur) augmente rapidement à mesure qu'approche la date où l'attribution des allocations cesse.

Lorsqu'il corrige pour l'hétérogénéité non observée (il emploi aussi une distribution gamma pour *G*(v)), les résultats changent à peine, l'élasticité de la fonction de hasard par rapport au niveau des allocations de chômage passant à -0.6. Les résultats ne changent également rien dans l'interprétation des *time until benefit exhaustion spline*.

On vient de passer un revue trois études empiriques qui emploient le modèle de hasard proportionnel afin de vérifier si, comme le prédit le modèle de recherche d'emploi, les allocations d'a-c accroissent la durée des épisodes de chômage. Il semble en effet que ce soit le cas. De plus, ces trois études ont recours au modèle de hasard proportionnel en spécifiant le hasard de base de trois façons différentes, respectivement paramétrique, non-paramétrique et semi-paramétrique. La seule faiblesse qui émane de ces études c'est la façon dont les auteurs traitent l'hétérogénéité non observée. Alors que Moffitt ignore tout simplement la question, Meyer suppose que $G(v)$ soit de type gamma alors qu'il aurait pu avoir recours à la méthode de Heckman et Singer (1984). Bien que ce soit effectivement une méthode assez compliquée à estimer, on a vu au point IV-1.b) que de l'avis de ces derniers, il serait plus prudent d'estimer des points de support dans la fonction de vraisemblance afin de détecter et contrôler pour la présence d'hétérogénéité non observée. Lancaster opte également pour une $G(v)$ de type gamma, mais son étude précède celle de Heckman et Singer.

Or, à la section IV-4), il est proposée une nouvelle méthode d'estimation du modèle de hasard proportionnel via le principe d'entropie généralisée. Il s'avère que cette méthode a potentiellement l'avantage d'offrir une façon assez simple afin de contrôler pour l'hétérogénéité non observée qui s'inspire de Heckman et Singer (1984), mais qui est beaucoup moins difficile à estimer que la méthode offerte par ces derniers. Avant d'y arriver, présentons le principe d'entropie généralisée.

IV-3) Principe d'entropie généralisée

Nous commençons par étudier au point IV-3.a) ce qu'on entend par entropie. Au point IV-3.b), on présente comment employer le concept d'entropie à des fins économétriques; c'est alors qu'on dérive le principe d'entropie généralisée.

IV-3.a) *Entropie: Concepts et définitions*

La notion d'entropie a vu le jour en physique, plus particulièrement en thermodynamique avec les travaux de Boltzmann. On la définie alors comme étant une grandeur qui permet d'évaluer la dégradation de l'énergie d'un système. Or, à travers ses travaux sur la théorie de l'information (par théorie de l'information, on entend la théorie mathématique de la communication), Shannon (1948) développa une formule qui servira à mesurer l'incertitude associée à un message ou à une expérience, qu'il appela également l'entropie, vue la grande similitude entre sa définition et celle de Bolzmann. Dérivons maintenant la formule d'entropie de Shannon.

Supposons une expérience X sur l'espace probabilisé $(\Omega, \Im, P)$ [1]. Étant donné qu'on ne connaît pas d'avance le résultat de cette expérience aléatoire, cette dernière est donc caractérisée par un certain degré d'incertitude face à ce qui en résultera. La question qui se pose est comment pourrions nous mesurer le degré, la grandeur de cette incertitude; cela nous permettrait entre autre de pouvoir comparer les incertitudes associées à différentes expériences.

Supposons deux expériences $(I = 2)$ X_i, $i = 1, 2$, ayant chacune d'elles deux événements $(N = 2)$ $_iA^n$ possibles, $n = 1, 2$. On voit bien que pour deux mesures de probabilité $P_i = (P(_iA^1), P(_iA^2))$, i=1, 2, données par $P_1 = (0.5, 0.5)$ et $P_2 = (0.9, 0.1)$, l'incertitude associée à l'expérience X_1 est plus élevée que l'incertitude de l'expérience X_2. Donc, pour des expériences avec deux seuls événements possibles, il est facile de déterminer l'incertitude de différentes expériences; on a qu'à comparer la plus petite des deux probabilités et la plus faible correspond à l'expérience la moins incertaine. Un corollaire à ceci est que l'incertitude est au maximum lorsque $P(_iA^n) = 0.5$.

Cependant, on comprend que les choses peuvent rapidement se compliquer lorsque le nombre d'événements est supérieur à deux. Ainsi, supposons une expérience X avec trois événements $(N = 3)$ possible ayant les probabilités $P(A^1) = p$, $P(A^2) = q$ et $P(A^3) = r$. Lorsque T répétitions indépendantes de l'expériences sont effectuées, on anticipe avoir approximativement Tp occurrences du premier événement et Tq et Tr respectivement pour les deux autres événements. Donc, on pourrait mesurer l'incertitude de l'expérience X par $p^{Tp}q^{Tq}r^{Tr}$; il s'agit de la probabilité d'obtenir un événement donné d'une expérience. Plus ce nombre est faible, plus l'incertitude sous-jacente à l'expérience X est grande. En prenant le logarithme de cette dernière expression, on trouve:

$$Tp \log p + Tq \log q + Tr \log r = T(p \log p + q \log q + r \log r).$$

En laissant tomber le T (on revient à une seule observation) et en changeant de signe pour tenir compte du fait que log $P(A^n)$ est négatif pour $P(A^n) < 1$, on obtient:

$$-\ p \log p - q \log q - r \log r$$

On peut maintenant généraliser pour N événements d'une expérience X; soit $P(A^n) = p_n$, $n = 1, 2, \ldots, N$:

$$H(p_1, p_2, \ldots, p_N) = \textstyle\sum_n - p_n \log p_n. \tag{8}$$

avec $0 \cdot \log 0 = 0$. Cette dernière expression[2] est l'*entropie* de l'expérience X; elle possède les propriétés suivantes (voir Ihara (1993) pour les preuves formelles):

---

[1] Voir Annexe A.1

[2] Voir Ihara (1993) pour une généralisation dans le cas continu

1) $H(p_1, p_2, \ldots, p_N) \geq 0$,

2) Si $p_n = 1$ pour un n donné, alors $H(p_1, p_2, \ldots, p_N) = 0$,

3) Pour toute répartition $p_1, p_2, \ldots, p_n$, on a : $H(p_1, p_2, \ldots, p_N) \leq H(1/N, 1/N, \ldots, 1/N)$,

4) $H(p_1, p_2, \ldots, p_N, 0) = H(p_1, p_2, \ldots, p_N)$,

5) Soit X_1 et X_2 deux expériences indépendantes. On a que $H(X_1, X_2) = H(X_1) + H(X_2)$.

À l'aide d'un certain ensemble de définitions et propriétés additionnelles, Shannon (1948) a montré que l'entropie (8) est la seule, à une constante près, à respecter ces propriétés (Ihara (1993)).

Ayant dérivé une mesure d'incertitude associée à une expérience en fonction des probabilités attribuées à chaque événements, on peut se donner une mesure qui nous informera de combien proche sont deux distributions de probabilités. Formellement, soit P et Q deux mesures de probabilités sur l'espace probabilisable $(\Omega, \Im)$; définissons $P(A^n) = p_n$ et $Q(A^n) = q_n$. Alors, l'*entropie relative* est définie comme suit:

$$H(P; Q) = \textstyle\sum_n p_n \log p_n/q_n \geq 0, \qquad (9)$$

avec égalité si $P = Q$. Donc, $H(P; Q)$ mesure en quelque sorte la "distance probabilistique" entre P et Q, c'est-à-dire combien proche sont les distributions P et Q entre elles. En d'autres mots, l'entropie relative peut être vue comme une grandeur qui nous indiquera dans quelle mesure P et Q se ressemblent. Par exemple, on peut réécrire (9) de la façon suivante:

$$H(P; Q) = \textstyle\sum_n p_n \log p_n/q_n = \sum_n p_n \log p_n - \sum_n p_n \log q_n = E[\log p_n] - E[\log q_n],$$

ce qui nous fourni une mesure naturelle de la distance entre P et Q. Cependant, il faut préciser que $H(P; Q)$ n'est pas un métrique en soit[1].

Une chose importante à remarquer en terminant, c'est que si on reprend l'expression (9) et qu'on pose $q_n = 1/N \; \forall n$, alors on montre que:

$$H(P; Q) = \textstyle\sum_n p_n \log p_n/q_n = -H(p_1, p_2, \ldots, p_N) - \sum_n p_n \log q_n = -H(p_1, p_2, \ldots, p_N) + \log N$$

Donc, si on ignore la constante additive $\log N$ et la constante multiplicative -1, alors:

$$H(P; Q) = H(p_1, p_2, \ldots, p_N),$$

ce qui est logique puisque si $q_n = 1/N \; \forall n$, alors cette dernière distribution n'amène aucune information supplémentaire étant donné que $H(q)$ est maximum. Ceci nous fournit une nouvelle façon d'interpréter $H(p_1, p_2, \ldots, p_N)$; l'entropie (8) mesure la "similitude" entre P et la distribution uniforme; si $H(p_1, p_2, \ldots, p_N) = 1$ (logarithme base 2), alors l'entropie de l'expérience sous-jacente est maximum, c'est-à-dire que $p_n = 1/N \; \forall n$.

---

[1] Voir Annexe A.3.

IV-3.b) *Analyse via l'entropie*

Ce qu'on veut faire maintenant est d'avoir recours au concept d'entropie à des fins économétriques; le prochain paragraphe résume le *principe d'entropie maximum* qui nous conduit plus loin au *principe d'entropie généralisée*.

Supposons qu'on cherche à obtenir un vecteur de probabilités associé à des événements d'une expérience aléatoire donnée. L'attrait du principe d'entropie maximum est qu'il permet de trouver un vecteur de probabilités estimé <u>en ne supposant rien d'autre que ce dont on est certain</u>. Autrement dit, on estime ces probabilités en employant que de l'information dont on ne peut douter la validité. Sur le plan opérationnel, l'idée est de supposer le moins de choses possibles (qui pourraient être fausses) et ça on y parviendra en maximisant l'entropie, c'est-à-dire l'incertitude (notre incertitude face à certains éléments du problème). S'il est convaincu de la validité de certains aspects du problème, le chercheur peut les introduire à la maximisation de l'entropie sous forme de contraintes. Donc, l'approche sera de maximiser l'entropie (8) sous contrainte à ce qu'on est absolument certain, formuler le Lagrangien et dériver un vecteur de probabilité qui sera représentatif de ce qu'on sait, rien de plus.

La présentation s'effectue en deux temps. Dans un premier temps, afin de bien comprendre le principe d'entropie maximum, de même que le principe d'entropie minimum relative (où l'idée sera plutôt de minimiser (9)), on l'emploi à l'intérieur d'un modèle très simple. Dans un deuxième temps, on généralise au modèle linéaire général et on présente le *principe d'entropie généralisée*.

(A) <u>Principe d'entropie maximum (PEM) et d'entropie minimum relative (PEMR)</u>

Afin de bien comprendre, supposons pour commencer qu'on s'intéresse à un problème du type $\mathbf{y} = \boldsymbol{X}\mathbf{p}$ où $\mathbf{y}$ est un vecteur d'observations $T \times 1$, $\mathbf{p}$ est un vecteur de probabilité inconnu $K \times 1$ et $\boldsymbol{X}$ est une matrice connue $T \times K$. Ce type de problème où il y a absence d'erreurs de mesures, bien que pas très intéressant en économétrie, nous aidera à comprendre l'idée de base. Par souci de cohérence avec le point IV-3.a), nous voyons successivement le principe d'entropie maximum et le principe d'entropie minimum relative.

- Principe d'entropie maximum (PEM)

Supposons qu'on souhaite maximiser l'entropie (8) sous contrainte à ce qu'on sait, mais sans imposer trop de contraintes. Donc, quelles devraient être ces contraintes? Choses certaines, la somme des probabilités doit être égale à un. Clairement, il nous faudra d'autres contraintes puisque si on ne s'en tient qu'à cette dernière, comme on a vu précédemment, alors le vecteur de probabilité $\mathbf{p}$ sera simplement uniforme. En effet, ce problème s'écrit comme suit:

$$\max_{\mathbf{p}} H(\mathbf{p}) = -\sum_k p_k \ln p_k \quad \text{s.c. } \sum_k p_k = 1$$

$$L(\mathbf{p}, \mu) = -\sum_k p_k \ln p_k + \mu\,(1 - \sum_k p_k)$$

<u>CPO</u>:

$$\frac{\partial L(\cdot)}{\partial p_k} = -\ln p^{\circ}_k - 1 - \mu^{\circ} = 0 \qquad \Rightarrow \qquad p^{\circ}_k = \exp(-\mu^{\circ} - 1), \qquad k = 1, 2, \dots, K$$

$$\downarrow$$

$$\frac{\partial L(\cdot)}{\partial \mu} = 1 - \sum_k p^{\circ}_k = 0 \qquad \rightarrow \qquad 1 = \sum_k \exp(-\mu^{\circ} - 1) = K \cdot \exp(-\mu^{\circ} - 1)$$

Donc, on peut réécrire $\quad p^{\circ}_k = \exp(-\mu^{\circ} - 1) / K \cdot \exp(-\mu^{\circ} - 1)$
$\qquad\qquad\qquad\qquad\qquad = 1 / K.$

Une deuxième chose dont on est certain, et qui fera l'objet de notre deuxième contrainte, c'est qu'on doit avoir $\mathbf{y} = X\mathbf{p}$; il s'agit de la condition de cohérence des données (*data* ou *moment consistency*). Le problème de maximisation devient alors le suivant:

$$\max_{\mathbf{p}} H(\mathbf{p}) = -\sum_k p_k \ln p_k \quad \text{sous contrainte } \sum_k p_k x_k = y_t \text{ et } \sum_k p_k = 1$$

$$L_1(\mathbf{p}, \lambda, \mu) = -\sum_k p_k \ln p_k + \sum_k \lambda_t [\,y_t - \sum_k p_k x_k\,] + \mu\,(1 - \sum_k p_k)$$

<u>CPO</u>:

$$\frac{\partial L_1(\cdot)}{\partial p_k} = -\ln p^{\circ}_k - 1 - \sum_t \lambda^{\circ}_t x_k - \mu^{\circ} = 0, \qquad k = 1, 2, \dots, K \qquad (10)$$

$$\frac{\partial L_1(\cdot)}{\partial \lambda_t} = y_t - \sum_k p^{\circ}_k x_k = 0, \qquad t = 1, 2, \dots, T \qquad (11)$$

$$\frac{\partial L_1(\cdot)}{\partial \mu} = 1 - \sum_k p^{\circ}_k = 0 \qquad (12)$$

En prenant l'exponentiel de (10), on trouve:

$$p^{\circ}_k = \exp(-\sum_t \lambda^{\circ}_t x_k - \mu^{\circ} - 1), \qquad k = 1, 2, \dots, K \qquad (10')$$

$$y_t = \sum_k \exp(-\sum_t \lambda^{\circ}_t x_k - \mu^{\circ} - 1)\, x_k, \qquad t = 1, 2, \dots, T \qquad (11')$$

$$1 = \sum_k \exp(-\sum_t \lambda^{\circ}_t x_k - \mu^{\circ} - 1) \qquad (12')$$

En prenant (10') / (12'), on obtient finalement que:

$$p^{\circ}_k = \frac{\exp(-\sum_t \lambda^{\circ}_t x_k - \mu^{\circ} - 1)}{\sum_k \exp(-\sum_t \lambda^{\circ}_t x_k - \mu^{\circ} - 1)} = \frac{\exp(-\sum_t \lambda^{\circ}_t x_k) \cdot \exp(-\mu^{\circ}) \cdot \exp(-1)}{\sum_k \exp(-\sum_t \lambda^{\circ}_t x_k) \cdot \exp(-\mu^{\circ}) \cdot \exp(-1)} = \frac{\exp(-\sum_t \lambda^{\circ}_t x_k)}{\Omega(\lambda^{\circ}_1, \lambda^{\circ}_2, \dots, \lambda^{\circ}_T)}.$$

où $\Omega(\lambda^{\circ}) = \sum_k \exp(-\sum_t \lambda^{\circ}_t x_k)$ est un facteur de normalisation (*partition function*). De plus, on montre que les éléments respectivement sur la diagonale principale et hors diagonale de la Hessienne sont:

$$\frac{\partial L_1(\cdot)^2}{\partial p_k^2} = -1 / p_k \quad \text{et} \quad \frac{\partial L_1(\cdot)^2}{\partial p_k \partial p_j} = 0.$$

de sorte qu'elle est définie négative pour $p_k > 0$ et satisfait donc à la condition suffisante d'un maximum globale unique.

Comme on peut le constater plus haut, la solution p°_k est fonction des $\{\lambda_t\}$. Cependant, l'autre source d'information qui nous reste issue du problème d'optimisation, (11), n'est pas fonction des λ. Donc, il n'y a pas de solution explicite (*closed-form*) à ce problème. Cependant, il est possible de construire un dual non contraint à ce problème d'optimisation et obtenir une solution explicite pour p°_k. Supposons $\mathbf{p}(\lambda)$ le vecteur de probabilité optimal p°_k (donc $\sum_k p^{\circ}_k = 1$); substituons-le dans L_1 ($\mathbf{p}$, λ, μ) pour obtenir la *minimal value function*:

$$\begin{aligned} L_1 (\mathbf{p}^{\circ}, \lambda, \mu^{\circ}) &= -\,\mathbf{p}(\lambda)' \ln \mathbf{p}(\lambda) + \lambda' [\mathbf{y} - X\,\mathbf{p}(\lambda)] \\ &= -\,\mathbf{p}(\lambda)' [-X\lambda' - \ln(\Omega(\lambda))] + [\mathbf{y}' - \mathbf{p}(\lambda)' X']\,\lambda \\ &= \mathbf{y}'\,\lambda + \mathbf{p}(\lambda)' \ln(\Omega(\lambda)) - \mathbf{p}(\lambda)' X'\lambda + \mathbf{p}(\lambda)' X'\,\lambda \\ &= \mathbf{y}'\,\lambda + \ln(\Omega(\lambda)) \equiv M_1(\lambda) \end{aligned}$$

Ainsi, minimiser $M_1(\lambda)$ par rapport à λ donne λ°, qui nous permet par la suite de trouver une solution explicite pour $\mathbf{p}^{\circ}$; en fait, $\nabla_\lambda M_1(\lambda) = \mathbf{y} - X\mathbf{p}$, c'est-à-dire que le gradient de $M_1(\lambda)$ correspond à la contrainte sur la cohérence des données. Maintenant, il est important de souligner que $M_1(\lambda)$ peut s'interpréter comme le logarithme de la fonction de vraisemblance, composée des données et de la distribution exponentielle pour $\mathbf{p}^{\dagger}$. Donc, la distribution $\mathbf{p}^{\circ}$ qui maximise l'entropie est la distribution qui maximise la vraisemblance d'obtenir $\mathbf{y}$.

- Principe d'entropie minimum relative (PEMR)

Il arrive parfois que le chercheur possède de l'information concernant le vecteur d'inconnu $\mathbf{p}$ et qui prenne la forme d'une distribution à priori $\mathbf{q}$. Dans ce contexte, l'objectif peut être formulé de façon à minimiser la "distance" entre $\mathbf{p}$ et $\mathbf{q}$. Donc, plutôt que de maximiser l'entropie (8), ici on minimisera l'entropie relative (9); on parle alors du *principe d'entropie minimum relative*. L'idée sera de trouver parmi l'ensemble des distributions de probabilités satisfaisant aux contraintes celle qui se rapproche le plus de $\mathbf{q}$. Le Lagrangien prend la forme suivante:

$$L_2 (\mathbf{p}, \mathbf{q}, \lambda, \mu) = \textstyle\sum_k p_k \ln(p_k/q_k) + \sum_t \lambda_t\, [\, y_t - \sum_k p_k\, x_k \,] + \mu\, (1 - \sum_k p_k)$$

Sans refaire les dérivations, on trouve procédant de la même méthode que le PEM:

$$p^{*}_k = \frac{q_k \cdot \exp(-\sum_t \lambda^{*}_t\, x_k)}{\Omega(\lambda^{*}_1, \lambda^{*}_2, \ldots, \lambda^{*}_T)}, \qquad k = 1, 2, \ldots, K$$

[†] La solution $p^{\circ}_k = \exp(-\sum_t \lambda^{\circ}_t\, x_k) \cdot \Omega(\lambda^{\circ})^{-1}$ de L_1 (p, λ, μ) est une version généralisée de la distribution de Maxwell, qui elle même fait parti de la famille des distributions exponentielles; Poirier (1995).

où $\Omega(\lambda^*) = \sum_k q_k \cdot \exp(-\sum_t \lambda^*_t x_k)$ est le facteur de normalisation[1]. À l'instar de la solution issu du principe d'entropie maximum, nous faisons encore face à un vecteur de probabilité $\mathbf{p}^*$ qui n'est pas une solution explicite; on peut cependant reformuler un dual non contraint et trouver un vecteur $\mathbf{p}^*$ indépendant des $\{\lambda^*_t\}$.

(B) PEM et PEMR généralisée

Maintenant, on se place dans le contexte plus familier du modèle linéaire général $\mathbf{y} = \beta\mathbf{X} + \mathbf{e}$ et on présente le *principe d'entropie* (maximum ou minimum relative) *généralisée*. Supposons donc qu'on souhaite obtenir une estimation pour le vecteur de coefficients β via le PEM. La difficulté qui se pose dans ce contexte est que comparativement à ce qu'on vient de voir, le vecteur de coefficients β ne prend pas nécessairement la forme de probabilités.

Afin de généraliser le principe d'entropie maximum dans ce contexte, on suppose que l'économètre est capable de borner les valeurs prises par les paramètres à estimer, de même pour les valeurs prises par les termes d'erreurs, afin de construire un support fini et discret pour β et $\mathbf{e}$. Par exemple, on peut avoir une idée sur le signe et/ou sur l'intervalle à l'intérieur duquel devrait s'arrêter l'estimation d'un paramètre; il sera alors possible de construire une nouvelle variable aléatoire avec poids et support qui reflètent cette information non échantillonalle sur β et $\mathbf{e}$. Le problème d'estimation se résumera donc à trouver la distribution de probabilité (les poids) pour β et $\mathbf{e}$ qui réconcilie notre information non-échantillonalle et l'information échantillonalle.

Concrètement, on veut représenter chaque β_k et e_t comme une variable aléatoire. Pour ce faire, on spécifie un support avec $2 \leq M$ valeurs possibles z_{km} et une distribution de poids $p_k \in [0,1]$ sur ces z_{km}. Par exemple, si on prend $M = 2$ tel quel (z_{k1}, z_{k2}) soient les bornes supérieures et inférieures de β_k, alors on représente ce dernier par la combinaison convexe $\beta_k = p_k \cdot z_{k1} + (1 - p_k) \cdot z_{k2}$. De la même façon, pour chaque e_t on spécifie un support avec $2 \leq L$ valeurs possibles v_{tl} et une distribution de poids $w_t \in [0,1]$ sur ces v_{tl}. Sous forme matricielle, la reparamétrisation des inconnus β et $\mathbf{e}$ donne $\mathbf{y} = \beta X + \mathbf{e} = \mathbf{XZp} + \mathbf{Vw}$, où les matrices $\mathbf{Z}$ et $\mathbf{V}$ sont respectivement d'ordre $(K \times KM)$ et $(T \times TL)$. Dans ce contexte, le problème d'optimisation se formule de la façon suivante:

$$\max_{\mathbf{p},\,\mathbf{w}} H(\mathbf{p}, \mathbf{w}) = -\mathbf{p}' \ln(\mathbf{p}) - \mathbf{w}' \ln(\mathbf{w}) \quad \text{s.c. } \mathbf{y} = \mathbf{XZp} + \mathbf{Vw},\ \mathbf{1}_K = (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} \text{ et } \mathbf{1}_T = (\mathbf{I}_T \otimes \mathbf{1}'_L)\mathbf{w}.$$

[1] On reconnaît une grande similitude entre la densité à posteriori qui découle de la méthode d'inférence bayesienne et le vecteur $\mathbf{p}^*$.

où $\otimes$ est le produit de Kronecker. Donc, l'idée est de sélectionner les distributions de probabilités **p** et **w** sur les supports **Z** et **V** qui sont les plus uniformes possibles sachant les observations à notre disposition. Les vecteurs de probabilités optimaux $\mathbf{p}^*$ et $\mathbf{w}^*$ qu'on dérive de ce problème d'optimisation nous permet d'obtenir un vecteur de coefficients estimés $\beta^* = \mathbf{Z}\mathbf{p}^*$ et un vecteur d'erreurs estimés $\mathbf{e}^* = \mathbf{V}\mathbf{w}^*$.

Plutôt que de présenter la dérivation de la solution du PEM généralisée, tournons-nous vers le PEMR généralisée qui, comme on le sait, l'inclus comme cas particulier. Donc, si le chercheur a à sa disposition de l'information à priori sur les valeurs que devraient prendre soit β_k ou e_t, on peut avoir recours au PEMR généralisée avec distributions à priori **q** et **u**. L'idée sera alors de minimiser $H(\mathbf{p}, \mathbf{w}; \mathbf{q}, \mathbf{u}) = \mathbf{p}' \ln(\mathbf{p}/\mathbf{q}) + \mathbf{w}' \ln(\mathbf{w}/\mathbf{u})$ sous les mêmes contraintes que dans le PEM généralisée. En passant directement au dual non contraint, on trouve:

$$\begin{aligned} L_3(\mathbf{p}^*, \mathbf{w}^*, \lambda, \mu^*, \tau^*) &= \mathbf{p}(\lambda)' \ln \mathbf{p}(\lambda) + \mathbf{w}(\lambda)' \ln \mathbf{w}(\lambda) + \lambda' [\mathbf{y} - \mathbf{XZp}(\lambda) - \mathbf{Vw}(\lambda)] \\ &= \mathbf{p}(\lambda)' [\mathbf{X'Z'}\lambda - \ln(\Omega(\lambda))] + \mathbf{w}(\lambda)' [\mathbf{V'}\lambda - \ln(\Psi(\lambda))] + [\mathbf{y}' - \mathbf{p}(\lambda)' \mathbf{Z'X'} - \mathbf{w}(\lambda)'\mathbf{V'}] \lambda \\ &= \mathbf{y}' \lambda - \mathbf{p}(\lambda)' \ln(\Omega(\lambda)) - \mathbf{w}(\lambda)' \ln(\Psi(\lambda)) \equiv M_3(\lambda) \end{aligned}$$

À l'instar du PEM, $M_3(\lambda)$ s'interprète comme une fonction de vraisemblance généralisée, composée des données et de deux distributions exponentielles pour **p** et **w**. Prendre $\partial M_3(\lambda)/\partial \lambda_r = \mathbf{y} - \mathbf{XZp}(\lambda) - \mathbf{Vw}(\lambda)$ va nous donner λ^* et une solution explicite pour $\mathbf{p}^*$ et $\mathbf{w}^*$.

Maintenant, on trouve la variance des coefficients estimés en passant par la dérivée seconde de $M_3(\lambda)$: $\partial^2 M_3(\lambda)/\partial\lambda_r \partial\lambda_s = -\sum_k X_{tk} \sum_m Z_{km} (\partial p_{km}/\partial \lambda_r) - \sum_l V_{tl} (\partial w_{tl}/\partial \lambda_r)$. Après quelques manipulations[1], on montre que $\mathrm{var}(\beta^*_k) = \sum_m \mathbf{Z}^2_{km} \mathrm{var}(p^*_k)$, où $\mathrm{var}(p^*_k) = \sum_m p^*_{km} \mathbf{Z}^2_{km} - [\sum_m p^*_{km} \mathbf{Z}_{km}]^2$.

<u>Remarques</u>:

- À première vue, le choix de M et L semble plutôt arbitraire. En fait, le choix du support pour chaque β_k est moins difficile envisager, dans la mesure où on a une idée sur la valeur que devrait prendre l'estimation de β_k. Si on nage dans l'inconnu, il est suggéré de spécifier un support centré à zéro avec deux larges bornes. Or, il semble que β^* soit assez robuste au choix du support. Cependant, plus on ajoute des points de supports à **Z**, plus $\mathrm{var}(\beta^*_k)$ diminue. La solution sera d'introduire, un peu dans le même ordre d'idée que dans le cadre d'estimation bayesienne, une fonction de perte et choisir le nombre de points de support en conséquence.

---

[1] Voir Golan, Judge et Miller (1996).

Par contre, intuitivement, le choix du support pour chaque e_t est moins évident. Or, si on suppose que **e** suit une distribution symétrique et centrée sur **0**, on peut spécifier un support symétrique $v_{t1} = -v_{tJ}$, pour chaque t. L'inégalité de Chebychev peut s'avérer une autre méthode afin de spécifier un ensemble de bornes pour le vecteur d'erreur **e**. De plus, on peut montrer qu'accroître inutilement le support de **e** diminue le contenu informatif de la contrainte sur la cohérence des données (mesuré par λ^*). En effet, $\partial\lambda^*/\partial v < 0$, $\forall|\lambda^*| > 0$ de sorte qu'on a plus de chance de trouver une solution intérieure au problème d'optimisation (le primal); cependant, à la limite, si $\lambda^* = \mathbf{0}$, alors $\mathbf{p} = \mathbf{q}$.

- Mentionnons tout d'abord que sous certaines hypothèses, $\text{plim}(\beta^*) = \beta$ et que β^* est asymptotiquement normal. Cependant, c'est pour ses propriétés en petit échantillon qu'il se démarque. En effet, on peut montrer que β^* est plus précis en petit échantillon (sur la base de l'erreur moyenne au carré) par rapport aux méthodes d'estimation plus traditionnelles. De plus, parmi les avantages de cette méthode, on note qu'il est possible d'obtenir un vecteur de coefficients estimés même si la matrice X a davantage de colonnes que de lignes, elle est beaucoup plus efficace que les approches traditionnelles d'estimations en présence de multicolinéarité et évidemment, il n'y a aucune hypothèse à faire concernant la loi de distribution du terme d'erreur (voir Golan, Judge et Miller (1996)).

IV-4) Nouvelle méthode d'estimation

C'est dans cette section-ci qu'on présente une nouvelle façon d'estimer le modèle de hasard proportionnel en employant le principe d'entropie (maximum) généralisée. Pour y parvenir, on doit d'une part faire appel à l'article de Han et Hausman (1990), qui nous permet de représenter un modèle de hasard proportionnel, avec hasard de base semi-paramétrique, sous forme d'un modèle à variable qualitative de type multinomial ordonné. Par la suite, on fais appel à l'article de Golan, Judge et Perloff (1997) qui emploient le principe de maximisation de l'entropie généralisée justement pour estimer un modèle à variable qualitative de type multinomial ordonné. Formellement, ces derniers procèdent comme suit:

Soit $\mathbf{y}^{\circ} = \mathbf{X}\beta + \mathbf{e} = \mathbf{XZp} + \mathbf{Vw}$, où y° est une variable latente. Dans le modèle multinomial ordonné classique, les observations sur y sont définies par:

$y_t = 0$ si $y^{\circ}_t \leq \delta_1$,
$y_t = 1$ si $\delta_1 < y^{\circ}_t \leq \delta_2$,
...
$y_t = J$ si $y^{\circ}_t > \delta_{J-1}$.

Supposons pour l'instant les δ_j comme étant connus. Alors, le PEM généralisée dans le cadre du modèle multinomial ordonné, appelons-le PEMG-MO, s'écrit comme suit:

$$\max_{\mathbf{p},\,\mathbf{w}} H(\mathbf{p},\mathbf{w}) = -\,\mathbf{p}'\ln(\mathbf{p}) - \mathbf{w}'\ln(\mathbf{w}) \quad \text{s.c} \quad \begin{array}{ll} \mathbf{X}_1'\mathbf{Z}\mathbf{p} + \mathbf{V}_1\mathbf{w}_1 \le \delta_1 & \text{pour } y_t = 0, \\ \delta_{j-1} < \mathbf{X}_j'\mathbf{Z}\mathbf{p} + \mathbf{V}_j\mathbf{w}_j \le \delta_{j-1} & \text{pour } y_t = 1, 2, \ldots, J-1, \\ \delta_{J-1} < \mathbf{X}_J'\mathbf{Z}\mathbf{p} + \mathbf{V}_J\mathbf{w}_J & \text{pour } y_t = J \\ \mathbf{1}'\mathbf{p} = 1 & \\ \mathbf{1}'\mathbf{w}_j = 1 & \end{array}$$

En formulant le Lagrangien et en solutionnant pour $\mathbf{p}$ et $\mathbf{w}$ comme on a fait plus haut, on trouve:

$$\hat{p}_{km} = \frac{\exp(-\mathbf{ZX}'_{1k}\hat{\lambda}_1 - \ldots - \mathbf{ZX}'_{Jk}\hat{\lambda}_J)}{\sum_h \exp(-\mathbf{ZX}'_{1k}\hat{\lambda}_1 - \ldots - \mathbf{ZX}'_{Jk}\hat{\lambda}_J)} \qquad \text{et} \qquad \hat{w}_{tl} = \frac{\exp(-\mathbf{V}_1\hat{\lambda}_1 - \ldots - \mathbf{V}_J\hat{\lambda}_J)}{\sum_l \exp(-\mathbf{V}_1\hat{\lambda}_1 - \ldots - \mathbf{V}_J\hat{\lambda}_J)}$$

où $\hat{\lambda}$ est la matrice des multiplicateurs de Lagrange; on trouve $\hat{\beta} = \mathbf{Z}\hat{\mathbf{p}}$ et $\hat{\mathbf{e}} = \mathbf{V}\hat{\mathbf{w}}$. Afin d'obtenir une solution explicite pour $\hat{\mathbf{p}}$ et $\hat{\mathbf{w}}$ qui ne soit pas fonction de $\hat{\lambda}$, on formule un dual non contraint comme à la section précédente; on parvient par la suite à dériver la matrice de variance/covariance pour $\hat{\beta}$.

Afin d'appliquer ce dernier résultat aux données de durée, on aurait qu'à prendre y comme les données observées sur la durée des épisodes de chômage et supposer, comme chez Han et Hausman (1990), que les δ_j soient inconnus. Pour ce faire, on peut spécifier un support $\mathbf{d}_j$ pour les δ_j, où $\mathbf{d}_j'\,\mathbf{r}_j = \delta_j$, avec $\mathbf{r}_j$ comme vecteur de probabilités qu'on dérive en même temps que $\mathbf{p}$ et $\mathbf{w}$. Ainsi, on substitut les δ_j par $\mathbf{d}_j'\,\mathbf{r}_j$ dans la formulation du problème d'optimisation PEMG-MO et en ajoutant la contrainte $\mathbf{1}'\mathbf{r}_j = 1$; on trouve alors $\hat{\mathbf{p}}$, $\hat{\mathbf{w}}$ et $\hat{\delta}$.

Donc, pour résumer la méthode d'estimation ici proposée, *c'est en traitant les données de durée comme ordonnée (par le résultat de Han et Hausman) qu'on peut estimer le modèle de hasard proportionnel via le principe d'entropie maximum généralisée (par Golan, Judge et Perloff)*. L'intérêt de cette méthode d'estimation est qu'elle devrait permettre de contrôler assez facilement pour l'hétérogénéité non observée, tout en ayant l'avantage d'être efficace au sens de Heckman et Singer (1984). Nous avons vu, au point IV-1.b), que la méthode de ces derniers consiste à approximer $G(v)$ semi-paramétriquement en estimant les probabilités associées à un nombre fini de points de support de la distribution de v dans la fonction de vraisemblance. Or, cette idée se prête parfaitement à l'estimation du modèle qui est suggéré dans cette section puisque le principe d'entropie généralisée consiste précisément, comme on a vu à la section précédente, à spécifier des points de supports et à trouver les probabilités qui y sont rattachées en maximisant l'entropie. Il suffirait donc d'ajouter deux autres contraintes au problème de maximisation PEMG-MO où y correspond aux données de durée, une pour spécifier les points de support pour l'hétérogénéité non observée et l'autre pour s'assurer que la somme des probabilités sur ce support égale à l'unité.

V- Conclusion

Comme on a pu le constater, le programme d'a-c est sujet à créer plusieurs distorsions sur le marché du travail, la plus importante étant sans doute celle associée au problème de hasard moral, bien que des gains de productivité puissent plus que contrebalancer les coûts du hasard moral, selon Acemoglu et Shimer (1999b).

Néanmoins, un fait demeure; les allocations d'a-c semblent retarder la réinsertion de chômeurs sur le marché du travail, ce sur la base de trois études empiriques qu'on a étudiées. Cependant, l'arbitraire par lequel les auteurs (aussi vrai pour d'autres études) choisissent de la méthode à adopter afin de contrôler pour l'hétérogénéité non observée dans l'estimation de leur modèle de hasard proportionnel peut nous laisser perplexe.

Or, on a vu qu'en passant successivement par les résultats de Han et Hausman (1990) et Golan, Judge et Perloff (1997), il est possible d'estimer le modèle de hasard proportionnel via le principe d'entropie maximum généralisée en traitant les données de durée comme ordonnées. Cette nouvelle façon d'estimer le modèle de hasard proportionnel offre potentiellement l'avantage d'être une façon de contrôler pour l'hétérogénéité non observée semblable à celle de Heckman et Singer (1984), mais en étant beaucoup plus simple à estimer.

On dit "potentiellement" puisque cette méthode d'estimation demeure exploratoire et n'est pas encore programmée. Dériver les propriétés sera un autre pas à franchir afin de juger efficacement du potentiel de cette nouvelle méthode d'estimation du modèle de hasard proportionnel. Tout ceci devrait être fait dans un avenir rapproché et l'emploi de données relativement à la durée des épisodes de chômage devrait servir comme base de données.

Référeces

Acemoglu D. et R. Shimer (1999a), "Efficient Unemployment Insurance", *Journal of Political Economy*, vol.107, 893-928.

Acemoglu D. et R. Shimer (1999b), "Productivity Gains From Unemployment Insurance", NBER WP-7352.

Atkinson A.B. et J. Micklewright (1991), "Unemployment Compensation in Labor Market Transitions: A Critical Review", *Journal of Economic Litterature*, vol. 29, 1679-1727.

Cahuc P. et A. Zilberberg (1996), *Économie du Travail: La formation des salaires et les déterminants du chômage*, De Boeck Université.

Chiu W.H. et E. Karni (1998), "Endogenous Adverse Selection and Unemployment Insurance", *Journal of Political Economy*, vol.106, 806-827.

Clark K.B. et L.H. Summers (1982), "Unemployment Insurance and Labor Market Transitions", dans: *Workers, Jobs, and Inflation*, édité par Matin N. Baily, Brookings Institution, 279-323.

Corak M. (1994), "Unemployment Insurance, Work Incentives, and the Canadian Labor Market: An Overview ", dans: *Unemployment Insurnance: How to make it work*, C.D. Howe Institute, 86-159.

Cullen J.B. et J. Gruber (2000), "Does Unemployment Insurance Crowd out Spousal Labor Supply?", *Joiurnal of Labor Economics*, vol. 18, p.546.

Dellas H. (1997), "Unemployment Insurance Benefits and Human Capital Accumulation", *European Economic Review*, vol. 41, 517-524.

Devine T.J. et N. Kiefer (1991), *Empirical Labor Economics: The Search Approach*, Oxford University Press.

Dingledine G. (1980), *Exposé chronologique: L'évolution de l'assurance chômage de 1940 à 1980*, préparé pour Emploi et Immigration Canada.

Feldstein M. (1976), "Temporary Layoffs in the Theory of Unemployment", *Journal of Political Economy*, vol.84, 937-957.

Golan A., G. Judge et D. Miller (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley.

Golan A., G. Judge et J. Perloff (1997), "Estimation and Inference with Censored and Ordered Multinomial Response Data", *Journal of Econometrics*, vol. 79, 23-51.

Gruber J. (1997), "The Consumption Smoothing Benefits of Unemployment Insurance", *American Economic Review*, vol.87, 192-205.

Han A. et J.A. Hausman (1990), "Flexible Parametric Estimation of Duration and Competing Risk Models", *Journal of Applied Econometrics*, vol.5, 1-28.

Heckman J.J. et B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, vol.52, 271-320.

Hopenhayn H.A. et J.P. Nicolini (1997), "Optimal Unemployment Insuranc ", *Journal of Political Economy*, vol.105, 412-438.

Ihara S. (1993), *Information Theory for Continuous Systems*, World Scientific.

Jones S. R.G. (1993), "Cyclical and Seasonal Properties of Canadian Gross Flows of Labor", *Canadian Public Policy*, vol.19, 1-17.

Kalbfleisch J. et R. Prentice (1980), *The Statistical Analysis of Failure Time Data*, John Wiley.

Kesselman J.R. (1986), "The Royal Commission's Proposals for Income Security reform", *Canadian Public Policy*, vol.12, 101-112.

Kiefer N.M. (1988), "Economic Duration Data and Hazard Functions", *Journal of Economic Literature*, vol.26, 646-679.

Klein E. (1973), *Mathematical Methods in Theoretical Economics: Topological and Vector Space Foundations of Equilibrium Analysis*, Academic Press.

Lancaster T. (1979), "Econometric Methods for the Duration of Unemployment", *Econometrica*, vol.47, 939-956.

Lazar F. (1994), "UI as a Redistributive Scheme and Automatic Fiscal Stabilizer", dans: *Unemployment Insurnance: How to make it work*, C.D. Howe Institute, 36-85.

Lemieux T. et W.B. MacLeod (1998), " Supply Side Hysteresis: The Case of the Canadian Unemployment Insurance System ", NBER WP-6732.

Lin Z. (1998), "Employment Insurance in Canada: Recent Trends and Policy Changes", Document de travail #125, Statistique Canada, Business and Labour Market Analysis.

Lucas R.E. (1978), "Unemployment Policy", *American Economic Review*, vol.68, 353-357

Mas-Colell A., M.D. Whinston et J.R. Green (1995), *Microeconomic Theory*, Oxford University Press.

Meyer B.D. (1990), "Unemployment Insurance and Unemployment Spells", *Econometrica*, vol.58, 757-782.

Moffitt R. (1985), "Unemployment Insurance and the Distribution of Unemployment Spells", *Journal of Econometrics*, vol.28, 85-101.

Mortensen D.T. (1986), "Job Search and Labor Market Analysis", dans : Handbook of Labor Economics, vol.2, édité par O.C. Ashenfelter et R. Layard, North Holland, 849-919.

Neumann G. (1997), "Search Models and Duration Data", dans : Handbook of Applied Econometrics; vol.2, édité par M.H. Peseran et P. Schmidt, Blackwell, chapitre 4.

Poirier D.J. (1995), *Intermediate Statistics and Econometrics*, MIT Press.

Ryu H.K. et D.J. Slottje (2000), "Estimating the Density of Unemployment Duration based on Contaminated Samples or Small Samples", *Journal of Econometrics*, vol.95, 131-156.

Shannon C.E. (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol.27, 379-423.

Stigler G.J. (1962), "Information in the Labor Market", *Journal of Political Economy*, vol.70, 94-105.

Torelli N. et U. Trivellato (1993), "Modelling Inacuracies in Job-Search Duration Data", *Journal of Econometrics*, vol.59, 187-211.

Annexes

A.1) Espace probabilisé (Ω, $\Im$, P) -- voir Poirier (1995).

Soit une expérience aléatoire X:

Ω : Univers associé à l'expérience (ensemble des résultats possibles de l'expérience)

ω : $\omega \in \Omega$ est un événement élémentaire (un et un seul événement élémentaire se réalise suite à une expérimentation)

A : $A \subset \Omega$ est un événement (A est considéré réalisé si l'événement élémentaire réalisé appartient à A)

Parmi ceux-ci, on note:
- Ω : L'événement certain
- $\varnothing$: L'événement incertain
- A^c : L'événement complémentaire ($A^c \equiv \{\omega \in \Omega; \omega \notin A\}$)

$\wp(\Omega)$: L'ensemble de tous les sous-ensembles de Ω ($\wp(\Omega)$: *power set*)

$\Im$: Un sous-ensemble $\Im$ de $\wp(\Omega)$ est une tribu (σ-algèbre); il doit posséder les propriétés suivantes:

1) $\Omega \in \Im$
2) si $A \in \Im$, alors $A^c \in \Im$
3) si $A_1, A_2, \ldots \in \Im$, alors $A_1 \cup A_2 \cup \ldots \in \Im$
4) si $A_1, A_2, \ldots \in \Im$, alors $A_1 \cap A_2 \cap \ldots \in \Im$

Donc, pour que la structure d'information soit cohérente, une tribu doit être un ensemble non vide et stable pour les trois opérations ensemblistes suivantes: complémentarité, union et intersection.

(Ω, $\Im$) : Espace mesurable (ou espace probabilisable)

μ : Une mesure sur un espace mesurable (Ω, $\Im$) est une fonction $\mu : \Im \to \Re$, où $\Re$ est l'ensemble des réels; elle doit satisfaire aux conditions suivantes:

1) $\mu(\varnothing) = 0$
2) $0 \leq \mu(A) \leq \infty$, $\forall A \in \Im$
3) Si $A_1, A_2, \ldots \in \Im$ sont des événements deux à deux disjoints, alors : $\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n)$

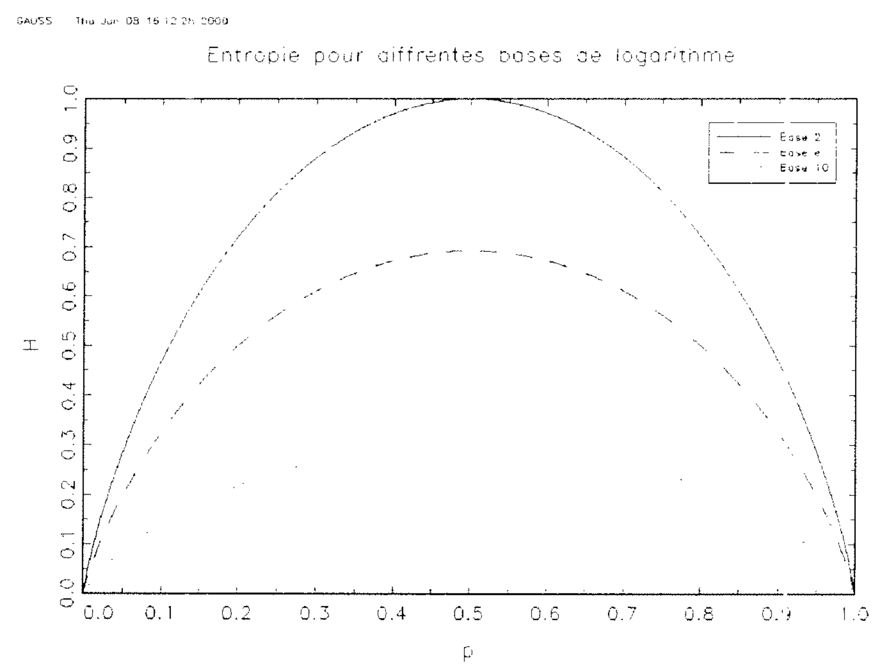
P : Une probabilité sur un espace mesurable (Ω, $\Im$) est une fonction μ: $\Im \to [0,1]$ et doit posséder les propriétés suivantes:

1) $P(\varnothing) = 0$ et $P(\Omega) = 1$
2) $P(A^c) = 1 - P(A^c)$
3) Soit A et $B \subset \Im$; si $A \subset B$, alors $P(A) \leq P(B)$
4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(Ω, $\Im$, P) : Espace probabilisé

A.2) Représentation graphique de l'entropie

Pour $N = 2$, l'expression (8) $H(p_1, p_2) = p_1 \log p_1 + (1 - p_1) \log (1 - p_1)$ a l'allure suivante:



Donc, indépendamment du choix de la base du logarithme, l'entropie est toujours à son maximum lorsque $p_1 = 0.5$.

Remarque:

Le choix de la base du logarithme n'affecte que l'unité de mesure de l'entropie. En théorie de l'information, alors que l'entropie mesure également la quantité d'information véhiculée à travers un message, on calcule généralement l'entropie en base 2; l'unité de mesure est alors les *binary digits*, ou plus simplement des *bits*. En économétrie, on emploie le logarithme népérien.

A.3) L'entropie relative versus métrique

Il est bon de préciser que l'entropie relative n'est pas un métrique en soit. En effet, soient μ et ν deux mesures de probabilités sur l'espace probabilisable (X, F). Alors, on a que: *i)* $H(\mu;\nu) \geq 0$ et *ii)* $H(\mu;\nu) = 0$ ssi $\mu = \nu$.

De l'autre côté, un métrique est défini comme suit (Klein (1973)): Soit X un ensemble non vide de points et R l'ensemble des réels. Une fonction en valeur réelle $d : X \times X \rightarrow R$ est appelée métrique ssi, $\forall\, x, y$ et $z \in X$, *i)* $d(x, y) \geq 0$, *ii)* $d(x, y) = 0$ ssi $x = y$, *iii)* $d(x, y) = d(y, x)$ et *iv)* $d(x, z) \leq d(x, y) + d(y, z)$; le nombre réel $d(x, y)$ est appelé distance entre les points x et y.

Donc, c'est en raison que l'entropie relative n'est pas symétrique (*iii*) et qu'elle ne respecte pas l'inégalité triangulaire (*iv*) qu'elle n'est pas un métrique.